

Penerapan Metode *Clustering* Dengan Algoritma *K-means* Pada Pengelompokan Data Calon Siswa Baru

Muhammad Norshahlan¹, Hendra Jaya², Rini Kustini³

^{1,2,3} Sistem Informasi, STMIK Triguna dharm

Email: ¹alansnulien@gmail.com, ²hendrajaya.tgd73@gmail.com, ³rinikustini.tgd@gmail.com

Email Penulis Korespondensi: alansnulien@gmail.com

Abstrak

Proses Penerimaan Siswa baru di setiap sekolah menghasilkan data siswa yang sangat berlimpah berupa data alamat dan data lainnya. Hal tersebut terjadi secara berulang dikarenakan penerimaan siswa baru di setiap tahunnya. Penelitian ini bertujuan untuk melakukan pengelompokan data siswa sekolah Harapan Bangsa dengan memanfaatkan proses data mining dengan menggunakan metode clustering. Adapun Algoritma yang digunakan adalah Algoritma K-Means Clustering. Algoritma K-Means merupakan Algoritma pengelompokan iterative yang melakukan partisi set data kedalam sejumlah K cluster yang sudah ditetapkan diawal. Implementasi dengan menggunakan aplikasi berbasis website digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan untuk tahapan clustering adalah jurusan, asal sekolah, dan tahun lahir. Cluster siswa yang terbentuk adalah 2 cluster, yaitu 47 items cluster pertama, 23 items cluster kedua. Adapun tujuan penelitian ini adalah salah satu dasar pengambilan keputusan untuk menentukan strategi dalam mempromosikan sekolah. Dapat disimpulkan bahwa proses clustering mencapai 3 kali iterasi untuk mendapatkan hasil clustering akhir, sehingga dapat disimpulkan bahwa peminat terbanyak di masing-masing jurusan dan asal sekolah, sehingga pihak kampus dapat memikirkan strategi promosi.

Kata Kunci: K-Means, Clustering, Siswa Baru, Data Mining, Promosi

Abstract

The process of accepting new students in each school produces very abundant student data in the form of address data and other data. This happens repeatedly due to the admission of new students every year. This study aims to group data of Harapan Bangsa school students by utilizing the data mining process using the clustering method. The algorithm used is the K-Means Clustering Algorithm. The K-Means algorithm is an iterative grouping algorithm that partitions datasets into a predefined number of K clusters. Implementation using web-based applications is used to help find accurate values. The attributes used for the clustering stage are major, school origin, and year of birth. The student cluster formed is 2 clusters, namely the first 47 cluster items, the second 23 cluster items. The purpose of this study is one of the basis for decision making to determine strategies in promoting schools. It can be concluded that the clustering process reached 3 iterations to get the final clustering results, so it can be concluded that there are the most enthusiasts in each department and school origin, so that the campus can think of a promotion strategy.

Keywords: K-Means, Clustering, New Students, Data Mining, Promotion (at least 5 words related to the research content separated by commas)

1. PENDAHULUAN

Penerimaan peserta didik baru merupakan kegiatan tahunan dari sekolah dan jumlah siswa baru yang dapat mendaftar di SMK Harapan Bangsa setiap tahun cenderung menurun. Oleh karena itu, diperlukan teknik untuk membantu mengubah data ini menjadi informasi yang berguna. Singkatnya, dengan menerapkan data penerimaan peserta didik baru sebagai strategi promosi dan menjadi dasar atau pedoman untuk strategi promosi. Penerapan metode Klasterisasi K-Means untuk strategi promosi dapat membantu dalam proses pengelompokan data dalam bentuk hasil pengelompokan data. Penerimaan siswa baru terjadi secara berulang dan tentunya mendapatkan data yang berlimpah, sehingga terjadi penumpukan data. Penumpukan data siswa secara terus menerus akan memperlambat pencarian informasi terhadap data tersebut. Berdasarkan berlimpahnya data siswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga informasi tersebut sangat berguna bagi pihak sekolah. Pengolahan data tersebut bisa diterapkan dengan data mining.

Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan teknik gambar dari statistik, mesin pembelajaran dan sistem manajemen database. Data mining yang disebut juga dengan *Knowledge Discovery in Database (KDD)* adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*) [1].

Output data mining dapat digunakan untuk membantu pengambilan keputusan di masadepan. "K-Means adalah metode clustering non-hierarchical yang mencoba untuk mempartisi data yang ada menjadi satu atau lebih cluster atau mungkin bertujuan untuk mempartisi datamenjadi beberapa cluster [2]

Data Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data clustering. K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster

yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster [3].

K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah clustering. Metode K-Means adalah metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik [4]

Aturan dan metode algoritma K-Means dapat diterapkan pada sebuah program bantu untuk mengelompokkan data siswa. Sekelompok data siswa dimasukkan ke dalam input program, kemudian program melakukan pengolahan data yang digunakan sebagai kriteria yang telah ditetapkan sesuai langkah algoritma K-Means, dan hasilnya berupa klaster data berdasarkan jurusan yang menjadi bahan untuk promosi sekolah guna meningkatkan jumlah siswa.

