

IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN SISWA BERMASALAH

(Studi Kasus : SMKS Taruna Bahari)

Oleh
AYIDIL KAI
T3121149

SKRIPSI

Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar sarjana



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN SKRIPSI

IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN SISWA BERMASALAH

(Studi Kasus : SMKS Taruna Bahari)

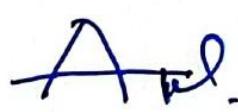
Oleh
AYIDIL KAI
T3121149

USULAN PENELITIAN

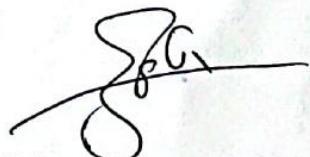
Untuk memenuhi salah satu syarat ujian
Guna memperoleh Gelar Sarjana
Program Studi Teknik Informatika
Ini telah disetujui oleh pembimbing

Gorontalo, April 2023

Pembimbing I


Amiruddin, S.Kom, M.Kom
NIDN 0910097601

Pembimbing II


Suhardi Rustam, S.Kom, M.Kom
NIDN 0915088403

PENGESAHAN SKRIPSI

IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN SISWA BERMASALAH

Oleh

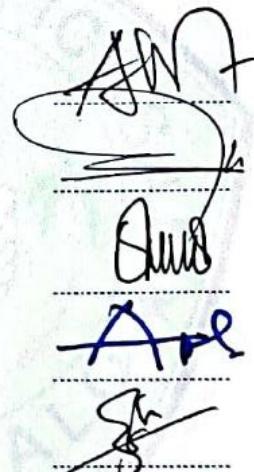
AYIDIL KAI

T3121149

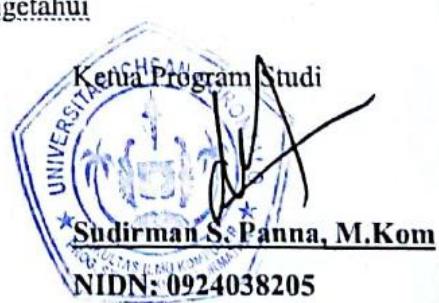
Diperiksa oleh panitia ujian Starata Satu (S1)

Universitas Ichsan Gorontalo

1. Ketua penguji
Yasin Aril Mustofa, M.Kom
2. Anggota
Sunarto Taliki, M.Kom
3. Anggota
Sumarni, M.Kom
4. Anggota
Amiruddin, M.Kom
5. Anggota
Suhardi Rustam, M.Kom



Mengetahui



PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, Mei 2023

Yang Membuat Pernyataan,



Ayidi Kai

ABSTRACT

AYIDIL KAI T3121149. IMPLEMENTATION OF K-MEANS ALGORITHM FOR GROUPING PROBLEMATIC STUDENTS

The purpose of this survey 1.) Knowing the optimal number of clusters to determine the number of student groups with problems, as a priority group that gets special treatment. 2.) Knowing the analysis of each clustering of students with severe problems in order to get special treatment. SMKS TARUNA BAHARI is one of the schools that became the place of research and to obtain information carried out by direct perception by directing direct perception, interviews with guidance counselors and documentation notes related to the object of research, especially on Student Problems. In this study obtained 179 student data information from SMKS TARUNA BAHARI to be used as test data in determining the grade level of problem students. On the basis of not sebandingya the number of problematic students who must be handled by the number of counseling guidance teachers (BK) in SMKS Taruna Bahari, So students who have problems need grouping to determine which groups are prioritized in handling problems that are right on target so that the problem does not experience a long time in handling and the same problem repeats again using the K-Means method. The results of clustering using the Elbow technique are 4 clusters using 2 The variables are the Absence Presentation variable and the Total Weight of Violations. So that the results of the analysis of each cluster are obtained, namely cluster 1 is a problem student who has a high score on the total weight of violations and low on the absence presentation with a total of 42 students, so that it becomes the 3rd priority, cluster 2 is a problem student who has a high score on the absence presentation and Low on the Total Weight of Violations with a Total of 34 Students, so it becomes the 2nd Priority, cluster 3 is a Troubled Student who has a high score on Absence Presentation and high on the Total Weight of Violations with a Total of 51 Students, so that it becomes the 1st Priority, cluster 4 is a Problem Students who has a Low Score on Presentation of Absence and Low on Total Weight of Violations with a Total of 52 Students, so that it becomes the 4th Priority.

Keywords: problematic students, Clustering, Elbow, K-Means



ABSTAK

AYIDIL KAI T3121149. IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN SISWA BERMASALAH

Tujuan dari penilitan ini 1.) Mengetahui Jumlah *Cluster* yang optimal untuk menentukan Jumlah kelompok Siswa yang bermasalah, sebagai kelompok prioritas yang mendapatkan penanganan Khusus. 2.) Mengetahui analisis setiap *Clustering* siswa bermasalah tingkat Berat guna untuk mendapatkan penanganan Khusus. SMKS TARUNA BAHARI merupakan salah satu sekolah yang menjadi tempat Penelitian dan untuk memperoleh informasi dilakukan dengan persepsi dengan mengarahkan persepsi langsung, wawancara dengan arahan pembimbing koseling dan dokumentasi mencatat yang berhubungan dengan objek penelitian, khususnya tentang masalah siswa. Dalam Penelitian ini diperoleh 179 informasi data siswa dari SMKS TARUNA BAHARI untuk digunakan sebagai bahan data uji dalam penentuan tingkat kelas siswa bermasalah. Atas dasar tidak sebandingnya jumlah siswa bermasalah yang harus di tangani dengan jumlah Guru Bimbingan Konseling (BK) di SMKS Taruna Bahari, Maka siswa yang bermasalah membutuhkan pengelompokan untuk menetapkan kelompok yang di prioritaskan dalam penanganan masalah yang tepat sasaran agar masalah tidak mengalami lama dalam penanganan dan masalah yang sama berulang lagi dengan menggunakan metode K-Means, Adapun hasil clustering menggunakan teknik Elbow sebanyak 4 cluster dengan menggunakan 2 Variabel yaitu variable Presentasi Ketidakhadiran dan Jumlah Bobot Pelanggaran. Sehingga mendapatkan hasil analisis dari setiap cluster yaitu cluster 1 merupakan Siswa Bermasalah yang memiliki Nilai tinggi pada Jumlah Bobot Pelanggaran dan Rendah Pada Presentasi Ketidakhadiran dengan Jumlah 42 Siswa, sehingga menjadi Proritas Ke 3, cluster 2 merupakan Siswa Bermasalah yang memiliki Nilai tinggi pada Presentasi Ketidakhadiran dan Rendah pada Jumlah Bobot Pelanggaran dengan Jumlah 34 Siswa, sehingga menjadi Proritas Ke 2, cluster 3 merupakan Siswa Bermasalah yang memiliki Nilai tinggi pada Presentasi Ketidakhadiran dan tinggi pada Jumlah Bobot Pelanggaran dengan Jumlah 51 Siswa, sehingga menjadi Proritas Ke 1, cluster 4 merupakan Siswa Bermasalah yang memiliki Nilai Rendah pada Presentasi Ketidakhadiran dan Rendah pada Jumlah Bobot Pelanggaran dengan Jumlah 52 Siswa, sehingga menjadi Proritas Ke 4.

Kata Kunci: Siswa Bermasalah, *Clustering*, *Elbow*, *K-Means*



KATA PENGANTAR

Assalamu 'alaikum wa rahmatullahi wabarakatuh.

Dengan memanjatkan puji syukur kehadirat Allah SWT, serta rahmat shalawat dan salam untuk junjungan besar Nabi Muhammad SAW. Sehingga penulis dapat menyelesaikan Skripsi yang berjudul “ **IMPLEMENTASI ALGORITMA K-MEANS UNTUK PENGELOMPOKAN SISWA BERMASALAH**” (**Studi kasus SMKS Taruna Bahari**) tepat pada waktunya. sebagai salah satu syarat Ujian Akhir guna memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa penyusunan Skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan setinggi-tingginya kepada:

1. Ibu Dr. Hj. Juriko Abdussamad, Msi, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan Dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, S.Kom, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, S.Kom, M.Kom, selaku Wakil Dekan II Fakultas Ilmu Komputer Universitas Ichsan Gorontalo, Sekaligus Pembimbing Utama;
6. Bapak Sudirman S. Panna, S.Kom, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;

7. Bapak Amiruddin, S.Kom, M.Kom, selaku Dosen Pembimbing 1 Pendamping Yang Juga Telah Memberikan Motivasi dan Penyusunan Program Aplikasi Kepada Penulis;
8. Bapak Suhardi Rustam, S.Kom, M.Kom, selaku Dosen Pembimbing 2 Pendamping Yang Juga yang Membantu Penyusunan Skripsi Kepada Penulis;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
10. Kepada Orang Tua saya yang tercinta, atas segala kasih sayang, jerit payah dan doa restunya dalam mendidik dan membesarkan penulis;
11. Kepada Istri dan Anak Yang Selalu Memberikan Spirit Baru Setiap kali Beraktivitas.
12. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
13. Kepada semua pihak yang ikut membantu dalam penyelesaian Skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua Aamiin.

Gorontalo, Mei 2023

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
ABSTRAK.....	iv
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	4
1.3 Rumusan Masalah.....	4
1.4 Tujuan Masalah.....	4
1.5 Manfatat Penelitian	5
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Studi.....	6
2.2 Tinjauan Pustaka.....	7
2.2.1 Kedipsilan	7
2.2.2 Konsep Dasar Peraturan Tata Tertib Sekolah.....	8
2.2.3 Pengertian Tata Tertib Sekolah.....	9
2.2.4 Tujuan Peraturan Sekolah.....	9
2.2.5 Tabel Pelanggaran.....	9
2.2.6 Pengertian Siswa.....	11
2.2.7 Data Set.....	11
2.2.8 Data Maining	12
2.2.9 Tahapan Data Maining.....	12
2.2.10 <i>Clustering</i>	14
2.2.11 Algoritma <i>K-Means</i>	14
2.2.12 Metode <i>Elbow</i>	16

2.2.13 Penerapan Algoritma <i>K-Means</i>	17
2.3 Perangkat Lunak Pendukung	22
2.4 Kerangka Pikir	24
BAB III METODE PENELITIAN	25
3.1 Metode, Subjek dan Objek Penelitian	25
3.2 Pengumpulan data	25
3.3 Pemodelan	26
3.4 Tahap <i>Clustering</i>	27
BAB IV HASIL PENELITIAN	29
4.1 Hasil Pengumpulan data	29
4.2 Hasil Pemodelan	31
4.2.1 Pra Pegolahan Data.....	31
4.2.2 Normalisasi Data.....	31
4.2.3 Hasil Hitungan K-Means	32
4.2.4 Hasil Clustering	39
BAB V PEMBAHASAN	36
5.1 Implementasi K-Means dengan Python	36
5.1.1 Import Library Phyton	36
5.1.2 Mengaktifkan Google Drive	36
5.1.3 Membaca Data Set	37
5.1.4 Konversi Data Set Excel ke CSV	38
5.1.5 Menghitung Banyak Data	38
5.1.6 Menentukan Kolom Sebaai Dataset.....	49
5.2 Kinerja Model	40
5.3 Pembahasan Model.....	41
5.3.1 Pemodelan Metode K-Means.....	41
5.3.2 Visualisasi Hasil Pemodelan K-Means.....	42
5.3.3 Pusal Cluster dan Jumlah Data	43
5.3.4 Pelabelan Data Hasil Custer	46
BAB VI Kesimpulan Dan Saran	52
6.1 Kesimpulan	52

6.2 Saran	53
DAFTAR PUSTAKA	54

DAFTAR TABEL

Tabel 1.1	Data Siswa	3
Tabel 2.1	Penelitian Terkait.....	7
Tabel 2.2	Tabel Pelanggaran	9
Tabel 2.3	Tabel Data Set	12
Tabel 2.4	Sampel Data Set Nilai Mata Kuliah Kalkulus	17
Tabel 2.5	Data Set Nilai Mata Kuliah Matematika Diskrit	18
Tabel 2.6	Sampel Hasil Penggabungan Data Set.....	18
Tabel 2.7	Sampel Hasil Konvesi Nilai Huruf Ke Nilai Angka.....	19
Tabel 2.8	<i>Centroid</i> dan Jumlah Data	21
Tabel 2.9	Sampel Hasil <i>Clustering</i>	22
Tabel 2.10	Perangkat Lunak Pendukung	22
Tabel 3.1	Variabel Data.....	26
Tabel 4.1	Hasil Pengumpulan Data	29
Tabel 4.2	penentuan awal cluster literasi 1	32
Tabel 4.3	Menentukan Jarak Literasi 1.....	33
Tabel 4.4	Menentukan Awal Pusat Cluster Literasi 2	34
Tabel 4.5	Awal Pusat Cluster Literasi 2	35
Tabel 4.6	Hasil Pengumpulan Data	35
Tabel 4.7	Literasi 3 Awal Centroid	36
Tabel 4.8	Mencari Jarak Terdekat Literasi 3	37
Tabel 4.9	Awal Pusat Cluster Literasi 4	37
Tabel 4.10	Awal Pusat Cluster Literasi 4	37
Tabel 4.11	Menentukan jarak literasi 4	38
Tabel 4.12	Cluster 1, 2, 3, dan 4.....	38
Tabel 4.13	Hasil Data Clustering.....	39
Tabel 4.14	penentuan awal cluster literasi 1	39
Tabel 4.15	Menentukan Jarak Literasi 1.....	40
Tabel 4.16	Menentukan Awal Pusat Cluster Literasi 2	41

