

***CLUSTERING* SMA NEGERI DI JAKARTA BARAT MENGGUNAKAN
ALGORITMA K-MEANS SEBAGAI REKOMENDASI UNTUK SUKU
DINAS PENDIDIKAN JAKARTA BARAT**



**ANDREAS AJI RANINDA
2110512045**

**S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2025**

***CLUSTERING* SMA NEGERI DI JAKARTA BARAT MENGGUNAKAN
ALGORITMA K-MEANS SEBAGAI REKOMENDASI UNTUK SUKU
DINAS PENDIDIKAN JAKARTA BARAT**

**ANDREAS AJI RANINDA
2110512045**

**Sebagai salah satu syarat untuk melaksanakan
penelitian oleh mahasiswa pada
Program Studi S1 Sistem Informasi**

**S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2025**

PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya saya sendiri serta semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Andreas Aji Raninda

NIM : 2110512045

Tanggal : 2 Februari 2026

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses dengan ketentuan yang berlaku.

Jakarta, 2 Februari 2026

Yang menyatakan,



A handwritten signature in black ink is written over a pink rectangular stamp. The stamp features the Garuda Pancasila emblem at the top, the word 'METERAI' in the center, and the word 'TEMPEL' below it. At the bottom of the stamp, the alphanumeric code 'AB176ANX268757710' is printed. The signature is a stylized, cursive script.

Andreas Aji Raninda

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMISI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Andreas Aji Raninda
NIM : 2110512045
Fakultas : Ilmu Komputer
Program Studi : S1 Sistem Informasi

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Eksklusif (Non-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul:

Clustering SMA Negeri di Jakarta Barat Menggunakan Algoritma K-Means sebagai Rekomendasi untuk Suku Dinas Pendidikan Jakarta Barat

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta

Pada tanggal : 2 Februari 2026

Yang menyatakan,



Andreas Aji Raninda

LEMBAR PENGESAHAN

Judul : Clustering SMA Negeri di Jakarta Barat Menggunakan Algoritma K-Means sebagai Rekomendasi untuk Suku Dinas Pendidikan Jakarta Barat
Nama : Andreas Aji Raninda
NIM : 2110512045
Program Studi : SI Sistem Informasi

Disetujui oleh :

Penguji 1:
Ika Nurlaili Isnainiyah, S.Kom., M.Sc.

Penguji 2:
Rifka Dwi Amalia, S.Pd., M.Kom

Pembimbing 1:
Dr. Tjahjanto, S.Kom., M.M.

Pembimbing 2:
Rudhy Ho Purabaya, SE.,MMSI



Diketahui oleh:

Koordinator Program Studi:
Anita Muliawati, S.Kom., MTI.
NIP. 197005212021212002

Dekan Fakultas Ilmu Komputer:
Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM.
NIP. 197605082003121002



Tanggal Ujian Tugas Akhir:
16 Desember 2025

ABSTRAK

Di masa kini, dimana kebutuhan akan pendidikan yang berkualitas semakin dibutuhkan, pemerataan sarana dan prasarana serta sumber daya pendukung di setiap sekolah menjadi kunci penting. Terlebih pada SMA Negeri, yang merupakan salah satu pintu masuk menuju tingkat Perguruan Tinggi Negeri yang dapat membantu menaikkan taraf hidup masyarakat. Dengan melakukan pengelompokkan terhadap kondisi fasilitas serta sumber daya pada setiap sekolah, diharapkan instansi yang bertanggung jawab dalam hal ini Suku Dinas Pendidikan Wilayah I dan II Jakarta Barat dapat memberikan keputusan atau kebijakan yang optimal dengan melihat pengelompokkan yang sudah dikumpulkan. Masalahnya adalah belum ada sistem informasi pada kedua instansi tersebut yang mengumpulkan data kondisi sekolah beserta pengelompokkannya. Data tersebut masih terpecah pada situs Dapodik, yang merupakan sebuah sumber data nasional yang memuat data-data mengenai pendidikan nasional. Pada penelitian ini akan dibangun sebuah sistem informasi yang tidak hanya memberikan informasi data mengenai kondisi sarana prasarana serta sumber daya manusia pada tiap sekolah namun juga terdapat hasil *clustering* menggunakan algoritma K-Means yang akan membantu dalam pengambilan keputusan.

Kata Kunci : *Clustering*, Jakarta Barat, K-Means, Sistem Informasi, SMA Negeri

ABSTRACT

In today's world, where the need for high quality education is increasingly important, the equitable distribution of facilities, infrastructure, and supporting resources across all schools is a key priority. This condition is particularly true for public high schools (SMA Negeri), which serve as a gateway to public universities and can help improve the standard of living for communities. By categorizing the condition of facilities and resources in each school, it is hoped that the responsible authorities, namely the West Jakarta Education Office Districts I and II (Suku Dinas Pendidikan Wilayah I dan II Jakarta Barat), can make optimal decisions or policies based on the categorized data that has been collected. The issue is that there is currently no information system in these two offices that collects data on school conditions and their categorization. This data is currently scattered across the Dapodik website, which is a national data source containing information about national education. In this study, an information system will be developed that not only provides data on the condition of infrastructure and human resources at each school but also includes clustering results using the K-Means algorithm, which will assist in decision-making.

Keywords: Clustering, West Jakarta, K-Means, Information System, SMA Negeri

KATA PENGANTAR

Dengan memanjatkan puji dan syukur atas kehadiran Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya sampai pada saat proposal skripsi ini telah dibuat. Berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan proposal skripsi ini dengan judul “*Clustering* SMA Negeri di Jakarta Barat Menggunakan Algoritma K-Means” dengan baik. Adapun proses penyelesaian penulisan proposal skripsi ini tidak lepas dari bantuan berbagai pihak kepada penulis. Untuk itu, penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa
2. Orang tua dan keluarga yang selalu menjadi penyemangat dan memberikan dukungan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik
3. Bapak Dr. Tjahjanto, S.Kom., M.M. dan Rudhy Ho Purabaya, SE., MMSI selaku dosen pembimbing proposal skripsi yang telah memberikan bimbingan dan saran selama berjalannya proposal skripsi.
4. Ibu Anita Muliawati, S.Kom., M.TI.. selaku Ketua Program Studi S1 Sistem Informasi
5. Kepada teman dan sahabat, saya yang telah bersama dan mendukung saya sejak pertama kali skripsi ini dibuat sampai dengan sekarang skripsi ini selesai.
6. Seluruh pihak yang terlibat dalam kelancaran penulisan tugas akhir ini dan yang belum disebutkan di atas, penulis ucapkan terima kasih.

