

Implementasi K-Means Clustering dalam Mengklasifikasi Pengaruh Les Terhadap Prestasi Siswa dengan Metode Elbow

Alief Fathul Habibie*, Rakhmat Kurniawan R

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}alieffathul.habibie@gmail.com, ²rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: alieffathul.habibie@email.com

Submitted: 05/11/2024; Accepted: 22/11/2024; Published: 22/11/2024

Abstrak—Prestasi dan kemampuan siswa merupakan sebuah sudut pandang yang sangat penting untuk diperhatikan. Dalam hal ini, Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan merupakan sebuah institusi pendidikan dasar yang telah memiliki reputasi baik dalam bidang ilmu agama maupun ilmu lainnya. Penelitian ini bertujuan untuk menganalisis pengaruh bimbingan belajar terhadap prestasi siswa di Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan dengan menggunakan metode K-Means *Clustering* dan teknik Elbow untuk menentukan jumlah cluster yang optimal. Data yang digunakan dalam penelitian ini melibatkan 514 siswa dari kelas 1A hingga 6B, dengan variabel yang dianalisis mencakup nilai ujian semester ganjil dan genap, partisipasi dalam les tambahan, serta kegiatan ekstrakurikuler. Hasil analisis menunjukkan bahwa siswa yang mengikuti les tambahan memiliki rata-rata nilai yang lebih tinggi dibandingkan dengan siswa yang tidak mengikuti les. Rata-rata nilai siswa pada ujian semester ganjil dan genap adalah 87,71, sedangkan rata-rata nilai untuk siswa yang mengikuti les dan kegiatan ekstrakurikuler adalah 87,12. Proses clustering menghasilkan empat kelompok siswa, dengan kelompok yang memiliki prestasi tertinggi berada pada *cluster* 2, sedangkan kelompok dengan prestasi terendah berada pada *cluster* 3 dan 4. Penelitian ini memberikan informasi penting bagi pihak sekolah dalam memahami pengaruh bimbingan belajar terhadap prestasi akademik siswa dan dapat digunakan untuk meningkatkan strategi pembelajaran di MIS Al-Falah Medan.

Kata Kunci: Pengaruh Les; Clustering; K-Means; Ekstrakurikuler

Abstract—Student achievement and ability are very important perspectives to consider. In this regard, Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan is an educational institution that has a good reputation in both religious and other sciences. This study aims to analyze the influence of tutoring on student achievement at Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan using the K-Means Clustering method and the Elbow technique to determine the optimal number of clusters. The data used in this research involves 514 students from classes 1A to 6B, with the analyzed variables including semester exam scores, participation in additional tutoring, and extracurricular activities. The analysis results show that students who participate in additional tutoring have higher average scores compared to those who do not. The average score of students in the semester exams is 87.71, while the average score for students participating in tutoring and extracurricular activities is 87.12. The clustering process results in four groups of students, with the highest performing group in cluster 2, while the lowest performing groups are in clusters 3 and 4. This research provides important information for the school in understanding the impact of tutoring on students' academic performance and can be used to improve learning strategies at MIS Al-Falah Medan.

Keywords: The Influence of Tutoring; Clustering; K-Means; Extracurricular Activities

1. PENDAHULUAN

Perkembangan teknologi informasi saat ini telah menunjukkan sebuah proses kemajuan besar yang memberikan dampak positif pada banyak bidang termasuk dalam bidang pendidikan, ekonomi, kesehatan, seni, dan lainnya. Namun, proses kemajuan ini harus dibarengi dengan peningkatan kualitas sumber daya manusia (SDM) yang merupakan salah satu komponen penting dalam keberhasilan perkembangan teknologi informasi. Teknologi yang berkembang pesat menuntut penggunaannya, mau tidak mau harus menggunakan teknologi dalam memaksimalkan pekerjaan salah satunya pada metode pendidikan. Pendidikan menjadi salah satu aspek esensial dalam peningkatan mutu SDM yang unggul. Kualitas pendidikan yang baik tentunya harus diimbangi dengan proses belajar mengajar yang merata [1]. Hal tersebut dapat dilakukan dengan menganalisis capaian hasil belajar agar dapat menentukan kompetensi dan kemampuan peserta didik dalam mengikuti kegiatan belajar mengajar (KBM) di sekolah. Perolehan hasil analisis tersebut diharapkan dapat membantu tenaga pendidik dalam mengoptimalkan metode kegiatan pembelajaran yang merata di sekolah khususnya pada sekolah Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan.

Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan merupakan sebuah yayasan pendidikan dasar bagi masyarakat yang telah memiliki reputasi baik dalam bidang ilmu agama maupun ilmu lainnya. Dikarenakan setiap lembaga pendidikan pasti menginginkan lulusan alumni yang unggul dalam berbagai disiplin ilmu. Kemampuan prestasi dan kompetensi hasil belajar peserta didik dalam proses kegiatan belajar mengajar yang telah dilaksanakan selama di kelas tentunya harus ada peningkatan pada setiap semesternya. Agar tercapai tujuan dari belajar mengajar itu, maka perlu suasana belajar yang menyenangkan dan interaktif supaya dapat memberikan pengaruh dalam peningkatan hasil belajar peserta didik. Dalam memenuhi pencapaian tujuan dari pendidikan itu, maka siswa berperan penting menjadi subjek di dalamnya untuk berorientasi pada suatu lembaga pendidikan baik formal maupun informal. Adapun pendidikan formal sendiri merupakan pendidikan yang sudah disusun dengan struktur dan kurikulum sedemikian rupa, di mana pendidikan formal ini adalah pendidikan yang memiliki jenjang tingkatan mulai dari tingkat dasar, menengah, sampai tingkat tinggi. Sedangkan untuk

pendidikan informal sendiri yaitu suatu pendidikan selain pendidikan formal, tentu pendidikan informal sendiri pelaksanaannya tidak diatur oleh struktur tertentu, tidak ada kurikulum khusus dan tidak ada jenjang tingkatan di dalamnya. Seperti contohnya, bimbingan belajar (Bimbel), kursus atau *Workshop*, Pelatihan, dls. Namun apabila dilihat dari keduanya sama-sama mengutamakan keberhasilan para peserta didiknya dalam menunjang kemampuan akademik maupun prestasi yang hendak dicapai. Pengaruh bimbingan belajar dapat mempengaruhi hasil belajar para peserta didik, dikarenakan pembelajaran yang dilakukan di sekolah dapat dimaksimalkan kembali bahkan di kembangkan serta dialami oleh para peserta didik sesuai preferensinya dalam memaknai pembelajaran guna mencapai kualitas hidupnya di masa mendatang [2].

Namun dalam hal ini tidak semua siswa memiliki kemampuan yang sama, terdapat tingkatan antara tinggi dan rendahnya kemampuan kognitif siswa dikarenakan minat belajar dari para peserta didik yang bisa dikatakan kurang [3], oleh karena itu bimbingan belajar atau kegiatan belajar tambahan di luar pendidikan formal dapat mendukung untuk meningkatkan dan memberikan pengaruh terhadap tingkat kognitif para siswa. Selain itu dalam melihat rendahnya minat belajar antara korelatif dengan kognitif siswa dipengaruhi karena pelajaran yang dianggap sulit oleh para siswa [4], penyampaian materi yang menjemukan sehingga siswa tidak merasa tertarik dalam menyimak dan mempelajarinya, selain itu implikasi materi dengan kehidupan siswa yang tidak dijelaskan membuat para siswa tidak dapat memahami secara utuh akan dampak pembelajaran yang dilakukan dengan kelangsungan hidupnya [5]. Sehingga dengan mengikuti kegiatan les tambahan dapat melatih siswa akan tanggung jawab serta mengatur waktu dalam melakukan pembelajaran. Beberapa penelitian yang sudah pernah dilakukan mengenai klasifikasi peserta didik dengan penerapan K-Means *Clustering* diantaranya adalah penelitian yang dilakukan oleh Adella Putri Riani et al., 2023 berjudul “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow” dengan Hasil dari penelitian ini adalah mengelompokkan peserta didik menjadi 3 *cluster* yaitu *cluster* 0 dengan nilai rata-rata cenderung rendah yang memiliki jumlah anggota cluster yaitu 57 peserta didik. Kemudian *cluster* 2 yaitu kelompok rata-rata nilai sedang dengan jumlah anggota klaster yaitu 42 peserta didik. Lalu *cluster* 1 merupakan kelompok nilai dengan rata-rata tinggi yang hanya berjumlah 1 peserta didik pada PKBM Sanggar Puri [1].

