

BAB I

PENDAHULUAN

1.1. Latar Belakang

Data mining merupakan suatu langkah dalam knowledge discovery in databases (KDD) yang memiliki teknik menganalisa data untuk digali informasi tersembunyi dalam jumlah besar dan kompleks, sehingga menghasilkan output berupa karakteristik atau pola dari data tersebut. Salah satu teknik analisa Data Mining adalah analisis kelompok (cluster analysis) yang lebih dikenal dengan Clustering (Darmi Y.,&Setiawan,A,2016).

Clustering adalah salah satu teknik analisis data yang digunakan untuk mengelompokkan data berdasarkan kemiripan atau jarak antar data. Teknik ini banyak digunakan di berbagai bidang, seperti pengelompokan wilayah geografis, analisis pasar, sistem zonasi pendidikan, hingga optimasi logistik (Xu,R.,&Wunsch,D, 2005). Algoritma clustering memungkinkan pembagian data menjadi beberapa kelompok sehingga pola tersembunyi dalam data dapat diidentifikasi. Namun, algoritma clustering klasik seperti K-Means, sering kali menghadapi beberapa keterbatasan, terutama dalam menangani data dengan distribusi yang tidak seimbang (MacQueen,J, 1967) dan K-Means mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal cluster. Keterbatasan ini menjadi masalah besar ketika clustering diterapkan dalam skenario seperti Zonasi Pendidikan.

Pendidikan yang berkualitas merupakan hak setiap warga negara sebagaimana diamanatkan dalam Pasal 31 UUD 1945. Namun, dalam praktiknya, pemerataan akses pendidikan masih menjadi tantangan di Indonesia. Penetapan sistem zonasi diberlakukan di setiap jenjang sekolah oleh pemerintahan daerah berdasarkan kewenangan dengan prinsip dasar yaitu domisili peserta didik dengan sekolah berdekatan berdasarkan Permendikbud No. 51 pasal 20 ayat 1. Sistem zonasi di Indonesia merupakan

langkah strategis jangka Panjang pemerintah dalam menentukan kebijakan sebagai upaya percepatan dalam penyelesaian permasalahan pendidikan, Adapun permasalahan utama yang dihadapi yaitu pemerataan kualitas pendidikan, pemerintah menjamin adanya layanan pendidikan bagi peserta didik serta menjadi sentral untuk kebijakan perumusan distribusi tenaga pengajar, ketersediaan sarana prasarana fasilitas sekolah serta wajib belajar 12 tahun bagi pelajar (Risal,2023).

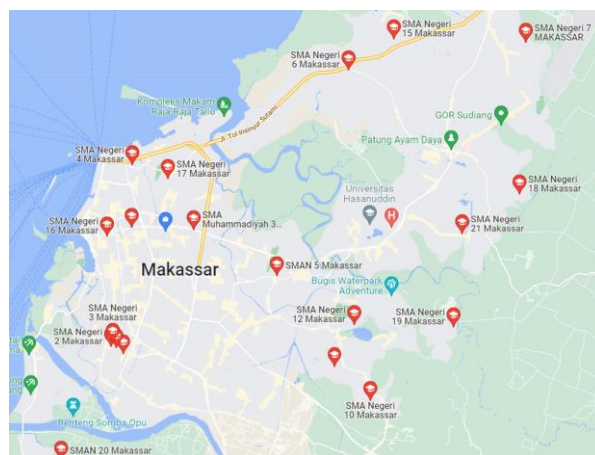
Sistem zonasi sekolah adalah proses mengidentifikasi wilayah geografis untuk tujuan mengalokasikan anak-anak usia sekolah yang tinggal di suatu pemukiman ke beberapa sekolah umum untuk tujuan sosial yang didukung oleh kekuatan politik. Selain itu, sistem zonasi sekolah memiliki peran dan kedudukan serta fungsi yang sama dalam memberikan pelayanan kepada setiap siswa. Seperti halnya siswa dengan prestasi akademik di atas rata-rata dapat ditempatkan di sekolah yang berbeda sehingga tidak berkumpul di beberapa sekolah tertentu yang disebut favorit atau sekolah unggulan, dengan demikian menghilangkan diskriminasi berbasis akademik yang menyatakan kasta atau pelabelan sekolah favorit (unggulan) dan tidak favorit terkhusus untuk sekolah negeri di Masyarakat (Risal,2023).

