

APPLICATION OF K-MEANS ALGORITHM TO CLUSTER STUDENTS' READING PATTERNS IN THE DIGITAL AGE

ALGORITMA K-MEANS UNTUK MENGELOMPOKKAN POLA MEMBACA SISWA DI ERA DIGITAL

Yongky Permana Putra¹, Reflan Nuari²

^{1,2} Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: yongky_permana_putra@teknokrat.ac.id¹, reflan@teknokrat.ac.id²

Abstract - This study aims to group students' reading patterns in the digital era using the K-Means algorithm. This algorithm divides data into clusters, such as reading duration, type of reading, reading frequency, and devices used. Data were obtained through questionnaires distributed to 224 students of SMK Negeri 4 Bandar Lampung, with 214 valid data analysed after the preprocessing stage. The selection of vocational high school students as this study was based on previous journal references that examined reading patterns in PAUD to SMA students, so special attention is paid to vocational high school students, understanding reading patterns that have different needs compared to references with other levels of education. The clustering process produced four clusters with unique characteristics, reflecting differences in reading patterns based on the type of media used, intensity, and digital devices. The results of the study showed that clusters with high digital reading intensity can be directed to utilise e-books and online learning platforms optimally, while clusters with a preference for printed books require strengthening physical reading habits through literacy activities. With a Davies-Bouldin index value of -2.224, the quality produced is proven to be very good. These findings provide guidance for educators to develop technology-based education policies and personal approaches to improving student literacy. Designing learning programs with methods and student reading patterns to support the quality of education in the digital era.

Keywords - Reading Patterns, K-Means Algorithm, Digital Era, Learning, Clustering.

Abstrak - Penelitian ini bertujuan untuk mengelompokkan pola membaca siswa di era digital menggunakan algoritma K-Means. Algoritma ini membagi data ke dalam klaster, seperti durasi membaca, jenis bacaan, frekuensi membaca, serta perangkat yang digunakan. Data diperoleh melalui kuesioner yang disebarakan kepada 224 siswa SMK Negeri 4 Bandar Lampung, dengan 214 data valid yang dianalisis setelah tahap preprocessing. Pemilihan siswa SMK sebagai penelitian ini didasarkan referensi jurnal sebelumnya yang mengkaji pola membaca pada siswa PAUD hingga SMA, sehingga perhatian khusus pada siswa SMK, memahami pola membaca yang memiliki kebutuhan berbeda dibandingkan referen dengan jenjang pendidikan lainnya. Proses klasterisasi menghasilkan empat klaster dengan karakteristik unik, mencerminkan perbedaan pola membaca berdasarkan jenis media yang digunakan, intensitas, dan perangkat digital. Hasil penelitian menunjukkan bahwa klaster dengan intensitas membaca digital tinggi dapat diarahkan untuk memanfaatkan e-book dan platform pembelajaran daring secara maksimal, sementara klaster dengan preferensi buku cetak memerlukan penguatan kebiasaan membaca fisik melalui kegiatan literasi. Dengan nilai indeks Davies-Bouldin sebesar -2.224, kualitas yang dihasilkan terbukti sangat baik. Temuan ini memberikan panduan bagi pendidik untuk mengembangkan kebijakan pendidikan berbasis teknologi dan pendekatan personal dalam meningkatkan literasi siswa. Merancangan program pembelajaran dengan metode dengan pola membaca siswa untuk mendukung kualitas pendidikan di era digital.

Kata Kunci - Pola Membaca, Algoritma K-Means, Era Digital, Pembelajaran, Klasterisasi.

I. PENDAHULUAN

Membaca merupakan keterampilan fundamental yang menjadi dasar bagi pengembangan pengetahuan dan kemampuan berpikir secara kritis. Bagi siswa, kemampuan membaca yang baik tidak hanya meningkatkan prestasi akademik, tetapi juga memperluas wawasan dan membentuk karakter yang gemar belajar. Aktivitas membaca menjadi sangat penting dalam proses belajar dan merupakan langkah awal untuk meningkatkan minat membaca. Dengan kata lain, membaca adalah kegiatan kognitif yang melibatkan proses pengetahuan, pemahaman, dan kemampuan mengevaluasi [1]. Selain itu, membaca dapat meningkatkan fungsi otak, kemampuan berbahasa, dan pemahaman terhadap konteks bacaan. Keterampilan berbahasa yang baik memberikan manfaat signifikan bagi siswa dalam berbagai aspek kehidupan, termasuk pendidikan dan karier [2]. Proses membaca menjaga otak tetap aktif, meningkatkan daya ingat, serta memperkuat konsentrasi. Hal ini menunjukkan bahwa aktivitas membaca mampu melatih kemampuan berpikir kritis dan menganalisis informasi yang ditemukan dalam bacaan [3].

Perkembangan era digital telah membawa perubahan signifikan terhadap kebiasaan membaca di kalangan siswa. Akses yang mudah ke berbagai sumber bacaan digital, seperti e-book, artikel online, dan media sosial, telah menggeser preferensi membaca dari media cetak ke digital. Menurut penelitian, media digital dapat memberikan dampak positif maupun negatif terhadap minat baca siswa, tergantung pada bagaimana teknologi tersebut dimanfaatkan [4]. Kemudahan akses informasi melalui perangkat digital dapat meningkatkan minat baca, namun juga dapat menurunkan konsentrasi dan pemahaman jika tidak digunakan dengan bijak. Pilihan antara membaca buku cetak dan digital menjadi pertimbangan penting bagi siswa di era digital. Buku cetak menawarkan pengalaman membaca yang bebas dari gangguan digital dan dapat meningkatkan konsentrasi. Sebaliknya, bacaan digital menawarkan interaktivitas dan aksesibilitas yang lebih tinggi, namun rentan terhadap distraksi seperti notifikasi atau kecenderungan untuk beralih ke aplikasi lain. Oleh karena itu, Perubahan ini juga menghadirkan tantangan bagi institusi pendidikan dalam memahami dan memanfaatkan pola membaca siswa sebagai bagian dari strategi pembelajaran yang efektif.