Akhir kata, penulis menyadari bahwa proposal skripsi ini masih jauh dari kata sempurna, hal tersebut disebabkan keterbatasan penulis dari segi pengetahuan dan kemampuan. Oleh karena itu, penulis mengharapkan saran dan kritik dari pembaca yang bersifat membangun agar penyusunan proposal skripsi ini dapat disempurnakan.

Jakarta 14 Mei 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMISI	iii
ABSTRAK	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	x
DAFTAR RUMUS.....	xi
DAFTAR LAMPIRAN	xi
BAB 1. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan dan Manfaat Penelitian	4
1.5. Sistematika Penulisan	4
BAB 2. TINJAUAN PUSTAKA.....	6
2.1. <i>Clustering</i>	4
2.2. Sekolah Menengah Atas Negeri	6
2.3. Suku Dinas Pendidikan Wilayah I dan Wilayah II Kota Jakarta Barat	6
2.4. Data Pokok Pendidikan (Dapodik)	7
2.5. Metode Elbow	7
2.6. Algoritma K-Means	8
2.7. Davies-Bouldin Index	9
2.8. Python	10
2.9. MySQL.....	10
2.10. HTML (Hyper Text Markup Language)	11
2.11. CSS (Cascade Style Sheet).....	11
2.12. <i>Blackbox Testing</i>	11

2.13.	Unified Modelling Language (UML)	8
2.14.	Data Mining	8
2.15.	Penelitian Terdahulu.....	8
BAB 3.	METODE PENELITIAN	15
3.1.	Tahapan Penelitian	15
3.2.	Alat dan Bahan Penelitian.....	27
3.3.	Jadwal Penelitian.....	28
BAB 4.	HASIL DAN PEMBAHASAN.....	29
4.1.	Struktur Organisasi Suku Dinas Pendidikan Wilayah I dan II.....	29
4.2.	Proses Pengumpulan Data.....	30
4.3.	Data Selection	33
4.4.	Penentuan Banyak <i>Cluster</i>	36
4.5.	Proses Clustering.....	38
4.6.	Evaluasi Davies-Boulden Index.....	52
4.7.	Implementasi <i>Dashboard</i>	53
4.8.	<i>Blackbox Testing</i>	68
BAB 5.	PENUTUP	71
5.1.	Kesimpulan	71
5.2.	Saran.....	71
DAFTAR PUSTAKA		72
LAMPIRAN.....		75

DAFTAR GAMBAR

Gambar 1.1 Misi Dinas Pendidikan DKI Jakarta.....	1
Gambar 2.1 Alur proses algoritma K-Means	6
Gambar 2.2 Alur proses Knowledge Discovery in Databases	12
Gambar 3.2 Alur penelitian.....	15
Gambar 3.2 Bagian profil halaman sekolah di Dapodik	17
Gambar 3.3 Bagian rekapitulasi Sumber Daya Manusia pada halaman sekolah di Dapodik	17
Gambar 3.4 Bagian rekapitulasi sarana prasarana pada halaman sekolah di Dapodik	18
Gambar 3.5 Tampilan halaman data referensi yang menampilkan data koordinat 20	
Gambar 3.6 Tampilan isi dari file_awal.xlsx	21
Gambar 3.7 Alur proses penentuan banyak cluster.....	23
Gambar 3.7 Pembagian fitur dalam proses clustering	23
Gambar 3.9 Alur proses clustering.....	25
Gambar 4.1 Struktur organisasi Suku Dinas Pendidikan Wilayah I dan II.....	30
Gambar 4.2 Tangkapan layar hasil proses pengumpulan data primer.....	32
Gambar 4.3 Tangkapan layar hasil proses pengumpulan data sekunder.....	33
Gambar 4.4 Tangkapan layar kode data selection.....	35
Gambar 4.5 Hasil visualisasi metode elbow variabel murid.....	36
Gambar 4.6 Hasil visualisasi metode elbow variabel SDM.....	37
Gambar 4.7 Hasil visualisasi metode elbow variabel fasilitas.....	37
Gambar 4.8 Hasil visualisasi scatter plot clustering klaster fasilitas dan murid....	38
Gambar 4.9 Hasil visualisasi scatter plot clustering variabel murid.....	41
Gambar 4.10 Hasil visualisasi scatter plot clustering variabel murid dan rombel.....	43
Gambar 4.11 Hasil evaluasi DBI variabel SDM	52
Gambar 4.12 Hasil evaluasi DBI variabel murid	52
Gambar 4.13 Hasil evaluasi DBI variabel fasilitas	52
Gambar 4.14 Use Case diagram.....	54
Gambar 4.15 Activity diagram dashboard utama.....	56
Gambar 4.16 Activity diagram data sinkronisasi	57
Gambar 4.17 Activity diagram data lengkap	57
Gambar 4.18 Activity diagram clustering	58
Gambar 4.19 Sequence diagram dashboard utama	58
Gambar 4.20 Sequence diagram data sinkronisasi.....	59
Gambar 4.21 Sequence diagram data lengkap	59
Gambar 4.22 Sequence diagram clustering.....	60
Gambar 4.23 Class diagram	61
Gambar 4.24 Rancangan antarmuka dashboard utama.....	64
Gambar 4.25 Rancangan antarmuka data sinkronisasi	65
Gambar 4.26 Rancangan antarmuka data lengkap	65
Gambar 4.27 Rancangan antarmuka clustering	66