Dengan hal tersebut tujuan dari pemerataan kualitas pendidikan melalui Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 dapat meningkatkan kualitas semua sekolah, dengan pemerataan peserta didik yang berprestasi sehingga dapat bersaing secara merata. Sistem zonasi yang diterapkan sejak 2017 melalui kebijakan Kementerian Pendidikan dan Kebudayaan bertujuan untuk memastikan bahwa siswa memiliki akses ke sekolah yang dekat dengan tempat tinggal mereka. Kebijakan ini sekaligus diharapkan dapat mengurangi ketimpangan antara sekolah favorit dan non-favorit (Kemendikbud RI, 2017). Meskipun bertujuan baik, pelaksanaan sistem zonasi menghadapi berbagai kendala yakni Ketidakseimbangan Distribusi Siswa, banyak sekolah yang berada di pusat kota kelebihan siswa, sedangkan sekolah di daerah pinggiran kekurangan siswa (Yusuf, R, 2021). Kualitas Sekolah yang Tidak Merata,

orang tua cenderung tetap menginginkan anaknya masuk sekolah dengan reputasi baik, meskipun lokasi jauh (Nugroho, T, 2019). Kurangnya Infrastruktur Pendukung, penentuan zonasi sering kali dilakukan tanpa memanfaatkan teknologi yang memadai, sehingga perhitungan jarak dan distribusi siswa tidak optimal (Wahyuni, A., et al. 2019).

Penelitian ini mengusulkan algoritma clustering *Fixed-Centroid* dengan *Balanced member* untuk mengatasi permasalahan ketidakseimbangan dalam sistem zonasi. Algoritma ini dirancang dengan pendekatan posisi centroid ditentukan berdasarkan lokasi sekolah yang telah ditetapkan dan keseimbangan cluster dalam proses clustering diatur agar jumlah anggota setiap cluster seimbang, sehingga dapat mengoptimalkan kapasitas sekolah di setiap zona.

Melalui algoritma ini, diharapkan sistem zonasi dapat menjadi lebih adil, efisien, dan mampu mengakomodasi kebutuhan distribusi siswa sesuai kapasitas sekolah. Wilayah sistem zonasi sekolah yang akan di gunakan adalah Kota Makassar.



Gambar 1. Lokasi Titik Sekolah Wilayah Makassar

Pada gambar 1 terlihat bahwa terdapat sekolah yang lokasinya berdekatan maka berdasarkan permasalahan tersebut maka dilakukan usulan model algoritma dengan nilai awal centroid ditetapkan berdasarkan kriteria pembentuk kelompok yang sudah diketahui di mana titik sekolah tidak dapat

berubah ataupun berpindah tempat, sehingga dapat dijadikan sebagai titik centroid (pusat cluster), sehingga akan diperoleh hasil yang optimal.

1.2. Rumusan Permasalahan

Berdasarkan latar belakang masalah di atas, maka rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana merancang algoritma clustering dengan centroid tetap (*fixed centroid*) yang dapat menghasilkan pembagian cluster dengan jumlah anggota yang seimbang?
2. Bagaimana mengevaluasi kinerja algoritma yang diusulkan ?