Berdasarkan latar belakang yang didapatkan dalam penelitian ini penulis mengelompokkan data siswa baru tahun ajaran 2022/2023 dengan teknik Clustering. Pengelompokkan yang penulis terapkan menggunakan algoritma K-Means Clustering. Pengelompokkan menggunakan data analisa seperti Nama Siswa Baru, Jurusan, Asal sekolah, dan Tahun Lahir.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini berfokus pada penerapan metode *clustering* dengan algoritma *K-Means* pada pengelompokan data calon siswa baru, terdapat tahapan-tahapan dalam penerapan algoritma K-Means pada aplikasi pengelompokan data calon siswa baru untuk memastikan bahwa penelitian ini mendapatkan hasil yang sesuai dengan tujuan yang merupakan penyelesaian dari masalah yang telah dijabarkan. Berikut adalah tahapan-tahapan yang digunakan dalam penelitian ini :

1. Studi Kepustakaan

Studi kepustakaan yang digunakan sebagai tinjauan dan acuan yang digunakan dalam penelitian ini, berisi teori-teori dan tinjauan oleh penelitian terdahulu

2. Pengumpulan Data

Pengumpulan data yang dilakukan dengan teknik observasi dan wawancara, yang mana data yang dikumpulkan berupa data siswa yang nantinya akan dilakukan proses pengelompokan

3. Penerapan Metode *K-Means*

Penerapan metode *Clustering* dilakukan dengan menerapkan algoritma *K-Means* pada data sampel yang telah dikumpulkan, mulai dari transformasi data, perhitungan, pengelompokan dan hasil

2.2 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar. (Turban et al, 2005). Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [5].

Menurut Narwati, data mining dapat diartikan sebagai proses mengekstrak atau “menggali” pengetahuan yang ada pada sekumpulan data. Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge Discovery in Database* atau yang biasa disebut KDD. Dari sudut pandang yang lain, data mining dianggap sebagai satu langkah yang penting didalam proses KDD. Menurut Han, J. and Kamber, M, 2001, proses KDD ini terdiri dari langkah-langkah sebagai berikut :

1. *Data Cleaning*, proses menghapus data yang tidak konsisten dan kotor *Data Integration*, penggabungan beberapa sumber data
2. *Data Selection*, pengambilan data yang akan dipakai dari sumber data
3. *Data Transformation*, proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining
4. *Data Mining*, suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data
5. *Pattern Evaluation*, proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada didalam data itu sendiri
6. *Knowledge Presentation*, proses visualisasi dan teknik menyajikan knowledge digunakan untuk menampilkan knowledge hasil mining kepada user [6].

2.3 Clustering

Clustering atau analisis kelompok merupakan proses mengelompokkan data (objek) berdasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Tujuan dari pengelompokan ini adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok yang lain. Tujuan pekerjaan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok yang untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari *prototipe* kelompok yang paling *representative* terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya [7].

2.4 Metode K-Means

K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat cluster awal. Pada tahap ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap *cluster*. Posisi pusat *cluster* akan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap pusat *cluster* dan terakhir akan terbentuk posisi pusat *cluster* yang baru.

Algoritma *K-Means* pada dasarnya melakukan dua proses, yakni proses pendeteksian lokasi pusat tiap *cluster* dan proses pencarian anggota dari tiap-tiap *cluster*. Cara kerja algoritma *K-Means*:

1. Tentukan *k* sebagai jumlah *cluster* yang ingin dibentuk.
2. Bangkitkan *k centroid* (titik pusat *cluster*) awal secara *random*.
3. Hitung jarak setiap data ke masing-masing *centroid*.
4. Setiap data memilih *centroid* yang terdekat.
5. Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama.
6. Kembali ke langkah-3 jika posisi *centroid* baru dengan *centroid* yang lama tidak sama [8].

2.5 UML (Unified Modeling Language)

Unified Modeling Language merupakan salah satu metode pemodelan visual yang digunakan dalam perancangan dan pembuatan sebuah *software* yang berorientasikan pada objek. UML merupakan sebuah standar penulisan atau semacam *blue print* dimana didalamnya termasuk sebuah bisnis proses, penulisan kelas-kelas dalam sebuah bahasa yang spesifik. Terdapat beberapa diagram UML yang sering digunakan dalam pengembangan sebuah sistem [9].

UML adalah salah satu *tool/model* untuk merancang pengembangan *software* yang berbasis *object-oriented*. UML sendiri juga memberikan standar penulisan sebuah sistem *blueprint*, yang meliputi konsep proses bisnis, penulisan kelas-kelas dalam bahasa program yang spesifik, skema *database*, dan komponen yang diperlukan dalam sistem *software* [10].

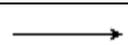
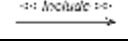
Menurut [11] "*Unified Modelling Language (UML)* merupakan kumpulan diagram-diagram yang sudah memiliki standar untuk membangun sebuah perangkat lunak berbasis objek. UML adalah salah satu standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan *requirement*, membuat analisis dan desain, serta menggambarkan arsitektur ke dalam pemrograman berorientasi objek".

2.5.1 Use Case Diagram

Use case diagram adalah satu dari berbagai jenis diagram UML (*Unified Modelling Language*) yang menggambarkan hubungan interaksi antara sistem dan aktor. *Use Case* dapat mendeskripsikan tipe interaksi antara si pengguna sistem dengan sistemnya. Merupakan gambaran dari fungsionalitas yang diharapkan dari sebuah sistem, dan merepresentasikan sebuah interaksi antara aktor dan sistem.

Tabel 1. Simbol-simbol *Use Case Diagram*.

