

Implementasi Data Mining Pengelompokan Siswa/Siswi Potensial Menggunakan Algoritma K-Means

Rahfi

Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia

Email: rahfi1789@gmail.com

Email Penulis Korespondensi: rahfi1789@gmail.com*

Submitted: 27/11/2024; Accepted: 20/12/2024; Published: 31/12/2024

Abstrak: Perkembangannya teknologi informasi telah mengubah dunia menjadi serba lebih mudah di semua bidang kehidupan dan mendorong untuk lebih kreatif dan inovatif terbukti bahwa mekanisme kerja yang panjang menjadi efektif dan efisien dalam menemukan masalah yang dihadapi. Dengan adanya berbagai teknologi saat ini membuat semua menjadi lebih mudah memperoleh solusi yang akurat dalam suatu masalah. Dalam proses pengelompokan siswa/siswi potensial pada suatu sekolah banyak terdapat berbagai kendala atau masalah yang dihadapi. Salah satunya adalah pengelompokan data siswa/siswi potensial pada MTS Ikhwanuts Tsalits Talun Kenas yang kurang maksimal atau efisien. Untuk mengatasi masalah tersebut, diperlukan data mining dengan metode algoritma K-means Clustering dalam pengelompokan siswa/siswi berdasarkan potensi yang dimiliki. Dengan hasil penelitian yang diperoleh di harapkan dapat memberikan informasi yang efektif kepada Kepala Sekolah dan guru yang mengajar dalam mengetahui atau menentukan potensi yang dimiliki siswa/siswinya. Pengelompokan data siswa/siswi potensial dilakukan dengan meliputi variabel, yaitu : NISN (Nomor Induk Siswa nasional), Nama dan Nilai siswa/siswi. Nilai siswa/siswi diambil berdasarkan jumlah nilai bahasa indonesia, matematika, dan ipa (ilmu pengetahuan alam).

Kata Kunci: Data Mining; Pengelompokan; K-Means, Siswa/Siswi; Potensial

Abstract: The development of information technology has transformed the world into a more accessible and convenient place across all aspects of life, encouraging greater creativity and innovation. It has been proven that lengthy work mechanisms have become more effective and efficient in identifying existing problems. With the advancement of current technologies, obtaining accurate solutions to various issues has become significantly easier. In the process of grouping potential students in a school, various challenges and obstacles are often encountered. One such issue is the inefficient and suboptimal grouping of potential students at MTS Ikhwanuts Tsalits Talun Kenas. To address this problem, data mining using the K-Means Clustering algorithm is required to classify students based on their individual potential. The results of this study are expected to provide effective information to the Principal and teachers, helping them to better understand and identify the potential of their students. The student data grouping is carried out using variables such as NISN (National Student Identification Number), Name, and Scores. The scores are based on the total marks in Bahasa Indonesia, Mathematics, and Natural Sciences (IPA).

Keywords: Data Mining; Clustering; K-Means; Students; Potential

1. PENDAHULUAN

Perkembangannya teknologi informasi telah mengubah dunia menjadi serba lebih mudah di semua bidang kehidupan dan mendorong untuk lebih kreatif dan inovatif terbukti bahwa mekanisme kerja yang panjang menjadi efektif dan efisien dalam menemukan masalah yang dihadapi. Dengan adanya berbagai teknologi saat ini membuat semua menjadi lebih mudah memperoleh solusi yang akurat dalam suatu masalah. Dalam proses pengelompokan siswa/siswi potensial pada suatu sekolah banyak terdapat berbagai kendala atau masalah yang dihadapi. Salah satunya adalah pengelompokan data siswa/siswi potensial pada MTS Ikhwanuts Tsalits Talun Kenas yang kurang maksimal atau efisien. Selama ini Kepala Sekolah dan guru yang mengajar di MTS Ikhwanuts tsalits Talun Kenas kadang mengalami kesulitan dalam menentukan siswa/siswi yang berpotensi, masalah tersebut disebabkan potensi yang dimiliki oleh setiap siswa/siswi berbeda-beda seperti di satu kelas sebagian siswa/siswi berpotensi di bidang matematika tetapi kurang di bidang bahasa indonesia dan sebaliknya. Salah satu penyebab terjadinya masalah seperti itu adalah karena pengelompokan siswa/siswi berasarkan peringkat seperti peringkat 1 sampai 20 kelas A, peringkat 21 sampai 40 kelas B, dan seterusnya. Untuk mengatasi masalah ini, dibutuhkan data mining dengan metode algoritma *K-means Clustering* dalam pengelompokan siswa/siswi berdasarkan potensi yang dimiliki.[1]

Ada dua jenis pengelompokan data yang biasa digunakan dalam proses pengelompokan data : pengelompokan data *hierarkis* dan pengelompokan data *non-hierarkis (non-hierarkis)*. Pengelompokan *non-hierarkis*. Data dipecah menjadi satu atau lebih *cluster/grup*, data dengan karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama, dan data dengan karakteristik yang berbeda dikelompokkan ke dalam grup yang berbeda. Sedangkan pengelompokan hierarki adalah metode analisis klaster yang membentuk level seperti pohon karena proses pengelompokan dilakukan secara bertahap dan berlapis. Hasil dari prosedur hirarki dapat divisualisasikan dalam bentuk dendrogram. Dendrogram adalah representasi visual dari hasil analisis klaster, yang menunjukkan bagaimana klaster terbentuk.[2]

Menurut penelitian terkait yang berjudul *Implementing Data Mining to Group Student Data Using the Kmeans Clustering Algorithm*, *Kmeans Clustering Algorithm* merupakan grup non-hierarchical yang dapat mengelompokkan data menjadi beberapa *cluster* berdasarkan kesamaan data dikatakan sebagai metode data *clustering*[3]. Algoritma *K-means* adalah algoritma *clustering* iteratif. Algoritma ini dimulai dengan K acak, yang

merupakan jumlah *cluster* yang terbentuk. Kemudian tetapkan nilai K secara acak[4]. Untuk sementara, nilai-nilai tersebut menjadi pusat *cluster*. Hal ini sering disebut sebagai *centroid*/rata-rata. Gunakan rumus yang tersedia untuk menghitung jarak setiap data di setiap *centroid* sampai Anda menemukan jarak terdekat dari semua data *centroid*. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Ikuti langkah-langkah ini sampai nilai *centroid* stabil.[5]