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut :

1. Mengembangkan algoritma clustering berbasis centroid tetap yang mampu menjaga keseimbangan jumlah anggota cluster.
2. Menguji performa algoritma yang diusulkan berdasarkan hasil yang diperoleh pada proses *Clustering*.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang algoritma clustering, dengan memperkenalkan metode *Balanced Fixed-Centroid Clustering* yang mampu menghasilkan cluster dengan anggota yang seimbang dan centroid tetap.
2. Penelitian ini memperkaya kajian dalam data mining, khususnya pada aplikasi algoritma clustering untuk data dengan distribusi tidak merata.
3. Menjadi referensi bagi peneliti selanjutnya yang ingin mengembangkan algoritma serupa untuk berbagai aplikasi dunia nyata.

1.5. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

1. Dataset yang digunakan dalam penelitian ini terdiri dari data simulasi atau data sekolah dan data siswa mencakup koordinat geografis, seperti lintang dan bujur. Penelitian ini tidak mencakup pengumpulan data baru tetapi menggunakan data yang sudah ada.
2. Lokasi centroid yang digunakan dalam algoritma telah ditentukan sebelumnya berdasarkan lokasi sekolah. Penelitian ini tidak membahas metode penentuan lokasi centroid secara dinamis atau optimalisasi lokasi centroid.
3. Clustering dalam penelitian ini sepenuhnya didasarkan pada jarak geografis antara data lokasi siswa dan centroid yakni lokasi sekolah.
4. Keseimbangan cluster diukur berdasarkan jumlah anggota dalam setiap cluster. Aspek lain, seperti keseragaman atribut anggota cluster (misalnya, karakteristik siswa), tidak dibahas dalam penelitian ini.

BAB II

TINJAUAN PUSTAKA

2.1. Kajian Pustaka

2.1.1. Sistem Zonasi

Sistem zonasi di Indonesia merupakan kebijakan dalam penerimaan peserta didik baru (PPDB) yang diterapkan oleh Kementerian Pendidikan dan Kebudayaan (Kemendikbud) sejak tahun 2017. Kebijakan ini bertujuan untuk meningkatkan pemerataan akses pendidikan bagi semua peserta didik tanpa diskriminasi, mengurangi stigma sekolah favorit dan non-favorit dan mendorong pemerintah daerah untuk meningkatkan kualitas sekolah di seluruh wilayah (Kemendikbud RI, 2017). Melalui sistem zonasi, siswa diutamakan diterima di sekolah yang berada dalam jarak terdekat dengan tempat tinggalnya, dengan pertimbangan kapasitas sekolah. Sistem zonasi diatur dalam Permendikbud No. 14 Tahun 2018 tentang Penerimaan Peserta Didik Baru (PPDB). Dalam kebijakan ini, kuota penerimaan dibagi menjadi 80% untuk jalur zonasi, di mana siswa diterima berdasarkan kedekatan domisili dengan sekolah 15% untuk jalur prestasi, yang mempertimbangkan nilai akademik atau non-akademik. 5% untuk jalur perpindahan orang tua/wali. Namun, pelaksanaan sistem ini sering kali menghadapi kendala di lapangan, seperti kesenjangan kualitas sekolah dan ketimpangan jumlah siswa (Anwar, M. K., & Safitri, R, 2020). Pelaksanaan sistem zonasi di Indonesia menghadapi beberapa tantangan salah satu diantaranya adalah ketidakseimbangan distribusi siswa. Zonasi sering kali menyebabkan beberapa sekolah kekurangan siswa, sementara sekolah lain kelebihan kapasitas. Hal ini diperparah oleh distribusi penduduk yang tidak merata di beberapa wilayah (Sudjana, T, 2021).

2.1.2. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data Knowledge Discovery in Databases (Alkhairi & Windarto, 2019).

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan (Rosmini et al., 2018), yaitu:

1. Deskripsi

Para peneliti dan analisis biasanya mencoba menemukan cara untuk menggambarkan pola dan trend yang tersembunyi dalam data.

2. Estimasi

Estimasi memiliki kemiripan dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi memiliki kemiripan dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa yang akan datang. Beberapa algoritma dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Sebagai contoh pengklasifikasian persediaan dalam tiga kelas, yaitu persediaan tinggi, persediaan sedang dan persediaan rendah.