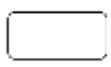
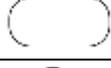
No.	Simbol	Nama Simbol	Keterangan
1		<i>Use Case</i>	Menerangkan "apa" yang dikerjakan sistem, bukan "bagaimana" sistem mengerjakannya.

No.	Simbol	Nama Simbol	Keterangan
2		Actor	Menggambarkan orang, sistem atau <i>External entitas/stakeholder</i> yang menyediakan atau menerima informasi dari sistem.
3		System Boundary	Menggambarkan jangkauan sistem.
4		Association	Menggambarkan bagaimana aktor terlibat dalam <i>use case</i> .
5		Generalization	Dibuat ketika ada sebuah keadaan yg lain/perlakuan khusus.
6		Extend	Perluasan dari <i>use case</i> lain jika kondisi atau syarat terpenuhi.
7		Include	Menjelaskan bahwa <i>use case</i> termasuk dalam <i>use case</i> lain.

2.5.2 Activity Diagram

Activity diagram atau diagram aktivitas adalah bentuk visual dari alur kerja yang berisi aktivitas dan tindakan, yang juga dapat berisi pilihan, atau pengulangan. Dalam *Unified Modeling Language (UML)*, diagram aktivitas dibuat untuk menjelaskan aktivitas komputer maupun alur aktivitas dalam organisasi. Selain itu diagram aktivitas juga menggambarkan alur kontrol secara garis besar. *Activity diagram* menggambarkan berbagai aliran aktivitas dalam sistem yang sedang di rancang, bagaimana masing-masing aliran berawal, *decision* yang mungkin terjadi dan bagaimana mereka berakhir.

Tabel 2. Simbol-simbol *Activity Diagram*.

No.	Simbol	Nama Simbol	Keterangan
1		Activity	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain
2		Action	<i>State</i> dari sistem yang mencerminkan eksekusi dari suatu aksi
3		Initial Node	Bagaimana objek dibentuk atau diawali.
4		Activity Final Node	Bagaimana objek dibentuk atau diakhiri.
5		Decision	Digunakan untuk menggambarkan suatu keputusan / tindakan yang harus diambil pada kondisi tertentu.
6		Line Connector	Digunakan untuk menghubungkan satu simbol dengan simbol lainnya.

2.5.3 Class Diagram

Class diagram adalah salah satu jenis diagram yang paling berguna di UML, hal ini karena dapat dengan jelas memetakan struktur sistem tertentu dengan memodelkan kelas, atribut, operasi serta hubungan antar objek. *Class diagram* juga dapat memperlihatkan hubungan antar kelas dan penjelasan detail tiap-tiap kelas di dalam model desain (*logical view*) dari suatu sistem. Selama proses desain, *class diagram* berperan dalam menangkap struktur dari semua kelas yang membentuk arsitektur sistem yang dibuat [12].

Tabel 3. Simbol-simbol *Class Diagram*.

No.	Simbol	Nama Simbol	Keterangan
1		Class	Kelas kepada struktur sistem.

2		<i>Interface</i>	Sama dengan konsep <i>interface</i> dalam pemrograman berorientasi objek.
3		<i>Association</i>	Relasi antarkelas dengan makna umum, asosiasi biasanya juga disertai dengan <i>multiplicity</i>
4		<i>Direction Association</i>	Relasi antarkelas dengan makna kelas yang satu digunakan oleh kelas yang lain.
5		<i>Generalisation</i>	Relasi antarkelas dengan makna generalisasi (umum khusus).
6		<i>Dependency</i>	Relasi antarkelas dengan makna kebergantungan antar kelas.
7		<i>Aggregation</i>	Relasi antarkelas dengan makna semua bagian.

2.6 Aplikasi Pengembangan Sistem

Dalam membangun sistem pengelompokan data pada siswa baru, maka dibutuhkan beberapa *software* yang digunakan dalam pembangunan aplikasi tersebut. Berikut aplikasi pengembangan sistem yang dibutuhkan.

2.6.1 Visual Studio Code

Visual Studio Code (VS Code) ini adalah sebuah teks editor ringan dan handal yang dibuat oleh *Microsoft* untuk sistem operasi *multiplatform*, artinya tersedia juga untuk versi *Linux*, *Mac*, dan *Windows*. Teks editor ini secara langsung mendukung bahasa pemrograman *JavaScript*, *Typescript*, dan *Node.js*, serta bahasa pemrograman lainnya dengan bantuan *plugin* yang dapat dipasang via *marketplace Visual Studio Code* (seperti *C++*, *C#*, *Python*, *Go*, *Java*, dst). Banyak sekali fitur-fitur yang disediakan oleh *Visual Studio Code*, diantaranya *Intellisense*. Pembaruan versi *Visual Studio Code* ini juga dilakukan berkala setiap bulan, dan inilah yang membedakan *VS Code* dengan *text editor-text editor* yang lain. *Text editor VS Code* juga bersifat *open source*, yang mana kode sumbernya dapat kalian lihat dan kalian dapat berkontribusi untuk pengembangannya. Kode sumber dari *VS Code* ini pun dapat dilihat di *link Github*. Hal ini juga yang membuat *VS Code* menjadi favorit para pengembang aplikasi, karena para pengembang aplikasi bisa ikut serta dalam proses pengembangan *VS Code* ke depannya [13].