Hipotesis dalam penelitian ini menyatakan bahwa pola membaca siswa dapat dikelompokkan ke dalam berbagai kategori, berdasarkan preferensi media, intensitas membaca, dan jenis bahan bacaan yang mereka pilih. Pertanyaan utama yang diajukan adalah: (1) Bagaimana K-Means dapat dimanfaatkan untuk mengelompokkan pola membaca siswa? (2) Apa saja faktor yang berpengaruh terhadap pengelompokan pola membaca siswa di era digital? Memahami pola membaca siswa di era digital membantu pendidik dalam menyesuaikan metode pengajaran yang sesuai dengan preferensi dan kebutuhan individu. Analisis pola membaca dapat memberikan wawasan penting untuk meningkatkan efektivitas pembelajaran [5], mengidentifikasi faktor-faktor yang mempengaruhi pemilihan bahan bacaan, seperti frekuensi bacaan siswa, jenis bacaan siswa, kemudahan akses, dan pengaruh lingkungan. Faktor-faktor tersebut meliputi intensitas penggunaan smartphone, yang dapat berdampak negatif terhadap minat baca jika tidak diimbangi dengan pengawasan yang tepat [6]. Dengan demikian, diperlukan metode yang mampu mengelompokkan pola membaca berdasarkan data yang ada untuk mendukung pembelajaran yang efektif.

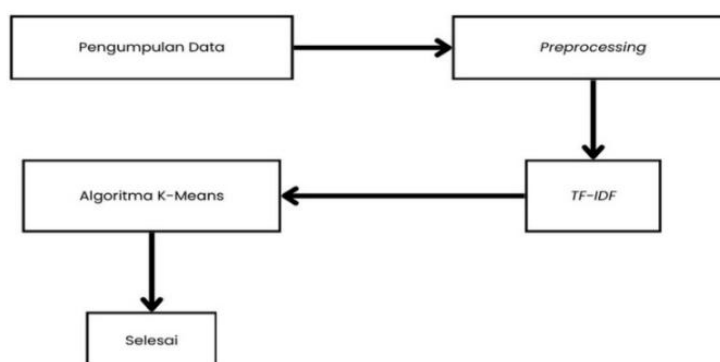
Algoritma K-Means merupakan metode pengelompokan yang membagi data menjadi kelompok-kelompok berdasarkan kesamaan ciri. Proses ini dimulai dengan menentukan jumlah kluster (K) yang diinginkan, kemudian memilih titik pusat (*centroid*) secara acak untuk setiap kluster. Setelah itu, setiap data akan dikelompokkan ke dalam kluster yang memiliki *centroid* terdekat. Langkah ini diulang hingga posisi *centroid* stabil dan tidak ada perubahan signifikan dalam pengelompokan data. Tujuan utama dari algoritma K-Means adalah meminimalkan variasi dalam kluster dan memaksimalkan variasi antar kluster, sehingga setiap kluster memiliki karakteristik yang homogen

dan berbeda dengan kluster lainnya. Metode ini banyak digunakan dalam berbagai bidang, seperti segmentasi pasar, analisis citra, dan pengelompokan dokumen, karena kesederhanaannya dan efisiensinya dalam menangani data besar[7].

Beberapa penelitian telah membahas pengelompokan pembelajaran siswa menggunakan algoritma K-Means. Salah satu penelitian berjudul “Penerapan Algoritma K-Means Clustering untuk Pengelompokan Performa Siswa pada Pembelajaran Bahasa Indonesia (Studi Kasus: SD Inpres Waingapu 3)” menggunakan aplikasi RapidMiner untuk uji coba. Hasil yang diperoleh menunjukkan performa *cluster* terhadap *centroid* 0 adalah 8.650, sementara *centroid* 1 adalah 8.281. Pernyataan ini menjelaskan rata-rata jarak tiap data ke *centroid* dari cluster masing-masing. Berdasarkan evaluasi menggunakan Davies-Bouldin Index (DBI), diperoleh nilai sebesar 0.078 pada cluster 2. Nilai DBI yang rendah tersebut mengindikasikan bahwa hasil pengelompokan cukup baik, dengan jarak antar-cluster yang terdefinisi dengan baik serta data dalam setiap cluster yang cenderung homogen [8]. Pada penelitian lain yang mengelompokkan mahasiswa berjudul “Implementasi Algoritma K-Means untuk Mengelompokkan Mahasiswa Program Studi Pendidikan Matematika Berdasarkan Sumber Belajarnya” menggunakan metode Elbow. Hasil penelitian ini menunjukkan pembagian karakteristik mahasiswa ke dalam tiga kelompok. Model ini memiliki nilai *cluster* sebesar 771.049 dan nilai *Silhouette* sebesar 0.615, yang menandakan bahwa hasil klasifikasi tersebut cukup baik [9]. Pembahasan ini memiliki tujuan untuk menerapkan algoritma K-Means dalam mengelompokkan pola membaca siswa di era digital. Dengan menganalisis data terkait preferensi membaca, intensitas, dan media yang digunakan, penelitian ini berkontribusi dalam menyediakan dasar empiris bagi pendidik dan pembuat kebijakan pendidikan. Informasi yang diperoleh diharapkan dapat membantu pendidik merancang intervensi pembelajaran yang lebih personal dan efektif, serta mendukung pengambilan kebijakan yang memperhatikan kebutuhan siswa di era digital.

II. SIGNIFIKANSI STUDI

Penelitian ini terdiri dari beberapa tahapan yaitu, kajian pustaka untuk mengumpulkan informasi terbaru, pengumpulan data kuisisioner dari siswa SMK Negeri 4 Bandar Lampung, dan melakukan persiapan awal untuk analisis data, termasuk seleksi, pembersihan, dan transformasi data. Selanjutnya, dilakukan pembobotan menggunakan metode TF-IDF. Tahap berikutnya adalah klasterisasi menggunakan metode Algoritma K-Means Clustering untuk mengelompokkan siswa berdasarkan pola membaca mereka. Penelitian ini bertujuan untuk mengembangkan solusi praktis dalam mengatasi permasalahan yang ada, khususnya dalam memahami dan mengelompokkan pola membaca siswa SMK Negeri 4 Bandar Lampung Berikut ini tahapan penelitian yang akan dilakukan, yaitu:



Gambar 1. Metode Penelitian

Penelitian ini berpeluang memberikan pemahaman mengenai cara menyesuaikan pendidikan digital dengan preferensi dan pola membaca masing-masing individu. Dengan memahami pola membaca siswa melalui klusterisasi, penelitian ini membuka peluang untuk pendekatan yang lebih personal dalam pendidikan digital, yang mendorong pengembangan metode pembelajaran berbasis data yang lebih efektif dan relevan di masa depan.