5. Clustering

Clustering merupakan teknik pengelompokan record data, pengamatan atau kasus dalam kelas yang memiliki kemiripan. Cluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record lain dalam cluster.

6. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu. Dalam dunia bisnis lebih umum disebut sebagai analisis keranjang belanja.

Menurut (Sihombing et al., 2012) Data mining memiliki kelebihan sebagai alat analisis seperti :

- a. Data mining ini mampu menangani data-data dalam jumlah besar dan juga kompleks.
- b. Data mining juga dapat menangani data-data dengan berbagai macam tipe atribut.
- c. Data mining mampu mencari dan mengolah data secara otomatis. Disebut semi otomatis karena dalam beberapa teknik data mining, diperlukan parameter yang harus di input oleh user secara manual.
- d. Data mining dapat menggunakan pengalaman ataupun kesalahan terdahulu untuk meningkatkan kualitas dan hasil Analisa sehingga mendapat hasil yang terbaik.

Menurut (Sihombing et al., 2012) Data mining memiliki kelemahan dalam pencarian data tidak mencari secara individualis, tetapi set individualis, atau dengan kata lain dikelompokkan dengan kriteria-kriteria tertentu.

2.1.3. *Clustering*

Clustering merupakan suatu proses pengelompokkan record suatu , observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan clustering dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan suatu pengelompokan pada proses clustering. *Clustering* sering dilakukan sebagai untuk langkah awal dalam proses data mining saat melakukan suatu metode analisis. Terdapat banyak algoritma *Clustering* yang telah digunakan oleh peneliti sebelumnya seperti *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS* dan *Fuzzy Subtractive*. Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, namun prinsip algoritma sama, yaitu mengelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok (P, Pedamkar, 2021).

2.1.4. *Fixed-Centroid Clustering*

Fixed-Centroid Clustering adalah pendekatan clustering di mana lokasi centroid tetap selama proses pengelompokan. Pendekatan ini sering digunakan dalam aplikasi berbasis lokasi, seperti Zonasi wilayah pendidikan (zonasi sekolah) dan Distribusi logistik berdasarkan depot atau pusat layanan. Penelitian sebelumnya menunjukkan bahwa *fixed-centroid clustering* dapat menghasilkan hasil yang stabil, tetapi metode ini sering mengabaikan keseimbangan jumlah anggota *cluster*, yang penting dalam banyak aplikasi dunia nyata (MacQueen, J, 1967).

2.1.5. *Balancing dalam clustering*

Balancing dalam *clustering* merujuk pada usaha untuk memastikan bahwa jumlah anggota dalam setiap cluster adalah relatif seimbang. Dalam konteks tertentu, seperti sistem zonasi pendidikan, *balancing* sangat penting

karena *cluster* yang terlalu kecil atau terlalu besar dapat menyebabkan ketidakadilan atau ketidakefisienan. Adapun rumus yang digunakan sebagai berikut :

$$M = \frac{\sum A_m}{\sum C_n} \quad (1)$$

Dimana A_m diinisialisasikan sebagai jumlah siswa dan C_n diinisialisasikan sebagai jumlah sekolah, dilakukan proses pembagian sehingga mendapatkan jumlah anggota atau siswa yang seimbang.

2.1.6. *Haversine distance*

Perhitungan jarak pada penelitian ini menggunakan *Haversine distance*. *Haversine Distance* adalah metode untuk menghitung jarak antara dua titik di permukaan bumi menggunakan koordinat geografis (lintang dan bujur). Metode ini memperhitungkan kelengkungan bumi, sehingga memberikan hasil yang lebih akurat dibandingkan jarak Euclidean pada bidang datar. Rumus *Haversine Distance* sebagai berikut :

$$d = 2r \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cdot \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (2)$$

Keterangan :

- d : jarak antar titik (dalam km atau mil, tergantung pada jari-jari r).
- r : jari-jari bumi, umumnya 6,371 km.
- φ_1, φ_2 : lintang (latitude) dari titik pertama dan kedua, dalam radian
- λ_1, λ_2 : bujur (longitude) dari titik pertama dan kedua, dalam radian
- $\Delta\varphi = \varphi_1 - \varphi_2$: perbedaan lintang
- $\Delta\lambda = \lambda_1 - \lambda_2$: perbedaan bujur.