2.6.2 Laragon

Laragon adalah perangkat lunak yang memiliki bahasa pemrograman *PHP*, *MySQL* sebagai tempat penyimpanan *database*, dan *apache* sebagai *web server* yang digunakan untuk membangun *local development environment* pada sistem operasi *windows*. Adapun kelebihan dari *Laragon* adalah sebagai berikut:

1. *Pretty URLs*, project dapat diakses dengan *app.test* tanpa harus menggunakan *localhost/app*.
2. *Portable*, project dapat dipindahkan dengan mudahnya tanpa merusak sistem.
3. *Isolated*, sistem pada *laragon* terisolasi langsung dengan sistem operasi sehingga apa yang pengguna lakukan pada aplikasi ini tidak mempengaruhi komputer lokal pengguna.
4. *Easy Operation*, aplikasi ini otomatis memiliki banyak konfigurasi sehingga sangat mudah untuk digunakan.
5. *Modern dan Powerful*, aplikasi ini memiliki arsitektur yang modern sehingga mudah digunakan saat membangun web yang modern [14].

2.6.3 Excel

Excel dari *Microsoft Corporation* membuat dan merilis *Excel*, alat spreadsheet yang kompatibel dengan *Windows* dan *Mac OS*. *Suite Microsoft Office* mencakup program ini.

Berdasarkan sistem operasi *windows*, *Microsoft Excel* adalah aplikasi dari keluarga *Microsoft Office*. *Microsoft Excel* menggunakan tabel dengan baris dan kolom untuk memproses data yang disajikan sebagai angka atau bilangan bulat. Alat pengolah data dan angka yang paling populer saat ini adalah *Microsoft Excel*, yang dapat digunakan dalam berbagai pengaturan pada perangkat seperti *desktop*, *tablet*, dan *smartphone*. Selain sistem operasi *Windows*, *Microsoft Excel* juga tersedia untuk *Macintosh*, *Android*, dan *Apple IOS*. *Excel* menjalankan rumus pada lembar bentang [15].

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Metode

Adapun tujuan dari Penelitian ini adalah untuk mengelompokkan data siswa baru tahun ajaran 2022/2023 dengan menggunakan Algoritma K-Means Clustering. Adapun variable yang digunakan untuk melakukan pengelompokan data siswa baru tahun ajaran 2022/2023 terdiri dari 3 yaitu:

1. Asal Sekolah
2. Prodi/Jurusan yang dipilih

3. Tahun Lahir

Untuk melakukan clustering menggunakan metode k-means ada beberapa langkah yang harus dilakukan adalah sebagai berikut:

2.1 Sumber Data

Data dalam penelitian ini bersumber dari Sekolah Harapan Bangsa. Adapun jumlah data yang diperoleh sebanyak 70 yang terdiri Nama Siswa, Jurusan, Asal sekolah dan Tahun Lahir.

Berikut contoh data Siswa baru tahun 2022/2023 yang diperoleh:

Tabel 4. Data siswa baru yang diperoleh

No.	Nama	Jurusan	Asal Sekolah	Tahun Lahir
1	ADAM ARYA	Teknik dan Bisnis Sepeda Motor	SMP Negeri 3 Galang	2007
2	ADITYA TRI WARDANA	Teknik dan Bisnis Sepeda Motor	SMP Negeri 4 Tanjung Morawa	2007
3	AKBAR KURNIAWAN	Teknik dan Bisnis Sepeda Motor	SMP Negeri 2 STM Hilir	2006
4	AKBAR NUGROHO	Teknik dan Bisnis Sepeda Motor	MTs Muhammadiyah 13	2007
5	ARDIANSYAH	Teknik dan Bisnis Sepeda Motor	SMP Negeri 3 Tanjung Morawa	2007
6	BALQIS NUR RAHMAN	Teknik dan Bisnis Sepeda Motor	SMP Tamora 2 Tanjung Morawa	2007
7	BRAHMANA KARTA NEGARA	Teknik dan Bisnis Sepeda Motor	SMP Wira Jaya Tanjung Morawa	2007
8	DIKA PRATAMA	Teknik dan Bisnis Sepeda Motor	MTs Sidratul Ulya Naga Timbul	2007
9	DIMAS RANGGA SATRIO	Teknik dan Bisnis Sepeda Motor	SMP Nur Azizi	2006
10	FAUZI ARDIANSYAH	Teknik dan Bisnis Sepeda Motor	SMP Wira Jaya Tanjung Morawa	2007
11	IKHWAN ARIEF PRAKOSO	Teknik dan Bisnis Sepeda Motor	SMP Nur Azizi	2008
12	JEKI PRANATA	Teknik dan Bisnis Sepeda Motor	SMP Negeri 3 Galang	2005
13	JUNIARDI AL VANZA	Teknik dan Bisnis Sepeda Motor	SMP Negeri 3 Galang	2007
14	MAHA RIZKY	Teknik dan Bisnis Sepeda Motor	SMP Sumatera	2006
15	MECKEL WIJAYA	Teknik dan Bisnis Sepeda Motor	SMP Negeri 5 Tanjung Morawa	2006

2.2 Transformasi Data

Transformasi Data dilakukan untuk mengubah data tujuannya adalah agar data dapat diolah dengan menggunakan metode K-Means Clustering. Adapun variabel yang digunakan yaitu Nama Siswa, Jurusan, Asal Sekolah dan Tahun Lahir, data tahun lahir bisa menggambarkan sudah berapa lama lulus sekolah.