Menurut penelitian terkait selanjutnya yg berjudul Implementasi Algoritma *K-means* Untuk Menentukan Kelompok Pengayaan Mata Pelajaran Ujian Nasional menyatakan bahwa Algoritma *K-means* bisa diimplementasikan buat membantu pengelompokkan kemampuan anak didik terhadap mata pelajaran Ujian Nasional.[6][7]

Menurut penelitian terkait selanjutnya yang berjudul Implementasi Algoritma *K-means Clustering* Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan menyatakan bahwa Metode *Clustering* dengan algoritma *K-means* dapat digunakan untuk mengelompokkan data siswa berdasarkan nilai ujian semester mata pelajaran Ujian Nasional, yaitu kemampuan siswa pintar, siswa sedang dan siswa kurang pintar. sehingga dapat mengetahui siswa yang mana saja yang akan diberi belajar tambahan agar dapat mencapai nilai standar kelulusan Ujian Nasional.[8]

Menurut penelitian terkait lainnya yang berjudul Implementasi Algoritma KMeans dalam Pengelompokan Penyandang Disabilitas, menurut Kabupaten. Simalungun, algoritma *K-means* mengelompokkan data berdasarkan pencarian iteratif dari pusat *cluster* dengan lokasi terkecil untuk setiap *cluster* data. membuatnya. Jarak dari data individu ke pusat *cluster*. [9]

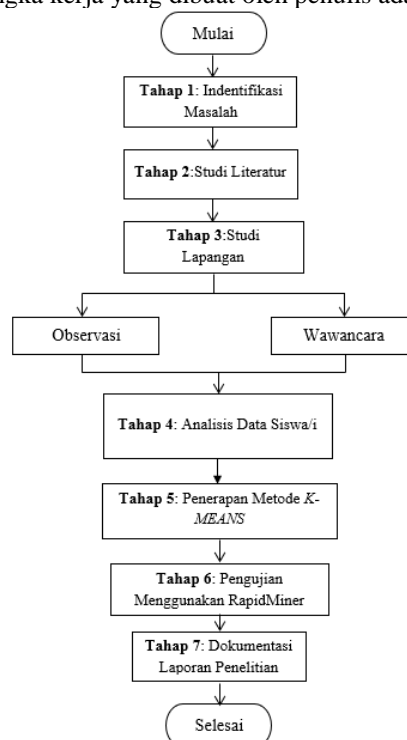
Penelitian yang akan dibuat penulis adalah Implementasi atau Penerapan Data Mining menggunakan Algoritma *K-means*. Penelitian mengenai penerapan Data Mining menggunakan Algoritma *K-means* sudah banyak dilakukan oleh orang-orang yang berada dibidang *Information Technology* atau pada bidang lainnya.

Dengan hasil penelitian yang diperoleh di diharapkan dapat memberikan informasi yang efektif kepada Kepala Sekolah dan guru yang mengajar dalam mengetahui atau menentukan potensi yang dimiliki siswa/siswinya. Pengelompokan data siswa/siswi potensial dilakukan dengan meliputi variabel, yaitu : NISN (Nomor Induk Siswa nasional), Nama dan Nilai siswa/siswi. Nilai siswa/siswi diambil berdasarkan jumlah nilai bahasa indonesia, matematika, dan ipa (ilmu pengetahuan alam).

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Kerangka kerja penelitian adalah suatu bentuk kerangka kerja yang dapat digunakan sebagai pendekatan untuk memecahkan masalah. Adapun kerangka kerja yang dibuat oleh penulis adalah sebagai berikut :



Gambar 1. Kerangka Penelitian

Adapun penjelasan dari tahapan penelitian adalah sebagai berikut:

1. Identifikasi Masalah

- Yaitu uraian masalah yang melatar belakangi pembuatan skripsi ini tentang permasalahan dalam menentukan penilaian siswa/siswi pada MTS Ikhwanuts Tsalits Talun Kenas.
2. Studi Literatur
Yaitu proses pengumpulan bahan-bahan referensi untuk memperoleh informasi yang dibutuhkan dalam pembuatan skripsi ini, baik dari buku, artikel, makalah, jurnal dan situs internet.
 3. Studi Lapangan
Yaitu pengumpulan data dengan meneliti langsung pada MTS Ikhwanuts Tsalits Talun Kenas dengan cara melibatkan pihak-pihak yang terkait. Hal ini dilakukan untuk mengumpulkan data dan informasi yang berhubungan dengan permasalahan. Yaitu dengan cara observasi dan wawancara:
 - a. Observasi (Pengamatan Langsung)
Yaitu penulis melakukan pengamatan secara langsung ke perusahaan atau tempat yang diteliti untuk mendapatkan informasi-informasi yang lebih banyak tentang perusahaan, yang nantinya akan digunakan dalam penelitian ini.
 - b. Interview (Wawancara)
Wawancara merupakan teknik pengumpulan data yang dilakukan dengan cara berkomunikasi atau mewawancarai secara langsung kepada pihak perusahaan yang menangani penentuan dalam penilaian siswa/siswi potensial pada MTS Ikhwanuts Tsalits Talun Kenas.
 4. Analisa Data Siswa/i
Analisis adalah serangkaian kegiatan, aktivitas, dan proses yang saling terkait yang digunakan untuk memecahkan masalah, memecahkan komponen secara lebih rinci, dan bergabung kembali untuk menarik kesimpulan.
 5. Penerapan Metode K-means
Pada tahap ini penulis melakukan cluster data pada siswa/siswi di MTS Ikhwanuts Tsalits Talun Kenas dengan metode K Means. Dalam clustering penilaian siswa/siswi potensial menggunakan algoritma K Means.
 6. Pengujian Menggunakan Rapidminer
Pengujian dilakukan dengan menggunakan software rapidminer. Pengujian ini dilakukan untuk mengetahui hasil dari penerapan metode K-means apakah hasil yang diperoleh sesuai dengan maksud dan tujuan penelitian dalam menyelesaikan suatu permasalahan yang dibuat.
 7. Kesimpulan dan Saran
Yaitu merupakan tahapan akhir dari penulisan skripsi ini, dengan memberi kesimpulan berdasarkan hasil yang diperoleh dari penelitian ini. Apakah penelitian ini mampu memecahkan permasalahan untuk mendukung suatu keputusan dalam menentukan siswa/siswi menjadi siswa/siswi berpotensi. Sedangkan saran yang dibuat dapat digunakan sebagai masukan bagi perusahaan untuk dapat dikembangkan lebih lanjut.