2.1.7. Perhitungan Jarak Rata-rata

Perhitungan jarak rata-rata digunakan untuk mengukur total jarak antara anggota cluster dengan centroidnya dalam setiap cluster. Berikut rumus yang digunakan :

$$\frac{1}{n_i} \sum_{j=1}^{n_i} distance(C_i, A_{i,j}) \quad (3)$$

Keterangan :

- n_i : Jumlah anggota (data) dalam klaster i
- $\sum_{j=1}^{n_i}$ = Operasi penjumlahan untuk semua anggota data (j) dalam klaster i
- $distance(C_i, A_{i,j})$ = Jarak antara centroid klaster C_i dan anggota data ke-j dalam klaster i.
- $\frac{1}{n_i}$ = Membagi total jarak dengan jumlah anggota klaster untuk mendapatkan rata-rata jarak

2.2. Metode Penyelesaian Masalah

2.2.1. State of the Art Penelitian

Penelitian terkait yang telah dilakukan sebelumnya dapat dilihat pada Tabel 2.1 di bawah ini.

Tabel 1. State of the Art Penelitian

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
1.	Judul : Usulan Algoritma Untuk <i>Distance Based Clustering</i> Dengan <i>Fixed Centroid</i> Dan <i>Balanced Member</i> Penulis : -	Objek : Sekolah Permasalahan : Bagaimana model algoritma yang diusulkan untuk <i>distance based clustering</i> dengan <i>fixed centroid</i> dan <i>balanced</i>	Usulan model algoritma clustering	Diharapkan algoritma clustering dapat mencari atau mengelompokkan keanggotaannya dengan <i>balanced</i> yang dimana titik pusat

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
		<i>member</i>		clusternya sudah ditentukan terlebih dahulu, sehingga akan diperoleh hasil yang optimal
2.	Judul : Tag clustering algorithm LMMSK: improved K-means algorithm based on latent semantic analysis Penulis : Jing Yang dan Jun Wang Tahun : 2017 Penerbit : IEEE	Objek : Tag Sosial Permasalahan : kurangnya semantik eksplisit dalam sistem penandaan sosial	MMSK-Means, Latent Semantic Analysis (LSA), LMMSK	Dari hasil pengujian ketiga metode, bahwa metode LMMSK memiliki kecenderungan asosiasi semantik dan signifikansi yang lebih praktis, pengelompokan yang lebih baik serta memiliki akurasi yang optimal.
3.	Judul : Cluster based Zoning of Crime Info Penulis : Thota et. Al Tahun :	Objek : Kriminal Permasalahan : Bagaimana membangun zona cluster kriminologi	K-Means	Dataset kejahatan diinput ke software WEKA untuk membangun zona cluster menggunakan metode K-

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
	2017 Penerbit : IEEE			means cluster. Algoritma pengelompokan Kmeans mengelompokkan objek tergantung pada karakteristiknya. a. Grup dikembangkan dengan meminimalkan jumlah kuadrat jarak antara data dan centroid grup yang bersangkutan.
4.	Judul : Urban Area Function Zoning Based on User Relationships in Location-Based Social Networks Penulis : Hao et al Tahun : 2020 Penerbit : IEEE	Objek : Wilayah perkotaan Permasalahan : Bagaimana mengelompokkan titik-titik wilayah diperkotaan untuk usulan pembangunan fasilitas pendukung	DBSCAN yang dimodifikasi	Algoritma pertama menghitung 'pusat pengguna di jejaring sosial, dan kemudian mengubah sentralitas pengguna menjadi pusat lokasi melalui data check-in pengguna. Setelah mendapatkan sentralitas