Jurusan dikelompokkan menjadi 3 kelompok. Dengan kelompok pertama yaitu Jurusan Teknik dan Bisnis Sepeda Motor ditransformasikan menjadi 1, Jurusan Teknik Komputer dan Jaringan ditransformasikan dengan nilai 2 dan Jurusan Otomatisasi dan Tata Kelola Perkantoran ditransformasikan menjadi 3. Untuk variabel asal sekolah dikelompokkan menjadi 3 kelompok. Adapun kelompok data pertama dengan asal Sekolah SMA ditransformasikan dengan nilai 1, asal sekolah SMK ditransformasikan dengan nilai 2, dan asal sekolah diluar SMK dan SMA ditransformasikan dengan nilai 3. Untuk variabel tahun lahir dikelompokkan menjadi 3 kelompok, yang pertama untuk tahun lahir 2000-2003 ditransformasikan dengan nilai 1, untuk tahun lahir 1997-1999 ditransformasikan dengan nilai 2, dan untuk tahun lahir >1997 ditransformasikan dengan nilai 3. Adapun hasil transformasi dapat dilihat pada tabel berikut:

Tabel 5. Data Hasil Transformasi

No.	Nama	Jurusan	Asal Sekolah	Tahun Lahir
P1	ADAM ARYA	1	1	2
P2	ADITYA TRI WARDANA	1	1	2
P3	AKBAR KURNIAWAN	1	1	2
P4	AKBAR NUGROHO	1	3	2
P5	ARDIANSYAH	1	1	2
P6	BALQIS NUR RAHMAN	1	2	2
P7	BRAHMANA KARTA NEGARA	1	2	2
P8	DIKA PRATAMA	1	3	2
P9	DIMAS RANGGA SATRIO	1	2	2
P10	FAUZI ARDIANSYAH	1	2	2
P11	IKHWAN ARIEF PRAKOSO	1	2	3
P12	JEKI PRANATA	1	1	1
P13	JUNIARDI AL VANZA	1	1	2
P14	MAHA RIZKY	1	2	2
P15	MECKEL WIJAYA	1	1	2

2.3 Pengolahan Data

Pengolahan data siswa baru dilakukan setelah proses transformasi sehingga data siswa baru bisa diolah menggunakan metode K-Mean Clustering. Adapun langkah-langkah proses Algoritma K-Mean Clustering adalah sebagai berikut:

1. Dilakukan k dari jumlah cluster baru yang ingin dibentuk. Adapun cluster yang akan dibuat adalah 2 cluster.
2. Tentukan titik pusat awal dari setiap cluster.

Adapun penentuan titik pusat awal dalam penelitian ini ditentukan dengan cara random dan titik pusat yang didapatkan dapat dilihat pada tabel berikut:

Tabel 6. Titik Pusat awal tiap cluster

Centroid Awal	K1	K2	K3
Centroid 1	1	3	2
Centroid 2	1	2	3

3. Hitung jarak setiap data ke pusat cluster antara objek ke centroid terdekat. Centroid terdekat akan menjadi cluster yang diikuti oleh data tersebut. Perhitungan jarak Euclidean dapat dilakukan dengan berikut:

$$\sqrt{(X_{ki} - X_{kj})^2 + (X_{ki} - X_{kj})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Persamaan diatas digunakan karena atribut yang digunakan berjumlah 3. Sebagai contoh, akan dihitung jarak dari data siswa pertama ke pusat cluster pertama dengan persamaan:

$$\begin{aligned} &= \sqrt{(X_{ki} - X_{kj})^2 + (X_{ki} - X_{kj})^2 + \dots + (X_{ki} - X_{kj})^2} \\ &= \sqrt{(1 - 1)^2 + (1 - 3)^2 + (2 - 2)^2} \\ &= 2.00 \end{aligned}$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data siswa baru pertama ke pusat cluster pertama adalah 2. Jarak data siswa baru pertama ke pusat cluster kedua dengan persamaan:

$$\begin{aligned} &= \sqrt{(X_{ki} - X_{kj})^2 + (X_{ki} - X_{kj})^2 + \dots + (X_{ki} - X_{kj})^2} \\ &= \sqrt{(1 - 1)^2 + (1 - 2)^2 + (2 - 3)^2} \\ &= 1.41 \end{aligned}$$

Berdasarkan hasil kedua perhitungan diatas dapat disimpulkan bahwa jarak data siswa baru pertama yang paling mendekati adalah cluster 2, sehingga siswa baru pertama dimasukkan kedalam cluster 2. Hasil perhitungan selengkapnya untuk 15 sampel data siswa baru dapat dilihat pada tabel berikut ini:

Tabel 7. Hasil perhitungan setiap data ke setiap pusat cluster Iterasi 1

Alt	Jarak titik ke M1	Jarak titik ke M2	Jarak Terdekat	Keterangan
P1	2.00	1.41	1.41	2
P2	2.00	1.41	1.41	2
P3	2.00	1.41	1.41	2
P4	0.00	1.41	0.00	1
P5	2.00	1.41	1.41	2
P6	1.00	1.00	1.00	1
P7	1.00	1.00	1.00	1
P8	0.00	1.41	0.00	1
P9	1.00	1.00	1.00	1
P10	1.00	1.00	1.00	1
P11	1.41	0.00	0.00	2
P12	2.24	2.24	2.24	1
P13	2.00	1.41	1.41	2
P14	1.00	1.00	1.00	1
P15	2.00	1.41	1.41	2

1. Setelah semua data ditempatkan kedalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut.
2. Jika centroid yang baru konvergen dengan centroid yang lama maka hentikan iterasi. Jika tidak maka iterasi dilanjutkan ke yang berikutnya.
3. Selanjutnya mengelompokkan hasil cluster pada iterasi pertama yang belum konvergen.

