

**SISTEM ZONASI SEKOLAH UNTUK PENERIMAAN SISWA  
MENGUNAKAN ALGORITMA  
*CONSTRAINED K-MEANS***

*School Zoning System for Student Admission Using  
Constrained K-Means Algorithms*

**ANDI ALVIADI NUR RISAL  
D032191016**



**PROGRAM STUDI MAGISTER TEKNIK ELEKTRO  
DEPARTEMEN TEKNIK ELEKTRO  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2023**

**PENGAJUAN TESIS**

**SISTEM ZONASI SEKOLAH UNTUK PENERIMAAN SISWA  
MENGUNAKAN ALGORITMA  
*CONSTRAINED K-MEANS***

Tesis

Sebagai Salah Satu Syarat untuk Mencapai Gelar Magister  
Program Studi Teknik Elektro

Disusun dan diajukan oleh:

ttd

**ANDI ALVIADI NUR RISAL  
D032191016**

Kepada

**FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2023**

# TESIS

## SISTEM ZONASI SEKOLAH UNTUK PENERIMAAN SISWA MENGGUNAKAN ALGORITMA *CONSTRAINED K-MEANS*

**ANDI ALVIADI NUR RISAL**  
**D032191016**

Telah dipertahankan dihadapan Panitia Ujian Tesis yang dibentuk dalam rangka penyelesaian studi pada Program Magister Teknik Elektro Fakultas Teknik Universitas Hasanuddin  
Pada tanggal 15 Juni 2023  
Dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama



**Dr. Ir. Zahir Zainuddin, M.Sc.**  
NIP. 19640427 198910 1 002

Pendamping Pembimbing



**Dr. Eng. Ir. Muhammad Niswar, S.T., M.IT.**  
NIP. 19730922 199903 1 001

Dekan Fakultas Teknik  
Universitas Hasanuddin



**Prof. Dr. Eng. Ir. Muhammad Isran Ramli, S.T., M.T., IPM.**  
NIP. 19730926 200012 1 002

Ketua Program Studi  
S2 Teknik Elektro



**Dr. Eng. Ir. Wardi, S.T., M.Eng.**  
NIP. 19720828 199903 1 003

## PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Yang bertanda tangan di bawah ini :

Nama : Andi Alviadi Nur Risal

Nomor Mahasiswa : D032191016

Program Studi : Magister Teknik Elektro

Dengan ini menyatakan bahwa, tesis yang berjudul “Sistem Zonasi Sekolah Untuk Penerimaan Siswa Menggunakan Algoritma *Constrained K-Means*” adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Ir. Zahir Zainuddin, M.Sc. dan Dr. Eng. Ir. Muhammad Niswar, S.T., M.IT.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal/Prosiding (*The 11<sup>th</sup> Institute of Electrical and Electronics Engineers (IEEE) International Conference on Communication, Networks, and Satellite (ComNetSat)* tahun 2022) sebagai artikel dengan judul “*School Zoning System for Student Admission using Constrained K-Means Algorithms*”.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Gowa, 20 Juni 2023

Yang Menyatakan



Andi Alviadi Nur Risal

## KATA PENGANTAR

Alhamdulillah rabbil'alamin, segala puji bagi Allah Subhanahu Wa Ta'ala Yang Maha Sempurna, yang telah memberikan rahmat, hidayah dan pertolongan-Nya sehingga penulis dapat menyelesaikan tesis dengan judul "**Sistem Zonasi Sekolah untuk Penerimaan Siswa menggunakan Algoritma *Constrained K-Means***". Tak lupa pula shalawat dan salam kepada Nabi Muhammad SAW yang telah menyinari dunia ini dengan keindahan ilmu dan akhlak yang diajarkan kepada seluruh umatnya.