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
				masing-masing lokasi, lokasi-lokasi diskrit dikelompokkan menurut empat batasan radius yang diberikan. Setelah dilakukan clustering, hasil zonasi fungsi kawasan perkotaan dapat dicapai. Reputasi titik-titik interior wilayah geografis dibagi adalah reputasi total wilayah geografis.
5.	Judul : School zoning system using K-Means algorithm for high school students in Makassar City Penulis : Febriana et al Tahun :	Objek : Sekolah Permasalahan : kurangnya pemerataan mutu Pendidikan	K-Means, Distance Euclidean	metode k-means dengan penyebaran data siswa yang relatif baik berdasarkan parameter jarak siswa dan sekolah, hasil klaster dari 22 sekolah dan

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
	2019 Penerbit : IEEE			1547 siswa. Hasil klaster menunjukkan bahwa setiap sekolah memiliki anggota klaster sendiri berupa data sebaran siswa.
6.	Judul : Application of k- means clustering to environmental risk zoning of the chemical industrial area Penulis : Shi, et.al Tahun : 2013 Penerbit : IEEE	Objek : Zonasi Wilayah Industri Kimia Permasalahan : Pencemaran lingkungan menjadi semakin sering karena ekonomi China terus berkembang. Kawasan industri kimia terdiri dari perusahaan kimia yang bertanggung jawab untuk melepaskan bahan kimia yang menumpuk dan menyebabkan pencemaran lingkungan, sehingga	K-Means Clustering	Kinerja sistem mampu mengklasterka n zona wilayah industri kimia

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
		menimbulkan risiko kesehatan bagi orang-orang di dalam dan di sekitar kawasan serta kerusakan lingkungan		
7.	Judul : Defining urban logistics profile zones in South American metropolis by combining functional and spatial clustering techniques Penulis : Regal, et.al Tahun : 2019 Penerbit : IEEE	Objek : Angkutan Barang Perkotaan Permasalahan : Bagaimana merancang sebuah framework untuk menangani masalah logistik pada kota metropolitan	UMAP dan HDBSCAN	Berhasil membuat framework untuk menangani masalah logistik pada kota metropolitan
8.	Judul : Efficiency Analysis of Machine Learning Intelligent Investment Based on K-Means Algorithm	Objek : Analisis Investasi Permasalahan : Bagaimana menganalisis data untuk	K-Means	Hasil algoritma K-Means dapat mengklusterkan data untuk memanfaatkan pada investasi cerdas

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
	Penulis : Liang Li, et.al Tahun : 2020 Penerbit : IEEE	pemanfaatan investasi cerdas		
9.	Judul : Zoning of Cultivated Land Intensive Use in Hubei Province Based on K-means Clustering Penulis : Cheng Qi Tahun : 2018 Penerbit : IEEE	Objek : Zonasi Lahan Budidaya Permasalahan : Bagaimana mengurangi kontradiksi antara penawaran dan permintaan lahan budidaya, tetapi juga merupakan langkah penting untuk mempromosikan pembangunan sosial dan ekonomi	K-Means Clustering	hasil zonasi yang baik dari penggunaan lahan budidaya secara intensif dengan menggunakan karakteristik singkat dan kecepatan tinggi telah dibuat. (2) Ada perbedaan spasial yang jelas dari penggunaan lahan budidaya intensif di Provinsi Hubei. Hasilnya dikategorikan oleh zona intensif rendah barat, zona intensif tengah selatan, zona intensif tinggi timur dan zona pertanian halus tengah.