A. Studi Pustaka

Penelitian ini diawali melakukan studi pustaka yang tinjauan mendalam terhadap studi-studi sebelumnya yang telah menggunakan algoritma K-Means untuk menganalisis pola membaca. Sebagian besar penelitian sebelumnya mengkaji pola membaca pada siswa PAUD hingga SMA, Penelitian ini memberikan perhatian khusus pada siswa SMK, mengingat pentingnya memahami pola membaca mereka yang memiliki kebutuhan berbeda dibandingkan referen dengan jenjang pendidikan lainnya. Studi pustaka dilakukan melalui jurnal ilmiah, buku referensi, dan penelitian terdahulu yang relevan untuk memahami penerapan algoritma K-Means, mengidentifikasi kelebihan dan keterbatasannya, serta merumuskan dasar teori yang kuat. Penelitian ini diharapkan dapat mengembangkan kebiasaan membaca siswa SMK sehingga minat bacanya meningkat, terutama terhadap bacaan yang relevan dengan keterampilan kerja, sehingga dapat mempersiapkan mereka lebih baik untuk menghadapi tuntutan dunia industri. Penelitian ini dapat mempengaruhi pendekatan masa depan terhadap pendidikan digital dengan mendorong pembelajaran berbasis teknologi yang lebih sesuai dengan pola membaca siswa, baik melalui media digital maupun media cetak. Dengan pemahaman yang lebih baik tentang pola membaca siswa, pendidikan dapat lebih bebas beradaptasi dengan perkembangan teknologi digital yang pesat, memastikan bahwa siswa terlibat dalam proses belajar dengan cara yang paling efektif dan relevan.

B. Pengumpulan Data

Proses pengumpulan data dalam penelitian ini dilakukan dengan menyebarkan kuesioner kepada siswa-siswi SMK Negeri 4 Bandar Lampung yang menjadi subjek penelitian. Penyebaran kuesioner dibagi ke dalam kelas 10 dan 11 serta mencakup 10 jurusan. Kuesioner dirancang secara sistematis untuk memperoleh informasi mendetail mengenai kebiasaan membaca siswa dalam konteks era digital. Informasi yang dikumpulkan meliputi durasi membaca, jenis bacaan (buku, e-book, artikel online, media sosial), frekuensi membaca (harian atau mingguan), dan perangkat yang digunakan. Kuesioner dirancang menggunakan bahasa yang sederhana dan mudah dimengerti, sehingga responden dapat memberikan jawaban secara jelas tanpa mengalami kebingungan. Pengumpulan data dilakukan secara langsung di lingkungan sekolah untuk memastikan partisipasi yang maksimal. Semua jawaban siswa akan diproses dengan menjaga kerahasiaan data, serta meningkatkan akurasi dan kejujuran respon. Pendekatan ini diharapkan dapat menghasilkan data yang tidak hanya relevan, tetapi juga mencerminkan pola membaca siswa secara menyeluruh dan terkini, sesuai dengan perkembangan teknologi dan kebiasaan belajar di era digital.

C. Pengolahan Data Set

Menjamin bahwa data yang digunakan dalam analisis memiliki kualitas yang baik dilakukan melalui pembersihan data. Proses ini meliputi penghapusan atribut yang kurang relevan atau tidak diperlukan, penanganan data yang tidak lengkap, serta langkah-langkah lain yang bertujuan untuk memastikan data berada dalam kondisi optimal sebelum dianalisis, sehingga hasil analisis dapat dipercaya dan dapat diandalkan [10]. Berikut tahapan yang akan di lalui:

1. Seleksi Data

Tahapan untuk menentukan data yang sesuai dari kumpulan data yang ada dengan memperhatikan kriteria tertentu seperti relevansi terhadap variabel penelitian, kelengkapan data, dan konsistensi.

Data yang tidak relevan atau tidak mendukung tujuan penelitian akan disaring untuk meningkatkan efisiensi analisis dan meminimalkan kesalahan.

2. Pembersihan Data

Pembersihan Data adalah tahap ini melibatkan pembersihan data dari gangguan yang dapat mempengaruhi kualitas analisis, Seperti data yang hilang (missing values), inkonsistensi data, atau adanya duplikasi. Data yang tidak valid akan diperbaiki atau dihapus agar dataset menjadi lebih akurat dan siap untuk dianalisis, menjamin hasil yang lebih reliabel.

3. Data Transformation

Data Transformation adalah tahap ini mengubah data ke format atau struktur yang lebih sesuai untuk analisis lebih lanjut. Proses ini bisa mencakup normalisasi, pengelompokan (binning), atau perubahan data ke bentuk yang lebih sederhana dan terorganisir. Transformasi ini penting untuk memastikan data dapat diproses secara efisien oleh algoritma seperti K-Means.

D. Klasterisasi

Klasterisasi adalah metode analisis data yang mengelompokkan objek-objek berdasarkan kesamaan karakteristik, sehingga objek dalam satu kelompok memiliki kesamaan yang lebih besar dibandingkan dengan objek di kelompok lain [11]. Metode pengelompokan K-Means menjadi salah satu algoritma yang paling banyak digunakan untuk tujuan ini. Algoritma ini berfungsi dengan membagi data ke dalam sejumlah klaster yang telah ditentukan, di mana setiap data dimasukkan ke klaster yang memiliki titik pusat terdekat, dan proses ini diulang hingga centroid mencapai kestabilan. [12]. Dalam penelitian ini, klasterisasi digunakan untuk mengelompokkan siswa berdasarkan pola bacaan mereka yang telah dianalisis menggunakan metode TF-IDF. Setiap klaster menggambarkan kelompok siswa dengan preferensi membaca yang serupa, yang memberikan informasi berharga bagi pendidik untuk menyesuaikan materi ajar dengan minat siswa. Keputusan klaster ini juga dapat digunakan oleh pembuat kebijakan untuk merancang kebijakan literasi yang lebih tepat sasaran, dan membantu siswa belajar dengan materi yang lebih sesuai dengan minat mereka.