Tabel 4.17	Awal Pusat Cluster Literasi 2	41
Tabel 4.18	Menghitung Jarak Literasi 2	41
Tabel 5.1	Tabel Pusat Cluster dan Jumlah Data	51
Tabel 5.2	Hasil Clustering Data Siswa Bermasalah SMK Taruna Bahari.55	

DAFTAR GAMBAR

Gambar 2.1 Proses Knowledge Deiscoveryin Database (KDD)	13
Gambar 2.2 Proses Algoritma <i>K-Means</i>	16
Gambar 2.3 Contoh Visualisasi Penggunaan Metode <i>Elbow</i>	17
Gambar 2.4 Visualisasi Hasil Metode <i>Elbow</i>	20
Gambar 2.5 Visualisasi Hasil <i>Cluster</i>	21
Gambar 2.6 Informasi Pusat <i>Cluster</i>	21
Gambar 2.7 Kerangka Pikir	24
Gambar 3.1 Pemodelan Clustering	27
Gambar 5.1 Mengaktifkan Google Drive	43
Gambar 5.2 Membaca Data Set	43
Gambar 5.3 Conversi Data Set Excel ke CSV	44
Gambar 5.4 Hasil Menghitung Banyak Data	45
Gambar 5.5 Menentukan Kolom sebagai Data Set.....	46
Gambar 5.6 Visualisasi Teknik Elbow	47
Gambar 5.7 Visualisasi Hasil Clustering	49
Gambar 5.8 Penentuan Pusat Cluster.....	50
Gambar 5.9 Hasil Perhitungan Jumlah Data	50
Gambar 5.10 Pelabelan Data Hasil Cluster	52
Gambar 5.11 Pelabelan Data Hasil Cluster Penambahan Keterangan.....	53
Gambar 5.12 Pelabelan Data Hasil Cluster Penambahan Status Cluster.....	54

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tidak sedikit tingkat kedisiplinan siswa pada umumnya masih sangat memprihatinkan. Jumlah pelanggaran yang dilakukan oleh siswa meningkat. Dari berbagai macam pelanggaran tata tertib sekolah, misalnya banyaknya siswa yang membolos atau pergi pada jam-jam pelajaran, tawuran, terlambat datang ke sekolah, kurang fokus belajar, sering tidak masuk kelas, dan belum tuntas di beberapa mata pelajaran.[1] Secara umum banyaknya pelanggaran yang dilakukan oleh siswa akan mempengaruhi kemajuan dan prestasi belajar di sekolah.

Diantara faktor tidak disiplinnya siswa adalah minimnya control Guru terhadap kesepakatan yang diberikan kepada siswa, minimnya pengawasan dari Guru untuk siswa yang menyalahgunakan kesepakatan. Oleh karena itu, upaya dilakukan untuk menerapkan kedisiplinan ini dengan memberikan wewenang dan manajemen yang sangat jelas dari Guru terkait dengan penanganan terhadap siswa yang menyalahgunakan kesepakatan. siswa yang bermasalah dinasehati dan di beri peringatan oleh guru mapel, setelah 3x diberi peringatan masih menyalahgunakan kesepakatan, maka tanggung jawab di alihkan ke Guru Perwalian (dinasehati dan diberi kesempatan sampai 3x Peringatan), apabila masalah masih belum terselesaikan maka tanggung jawab dialihkan ke Wakil Kepala Sekolah di Bidang Kesiswaaan, dan apabila masalah masih belum terselesaikan maka tanggung jawab di alihkan ke Guru BK (Bimbingan Konseling). Apabila dalam pembinaan Guru BK, seorang siswa ditemukan memiliki masalah yang mendasar penyebab terjadinya penyalagunaan kesepakatan, maka siswa tersebut akan ditindaklanjuti melalui layanan pembinaan oleh Guru BK (Bimbingan Konseling). hanya saja keberadaan jumlah Guru BK tidak sebanding dengan jumlah Siswa, Jumlah Guru BK di sekolah hanya 1 Orang dan jumlah siswa yang harus mendapatkan pendampingan bimbingan konseling sejumlah 178 siswa, segingga metode Pengelompokan ini

sangat di butuhkan Guna untuk mengontrol siswa agar bisa di tentukan mana yang akan lebih dulu di tangani. sehingga sekolah khususnya pelayanan BK (Bimbingan Konseling), dapat lebih efektif membedakan kelompok siswa yang membutuhkan Penanganan khusus. Dari permasalahan tersebut Penulis memanfaatkan metode *K-means* untuk mengelompokan siswa yang bermasalah di SMKS Taruna Bahari Gorontalo Utara.

Tabel 1.1: Data Siswa

No	Nama Peserta Didik	Jumlah Mata pelajaran yang belum Tuntas	Presentase Ketidak Kehadiran	Jumlah Bobot Pelanggaran
1	Sandra Hasan	3	28	73 Point
2	SELISTI NUKU	0	11	21 Point
3	Sri Hiola Tonang	0	38	72 Point
5	AKSEL AGUSTIAN DUKI	2	80	38 Point
6	Alwin Ishak	1	43	83 Point
7	Anisa Mokodongan	3	53	60 Point
8	AULIA LIBUNEOLO	0	22	17 Point

Sumber Data : SMKS Taruna Bahari, 2022

Penggalian informasi dapat diterapkan untuk menyelidiki nilai tambahan dari suatu kumpulan informasi sebagai informasi yang belum diketahui secara fisik. Ada beberapa strategi yang digunakan untuk Penggalian informasi, salah satu prosedur Penggalian informasi adalah pengelompokan.

Pengelompokan informasi merupakan tahapan untuk mengkarakterisasi Kumpulan informasi yang kriteria kelasnya belum tergambar, ide pengelompokan adalah untuk meningkatkan dan membatasi kemiripan antar kelas. [2] Misalnya, ada sekelompok artikel, siklus primer dapat dikelompokkan ke dalam beberapa pengaturan kelas kemudian ke dalam himpunan normal

sehingga dapat ditentukan dengan baik berdasarkan kumpulan karakterisasi tertentu. Maka hasil dari analisis *Clustering* pada suatu data yaitu sejumlah *Cluster* (Kelompok).

K-means adalah salah satu strategi *Clustering* di Data Mining, sistem tampilan tanpa pengawasan dan teknik pengumpulan informasi per parsel. Informasi tersebut dirangkai dengan teknik *K-means* menjadi beberapa kelompok dan setiap kumpulan memiliki kualitas yang sama atau setara dengan yang lain namun dengan Kelompok yang berbeda memiliki atribut yang berbeda-beda. Dengan tujuan untuk membatasi perbedaan antara setiap informasi dalam satu kelompok dan memperluas perbedaan dengan kelompok yang lainnya. [3]

Dalam beberapa penelitian sebelumnya yang dilakukan oleh Bapak Amiruddin dengan judul “*Clustering Tingkat Pemahaman Dasar Mahasiswa pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-means*” berhasil mendapatkan klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih efektif. Hal ini membuktikan jika metode *K-means* dapat digunakan dalam proses peng*Clusteran* Siswa Bermasalah yang butuh Penanganan Khusus.

Maka dengan demikian teknik *K-means* sangat bagus di gunakan untuk mengelompokan siswa yang bermasalah berdasarkan pembagian kelas tertentu yaitu kelas berat, sedang, dan ringan. Dimana kelas berat adalah kelas yang terdapat beberapa siswa bermasalah serius dan mendapatkan penanganan khusus, sedangkan kelas sedang dan ringan adalah kelas yang terdapat beberapa siswa dengan penangan standar seperti pembinaan dan sangsi-sangsi ringan. [4]

Dalam beberapa penelitian sebelumnya yang dilakukan oleh Amiruddin dengan judul “*Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika dengan Metode K-means*” berhasil mendapatkan klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih akurat. Hal ini membuktikan jika metode *K-means* dapat digunakan dalam proses peng*Clusteran* Siswa Bermasalah.

Dengan adanya masalah yang dihadapi oleh Sekolah SMKS Taruna Bahari, maka penulis tertarik melakukan penelitian dengan mengimplementasikan

algoritma *Clustering* pada siswa yang bermasalah dengan judul **“IMPLEMENTASI ALGORITMA K-MEANS UNTUK PENGELOMPOKAN SISWA BERMASALAH”**.

Sangat diharapkan agar penelitian ini di pilih guna untuk bisa membantu mengelompokan siswa yang bermasalah di SMKS Taruna Bahari, Sehingga masalah yang ada di sekolah bisa lebih mudah diselesaikan.

1.2 Identifikasi Masalah

Berdasarkan Latar Belakang diatas dapat di identifikasi Bahwa sulitnya kontrol Guru BK terhadap siswa yang bermasalah di sebabkan karena keberadaan jumlah Guru BK tidak sebanding dengan Jumlah siswa yang butuh penanganan Khusus, sehingga Ketika adanya masalah siswa yang datang secara berurutan menyebabkan masalah siswa yang pertama datang tidak lagi di prioritaskan. Dampaknya masalah yang sama terulang kembali dengan penanganan yang belum terselesaikan.

1.3 Rumusan Masalah

Adapun Rumusan Masalah yang dapat kita lihat dari identifikasi diatas yaitu

1. Bagaimana cara menentukan jumlah *Cluster* yang optimal untuk menentukan Jumlah kelompok Siswa yang bermasalah dengan kategori Kelas Ringan, Sedang, dan Berat. sebagai kelompok prioritas yang mendapatkan penanganan Khusus.
2. Bagaimana hasil analisis setiap *Clustering* siswa bermasalah tingkat Berat gunak untuk mendapatkan penanganan Khusus.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui Jumlah *Cluster* yang optimal untuk menentukan Jumlah kelompok Siswa yang bermasalah dengan kategori Kelas Ringan, Sedang, dan Berat. sebagai kelompok prioritas yang mendapatkan penanganan Khusus.

2. Mengetahui analisis setiap *Clustering* siswa bermasalah tingkat Berat guna untuk mendapatkan penanganan Khusus.

1.5 Manfaat Penelitian

1. Pengembangan Ilmu

Peneliti diharapkan dapat menambah ilmu pada bidang teknologi informasi, dalam hal pengimplementasian Algoritma *Clustering* pada pengelompokan siswa yang bermasalah di SMKS Taruna Bahari.

2. Manfaat Praktisi

Harapan dari penelitian ini dapat membantu sekolah SMKS Taruna Bahari untuk menyelesaikan masalah pada siswa, demi untuk memajukan kecerdasan bangsa.

3. Peneliti

Semoga dengan penelitian ini dapat menambah wawasan dan pengalaman yang lebih nyata dalam hal pengimplementasian Algoritma *Clustering*.

BAB II

LANDASAN TEORI

2.1 Tinjauan studi

Adapun penelitian terkait menggunakan metode *Clustering* sebagai berikut :

Tabel 2.1 Penelitian Terkait

NO	PENELITI	JUDUL	TAHUN	METODE	HASIL
1	Rezqiwati dan Amiruddin	<i>Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahinan Probabilitas Statistika Dengan Metode K-means</i>	2022	<i>K-means</i>	Pengelompokan tingkat pemahaman dasar mahasiswa berdasarkan nilai mata kuliah prasyarat Metode yang digunakan yaitu K-Means untuk <i>Clustering</i> .
2	Bagus Setya Ryantiarna, ST.M,Kom	Klasterisasi siswa smk negeri 6 jember dengan tingkat kedisiplinannya	2018	<i>K-means</i>	Mengelompokan siswa yang bermsalah dari berbagai

		menggunakan algoritma <i>K-means</i>			tingkat kedisiplinan, dengan menerapkan metode <i>Clustering</i> algoritma <i>K-</i> <i>means</i>
3	Yuhandri Yunus	Klasterisasi Penempatan Siswa Yang Optimal untuk meningkatkan nilai rata-rata kelas menggunakan <i>K-</i> <i>means</i> menggunakan <i>K-</i> <i>means</i>	2021	<i>K-means</i>	Mengelomp okan siswa menjadi beberapa tingkatan dengan dasar ukur nilai rata- rata dengan mengunk an metode <i>Clustering</i> algoritma <i>K-</i> <i>means</i>

2.2 Tinjauan Pustaka

2.2.1 Kedisiplinan

Kata disiplin berasal dari bahasa Latin *discipulus*, dan itu berarti mendidik atau mengikuti yang dihormati. Menurut Kamus Besar Bahasa Indonesia, menyatakan bahwa disiplin adalah:[3]

- Sebuah Permintaan (di sekolah, di tempat kerja, di militer, dll).
- Konsistensi (konsistensi) dengan standar dan pedoman.

c. Bidang studi yang memiliki item dan kerangka kerja tertentu.

Disiplin adalah suatu kondisi yang dibuat dan dibingkai melalui perjalanan perkembangan cara berperilaku yang menunjukkan sisi positif dari ketaatan, kepatuhan, kesetiaan, konsistensi atau permintaan menjadi penting untuk perilaku sepanjang kehidupan sehari-hari. Perilaku dilakukan melalui siklus yang dibudayakan melalui keluarga, pelatihan, dan pengalaman (Prijodarminto dalam Tu“u 2004, hlm.31).