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
				Strategi pengembangan yang sesuai juga diusulkan sesuai dengan kondisi produktif pertanian masing-masing zona.
10.	Judul : A Novel Approach to the Optimization of a Public Bus Schedule Using K-Means and a Genetic Algorithm Penulis : Shima et.al Tahun : 2021 Penerbit : IEEE	Objek : Jadwal Bus Permasalahan : Bagaimana optimalisasi jadwal bus menggunakan metode berdasarkan K-means dan algoritme genetika Dengan tujuan memberikan layanan yang andal kepada penumpang dengan biaya yang wajar.	K-Means Clustering	Hasil dari algoritma K-means digunakan sebagai nilai fitness objektif untuk optimasi jadwal bus menggunakan algoritma genetika. Hasil percobaan menunjukkan bahwa model optimasi yang diusulkan dapat meningkatkan waktu tinggal sementara mempertahankan biaya operasi pada tingkat saat ini atau kurang, dan peningkatan yang luar biasa dalam tingkat operasi dicapai dalam studi

No	Judul Karya Ilmiah, Nama, Tahun Terbit dan Penerbit	Objek dan Permasalahan	Metode Penyelesaian	Kinerja
				kasus. Model yang diusulkan mampu secara efektif mengoptimalkan pengeluaran keuangan dan memungkinkan operator bus untuk memenuhi permintaan penumpang dengan cara yang saling memuaskan.
11.	Judul : GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis Penulis : Praene et.al Tahun : 2019 Penerbit : Elsevier	Objek : Zonasi Iklim Permasalahan : Bagaimana memperoleh pengetahuan yang akurat mengenai kondisi iklim di setiap wilayah	HCPC, K-means, PCA, GIS	Hasil yang diperoleh menunjukkan beberapa konsistensi karakteristik daerah dengan variasi lokasi.

Tabel 2.1 merupakan tabel ringkasan yang memuat penelitian-penelitian sebelumnya yang memiliki keterkaitan dengan penelitian yang akan dilakukan.

Tabel 2.1 memaparkan keterangan mengenai judul, penulis, tahun terbit, objek dan permasalahan, metode penyelesaian dan hasil penelitian.

Berdasarkan *State of The Art* yang ditunjukkan pada Tabel 2.1, maka pada penelitian ini akan mengusulkan sebuah algoritma clustering yang dapat mencari atau mengelompokkan keanggotaannya dengan balanced yang dimana titik pusat clusternya sudah ditentukan terlebih dahulu, sehingga akan diperoleh hasil yang optimal.