E. TF-IDF

Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik dalam pengambilan informasi dan penambangan teks untuk mengambil vektorisasi atau ekstraksi teks [13]. *Term Frequency* (TF) mengukur seberapa sering suatu kata muncul dalam dokumen tertentu, sementara *Inverse Document Frequency* (IDF) menilai kelangkaan kata tersebut di seluruh dokumen yang ada. Dengan menggabungkan nilai TF dan IDF, metode ini memberikan skor tinggi pada kata-kata yang sering muncul dalam satu dokumen namun jarang muncul di dokumen lain. Hal ini memungkinkan identifikasi kata-kata yang memiliki peran penting dalam konteks dokumen tersebut. Dalam penelitian ini, TF-IDF diterapkan untuk menganalisis teks atau bacaan yang dikonsumsi siswa berdasarkan pertanyaan yang terdapat dalam kuisisioner, guna mengidentifikasi preferensi membaca mereka. Berbagai penelitian sebelumnya telah membuktikan keefektifan metode ini dalam klasifikasi teks dan analisis sentimen, seperti pada klasterisasi buku untuk mengevaluasi peningkatan minat baca siswa [14] serta penerapan text mining berbasis TF-IDF untuk klasifikasi genre novel [15]. Selain itu, metode ini juga diterapkan dalam pemberian bobot pada judul penelitian dosen sebagai dasar untuk klasifikasi dengan algoritma K-NN. Dengan demikian, TF-IDF terbukti sebagai cara yang efektif dalam memahami dan menganalisis pola membaca siswa di era digital.

F. K-Means Clustering

Metode K-Means *Clustering* merupakan teknik pengelompokan data yang membagi data kedalam kelompok berdasarkan kesamaan antara objek-objek data [16]. Tujuan dari K-Means adalah untuk mengecilkan fungsi objektif yang dirancang dalam proses pengelompokan, berusaha untuk mengurangi perubahan dalam satu kelompok dan meningkatkan perbedaan antara kelompok-kelompok yang ada [17]. Dalam penelitian ini, peneliti telah memilih untuk menggunakan 4 kluster. Proses K-Means Clustering terdiri dari tiga tahap, yaitu: pertama, pengumpulan informasi mengenai pola baca siswa dari kelas 10 dan 11; kedua, tahap pra-pemrosesan data yang mencakup pemilihan, pembersihan, dan transformasi data; dan ketiga, pemodelan dengan algoritma K-Means, yang meliputi penerapan algoritma K-Means di aplikasi RapidMiner, perancangan diagram proses, serta pelaksanaan diagram proses tersebut.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh melalui pengumpulan data menggunakan kuesioner yang disebarikan kepada siswa-siswi SMK Negeri 4 Bandar Lampung sebagai subjek penelitian. Penyebaran kuesioner dilakukan di kelas 10 dan 11, mencakup 10 jurusan. Dataset ini terdiri dari 224 baris data yang mencerminkan jawaban responden terhadap lima subjek utama pertanyaan. Pertanyaan tersebut meliputi durasi membaca, yaitu seberapa lama siswa menghabiskan waktu untuk membaca; jenis bacaan yang diakses, seperti buku cetak, e-book, artikel online, atau konten media sosial; serta frekuensi membaca, apakah dilakukan secara harian atau mingguan. Selain itu, dataset juga mencakup informasi mengenai perangkat yang digunakan untuk membaca, seperti buku cetak, smartphone, tablet, atau laptop. Data yang terkumpul melalui kuesioner ini telah dirancang untuk menangkap pola membaca siswa secara mendalam dan mencakup berbagai aspek yang relevan dengan penelitian. Hasil dari pengumpulan data ini diharapkan dapat memberikan kontribusi signifikan terhadap penelitian terkait pengelompokan pola membaca siswa di era digital.

B. Preprocessing

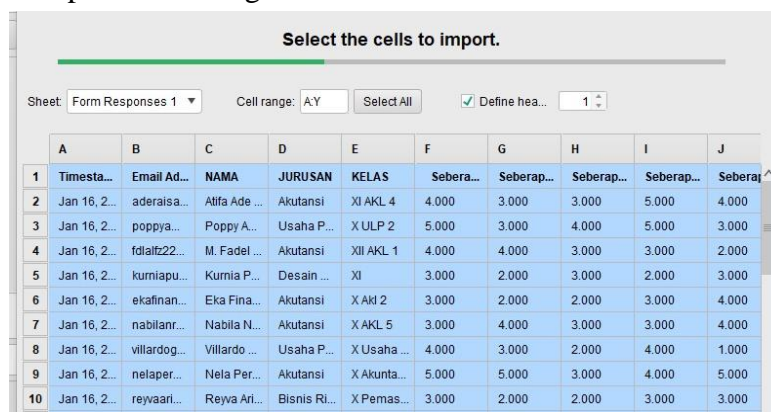
data adalah langkah pertama dalam proses data mining yang bertujuan untuk mengubah data mentah menjadi format yang lebih terstruktur dan siap digunakan dalam tahap analisis berikutnya. Tahap ini melibatkan berbagai proses seperti pembersihan, penyaringan, dan pengintegrasian data untuk memastikan kualitas data yang optimal. Preprocessing bertujuan untuk menangani tiga masalah utama yang sering ditemukan dalam data: *missing values*, *data noise*, dan ketidakkonsistenan data. *Missing values* merujuk pada data yang tidak lengkap atau hilang, yang sering kali disebabkan oleh kesalahan input data atau kegagalan dalam proses pengumpulan informasi. *Data noise* adalah data yang tidak valid atau tidak relevan, yang dapat muncul akibat kesalahan pencatatan, pelabelan, atau masalah teknis lainnya selama pengumpulan data. Selain itu, data yang tidak konsisten, seperti perbedaan format atau skema dalam penyajian informasi, juga dapat menjadi tantangan yang mempengaruhi hasil analisis. Dengan preprocessing, data dapat disiapkan untuk memberikan hasil analisis yang lebih akurat dan relevan.