Tabel 8. Titik pusat *cluster* baru

Centroid	K1	K2	K3
M1	1.36	2.19	1.87
M2	1.43	1.04	2.04

Karena centroid baru yang digunakan belum konvergen maka iterasi harus dilanjutkan. Dalam penelitian ini, iterasi terjadi sebanyak 3 kali iterasi karena kondisi cluster sudah mencapai konvergen dan proses iterasi pun berhenti. Adapun hasil akhir clustering dapat dilihat pada tabel berikut:

Tabel 9. Hasil perhitungan data ke setiap Cluster pada iterasi-3

Alt	Jarak Ke Titik M1	Jarak Ke Titik M2	Jarak Terdekat	Keterangan
P1	1.27	0.44	0.44	2
P2	1.27	0.44	0.44	2
P3	1.27	0.44	0.44	2

Tabel 10. Hasil perhitungan data ke setiap Cluster pada iterasi-3 (Lanjutan)

Alt	Jarak Ke Titik M1	Jarak Ke Titik M2	Jarak Terdekat	Keterangan
P4	0.87	2.05	0.87	1
P5	1.27	0.44	0.44	2
P6	0.43	1.09	0.43	1
P7	0.43	1.09	0.43	1
P8	0.87	2.05	0.87	1
P9	0.43	1.09	0.43	1

P10	0.43	1.09	0.43	1
P11	1.16	1.51	1.16	1
P12	1.56	1.05	1.05	2
P13	1.27	0.44	0.44	2
P14	0.43	1.09	0.43	1
P15	1.27	0.44	0.44	2

Pada Iterasi Keempat, titik pusat dari setiap Cluster tidak berubah dan tidak ada lagi data yang berpindah dari satu Cluster ke Cluster yang lain.

Adapun hasil analisa *cluster* pada pada tabel 3.17 dan 3.18, berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data siswa pada setiap atribut.

Tabel 11. Hasil Analisa *Cluster* (*Cluster 1*)

Cluster 1 Terdiri dari 47 Siswa		
Jurusan :	Asal Sekolah :	Tahun Lahir :
TBSM = 33 Siswa	SMP Negeri = 0	2004-2005 = 5 Siswa
TKJ = 11 Siswa	SMP Swasta = 37 Siswa	2006-2007 = 41 Siswa
OTKP = 3 Siswa	MTs = 10 Siswa	> 2007 = 1 Siswa

Tabel 12. Hasil Analisa *Cluster* (*Cluster 2*)

Cluster 2 Terdiri dari 23 Siswa		
Jurusan :	Asal Sekolah :	Tahun Lahir :
TBSM = 15 Siswa	SMP Negeri = 23 Siswa	2004-2005 = 1 Siswa
TKJ = 6 Siswa	SMP Swasta = 0	2006-2007 = 22 Siswa
OTKP = 2 Siswa	MTs = 0	> 2007 = 0

Dari data hasil *clustering* yang telah dilakukan diatas, maka dapat dilihat data tersusun rapi berdasarkan peminat setiap jurusan, asal sekolah, beserta tahun lahir. Dapat disimpulkan bahwa pada *cluster 1* Jurusan yang diminati adalah Teknik dan Bisnis Sepeda Motor, dengan asal sekolah dominan SMP swasta, dan tahun lahir yaitu dominan dari tahun 2006-2007. Pada *cluster 2* yang diminati adalah Jurusan Teknik dan Bisnis Sepeda Motor, dengan asal sekolah secara keseluruhan dari SMP negeri, sementara tahun lahir yaitu dominan dari tahun 2006-2007. Dari hasil *clustering* diatas yang paling efektif adalah *cluster 1*, karena merupakan jumlah yang terbanyak dari hasil *clustering*. Dari hasil tersebut dapat dilihat bahwa untuk peminat setiap jurusan dominan dari SMP swasta walaupun dari *cluster* lain ada beberapa jurusan yang mendominasi, dan untuk tahun lahir secara umum dari setiap *cluster* adalah tahun 2006-2007. Dari hasil analisa tersebut dapat dipikirkan strategi promosi sekolah, yaitu dimaksimalkan pada sekolah SMP swasta dan SMP negeri.

3.2 Hasil Tampilan Antarmuka

Berikut merupakan hasil dari tampilan antarmuka pada aplikasi yang telah dibangun.

1. Tampilan Halaman *Login*

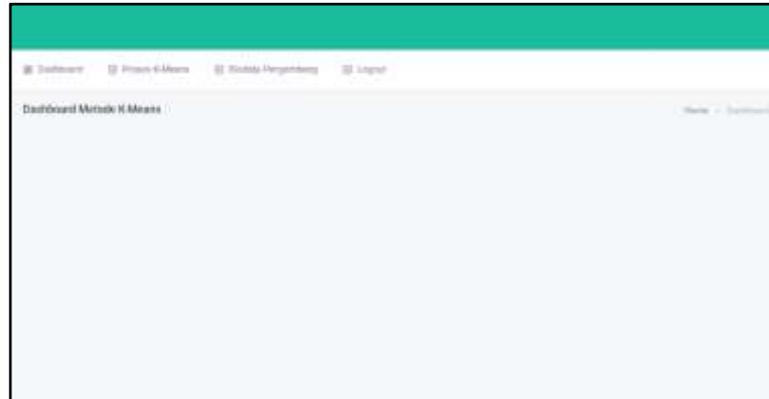
Halaman *Login* adalah halaman yang digunakan admin untuk masuk ke dalam sistem dengan menggunakan *username* dan *password* yang sudah terdaftar pada sistem *database*. Berikut ini tampilan halaman *login* yang telah selesai dibangun.