Gambar 4.28 Tampilan dashboard halaman dashboard utama.....	66
Gambar 4.29 Tampilan dashboard halaman data sinkronisasi	67
Gambar 4.30 Tampilan dashboard halaman data lengkap.....	67
Gambar 4.31 Tampilan dashboard halaman clustering	68

DAFTAR TABEL

Tabel 2.1 Ringkasan penelitian terdahulu	13
Tabel 3.1 Data primer yang akan dikumpulkan	18
Tabel 3.2 Data sekunder yang akan dikumpulkan	20
Tabel 3.3 Daftar halaman yang akan dirancang untuk <i>dashboard</i>	27
Tabel 3.4 Jadwal penelitian	28
Tabel 4.1 Pembagian wilayah Suku Dinas Pendidikan Jakarta Barat	29
Tabel 4.2 Hasil dari proses pengumpulan data primer	30
Tabel 4.3 Hasil dari proses pengumpulan data sekunder	33
Tabel 4.4 Hasil dari proses data selection	34
Tabel 4.5 Hasil penentuan banyak klaster 3 variabel	38
Tabel 4.6 Kesimpulan isi tiap klaster untuk variabel fasilitas dan murid	39
Tabel 4.7 Hasil clustering variabel fasilitas dan murid	39
Tabel 4.8 Kesimpulan isi tiap klaster untuk variabel fasilitas dan murid	41
Tabel 4.9 Hasil clustering variabel SDM dan murid	42
Tabel 4.10 Kesimpulan isi tiap klaster untuk variabel rombel dan murid	44
Tabel 4.11 Hasil clustering variabel rombel dan murid	44
Tabel 4.12 Rangkuman hasil clustering ketiga variabel	46
Tabel 4.13 Kesimpulan dan rekomendasi berdasarkan hasil clustering	47
Tabel 4.14 Rangkuman hasil evaluasi ketiga klaster	53
Tabel 4.15 Use Case skenario dashboard utama	54
Tabel 4.16 Use Case skenario data sinkronisasi	54
Tabel 4.17 Use Case skenario data lengkap	55
Tabel 4.18 Use Case skenario clustering	56
Tabel 4.19 Rancangan table identitas sekolah	61
Tabel 4.20 Rancangan table fasilitas	62
Tabel 4.21 Rancangan table SDM	62
Tabel 4.22 Rancangan table murid	63
Tabel 4.23 Tabel blackbox testing	68

DAFTAR RUMUS

Rumus 2.1 Rumus untuk menghitung SSE pada Metode Elbow.....	5
Rumus 2.2 Rumus untuk menghitung metrik DBI.....	7

DAFTAR LAMPIRAN

Lampiran 1 Surat izin riset.....	65
Lampiran 2 Dokumentasi kegiatan wawancara UP SMA Suku Dinas Pendidikan Wilayah II Kota Jakarta Barat.....	69
Lampiran 3 Transkrip wawancara dengan UP SMA Suku Dinas Pendidikan Wilayah II Kota Jakarta Barat.....	69
Lampiran 4 Dokumentasi kegiatan wawancara UP SMA Suku Dinas Pendidikan Wilayah II Kota Jakarta Barat.....	67
Lampiran 5 Transkrip wawancara dengan UP SMA Suku Dinas Pendidikan Wilayah I Kota Jakarta Barat.	67
Lampiran 6 Hasil cek plagiarisme.....	68

BAB 1. PENDAHULUAN

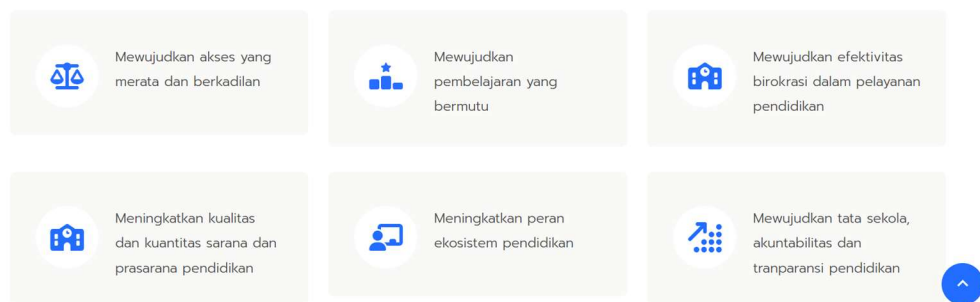
1.1. Latar Belakang

Di era perkembangan dunia yang semakin cepat dan pesat, kebutuhan akan pendidikan yang berkualitas semakin dibutuhkan oleh semua lapisan masyarakat. Kebutuhan akan fasilitas serta sumber daya pendidik yang layak dan sesuai dengan kondisi setiap sekolah merupakan hak setiap warga negara. (Nurahman et al. 2022). Kondisi fasilitas serta sumber daya yang sesuai akan sangat mendukung proses belajar mengajar menjadi lebih baik.

Ditambah dengan masyarakat yang berlomba memperebutkan masuk ke sekolah negeri dikarenakan sekolah negeri bersifat gratis untuk semua masyarakat serta kualitasnya terjaga karena dikelola langsung oleh pemerintah melalui Dinas Pendidikan. Terutama SMA Negeri dikarenakan merupakan pintu masuk besar untuk masyarakat menuju ke perguruan tinggi negeri.

Kondisi sarana dan prasarana yang mendukung dengan baik serta merata di setiap sekolah juga merupakan hal yang diperlukan. Hal ini selaras dengan beberapa misi dari Dinas Pendidikan DKI Jakarta yang dapat dilihat pada gambar di bawah ini:

Misi Dinas Pendidikan



Gambar 1.1 Misi Dinas Pendidikan DKI Jakarta

Pengelolaan data pendidikan di wilayah Jakarta Barat dilaksanakan oleh dua instansi yang berbeda. Pengelolaan data oleh dua instansi yang berbeda ini sering menimbulkan permasalahan ketika membuat laporan atau kebijakan di tingkat yang lebih tinggi yang butuh untuk data disatukan. Ditambah data yang mereka butuhkan mengenai informasi sekolah masih terpecah pada situs Dapodik.