Tesis ini disusun untuk memenuhi persyaratan untuk memperoleh gelar Magister Teknik (M.T.) pada Program Pascasarjana Departemen Teknik Elektro, Konsentrasi Teknik Informatika Universitas Hasanuddin Makassar. Tentunya penyelesaian tesis ini tidak terlepas dari dukungan dan bantuan dari semua pihak. Untuk itu, dengan penuh kerendahan hati penulis menyampaikan terima kasih setulus-tulusnya dan setinggi-tingginya kepada:

1. Bapak Dr. Ir. Zahir Zainuddin, M.Sc. sebagai pembimbing utama dan bapak Dr. Eng. Ir. Muhammad Niswar, S.T., M.IT. selaku dosen pembimbing pendamping yang telah meluangkan waktunya kepada penulis untuk membimbing, memberikan masukan, memotivasi tiada henti-hentinya hingga tahap penyelesaian tesis ini.
2. Bapak dan ibu Merna Baharuddin, S.T., M.Tel.Eng., Ph.D., Dr. Hasniaty A, S.T., M.T., dan Dr. Ir. Yustinus Upa Sombolayuk, M.T. selaku penguji yang memberikan masukan dan saran yang membangun selama proses penelitian berlangsung.
3. Bapak dan Ibu dosen serta Staf Program Studi Program Studi S2 Departemen Teknik Elektro Fakultas Teknik Universitas Hasanuddin atas bimbingan, bantuan dan arahnya selama menempuh perkuliahan.
4. Bapak Dr. Eng. Ir. Wardi, S.T., M.Eng. selaku Ketua Program Studi S2 Departemen Teknik Elektro, yang telah banyak mendukung dan membantu selama penulis menempuh pendidikan pascasarjana di Universitas Hasanuddin.

5. Rektor Universitas Hasanuddin dan Dekan Fakultas Teknik Universitas Hasanuddin yang telah memfasilitasi saya menempuh program Program Studi Magister.
6. Ayahanda penulis H. Andi Ahmad Risal, S. SIT. dan ibunda tercinta Hj. Nuraeni Nur, AMK. yang telah memberikan dukungan materil, doa dan motivasi yang kuat kepada penulis, sehingga penulis dapat menyelesaikan penelitian ini.
7. Rekan-rekan Lab *Computer Based System* Departemen Teknik Informatika yang selalu saling mendukung dalam suka maupun duka dalam proses penyelesaian tesis ini.
8. Rekan-rekan Mahasiswa S2 Departemen Teknik Elektro angkatan 2019 yang selalu mendukung dalam proses penyelesaian tesis ini.

Penulis menyadari bahwa tesis masih jauh dari kata sempurna dan di dalam penyelesaiannya masih menemui kesulitan dan hambatan, sehingga penulis tetap mengharapkan saran dan kritik untuk pengembangan lebih lanjut, agar dapat memberikan manfaat yang banyak bagi semua pembaca.

Gowa, 20 Juni 2023

Penulis

## ABSTRAK

**ANDI ALVIADI NUR RISAL.** *Sistem Zonasi Sekolah untuk Penerimaan Siswa menggunakan Algoritma Constrained K-Means.* (dibimbing oleh **Zahir Zainuddin** dan **Muhammad Niswar**)