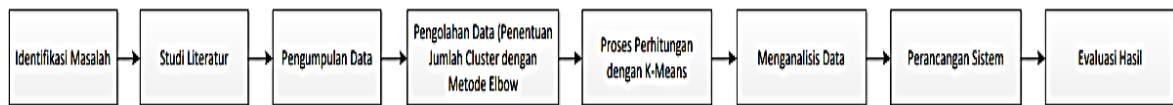
Penelitian ini akan memberikan informasi berharga bagi pihak sekolah maupun pihak terkait dalam memahami perbedaan signifikan yang dapat ditimbulkan untuk para siswa dari mengikuti kegiatan bimbingan belajar atau les tambahan, baik di lingkungan rumah maupun sekolah secara sistematis dengan menggunakan teknik klasterisasi K-Means *Clustering* dibantu dengan menggunakan metode elbow pada objek siswa peserta didik di Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan dengan menggunakan data sebanyak 514 siswa yang terdiri dari kelas 1A s.d 6B peneliti mengambil beberapa tolak ukur untuk menentukan jumlah *cluster* terbaik dalam penelitian ini seperti, menggunakan faktor usia dari para siswa lalu waktu belajar di rumah dan di sekolah, kemudian dukungan dari para orang tua di rumah selanjutnya peneliti juga mengambil data berupa interaksi sosial siswa baik di rumah dan dengan teman sebaya. Kemudian peneliti juga melihat kondisi kesehatan siswa, nilai kehadiran, dan yang paling penting perolehan rata-rata nilai para siswa yang didapat dari semester ganjil dan genap, nilai para siswa dalam mengikuti kegiatan ekstrakurikuler, serta nilai yang akan dijadikan perbandingan yaitu nilai siswa yang mengikuti les tambahan dengan siswa yang tidak mengikuti les tambahan. Proses yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok tertentu sehingga setiap kelompok berisi data dengan kemiripan paling besar sering disebut *clustering*. Sedangkan K-Means merupakan metode yang banyak digunakan untuk mengidentifikasi sekumpulan kasus [6] K-Means sendiri mengacu pada proses pengelompokan data berulang yang dapat melakukan pembatasan tertentu pada data untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek. Proses penyelesaian K-Means merupakan salah satu metode data *mining* yang banyak digunakan untuk mengidentifikasi kelompok kasus alami yang dilihat berdasarkan pengelompokan data yang mempunyai kemiripan, sehingga hasil pengelompokan tersebut dapat dianalisis [7].

Sehingga dengan mengacu pada beberapa penjelasan serta landasan penelitian terdahulu maka algoritma K-Means dipilih dan diimplementasikan dengan bantuan menggunakan metode elbow untuk menyelesaikan permasalahan yang ada. Diharapkan dari hasil penelitian ini dapat memberikan informasi dan membantu pihak sekolah khususnya guru agar dapat mengerti dengan kondisi para siswa serta melihat seberapa pengaruhnya les tambahan bagi para siswa dalam peningkatan kualitas belajar terhadap pola kognitif mereka. Selain itu akan dikaji kembali dari hasil penelitian ini untuk menentukan kategori siswa yang masuk ke dalam pengelompokan beberapa cluster yaitu (unggulan, baik, dan cukup baik) pada sekolah Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan.

2. METODOLOGI PENELITIAN

Pada penelitian ini kegiatan pengumpulan data dilakukan pada sekolah MIS Al-Falah Medan. Di mana data yang digunakan oleh peneliti sebanyak 514 Siswa/i yang dihitung mulai dari kelas 1A s.d 6B. Proses kegiatan penelitian ini dilakukan dengan menggunakan data numerik seperti nilai siswa, pengaruh les terhadap prestasi siswa yang dapat diukur dan dihitung. Variabel-variabel dalam penelitian ini diukur secara kuantitatif. Misalnya, prestasi siswa dapat diukur dengan perolehan hasil nilai ujian pada rapor mereka, sedangkan pengaruh les

tambahan dapat diukur dengan membandingkan nilai siswa sebelum dan sesudah mengikuti les. Tujuannya untuk mengklasifikasikan para siswa berdasarkan pengaruh les terhadap prestasi mereka. Data tersebut dikumpulkan oleh peneliti dengan melakukan koordinasi terlebih dahulu dengan penanggung jawab bidang kurikulum dan kesiswaan pada sekolah MIS Al-Falah Medan Ibu Titin Hartini, S.Pd. Sehingga, berdasarkan penjelasan diatas penelitian ini masuk sebagai penelitian kuantitatif dengan menggunakan metode-metode statistik yang digambarkan ke dalam kerangka kerja penelitian untuk penyelesaian masalah sebagai berikut.



Gambar 1. Bagan alur proses penelitian di Sekolah MIS Al-Falah Medan

Berdasarkan gambar 1 dari kerangka kerja penelitian tersebut maka dapat diuraikan pembahasan dalam masing-masing tahapan pada penelitian ini yaitu:

2.1 Identifikasi Masalah

Pada tahapan ini peneliti melakukan proses identifikasi masalah terhadap pengaruh program les privat atau bimbingan belajar yang berdampak pada peningkatan prestasi belajar siswa. Program les yang ada saat ini sering kali bersifat umum dan tidak disesuaikan dengan kebutuhan individu siswa. Tentunya karena setiap siswa memiliki karakteristik, kemampuan, dan gaya belajar yang berbeda-beda. Evaluasi terhadap kegiatan les seringkali hanya berfokus pada peningkatan nilai rata-rata kelas, tanpa mempertimbangkan perbedaan individu. Sehingga dengan mengidentifikasi masalah secara komprehensif terhadap para siswa di sekolah MIS Al-Falah Medan, penelitian ini diharapkan dapat memberikan solusi yang lebih tepat dalam meningkatkan kualitas pembelajaran siswa di MIS Al-Falah Medan.

2.2 Studi Literatur

Pada tahap ini peneliti akan melakukan pencarian untuk menyusun landasan teori yang diperoleh dari berbagai macam buku baik yang ada di perpustakaan maupun di internet serta sumber rujukan berupa artikel jurnal. Tujuannya untuk membantu peneliti dalam melakukan penerapan dan mempelajari tutorial tentang K-Means clustering dengan bantuan metode elbow dan juga membantu peneliti dalam menyusun landasan teori yang baik mengenai penelitian yang dilakukan.

2.3 Pengumpulan Data

Sebagai media bahan pendukung penelitian yang sangat berguna bagi peneliti untuk mencari dan mengumpulkan data yang diperlukan untuk penelitian ini, maka peneliti menggunakan dua cara pengumpulan data yaitu;

2.3.1 Pengamatan (Observation)

Metode ini dilakukan dengan cara mengamati secara langsung mengenai objek sasaran penelitian dengan cara mengumpulkan data konkrit yang diperlukan dalam pengerjaan penelitian. Sehingga untuk mendapatkan data-data tersebut, peneliti melakukan pengamatan secara langsung pada sekolah Madrasah Ibtidaiyah Swasta (MIS) Al-Falah Medan.

2.3.2 Wawancara (Interview)

Pada metode ini peneliti melakukan kegiatan wawancara atau bertanya langsung kepada pihak sekolah yaitu guru-guru maupun bagian akademik kesiswaan mengenai permasalahan yang akan diteliti

2.4 Pengolahan Data

Pada tahap ini peneliti akan melakukan pengolahan terhadap data yang telah dikumpulkan dengan proses penentuan jumlah *cluster* dengan bantuan metode elbow. Dimana tujuan dari metode elbow sendiri adalah untuk memilih berupa nilai (k) yang kecil dan masih memiliki nilai *whitins* yang rendah terhadap perolehan data siswa untuk mengukur tingkat pengaruh les tambahan kepada siswa di sekolah MIS Al-Falah Medan.

2.5 Proses Perhitungan dengan K-Means

Setelah memperoleh hasil dari perhitungan menggunakan metode elbow, maka langkah selanjutnya pada tahap ini peneliti melakukan perhitungan *clustering* dengan algoritma K-Means. Dimana langkah dalam melakukan perhitungan ini di antaranya adalah dengan memosisikan setiap data objek ke dalam *cluster* terdekat yang ditentukan berdasarkan jarak kedua objek tersebut. Sehingga diharapkan dengan menggunakan metode K-Means *clustering* dengan dibantu metode elbow diharapkan dapat membantu dalam mengolah data agar memperoleh hasil pengklasifikasi yang baik [8]. Adapun di bawah ini merupakan uraian dari algoritma menggunakan metode K-Means clustering.