Data mengenai kondisi sekolah pada situs Dapodik tertuang dalam halaman khusus untuk setiap sekolah. Ditambah pada situs Dapodik sering dilakukan sinkronisasi data, yang membuat data harus selalu *diupdate* berdasarkan sinkronisasi tersebut. Kondisi tersebut yang membuat sampai saat ini belum ada sebuah sistem informasi pada kedua instansi tersebut yang merangkum mengenai

kondisi sekolah. Ditambah, pada situs Dapodik, data mengenai atribut setiap sekolah masih tersebar hanya pada situs masing-masing sekolah dan belum ada suatu sistem yang mengumpulkan data-data tersebut sehingga memudahkan untuk melakukan perbandingan.

Kondisi ini membuat dibutuhkan sistem yang dapat digunakan oleh dua instansi pendidikan tersebut secara bersamaan untuk dapat melihat kondisi SMA Negeri di Jakarta Barat. Sebuah sistem yang dapat membantu melihat apakah kondisi murid, fasilitas serta sumber daya manusianya pada setiap sekolah di Jakarta Barat sudah berada pada kondisi yang merata. Kedua instansi juga dapat menggunakan sistem yang ada sebagai sarana untuk bekerja sama agar dapat membuat kebijakan yang optimal dan menyeluruh. Solusi untuk masalah ini adalah dengan membuat sebuah sistem informasi berbentuk *dashboard* yang bisa menyajikan data mengenai kondisi SMA Negeri di Jakarta Barat disertai dengan hasil *clustering* mengenai kondisi setiap sekolah.

Clustering dapat dengan mudah dikatakan sebagai salah satu metode yang paling umum digunakan untuk membentuk suatu kelompok data yang ada menjadi kelompok-kelompok lain yang terpisah yang memiliki kesamaan antar anggota kelompok dan perbedaan antar kelompok lain. (Nahjan *et al.* 2023)

Dalam penelitian ini, proses *clustering* akan menggunakan suatu algoritma yang umum digunakan yakni algoritma K-Means. Inti dari penggunaan algoritma K-Means adalah membagi suatu kelompok data besar menjadi kelompok-kelompok lainnya yang memiliki kedekatan antara objek data tersebut dengan pusat kelompoknya.

Banyaknya *cluster* yang digunakan pada algoritma K-Means dapat ditentukan menggunakan metode Elbow yang akan memberikan banyaknya *cluster* optimal yang dapat digunakan. Dan hasil akhir dari algoritma K-Means dapat diuji keefektifitasnya dengan Davies-Boulden Index yang menghitung rasio jarak antar *cluster*, semakin kecil nilainya maka semakin jauh atau berbeda tiap *cluster* tersebut.

Proses pembuatan sistem informasi berbentuk *dashboard* pada penelitian kali ini akan menggunakan bahasa MySQL sebagai basis datanya serta menggunakan HTML dan CSS untuk tampilan depannya dan juga menggunakan Python untuk *clustering* dan menghubungkan hasil *clustering* dengan *dashboard* (menggunakan *framework flask*).

Pada penelitian kali ini akan membuat 3 fitur pengelompokkan yakni, berdasarkan hubungan antara jumlah murid dengan jumlah fasilitas, jumlah murid dengan jumlah SDM dan jumlah murid dengan jumlah rombongan belajar.

Pemilihan fitur-fitur tersebut dipilih bukan hanya berdasarkan ketersediaan data yang terbatas pada situs Dapodik, tetapi juga karena pentingnya ketiga fitur tersebut

dalam mendukung pembelajaran. Menurut (Candra Kartika Sari *et al.* 2024), mengungkapkan bahwa murid yang di sekolahnya memiliki kemudahan akses terhadap fasilitas yang lebih lengkap dan lengkap baik cenderung lebih termotivasi untuk belajar. Sebaliknya, kurangnya fasilitas pendidikan dapat mengurangi keinginan siswa untuk berprestasi secara akademis.

Peran SDM (guru dan tenaga pendidik) juga menjadi hal yang dikelompokkan pada penelitian ini. Hal ini dikarenakan peran guru dan tenaga pendidik sangat penting dalam hal peningkatan mutu pendidikan. Proses pendidikan yang ada sekarang tidak bisa lepas dari fungsi guru dan tenaga pendidik dikarenakan mereka adalah titik vital dari proses tersebut. Tercukupinya kebutuhan SDM di sekolah dengan murid menjadi hal yang penting untuk mencapai hal diatas. (Kirana dan Rosita 2024)

Sedangkan untuk rombongan belajar, menurut penelitian yang dilakukan (Suryaningtyas dan Amin 2024), didapati hasil bahwa semakin kecil jumlah murid per rombongan belajar didapati bahwa sekolah tersebut memiliki nilai akreditasi yang lebih tinggi. Hal ini menunjukkan bahwa semakin kecil jumlah murid per rombongan belajar mendukung kegiatan pembelajaran menjadi lebih baik dan kondusif.

Dengan memanfaatkan metode ini, penelitian ini bertujuan untuk memberikan gambaran mengenai pengelompokkan sekolah-sekolah di Jakarta Barat berdasarkan atribut-atribut yang ada di dalam sekolah tersebut. Diharapkan juga melalui hadirnya sistem informasi beserta pengelompokkan dalam kedua instansi tersebut dapat memberikan bantuan bagi kedua instansi tersebut dalam mewujudkan pendidikan berkualitas yang merata yang tercermin pada meratanya fasilitas serta sumber daya pada setiap sekolah. Diharapkan penelitian ini dapat menjadi awal bagi pemetaan-pemetaan atribut-atribut lainnya atau menjadi contoh bagi wilayah lain.