Gagasan disiplin siswa berhubungan dengan teknik, keputusan atau standar yang berlaku di sekolah. Sekolah adalah sarana untuk mendapatkan informasi edukatif, sahabat, daerah belajar. Di sekolah, ada keputusan yang telah ditetapkan untuk kelancaran siswa dalam mengerjakan tugas belajar. [5]

2.2.2 Konsep dasar Peraturan Tata Tertib Sekolah

Dalam Permendikbud Nomor 19 Tahun 2007 mengatur tentang Pedoman Pelaksanaan Kode Etik dalam point c dan d sebagai berikut: c. Sekolah/Madrasah menetapkan tata tertib disiplin yang memuat:

- 1) Peraturan bagi pendidik, tenaga kependidikan, dan siswa, dengan tetap memperhatikan pengaturan penggunaan dan pengendaliannya. kantor dan panggung pencerahan;
- 2) Instruksi, peringatan, dan pemecatan perilaku di sekolah/madrasah, misalnya memberikan persetujuan kepada penghuni yang menyalahgunakan standar. D. Sekolah/madrasah tidak seluruhnya diselesaikan oleh kepala/madrasah melalui rapat kepala sekolah dengan mempertimbangkan komitmen dewan sekolah/madrasah, dan siswa.1 Berdasarkan Permendikbud di atas, menjadi acuan aturan dalam menyusun dan melaksanakan suatu standar yang dihubungkan dengan lingkungan sekolah dengan mempertimbangkan kebutuhan dasar penjabaran tata tertib sekolah.[6]

2.2.3 Pengertian Tata Tertib Sekolah

Sebagai aturan umum, peraturan sekolah dapat diartikan sebagai signifikansi atau keputusan yang harus dipatuhi oleh setiap individu dari sekolah tempat pendidikan dan pengalaman pendidikan itu terjadi. Pelaksanaan tata tertib sekolah akan benar-benar ingin berjalan dengan baik jika para pendidik, pihak sekolah dan bawahannya. Merupakan standar untuk mempertahankan sekolah yang sebenarnya, kekurangan bantuan dari siswa akan menyebabkan kurang kritisnya pedoman sekolah yang diterapkan di sekolah.[7]

2.2.4 Tujuan Peraturan Sekolah

Secara umum, pedoman sekolah memiliki tujuan aturan agar semua siswa mengetahui apa komitmen, hak dan tanggung jawab mereka dan menyelesaiakannya dengan benar [8] sehingga persiapan sekolah dapat berjalan sesuai dengan bentuknya. Norma disiplin sekolah diperlukan, disarankan, dan ada hal-hal yang tidak boleh dilakukan melihat seseorang di dalam iklim sekolah

2.2.5 Tabel Pelanggaran

Dan adapun tabel pelanggaran beserta point-point pelanggaran yang ada di SMKS TARUNA BAHARI GORONTALO UTARA. Sebagai berikut :

Tabel 2.2 Pelanggaran

No	Klasifikasi Pelanggaran	Bobot Poin Pelanggaran
1	Datang terlambat masuk sekolah satu kali	5
2	Keluar tanpa izin	5
3	Tidak melaksanakan tugas piket kelas	5
4	Berpakaian seragam tidak lengkap	5
5	Makan di kelas waktu pelajaran sedang berjalan	10
6	Membuang sampah tidak pada tempatnya	5
7	Bermain di tempat parkir	5

No	Klasifikasi Pelanggaran	Bobot Poin Pelanggaran
8	Memakai gelang, kalung, anting-anting bagi pria dan berhias berlebihan bagi perempuan	10
9	Tidak memperhatikan panggilan guru dan pegawai	10
10	Berada di lantai pada waktu pergantian pelajaran	10
11	Membawa HP ke sekolah	10
12	Berpakaian tidak rapi	5
13	Siswa tidak masuk tanpa keterangan	5
14	Membuat izin palsu	10
15	Membolos/keluar/meninggalkan kelas	10
16	Membawa buku, majalah, kaset terlarang/gambar porno/HP bergambar porno	20
17	Melindungi teman yang salah/provokator	20
18	Melompat pagar/tidak mengikuti upacara	20
19	Mengganggu/mengacau kelas lain	10
20	Bersikap tidak sopan/menentang guru/karyawan	50
21	Mencoret-coret tembok, pintu, meja, kursi, dan memecahkan kaca yang tidak semestinya	10
22	Melakukan tindakan amoral	50
23	Melakukan tindakan pemerasan dan pemalakan terhadap teman dan orang lain	50
24	Membawa dan merokok pada lingkungan sekolah	20
25	Berbahasa yang kotor/kasar kepada teman/orang lain	20
26	Menyampaikan berita yang tidak benar (memprovokasi) orang tua, wali, masyarakat sehingga melakukan tindakan anarkis atau campur tangan orang tua tanpa melalui proses	70
27	Rambut gondrong, mengecat rambut dan tidak ropsi	20
28	Membuat tato pada anggota badan	20
29	Memalsu tanda tangan walikelas, guru dan kepala sekolah	80
30	Membawa/minum minuman keras	80
31	Berkelahi atau main hakim sendiri	50

No	Klasifikasi Pelanggaran	Bobot Poin Pelanggaran
32	Merusak sarana/prasarana sekolah	50
33	Mengambil milik orang lain (mencuri)	70
34	Membawa/menyebarkan selebaran yang menimbulkan keresahan	70
35	Berurusan dengan pihak yang berwajib karena melakukan kejahatan	80
36	Membawa senjata tajam tanpa sepenuhnya sekolah	50
37	Merubah atau memalsukan raport	80
38	Mengikuti organisasi terlarang	80
39	Terlibat dalam menyalahgunakan narkoba/zat aditif lainnya	80
40	Melakukan tindakan seksual di luar nikah	100
41	Memprovokasi orang tua/masyarakat yang dapat merongrong kewibawaan sekolah	80
42	Berjudi di sekolah	50
43	Nikah/kawin dalam Pendidikan	100

Sumber Data : SMKS Taruna Bahari

2.2.6 Pengertian Siswa

Siswa adalah individu yang sengaja datang ke sekolah. Menilik penjelasan dari pengertian di atas, dapat dikatakan bahwa siswa adalah individu yang dengan sengaja mensurvei di sekolah untuk mengembangkan batas sejatinya melalui kerangka pembelajaran yang edukatif, baik persiapan yang tepat (dari SD, SMP, dan SMA), seperti persiapan nonformal

2.2.7 Data Set

No	Nama Peserta Didik	Jumlah Mata pelajaran yang belum Tuntas	Presentase Ketidak Kehadiran	Jumlah Bobot Pelanggaran
1	Sandra Hasan	3	28	73 Point
2	SELISTI NUKU	0	11	21 Point

No	Nama Peserta Didik	Jumlah Mata pelajaran yang belum Tuntas	Presentase Ketidak Kehadiran	Jumlah Bobot Pelanggaran
3	Sri Hiola Tonang	0	38	72 Point
5	AKSEL AGUSTIAN DUKI	2	80	38 Point
6	Alwin Ishak	1	43	83 Point
7	Anisa Mokodongan	3	53	60 Point
8	AULIA LIBUNELO	0	22	17 Point

Tabel 2.3 Data Set

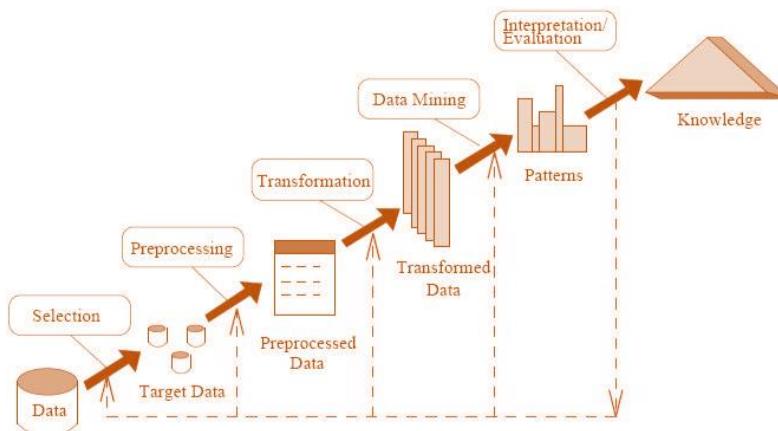
Mengenai dataset yang diperoleh dari lokasi eksplorasi untuk digunakan sebagai informasi dalam memutuskan penelitian yang berbahaya.

2.2.8 Data Mining

Data Mining Merupakan prosedur yang disebut "penambangan data" melihat bagaimana pengetahuan dapat menambah nilai pada kumpulan data. Data Mining adalah proses mengekstraksi pengetahuan dari sejumlah besar data, Mengklasifikasikan pola yang perlu ditemukan dalam penambangan data adalah tujuan Data Mining.

2.2.9 Tahapan Data Mining

Pemilihan data dari sumber hingga target merupakan langkah pertama dalam proses data mining. Berikut tahapan pengolahan untuk meningkatkan kualitas data, transfer data, data mining, dan tahapan interpretasi [4] dan pengujian atau evaluasi. Diharapkan hasilnya berupa pengetahuan baru, akan meningkatkan kontribusi .Berikut ini adalah spesifikasinya:



Gambar 2.1. Proses Knowledge Discoveryin Database (KDD)

1. Data Selection

Satu set data operasional harus sebelum Penemuan Pengetahuan di Database tahap ekstraksi informasi dapat dimulai. Salah satunya digunakan untuk menyimpan data yang dipilih untuk proses data mining. File yang berbeda dari database yang sebenarnya.[9]

2. Pre-processing / cleaning

Data yang menjadi fokus KDD harus dibersihkan sebelum proses data mining. Kesalahan data, penghapusan data duplikat, dan pemeriksaan konsistensi data adalah bagian dari proses pembersihan.

3. Transformation

Data yang dipilih dapat digunakan dalam proses data mining karena coding adalah proses mengubahnya. Dalam KDD 11, coding adalah proses kreatif yang sebenarnya ditentukan oleh jenis atau modus informasi yang akan dicari dalam database. [9]

4. Data Maining

Proses menggunakan metode atau teknik tertentu untuk menemukan pola atau informasi yang menarik dalam data yang dipilih dikenal sebagai data mining. Dalam data mining, ada banyak algoritma, teknik, atau metode yang berbeda. Tujuan dan proses KDD secara keseluruhan memainkan peran penting berperan dalam menentukan pendekatan atau algoritma. [4]

5. *Interpretation / evalution*

Pola informasi proses data mining harus disajikan sedemikian rupa sehingga peserta yang tertarik dapat dengan mudah memahami. Prosedur KDD (juga disebut interpretasi) termasuk tahap ini. Memeriksa untuk melihat apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau asumsi sebelumnya adalah bagian dari tahap ini.

2.2.10 *Clustering*

Proses pengelompokan set data yang atribut kelasnya belum diketahui disebut *Clustering*. Secara konseptual, *Clustering* berarti memaksimalkan dan meminimalkan kesamaan kelas. Misalnya, ada satu set objek, dan proses pertama dapat dikelompokkan menjadi beberapa set kelas-kelas dan akhirnya menjadi himpunan beraturan sehingga dapat diturunkan dari kelompok-kelompok klaster tertentu. Suatu kelompok juga dapat dipahami sebagai sebuah klaster.[11] Dengan demikian, sejumlah klaster (grup) akan dihasilkan dengan analisis pengelompokan. memahami bahwa sekumpulan data tertentu sebenarnya sudah mengandung kesamaan di antara anggota lain sebelum melakukan analisis. Akibatnya, anggota yang memiliki sifat yang sama dimasukkan ke dalam satu atau lebih kelompok. Dalam proses pengelompokan data, tujuannya adalah untuk meminimalkan himpunan fungsi tujuan, biasanya dengan meminimalkan variasi dalam sebuah *Cluster* dan memaksimalkan variasi antar *Cluster*.[10]

2.2.11 *Algoritma K-means*

Salah satu teknik *Clustering* dalam proses pemodelan Data *Mining* tanpa pengawasan dan metode pengelompokan data secara berkelompok yang demikian itu adalah algoritma *K-means* disebut juga dengan “*Clustering*.” Dengan menggunakan metode *K-means*, data dibagi menjadi beberapa kelompok dengan karakteristik yang mirip dengan kelompok lain tetapi berbeda dari kelompok lain. [6]Dengan maksud untuk memaksimalkan perbedaan dengan kelompok lain dan meminimalkan perbedaan dalam satu kelompok. Algoritma pengelompokan *K-means* menggunakan istilah:

1. *Cluster* adalah kumpulan dari grup.
2. Titik pusat untuk menghitung jarak Auclidian adalah Cendroid.
3. Iterasi adalah proses pengulangan sampai hasil konvergen.