1. Tentukan sebanyak nilai (k) kelompok yang kita inginkan.

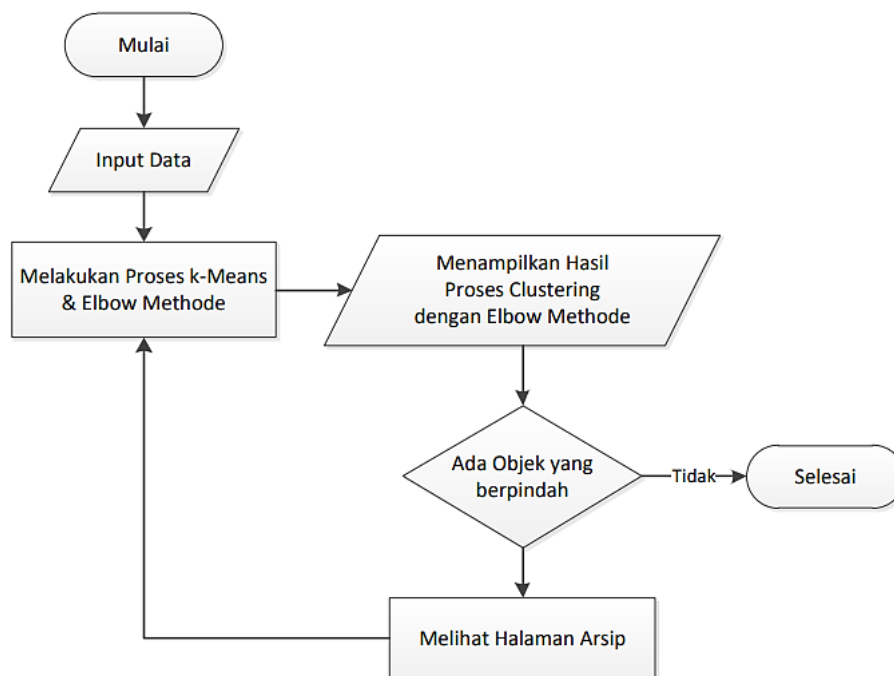
2. Kemudian inisiasi centroid untuk setiap kelompok. Centroid ibarat seperti “*leader group*”, yang mempresentasikan kelompok.
3. Hitung kedekatan antara suatu data dengan menggunakan centroid, selanjutnya masukkan data tersebut pada kelompok centroid yang memiliki sifat terdekat dengan dirinya.
4. Pilih centroid baru untuk masing-masing kelompok dari anggota kelompok itu sendiri dan lakukan perhitungan kedekatan antara suatu data dengan centroid.
5. Ulangi langkah-langkah sebelumnya hingga tidak ada perubahan anggota untuk semua kelompok [9].

2.6 Menganalisis Data

Pada tahapan ini peneliti akan menganalisis hasil yang telah diperoleh untuk melihat pencapaian dari tahapan-tahapan sebelumnya dan pada tahap ini peneliti juga melakukan serangkaian menggunakan software yang telah dibangun oleh peneliti untuk melihat pengaruh dan kesesuaian hasil terhadap proses perhitungan K-Means menggunakan cara manual yang sebelumnya telah dilakukan oleh peneliti menggunakan Microsoft Excel.

2.7 Perancangan Sistem

Dalam proses penelitian ini, dilakukan juga tahapan yaitu membangun sebuah aplikasi berupa tampilan *interface* berbasis *website*. Dalam pengembangan model *clustering* di Bahasa pemrograman Python, peneliti akan melakukan beberapa pemodelan dengan nilai konfigurasi yang berbeda, nilai parameter konfigurasi yang akan diuji pada pemodelan diantaranya jumlah iterasi, nilai sentroid dan jumlah klaster. Dari pengujian tersebut akan dipilih model dengan nilai akurasi terbaik yang akan diimplementasikan pada sistem informasi berbasis *website* [10]. Dimana nantinya digunakan untuk melakukan proses klasterisasi serta membandingkan dari hasil proses sebelumnya yang telah dilakukan secara manual. Sehingga nantinya akan ada perbandingan antara hasil proses perhitungan K-Means yang Microsoft Excel dan Google *Collaboratory*, serta dengan menggunakan hasil perancangan aplikasi K-Means berbasis web yang didukung dengan menggunakan bahasa pemrograman Python [11].



Gambar 2. Flowchart Proses Penelitian Sistem

Pada gambar 2 tersebut adapun proses sistem dimulai dengan menginput data dimana data yang akan di *clustering* dimasukkan ke dalam sistem. Data ini bisa berupa data numerik atau data kategori yang sudah diubah menjadi bentuk numerik., dilanjutkan dengan penerapan algoritma k-means dan metode elbow untuk menentukan jumlah *cluster* yang optimal, dan diakhiri dengan visualisasi hasil *clustering*. Nantinya sistem akan mengecek apakah ada data yang berpindah *cluster* setelah satu iterasi. Jika ada, maka proses k-means akan diulang dari langkah 3. Jika tidak ada yang berpindah, maka proses dianggap sudah konvergen. Iterasi terus dilakukan hingga tidak ada lagi perubahan signifikan dalam anggota *cluster*.

2.8 Evaluasi Hasil

Kegiatan akhir yang dilakukan dalam penelitian ini adalah berupa evaluasi hasil yang bertujuan untuk memberikan keputusan akhir atau menarik hasil akhir dari proses tahapan penelitian dengan garis besar.

Tujuannya untuk memperoleh hasil klasifikasi terhadap pengaruh siswa dalam mengikuti kegiatan les tambahan yang diberikan baik di lingkungan sekolah maupun di lingkungan rumah terhadap hasil nilai pada ujian mereka.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Adapun sumber data yang diperoleh dalam penelitian ini dilakukan dengan metode pengumpulan data pada sekolah MIS Al-Falah Medan yaitu:

- Sumber data primer diperoleh dari nilai rapor siswa di sekolah MIS Al-Falah Medan dari mulai kelas 1A s.d 6B dengan perhitungan periode pada Semester Ganjil dan Genap T.A 2023/2024. Di mana data yang diperoleh sebanyak 514 orang siswa/i.
- Menyebarkan kuesioner sebagai perolehan informasi tambahan dan juga bahan penelitian untuk mencari tolak ukur terhadap data berupa nilai yang diperlukan dari para siswa/i di sekolah yaitu, usia para siswa, waktu belajar di rumah, waktu belajar di sekolah, dukungan orang tua, interaksi sosial dengan teman sebaya, interaksi sosial dengan teman di rumah, ketersediaan fasilitas belajar di rumah, suasana belajar di kelas, kondisi kesehatan siswa, nilai kehadiran, nilai dalam mengikuti ekstrakurikuler di sekolah, nilai dalam mengikuti kegiatan les tambahan baik di rumah dan di sekolah, nilai untuk siswa yang tidak mengikuti les tambahan, serta nilai kepribadian dan akhlak.
- Pengumpulan data berupa kegiatan dokumentasi yang berhubungan dengan penelitian ini, misalnya berupa foto maupun data pendukung lainnya sebagai pendukung dalam melakukan analisis terhadap proses pembelajaran serta melihat interaktif pola keaktifan dari para siswa di sekolah MIS Al-Falah Medan [12].

3.2 Pemrosesan Data dan Transformasi Data

Pada tahapan ini ketika selesai melakukan pengumpulan data dan menyeleksi untuk digunakan sebagai data penelitian, maka selanjutnya dilakukan pemrosesan data dan transformasi data. Dalam melakukan pemrosesan data dapat dilakukan dengan melakukan *cleaning* dari data yang bersifat *missing value* serta dilakukan juga proses integrasi data. Namun, dikarenakan data yang diperoleh sudah lulus ke dalam tahap *cleaning* dan integrasi, maka selanjutnya dilakukan proses transformasi data. Proses ini merupakan penggabungan antara data yang sudah sesuai dengan data yang digunakan dalam proses data *mining* nantinya [13]. Dalam proses transformasi dalam KDD (*Knowledge Discovery in Database*) bersifat kreatif dan sangat bergantung pada jenis data atau pola informasi yang ingin diambil dalam database. Meskipun demikian, pada tahap ini penulis memutuskan untuk tidak melakukan transformasi data karena data yang diperoleh dari sekolah MIS Al-Falah Medan sudah sesuai dan siap untuk dilakukan proses data *mining* [14].

Dapat dilihat pada Tabel 1 berikut merupakan data yang sudah ditransformasikan dan akan dilakukan proses data *mining*.