2.2 Data Mining

Data mining adalah sebuah konsep yang digunakan untuk menemukan nilai tambah yang tersembunyi dalam sebuah database. Data mining adalah proses semi-otomatis yang menggunakan statistik, matematika, kecerdasan buatan, dan teknik pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi pengetahuan yang berpotensi berguna yang disimpan dalam database besar.[10] Data mining adalah bagian dari proses Knowledge Discovery Databases (KDD) dan terdiri dari beberapa fase seperti pemilihan data, pemrosesan, transformasi, penambangan data, dan evaluasi hasil. Teknologi data mining secara garis besar dapat dibagi menjadi dua kelompok: validasi dan penemuan[11]. Metode validasi mencakup metode statistik seperti Goodness-of-Fit dan ANOVA, tetapi metode deteksi dapat dibagi lagi menjadi model prediktif dan deskriptif. Model prediktif menggunakan hasil yang diketahui dari berbagai data untuk memprediksi data. Model deskriptif dapat dibangun di atas penggunaan data historis lainnya. Model deskriptif, di sisi lain, dimaksudkan untuk menyediakan cara untuk mengidentifikasi pola atau hubungan antara data dan untuk menyelidiki sifat-sifat data yang sedang diselidiki. [12] Data mining adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam database. Data mining adalah proses mengekstraksi dan mengidentifikasi informasi dan pengetahuan yang berguna yang dikumpulkan dari berbagai database besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Menurut Grup Gatner, penambangan data menggunakan teknik pengenalan pola seperti teknik statistik dan matematika untuk memeriksa kumpulan data besar yang disimpan dalam memori untuk mengidentifikasi hubungan, pola, dan tren yang bermakna.[13], [14]

2.3 Clustering

Clustering adalah proses pembentukan data group dari data set yang grupnya tidak diketahui karena kemiripannya. Artinya ada data dalam *cluster* yang memiliki sifat yang hampir sama. Analisis *clustering* adalah teknik analisis multivariat yang digunakan untuk mengelompokkan objek (variabel atau data), menghasilkan informasi yang berguna untuk melaksanakan pengujian objek, dan pada akhirnya membuat hipotesis berdasarkan hubungan yang terjadi. Tujuan dari teknik ini adalah untuk mengelompokkan berdasarkan kriteria tertentu. Hal ini menyebabkan objek-objek tersebut memiliki variasi yang relatif kecil di dalam *cluster* dibandingkan dengan variasi antar *cluster*. Untuk mengukur kesamaan, gunakan rumus untuk menghitung jarak. Rumus yang paling umum digunakan adalah rumus jarak *Euclidean*. [5] *Clustering*, atau *clustering*, adalah teknik penambangan data yang digunakan untuk

memecahkan masalah seperti menganalisis data untuk mengelompokkan data atau membagi *record* menjadi subset. Dalam teknik *clustering*, tujuan dari kasus ini adalah untuk mendistribusikan kelompok (objek, orang, peristiwa, dan lainnya) ke dalam kelompok untuk meningkatkan konektivitas antar anggota *cluster* yang sama dan melemahkan konektivitas antar anggota *cluster* yang berbeda.[15] Ada dua metode pengelompokan untuk pengelompokan: pengelompokan hierarkis dan pengelompokan *non-hierarkis*. *Hierarchical clustering* adalah suatu metode pengelompokan data yang bekerja dengan cara mengelompokkan dua atau lebih bagian data yang memiliki persamaan atau persamaan kemudian mengkonstruksinya dengan objek lain yang mendekati dua. Proses ini melibatkan hierarki *cluster* atau level yang jelas antara objek mulai dari yang paling mirip hingga yang paling tidak mirip. Namun secara logika, semua objek tersebut akhirnya hanya membentuk *cluster*. [2] *Clustering non-hierarchical* dalam teknik ini dimulai dengan menentukan jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, empat atau lebih *cluster*), setelah jumlah *cluster* yang diinginkan, proses *clustering* dimulai tanpa mengikuti aturan. juga biasa disebut sebagai metode *K-means Clustering*. [15]

2.4 Algoritma K-Means

K-means adalah algoritma yang digunakan dalam partisi yang membagi data ke dalam kelompok yang berbeda. Algoritma ini dapat meminimalkan jarak antara data dan *cluster*. Pada dasarnya, penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang diterima dan kesimpulan yang diambil di akhir proses. [15] *K-Means* adalah algoritma pengelompokan berulang yang membagi kumpulan data menjadi sejumlah *K cluster* yang telah ditentukan. Algoritma KMeans mudah diimplementasikan dan dieksekusi, relatif cepat, mudah beradaptasi, dan pada kenyataannya banyak digunakan. Secara historis, *K-means* telah menjadi salah satu algoritma terpenting dalam penambahan data. Berikut adalah langkah-langkah untuk algoritma *K-means S*. [16] Metode *K-means* terkenal karena kesederhanaannya dan kemampuannya untuk mengelompokkan data besar dan outlier dengan sangat cepat. Pengelompokan *K-means* mensyaratkan bahwa semua data milik *cluster* tertentu dalam satu fase proses dan dapat dipindahkan ke *cluster* lain di fase proses berikutnya. [17] Metode *K-means* adalah fungsi *clustering* atau metode pengelompokan. Menurut Larose, pengelompokan mengacu pada pengelompokan data, pengamatan, atau kasus berdasarkan kesamaan objek yang diselidiki. *Cluster* adalah kumpulan data yang mirip atau berbeda dari data dalam kelompok lain. Xu & Wunsch menjelaskan bahwa *clustering* membagi objek data (bentuk, entitas, contoh, ketaatan, unit) menjadi beberapa kelompok (grup, bagian, atau kategori). Tujuan dari proses *clustering* adalah untuk meminimalkan terjadinya himpunan fungsi yang diinginkan dalam proses *clustering*. Ini biasanya digunakan untuk meminimalkan variabilitas dalam sebuah *cluster* dan memaksimalkan variabilitas antar *cluster*. [18]

Tahapan melakukan *clustering* atau pengelompokan dengan metode *K-means S* adalah sebagai berikut :

1. Menentukan berapa banyak *cluster* yang ingin yang ingin dibentuk, di mana nilai *K* adalah banyaknya *cluster*/ jumlah *cluster*.
2. Menentukan pusat *cluster* (*centroid*) awal.
3. Setelah menentukan *centroid* awal, maka setiap data akan menemukan *centroid* terdekatnya
4. Setelah menghitung jarak data ke *centroid* nya.
5. Mencari *centroid* baru berdasarkan membership dari masing-masing *cluster*.
6. Kembali ke tahap 3.
7. Perulangan berhenti apabila tidak ada data yang berpindah.