C. Pemodelan Menggunakan Algoritma K-Means

Tahap ini menggunakan aplikasi RapidMiner untuk mengelompokkan pola membaca siswa. Prosesnya mencakup impor data, perancangan diagram proses, dan penerapan algoritma hingga menghasilkan cluster yang relevan. Setiap langkah dirancang secara sistematis untuk memastikan analisis yang akurat.

1. Penerapan Algoritma K-Means menggunakan Aplikasi RapidMiner

Proses pemodelan menggunakan Algoritma K-Means dalam aplikasi RapidMiner dilakukan melalui beberapa tahap, termasuk perancangan diagram proses dan pengolahan output dari eksekusi. Langkah pertama merancang diagram alur, dengan mengimpor data. Pengguna dapat memanfaatkan fitur "Import Data" pada Repository View untuk memuat data, memilih file yang diinginkan dari My Computer, serta menentukan kolom yang akan diproses dengan menetapkan rentang sel. Selanjutnya, format kolom diatur dengan memilih tipe data atau atribut yang tepat. Setelah semua pengaturan selesai, data disimpan dalam local repository. Proses impor data ke dalam RapidMiner dapat dilihat pada ilustrasi gambar 2.



	A	B	C	D	E	F	G	H	I	J
1	Timesta...	Email Ad...	NAMA	JURUSAN	KELAS	Sebera...	Seberap...	Seberap...	Seberap...	Seberap...
2	Jan 16, 2...	aderaisa...	Atifa Ade ...	Akutansi	XI AKL 4	4.000	3.000	3.000	5.000	4.000
3	Jan 16, 2...	poppya...	Poppy A...	Usaha P...	X ULP 2	5.000	3.000	4.000	5.000	3.000
4	Jan 16, 2...	fdlalfz22...	M. Fadel ...	Akutansi	XII AKL 1	4.000	4.000	3.000	3.000	2.000
5	Jan 16, 2...	kumiapu...	Kurnia P...	Desain ...	XI	3.000	2.000	3.000	2.000	3.000
6	Jan 16, 2...	ekalfinan...	Eka Fina...	Akutansi	X AKI 2	3.000	2.000	2.000	3.000	4.000
7	Jan 16, 2...	nabilanr...	Nabila N...	Akutansi	X AKL 5	3.000	4.000	3.000	3.000	4.000
8	Jan 16, 2...	villardog...	Villardo ...	Usaha P...	X Usaha ...	4.000	3.000	2.000	4.000	1.000
9	Jan 16, 2...	nelaper...	Nela Per...	Akutansi	X Akunta...	5.000	5.000	3.000	4.000	5.000
10	Jan 16, 2...	reyyaari...	Reyva Ari...	Bisnis Ri...	X Pemas...	3.000	2.000	2.000	3.000	3.000

Gambar 2. Tampilan Import Data

2. Membuat Diagram

Penyusunan diagram proses dalam analisis data menggunakan RapidMiner dilakukan dengan cara menyisipkan operator dan menghubungkannya di lokasi *Process View*. Diagram ini bertujuan untuk menggambarkan secara visual alur kerja analisis data, mulai dari tahap awal hingga pelaksanaan algoritma K-Means.

a. Menambahkan Operator

Langkah pertama dalam penyusunan diagram proses adalah menambahkan operator yang dibutuhkan ke area Process View di RapidMiner. Operator-operator ini dipilih dari panel operator yang menyediakan berbagai fungsi untuk analisis data. Operator yang relevan untuk penerapan algoritma K-Means, seperti operator untuk membaca data, membersihkan data, atau menerapkan algoritma pengelompokan, dipilih secara selektif dan diseret ke dalam area Process View. Penambahan operator dilakukan secara hati-hati untuk memastikan bahwa setiap langkah dalam alur analisis terwakili dengan baik. Operator ini berfungsi sebagai komponen utama dalam membangun alur kerja yang terstruktur, memungkinkan proses analisis berjalan sesuai dengan kebutuhan.

b. Menghubungkan Operator

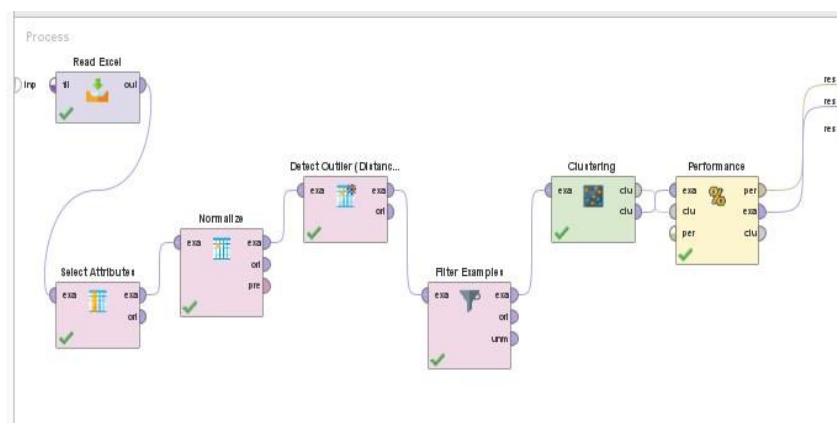
Setelah menambahkan operator yang diperlukan ke dalam area Process View, tahap selanjutnya adalah menyambungkan operator tersebut sesuai dengan alur analisis yang telah dirancang. Penghubungan dilakukan dengan menarik garis koneksi antara operator satu dengan lainnya. Garis koneksi ini menunjukkan aliran data atau proses dari satu tahap ke tahap berikutnya. Proses penghubungan ini memastikan bahwa data diproses secara berurutan dan logis, mulai dari tahap awal hingga tahap akhir analisis dengan algoritma K-Means. Dengan penghubungan yang tepat, diagram proses akan menjadi panduan yang jelas dalam pelaksanaan analisis data pola membaca siswa.