Tabel 1. Dataset Penilaian Siswa (a)

No	Kelas	Nama Siswa	Uraian	Waktu Belajar di Rumah	Waktu Belajar di Sekolah	Dukungan Orang Tua	Interaksi Sosial di Rumah	Interaksi Sosial dengan Teman	Ketersediaan Fasilitas Belajar di Rumah	Suasana Belajar di Kelas	Kondisi Kesehatan Siswa	Nilai Kehadiran
1	1A	Alexa Putri Suroyo	7	20	6	Baik	4	4	Baik	3	Baik	93
2	1A	Arsya Nugraha	7	20	6	Baik	4	4	Baik	3	Baik	89
3	1A	Alya Fathyah Raihana Albizar	7	20	6	Baik	4	4	Cukup Baik	3	Baik	84
4	1A	Ramhmad Siregar	7	20	6	Baik	4	4	Baik	3	Baik	90
5	1A	Alisha Kurniawan Azzizan	7	20	6	Baik	4	4	Cukup Baik	3	Baik	85
6	1A	Azka Taswar	6	20	6	Baik	4	4	Baik	3	Baik	83
7	1A	Azizah Putri	7	20	6	Baik	4	4	Cukup Baik	3	Baik	88

No	Kelas	Nama Siswa	Usia	Waktu Belajar di Rumah	Waktu Belajar di Sekolah	Dukungan Orang Tua	Interaksi Sosial di Rumah	Interaksi Sosial dengan Teman	Ketersediaan Fasilitas Belajar di Rumah	Suasana Belajar di Kelas	Kondisi Kesehatan Siswa	Nilai Kehadiran
8	1A	Baim Putra Zunaidi	7	20	6	Baik	4	4	Cukup Baik	3	Baik	88
9	1A	Bahri Hernandes Chaniya	7	20	6	Baik	4	4	Cukup Baik	3	Baik	89
10	1A	Alshad Pratiwi	7	20	6	Baik	4	4	Cukup Baik	3	Baik	91
...
331	4B	Alvaro Azka Maiza	10	50	6	Baik	4	4	Baik	3	Baik	85
332	4B	Alya Nayla	10	60	6	Sangat Baik	4	4	Baik	3	Baik	81
333	4B	Azmi Dwiansyah Lubis	10	35	6	Baik	4	4	Cukup Baik	3	Baik	90
334	4B	Dimas Prayogi	10	55	6	Baik	4	4	Baik	3	Baik	86
335	4B	Dzakir Khafadi	10	45	6	Baik	4	4	Baik	3	Baik	77
...
510	6B	Willy Ashadi P R Daulay	12	70	7	Sangat Baik	4	4	Baik	3	Baik	83
511	6B	Yasmin Naura Izzati	12	75	7	Sangat Baik	4	4	Baik	3	Baik	77
512	6B	Yashvin Rafiq Zaidan	12	65	7	Sangat Baik	4	4	Cukup Baik	3	Baik	79
513	6B	Zain Ahmad Satria	12	60	7	Baik	4	4	Baik	3	Baik	93
514	6B	Zakiyah Putri	12	90	7	Sangat Baik	4	4	Baik	3	Baik	89

Tabel 1. Dataset Penilaian Siswa (b)

No	Kelas	Nama Siswa	Usia	Rata-rata Nilai Semester Genap	Rata-rata Nilai Semester Ganjil	Mengikuti Ekskul	Mengikuti Les	Nilai Kepribadian dan Akhlak	Hasil Rata-rata Nilai Ujian	Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les
1	1A	Alexa Putri Suroyo	7	80	83	90	80	100	82	85
2	1A	Arsya Nugraha Alya	7	79	78	80	80	100	79	80
3	1A	Fathyah Raihana	7	88	86	90	90	100	87	90
4	1A	Albizar Ramhma	7	83	84	90	80	100	84	85

No	Kelas	Nama Siswa	Usia	Rata-rata Nilai Semester Genap	Rata-rata Nilai Semester Ganjil	Mengikuti Ekskul	Mengikuti Les	Nilai Kepribadian dan Akhlak	Hasil Rata-rata Nilai Ujian	Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les
5	1A	d Siregar Alisha Kurniawan	7	87	84	80	80	100	86	80
6	1A	Azzizan Azka	6	81	83	90	80	100	82	85
7	1A	Taswar Azizah Putri	7	79	80	80	80	100	80	80
8	1A	Baim Putra	7	85	84	80	95	100	85	88
9	1A	Zunaidi Bahri	7	83	83	80	80	100	83	80
10	1A	Hernandes Chaniya Alshad Pratiwi	7	92	93	80	90	100	93	85
...
331	4B	Alvaro Azka Maiza	10	86	92	90	90	100	89	90
332	4B	Alya Nayla Azmi	10	82	93	90	95	100	88	93
333	4B	Dwiansyah Lubis	10	93	93	90	80	100	93	85
334	4B	Dimas Prayogi	10	97	95	90	95	100	96	93
335	4B	Dzakir Khafadi	10	92	90	90	80	100	91	85
...
510	6B	Willy Ashadi P R Daulay	12	87	90	100	95	100	89	98
511	6B	Yasmin Naura Izzati	12	96	96	100	95	100	96	98
512	6B	Yashvin Rafiq Zaidan	12	78	81	100	95	100	80	98
513	6B	Zain Ahmad Satria	12	83	87	90	95	100	85	93
514	6B	Zakiah Putri	12	85	85	90	95	100	85	93

Pada Tabel 1 (a) dan 1 (b) tersebut merupakan satu dataset yang memiliki 17 variabel dan 514 data yang nantinya akan diolah menggunakan algoritma K-Means *clustering* dibantu dengan menggunakan metode elbow dalam mengelompokkan kategori siswa terhadap pengaruh mengikuti bimbingan belajar tambahan terhadap nilai dan persentase hasil terbaik mereka [15].

3.3 Metode K-Means Clustering

Setelah melakukan proses pembersihan data maka langkah selanjutnya data tersebut diproses dengan melakukan tahapan data *mining* menggunakan algoritma K-Means clustering. Dimana data tersebut akan dikelompokkan dengan nilai para siswa dari beberapa kelompok kedalam satuan bentuk *cluster*. Hal tersebut dapat dilakukan dengan menganalisis terhadap pola pembentukan *cluster* yang nantinya hasil tersebut dapat menetapkan nilai *centroid* secara acak. Kemudian setelah melakukan tahap penentuan nilai *centroid* maka dilakukanlah proses mengestimasi jarak pada data ke setiap *centroid* dengan menggunakan metode korelasi antara dua buah objek yang menggunakan rumus metode *Euclidean Distance* pada Pers (1) berikut [16].

$$d(i, j) = \sqrt{(x1i - x1j)^2 + (x2i - x2j)^2 + (xki - xkj)^2} \tag{1}$$

Dimana $d(i, j)$ merupakan jarak data ke i pada pusat *cluster* j , dengan X_{ki} adalah data ke i pada atribut data pada k dan X_{kj} adalah titik pusat ke j pada atribut ke k . Sehingga dalam proses menentukan nilai pada metode *K-Means Clustering* terdapat beberapa tahapan seperti yang telah dijabarkan pada metode penelitian yaitu:

1. Menentukan (k) kelompok :

Dalam menentukan nilai pengujian K dari penentuan nilai K *cluster* awal pada penelitian ini yaitu $k = 4$

2. Inisiasi nilai *centroid*

Dalam hal ini inisiasi nilai *centroid* awal pada setiap *cluster* ditentukan secara acak, dimana Nilai X merupakan Hasil penjumlahan nilai rata-rata pada ujian semester ganjil dan genap sedangkan Nilai Y merupakan Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les yang terdapat didalam dataset. seperti pada Tabel 1 berikut;

Tabel 2. Inisiasi *Centroid*

Data Ke	Centroid	nRata-rata Ujian Sem Ganjil & Genap	nRata-rata Mengikuti Ekskul & Les
2	0	79	80
100	1	90	93
431	2	93	85
512	3	80	98

Pada Tabel 2 merupakan penentuan nilai *centroid* secara acak yang diambil dari data ke 2, 100, 431, dan 512. Dimana dalam menentukan nilai *centroid* ini peneliti menggunakan dua variabel yang sebelumnya telah dijumlahkan yaitu Rata-rata nilai ujian pada semester ganjil dan genap serta Rata-rata nilai siswa dalam mengikuti kegiatan ekskul dan les.

3. Menghitung jarak setiap titik data dengan setiap *centroid* dengan *Euclidean Distance* menggunakan rumus yang tertera. Pada Tabel 3 merupakan contoh hasil perhitungan sampai data ke-3

Tabel 3. Jarak Pusat *Cluster* Dengan *Euclidean Distance*

	D(1,1)	D(1,2)	D(1,3)	D(1,4)
	5,5902	11,6726	11,5000	13,0863
	0,5000	17,3566	15,3379	18,0624
	12,8062	4,2426	7,8102	10,6301
	6,7268	10,3078	9,5000	13,4629

Pada Tabel 3 merupakan hasil representasi nilai jarak *Euclidean* antara dua pusat *cluster*. Misalnya, nilai 5,5902 pada baris pertama kolom pertama (D(1,1)) menunjukkan jarak antara pusat cluster 1 dengan dirinya sendiri. Selanjutnya dalam proses perhitungan dimulai dari data ke-2 sampai data ke-514 dengan menggunakan rumus perhitungan pada Microsoft Excel sehingga diperoleh hasil seperti pada Tabel 4 berikut.

Tabel 4. Hasil 3 Data Pertama C0 Iterasi Ke-1

Data Ke	(X)	(Y)	C0	C1	C2	C3
1	82	85	5,5902	11,6726	11,5000	13,0863
2	79	80	0,5000	17,3566	15,3379	18,0624
4	84	85	6,7268	10,3078	9,5000	13,4629

Berdasarkan hasil yang masuk kedalam kelompok *cluster* 0 yang diperoleh pada perhitungan jumlah data saat iterasi ke-1 terdapat sebanyak 120 data untuk jumlah keseluruhan *cluster* 0. Kemudian selanjutnya proses perhitungan dari jumlah 3 data pertama pada kelompok *cluster* 1 yang dilakukan pada perhitungan iterasi ke-1 terdapat sebanyak 174 data untuk keseluruhan jumlah pada kelompok *cluster* 1. Seperti pada Tabel 5 di bawah.