Tahapan algoritma *K-means* secara umum adalah sebagai berikut:

1. Temukan k *Cluster* pada tahap awal
2. Pilih sejumlah k objek secara acak untuk dijadikan sebagai pusat Centoid *Cluster*.
3. Temukan k centroid, atau titik tengah.
4. Menggunakan jarak Euclidean untuk mengelompokkan objek ke dalam *Cluster* centroid terdekat:

$$dik = \sqrt{\sum (C_{ij} - C_{kj})^2} \quad n \quad i=1 \quad (2.1)$$

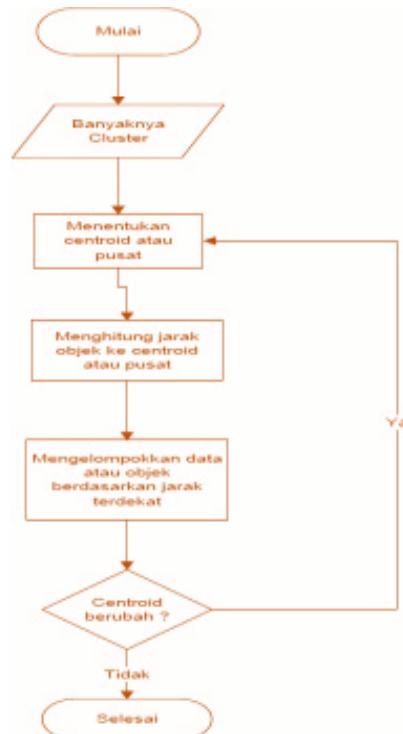
Dimana:

C_{ij} = pusat cluster

C_{kj} = data

5. Semua titik Centroid di hitung kembali.

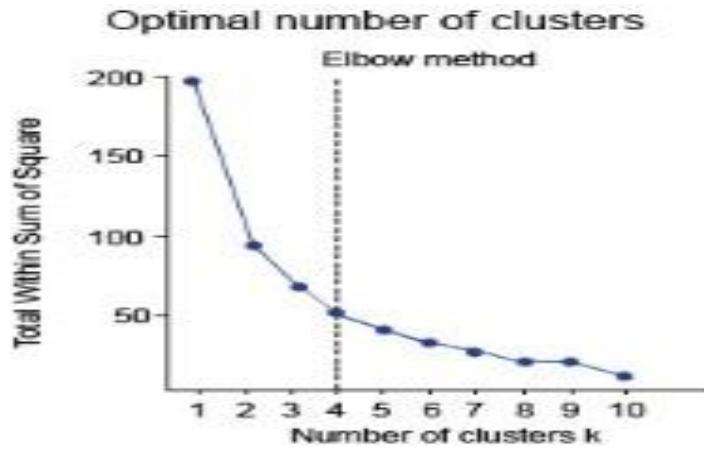
6. Pada tahap ini mengulang kembali tahap 3-5 sampai tidak terjadi perubahan nilai pada centroid.



Gambar 2.2. Proses Algoritma *K-means*

2.2.12 Metode *Elbow*

Metode *Elbow* merupakan salah satu teknik yang dapat digunakan dengan algoritma *K-means* untuk mengetahui besaran K yang terbaik. Metode *Elbow* hanya menggunakan titik pada graf dimana penurunan inersia tidak lagi signifikan untuk mengetahui jumlah K terbaik.[4] Hal ini ditunjukkan pada gambar di bawah ini, di mana K = 4 adalah angka optimal.



Gambar 2.3. Contoh Visualisasi Penggunaan Metode *Elbow*

2.2.13 Penerapan Algoritma *K-means*

Algoritma *K-means* di bawah ini merupakan contoh yang berhasil di terapkan dengan judul “*Clustering* Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika dengan Metode *K-means*”. Adapun proses *Clustering* ini [4] menggunakan tools yaitu bahasa pemrograman *Python* dan beberapa *library* seperti *Numpy*, *Pandas*, *Matplotlib* dan *Scikit Learn*.[8]

1. Tahapan *Selection*

Pilih dataset yang akan digunakan pada saat ini, yaitu nilai siswa pada mata kuliah prasyarat, seperti matematika diskrit pada tabel 2.5 dan kalkulus pada tabel 2.4.

Tabel 2.4. Sample Dataset Nilai Mata Kuliah Kalkulus

No	NIM	Nilai Huruf
1	T31 19 101	D
2	T31 20 002	A
3	T31 20 003	B
4	T31 20 004	A
5	T31 20 005	A
6	T31 20 006	B
7	T31 20 007	A

No	NIM	Nilai Huruf
8	T31 20 008	A
9	T31 20 009	A
10	T31 20 010	A

Tabel 2.5. Dataset Nilai Mata Kuliah Matematika Diskrit

No	NIM	Nilai Huruf
1	T31 14 097	B
2	T31 16 126	D
3	T31 16 342	D
4	T31 17 006	B
5	T31 17 057	B
6	T31 17 069	D
7	T31 17 149	B
8	T31 17 168	C
9	T31 18 003	D
10	T31 20 010	A

2. Tahapan *Processing*

Proses penggabungan data nilai antara nilai mata kuliah Kalkulus dan nilai mata kuliah Matematika Diskrit selesai pada titik ini, dan hasil penggabungan dataset ditampilkan pada tabel 2.6 di bawah ini:

Tabel 2.6. Sample Hasil Penggabungan Dataset

No	NIM	Nilai Kalkulus	Nilai M_Diskrit
1	T31 20 002	A	A
2	T31 20 003	B	C
3	T31 20 004	A	D
4	T31 20 005	A	A
5	T31 20 006	B	D
6	T31 20 007	A	A

No	NIM	Nilai Kalkulus	Nilai M_Diskrit
7	T31 20 008	A	B
8	T31 20 009	A	B
9	T31 20 010	A	A
10	T31 20 011	B	A

3. Tahapan *Transformation*

Pada tahapan ini menggunakan konversi data nilai yaitu nilai mata kuliah bentuk huruf dikonversikan ke bentuk nilai angka yaitu nilai A = 4, B = 3, C = 2, D = 1, hal ini dilakukan agar mudah Mengelompokan data.

Tabel 2.7. Sample Hasil Konversi Nilai Huruf Ke Nilai Angka

No	NIM	Nilai Kalkulus	Nilai M_Diskrit
1	T31 20 002	4	4
2	T31 20 003	3	2
3	T31 20 004	4	1
4	T31 20 005	4	4
5	T31 20 006	3	1
6	T31 20 007	4	4
7	T31 20 008	4	3
8	T31 20 009	4	3
9	T31 20 010	4	4
10	T31 20 131	3	4

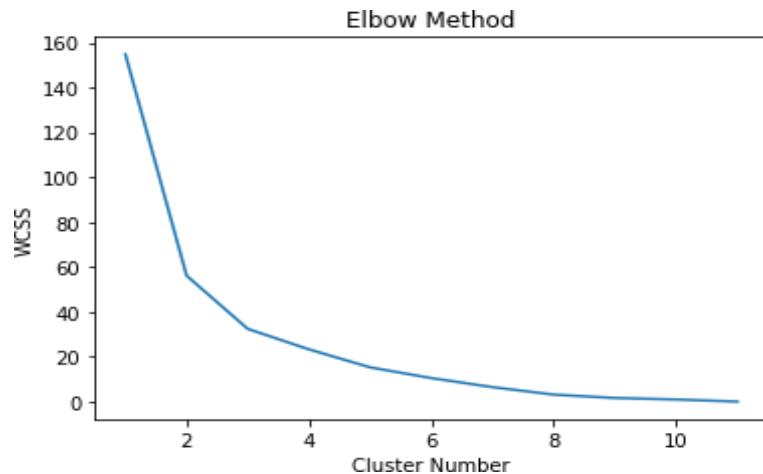
4. Tahapan *Data Mining*

Menggunakan model *K-means* untuk pengelompokan, model data mining dikembangkan pada titik ini. Library dari *Scikit Learn* digunakan dalam perintah

yang digunakan dalam penelitian ini untuk proses pemodelan metode *K-means*. [10]

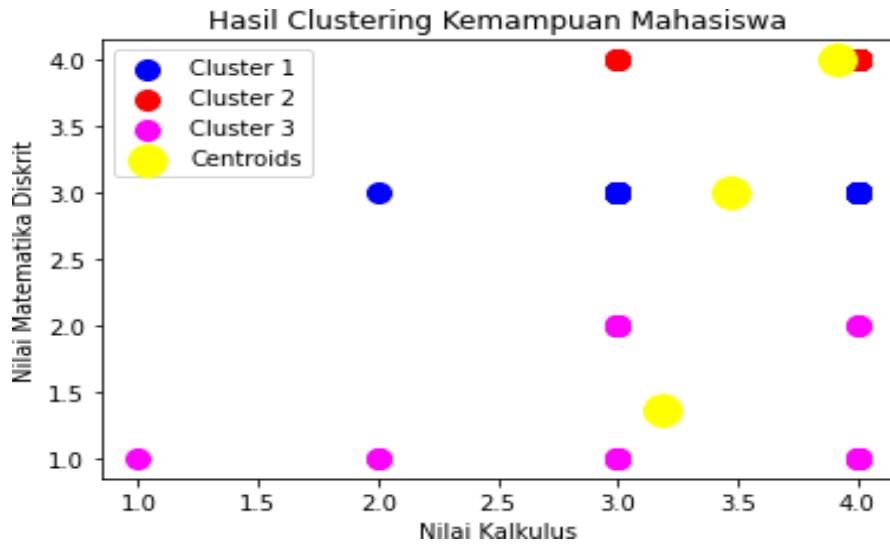
5. Tahapan *Interpretation/Evaluation*:

Metode atau teknik *Elbow* digunakan untuk menentukan jumlah *Cluster* yang optimal pada titik ini. Library Matplotlib digunakan untuk membuat grafik atau visualisasi dari hasil metode, yang dapat dilihat pada Gambar 4 di bawah ini:



Gambar 2.4. Visualisasi Hasil Metode *Elbow*

Jumlah *Cluster* yang optimal untuk penelitian ini adalah tiga, seperti yang ditunjukkan pada Gambar 2.2 di atas. Tanpa harus menguji jumlah *Cluster* berulang kali, metode *Elbow* ini memungkinkan identifikasi langsung jumlah optimal. Setelah menentukan jumlah *Cluster* yang optimal, jumlah dari *Cluster* dimasukkan ke dalam fungsi Scikit Learn *K-means*, yang digunakan:



Gambar 2.5. Visualisasi Hasil *Cluster*

Posisi koordinat pusat *Cluster* kuning dapat digunakan untuk menentukan titik centroid untuk setiap *Cluster*, seperti yang ditunjukkan pada Gambar 2.3, atau perintah "`(kmeans.Cluster_centers_)`" dapat digunakan secara langsung untuk menentukan titik *centroid*, menghasilkan hasil sebagai berikut :

```
[[3.46666667 3.]
 [3.9137931 4.]
 [3.18181818 1.36363636]]
```

Gambar 2.6. Informasi Pusat *Cluster*

Hasilnya ditunjukkan pada tabel 2.6 di bawah ini, dan perintah "`dfData['Cluster'].value_counts()`" dapat digunakan untuk menentukan jumlah anggota untuk setiap *Cluster* yang telah dihasilkan:

Tabel 2.8. *Centroids* Dan Jumlah Data

Cluster	Centroids		Jumlah Mahasiswa
	Kalkulus	Matematika Diskrit	
1	3,46	3	30
2	3,91	4	58
3	3,18	1,36	22

Berdasarkan titik-titik *Centroids* pada masing-masing *Cluster* pada tabel 2.8 sebelumnya, dapat ditentukan bahwa klaster 1 berkategori baik, klaster 2 berkategori sangat baik, dan klaster 3 berkategori buruk. Tabel 2.9 menampilkan hasil pengelompokan dan pelabelan untuk setiap siswa:

Tabel 2.9. Sampel Hasil *Clustering*

NIM	Kalkulus	M_Diskrit	Kluster
T31 20 002	4	4	2
T31 20 003	3	2	3
T31 20 004	4	1	3
T31 20 005	4	4	2
T31 20 006	3	1	3
T31 20 007	4	4	2
T31 20 008	4	3	1
T31 20 009	4	3	1
T31 20 010	4	4	2
T31 20 011	3	4	1

2.3 Perangkat Lunak Pendukung

Adapun perangkat lunak sebagai pendukung dari peneliti dalam melakukan *Clustering* diantaranya adalah:

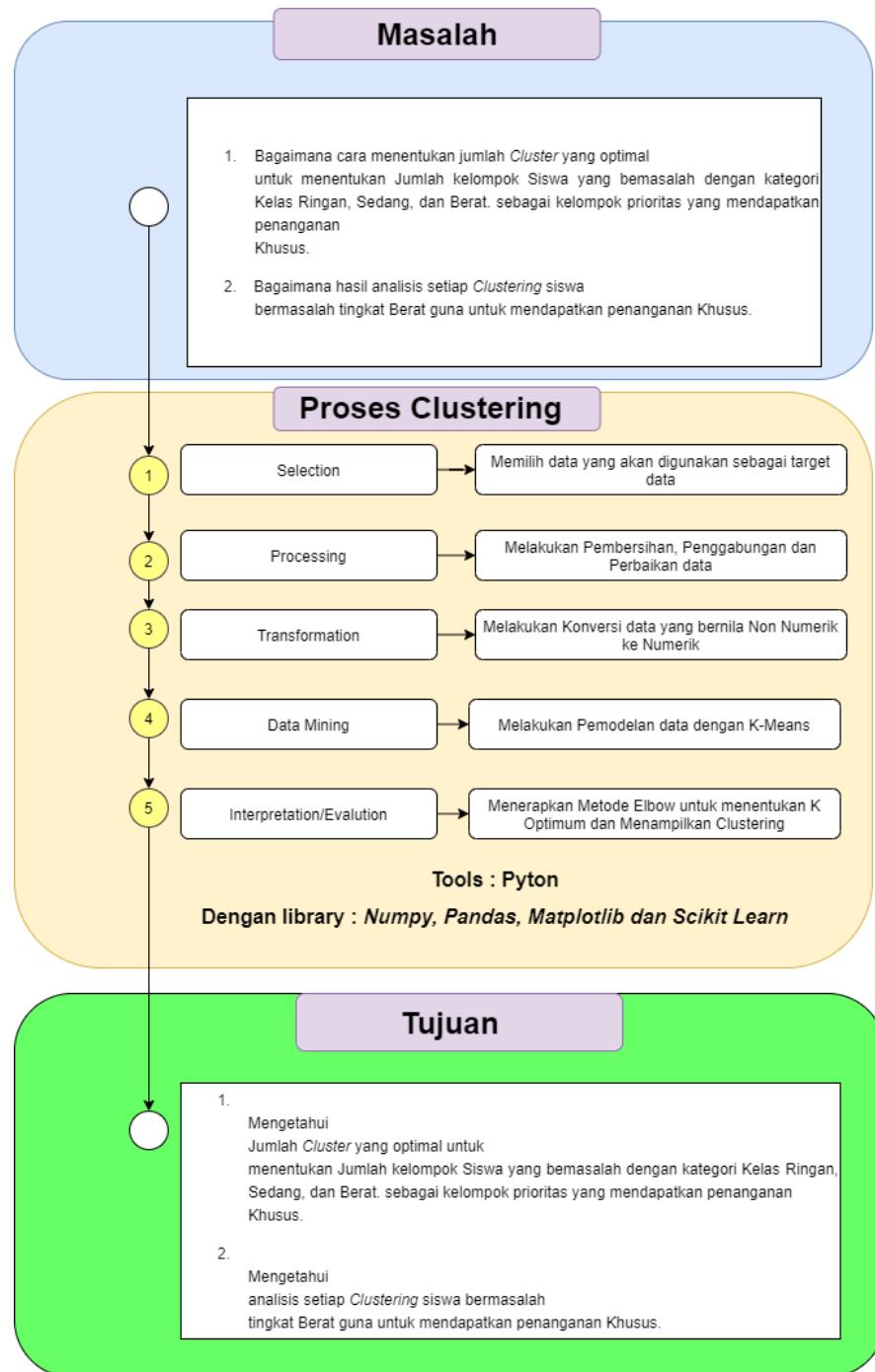
Tabel 2.10. Perangkat Lunak Pendukung

No.	Perangkat Lunak Pendukung	Keterangan
1.	<i>Google Colaboratory</i> (<i>Google Colab</i>)	<i>Integrated Development Environments (IDE)</i> menuliskan coding Bahasa <i>Python</i> berbasis Web yang akan digunakan.

No.	Perangkat Lunak Pendukung	Keterangan
2.	<i>Python</i>	<i>Open source, lintas platform, cocok untuk Ilmuwan Data, dan didukung oleh sejumlah besar Library, termasuk: Matplotlib, Numpy, SciPy, Pandas, dan Scikit-Learn</i>

2.4 Kerangka Pikir

Adapun kerangka pikir penelitian dalam mengelompokan siswa berasalah yaitu sebagai berikut.



Gambar 2.1: Kerangka Pikir

BAB III

METODE PENELITIAN

3. 1 Metode, Subjek dan Objek Penelitian

Penelitian ini termasuk penelitian terapan karena tingkat strateginya. Penelitian ini merupakan penelitian kuantitatif berdasarkan jenis informasi yang diproses. Secara simultan, dari perilaku hingga informasi, penelitian ini merupakan laporan yang menguatkan sehingga mudah di proses.

Di SMK Taruna Bahari, pendekatan penelitian studi kasus digunakan dalam penelitian ini. Oleh karena itu jenis penelitian ini deskriptif. Pengelompokan Siswa Bermasalah Menggunakan Metode *K-means* adalah fokus penelitian ini. Ujian ini dimulai dari Oktober hingga Desember 2022 bertempat di SMKS Taruna Bahari.

3.2 Pengumpulan Data

Data primer dan sekunder adalah dua jenis data yang digunakan untuk pengumpulan data. Data primer adalah informasi yang dikumpulkan di lapangan atau dengan terjun langsung ke Lokasi. Data sekunder, di sisi lain, adalah data yang berasal dari penelitian atau literatur yang telah dipublikasikan sebelumnya. Data yang digunakan dalam penelitian ini ada 2 jenis yaitu sebagai berikut:

1. Data Primer

Data primer adalah data yang dikumpulkan langsung oleh Peneliti di SMKS Taruna Bahari yang akan digunakan sebagai informasi dalam mengelompokan siswa yang bermasalah.

- a. Observasi : Dilakukan pencarian informasi langsung dilapangan mengenai siswa yang bermasalah di SMKS Taruna Bahari
- b. Wawancara : Dilakukan wawancara pada guru bimbingan konseling yang sedang bertugas di SMKS Taruna Bahari

Adapun variabel dengan tipe datanya masing-masing ditunjukan pada Tabel 3.1 berikut ini:

Tabel 3.1 Variabel Data

No	Variabel/Atribut	Type	Nilai	Ket
1	Jumlah Mata pelajaran yang belum Tuntas	Rasio	0 s/d 15	Variabel Input
2	Presentase Ketidak Kehadiran	Rasio	0 s/d 100	Variabel Input
3	Jumlah Bobot Pelanggaran	Rasio	0 s/d 1500	Variabel Input

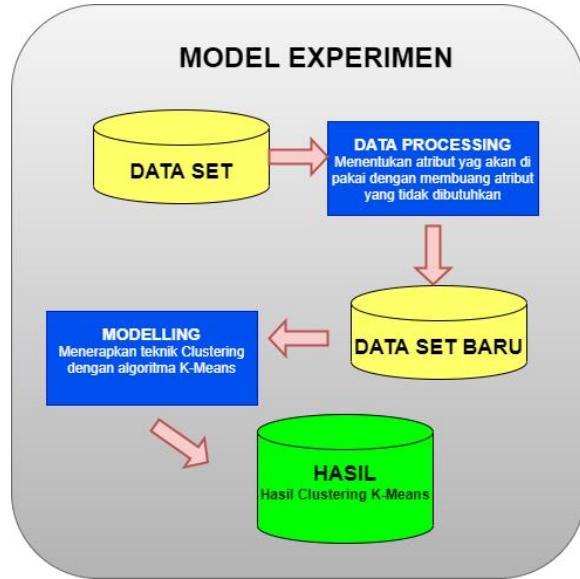
2. Penelitian Data Sekunder

Landasan teoritis penelitian kepustakaan adalah sumber data sekunder, yang dapat digunakan selain data primer. Analis sistem menggunakan pendekatan ini untuk menemukan contoh catatan atau dokumen yang terkait dengan penelitian. Analis sistem mencari buku, majalah, dan publikasi lainnya untuk data terkait penelitian atau item lainnya.

3.3 Pemodelan

Metode yang digunakan dalam penelitian ini adalah metode *Clustering*. *Clustering* adalah pengelompokan dengan pendekatan parsel dan pengelompokan dengan metodologi progresif. Pengelompokan dengan pendekatan persil atau sering disebut *segment based grouping information* dengan cara menyusun informasi yang sudah dipecah ke dalam kelompok-kelompok yang sudah ditentukan.

Pengelompokan dengan metodologi progresif juga sering di sebut sebagai pengelompokan berjenjang dengan membuat data sebagai dendogram di mana informasi yang sebanding akan ditempatkan pada jarak data terdekat dan tidak pada data yang jauh.



Gambar 3.1: Pemodelan Clustering

Sebelum data diolah, terlebih dahulu mencari infomasi demi untuk mendapatkan data yang dibutuhkan, hal ini dilakukan menganalisis apakah terdapat masalah sesuai dengan judul yang akan diteliti.

Hasil clustering dapat dari hasil penelusuran data yang paling dekat dengan pusat cluster, kemudian mengelompokan data tersebut pada satu cluster, dan mengulangi lagi dari awal sampai semua cluster saling terhubung.

Dari hasil yang didapatkan kemudian dievaluasi agar dapat mendapatkan kinerja dari metode clustering yang digunakan, kemudian dilakukan evaluasi lagi ke dalam data centroid untuk mendapatkan nilai yang lebih akurat.

3.4 Tahap *Clustering*

Mengikuti kerangka kerja berikut, serangkaian prosedur dilakukan pada tahap berikut ini:

1. Tahap *Selection* : pada tahapan ini akan dilakukan seleksi data terhadap dataset yang dikumpulkan dari SMKS Taruna Bahari untuk dijadikan sebagai target data (dataset)
2. Tahap *Preprocessing* : Pada titik ini, sejumlah proses dilakukan, termasuk penghapusan atribut yang tidak diperlukan untuk pemodelan,

verifikasi data yang tidak konsisten, penyelesaian data kosong, dan penghapusan data duplikat.

3. Tahap *Transformation* : Pada titik ini terjadi perubahan atau konversi data, khususnya konversi nilai yang belum sesuai akan di sesuaikan untuk digunakan dalam pemodelan *K-means*.

4. Tahap *Data Mining* : Pada titik ini dibangun sebuah model yang menggunakan pendekatan *K-means* untuk mendapatkan hasil *Cluster*.

5. Tahap *Interpretation/Evaluation* : Pada titik ini, menggunakan metode *Elbow* untuk menentukan jumlah *Cluster* yang ideal. Metode *K-means* digunakan sekali lagi untuk menerapkan hasil nilai K optimal metode *Elbow*. Hasilnya adalah divisualisasi jumlah *Cluster* dan individu *Cluster*, sehingga Dilihat dari representasinya, maka secara langsung di ketahui masing-masing *Cluster* yang kemudian digunakan sebagai informasi atau pilihan untuk mengelompokkan Siswa bermasalah pada kelas Berat, Sedang, dan Ringan..

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Penulis penelitian ini mengelompokkan siswa dengan permasalahan pada Satuan Pendidikan di Kabupaten Gorontalo Utara dengan menggunakan Metode *K-Means*. Di SMK TARUNA BAHARI, sebuah lembaga pendidikan menengah, kapasitas siswa untuk berfungsi di bidang Kejuruan, kapasitas mereka untuk menyesuaikan diri dengan lingkungan kerja, dan kapasitas mereka untuk tumbuh sebagai individu di masa depan agar diberikan prioritas utama. Untuk mengidentifikasi Siswa bermasalah, penulis menggunakan data dari SMK Taruna Bahari. Untuk melakukan ini, penulis mengunjungi sekolah secara langsung, mengamati, dan melakukan wawancara.

Adapun Hasil pengumpulan informasi dari lokasi penelitian adalah sebagai berikut:

Tabel 4.1 Hasil Pengumpulan Data

No	Nama	Tingkat	Jurusan	JNT	PKH	JPB
1	Sandra Hasan	1	Nautika Kapal Niaga	3	28	73
2	SELISTI NUKU	1	Nautika Kapal Niaga	0	11	21
3	Sri Hiola Tonang	1	Nautika Kapal Niaga	0	38	72
4	AKSEL AGUSTIAN DUKI	1	Nautika Kapal Niaga	2	80	38
5	Alwin Ishak	1	Nautika Kapal Niaga	1	43	83

No	Nama	Tingkat	Jurusan	JNT T	PKH	JPB
6	Anisa Mokodongan	1	Nautika Kapal Niaga	3	53	60
7	AULIA LIBUNEOLO	1	Nautika Kapal Niaga	0	22	17
8	FIRMAN MONGGI	1	Nautika Kapal Niaga	2	48	10
9	Ilham Noe	1	Nautika Kapal Niaga	1	67	77
10	Indriyati T.D Hasim	1	Nautika Kapal Niaga	0	76	36
11	IRHAM R. HASAN	1	Nautika Kapal Niaga	2	25	11
12	MICKY ISMAIL	1	Nautika Kapal Niaga	1	51	55
13	MOHAMAD ICHSAN S. USMAN	1	Nautika Kapal Niaga	3	34	14
14	Mohammad Akbar Nurdin	1	Nautika Kapal Niaga	2	65	50
15	MUHAMAD KEFIN KARIM	1	Nautika Kapal Niaga	1	47	39
16	PUTRI AYU ALFIANE LAMATO	1	Nautika Kapal Niaga	0	57	57
17	RAHMIN BALALANGI	1	Nautika Kapal Niaga	0	28	66
18	Ilham Noe	1	Nautika Kapal Niaga	2	76	37

No	Nama	Tingkat	Jurusan	JNT T	PKH	JBK
19	Indriyati T.D Hasim	1	Nautika Kapal Niaga	4	86	10
20	FAISAL LAKU	1	Nautika Kapal Niaga	2	17	18
...
179	Pian Datau	3	Rekayasa Perangkat Lunak	0	3	61

Sumber Data : SMKS TARUNA BAHARI Tahun 2022

Tabel 4.1 adalah hasil peneliti dari pendataan di SMKS TARUNA BAHARI. Selain itu, penulis hanya menggunakan 179 data dan beberapa karakteristik dalam penelitian, adapun data tersebut merupakan data dari kelas X, XI, dan XII. Untuk mengkategorikan siswa bermasalah, penulis membutuhkan tiga kriteria, yaitu jumlah nilai tidak tuntas, persentase ketidakhadiran, dan bobot pelanggaran secara keseluruhan.

4.2 Hasil Pemodelan

4.2.1 Pra Pengolahan Data

Lokasi penelitian yaitu di SMK Taruna Bahari Gorontalo Utara dimana data yang diperoleh akan diolah untuk mengidentifikasinya; data yang sebelumnya tidak teridentifikasi harus diproses untuk dikenali. Dari sinilah data yang digunakan dalam penelitian ini berasal. Ini dilakukan untuk dapat memperkirakan volume data yang akan dikelompokkan.

4.2.2 Normalisasi Data

Peneliti Menemukan 3 Variabel untuk digunakan dalam penelitian ini guna utnuk mengidentifikasi cluster yang dimaksud. Variabel tersebut adalah (Jumlah Nilai yan Tidak Tuntas, Presentase Ketidak Hadiran, dan Jumlah Bobot Pelanggaran).