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} x_{ki} \quad (1)$$

Dimana :

V_{ij} = *centroid* rata-rata *cluster* ke-*i* untuk variable ke-*j*.

N_i = jumlah anggota *cluster* ke-*i*.

i, k = indeks dari *cluster*.

j = indeks dari variable.

X_{kj} = nilai data ke-*k* variable ke-*j* dalam *cluster* tersebut.

Menggunakan rumus *Euclidean Distance* berikut :

$$d(x, y) = ||x - y||^2 = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

Dimana :

$d(x, y)$ = jarak *ecludid*

X = data *record*

Y = data *centroid*

$||$ = nilai *absolute*

Demikianlah rumus yang digunakan untuk menentukan korelasi antara dua objek yaitu dengan menggunakan rumus *Euclidean Distance*. Sementara itu, Theodoridis dan Koitroumbas menyatakan bahwa *K-means* adalah salah satu metode algoritma *clustering* yang paling populer. Di sini, *K-means* menemukan partisi terbaik dari data dengan meminimalkan ukuran jumlah kesalahan kuadrat dengan metode iteratif terbaik. Itu termasuk dalam kategori algoritma pendakian gunung. Algoritma dasar untuk pengelompokan data menggunakan metode *K-means* dapat dilakukan sebagai berikut :

1. Tentukan jumlah kelompok

2. Alokasikan data kedalam kelompok secara acak
3. Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata (mean) semua nilai pada setiap fitur nya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan 1.

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j \tag{3}$$

Persamaan 1 dilakukan sebanyak p . Dimensi dari $i=1$ sampai dengan $i=p$

4. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok. Pengukuran jarak pada ruang jarak (distance space) euclidean dapat dicari menggunakan persamaan 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{4}$$

Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode *K-means* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai *centroid* dengan jarak terdekat dari data tersebut. Pengalokasian data ini dapat ditentukan menggunakan persamaan sebagai berikut.

$$a_{i1} = \begin{cases} 1 & d = \min\{D(x, e_1)\} \\ 0 & \text{lainya} \end{cases} \tag{5}$$

a_{i1} adalah keanggotaan titik x_i ke pusat kelompok setelah dibandingkan, dan c_1 adalah *centroid* (pusat kelompok) ke-1. Fungsi objektif yang digunakan untuk metode *K-means* ditentukan berdasarkan jarak dan nilai ke anggotaan dalam kelompok. Fungsi objektif dapat ditentukan dengan persamaan 4.

$$\sum_{i=1}^n \sum_{k=1}^k a_{ik} D(x_i, c_k)^2 \tag{6}$$

n adalah jumlah data, k adalah jumlah kelompok, a_{i1} adalah nilai keanggotaan titik data x_i ke kelompok c_1 yang diikuti. a mempunyai nilai 0 atau 1. Apabila data merupakan anggota suatu kelompok, nilai $a_{i1} = 1$, jika tidak nilai $a_{i1} = 0$.

5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

Pada dasarnya ada dua cara untuk mengasosiasikan data dengan setiap *cluster* selama proses iteratif *clustering*. Metode pertama adalah pemetaan keras. Artinya, item data secara eksplisit ditentukan sebagai anggota dari satu *cluster*, bukan anggota *cluster* lain. Metode kedua adalah metode *fuzzy* dimana setiap elemen data diberi nilai probabilitas sehingga dapat bergabung dengan *cluster* yang ada.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Dan Penerapan Metode

Data siswa/siswi MTS Ikhwanuts Tsalits Talun Kenas saat ini sudah terkumpul cukup banyak tetapi dalam pengumpulan datanya kurang maksimal atau efisien, sehingga memakan waktu yang relative lama dan menyulitkan Kepala Sekolah dan guru yang mengajar dalam melakukan pemantauan terhadap perkembangan siswa/siswinya. Adapun data yang ada atau yang sedang berjalan di MTS Ikhwanuts Tsalits Talun Kenas adalah sebagai berikut:

Tabel 1. Daftar Nama-Nama siswa/siswi di MTS Ikhwanuts Tsalits Talun Kenas

No	NISN (Nomor Induk Siswa Nasional)	Nama	Kriteria dan Penilaian		
			Bahasa Indonesia	Matematika	IPA
1	0094221068	Adithia pratama tambun	92	90	89
2	0093033682	Aldi ramadhani sitepu	88	89	90
3	0096313567	Adelia salsal nabila	89	90	85
4	0098714480	Aulia apri sipani	90	86	88
5	0091015583	Amini arun dati	86	84	90
6	0098548806	Ewi sindi aulia	84	88	90
7	0092468953	Ferry ardyansyah	90	87	88
8	0092052399	Goviani tarigan	85	90	89
9	0083275754	Hendra lesmana	88	89	90
10	0092425607	Hanun alliansyah	90	89	88
11	1081466674	Iis erniwati damanik	90	89	87
12	0092322072	Muhammad ilham zahri	90	90	84
13	3095671845	M arafah ramadhan	90	85	87
14	0064749404	M. yusuf	90	89	88

No	NISN (Nomor Induk Siswa Nasional)	Nama	Kriteria dan Penilaian		
			Bahasa Indonesia	Matematika	IPA
15	0097263018	Mhd. Ikhsan	88	87	90
16	0108574012	Muhnizar syahputra	87	90	88
17	0093033585	Nadya Carissa putri tarigan	90	89	86
18	0092001270	Nadia sapitri	89	88	90
19	0098999150	Novitasari nasution	87	86	87
20	0087156445	Nur azizah	89	88	88

Dengan adanya *Algoritma K-means* diharapkan akan dapat membantu pekerjaan Kepala sekolah dan guru yang mengajar untuk mengetahui perkembangan siswa/siswinya di sekolah dan dapat mempermudah dalam proses pengelompokan data siswa/siswi yang berpotensi secara maksimal atau efisien.