Terbitnya Peraturan Menteri Pendidikan dan Kebudayaan nomor 51 tahun 2018 mengatur penerimaan peserta didik baru dengan menerapkan sistem zonasi untuk mencapai pemerataan mutu pendidikan di setiap sekolah khususnya pada jenjang SMA di Kota Makassar. Penelitian ini bertujuan untuk mengelompokkan zonasi sekolah berdasarkan jarak terdekat antara domisili siswa dengan lokasi sekolah. Dataset yang digunakan adalah 22 lokasi sekolah dan 2248 data lokasi siswa. Dalam penelitian ini, metode yang digunakan adalah *Constrained K-Means* untuk mengelompokkan calon siswa baru ke masing-masing sekolah. Metode *Constrained K-Means* bekerja berdasarkan nilai  $K$  sebagai pusat *cluster* yang paling dekat dengan nilai  $N$  (anggota *cluster*) dengan pendekatan *Linear Programming Algorithm* (LPA) sehingga setiap *cluster* memiliki anggota  $N$  yang seimbang. Hasil penelitian ini dapat mengatasi persebaran data yang tidak merata, dengan rata-rata anggota *cluster* sebanyak 103 dan tidak ada *cluster* kosong di setiap sekolah atau pusat *cluster*. Tingkat akurasi yang diperoleh adalah 98,35%. Teknik validasi sistem membandingkan kinerja algoritma *Constrained K-Means* dengan algoritma *K-Means* tradisional. Performansi algoritma *K-Means* tradisional menunjukkan bahwa sebaran siswa tidak merata menuju pusat *cluster*, terutama berdasarkan parameter jarak domisili siswa ke sekolah dan adanya tumpang tindih. Akurasi yang dicapai antara anggota *cluster* dengan pusat *cluster* hanya 73,93%. Dengan demikian, algoritma *Constrained K-Means* dapat diimplementasikan pada penerimaan siswa baru untuk menentukan zonasi sekolah yang optimal dan akurat berdasarkan pusat *cluster*.

Kata Kunci: Penerimaan Siswa, Sistem Zonasi Sekolah, *Clustering*, *Constrained K-Means*

## ABSTRACT

**ANDI ALVIADI NUR RISAL.** *School Zoning System for Student Admission using Constrained K-Means Algorithms.* (supervised by **Zahir Zainuddin** and **Muhammad Niswar**)

The issuance of Minister of Education and Culture Regulation number 51 of 2018 regulates the acceptance of new students by implementing a zoning system to achieve equal distribution of quality education in every school, especially at the high school level in Makassar City. This study aims to classify school zoning based on the shortest distance between the student's domicile and the school's location. The dataset used is 22 school locations and 2248 student location data. In this study, the method used is Constrained K-Means to group prospective new students into each school. The Constrained K-Means method works based on the value of K as the center of the cluster that is closest to the value of N (cluster members) with the Linear Programming Algorithm (LPA) approach so that each cluster has balanced N members. The results of this study can overcome the uneven distribution of data, with an average of 103 cluster members and no empty clusters in each school or cluster center. The accuracy level obtained is 98.35%. The system validation technique compares the performance of the Constrained K-Means algorithm to that of the traditional K-Means algorithm. The performance of the traditional K-Means algorithm shows that the distribution of students is uneven towards the center of the cluster, mainly based on the parameters of the student's domicile distance to school and the presence of overlap. The accuracy achieved between cluster members and the cluster center is only 73.93%. Thus, the Constrained K-Means algorithm can be implemented on new student admissions to determine optimal and accurate school zoning based on cluster centers.

Keywords: Student Admission, School Zoning System, Clustering, Constrained K-Means

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>PENGAJUAN TESIS.....</b>	<b>ii</b>
<b>PERSETUJUAN TESIS .....</b>	<b>iii</b>
<b>PERNYATAAN KEASLIAN TESIS .....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	12
1.3 Tujuan Penelitian .....	12
1.4 Manfaat Penelitian .....	13
1.5 Batasan Masalah .....	13
1.6 Sistematika Penulisan .....	14
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>16</b>
2.1 Landasan Teori.....	16
2.1.1 Sistem zonasi penerimaan peserta didik baru.....	16
2.1.2 <i>Data mining</i> .....	17
2.1.3 <i>Clustering</i> .....	19
2.1.4 Algoritma <i>constrained k-means clustering</i> .....	20
2.1.5 <i>Imbalanced data</i> .....	24