4.2.3 Hasil Hitungan K-Means

Pada penelitian ini peneliti menggunakan data pelanggaran siswa yang diperoleh dari lokasi penelitian yaitu SMK Taruna Bahari :

Jumlah Data : 179

Jumlah Cluster : Untuk Perhitungan Manual Peneliti menggunakan 4 Cluster, Untuk menggunakan Program Python peneliti belum bisa menentukan dari awal Jumlah Cluster yang optimum, karna jumlah cluster yang optimum akan di ketahui setelah menggunakan Metode Elbow pada Program Python.

Jumlah Atribut : 6

Jumlah Variable : 2

Ket : C1 = Berat Sekali

C2 = Berat

C3 = Sedang

C4 = Ringan

Setelah melakukan beberapa analisis, peneliti mendapatkan kesulitan dalam menentukan pusat cluster, di karenakan pada variable Nilai tidak tuntas banyak terdapat angka nilai Nol (0), sehingga peneliti melakukan percobaan kembali dengan perhitungan K-Means dengan menggunakan 2 Variabel yaitu variable Presentase Ketidak Hadiran dan Jumlah Bobot Pelanggaran. Sehingga peneliti menemukan hasilnya Cluster yang lebih mudah untuk di analisis. Berikut hasil analisis dengan menggunakan 2 variabel.

data di ambil secara acak, untuk menentukan Nilai Centroid.

Tabel 4.14 penentuan awal cluster literasi 1

cluster	PKH	JBP
c1	28	73
c2	53	60
c3	76	36
c4	17	18

Ket : PKH = Presentase Ketidak Hadiran.

JBP = Jumlah Bobot Pelanggaran.

Penentuan Cluster Awal

Data Ke - 1 diambil sebagai pusat Cluster 1

Data Ke - 6 diambil sebagai pusat Cluster 2

Data Ke – 10 di ambil sebagai pusat Cluster 3

Data Ke-20 diambil sebagai cluster 4

Tabel 4.15 Menentukan Jarak Literasi 1

No	NAMA	PKH	JBK	C1	C2	C3	C4	Klmpok Kluster
1	Sandra Hasan	28	73	0	28.17800561	60.6052803	56.08921465	C1
2	SELISTI NUKU	11	21	54.70831747	57.31491952	66.70832032	6.708203932	C4
3	Sri Hiola Tonang	38	72	10.04987562	19.20937271	52.34500931	57.93962375	C1
4	AKSEL AGUSTIAN DUKI	80	38	62.68173578	34.82814953	4.472135955	66.09841148	C3
5	Alwin Ishak	43	83	18.02775638	25.07987241	57.42821606	70.00714249	C1
6	Anisa Mokodongan	53	60	28.17800561	0	33.24154028	55.31726674	C2
7	AULIA LIBUNEO	22	17	56.32051136	53.00943312	57.24508713	5.099019514	C4
8	FIRMAN MONGGI	48	10	66.09841148	50.24937811	38.20994635	32.01562119	C4
9	Ilham Noe	67	77	39.20459157	22.02271555	41.97618372	77.3369252	C2
10	Indriyati T.D Hasim	76	36	60.6052803	33.24154028	0	61.68468205	C3
11	IRHAM R. HASAN	25	11	62.07253821	56.43580424	56.79788728	10.63014581	C4
12	MICKY ISMAIL	51	55	29.20616373	5.385164807	31.40063694	50.24937811	C2
13	MOHAMAD ICHSAN S. USMAN	34	14	59.30430001	49.76946855	47.41307836	17.4642492	C4
14	Mohammad Akbar Nurdin	65	50	43.56604182	15.62049935	17.80449381	57.68882041	C2
15	MUHAMAD KEFIN KARIM	47	39	38.94868419	21.84032967	29.154775947	36.61966685	C2
16	PUTRI AYU ALFIANE LAMATO	57	57	33.12099032	5	28.31960452	55.86591089	C2
17	RAHMIN BALALANGI	28	66	7	25.70992026	56.60388679	49.24428901	C1
18	Ilham Noe	76	37	60	32.52691193	1	61.98386887	C3
19	Indriyati T.D Hasim	86	10	85.63293759	59.9082632	27.85677655	69.46221995	C3
20	FAISAL LAKU	17	18	56.08921465	55.31726674	61.68468205	0	C4
21	MOHAMAD ICHSAN S. USMAN	26	37	36.05551275	35.4682957	50.009999	21.02379604	C4
22	Mohammad Akbar Nurdin	19	22	51.7880295	50.99019514	58.69412236	4.472135955	C4
23	MUHAMAD KEFIN KARIM	29	27	46.01086828	40.80441153	47.85394446	15	C4
24	PUTRI AYU ALFIANE LAMATO	1	38	44.20407221	56.46237685	75.02666193	25.61249695	C4
25	RAHMIN BALALANGI	33	72	5.099019514	23.32380758	56.08029957	56.32051136	C1
26	RANI KAJELE	21	21	52.46903849	50.44799302	57.00877125	5	C4
27	RENDI KINO	24	15	58.13776741	53.53503526	56.08029957	7.615773106	C4
28	SINDI DUKI	16	57	20	37.12142239	63.5688603	39.01281841	C1
29	TIARANDA CANTIKA SAHUPALA	35	81	10.63014581	27.65863337	60.87692502	65.520989	C1
30	Sandri Hasan	47	14	61.98386887	46.38965402	36.40054945	30.2654919	C4

Pada klaster yang telah di lakukan sama dengan proses perhitungan dengan menggunakan 3 variabel, sehingga inilah hasil klister dengan menggunakan 2 variabel

Tabel 4.16 Menentukan Awal Pusat Cluster Literasi 2

No	NAMA	PKH	JBP	Cluster	Cluster Baru			
					C1	C2	C3	C4
1	Sandra Hasan	28	73	C1	31.57142857	56.66666667	79.5	24.92307692
2	SELISTI NUKU	11	21	C4		72	56.33333333	30.25
3	Sri Hiola Tonang	38	72	C1				
4	AKSEL AGUSTIAN DUKI	80	38	C3				
5	Alwin Ishak	43	83	C1				
6	Anisa Mokodongan	53	60	C2				
7	AULIA LIBUNEO	22	17	C4				
8	FIRMAN MONGGI	48	10	C4				
9	Ilham Noe	67	77	C2				
10	Indriyati T.D Hasim	76	36	C3				
11	IRHAM R. HASAN	25	11	C4				
12	MICKY ISMAIL	51	55	C2				
13	MOHAMAD ICHSAN S. USMAN	34	14	C4				
14	Mohammad Akbar Nurdin	65	50	C2				

Tabel 4.17 Awal Pusat Cluster Literasi 2

cluster	PKH	JBP
c1	31.57142857	72
c2	56.66666667	56.33333333
c3	79.5	30.25
c4	24.92307692	20.38461538

Tabel 4.18 Menghitung Jarak Literasi 2

No	NAMA	PKH	JBP	C1	C2	C3	C4	Cluster
1	Sandra Hasan	28	73	3.708787139	33.15954697	66.93140145	52.70527634	C1
2	SELISTI NUKU	11	21	54.99257835	57.73983797	69.12172235	13.93666995	C4
3	Sri Hiola Tonang	38	72	6.428571429	24.36983564	58.86690496	53.24616274	C1
4	AKSEL AGUSTIAN DUKI	80	38	59.172008	29.67415636	7.766112284	57.82533381	C3
5	Alwin Ishak	43	83	15.86229003	29.96479416	64.14680429	65.17255203	C1
6	Anisa Mokodongan	53	60	24.55979791	5.185449729	39.8410906	48.55607385	C2
7	AULIA LIBUNEO	22	17	55.82662667	52.42984731	59.00688519	4.472135955	C4
8	FIRMAN MONGGI	48	10	64.13967539	47.13691641	37.4474632	25.30582177	C4
9	Ilham Noe	67	77	35.77965446	23.10603577	48.39227728	70.53913262	C2
10	Indriyati T.D Hasim	76	36	57.1830216	28.05748068	6.731456009	53.41060108	C3
11	IRHAM R. HASAN	25	11	61.35294348	55.29818161	57.79976211	9.384930638	C4
12	MICKY ISMAIL	51	55	25.8160684	5.821416399	37.7466886	43.33855984	C2
13	MOHAMAD ICHSAN S. USMAN	34	14	58.05082221	48.01967189	48.31472343	11.09747026	C4
14	Mohammad Akbar Nurdin	65	50	40.01836313	10.46687898	24.50127548	49.83202554	C2
15	MUHAMAD KEFIN KARIM	47	39	36.42857143	19.8466342	33.65728004	28.87772631	C2
16	PUTRI AYU ALFIANE LAMATO	57	57	29.52307987	0.745355992	34.95443463	48.67869539	C2
17	RAHMIN BALALANGI	28	66	6.982485377	30.25263992	62.69220446	45.71904165	C1
18	Ilham Noe	76	37	56.5588009	27.34146221	7.603453163	53.71147994	C3
19	Indriyati T.D Hasim	86	10	82.50132961	54.83814569	21.26763974	61.95345648	C3
20	FAISAL LAKU	17	18	55.93144492	55.16238654	63.68918668	8.274148806	C4
21	MOHAMAD ICHSAN S. USMAN	26	37	35.44066614	36.252203	53.92413653	16.65024832	C4
22	Mohammad Akbar Nurdin	19	22	51.55619086	50.96621975	61.05990911	6.139406135	C4
23	MUHAMAD KEFIN KARIM	29	27	45.07340951	40.32231254	50.60447115	7.770753849	C4
24	PUTRI AYU ALFIANE LAMATO	1	38	45.72321341	58.60792514	78.88163601	29.70884354	C4
25	RAHMIN BALALANGI	33	72	1.428571429	28.38231061	62.49249955	52.24351266	C1
26	RANI KAJELE	21	21	52.08411564	50.20513475	59.2267887	3.971049077	C4
27	RENDI KINO	24	15	57.50066548	52.6835416	57.55703693	5.463163355	C4
28	SINDI DUKI	16	57	21.62104039	40.67213078	68.90437214	37.68696979	C1
29	TIARANDA CANTIKA SAHUPAL	35	81	9.630945023	32.83121821	67.49675918	61.44728823	C1
30	Sandri Hasan	47	14	60.01700439	43.42298419	36.33610463	22.98159799	C4

Pada Perhitungan jarak literasi 2 sudah mendapekatan kluster yang same dengan hasil klister 1, sehingga peneliti mencukupkan uji coba pada tahp literasi 2 ini.

BAB V

PEMBAHASAN

5.1 Implementasi K-Means Dengan Python

Implementasi pemodelan K-means ini menggunakan Bahasa pemrograman Python dengan menggunakan Notebook Colaboratory, berikut adalah Tahapannya.

5.1.1 Import Library Python

Berikut Library Python yang perlu di impor agar proses pemodelan K-Means dapat di jalankan tanpa Kendala.

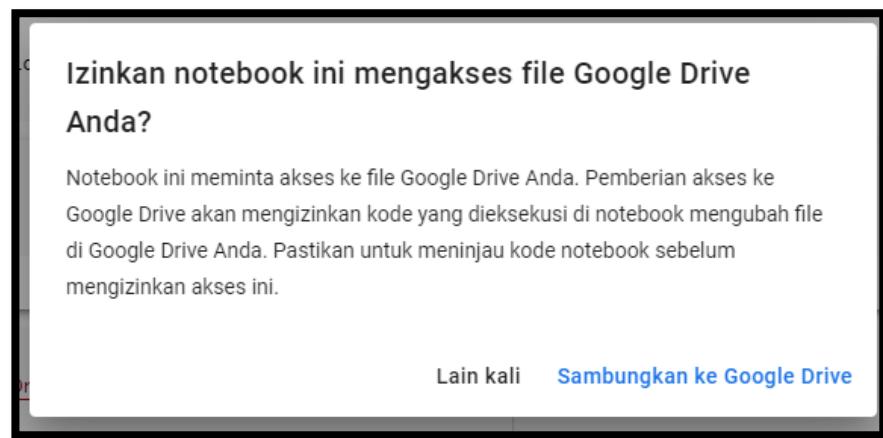
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
import plotly.graph_objs as go
from plotly import tools
from plotly.subplots import make_subplots
import plotly.offline as py
```

5.1.2 Mengaktifkan Google Drive

Agar program dapat mengakses data set yang kan di olah, maka Google Drive perlu untuk di aktifkan dengan menggunakan perintah berikut:

```
from google.colab import drive
drive.mount('/content/drive')
```

Jika Program ini di jalankan, maka akan muncul Tampilan Konfirmasi untuk izin akses penggunaan Email.



Gambar 5.1 Mengaktifkan Google Drive

5.1.3 Membaca Data Set

Dataset yang sudah dibersihkan yang akan digunakan untuk clustering terlebih dahulu harus diupload ke Google Drive dalam bentuk file excel, setelah itu dataset dibaca menggunakan perintah berikut:

```
df = pd.read_excel('/content/drive/MyDrive/UJI/program_skripsi/data_siswa.xlsx')
df.head()
```

Berikut hasil menjalankan perintah di atas.

	JNTT	PKH	JPB
0	3	28	73
1	0	11	21
2	0	38	72
3	2	80	38
4	1	43	83

Gambar 5.2 Membaca Data Set

5.1.4 Konversi Data Set Excel ke CSV

Perintah berikut harus digunakan untuk mengubah dataset dalam format excel menjadi CSV (COMMA Separated Values) agar dapat digunakan dalam pemodelan data.