Tabel 5. Hasil 3 Data Pertama C1 Iterasi Ke-1

Data Ke	(X)	(Y)	C0	C1	C2	C3
3	87	90	12,8062	4,2426	7,8102	10,6301
8	85	88	9,3005	7,7782	8,86	11,4237
12	85	90	11,4127	6,265	9,8615	9,1788

Selanjutnya proses perhitungan iterasi ke-1 yang dilakukan dalam memperoleh tiga data pertama untuk menentukan kelompok *cluster 2* menghasilkan sebanyak 190 data untuk jumlah keseluruhan yang masuk pada *cluster 2*. Seperti pada Tabel 6 berikut.

Tabel 6. Hasil 3 Data Pertama C2 Iterasi Ke-1

Data Ke	(X)	(Y)	C0	C1	C2	C3
10	93	85	14,3962	8,3815	0,5	18,0347
26	87	85	9,0139	8,7321	6,5	14,5344
30	89	88	12,5	5,5902	4,717	13,8293

Kemudian pada proses perhitungan iterasi ke-1 yang dilakukan terhadap data untuk mengelompokkan *cluster 3* diperoleh sebanyak 30 data yang masuk kedalam kategori kelompok *cluster 3*. Dapat dilihat seperti pada Tabel 7 berikut.

Tabel 7. Hasil 3 Data Pertama C3 Iterasi Ke-1

Data Ke	(X)	(Y)	C0	C1	C2	C3
19	83	93	13,1244	7,0178	12,5000	6,2650
86	88	88	11,3358	6,0415	6,0415	12,9035
224	88	88	11,3358	6,0415	6,0415	12,9035

Sehingga setelah melakukan proses perhitungan iterasi pertama, maka selanjutnya adalah mencari kelompok *cluster* baru dari hasil proses iterasi ke-1.

4. Menentukan kelompok *cluster* baru dari hasil iterasi ke-1

Dalam menentukan *cluster* baru dapat dilakukan dengan menjumlahkan data kemudian dirata-ratakan oleh jumlah seluruh data yang terdapat pada setiap *cluster* yaitu, *cluster 0*, *cluster 1*, *cluster 2*, dan *cluster 3*. Sehingga hasil perolehannya seperti pada Tabel 7 berikut.

Tabel 8. Penentuan Hasil Nilai *Centroid* Baru

Centroid	(X)	(Y)
0	82	81
1	89	92
2	89	84
3	83	95

Sehingga apabila sudah dihasilkan nilai *centroid* baru, maka setelah itu akan dilakukan proses perhitungan kembali untuk melanjutkan tahapan proses iterasi ke-2. Jika iterasi ke-2 sudah dilakukan, maka akan dilakukan kembali proses perhitungan ulang dengan cara melakukan pengelompokan *cluster* dari iterasi ke-2 dikurangi dengan kelompok *cluster* pada hasil perhitungan iterasi ke-1. Proses iterasi ini akan terus dilakukan berulang kali hingga seluruh nilai yang terdapat di dalam data sama dengan nol.

Maksudnya dalam artian semua data telah masuk ke dalam kelompoknya masing-masing dan tidak ada lagi perubahan pada kelompok data yang ada. Sehingga dari hasil proses iterasi yang telah dilakukan diatas, maka proses perhitungan iterasi dilakukan proses iterasi sebanyak 7 kali untuk memperoleh hasil data sama dengan nol. Hasil akhir pada proses perhitungan *K-Means Clustering* dapat dilihat pada Tabel 9 berikut.

Tabel 9. Hasil Perhitungan *K-Means Clustering* Pada Iterasi Ke-8 terhadap Pengaruh Bimbel dan Tingkat Prestasi Siswa di Sekolah

No	D(8,1)	D(8,2)	D(8,3)	D(8,4)	Cluster
1	3,5707	12,0298	7,4101	9,169	0
2	5,2923	17,4799	11,4564	14,9555	0
3	8,7865	4,8075	5,5484	4,0759	3
4	3,0199	10,4259	5,4112	8,6245	0
5	2,8819	13,1335	5,8778	13,5189	0
6	3,3305	11,6173	6,9103	8,9944	0
7	4,3811	16,7493	10,5562	14,5542	0
8	5,6256	8,157	5,1739	6,0213	2

No	D(8,1)	D(8,2)	D(8,3)	D(8,4)	Cluster
9	2,0229	14,4485	7,6049	13,6426	0
10	9,5805	6,7545	3,5991	11,3923	2
...
510	16,3328	6,549	12,7169	5,3619	3
511	19,9849	7,3537	14,5555	11,7758	1
512	16,0029	13,3404	15,8129	6,734	3
513	10,6384	6,5556	8,6438	1,0095	3
514	10,6384	6,5556	8,6438	1,0095	3

Pada Tabel 9 merupakan hasil akhir proses iterasi ke-8 dengan jumlah data setiap cluster yaitu, pada cluster 0 terdapat sebanyak 135 data yang masuk kedalam cluster ini, kemudian pada kelompok cluster 1 terdapat sebanyak 119 data lalu pada kelompok cluster 2 terdapat sebanyak 183 data dan kelompok data yang masuk kedalam cluster 3 sebanyak 77 data. Sehingga total data yang dikelompokkan kedalam empat cluster sebanyak 514 data, dimana jumlah perolehan hasil tersebut berbeda dengan proses perhitungan pada iterasi ke-1 yaitu, data yang masuk kedalam cluster 0 sebanyak 120 data, kemudian untuk kelompok cluster 1 sebanyak 174 data lalu cluster 2 terdapat sebanyak 190 data dan cluster 3 terdapat sebanyak 30 data untuk jumlah keseluruhan data yang sama.

3.4 Implementasi Hasil K-Means Clustering

Pada proses ini data dari hasil proses K-Means yang sebelumnya telah dilakukan secara manual menggunakan Microsoft Excel kemudian diaplikasikan menggunakan program aplikasi Google Collaboratory, dalam hal implementasi ini juga peneliti membangun sebuah aplikasi berupa tampilan interface berbasis website yang nantinya dibuat bertujuan untuk memproses dan mengolah dataset kemudian menampilkannya dalam bentuk sebuah grafik dari hasil akhir proses iterasi yang terdapat didalam web tersebut. Hal ini dilakukan untuk melihat keakurasian data dengan membandingkan proses yang sebelumnya telah dilakukan. Adapun langkah awal dalam melakukan proses K-Means clustering menggunakan Google Collaboratory yaitu, mengimport terlebih dahulu library seperti yang terlihat pada Gambar 3 berikut.

Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import pylab as pl
import seaborn as sns
from scipy.spatial import distance
from sklearn.cluster import KMeans
from scipy.spatial.distance import euclidean
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import silhouette_score
```

Gambar 3. Perintah Import Library pada Google Collab

Kemudian langkah selanjutnya melakukan import data siswa ke dalam Dataset Google Collaboratory dengan format .csv dapat dilihat seperti pada gambar 4 berikut.