1.2. Rumusan Masalah

Rumusan masalah berdasarkan latar belakang di atas adalah:

1. Bagaimana cara melakukan *clustering* SMA Negeri di Jakarta Barat berdasarkan hubungan antara jumlah murid dengan jumlah SDM, fasilitas dan rombongan belajar?
2. Bagaimana cara mengembangkan sistem informasi yang dapat memberikan informasi mengenai SMA Negeri di Jakarta Barat dan juga memberikan rekomendasi berdasarkan *clustering* hubungan jumlah murid dengan jumlah SDM, fasilitas dan rombongan belajar?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian berdasarkan SMA Negeri di wilayah Jakarta Barat.
2. Penelitian hanya berdasarkan atribut yang dikelompokkan menjadi berdasarkan jumlah fasilitas, jumlah murid dan jumlah sumber daya manusia.

1.4. Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui cara melakukan *clustering* SMA Negeri di Jakarta Barat.
2. Mengetahui cara membuat sebuah sistem informasi mengenai *clustering* SMA Negeri di Jakarta Barat.

Berdasarkan tujuan-tujuan diatas, penelitian ini juga memiliki manfaat penelitian untuk berbagai pihak. Manfaat penelitian dibagi dua yakni manfaat praktis yang dapat dirasakan langsung dan manfaat teoritis yang dapat digunakan sebagai pengembangan ilmu pengetahuan. Manfaat praktis dari penelitian ini adalah sebagai berikut :

Untuk mahasiswa:

1. Memberikan pengalaman praktis mengenai penerapan algoritma K-Means dalam *clustering*.
2. Menjadi sarana pembelajaran untuk mengembangkan kemampuan pengembangan sistem informasi dan visualisasi data.

Untuk instansi:

1. Menyediakan sistem informasi yang berisi data terkait kondisi SMA Negeri di Jakarta Barat beserta pemetaannya dan pengelompokkannya berdasarkan fasilitas, sumber daya manusia dan murid.
2. Memberikan informasi berdasarkan data yang dapat membantu mengambil keputusan untuk pemerataan kualitas setiap SMA Negeri yang ada di Jakarta Barat.

Manfaat teoritis dari penelitian ini adalah sebagai berikut :

1. Menambah pemahaman teoritis mengenai konsep *clustering*, khususnya K-Means.
2. Menambah pemahaman teoritis mengenai pembuatan sistem informasi yang menerapkan algoritma K-Means.

1.5. Sistematika Penulisan

Sistematika penulisan disusun agar pembaca dapat memahami isi penelitian secara sistematis dan teratur. Sistematika penulisan untuk penelitian ini disusun seperti berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan secara singkat mengenai latar belakang permasalahan mengapa memilih penelitian ini serta rumusan dan batasan masalah hingga tujuan dan manfaat penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan teori-teori yang mendasari pembahasan secara detail, dapat berupa definisi-definisi atau model matematis yang langsung berkaitan dengan ilmu atau masalah yang diteliti dan penelitian terdahulu.

Bab ini menjelaskan tentang teori-teori yang digunakan menjadi dasar dalam penelitian ini serta berisi pengertian-pengertian dan model matematis yang dipakai dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini menguraikan tentang tahapan penelitian, alat dan bahan serta jadwal penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang proses pengolahan data serta pembuatan sistem berdasarkan penelitian ini

BAB V PENUTUP

Bab ini menguraikan tentang kesimpulan dan saran penelitian berikutnya.

BAB 2. TINJAUAN PUSTAKA

2.1. *Clustering*

Clustering (klasterisasi) dapat diartikan sebagai pengelompokan objek data dalam data mining sehingga setiap anggota kelompok objek (*cluster*) memiliki kesamaan homogen yang lebih besar daripada dengan *cluster* lain. (Sutanto dan Susilawaty 2023). Ada banyak jenis pengaplikasian *clustering* berdasarkan metodenya, seperti Partisional, Hierarkis, Density-Based, Grid-Based dan Model-Based.

Pada penelitian ini, algoritma *clustering* yang dipakai adalah K-Means yang menerapkan *clustering* partisional. Menurut (Mustikasari dan Salman 2023), *Clustering* partisional adalah *clustering* yang mempartisi data ke dalam banyaknya kelompok k dengan suatu keunikan tertentu dan suatu objek data akan masuk ke dalam suatu kelompok k yang memiliki kedekatan dengannya.

2.2. Sekolah Menengah Atas Negeri

Sekolah Menengah Atas (SMA) menurut Peraturan Pemerintah Nomor 17 tahun 2010 adalah salah satu bentuk pendidikan formal di masyarakat yang bertugas untuk menyelenggarakan suatu kegiatan pendidikan umum pada jenjang pendidikan menengah sebagai suatu kelanjutan dari jenjang Sekolah Menengah Pertama, MTs atau bentuk lain dari jenjang tersebut yang sederajat.

SMA Negeri merupakan SMA yang dibangun dan dikelola langsung oleh pemerintah pusat maupun pemerintah daerah yang menaunginya. Hadirnya SMA Negeri bertujuan untuk menjadi suatu lembaga pendidikan yang diawasi langsung oleh pemerintah untuk menjadi penyedia layanan pendidikan tingkat menengah akhir yang kualitasnya terjamin.

Pembentukan sekolah negeri di semua jenjang termasuk SMA Negeri merupakan suatu bentuk pemenuhan kewajiban negara dalam hal ini pemerintah untuk memenuhi salah satu kebutuhan dasar masyarakatnya yakni pendidikan. Landasan hukum terkait SMA Negeri adalah UUD 1945 pasal 31 ayat 3 dan 4 dan juga UU Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional.

2.3. Suku Dinas Pendidikan Wilayah I dan Wilayah II Kota Jakarta Barat

Menurut halaman resmi Suku Dinas Pendidikan DKI Jakarta, setiap kota administrasi di Jakarta terdapat 2 Suku Dinas Pendidikan yang bertanggung jawab untuk setiap kecamatan yang berbeda di kota tersebut. Untuk wilayah kota Jakarta Barat terdapat Sudin Pendidikan Jakarta Barat Wilayah I yang membawahi kecamatan Cengkareng, Kalideres, Taman Sari dan Tambora serta untuk Sudin Pendidikan Jakarta Barat Wilayah II membawahi kecamatan Kembangan, Kebon Jeruk, Palmerah dan Grogol Petamburan.