```
df.to_csv("/content/drive/MyDrive/UJI/program skrip
si/data_siswa.csv", index=False)
df.info()
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179 entries, 0 to 178
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype  
 ---  --      --      --      --      
 0   JNTT    179 non-null    int64  
 1   PKH     179 non-null    int64  
 2   JBP     179 non-null    int64  
 dtypes: int64(3)
memory usage: 4.3 KB
```

Gambar 5.3 Conversi Data Set Excel ke CSV

5.1.5 Menghitung Banyak Data

Menghitung jumlah data yang dibutuhkan untuk setiap nilai atribut penting sebelum mengganti nilai atribut.

```
df.describe()
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

	JNTT	PKH	JBP
count	179.000000	179.000000	179.000000
mean	0.743017	27.223464	54.949721
std	1.071014	17.597411	28.496653
min	0.000000	0.000000	10.000000
25%	0.000000	12.000000	28.500000
50%	0.000000	27.000000	57.000000
75%	1.000000	39.000000	79.000000
max	5.000000	86.000000	100.000000

Gambar 5.4 Hasil Menghitung Banyak Data

5.1.6 Menentukan Kolom Sebagai Dataset

Atribut yang akan digunakan dalam pemodelan k-means perlu didefinisikan terlebih dahulu, yaitu atribut: Penilaian Ketidak Hadiran dan Jumlah Bobot Pelanggaran.

```
X = df[['PKH', 'JBP']].values
X
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

```
array([[ 28,  73],
       [ 11,  21],
       [ 38,  72],
       [ 80,  38],
       [ 43,  83],
       [ 53,  60],
       [ 22,  17],
       [ 48,  10],
       [ 67,  77],
       [ 76,  36],
       [ 25,  11],
       [ 51,  55],
       [ 34,  14],
       [ 65,  50],
       [ 47,  39],
       [ 57,  57]]
```

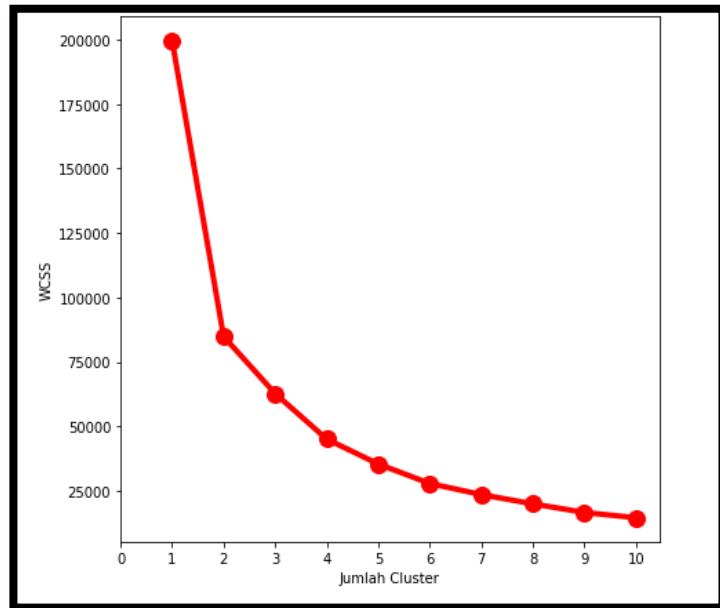
Gambar 5.5 Menentukan Kolom sebagai Data Set

5.2 Kinerja Model

Agar tidak membutuhkan banyak waktu untuk menguji jumlah klaster yang tepat, sehingga diperlukan suatu metode untuk mengidentifikasi jumlah klaster yang ideal dengan memanfaatkan metodologi Elbow. Performansi metode K-Means dalam pengelompokan data sangat bergantung pada jumlah cluster yang ditentukan. Berikut adalah perintah yang di gunakan untuk menentukan Jumlah Cluster yang Optimum.

```
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = 'k-
means++')
    model.fit(X)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4, markersize=
12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:



Gambar 5.6 Visualisasi Teknik Elbow

Jelas dari Gambar 5.1 di atas bahwa 4 cluster adalah jumlah cluster yang ideal. Sehingga jumlah cluster = 4 dapat langsung digunakan untuk aplikasi teknik K-Means.

5.3 Pembahasan Model

5.3.1 Pemodelan Metode K-Means

Langkah selanjutnya adalah menerapkan pemodelan K-Means menggunakan potongan kode berikut setelah menentukan jumlah cluster yang ideal:

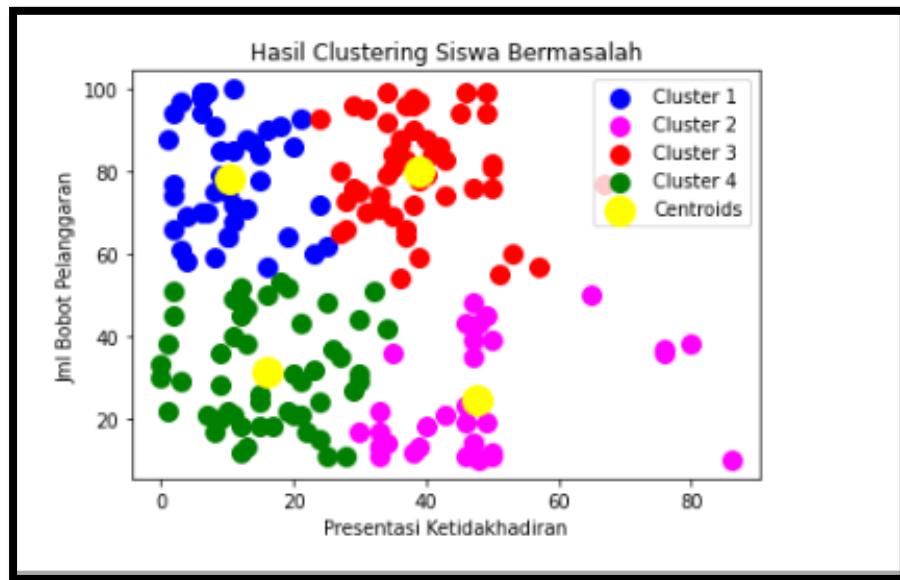
```
model = KMeans(n_clusters = 4, init = "k-means++",
                max_iter = 300, n_init = 10, random_
                state = 0)
y_clusters = model.fit_predict(X)
```

5.3.2 Visualisasi Hasil Pemodelan K-Means

Setelah melakukan pemodelan K-Means, berikut ini dapat dilakukan untuk membuat visualisasi setiap cluster dengan menggunakan potongan kode seperti berikut:

```
plt.scatter(X[y_clusters == 0, 0], X[y_clusters == 0, 1],  
           s = 100, c = 'blue', label = 'Cluster 1')  
plt.scatter(X[y_clusters == 1, 0], X[y_clusters == 1, 1],  
           s = 100, c = 'magenta', label = 'Cluster 2')  
plt.scatter(X[y_clusters == 2, 0], X[y_clusters == 2, 1],  
           s = 100, c = 'red', label = 'Cluster 3')  
plt.scatter(X[y_clusters == 3, 0], X[y_clusters == 3, 1],  
           s = 100, c = 'green', label = 'Cluster 4')  
plt.scatter(X[y_clusters == 4, 0], X[y_clusters == 4, 1],  
           s = 100, c = 'cyan', label = 'Cluster 5')  
plt.scatter(model.cluster_centers_[:, 0], model.cluster_ce  
           nters_[:, 1],  
           s = 250, c = 'yellow', label = 'Centroids')  
plt.title('Hasil Clustering Siswa Bermasalah')  
plt.xlabel('Presentasi Ketidakhadiran')  
plt.ylabel('Jml Bobot Pelanggaran')  
plt.legend()  
plt.show()
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:



Gambar 5.7 Visualisasi Hasil Clustering

Seperti yang terlihat pada gambar di atas, setiap cluster diwakili oleh warna yang berbeda, dengan cluster 1 berwarna biru, cluster 2 berwarna ungu, cluster 3 berwarna merah, cluster 4 hijau dan pusat cluster berwarna kuning. Hubungan antara ketiga kualitas yang digunakan dalam pengelompokan juga dapat dilihat pada gambar di atas.

5.3.3 Pusat Cluster dan Jumlah Data

Dalam proses analisis, selain melihat langsung gambar di atas, penting juga untuk mengetahui nilai pusat klaster setiap klaster serta jumlah data pada masing-masing klaster agar dapat diambil kesimpulan tentang masing-masing klaster. Pusat setiap kluster dapat ditentukan menggunakan potongan kode berikut.

```
print(model.cluster_centers_)
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

```
[[10.45238095 78.30952381]
 [47.55882353 24.35294118]
 [38.78431373 80.03921569]
 [16.13461538 31.48076923]]
```

Gambar 5.8 Penentuan Pusat Cluster

Adapun untuk menghitung jumlah data per kluster dengan menggunakan perintah di bawah ini:

```
df['Kluster'].value_counts()
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

```
3    52
2    51
0    42
1    34
Name: Kluster, dtype: int64
```

Gambar 5.9 Hasil Perhitungan Jumlah Data

Tabel 5.1 Tabel Pusat Cluster dan Jumlah Data

Cluster	Pusat Cluster		Jumlah Data
	Presentasi Ketidakhadiran	Jumlah Bobot Pelanggaran	
1	10	78	42
2	47	24	34
3	38	80	51
4	16	31	52

Dari tabel di atas, dapat dianalisis atau disimpulkan bahwa :

Cluster 1 adalah kelompok Siswa yang Bermasalah tingkat **Sedang** dan menjadi prioritas Ke Tiga untuk mendapatkan penanganan, dengan Nilai Pelanggaran Presentasi Kehadiran 10 dan Jumlah Bobot Pelanggaran 78 dengan jumlah data 42 Orang, Hasil ini menunjukan Bawa yang termasuk di Cluster Sedang yaitu siswa yang memiliki nilai tinggi pada JBP.

Cluster 2 adalah kelompok Siswa yang Bermasalah tingkat **Berat** dan menjadi prioritas Ke Dua untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 47 Keatas dan Jumlah Bobot Pelanggaran 24 Keatas dengan jumlah data 34 Orang. Hasil ini menunjukan Bawa yang termasuk di Cluster Berat yaitu siswa yang memiliki nilai Tinggi Pada PKH

Cluster 3 adalah kelompok Siswa yang Bermasalah tingkat **Berat Sekali** dan menjadi prioritas Pertama untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 38 dan Jumlah Bobot Pelanggaran 80 dengan jumlah data 51 Orang. Hasil ini menunjukan Bawa yang termasuk di Cluster Berat Sekali yaitu siswa yang memiliki nilai tinggi pada PKH dan JBP.

Cluster 4 adalah kelompok Siswa yang Bermasalah tingkat **Ringan** dan menjadi prioritas ke Empat untuk mendapatkan penanganan, dengan Nilai Pelanggaran Presentasi Kehadiran 16 Keatas dan Jumlah Bobot Pelanggaran 31 Keatas dengan jumlah data 52 Orang.

Hasil ini menunjukkan Bawa yang termasuk di Cluster Ringan yaitu siswa yang memiliki nilai Rendah pada JBP dan Rendah pada PKH.

5.3.4 Pelabelan Data Hasil Custer

Cuplikan kode berikut dapat digunakan untuk menentukan cluster mana yang dimiliki setiap data:

```
print(model.labels_)
# Menambahkan kolom "kluster" dalam data frame Data Nilai
df["Kluster"] = model.labels_
df
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

	JNTT	PKH	JBP	Kluster
0	3	28	73	2
1	0	11	21	3
2	0	38	72	2
3	2	80	38	1
4	1	43	83	2
...
174	0	19	64	0
175	0	25	62	0
176	0	43	74	2
177	0	46	11	1
178	0	3	61	0

Gambar 5.10 Pelabelan Data Hasil Cluster

Dengan potongan kode di bawah ini, informasi tambahan untuk nilai cluster dapat dibuat dari ketentuan:

```
conditions = [
    (df['Kluster']==0),
    (df['Kluster']==1),
    (df['Kluster']==2),
    (df['Kluster']==3)]
choices = ['C1', 'C2', 'C3', 'C4']
df['Ket'] = np.select(conditions, choices)
df
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

	JNTT	PKH	JBP	Kluster	Ket
0	3	28	73	2	C3
1	0	11	21	3	C4
2	0	38	72	2	C3
3	2	80	38	1	C2
4	1	43	83	2	C3

Gambar 5.11 Pelabelan Data Hasil Cluster Penambahan Keterangan

Dengan potongan kode di bawah ini, informasi Status untuk Setiap Keterangan Cluster.