```
data = pd.read_csv('datasetMIS.csv', index_col=0)
data.head()
```

No	Kelas	Nama Siswa	Usia	Waktu Belajar di Rumah	Waktu Belajar di Sekolah	Dukungan Orang Tua	Interaksi Sosial di Rumah	Interaksi Sosial dengan Teman Sebayu	Ketersediaan fasilitas Belajar di Rumah	Suasana Belajar di Kelas	Kondisi Kesehatan Siswa	Nilai Kehadiran	Rata-rata nilai semester genap	Rata-rata nilai semester ganjil	Mengikuti Ekskul	Mengikuti Les	Nilai Kepribadian dan Akhlak	Hasil Rata-rata Nilai Ujian	Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les
1	1A	Alexa Putri Suroyo	7	20	6	Baik	4	4	Baik	3	Baik	93	80	83	90	80	100	82	85
2	1A	Arsya Nugraha	7	20	6	Baik	4	4	Baik	3	Baik	89	79	78	80	80	100	79	80
3	1A	Alya Fathiyah Raihana	7	20	6	Baik	4	4	Cukup Baik	3	Baik	84	88	86	90	90	100	87	90
4	1A	Albizar Ramhmad Siregar	7	20	6	Baik	4	4	Baik	3	Baik	90	83	84	90	80	100	84	85
5	1A	Alisha Kurniawan	7	20	6	Baik	4	4	Cukup Baik	3	Baik	85	87	84	80	80	100	86	80

Gambar 4. Perintah Import file .csv ke dalam Dataset Google Collab

Kemudian langkah selanjutnya, untuk melihat deskripsi pada dataset yang kita punya maka kita harus melakukan import perintah deskripsi data seperti pada Gambar 5 berikut.

data.describe()

	Usia	Waktu Belajar di Rumah	Waktu Belajar di Sekolah	Interaksi Sosial di Rumah	Interaksi Sosial dengan Teman Sebayanya	Suasana Belajar di Kelas	Nilai Kehadiran	Rata-rata nilai semester genap	Rata-rata nilai semester ganjil	Mengikuti Ekskul	Mengikuti Les	Nilai Kepribadian dan Akhlak	Hasil Rata-rata Nilai Ujian	Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les
count	514.000000	514.000000	514.000000	514.0	514.0	514.0	514.000000	514.000000	514.000000	514.000000	514.000000	514.0	514.000000	514.000000
mean	9.262646	45.560311	6.128405	4.0	4.0	3.0	83.081712	86.813220	88.120623	87.836576	86.070039	100.0	87.714008	87.128405
std	1.715395	13.017654	0.334866	0.0	0.0	0.0	5.436245	4.303784	3.851159	6.412748	7.065756	0.0	3.887542	5.247022
min	6.000000	20.000000	6.000000	4.0	4.0	3.0	68.000000	77.000000	78.000000	80.000000	80.000000	100.0	78.000000	80.000000
25%	8.000000	35.000000	6.000000	4.0	4.0	3.0	79.000000	83.250000	86.000000	80.000000	80.000000	100.0	85.000000	85.000000
50%	9.000000	45.000000	6.000000	4.0	4.0	3.0	83.000000	87.000000	88.000000	90.000000	80.000000	100.0	88.000000	85.000000
75%	11.000000	55.000000	6.000000	4.0	4.0	3.0	87.500000	90.000000	91.000000	90.000000	95.000000	100.0	90.000000	93.000000
max	13.000000	90.000000	7.000000	4.0	4.0	3.0	96.000000	97.000000	98.000000	100.000000	95.000000	100.0	97.000000	98.000000

Gambar 5. Tampilan Deskripsi pada Dataset

Dapat dilihat pada Gambar 5 tersebut bahwa, hasil deskripsi pada dataset yang dipunya terlihat bahwa rata-rata nilai perolehan ujian pada semester ganjil dan genap yaitu 87,71 sedangkan untuk hasil rata-rata pada penjumlahan nilai siswa dalam mengikuti kegiatan Bimbingan belajar dan ekstrakurikuler yaitu 87,12. Tentunya nilai tersebut tidak jauh berbeda dengan hasil yang didapat para siswa saat mengikuti ujian di sekolah. Selanjutnya, untuk melihat nilai cluster yang paling optimal terhadap dataset yang kita punya maka kita perlu melakukan proses yaitu klusterisasi menggunakan elbow method dengan WCSS [17]. Dapat dilihat pada Gambar 6 berikut.

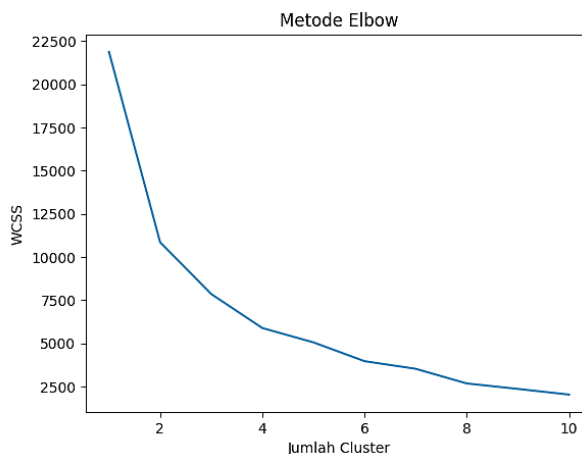
```
# Membaca dataset
dataset = pd.read_csv('datasetMIS.csv', index_col=0)
X = dataset.iloc[:, [17, 18]].values

# Metode Elbow untuk menentukan jumlah cluster yang optimal
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

# Plot Metode Elbow
plt.plot(range(1, 11), wcss)
plt.title('Metode Elbow')
plt.xlabel('Jumlah Cluster')
plt.ylabel('WCSS')
plt.show()
```

Gambar 6. Perintah untuk Menentukan Jumlah Cluster Optimal dengan Elbow Method

Sehingga setelah perintah tersebut dijalankan, maka kita tinggal menggambarkan proses metode elbow kedalam plot. Dari data yang diperoleh, jumlah cluster yang direkomendasikan adalah sebanyak 4 cluster. Berikut tampilan pada hasil menentukan jumlah cluster terbaik yang terdapat pada Gambar 7 dibawah.



Gambar 7. Hasil Penentuan Cluster Berdasarkan Metode Elbow

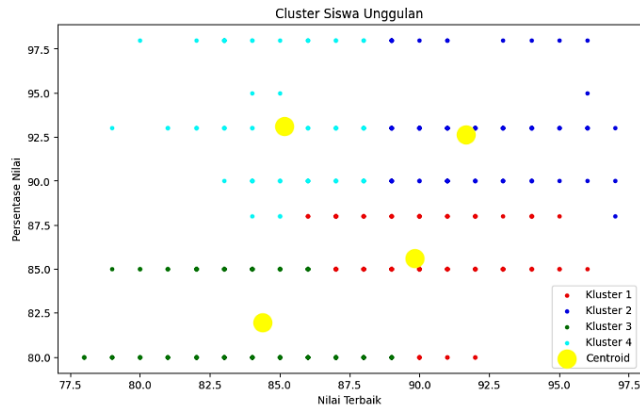
Langkah selanjutnya adalah, melakukan proses pengelompokan data pada kolom hasil penjumlahan rata-rata nilai ujian semester ganjil dan genap serta hasil penjumlahan siswa yang mengikuti les dan kegiatan ekstrakurikuler kedalam bentuk 4 cluster yang sebelumnya sudah dibentuk. Maka hasil dari perintah ini berupa bentuk bilangan array yang menampilkan angka kelompok cluster yang dibentuk. Bilangan array sendiri antara lain yaitu 0, 1, 2, dan 3. Kemudian setelah kita menjalankan hasil pengelompokan tersebut, maka selanjutnya

kita juga akan memvisualisasikan data tersebut kedalam sebuah gambar plot *scatter* seperti pada Gambar 8 serta Gambar 9 berikut [18].

```
# Memberikan cluster untuk setiap titik data
labels = kmeans.labels_
# Memberikan titik koordinat pada pusat cluster
centers = kmeans.cluster_centers_
print("Cluster labels:", labels)
print("Cluster centers:", centers)

Cluster labels: [2 2 3 2 2 2 2 3 2 0 2 3 2 3 3 3 2 2 3 0 0 1 2 2 2 0 2 2 1 0 0 1 0 3 3 0 1
2 2 0 2 2 2 2 0 2 0 2 1 2 2 1 1 2 2 3 2 3 0 2 1 1 2 3 2 3 0 0 1 2 2 0 2
1 1 1 1 2 0 1 2 0 3 1 0 1 1 1 0 2 0 2 2 0 2 1 0 2 1 0 2 1 1 0 1 2 0 0 2 0
2 2 0 3 1 2 0 2 1 2 0 0 0 2 1 2 2 0 3 1 0 2 1 3 2 0 2 2 1 0 2 2 1 2 1 1
2 0 0 0 1 0 2 0 0 0 3 0 2 1 0 2 0 1 0 2 1 3 2 2 0 0 3 2 0 0 1 2 1 0 0 1 1
2 0 2 0 1 0 0 2 1 3 2 1 2 3 0 0 1 2 1 2 1 0 1 0 2 1 2 0 0 0 2 0 0 0 2
0 0 1 2 0 2 2 2 0 0 1 2 1 2 2 2 2 1 2 0 0 0 1 0 0 3 2 2 3 1 2 2 2 2 0 2
0 1 0 0 2 2 2 0 2 2 1 2 1 0 2 0 0 2 2 0 1 2 2 1 2 2 2 0 2 2 0 2 1 2 2
2 0 1 0 3 0 2 2 0 2 3 2 1 0 0 2 3 3 1 1 2 3 3 3 1 0 1 1 2 1 3 3 3 2 1 3 0
1 0 1 0 1 3 3 1 3 1 0 2 0 3 2 1 0 0 0 3 1 0 2 1 2 3 3 3 3 0 2 1 2 0 2 0 0
2 0 1 2 2 0 2 0 2 1 1 2 0 2 0 2 0 1 2 3 1 2 0 0 2 2 0 0 0 0 0 2 0 0 2 2 2
0 1 2 2 0 0 2 0 1 2 2 2 0 2 2 0 0 0 1 0 0 1 0 2 0 0 2 0 0 1 2 2 0 0 0 0
1 0 2 0 0 1 1 1 1 1 1 3 1 1 0 0 0 0 0 1 0 0 3 3 3 3 3 3 3 3 3 1 1 1 1 3
3 3 1 3 3 3 1 3 3 3 3 3 3 3 3 1 3 3 1 1 3 3 3 3 3 1 3 1 1 3 3 3 3]
Cluster centers: [[89.83766234 85.58441558]
[91.66666667 92.64864865]
[84.38323353 81.94610778]
[85.15853659 93.1097561 ]]
```

Gambar 8. Hasil Perintah Mengelompokkan dan Memberikan Titik Koordinat Pada Pusat Cluster



Gambar 9. Hasil Visualisasi Data Terhadap Hasil Rata-rata Nilai Ujian dengan Pengaruh Les dan Ekskul

Sehingga pada Gambar 9 dapat kita lihat bahwa, kelompok nilai siswa dengan rata-rata tertinggi masuk kedalam kategori *cluster* ke-2 sedangkan siswa dengan rata-rata nilai terendah berada dalam kelompok *cluster* ke-3 dan ke-4. Setelah itu, langkah selanjutnya adalah membuat penambahan kolom baru pada tabel dataset yang berfungsi untuk menampung nilai hasil prediksi yang telah dilakukan pada proses sebelumnya. Dalam penelitian ini, kolom tambahan tersebut diberi nama “Level” yang bertujuan untuk memberikan status berupa keterangan pada dataset yang kita punya. Hasilnya terdapat seperti pada Gambar 10 berikut ini.