2.1.6	Sistem informasi geografis .....	25
2.2	Penelitian Terkait .....	27
2.3	<i>State of The Art</i> .....	32
2.4	Kerangka Pikir .....	38
<b>BAB III METODE PENELITIAN .....</b>		<b>41</b>
3.1	Tahapan Penelitian .....	41
3.2	Waktu Dan Lokasi Penelitian .....	43
3.3	Jenis Penelitian.....	43
3.4	Rancangan Sistem.....	43
3.4.1	<i>Input data</i> .....	44
3.4.2	<i>Pre - processing</i> .....	46
3.4.3	<i>Calculate distance using euclidean distance</i> .....	47
3.4.4	<i>Clustering using constrained k-means</i> .....	48
3.5	Sumber Data.....	50
3.6	Instrumen Penelitian .....	50
3.7	Teknik Evaluasi Kinerja Sistem.....	51
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>52</b>
4.1	Akuisisi Data.....	52
4.2	Hasil <i>Clustering</i> Sistem Penentuan Zonasi Sekolah.....	53
4.3	Analisi Kinerja Sistem .....	55
4.4	Validasi Sistem .....	58
<b>BAB V PENUTUP.....</b>		<b>63</b>
5.1	Kesimpulan .....	63
5.2	Saran .....	64
<b>DAFTAR PUSTAKA .....</b>		<b>65</b>
<b>LAMPIRAN.....</b>		<b>70</b>

**DAFTAR TABEL**

Nomor	Halaman
<b>Tabel 1.</b> <i>State of the art</i> penelitian.....	32
<b>Tabel 2.</b> Rincian data siswa .....	45
<b>Tabel 3.</b> Rincian data sekolah.....	45
<b>Tabel 4.</b> Rincian hasil <i>clustering</i> .....	54
<b>Tabel 5.</b> Perbandingan kinerja hasil <i>clustering</i> pada algoritma <i>k-means</i> tradisional dengan <i>constrained k-means</i> .....	56
<b>Tabel 6.</b> Hasil <i>clustering members</i> menggunakan algoritma <i>constrained k-means</i> .....	58
<b>Tabel 7.</b> Hasil <i>clustering members</i> menggunakan algoritma <i>k-means</i> .....	60

## DAFTAR GAMBAR

Nomor	Halaman
<b>Gambar 1.</b> Ilustrasi sistem zonasi sekolah.....	16
<b>Gambar 2.</b> Langkah-langkah menyusun <i>knowledge discovery in databases</i> (KDD) .....	18
<b>Gambar 3.</b> Proses <i>knowledge discovery in databases</i> (KDD).....	19
<b>Gambar 4.</b> Karakteristik <i>cluster</i> .....	20
<b>Gambar 5.</b> Keterbatasan algoritma <i>k-means</i> tradisional.....	21
<b>Gambar 6.</b> <i>Model minimum cost flow</i> (MCF).....	22
<b>Gambar 7.</b> Ilustrasi proses <i>constrained clustering</i> .....	23
<b>Gambar 8.</b> Ilustrasi sifat kompleksitas <i>imbalanced data: overlap</i> (I), <i>small disjunct</i> (II), <i>outlier</i> (III) .....	24
<b>Gambar 9.</b> Ilustrasi subsistem SIG .....	26
<b>Gambar 10.</b> Kerangka pikir penelitian .....	40
<b>Gambar 11.</b> Tahapan penelitian.....	41
<b>Gambar 12.</b> Alur desain sistem .....	44
<b>Gambar 13.</b> Letak situs sekolah.....	46
<b>Gambar 14.</b> Map kota Makassar (area penelitian).....	53
<b>Gambar 15.</b> <i>Mapbox</i> hasil <i>clustering</i> menggunakan algoritma <i>constrained k-means</i> .....	53
<b>Gambar 16.</b> <i>Plotting</i> dari hasil <i>clustering</i> menggunakan algoritma <i>constrained</i> <i>k-means</i> .....	54

**DAFTAR LAMPIRAN**

Nomor	Halaman
<b>Lampiran 1.</b> Data distribusi siswa pada setiap pusat <i>cluster</i> .....	70
<b>Lampiran 2.</b> Publikasi artikel ilmiah .....	198