```
conditions = [
    (df['Kluster']==0),
    (df['Kluster']==1),
    (df['Kluster']==2),
    (df['Kluster']==3)]
choices = ['Sedang', 'Berat', 'Berat Sekali', 'Ringan']
df['Status'] = np.select(conditions, choices)
df
```

Hasil yang diperoleh setelah menjalankan perintah tersebut di atas adalah sebagai berikut:

	JNTT	PKH	JBP	Kluster	status	Ket	edit
0	3	28	73	2	Berat Sekali	C3	
1	0	11	21	3	Ringan	C4	
2	0	38	72	2	Berat Sekali	C3	
3	2	80	38	1	Berat	C2	
4	1	43	83	2	Berat Sekali	C3	
...	
174	0	19	64	0	Sedang	C1	
175	0	25	62	0	Sedang	C1	
176	0	43	74	2	Berat Sekali	C3	
177	0	46	11	1	Berat	C2	
178	0	3	61	0	Sedang	C1	

Gambar 5.12 Pelabelan Data Hasil Cluster Penambahan Status Cluster

```
df.to_excel("/content/drive/MyDrive/UJI/program_skripsi/hasil2v_data_siswa.xlsx", index=False)
```

Tabel berikut menunjukkan hasil setelah menjalankan dua perintah yang dituliskan di atas:

Tabel 5.2 Hasil Clustering Data Siswa Bermasalah SMK Taruna Bahari

No	NAMA	JNTT	PKH	JBP	Kluster	Ket	Status
1	Sandra Hasan	3	28	73	2	C3	Berat Sekali
2	SELISTI NUKU	0	11	21	3	C4	Ringan
3	Sri Hiola Tonang	0	38	72	2	C3	Berat Sekali
4	AKSEL AGUSTIAN DUKI	2	80	38	1	C2	Berat
5	Alwin Ishak	1	43	83	2	C3	Berat Sekali
6	Anisa Mokodongan	3	53	60	2	C3	Berat Sekali
7	AULIA LIBUNELO	0	22	17	3	C4	Ringan
8	FIRMAN MONGGI	2	48	10	1	C2	Berat
9	Ilham Noe	1	67	77	2	C3	Berat Sekali
10	Indriyati T.D Hasim	0	76	36	1	C2	Berat
11	IRHAM R. HASAN	2	25	11	3	C4	Ringan
12	MICKY ISMAIL	1	51	55	2	C3	Berat Sekali
13	MOHAMAD ICHSAN S. USMAN	3	34	14	1	C2	Berat
14	Mohammad Akbar Nurdin	2	65	50	1	C2	Berat
15	MUHAMAD KEFIN KARIM	1	47	39	1	C2	Berat
16	PUTRI AYU ALFIANE LAMATO	0	57	57	2	C3	Berat Sekali
17	RAHMIN BALALANGI	0	28	66	2	C3	Berat Sekali
18	Ilham Noe	2	76	37	1	C2	Berat
19	Indriyati T.D Hasim	4	86	10	1	C2	Berat

No	NAMA	JNTT	PKH	JBK	Kluster	Ket	Status
20	FAISAL LAKU	2	17	18	3	C4	Ringan
21	MOHAMAD ICHSAN S. USMAN	3	26	37	3	C4	Ringan
22	Mohammad Akbar Nurdin	0	19	22	3	C4	Ringan
23	MUHAMAD KEFIN KARIM	0	29	27	3	C4	Ringan
24	PUTRI AYU ALFIANE LAMATO	2	1	38	3	C4	Ringan
25	RAHMIN BALALANGI	1	33	72	2	C3	Berat Sekali
26	RANI KAJELE	1	21	21	3	C4	Ringan
27	RENDI KINO	0	24	15	3	C4	Ringan
28	SINDI DUKI	0	16	57	0	C1	Sedang
29	TIARANDA CANTIKA SAHUPALA	0	35	81	2	C3	Berat Sekali
30	Sandri Hasan	0	47	14	1	C2	Berat
...
179	Pian Datau	0	3	61	0	C1	Sedang

Hasil Analisa clustering di atas yaitu siswa bermasalah terbagi menjadi 4 Cluster, adapun masing-masing cluster yaitu **Cluster 1** adalah kelompok Siswa yang Bermasalah tingkat **Sedang** dengan Nilai Pelanggaran Presentasi Kehadiran 10 dan Jumlah Bobot Pelanggaran 78 dengan jumlah data 42 Orang, Hasil ini menunjukan Bawa yang termasuk di Cluster Sedang yaitu siswa yang memiliki nilai tinggi pada JBP.

Cluster 2 adalah kelompok Siswa yang Bermasalah tingkat **Berat** dengan Nilai Pelanggaran Presentasi Kehadiran 47 Keatas dan Jumlah Bobot Pelanggaran 24 Keatas dengan jumlah data 34 Orang. Hasil ini menunjukan Bawa yang termasuk di Cluster Ringan yaitu siswa yang memiliki nilai Tinggi Pada PKH.

Cluster 3 adalah kelompok Siswa yang Bermasalah tingkat **Berat** dengan Nilai

Pelanggaran Presentasi Kehadiran 38 dan Jumlah Bobot Pelanggaran 80 dengan jumlah data 51 Orang. Hasil ini menunjukan Bahwa yang termasuk di Cluster Berat Sekali yaitu siswa yang memiliki nilai tinggi pada PKH dan JBP.

Cluster 4 adalah kelompok Siswa yang Bermasalah tingkat **Ringan** dengan Nilai Pelanggaran Presentasi Kehadiran 16 Keatas dan Jumlah Bobot Pelanggaran 31 Keatas dengan jumlah data 52 Orang. Hasil ini menunjukan Bahwa yang termasuk di Cluster Ringan yaitu siswa yang memiliki nilai Rendah pada JBP dan Rendah pada PKH.

BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan temuan penelitian di atas mengenai pengelompokan siswa bermasalah dengan metode K-Means, dapat ditarik kesimpulan sebagai berikut:

1. Jumlah cluster yang Optimal untuk melakukan clustering tingkat Siswa Bermasalah yaitu sebanyak 4 cluster dengan menggunakan teknik Elbow.
2. Hasil analisis clustering untuk masing-masing cluster yaitu Cluster tingkat Berat Sekali terdapat pada cluster 3 merupakan prioritas Pertama untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 38 dan Jumlah Bobot Pelanggaran 80 dengan jumlah data 51 Orang. siswa bermasalah tingkat Berat berada pada Cluster 2 merupakan prioritas Ke Dua untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 47 Keatas dan Jumlah Bobot Pelanggaran 24 Keatas dengan jumlah data 34 Orang.. siswa bermasalah tingkat Sedang berada pada Cluster 1 merupakan prioritas 3 untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 10 dan Jumlah Bobot Pelanggaran 78 dengan jumlah data 42 Orang. siswa bermasalah tingkat Ringan berada pada Cluster 4 merupakan prioritas Terakhir untuk mendapatkan penanganan dengan Nilai Pelanggaran Presentasi Kehadiran 16 dan Jumlah Bobot Pelanggaran 31 dengan jumlah data 52 Orang.. dari hasil analisis sekolah bisa mengambil tindakan baik berupa skorsing atau mengundang ke ruangan bimbingan konseling agar mendapatkan nasehat terkait perubahan dalam menjalani aktivitas Pendidikan jadi lebih baik.

6.2 Saran

Berdasarkan kesimpulan di atas, peneliti dapat memberikan rekomendasi untuk penelitian selanjutnya sebagai berikut:

1. Penelitian selanjutnya dapat menggunakan fitur lain atau menambah dataset agar hasil clustering dapat dibandingkan.
2. Penelitian selanjutnya dapat dibuat menggunakan teknik pengelompokan yang berbeda sehingga hasilnya dapat dibandingkan untuk melihat mana yang memberikan hasil terbaik.

DAFTAR PUSTAKA

- [1]. Kom Daryoto, M, “Implementasi Metode *K-means Clustering* Dalam Penilaian Kedisiplinan Siswa (Studi Kasus: Smk Sunan Kalijogo Lumajang),” 2018.
- [2] J. Waluyo, F. Teknik, and U. M. Magelang, *Penerapan Metode K-means Clustering Dalam Penilaian Kedisiplinan Siswa Untuk Layanan Bimbingan Dan Konseling Di SMP Negeri 3 Temanggung*. 2019.
- [3] A. Syahrin, “Implementasi algoritma *K-means* untuk klasterisasi mahasiswa berdasarkan prediksi waktu kelulusan skripsi,” *UPN “Veteran” Jatim*, vol. 1–23, 2013.
- [4] R. Ishak and A. Bengnga, “*Clustering* Tingkat Pemahaman Mahasiswa Pada Perkuliahan Probabilitas Statistika Dengan Metode *K-means*,” *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 1, pp. 65–69, 2022, doi: 10.37905/jjeee.v4i1.11997.
- [5] A. Afandi, “Identifikasi Masalah Belajar Siswa Smk Negeri 2 Jember Menggunakan Metode Naive Bayes Classifier,” 2016.
- [6] Lutfia Miftahur Rahma, “Implementasi Algoritma *K-means* Berbasis Android Untuk *Clustering* Tingkat Pemahaman Mahasiswa Skripsi Oleh : Lutfia Miftahur Rahma,” 2020.
- [7] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, “Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan *K-means*,” *J. Inf. dan Teknol.*, pp. 103–108, Sep. 2021, doi: 10.37034/jidt.v3i3.130.
- [8] Luz Yolanda Toro Suarez, *Klasterisasi Siswa SMK Negeri 6 Jember dengan tingkat kedisiplinannya menggunakan Algoritma K-means*. 2015. [Online]. Available: <http://repository.unmuhjember.ac.id/528/>

- [9] K. I. P. Kuliah, D. Metode, and M. S. Sompa, “*Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar Program Sarjana*,” 2022.
- [10] F. B. Abay, “*Clustering dengan algoritma K-means terhadap cal...*,” 2021.
- [11] L. Wicaksono Program Studi Bimbingan dan Konseling FKIP Untan Pontianak, “Persepsi Terhadap Perilaku Bermasalah Pada Siswa Kelas Viii Smpn 19 Pontianak,” 2017.

LAMPIRAN

KODE PROGRAM

1. Koversi Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179 entries, 0 to 178
Data columns (total 3 columns):
 #   Column   Non-Null Count   Dtype  
---  --  
 0   JNTT      179 non-null    int64  
 1   PKH       179 non-null    int64  
 2   JBP       179 non-null    int64  
dtypes: int64(3)
memory usage: 4.3 KB
```

2. Menentukan Jumlah Cluster dengan Teknik Elbow

```
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = 'k-means++')
    model.fit(X)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4,
         markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

3. Pemodelan K-Means Dengan Jumlah Cluster Yang Optimum

```
model = KMeans(n_clusters = 4, init = "k-means++",
               max_iter = 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(X)
```

4. Visualisasi Hasil Cluster

```
plt.scatter(X[y_clusters == 0, 0], X[y_clusters == 0, 1], s =
           100, c = 'blue', label = 'Cluster 1')
plt.scatter(X[y_clusters == 1, 0], X[y_clusters == 1, 1], s =
           100, c = 'magenta', label = 'Cluster 2')
plt.scatter(X[y_clusters == 2, 0], X[y_clusters == 2, 1], s =
           100, c = 'red', label = 'Cluster 3')
plt.scatter(X[y_clusters == 3, 0], X[y_clusters == 3, 1], s =
           100, c = 'green', label = 'Cluster 4')
```

```
#plt.scatter(X[y_clusters == 4, 0], X[y_clusters == 4, 1], s = 100, c = 'cyan', label = 'Cluster 5')
plt.scatter(model.cluster_centers_[:, 0],
model.cluster_centers_[:, 1],
s = 250, c = 'yellow', label = 'Centroids')
plt.title('Hasil Clustering Siswa Bermasalah')
plt.xlabel('Presentasi Ketidakhadiran')
plt.ylabel('Jml Bobot Pelanggaran')
plt.legend()
plt.show()
```

HASIL TURNITIN

 Similarity Report ID: oid:25211:34113227

PAPER NAME	AUTHOR
SKRIPSI_T3121149_AYIDIL KAI.docx	AYIDIL KAI
<hr/>	
WORD COUNT	CHARACTER COUNT
7619 Words	44627 Characters
PAGE COUNT	FILE SIZE
63 Pages	792.8KB
SUBMISSION DATE	REPORT DATE
Apr 17, 2023 10:22 AM GMT+8	Apr 17, 2023 10:23 AM GMT+8
<hr/>	
● 4% Overall Similarity	
The combined total of all matches, including overlapping sources, for each database.	
• 4% Internet database	• 1% Publications database
• Crossref database	• Crossref Posted Content database
• 0% Submitted Works database	
● Excluded from Similarity Report	
• Cited material	• Small Matches (Less than 30 words)

Summary

● 4% Overall Similarity

Top sources found in the following databases:

- 4% Internet database
- Crossref database
- 0% Submitted Works database
- 1% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	mkks-smplotim.org	2%
	Internet	
2	Aria Teimourzadeh, Samaneh Kakavand, Benjamin Kakavand. "APPLIC...	<1%
	Crossref	
3	farbenklick.de	<1%
	Internet	
4	kaggle.com	<1%
	Internet	
5	slideshare.net	<1%
	Internet	
6	jurnal.stmikroyal.ac.id	<1%
	Internet	
7	journal.stekom.ac.id	<1%
	Internet	
8	ojs.unm.ac.id	<1%
	Internet	

Sources overview

Daftar Riwayat Hidup

Nama : Ayidil Kai

NIM : T3121149

Jenis Kelamin : Laki - Laki

Tempat Tanggal Lahir: Tibawa, 21 Mei 1994



Status : Menikah

Alamat : Desa Molingkapoto, Kec, Kwandang, Gorontalo Utara

Kewarganegaraan : Indonesia

Agama : Islam

Email : ayidilkai@gmail.com

Riwayat Pendidikan :

1. Tamatan SD Negeri 1 Labanu pada tahun 2006
2. Tamatan SMP Negeri 2 Tibawa pada tahun 2009
3. Tamatan SMK Negeri 1 Limboto pada Tahun 2012
4. Tamata D3 Politeknik Gorontalo Pada Tahun 2015
5. Masuk dan diterima di Universitas Ichsan Gorontalo pada tahun 2021