Melalui penyusunan diagram proses ini, alur kerja analisis dapat diatur secara sistematis dan mudah dipahami. Diagram yang dihasilkan menjadi panduan untuk mengimplementasikan algoritma K-Means pada dataset pola membaca siswa. Detail mengenai operator-operator yang digunakan untuk membangun desain proses metode K-Means dapat dilihat pada Tabel 1, yang memberikan deskripsi lebih rinci tentang fungsi dan peran masing-masing operator dalam proses

TABEL 1
OPERATOR PROSES KLASTERISASI ALGORITMA K-MEANS

Operator dan Fungsi	Parameter
Import Data atau Read XLS digunakan untuk memuat dataset yang ada di file Ms Excel agar dapat digunakan sebagai data dalam proses.	Lokasi file
Select Attributes untuk mengidentifikasi atribut dari tabel pertanyaan	Select Attributes
Normalize berfungsi menormalisasikan data agar data yang di ambil sesuai dengan kebutuhan pengguna.	Normalize
Detect Outliner (Distances) berfungsi mengidentifikasi N pencilan dalam ExampleSet yang diberikan berdasarkan jarak ke k tetangga terdekat mereka. Variabel N dan K dapat ditentukan melalui parameter.	Detect Outliner (Distances)
Filter Example berfungsi memilih Contoh mana dari ExampleSet yang disimpan dan Contoh mana yang dihapus.	Filter Example = outlier (false)
K-Means berfungsi untuk penilaian performa pada suatu model.	K = 4 atau mengikuti parameter yang diperlukan.
Performance untuk evaluasi kinerja metode pengelompokan berbasis centroid. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan centroid cluster.	Performance

Diagram ini menjadi panduan utama dalam mengimplementasikan algoritma K-Means pada dataset pola membaca siswa. Detail mengenai operator-operator yang akan digunakan untuk membangun desain proses metode K-Means dapat dilihat melalui Tabel 1, yang memberikan deskripsi lebih rinci tentang fungsi dan peran masing-masing operator dalam proses analisis. Hasil dari pembentukan diagram proses yang mencerminkan alur kerja analisis ini dapat dilihat pada Gambar 3, yang memberikan gambaran visual lengkap mengenai tahapan dan koneksi antaroperator secara menyeluruh.



Gambar 3. Hasil pembuatan diagram alur proses

3. Eksekusi Diagram Proses

Setelah diagram proses selesai dibuat, langkah berikutnya adalah menjalankan diagram tersebut. Proses ini bertujuan untuk melakukan analisis menggunakan algoritma K-Means dan memfilter data berdasarkan nilai outlier, di mana hanya data dengan status *outlier* "false" yang dipertimbangkan. Hasil dari proses eksekusi ini mencakup informasi penting yang digunakan untuk mengevaluasi dan memahami hasil klusterisasi.

Hasil yang diperoleh mencakup data klusterisasi, yang menunjukkan pengelompokan objek dalam cluster 0-3 menggunakan algoritma K-Means. Selain itu, analisis statistik yang relevan, seperti nilai rata-rata, deviasi standar, dan statistik lainnya, juga disertakan [18]. Proses ini juga menghasilkan cluster 0 yang memiliki intensitas membaca melalui perangkat digital dengan frekuensi sedang, cluster 1 yang berupa siswa memiliki intensitas membaca tinggi melalui perangkat digital, cluster 2 yang berupa siswa memiliki intensitas membaca menggunakan kombinasi media cetak dan digital dengan durasi tinggi, dan cluster 3 yang memiliki intensitas membaca melalui buku cetak dengan durasi rendah. Tampilan hasil eksekusi proses ini dapat dilihat pada gambar 4.

Row No.	id	cluster	outlier	Seberapa c...	Seberapa k...	Seberapa ef...	Seberapa p...	Seberapa s...	Sebi
1	1	cluster_0	false	0.566	-0.351	-0.094	1.745	0.819	-0.06
2	2	cluster_0	false	1.739	-0.351	1.077	1.745	-0.235	2.03
3	3	cluster_2	false	0.566	0.878	-0.094	-0.521	-1.289	0.98
4	4	cluster_1	false	-0.608	-1.580	-0.094	-1.654	-0.235	-1.11
5	5	cluster_2	false	-0.608	-1.580	-1.265	-0.521	0.819	-0.06
6	6	cluster_0	false	-0.608	0.878	-0.094	-0.521	0.819	-0.06
7	7	cluster_1	false	0.566	-0.351	-1.265	0.612	-2.343	-0.06

Gambar 4. Visualisasi data untuk algoritma K-Means menggunakan RapidMiner

Setelah menyelesaikan proses klusterisasi menggunakan metode K-Means, langkah selanjutnya adalah menganalisis data statistik yang telah diperoleh [19]. Analisis statistik yang dihasilkan oleh algoritma K-Means memberikan pemahaman penting mengenai karakteristik dari setiap kluster yang terbentuk. Informasi ini sangat membantu dalam memahami struktur data di dalam kluster, sebagaimana yang ditampilkan pada Gambar 5.

Name	Type	Missing	Statistics	Filter (23 / 23 attributes):
id	Integer	0	Min: 1, Max: 214, Average: 107.500	
cluster	Nominal	0	Least: cluster_3 (27), Most: cluster_0 (63), Values: cluster_0 (63)	
outlier	Binominal	0	Negative: false, Positive: true, Values: false (214), true	

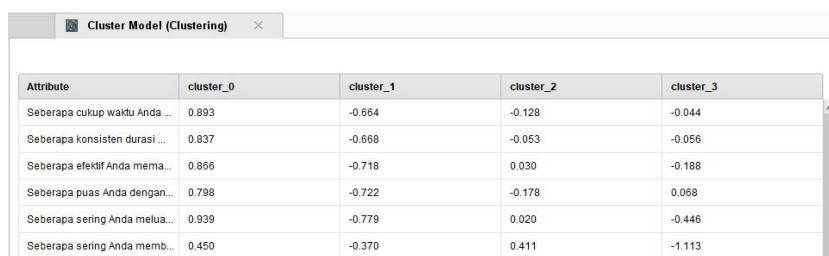
Gambar 5. Tampilan statistic data Algoritma K-Means

Setelah tampilan statistik data dari algoritma K-Means, tahap berikutnya adalah mengevaluasi kinerja model klusterisasi dengan menggunakan vektor kinerja K-Means [20]. Tampilan ini menyajikan data penting mengenai penilaian kinerja dari model klusterisasi yang dihasilkan. Penilaian performa ini dapat diamati pada gambar 6 (a).