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan dunia pendidikan semakin pesat tidak lepas dari peran masyarakat. Pendidikan merupakan hal yang sangat penting serta tidak bisa dipisahkan dari kehidupan. Indonesia merupakan negara berkembang dalam berbagai aspek kehidupan, termasuk pada bidang pendidikan yang masih mengalami perkembangan, baik dalam kebijakan maupun sistem yang berlaku, berdasarkan undang-undang dasar 1945 pendidikan adalah hak setiap bangsa. Dengan demikian pemerintah memiliki kewajiban untuk memenuhi hak warga negaranya atas pendidikan untuk mendapatkan taraf kualitas hidup bangsa di masa depan. Pendidikan merupakan dasar landasan yang sangat berarti untuk mencapai kemajuan bangsa serta diperlukan sebagai langkah preventif untuk mengikuti perkembangan zaman dalam setiap prosesnya.

Pendidikan di negeri ini telah menjadi salah satu perhatian khusus di setiap elemen masyarakat untuk melihat dan mengetahui perkembangannya mengingat bagaimana pemerintah memberikan akses pemerataan yang sama bagi seluruh warga Indonesia.

Pendaftaran penerimaan siswa baru merupakan program kegiatan yang dilaksanakan setiap tahun pada institusi pendidikan sekolah untuk menjaring calon siswa baru.

Terkait penerimaan siswa baru, pemerintah melalui Kementerian Pendidikan dan Kebudayaan telah menerbitkan aturan kebijakan, yaitu Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 tentang Penerimaan Peserta Didik Baru (PPDB).

Pemerintah menerbitkan Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 pada prinsip dasarnya pemerintah bertujuan untuk merevitalisasi pelaksanaan Penerimaan Peserta Didik Baru (PPDB) di setiap instansi lembaga pendidikan khususnya pada sekolah berlangsung secara lebih objektif, akuntabel, transparan, berkeadilan, serta non-diskriminatif sehingga dapat meningkatkan kualitas pelayanan serta mutu dalam bidang pendidikan.

Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) Nomor 51 Tahun 2018 merupakan kebijakan pemerintah yang mengatur sistematisa penerimaan peserta didik baru pada semua jenjang sekolah mulai dari taman kanak-kanak, sekolah dasar, sekolah menengah pertama, maupun sekolah menengah atas dan sekolah menengah kejuruan.

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.

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.

Negara-negara maju dalam bidang pendidikan seperti Eropa, Amerika Serikat dan Finlandia telah lama menerapkan sistem zonasi dengan tujuan untuk memutuskan tingkat angka kemiskinan. Peningkatan pendidikan merupakan upaya yang efisien untuk meningkatkan kualitas sumber daya manusia. Dengan memperhatikan bahwa negara-negara yang memiliki taraf kualitas pendidikan yang baik serta maju dalam bidang pendidikan berhasil meminimalkan permasalahan pendidikannya. Pada tahun ajaran baru 2018 / 2019 Indonesia menerapkan Penerimaan Peserta Didik Baru (PPDB) dengan sistem zonasi.

Implementasi dari Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 menerapkan sistem zonasi dalam Penerimaan Peserta Didik Baru (PPDB) pada jalur zonasi berdasarkan radius jarak

terdekat antara rumah dan sekolah untuk mendekatkan lingkungan sekolah dengan lingkungan keluarga.

Aturan sistem zonasi dalam Penerimaan Peserta Didik Baru (PPDB) menetapkan bahwa satuan pendidikan sekolah yang diselenggarakan oleh pemerintah daerah wajib menerima pendaftaran siswa dengan jumlah ketentuan paling sedikit 90 % dari jumlah domisili siswa di sekitar wilayah terdekat dengan lokasi sekolah.

Penerapan dari sistem zonasi ini dapat mengurangi disparitas dan persaingan antara lembaga satuan pendidikan sekolah negeri maupun sekolah swasta. Dengan demikian menjamin pemerataan akses pendidikan bagi semua kalangan masyarakat.