2.2.2. Metode Penelitian

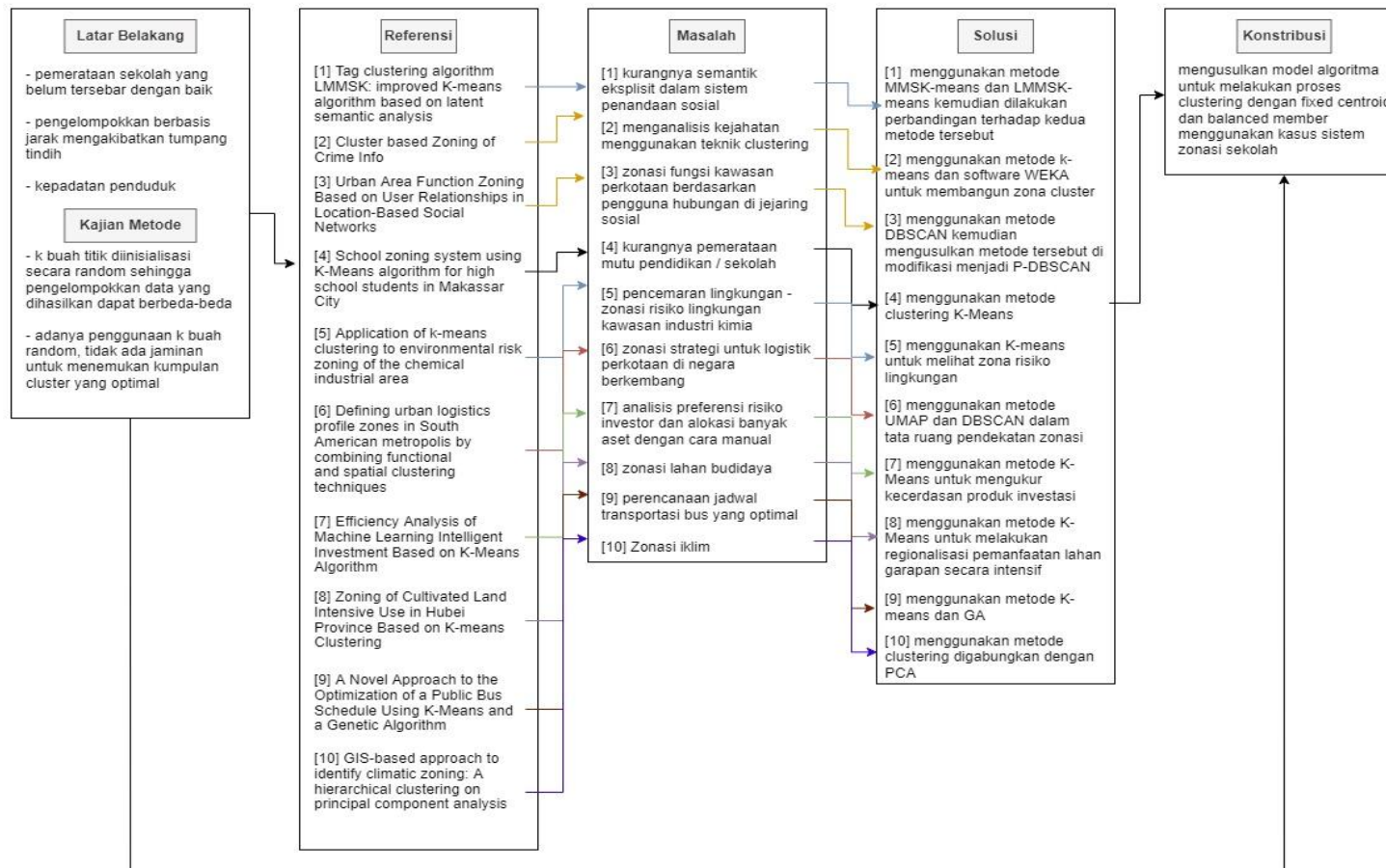
Penelitian ini mengusulkan tiga model algoritma clustering berbasis jarak. Lokasi sekolah sebagai centroid tetap, dengan data alamat siswa sebagai titik data. Kemudian dilakukan proses balanced member dan melakukan proses *clustering* berdasarkan jarak terdekat dengan centroid. Evaluasi terhadap algoritma dilakukan dengan cara menghitung total rata-rata jarak antar cluster. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam meningkatkan pemahaman dalam bidang clustering berbasis jarak dengan memperkenalkan algoritma baru yang mempertimbangkan aspek centroid tetap dan keseimbangan anggota serta menyediakan pendekatan yang lebih efisien dan adil untuk pengelompokan data berbasis lokasi, khususnya dalam konteks sistem zonasi pendidikan.

2.3. Target Hasil Penelitian

Berdasarkan tabel *State of The Art* penelitian yang telah dilakukan, target hasil dalam penelitian ini adalah penelitian ini mampu menemukan titik centroid yang tidak berubah-ubah serta melakukan *clustering* sehingga mendapatkan anggota cluster yang seimbang. Penelitian ini juga diharapkan mampu memberi pemahaman kepada pembaca dalam penerapan usulan algoritma *Clustering* dalam melakukan pengelompokan data berdasarkan titik pusat *cluster* yang telah ditentukan.

2.4. Kerangka Pikir

Kerangka pikir dapat dilihat pada Tabel, yang menjelaskan mengenai alur penelitian yang akan dilakukan



Gambar 2. Kerangka Pikir