PerformanceVector (Performance)	PerformanceVector (Performance)
<p>Avg. within centroid distance</p> <p>Avg. within centroid distance: -11.658</p>	<p>Davies Bouldin</p> <p>Davies Bouldin: -2.224</p>

Gambar 6. (a) Tampilan performance vector K-Means (b) Tampilan Hasil dari Davies Bouldin

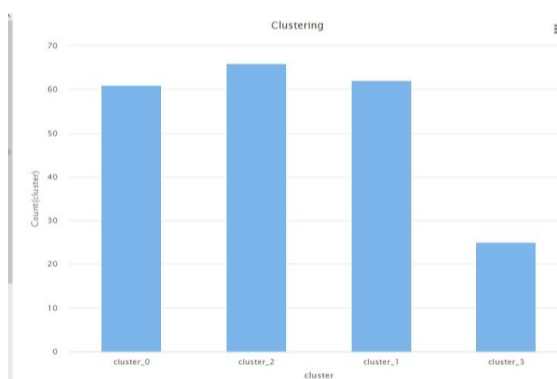
Hasil evaluasi kualitas kluster berdasarkan indeks Davies-Bouldin menghasilkan nilai sebesar -2.224. Indeks ini berfungsi untuk mengukur kualitas kluster dengan cara memperhatikan jarak antar elemen dalam satu kluster yang kecil (menunjukkan tingkat kesamaan yang tinggi di dalam kluster) dan jarak antar kluster yang besar (menunjukkan perbedaan yang signifikan antar kluster). Nilai indeks yang semakin rendah menunjukkan kualitas kluster yang semakin baik. Nilai -2.224 yang diperoleh, sebagaimana ditunjukkan pada Gambar 6 (b), mengindikasikan bahwa kluster yang terbentuk memiliki kesamaan yang tinggi di dalam kluster dan perbedaan signifikan antar-kluster. Hal ini menandakan bahwa kualitas kluster yang dihasilkan dalam analisis pengaduan masyarakat sangat baik, karena nilainya mendekati nol.



Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Seberapa cukup waktu Anda ...	0.893	-0.664	-0.128	-0.044
Seberapa konsisten durasi ...	0.837	-0.668	-0.053	-0.056
Seberapa efektif Anda mema...	0.866	-0.718	0.030	-0.188
Seberapa puas Anda dengan...	0.798	-0.722	-0.178	0.068
Seberapa sering Anda melua...	0.939	-0.779	0.020	-0.446
Seberapa sering Anda memb...	0.450	-0.370	0.411	-1.113

Gambar 7. Tampilan Centroid Table

Gambar 7 menampilkan tabel berisi nilai-nilai centroid untuk setiap cluster dalam tabel. Tabel ini memberikan informasi mengenai lokasi pusat kluster yang diperoleh melalui proses klusterisasi dengan menggunakan algoritma K-Means [21]. Informasi ini berfungsi untuk mengenali karakteristik unik dari setiap kluster dan mendukung analisis hasil pengelompokan secara lebih detail. Dengan memahami posisi pusat dari masing-masing kluster, pengguna dapat melihat distribusi data di sekitar pusat tersebut serta mengidentifikasi perbedaan sifat antar kluster. Hal tersebut akan membantu pengguna dalam membuat keputusan yang lebih tepat dalam menganalisis data dan mengidentifikasi pola di setiap kluster.



Gambar 8. Tampilan Diagram Batang Cluster

Pada gambar 8, terdapat diagram batang dari setiap cluster yang telah di proses. Dari tabel tersebut, terlihat bahwa cluster 2 memiliki intensitas tertinggi dari yang lain, sementara untuk cluster 3 memiliki intensitas paling rendah dari cluster lainnya. Grafik diagram batang ini memberikan gambaran jelas tentang distribusi daya ke dalam masing masing kluster, sehingga membantu sekolah untuk pola membaca yang efektif dan distribusi data dalam analisis kluster.

Berdasarkan hasil evaluasi kualitas kluster menggunakan indeks Davies-Bouldin menunjukkan nilai -2.224, dan tampilan grafik diagram yang menunjukan cluster 2 memiliki intensitas tertinggi. Hal memberi kesempatan kepada sekolah untuk mengambil pilihan yang lebih akurat saat mengevaluasi informasi dan menemukan tren yang muncul dalam setiap kelompok.

IV. KESIMPULAN

Berdasarkan penelitian ini, penerapan algoritma K-Means terbukti efektif dalam mengelompokkan pola membaca siswa di era digital. Dari total 224 data kuisioner, proses preprocessing menghasilkan 214 data valid, sementara 10 data lainnya dianggap outlier dan dikeluarkan untuk menjaga akurasi analisis. Hasil klasterisasi membagi siswa ke dalam empat klaster dengan karakteristik unik. Klaster 0 mencakup siswa dengan intensitas membaca melalui perangkat digital yang sedang, klaster 1 mencakup siswa dengan intensitas membaca digital yang tinggi, klaster 2 terdiri dari siswa yang menggunakan kombinasi media cetak dan digital dengan durasi tinggi, dan klaster 3 mencakup siswa yang membaca buku cetak dengan durasi rendah. Setiap klaster memberikan wawasan penting untuk mendukung kebutuhan spesifik siswa. Siswa di klaster 0 dapat diberikan pelatihan literasi digital untuk meningkatkan efektivitas membaca melalui perangkat. Klaster 1 dapat difasilitasi dengan akses lebih luas ke konten digital berkualitas seperti e-book atau platform pembelajaran daring. Sementara itu, klaster 2 dapat didukung dengan perpustakaan hybrid yang mengintegrasikan media cetak dan digital, dan klaster 3 memerlukan program peningkatan kebiasaan membaca buku cetak melalui kegiatan seperti sesi membaca bersama atau penyediaan buku fisik yang menarik. Hasil analisis ini memberikan wawasan strategis untuk pengambilan keputusan berbasis data, seperti merancang program literasi, meningkatkan sarana membaca, dan mengidentifikasi kelompok siswa yang memerlukan intervensi khusus. Efektivitas algoritma K-Means juga dibuktikan dengan nilai indeks Davies-Bouldin sebesar -2.224, yang menunjukkan kualitas klaster yang sangat baik. Penelitian ini menyimpulkan bahwa algoritma K-Means mampu membantu sekolah merancang kebijakan yang tepat sasaran dalam mendukung pola membaca siswa. Penelitian selanjutnya disarankan untuk mengembangkan model klasterisasi menggunakan algoritma lain, terutama dalam konteks siswa SMK.