Gambar 1 Tampilan Halaman Utama

2. Tampilan Halaman Utama

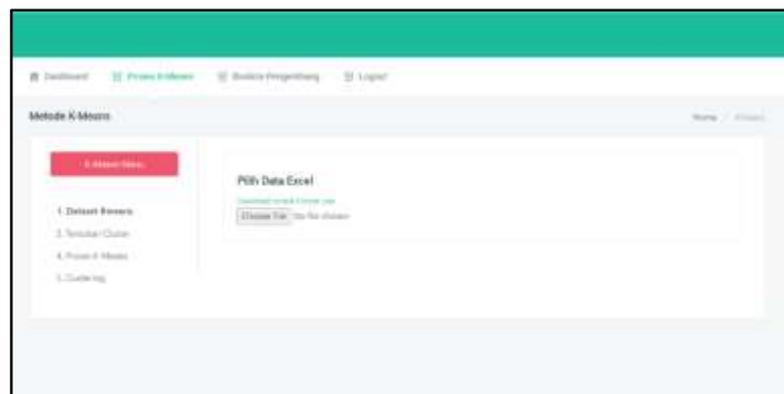
Halaman Utama ini berisi tentang *dashboard*, proses *K-Means*, biodata pengembang dan *logout*. Berikut ini tampilan halaman menu utama admin yang telah selesai dibangun.



Gambar 2 Tampilan Halaman Utama

3. Tampilan Proses *K-Means* (*Dataset K-Means*)

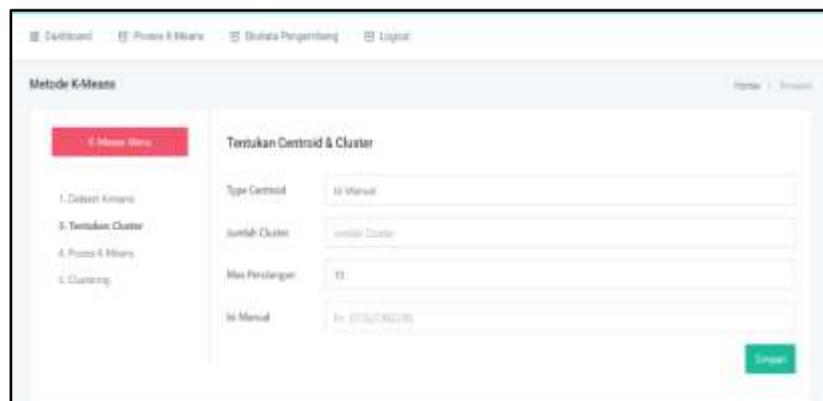
Halaman proses *K-Means* (*Dataset K-Means*) ini berisi *form upload* data siswa sebelum melakukan perhitungan. Berikut halaman proses *K-Means*.



Gambar 3 Tampilan Halaman Proses *K-Means* (*Dataset K-Means*)

4. Tampilan Proses *K-Means* (*Tentukan cluster*)

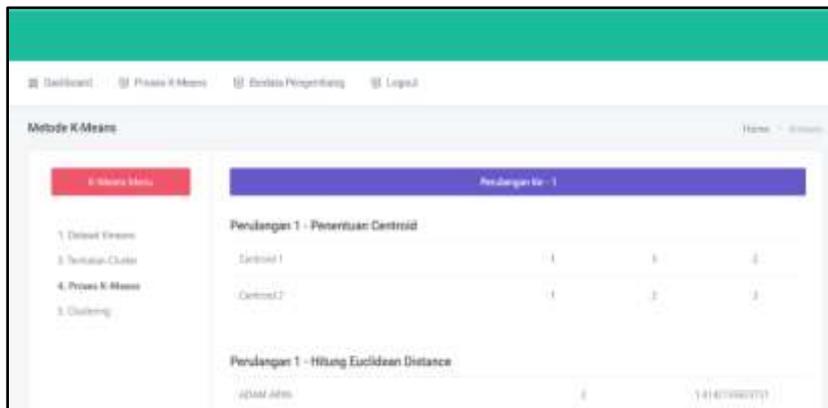
Halaman proses *K-Means* (*Tentukan cluster*) ini terdiri dari tipe *centroid*, penentuan jumlah *cluster*, penentuan maksimal perulangan dan penentuan *centroid* beserta tombol simpan. Berikut ini tampilan halaman proses *K-Means* (*Tentukan cluster*) admin yang telah selesai dibangun.