Kebijakan sistem zonasi yang tertuang dalam Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 merupakan inisiatif pemerintah untuk mendistribusikan kualitas pendidikan di berbagai daerah di Indonesia.

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. 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.

Manfaat dari kebijakan pemerintah mengenai sistem zonasi sekolah adalah dapat memonitoring kebutuhan dan penempatan guru. Guru yang berkualitas dan berprestasi sebagai pendidik tidak lagi dikelompokkan dalam satu sekolah akan tetapi didistribusikan ke semua sekolah, situasi siswa yang beragam menuntut guru untuk terlibat dalam kegiatan belajar mengajar yang lebih kreatif.

Guru yang berpotensi dapat dipindahkan ke sekolah daerah untuk mempromosikan pemerataan kualitas pendidikan.

Peraturan pemerintah tentang pemerataan kualitas taraf pendidikan di Indonesia untuk mempermudah akses pendidikan serta menghilangkan eksklusivitas pada satuan lembaga pendidikan sekolah dan meningkatkan kualitas pelayanan dan mutu sekolah sebagaimana yang termuat pada permendikbud nomor 51 tahun 2018 dalam implementasinya memicu permasalahan sebagaimana dikemukakan oleh Federasi Guru Indonesia (FSGI) yaitu persoalan kapasitas daya tampung sekolah dengan jumlah siswa yang tidak seimbang. Dampaknya, ditemukan banyak calon siswa yang pendaftarannya tidak diakomodir oleh pihak sekolah. Sementara itu, jarak tempuh rumah calon siswa dengan sekolah tujuan tidak saling berjauhan.

Persoalan lain yang ditimbulkan dari sistem zonasi ini terdapat pada modifikasi domisili yang dilakukan masyarakat secara tiba-tiba. Seorang calon peserta didik baru mencantumkan namanya pada identitas kartu keluarga kerabatnya agar dapat diakomodir mendaftar di sekolah daerah perpindahan tersebut, artinya dari kebijakan pemerintah yang tertuang dari permendikbud nomor 51 tahun 2018 terkait sistem penerimaan siswa dengan jalur zonasi ini bisa dicurangi.

Berdasarkan penelitian yang dilakukan Wahyuni (2018), penerapan sistem zonasi sekolah banyak ditemukan permasalahan, diantaranya adalah prioritas jarak tempat tinggal calon siswa dengan sekolah sebagai penentu utama pada proses Penerimaan Peserta Didik Baru (PPDB). Dalam hal ini, menjadi permasalahan karena sistem zonasi belum dapat mengakomodasi pendaftaran calon peserta didik baru terutama calon peserta didik yang berdomisili di area *blind spot* (tidak terjangkau pada radius zona sekolah).

Kompleksitas yang terjadi di masyarakat adalah perdebatan dalam penafsiran jarak dari rumah ke sekolah, hal ini didasarkan pada tempat tinggal calon siswa yang mengikuti identitas kependudukan yang berupa kartu keluarga orang tua ke sekolah terdekat.

Masalah lainnya adalah persebaran satuan pendidikan sekolah yang tidak merata, terbukti banyaknya sekolah yang letak geografisnya saling berapit satu sama lain atau berada dalam satu radius yang mengakibatkan area tersebut akan tumpang tindih dalam hal kapasitas.

Oleh karena itu, sangat penting untuk mencari area yang optimal dan efisien agar daya tampung sekolah merata sesuai dengan jumlah pendaftar calon peserta didik.