REFERENSI

- [1] L. F. Az Zahroh, N. Rahaningsih, and R. D. Dana, "Klasterisasi Data Kegemaran Membaca Menggunakan Algoritma K-Means Di Sma Al-Islam Cirebon," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 3, pp. 2692–2698, 2024, doi: 10.36040/jati.v8i3.9543.
- [2] M. E. Simbolon, A. Marini, and M. Nafiah, "PENGARUH LITERASI DIGITAL TERHADAP MINAT BACA SISWA SEKOLAH DASAR," *J. Cakrawala Pendas,* vol. 8, no. 2, pp. 532–542, 2022, [Online]. Available: <http://dx.doi.org/10.31949/jcp.v8i2.2449>
- [3] S. Arifin, "PENGARUH BUDAYA MEMBACA TERHADAP KEMAMPUAN BERPIKIR KRITIS SISWA DI SMP ISLAM AL SYUKRO UNIVERSAL CIPUTAT TAHUN 2019," *Andragogi J. Pendidik. Islam,* vol. 1, no. 1, pp. 1–21, 2019, [Online]. Available: <https://jurnalptiq.com/index.php/andragogi/article/view/54/54>
- [4] A. F. Zuhria, M. D. Kurnia, J. Jaja, and C. Hasanudin, "Dampak Era Digital terhadap Minat Baca Remaja," *JUBAH RAJA J. Bahasa, Sastra, dan Pengajaran,* vol. 1, no. 2, p. 22, 2020.
- [5] M. Anak, F. M. Sarimole, and M. A. Septiansyah, "Analisis Pola Kinerja Anak dalam Tes Membaca untuk Dini Menggunakan Algoritma K-Means Clustering di PAUD," vol. 5, no. 3, pp. 3054–3064, 2024.
- [6] S. Ilmu, "DAMPAK INTENSITAS PENGGUNAAN SMARTPHONE TERHADAP MINAT BACA PESERTA DIDIK," vol. 3157, pp. 1–14.
- [7] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics,* vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

- [8] A. Of *et al.*, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PERFORMA SISWA PADA PEMBELAJARAN BAHASA INDONESIA (Studi Kasus : SD INPRES WAINGAPU 3),” pp. 523–537, 2024.
- [9] N. A. Rizki, K. Kurniawan, I. K. Hasan, and N. Sampe, “Implementasi Algoritma K-Means Untuk Mengelompokkan Mahasiswa Berdasarkan Sumber Belajarnya,” *Metik J.*, vol. 7, no. 2, pp. 62–67, 2023, doi: 10.47002/metik.v7i2.584.
- [10] M. Ondang and O. Kembuan, “Analisis Pola Lapangan Usaha Per Daerah Menggunakan Algoritma K-Means Clustering Pada Kabupaten Minahasa Selatan,” *Innov. J. Soc. Sci. ...*, vol. 3, pp. 7672–7686, 2023, [Online]. Available: <http://j-innovative.org/index.php/Innovative/article/view/7108%0Ahttps://j-innovative.org/index.php/Innovative/article/download/7108/5015>
- [11] N. H. Ely Kurniawan, “PENERAPAN ALGORITMA K-MEANS UNTUK MELAKUKAN KLASTERISASI PADA PERINGKASAN TEKS,” pp. 514–520, 2024.
- [12] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, “Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means,” *J. Inf. dan Teknol.*, vol. 3, pp. 103–108, 2021, doi: 10.37034/jidt.v3i3.130.
- [13] R. K. Bania, “COVID-19 Public Tweets Sentiment Analysis using TF-IDF and Inductive Learning Models,” *Infocomp*, vol. 19, no. 2, pp. 23–41, 2020.
- [14] D. L. Karputri and W. Yustanti, “Analisis Klastering Buku sebagai Evaluasi untuk Peningkatan Minat Baca Perpustakaan SMAN 1 Grogol,” *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 3, no. 3, pp. 94–101, 2022.
- [15] T. W. D. Sari, “Penerapan Text Mining Dengan Menggunakan Algoritma TF-IF Untuk Klasifikasi Genre Novel,” *Pelita Inform. Inf. dan Inform.*, vol. 10, no. 1, pp. 29–37, 2021, [Online]. Available: <http://www.stmik-budidarma.ac.id/ejurnal/index.php/pelita/article/view/3142>
- [16] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, “Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang),” *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [17] K. Ademariana, A. Aristoteles, F. R. Lumbanraja, and R. Andrian, “Clustering K-Means Jenis Kata Pada Laporan Kegiatan Kuliah Kerja Nyata (Kkn) Universitas Lampung Menggunakan Word2Vec,” *J. Pepadun*, vol. 2, no. 2, pp. 221–228, 2021, doi: 10.23960/pepadun.v2i2.64.
- [18] Sumarjono and M. A. Saputra, “Penerapan Data Mining Untuk Prediksi Ujuk Kerja Operasional Penambangan Batubara,” *Tpt Perhapi*, pp. 245–258, 2022, [Online]. Available: <https://www.prosiding.perhapi.or.id/index.php/prosiding/article/view/293%0Ahttps://www.prosiding.perhapi.or.id/index.php/prosiding/article/view/293/369>
- [19] V. Virtusena, A. Johar, and A. Wijanarko, “Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan Algoritme K-Means (Studi Kasus: Fakultas Teknik Universitas Bengkulu),” *Rekursif J. Inform.*, vol. 9, no. 2, pp. 206–225, 2021, doi: 10.33369/rekursif.v9i2.17073.
- [20] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020, doi: 10.25105/jti.v10i3.8428.
- [21] A. Primandana, S. Adinugroho, and C. Dewi, “Optimasi Penentuan Centroid pada Algoritme K-Means Menggunakan Algoritme Pillar (Studi Kasus: Penyandang Masalah Kesejahteraan Sosial di Provinsi ...),” ... *Teknol. Inf. dan Ilmu ...*, vol. 3, no. 11, pp. 10678–10683, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/6748/3264>