Sampai tulisan ini dibuat, kedua instansi tersebut sementara dipimpin oleh orang yang sama yakni bapak Diding Wahyudin yang merangkap sebagai Kepala Sudin Pendidikan Wilayah I Jakarta Barat serta Plt. Kepala Sudin Pendidikan Wilayah II Jakarta Barat. Di kedua instansi tersebut juga dibagi berdasarkan jenjang pendidikan yang ada.

2.4. Data Pokok Pendidikan (Dapodik)

Dapodik adalah suatu sistem yang bertujuan untuk mengintegrasikan data dari setiap jenjang pendidikan yang dibuat oleh Kementerian Pendidikan karena kebutuhan data yang semakin kompleks di masa kini. Dapodik memiliki dasar hukum dari Instruksi Menteri Nomor 2 Tahun 2011 tentang Pengelolaan Data Pokok Pendidikan dan Permendikbud Nomor 79 tahun 2015 tentang Data Pokok Pendidikan.

Data yang dikelola oleh Dapodik menjadi suatu acuan data resmi untuk semua instansi pendidikan dan datanya dapat dilihat secara real time melalui website resmi Dapodik yakni dapo.kemendikdasmen.go.id. Data yang dikelola oleh Dapodik yang menggambarkan kondisi sekolah diantaranya adalah data sarana dan prasarana sekolah serta data sumber daya manusia dalam sekolah tersebut (murid, guru dan tenaga kependidikan).

2.5. Metode Elbow

Metode Elbow adalah salah satu metode yang digunakan dalam penggunaan algoritma K-Means yang bertujuan untuk menentukan jumlah kluster yang akan digunakan secara optimal. Cara kerja dari metode Elbow adalah dengan menerapkan perhitungan nilai SSE (Sum of Square Error) pada berbagai jumlah kluster (Maori 2023). SSE sendiri adalah perhitungan jarak antar setiap objek data di dalam sebuah *cluster* berbanding dengan titik pusat (centroid) dari *cluster* tersebut. Cara penggunaan metode Elbow adalah dengan memploting nilai SSE yang ada terhadap jumlah kluster yang tersedia. Metode Elbow akan dilakukan sampai beberapa iterasi yang dibutuhkan. Grafik yang menampilkan hasil perhitungan SSE ketika sudah menunjukkan adanya titik siku (perubahan nilai secara besar) dan disertai dengan mulai konsistennya data, maka iterasi metode Elbow dapat dihentikan.

Titik dimana terbentuknya siku serta diikuti dengan hasil selanjutnya yang mulai konsisten dalam grafik yang ditampilkan merupakan jumlah kluster yang optimal yang dapat digunakan pada proses *clustering* berikutnya. Menurut (Yuliana Sari *et al.* 2022), mengapa titik tersebut dipilih menjadi banyaknya kluster yang optimal, dikarenakan bahwa menambah lebih banyak kluster tidak secara optimal meningkatkan kualitas hasil *clustering*. SSE dapat dirumuskan sebagai berikut:

$$SSE = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \varphi_k)^2 \quad (2.1.)$$

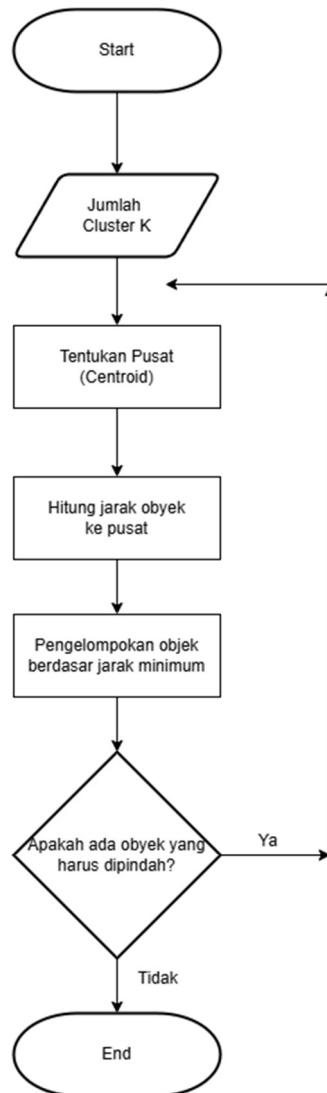
Rumus 2.1 adalah rumus untuk menghitung nilai SSE pada sebuah *cluster*. k melambangkan banyaknya jumlah *cluster*. x_i adalah titik objek data pada sebuah *cluster*. C_k melambangkan klaster ke- i . φ_k melambangkan titik centroid dari *cluster*. Perhitungan SSE dilakukan dengan mencari nilai kuadrat dari jarak titik objek dalam suatu *cluster* ke titik centroidnya.

2.6. Algoritma K-Means

K-Means adalah salah satu algoritma *clustering* partisional yang umum digunakan dalam data mining yang umumnya digunakan untuk menempatkan objek data ke dalam suatu *cluster* berdasarkan dengan kedekatannya dengan centroid dari *cluster* tersebut. (Sekar Setyaningtyas et al. 2022). Data yang berada dalam satu *cluster* yang sama memiliki kesamaan dibandingkan dengan data lain di *cluster* yang berbeda. Banyaknya *cluster* yang digunakan dalam proses iterasi pada algoritma K-Means dapat dilakukan dengan penentuan menggunakan metode elbow untuk hasil yang lebih optimal.