Sistem pengelompokan data siswa/siswi potensial yang menggunakan metode *K-mean Clustering* yang mana metode ini menghitung jarak kedekatan data dengan *centroid* yang sudah ditentukan lalu dilihat seberapa dekat jarak kemiripannya.

1. Menentukan pusat *cluster (centroid)* awal.

Hal yang dilakukan dalam iterasi pertama adalah menentukan *centroid* awal. Nilai *centroid* awal dipilih secara acak dengan melihat bagaimana *range* dari tiap nilai pada data siswa. Untuk *centroid* awal diambil dari data ke-6 sebagai pusat *cluster* 1, data ke-4 sebagai pusat *cluster* 2, dan data ke-7 sebagai pusat *cluster* 3 yang diambil dari sampel data pada bab III Tabel 3.1 Daftar Nama-Nama siswa/siswi di MTS Ikhwanuts Tsalits Talun Kenas.

Tabel 2. Nilai *centroid* awal

No	Nama	Bahasa Indonesia	Matematika	Ipa	<i>Centroid</i>
6	Ewi sindi aulia	84	88	90	C1
4	Aulia apri sipani	90	86	88	C2
7	Ferry ardyansyah	90	87	88	C3

Rumus yang digunakan diambil dari bab II dengan nomor persamaan (2.2) yaitu Menggunakan rumus *Euclidean Distance*. Berikut ini adalah perhitungan iterasi 1 (satu) jarak data terhadap pusat *cluster* :

Cluster 1

- $\sqrt{((92 - 84))^2 + ((90 - 88)^2) + (89 - 90)^2} = 8,30$
- $\sqrt{((88 - 84))^2 + ((89 - 88)^2) + (90 - 90)^2} = 4,12$
- $\sqrt{((89 - 84))^2 + ((90 - 88)^2) + (85 - 90)^2} = 7,34$
- $\sqrt{((90 - 84))^2 + ((86 - 88)^2) + (88 - 90)^2} = 6,63$
- $\sqrt{((86 - 84))^2 + ((84 - 88)^2) + (90 - 90)^2} = 4,47$
- $\sqrt{((84 - 84))^2 + ((88 - 88)^2) + (90 - 90)^2} = 0$
- $\sqrt{((90 - 84))^2 + ((87 - 88)^2) + (88 - 90)^2} = 6,40$
- $\sqrt{((85 - 84))^2 + ((90 - 88)^2) + (89 - 90)^2} = 2,44$
- $\sqrt{((88 - 84))^2 + ((89 - 88)^2) + (90 - 90)^2} = 4,12$
- $\sqrt{((90 - 84))^2 + ((89 - 88)^2) + (88 - 90)^2} = 6,40$
- $\sqrt{((90 - 84))^2 + ((89 - 88)^2) + (87 - 90)^2} = 6,78$
- $\sqrt{((90 - 84))^2 + ((90 - 88)^2) + (84 - 90)^2} = 8,71$
- $\sqrt{((90 - 84))^2 + ((85 - 88)^2) + (87 - 90)^2} = 7,34$
- $\sqrt{((90 - 84))^2 + ((89 - 88)^2) + (88 - 90)^2} = 6,40$
- $\sqrt{((88 - 84))^2 + ((87 - 88)^2) + (90 - 90)^2} = 4,12$
- $\sqrt{((87 - 84))^2 + ((90 - 88)^2) + (88 - 90)^2} = 4,12$
- $\sqrt{((90 - 84))^2 + ((89 - 88)^2) + (86 - 90)^2} = 7,28$
- $\sqrt{((89 - 84))^2 + ((88 - 88)^2) + (90 - 90)^2} = 5$
- $\sqrt{((87 - 84))^2 + ((86 - 88)^2) + (87 - 90)^2} = 4,69$
- $\sqrt{((89 - 84))^2 + ((88 - 88)^2) + (88 - 90)^2} = 5,38$

Lakukan hingga 3 kali *cluster*. Dari perhitungan diatas, maka didapatkan hasil perhitungan iterasi 1 pada tabel berikut ini dengan terbentuk *cluster* 1 (C1), *cluster* 2 (C2), dan *cluster* 3 (C3).

Tabel 3. hasil perhitungan iterrasi 1

No	NISN	NAMA	C1	C2	C3	Hasil
1	0094221068	Adithia pratama tambun	8,30	4,58	3,74	C3
2	0093033682	Aldi ramadhani sitepu	4,12	4,12	3,46	C3

3	0096313567	Adelia salsal nabila	7,34	5,09	4,35	C3
4	0098714480	Aulia apri sipani	6,63	0	1	C2
5	0091015583	Amini arun dati	4,47	4,89	5,38	C2
6	0098548806	Ewi sindi aulia	0	6,63	6,40	C1
7	0092468953	Ferry ardyansyah	6,40	1	0	C3
8	0092052399	Goviani tarigan	2,44	6,48	5,91	C1
9	0083275754	Hendra lesmana	4,12	4,12	3,46	C3
10	0092425607	Hanun allianyah	6,40	3	2	C3
11	1081466674	Iis erniwati damanik	6,78	3,16	2,23	C3
12	0092322072	Muhammad ilham zahri	8,71	5,65	5	C3
13	3095671845	M arafah ramadhan	7,34	1,41	2,23	C2
14	0064749404	M. yusuf	6,40	3	2	C3
15	0097263018	Mhd. ikhsan	4,12	3	3,82	C2
16	0108574012	Muhnizar syahputra	4,12	5	4,24	C3
17	0093033585	Nadya Carissa putri tarigan	7,28	3,60	2,82	C3
18	0092001270	Nadia sapitri	5	3	2,44	C3
19	0098999150	Novitasari nasution	4,69	3,16	3,31	C2
20	0087156445	Nur azizah	5,38	2,23	1,41	C3

2. Menentukan pusat *cluster* (*centroid*) baru.

Setelah anggota dari setiap *cluster* diketahui, maka *centroid* baru dihitung berdasarkan rata-rata data untuk setiap anggota *cluster*, untuk *cluster* pertama ada 2 data, *cluster* kedua ada 5 data dan *cluster* ketiga ada 13 data untuk dihitung menjadi *cluster* baru.