```
dataset['level'] = y_predict
dataset
```

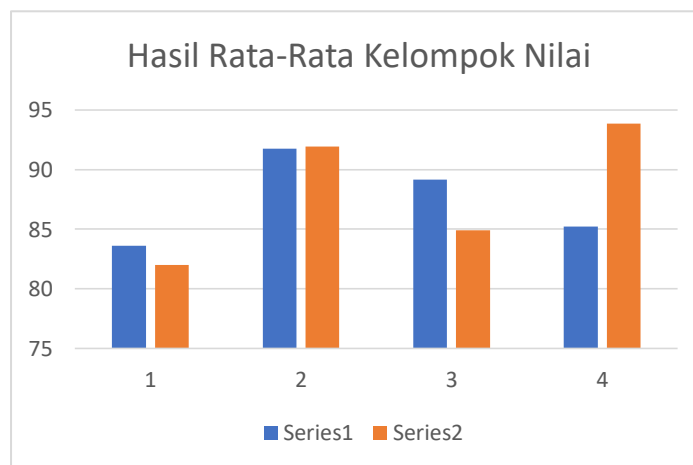
No	Kelas	Nama Siswa	Usia	Waktu Belajar di Rumah	Waktu Belajar di Sekolah	Dukungan Orang Tua	Interaksi Sosial di Rumah	Interaksi Sosial dengan Teman Sebaya	Ketersediaan Fasilitas Belajar di Rumah	Kondisi Kesehatan Siswa	Nilai Kehadiran	Rata-rata nilai semester genap	Rata-rata nilai semester ganjil	Mengikuti Ekskul	Mengikuti Les	Nilai Kepribadian dan Akhlak	Hasil Rata-rata Nilai Ujian	Hasil Rata-rata Penjumlahan Nilai Mengikuti Ekskul dan Les	Level
0	1	Alexa Putri Suryo	7	20	6	Baik	4	4	Baik	Baik	93	80	83	90	80	100	82	85	2
1	2	Arya Nugraha	7	20	6	Baik	4	4	Baik	Baik	89	79	78	80	80	100	79	80	1
2	3	Alya Fathyah Raihana	7	20	6	Baik	4	4	Cukup Baik	Baik	84	88	86	90	90	100	87	90	2
3	4	Albizar Ramhmad Siregar	7	20	6	Baik	4	4	Baik	Baik	90	83	84	90	80	100	84	85	2
4	5	Alisha Kurniawan	7	20	6	Baik	4	4	Cukup Baik	Baik	85	87	84	80	80	100	86	80	1
...
09	510	6B Willy Ashadi P R Daulay	12	70	7	Sangat Baik	4	4	Baik	Baik	83	87	90	100	95	100	89	98	3
10	511	6B Yasmin Naura Izzati	12	75	7	Sangat Baik	4	4	Baik	Baik	77	96	96	100	95	100	96	98	0

Gambar 10. Menambahkan Kolom Level untuk Menampung Keterangan Hasil Prediksi

3.5 Menganalisis Hasil Pengujian

Setelah melakukan proses dalam pengujian data menggunakan K-Means *Clustering*, maka proses selanjutnya peneliti melakukan analisis terhadap data yang telah berhasil diolah dengan menggunakan metode elbow. Maka dalam melakukan proses ini kita dapat menganalisis dan memahami pengaruh les tambahan dan kegiatan ekstrakurikuler terhadap nilai siswa secara lebih mendalam. Analisis ini tidak hanya memberikan wawasan tentang faktor-faktor yang mempengaruhi prestasi akademik, tetapi juga dapat membantu dalam pengambilan keputusan yang lebih baik dalam konteks pendidikan. Sehingga diperoleh bahwa siswa yang mengikuti Les tambahan dan kegiatan Ekstrakurikuler cenderung memperoleh nilai rata-rata yang tinggi. Dimana dalam bentuk persentase yang dilakukan oleh peneliti menemukan bahwa dari 514 orang siswa di sekolah MIS Al-Falah Medan, tingkat persentase tertinggi berada dalam nilai 94,05 sedangkan persentase nilai terendah berada pada nilai 78,8.

Dalam menghitung persentase nilai ini, peneliti menggunakan beberapa tolak ukur yang menjadi perhitungan yaitu, Nilai Rata-rata Ujian pada Semester Ganjil dan Genap dikali 30%, kemudian Nilai Kehadiran dikali 15%, kemudian ditambah Nilai apabila siswa tersebut mengikuti Les maka hasilnya akan dikali 15%, dan terakhir Nilai Kegiatan Ekstrakurikuler dikali 10%. Sehingga total persentasenya berjumlah 100%. Kemudian perlu diketahui juga bahwa dalam memberikan nilai siswa bagi yang mengikuti les, di sini peneliti berkoordinasi dengan pihak sekolah bahwasanya apabila siswa tersebut tidak mengikuti Les atau hanya belajar di rumah saja, maka tetap diberi nilai yaitu 80. Sedangkan apabila siswa tersebut mengikuti kegiatan Les di luar sekolah baik itu berupa *Private* Bimbingan Belajar maka akan diberi nilai 90, sedangkan untuk siswa yang mengikuti Les tambahan di Sekolah dan Keduanya maka akan diberi nilai 95. Tentunya hal ini bertujuan untuk memberikan nilai berupa angka agar dalam proses perhitungan K-Means *Clustering* dapat digunakan sebagai bentuk integer.



Gambar 11. Visualisasi Hasil Perhitungan Rata-rata Nilai Siswa

Dapat dilihat pada Gambar 11 bahwasannya, untuk bagian yang berwarna biru menunjukkan variabel (X) yaitu berupa Nilai Rata-rata siswa pada Ujian Semester Ganjil dan Genap. Sedangkan untuk bagian yang berwarna orange menunjukkan variabel (Y) yaitu berupa Hasil Penjumlahan Nilai mengikuti Les dan Ekskul. Sehingga dapat kita lihat bahwa siswa yang memiliki Rata-rata nilai ujian pada semester ganjil dan genap tertinggi masuk kedalam kelompok cluster ke-2 sedangkan untuk siswa yang lebih banyak dan cenderung aktif dalam mengikuti kegiatan Ekskul dan Les berada pada kelompok cluster ke-4. Tentunya hal ini sama dengan yang telah kita hasilkan pada proses sebelumnya dengan menggunakan visualisasi pada Google *Collaboratory*.

3.6 Evaluasi Hasil Pengujian

Untuk mengoptimalkan hasil perolehan nilai yang telah didapat pada proses sebelumnya. Maka dalam hal ini peneliti melakukan tahapan evaluasi hasil pengujian terhadap proses hasil berupa perbandingan data tolak ukur yang dihitung terhadap tingkat pengaruh les dengan nilai yang didapat siswa dalam mengikuti ujian di sekolah baik disaat semester ganjil maupun genap. Hal ini dilakukan untuk mengetahui hasil optimal dari pengelompokan data terhadap pengaruh siswa dalam mengikuti kegiatan Les dan Ekskul terhadap Nilai mereka pada saat ujian di sekolah agar menghasilkan data yang ideal. Dalam penelitian ini peneliti menggunakan *Davies Bouldin Index* (DBI) sebagai metode dalam mencari nilai yang paling optimal dalam menentukan klasterisasi data [19]. Nantinya nilai DBI yang diperoleh dalam model ini menggunakan algoritma K-Means *Clustering* yang bernilai 1,347 dan berikut merupakan perolehan hasilnya pada Gambar 12 dibawah.

Davies Bouldin

Davies Bouldin: 1.347

Gambar 12. Hasil Nilai DBI dengan Jumlah Cluster n = 4

Sehingga untuk mengetahui apakah *clustering* $n = 4$ merupakan jumlah *cluster* yang paling optimal maka dalam hal ini peneliti melakukan perbandingan terhadap hasil perhitungan DBI yang dimulai dari *cluster* ke-2 sampai *cluster* ke-5 seperti pada Tabel 10 berikut.

Tabel 10. Perhitungan Jumlah Cluster (C) dengan DBI

Percobaan	Jumlah <i>Cluster</i>	Hasil <i>Davies Bouldin Index</i> (DBI)
1	2	1,351
2	3	1,418
3	4	1,347
4	5	1,372

Sehingga berdasarkan hasil nilai DBI pada Tabel 9 diatas maka nilai yang paling kecil yang berada pada saat jumlah *cluster* $n = 4$. Karena semakin kecil nilai DBI dalam suatu kelompok maka jarak antar data akan semakin dekat. Sedangkan sebaliknya, apabila nilai DBI dalam suatu data semakin besar maka kemiripan antar data dalam suatu kelompok akan semakin berjauhan [20].

4. KESIMPULAN