Cara kerja dari algoritma K-Means salah satunya dengan menggunakan penghitungan Euclidean Distance (Jarak Euclidean). Hasil dari jarak euclidean suatu objek data terhadap centroid dari *cluster-cluster* yang sudah ditentukan akan diurutkan dari yang terkecil dan *cluster* yang memiliki jarak antar centroidnya dengan objek datanya yang terkecil maka dipilih sebagai *cluster* bagi objek data tersebut. (Noor et al. 2021)

Proses iterasi dari algoritma K-Means terus berlanjut sampai di titik dimana anggota *cluster* dari iterasi sebelumnya tidak mengalami perubahan dengan anggota *cluster* dari iterasi terakhir (tidak ada objek data yang berpindah). *Cluster* serta anggota *cluster* pada iterasi terakhir menjadi *cluster* tetap yang menjadi bagian final dari proses algoritma K-Means. Alur kerja dari algoritma K-Means digambarkan pada alur di bawah:



Gambar 2.1 Alur proses algoritma K-Means

2.7. Davies-Bouldin Index

Davies-Bouldin Index (DBI) adalah sebuah metrik yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 yang digunakan sebagai sebuah cara untuk mengevaluasi hasil akhir dari suatu klasterisasi. DBI bekerja dengan cara mencari rata-rata dari rasio antara jarak titik-titik anggota tiap *cluster* dengan centroidnya (*inner-cluster distance*) dengan jarak antara centroid dengan centroid lain terdekatnya (*inter-cluster distance*). (Sopyan *et al.* 2022)

Penghitungan jarak tersebut dihitung dengan menggunakan perhitungan Euclidean Distance (Jarak Euclidean). Perhitungan dilakukan untuk setiap *cluster* yang ada dan nilai DBI merupakan hasil akhir dari rata-rata dari rasio maksimum untuk semua *cluster*. Semakin kecil nilai DBI berarti anggota tiap *cluster* memiliki persamaan yang kompak dan antar *cluster* yang lain memiliki perbedaan yang jauh

terpisah, yang berarti hasil *clusterisasi* semakin baik. (Umagapi *et al.* 2023) Rumus perhitungan DBI dapat dilihat di bawah:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (2.2.)$$

Rumus 2.2 merupakan rumus untuk menghitung nilai DBI pada suatu *cluster*. K menunjukkan jumlah *cluster*. S_i adalah rata-rata jarak dari tiap titik yang ada dalam *cluster i* dengan centroidnya. S_j adalah rata-rata jarak dari tiap titik yang ada dalam *cluster j* dengan centroidnya. M_{ij} jarak antar titik centroid dari *cluster i* dengan *cluster j*. Perhitungan DBI berdasarkan pada rata-rata dari nilai maksimum yang didapatkan dari rasio tersebut.

2.8. Python

Python adalah suatu bahasa pemrograman yang umum digunakan pada analisis data, machine learning dan juga pada deep learning. Python termasuk ke dalam high-level programming language (bahasa pemrograman tingkat tinggi) yang sedang populer saat ini karena penulisan kodenya yang terbilang mudah untuk digunakan. (Riziq sirfatullah Alfarizi *et al.* 2023)

Alasan lain mengapa python menjadi salah satu bahasa pemrograman yang populer digunakan karena menggunakan banyak konsep dari bahasa pemrograman lain lalu disederhanakan penggunaan serta penulisan kodenya. Python mengadopsi bahasa C sebagai bahasa yang bersifat prosedural, bahasa Java yang bersifat berorientasi objek dan bahasa Lisp sebagai bahasa yang fungsional. (Rahman *et al.* 2023)

Selain itu, kehandalan python dalam menangani data besar membuatnya umum digunakan untuk berbagai bidang dalam dunia komputer yang berhubungan dengan data seperti big data, data mining, data analytics maupun machine learning. Ditambah, banyaknya *library* yang tersedia juga memudahkan penggunaannya dalam mengeksplorasi penggunaan bahasa pemrograman ini.

2.9. MySQL

MySQL adalah salah satu server database yang termasuk ke dalam RDBMS (Relational Database Management System) yang dapat diartikan bahwa data pada setiap tabel pada database tersebut dapat dikelola secara berrelasi dengan tabel lain maupun database yang lain. MySQL umumnya digunakan sebagai sarana pembelajaran untuk mengenal RDBMS karena ukurannya yang terbilang ringan dan memiliki fitur tampilan yang bukan hanya kode tapi bisa dilakukan secara drag and drop menggunakan tools phpMyAdmin. MySQL juga umum dan cocok digunakan dengan bahasa pemrograman PHP membuatnya menjadi salah satu RDBMS yang paling banyak digunakan. (Hermiati Reza *et al.* 2021).

2.10. HTML (Hyper Text Markup Language)

HTML (Hyper Text Markup Language) dapat disebut sebagai sebuah kode program yang terdiri dari kumpulan tag-tag informasi yang merupakan sebuah format terhadap suatu informasi yang dimaksud. HTML juga merupakan dasar dari tampilan visual yang akan ditampilkan. HTML secara umum digunakan di mayoritas situs web yang ada di internet dikarenakan penggunaannya yang menggunakan metode semantik yang memudahkan para programmer dan pengembang dalam proses Development maupun dalam proses Maintenance. (Pranita Nasution *et al.* 2022)

2.11. CSS (Cascade Style Sheet)

CSS adalah kumpulan kode yang didalamnya merepresentasikan warna, layout maupun font pada tampilan halaman web. Hadirnya CSS membuat tampilan website menjadi lebih menarik dan bervariasi karena tidak terbatas pada tag-tag di HTML. Penerapan CSS dapat dengan membuat suatu file khusus CSS atau dapat disisipkan langsung dalam file HTML. CSS merupakan pasangan yang pas untuk HTML agar tampilan website semakin menarik dan bervariasi. (Ramadhani *et al.* 2023)

2.12. Blackbox Testing

Menurut (Nur Ichsanudin *et al.* 2022) , *black box testing* adalah sebuah tes yang dilakukan pada aplikasi atau system untuk melihat apakah sebuah aplikasi atau sistem bekerja dengan baik tanpa harus mengetahui bagaimana pemograman di baliknya. Dari 3 jenis pengujian yang ada di *black box testing* (Fungsional, Non-Fungsional dan Regulation Testing), yang akan dipakai di dalam penelitian ini adalah fungsional testing, dikarenakan sistem yang akan dibangun sangat bergantung pada fungsinya yakni menampilkan informasi dan hasil clustering, maka keberhasilan fungsi itu ditampilkan sangat penting.