Tabel 4. Perhitungan *centroid* baru *cluster* 1

6	Ewi sindi aulia	84	88	90
8	Goviani tarigan	85	90	89
Nilai Rata-rata		84,5	88,2	89,38

Tabel 5. perhitungan *centroid* baru *cluster* 2

4	Aulia apri sipani	90	86	88
5	Amini arun dati	86	84	90
13	M arafah ramadhan	90	85	87
15	Mhd. ikhsan	88	87	90
19	Novitasari nasution	87	86	87
Nilai Rata-rata		89	85,6	89

Tabel 6. perhitungan *centroid* baru *cluster* 3

1	Adithia pratama tambun	92	90	89
2	Aldi ramadhani sitepu	88	89	90
3	Adelia salsal nabila	89	90	85
7	Ferry ardyansyah	90	87	88
9	Hendra lesmana	88	89	90
10	Hanun allianyah	90	89	88
11	Iis erniwati damanik	90	89	87
12	Muhammad ilham zahri	90	90	84
14	M. yusuf	90	89	88
16	Muhnizar syahputra	87	90	88
17	Nadya Carissa putri tarigan	90	89	86
18	Nadia sapitri	89	88	90
20	Nur azizah	89	88	88
Nilai Rata-rata		89,5	88,4	87,76

Setelah dilakukan perhitungan maka didapatkan *centroid* baru yang dapat dilihat pada tabel dibawah ini :

Tabel 7. *centroid* baru

No	Bahasa Indonesia	Matematika	Ipa	<i>Centroid</i>
1	84,5	88,2	89,38	C1
2	89	85,6	89	C2
3	89,5	88,4	87,76	C3

Setelah diketahui *centroid* baru lakukan perhitungan iterasi ke-2 (dua)

Cluster 1

1. $\sqrt{((92 - 84,5))^2 + ((90 - 88,2))^2 + (89 - 89,38)^2} = 7,72$
2. $\sqrt{((88 - 84,5))^2 + ((89 - 88,2))^2 + (90 - 89,38)^2} = 3,64$
3. $\sqrt{((89 - 84,5))^2 + ((90 - 88,2))^2 + (85 - 89,38)^2} = 6,53$
4. $\sqrt{((90 - 84,5))^2 + ((86 - 88,2))^2 + (88 - 89,38)^2} = 6,08$
5. $\sqrt{((86 - 84,5))^2 + ((84 - 88,2))^2 + (90 - 89,38)^2} = 4,50$
6. $\sqrt{((84 - 84,5))^2 + ((88 - 88,2))^2 + (90 - 89,38)^2} = 0,82$
7. $\sqrt{((90 - 84,5))^2 + ((87 - 88,2))^2 + (88 - 89,38)^2} = 5,79$
8. $\sqrt{((85 - 84,5))^2 + ((90 - 88,2))^2 + (89 - 89,38)^2} = 1,90$
9. $\sqrt{((88 - 84,5))^2 + ((89 - 88,2))^2 + (90 - 89,38)^2} = 3,64$
10. $\sqrt{((90 - 84,5))^2 + ((89 - 88,2))^2 + (88 - 89,38)^2} = 5,72$
11. $\sqrt{((90 - 84,5))^2 + ((89 - 88,2))^2 + (87 - 89,38)^2} = 6,04$
12. $\sqrt{((90 - 84,5))^2 + ((90 - 88,2))^2 + (84 - 89,38)^2} = 7,90$
13. $\sqrt{((90 - 84,5))^2 + ((85 - 88,2))^2 + (87 - 89,38)^2} = 6,70$
14. $\sqrt{((90 - 84,5))^2 + ((89 - 88,2))^2 + (88 - 89,38)^2} = 5,72$
15. $\sqrt{((88 - 84,5))^2 + ((87 - 88,2))^2 + (90 - 89,38)^2} = 3,75$
16. $\sqrt{((87 - 84,5))^2 + ((90 - 88,2))^2 + (88 - 89,38)^2} = 3,37$
17. $\sqrt{((90 - 84,5))^2 + ((89 - 88,2))^2 + (86 - 89,38)^2} = 6,50$
18. $\sqrt{((89 - 84,5))^2 + ((88 - 88,2))^2 + (90 - 89,38)^2} = 4,54$
19. $\sqrt{((87 - 84,5))^2 + ((86 - 88,2))^2 + (87 - 89,38)^2} = 4,09$
20. $\sqrt{((89 - 84,5))^2 + ((88 - 88,2))^2 + (88 - 89,38)^2} = 4,71$

Tabel 8. Hasil perhitungan iterasi 2

No	NISN	NAMA	C1	C2	C3	Hasil
1	0094221068	Adithia pratama tambun	7,72	5,32	3,21	C3
2	0093033682	Aldi ramadhani sitepu	3,64	3,68	2,76	C3
3	0096313567	Adelia salsal nabila	6,53	5,94	3,22	C3
4	0098714480	Aulia apri sipani	6,08	1,46	2,46	C2
5	0091015583	Amini arun dati	4,50	3,54	6,05	C2
6	0098548806	Ewi sindi aulia	0,82	5,63	5,95	C1
7	0092468953	Ferry ardyansyah	5,79	1,98	1,50	C3
8	0092052399	Goviani tarigan	1,90	5,94	4,93	C1
9	0083275754	Hendra lesmana	3,64	3,68	2,76	C3
10	0092425607	Hanun allianyah	5,72	3,68	0,81	C3
11	1081466674	Iis erniwati damanik	6,04	4,06	1,08	C3
12	0092322072	Muhammad ilham zahri	7,90	6,73	4,11	C3
13	3095671845	M arafah ramadhan	6,7	2,31	3,51	C2
14	0064749404	M. yusuf	5,72	3,68	0,81	C3
15	0097263018	Mhd. ikhsan	3,75	1,98	3,03	C2
16	0108574012	Muhnizar syahputra	3,37	4,93	2,97	C3
17	0093033585	Nadya Carissa putri tarigan	6,50	4,64	1,92	C3
18	0092001270	Nadia sapitri	4,54	2,6	2,32	C3
19	0098999150	Novitasari nasution	4,09	2,85	3,54	C2
20	0087156445	Nur azizah	4,71	2,6	0,68	C3

Pada iterasi 2, tidak ada perbedaan hasil *clustering* pada iterasi 1 dengan iterasi 2. Oleh karena itu, maka proses berhenti. Berdasarkan perhitungan jarak dekat, maka siswa/siswi potensial baru dapat dikelompokkan sebagai berikut :