Hasil dari perhitungan algoritma K-Means Clustering menggunakan metode elbow dapat menghasilkan pengelompokan data yang baik dan akurat. Dalam hal ini algoritma K-Means telah berhasil mengelompokkan data para siswa di sekolah MIS Al-Falah Medan berdasarkan jarak nilai antar data dengan *centroid*. Perolehan hasil yang didapatkan dari proses clustering ini menghasilkan nilai *cluster* yang optimal yaitu 4 *cluster*. Dimana dari proses klasterisasi yang dihitung menggunakan hasil penjumlahan nilai rata-rata ujian pada semester ganjil dan genap serta hasil penjumlahan rata-rata nilai siswa dalam mengikuti kegiatan ekstrakurikuler serta kegiatan Les baik di rumah dan di sekolah. Sehingga hasil yang diperoleh dalam penelitian ini ialah, dari total data sebanyak 514 orang ditemukan bahwa siswa yang masuk kedalam persentasi nilai terbaik yaitu 94,05 dimana persentasi tersebut masuk kedalam kelompok *cluster* ke-2. Sedangkan untuk persentasi nilai terendah yaitu 78,8 dan masuk kedalam kelompok *cluster* ke-1. Kemudian untuk rata-rata kelompok siswa yang masuk kedalam kategori siswa “Unggulan” berada kedalam kelompok *cluster* ke-2 dengan total siswa sebanyak 122 orang dimana perolehan rata-rata nilai ujian siswanya 91,75 dan hampir Sebagian siswa dalam kelompok ini mengikuti kegiatan Les dan Ekskul dengan baik. Sedangkan untuk siswa yang masuk kedalam kategori siswa “Baik” berada dalam kelompok *cluster* ke-3 dengan total 180 orang siswa dan perolehan rata-rata nilai ujian siswa pada kelompok ini yaitu 89,16 dan sebagian siswa juga mengikuti kegiatan Les serta Ekskul dengan baik. Lalu kelompok *cluster* ke-4 dimana total siswa pada *cluster* ini sebanyak 77 orang dengan rata-rata nilai ujian yaitu 85,19 dan pada kelompok ini hanya sebagian saja siswa yang mengikuti kegiatan Les dan Ekskul dan terakhir kelompok *cluster* ke-1 dengan total sebanyak 135 orang siswa yang nilai rata-rata ujiannya 83,62 dan pada kelompok ini menjadi kelompok paling kecil kategori siswa sedikit siswa yang mengikuti Les baik di sekolah maupun di rumah. Sehingga dari hasil pemodelan yang telah didapat dari proses perhitungan K-Means Clustering ini dapat difokuskan untuk melihat seberapa besar pengaruh dan perkembangan siswa dari yang tidak mengikuti Les maupun yang mengikuti Les secara keseluruhan agar pihak sekolah dapat melakukan pemerataan proses kegiatan belajar mengajar yang lebih baik. Kemudian yang terakhir, saran dalam melakukan penelitian selanjutnya ialah dengan menambahkan lebih banyak lagi atribut variable serta menggunakan algoritma dan Teknik evaluasi *clustering* yang lebih baik seiring waktu berjalan sehingga informasi yang diperoleh dapat lebih rinci dan optimal.

REFERENCES

- [1] A. P. Riani, A. Voutama, and T. Ridwan, “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow,” J-SISKO TECH, vol. 6, no. 1, pp. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [2] H. Pancawardana, I. I. Al Ayyubi, R. Rohmatulloh, and M. Murharyana, “The Influence of Nonformal Education on Students’ Cognitive Formation,” Kolok. J. Pendidik. Luar Sekol., vol. 11, no. 2, pp. 236–243, 2023, doi: 10.24036/kolokium.v11i2.612.
- [3] N. A. Agustin, Kanom, and R. N. Darmawan, “Tingkat Minat Belajar Siswa Kelas IX SMPN 11 Kota Sungai Penuh,” J. Inov. Penelit., vol. 1, no. 3, pp. 1–4, 2020, doi: 10.47492/jip.v1i11.496
- [4] S. Nurhayati, H. Anggidesialamia, D. Suryadi, and M. I. Fasa, “Reading Review Program as an Innovative Effort to Increase Community’s Reading Interest,” vol. 491, no. Ijcah, pp. 184–190, 2020, doi: 10.2991/assehr.k.201201.032.
- [5] A. A. Maiyanti and U. F. Laili, “Pengaruh Bimbingan Belajar (Les di Luar Sekolah) Pada Prestasi Belajar IPA Siswa Kelas VIII MTs Jabal Noer Geluran Sidoarjo,” J. Pendidik. Sains dan Komput., vol. 2, no. 01, pp. 30–40, 2022, doi: 10.47709/jpsk.v2i01.1350.
- [6] Z. S. Barly, F. F. Coastera, and M. Yusa, “Implementasi K-Means Clustering Dan Cosine Similarity Pada

- Case Base Reasoning Untuk Diagnosa Penyakit Sapi,” *Pseudocode*, vol. 9, no. 2, pp. 88–94, 2022, doi: 10.33369/pseudocode.9.2.88-94.
- [7] M. Ulfah and A. Sri Irtwaty, “Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan,” *Fidel. J. Tek. Elektro*, vol. 4, no. 3, pp. 62–68, 2022, doi: 10.52005/fidelity.v4i3.126.
- [8] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [9] F. P. Dewi, P. S. Aryni, and Y. Umidah, “Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran,” *JISKA*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.
- [10] S. Ningsih and A. Dhieka Lantana, “Design Aplikasi Kinerja Dosen Pada Unit Penjaminan Mutu Fakultas Menggunakan Algoritma K-Means Clustering Berbasis Phyton,” *IBI Kosgoro*, vol. 4, no. 1, pp. 23–33, 2023, doi: 10.55122/junsibi.v4i1.693.
- [11] K. Neighbours, “Pengenalan Pola untuk Identifikasi Jenis Kain Tenun Sibolga Menggunakan Metode Principal Component Analysis,” vol. 5, no. 4, pp. 796–805, 2024, doi: 10.47065/josyc.v5i4.
- [12] S. J. Arditna and R. Kurniawan, “Prediksi Harga Tandan Buah Segar dengan Algoritma K-Nearest Neighbor,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 1, pp. 92–101, 2023, doi: 10.30865/json.v5i1.6818.
- [13] M. Sholeh and K. Aeni, “Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means,” *STRING*, vol. 8, no. 1, pp. 56, 2023, doi: 10.30998/string.v8i1.16388.
- [14] R. P. Kurniawan and Ferdiansyah, “Penerapan Algoritme K-Means Clustering Untuk Mengelompokkan Siswa Berdasarkan Nilai Akademik Di SMP Negeri 207 SSN,” *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 2, pp. 530–538, 2023.
- [15] A. A. Muhartini, T. Febriati, and S. Sukmawati, “Analisis Cluster Untuk Mengelompokkan Penggunaan Kartu Perdana Seluler Di Universitas Bina Bangsa,” *J. Bayesian J. Ilm. Stat. dan Ekon.*, vol. 2, no. 1, pp. 15–30, 2022, doi: 10.46306/bay.v2i1.25.
- [16] D. O. Dacwanda and Y. Nataliani, “Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan,” *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [17] A. Harmain, H. Kurniawan, D. Maulina, and M. Universitas Amikom Yogyakarta, “Data Normalization for K-Means Efficiency on Groups of Areas With Potential Forest and Land Fire Based on Heat Spots Distribution,” *Teknimedia*, vol. 2, no. 2, pp. 83–89, 2021, doi: 10.46764/teknimedia.v2i2.49
- [18] D. A. Ofori et al., “Beauty and Skincare E-Commerce Mobile Application with Advanced Searching Module Using Image Processing,” *Molecules*, vol. 2, no. 1, pp. 1–12, 2020.
- [19] S. Butsianto and N. Saepudin, “Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 51–59, 2020, doi: 10.32672/jnkti.v3i1.2008.
- [20] A. W. Chandra, R. Primartha, A. Desiani, and A. Wijaya, “The Effect of Preliminary Centroid Determination Using Particle Swarm Optimization Algorithm in High Dimension Data Clustering,” vol. 172, no. Siconian 2019, pp. 347–352, 2020, doi: 10.2991/aisr.k.200424.052.