Gambar 4 Halaman Tampilan proses *K-Means* (*Tentukan cluster*)

5. Tampilan Proses *K-Means* (Proses perhitungan)

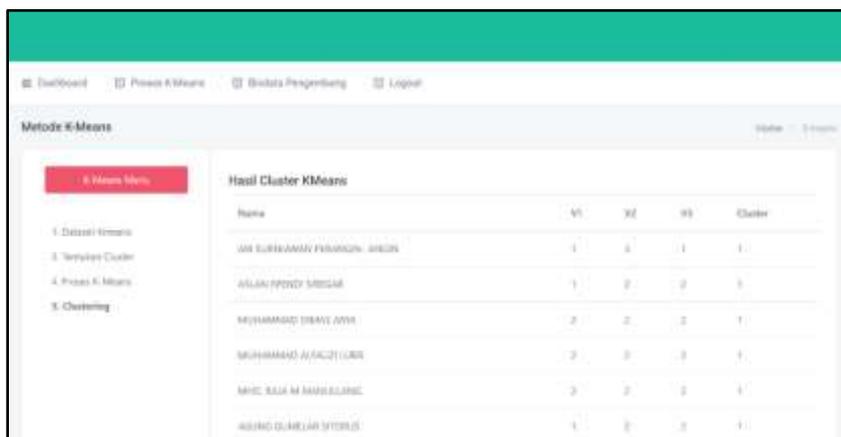
Halaman proses *K-Means* (proses perhitungan) ini berisikan perhitungan data mulai dari perulangan 1 dengan *centroid* yang sudah ditentukan sampai dengan selesai. Berikut halaman proses *K-Means* (Proses perhitungan).



Gambar 5 Halaman Tampilan proses *K-Means* (Proses perhitungan)

6. Tampilan Proses *K-Means* (Clustering)

Halaman proses *K-Means* (Clustering) ini berisikan hasil cluster dari perhitungan data yang sudah. Halaman ini juga dilengkapi dengan fitur *export* berekstensi *word* data hasil proses perhitungan. Berikut ini tampilan halaman proses *K-Means* (Clustering) admin yang telah selesai dibangun.



Gambar 6 Halaman Tampilan proses *K-Means* (Clustering)

4. KESIMPULAN

Berdasarkan hasil pembahasan tentang implementasi *data mining* untuk pengelompokan data siswa baru pada sekolah menggunakan Metode *K-Means* yang telah dikemukakan, maka dapat diperoleh beberapa kesimpulan. Adapun kesimpulan tersebut adalah sebagai berikut, Penentuan *centroid* (titik pusat) pada tahap awal algoritma *K-Means* sangat berpengaruh pada hasil *cluster* seperti pada hasil pengujian yang dilakukan dengan menggunakan 70 dataset dengan *centroid* yang berbeda menghasilkan hasil *cluster* yang berbeda juga, lalu Setelah dilakukan pengelompokan data siswa baru menggunakan metode *k-means*, *Clustering* yang terbentuk dua cluster yaitu *cluster* satu dengan jumlah 47 items dan *cluster* dua dengan jumlah 23 items.

UCAPAN TERIMAKASIH

Puji dan Syukur dipanjatkan kepada Allah SWT yang memberikan rahmat dan karunia sehingga mampu menyelesaikan jurnal ini. Kemudian kepada Bapak Hendra Jaya dan Rini Kustini atas segala waktu dan ilmunya yang telah memberikan bimbingan selama masa pengerjaan hingga menyelesaikan jurnal ini.

DAFTAR PUSTAKA

- [1] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [2] G. Aprilianur, E. L. Hadisaputro, J. A. Syahranie, B. Ampar, K. Balikpapan, and K. Timur, "Penerapan Data Mining menggunakan metode"
- [3] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [4] W. Mega, "Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita," 2015.
- [5] D. T. Larose and C. D. Larose, "DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining."
- [6] J. Han *et al.*, "Designing Data-Intensive Web Applications."
- [7] dan Marlindawati, "Pengelompokan Minat Belajar Mahasiswa Menggunakan Teknik Data Mining Dengan Metode Clustering."
- [8] Y. Darmi, A. Setiawan, J. Bali, K. Kampung Bali, K. Teluk Segara, and K. Bengkulu, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," 2016.
- [9] M Teguh Prihandoyo, "Unified Modeling Language (UML) Model Untuk Pengembangan Sistem Informasi Akademik Berbasis Web".
- [10] F.- Sonata, "Pemanfaatan UML (Unified Modeling Language) Dalam Perancangan Sistem Informasi E-Commerce Jenis Customer-To-Customer," *Jurnal Komunika: Jurnal Komunikasi, Media dan Informatika*, vol. 8, no. 1, p. 22, Jun. 2019, doi: 10.31504/komunika.v8i1.1832.
- [11] F. Ayu, N. Permatasari, M. Informatika, M. Riau, J. H. Soebrantas, and N. 77 Panam, "Perancangan Sistem Informasi Pengolahan Data Praktek Kerja Lapangan (Pkl) Pada Devisi Humas Pt. Pegadaian," vol. 2, no. 2, 2018.
- [12] T. Bayu Kurniawan, "Perancangan Sistem Aplikasi Pemesanan Makanan Dan Minuman Pada Cafeteria No Caffe Di Tanjung Balai Karimun Menggunakan Bahasa Pemograman".
- [13] P. R. A. Yudi Permana1), "Perancangan Sistem Informasi Penjualan Perumahan Menggunakan Metode Sdlc Pada Pt. Mandiri Land Prosperous Berbasis Mobile".
- [14] I. Meidina, Y. Siradj, and E. Insanudin, "Pembangunan Web Administrator Pada Aplikasi Media Informasi Dan Perdagangan Untuk Petani Satur Di Nagari Alahan Panjang Kabupaten Solok Web-Based Administrator Development In The Application Of Information And Trade Media For Vegetable Farmers In Nagari Alahan Panjang Solok Regency."
- [15] A. Firman, H. F. Wowor, and X. Najoran, "Sistem Informasi Perpustakaan Online Berbasis Web," 2016.