Tabel 9. kelompok 1

No	NISN	NAMA
6	0098548806	Ewi sindi aulia
8	0092052399	Goviani tarigan

Siswa/siswi yang berpotensi dalam bidang bahasa indonesia

Tabel 10. kelompok 2

No	NISN	NAMA
4	0098714480	Aulia apri sipani

5	0091015583	Amini arun dati
13	3095671845	M arafah ramadhan
15	0097263018	Mhd. ikhsan
19	0098999150	Novitasari nasution

Siswa/siswi yang berpotensi dalam bidang matematika

Tabel 11. kelompok 3

No	NISN	NAMA
1	0094221068	Adithia pratama tambun
2	0093033682	Aldi ramadhani sitepu
3	0096313567	Adelia salsal nabila
7	0092468953	Ferry ardyansyah
9	0083275754	Hendra lesmana
10	0092425607	Hanun allianyah
11	1081466674	Iis erniwati damanik
12	0092322072	Muhammad ilham zahri
14	0064749404	M. yusuf
16	0108574012	Muhnizar syahputra
17	0093033585	Nadya Carissa putri tarigan
18	0092001270	Nadia sapitri
20	0087156445	Nur azizah

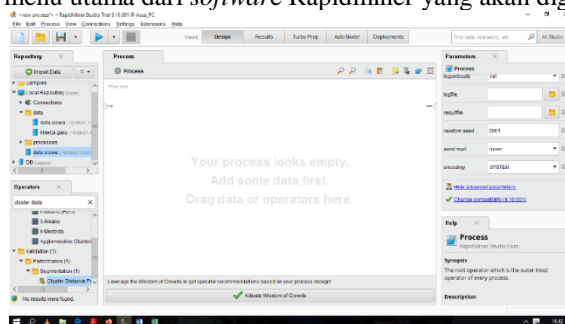
Siswa/siswi yang berpotensi dalam bidang ipa (ilmu pengetahuan alam)

3.2 Hasil Pengujian

Pengujian dilakukan dengan menggunakan *software* rapidminer. Pengujian ini dilakukan untuk mengetahui hasil dari penerapan metode *K-MEANS*. Berikut tampilan pengujian dari sistem pengelompokan siswa/siswi potensial dengan menggunakan metode *K-MEANS clustering software* rapidminer :

1. Menu Utama

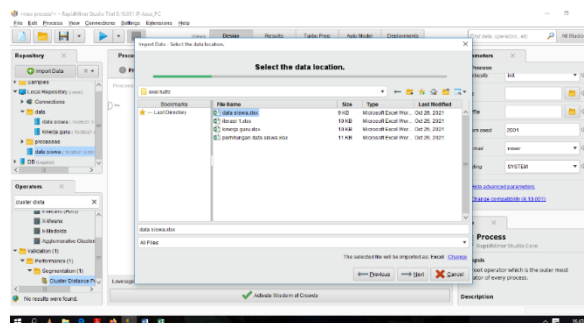
Berikut ini adalah tampilan menu utama dari *software* Rapidminer yang akan digunakan dalam pengujian.



Gambar 2. Tampilan menu utama

2. Import Data

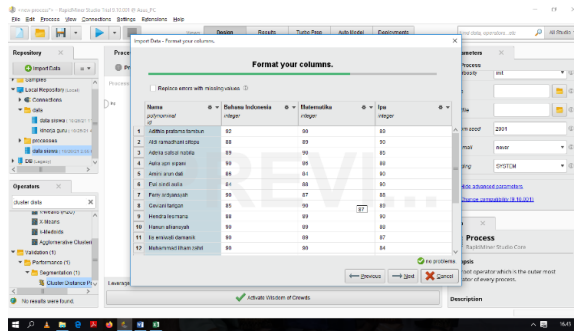
Tahapan ini merupakan proses memasukkan (*Import*) data yang akan diolah dalam pengujian menggunakan *software* rapidminer.



Gambar 3. Tampilan Import Data

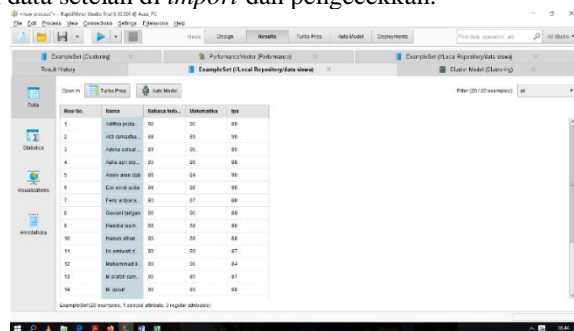
3. Pengecekan Data Yang di Import

Setelah *import* data tahap selanjutnya adalah pengecekan data, pengecekan ini dilakukan untuk memastikan data yang akan diolah sesuai dengan yang telah dibuat.



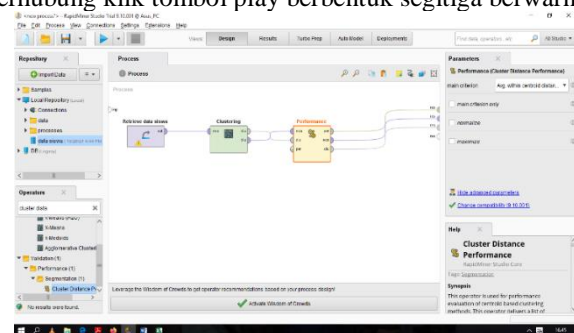
Gambar 4. Tampilan pengecekan data yang di import

4. Data Setelah di import
Berikut ini adalah tampilan data setelah di *import* dan pengecekan.



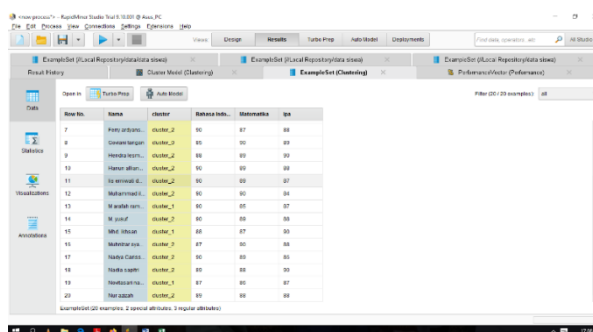
Gambar 5. Tampilan setelah data di import

5. Proses K-Means
Proses ini adalah tahap menghubungkan data dengan algoritma *k-mean Clustering* untuk mengetahui hasil dari pengujian. Apabila sudah terhubung klik tombol play berbentuk segitiga berwarna biru.