2.13 Unified Modelling Language (UML)

Unified Modelling Language merupakan bahasa yang berfungsi untuk menggambarkan dan menunjukkan spesifikasi pada pengembangan suatu system yang divisualisasikan dalam bentuk gambar. UML juga berguna dalam proses dokumentasi pengerjaan suatu system. UML memberikan sebuah standar blue print dalam proses pembuatan system karena dapat menggambarkan proses bisnis, pembuatan basis data serta atribut-atribut yang dibutuhkan. (Siska Narulita *et al.* 2024). Berikut penjelasan mengenai diagram dalam UML yang digunakan pada penelitian ini :

a. Use Case Diagram

Use case diagram adalah sebuah diagram yang menggambarkan jalannya fungsi pada suatu system yang akan dibangun. Use case diagram menggambarkan scenario yang melibatkan hubungan antara

system dengan actor disertai dengan Batasan pada system tersebut yang dihubungkan dengan notasi garis.

b. Activity Diagram

Activity diagram adalah sebuah diagram yang berfungsi untuk menunjukkan alur proses bisnis suatu system yang dibangun. Activity diagram ditampilkan dengan menyajikan aktivitas pada sebuah system secara terstruktur dari awal hingga akhir proses yang menggambarkan juga peran dan fungsinya yang ditunjukkan dengan symbol dan notasi yang sesuai untuk setiap aktivitas.

c. Sequence Diagram

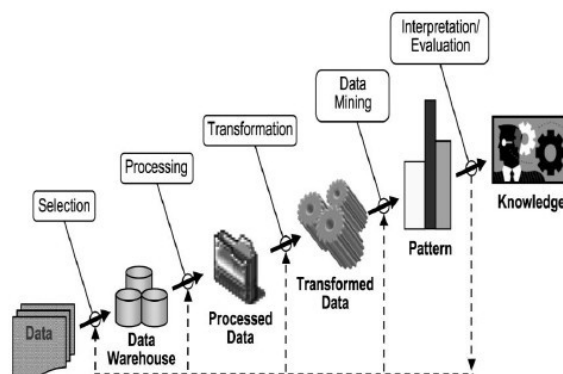
Sequence diagram adalah diagram dalam UML yang berfungsi untuk menunjukkan interaksi antar objek objek dalam sebuah system. Diagram ini menunjukkan bagaimana pesan dikirimkan dari setiap objek yang ada beserta hierarki serta urutan prosesnya.

d. Class Diagram

Class diagram adalah diagram yang menunjukkan hubungan antar kelas pada suatu system serta rincian mengenai aturan, atribut dan tanggung jawab masing-masing entitas yang menunjukkan perilaku berjalannya system. Class diagram berguna juga sebagai visualisasi dari struktur system yang akan dibangun berdasarkan jenis kelas yang ada.

2.14 Data Mining

Data mining dapat diartikan sebagai suatu proses menambang informasi penting yang didapatkan dari suatu data yang menerapkan beberapa disiplin ilmu seperti kecerdasan buatan, statistika, ilmu matematika dan sebagainya. Proses data mining nantinya akan mengekstrasi dan mengidentifikasi informasi yang bermanfaat dari sebuah basis data dengan menemukan pola pada kumpulan data tersebut. (Sudarsono *et al.* 2021). Data mining termasuk ke dalam proses Knowledge discovery yang memiliki tahapan seperti berikut :



Gambar 2.2 Alur proses Knowledge Discovery in Databases

Berdasarkan gambar diatas yang menjelaskan mengenai tahapan-tahapan pada knowledge discovery, yang terdiri dari data selection, preprocessing, transformation, data mining dan pattern evaluation. Penjelasan mengenai setiap tahapan pada knowledge discovery dapat dilihat di bawah ini :

- a. Data Selection, merupakan Langkah pertama yang dilakukan dalam proses Knowledge Discovery in Databases yang dimana pada proses ini dilakukan proses penyeleksian terhadap data yang dianggap relevean dari kumpulan data yang besar. Kemudian data yang telah diseleksi dari proses ini kemudian akan disiapkan untuk proses data mining ke depannya. Tujuan dari tahapan Data Selection ini adalah agar mempersiapkan data yang memiliki keterkaitan langsung dengan penelitian.
- b. Preprocessing, merupakan sebuah proses yang bertujuan agar memastikan data yang akan dipakai pada proses selanjutnya memiliki akurasi serta kualitas yang optimal. Pada proses preprocessing ini, dilakukan identifikasi terhadap baris data yang kosong, tidak lengkap ataupun tidak valid. Langkah ini penting dikarenakan dapat memastikan proses selanjutnya memproses serta menghasilkan output yang valid dan optimal.
- c. Transformation, merupakan sebuah langkah dimana data yang telah dibersihkan pada proses sebelumnya selanjutnya akan melalui proses dimana dilakukan perubahan struktur serta tipe data untuk disesuaikan dengan algoritma yang akan dijalankan. Langkah ini bertujuan agar setiap data yang diolah sesuai dengan kebutuhannya dan dapat memastikan proses data mining yang dijalankan dapat memberikan hasil yang optimal.
- d. Data mining, merupakan Langkah Utama dari proses KDD ini, dimana Teknik atau algoritma digunakan untuk menganalisis sekumpulan data yang digunakan. Tujuan Utama dari Langkah ini adalah mencari hubungan, pola atau hubungan yang tersembunyi dalam data yang kemudian akan dijadikan sebuah informasi penting.
- e. Pattern Evaluation, merupakan tahapan Ketika hasil yang diperoleh dari data mining akan dianalisis untuk mengidentifikasi pola pola yang bermanfaat untuk penelitian. Informasi yang dianggap sesuai dengan penelitian kemudian disimpan didalam basis pengetahuan

2.15 Penelitian Terdahulu

Tabel 2.1 Ringkasan penelitian terdahulu

No.	:	1s
Penulis	:	Sigrid Tamriesfatno, Riska Kherani, Sarwono Sarwono
Judul	:	Penerapan Metode <i>Clustering</i> dengan Algoritma K-Means untuk Pengelompokkan Data Sekolah Menengah di Kabupaten Muna Barat