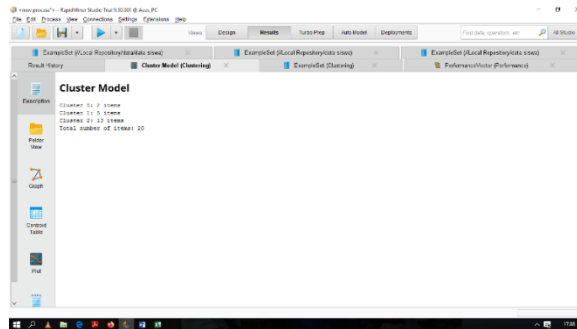
Gambar 6. Tampilan proses K-Means

6. Running Data
Pada bagian ini terdapat beberapa bentuk tampilan hasil cluster yaitu : Tampilan hasil data view, Folder view dan Text view. Data view merupakan tampilan hasil cluster data keseluruhan sesuai dengan data yang telah di import.



Gambar 7. Tampilan folder view

Text view merupakan tampilan hasil pengelompokan berdasarkan *cluster* dan jumlah anggotanya.



Gambar 8. Tampilan text view

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma clustering K-Means terbukti efektif dalam proses pengelompokan siswa dan siswi yang berpotensi, khususnya di lingkungan MTS Ikhwanuts Tsalits Talun Kenas. Metode ini mampu mengelompokkan data siswa secara sistematis berdasarkan kemampuan mereka dalam mata pelajaran utama seperti Bahasa Indonesia, Matematika, dan IPA (Ilmu Pengetahuan Alam). Penerapan algoritma K-Means dalam sistem ini menjadi salah satu langkah strategis dalam membantu pihak sekolah mengidentifikasi siswa-siswi berpotensi secara lebih objektif dan efisien. Dengan pengelompokan berbasis data, proses pengambilan keputusan oleh guru maupun kepala sekolah menjadi lebih terarah dan berbasis bukti, bukan hanya berdasarkan pengamatan subjektif. Selain itu, sistem ini memberikan kemudahan dalam pengolahan data akademik yang sebelumnya dilakukan secara manual. Guru dapat dengan cepat mengetahui kelompok siswa yang menunjukkan potensi unggul dalam bidang tertentu dan memberikan perhatian atau pembinaan lebih lanjut kepada mereka sesuai kebutuhan. Hal ini diharapkan dapat mendukung pengembangan bakat dan prestasi siswa secara lebih optimal. Secara keseluruhan, implementasi algoritma clustering K-Means pada sistem pengelompokan siswa ini membawa dampak positif dalam dunia pendidikan, terutama dalam proses evaluasi dan pengembangan potensi peserta didik. Sistem ini dapat menjadi salah satu inovasi yang relevan untuk diterapkan di sekolah-sekolah lain guna meningkatkan kualitas pembelajaran dan pemantauan perkembangan siswa secara berkelanjutan.

REFERENCES

- [1] D. I. Jenjang and S. Dasar, "PENERAPAN PENGELOMPOKAN SISWA BERDASARKAN PRESTASI DI JENJANG SEKOLAH DASAR Doddy Hendro Wibowo," vol. 14, no. 2, pp. 148–159, 2015.
- [2] P. Ekonomi, D. I. Provinsi, and S. Selatan, "1, 2, 3," pp. 1–9, 2018.
- [3] R. S. Kenett and E. Baker, "Software Process Quality," *Softw. Process Qual.*, pp. 448–475, 1999, doi: 10.1201/9780203909782.
- [4] H. Yuwafi, F. Marisa, and I. D. Wijaya, "Implementasi Data Mining Untuk Menentukan Santri Berprestasi Di Pp . Manaaruhuda Dengan Metode Clustering Algoritma K-Means," *J. SPIRIT*, vol. 11, no. 1, pp. 22–29, 2019.
- [5] A. E. Wicaksono, "... Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus Sman 16 ...)," *J. Ilm. Teknol. dan Rekayasa*, vol. 21, no. 3, 2017.
- [6] F. E. M. Agustin, A. Fitria, and A. H. S, "(STUDI KASUS : SMP NEGERI 101 JAKARTA) Program Studi Teknik Informatika , Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah," pp. 73–78.
- [7] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [8] D. Mining and K. Belajar, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN KELAS KELOMPOK BIMBINGAN BELAJAR TAMBAHAN (STUDI KASUS : SISWA SMA NEGERI 1 RANAH PESISIR) Abstrak," 2017.
- [9] A. I. Sari, H. S. Tambunan, W. Saputra, and I. Sudahri, "Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten Simalungun," vol. 2, pp. 54–61, 2020.
- [10] N. Rofiqo *et al.*, "PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS," vol. 2, pp. 216–223, 2018.
- [11] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- [12] I. Lestari, L. Gaol, S. Sinurat, and E. R. Siagian, "IMPLEMENTASI DATA MINING DENGAN METODE REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI DATA PERSEDIAAN BUKU PADA PT . YUDHISTIRA," vol. 3, pp. 130–133, 2019, doi: 10.30865/komik.v3i1.1579.
- [13] M. Faid, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," vol. 8, 2019, doi: 10.34148/teknika.v8i1.95.
- [14] I. Budiman and R. Ramadina, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *Ijccs*, vol. x, No.x, no. 1, pp. 1–5, 2015.
- [15] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen



- Pekalongan,” *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [16] R. A. Margolang, S. R. Andani, and M. R. Lubis, “Implementasi Data Mining dalam Mengelompokkan Rumah Tangga Kumuh di Perkotaan Berdasarkan Provinsi Menggunakan Algoritma K-Means,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 602, 2019, doi: 10.30645/senaris.v1i0.66.
- [17] G. Abdillah *et al.*, “PENERAPAN DATA MINING PEMAKAIAN AIR PELANGGAN UNTUK MENENTUKAN KLASIFIKASI POTENSI PEMAKAIAN AIR PELANGGAN BARU DI PDAM TIRTA RAHARJA MENGGUNAKAN ALGORITMA K-MEANS,” vol. 2016, no. Sentika, pp. 18–19, 2016.
- [18] V. No, W. Purba, and W. Siawin, “IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DAN PREDIKSI KARYAWAN YANG BERPOTENSI PHK DENGAN ALGORITMA K-MEANS CLUSTERING,” vol. 2, no. 2, 2019.