Permasalahan lain yang ada saat ini dimasyarakat mengenai teknis dalam menentukan zonasi calon peserta didik baru dengan sekolah tujuan hanya beracuan pada perhitungan *platform Google Maps* dengan fitur tarikan garis lurus antara jarak domisili calon peserta didik ke titik sekolah untuk menentukan radius dan jarak terdekat (*measure distance*) sebagai syarat administrasi pada saat melakukan pendaftar, hal ini dinilai tidak akurat dalam proses pelaksanaan penerimaan calon peserta didik baru (PPDB) karena dapat dicurangi, sebab dilakukan dengan cara konvensional sehingga dalam menentukan jumlah kapasitas daya tampung siswa pada setiap sekolah mengalami ketimpangan yang menghasilkan *output* data yang tumpang tindih (*overlap*).

Untuk menyelesaikan permasalahan diatas adalah dengan pemanfaatan metode *clustering*. *Clustering* adalah suatu aktivitas proses pengelompokan data yang bertugas menunjukkan kemiripan antara satu data dengan data lainnya ke dalam *cluster* atau kelompok sehingga data dalam suatu *cluster* tersebut memiliki tingkat kemiripan karakteristik (*similarity*) yang paling tinggi sedangkan dengan data lain di antara *cluster* memiliki kemiripan paling rendah.

*Clustering* juga dapat diartikan metode segmentasi data yang diimplementasikan dalam beberapa bidang, diantaranya *marketing*, analisa masalah prediksi segmentasi pasar dan bisnis, pola dalam bidang *computer vision*, zonasi wilayah hingga identifikasi objek dan pengolahan citra, serta bioinformatika.

Tujuan dari analisis *cluster* adalah untuk menemukan kelompok objek dimana objek-objek dalam kelompok tersebut sama (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek-objek dalam kelompok lain.

*Constrained clustering* adalah salah satu metode *clustering* yang berperan penting untuk proses *machine learning*, pemrosesan sinyal, dan *data mining* untuk menemukan pola data di dalam suatu *cluster* ketika ada tambahan informasi domain.

*K-Means* merupakan algoritma partisi terpopuler untuk *clustering* data karena diasumsikan efisien. Hal ini didasarkan pada konsep *centroid* yang digunakan untuk mendefinisikan *cluster* dalam berkinerja. Namun dalam beberapa kasus ditemukan pokok permasalahan pada batasan ukuran *cluster* yang mengharuskan setiap *cluster* berada dalam kisaran sama, dalam kasus tersebut *K-Means* sering digabungkan dengan satu atau lebih *cluster* yang bernilai kosong dengan demikian *K-Means* merangkum sejumlah data yang berukuran, atau dalam artian lain *K-Means* menghasilkan konvergen yang berlawanan untuk minimum data objek di setiap *cluster*, terutama pada permasalahan keacakan inisialisasinya serta pertimbangan kontrol yang tidak efisien atas ukuran *cluster* satu dengan *cluster* lainnya.

Mengingat atas kelemahan dari penggunaan algoritma *K-Means clustering*, penulis memfokuskan *K-Means* dapat digabungkan dengan *constraint* pada studi kasus Penerimaan Peserta Didik Baru (PPDB) pada jalur zonasi sehingga penyebaran calon peserta didik merata terutama pada wilayah dengan kondisi sekolah saling berapit satu sama lain di dalam satu radius jangkauan.

Penelitian yang dilakukan oleh (Bradley *et al.*, 2000) mengembangkan algoritma *K-Means clustering* dengan melakukan usulan pendekatan terhadap metode *Linier Pemrogramming Algorithm* (LPA), mengingat dimana nilai “*k*” dalam algoritma tradisional *K-Means* sesuai dengan dengan jumlah pusat (*means*) *cluster* di sekitar *cluster* yang akan terbentuk dengan demikian tidak ada *cluster* yang terbentuk dari jumlah nilai “*k*”. Dengan metode yang diusulkan oleh (Bradley *et al.*, 2000) mengharuskan setiap *cluster* memiliki setidaknya 2 atau lebih subjek di setiap *cluster* yang terbentuk serta menghasilkan optimasi *cluster* yang lebih baik terhadap data yang tidak seimbang.

Pada penelitian lainnya yang pernah dilakukan oleh (Wagstaff *et al.*, 2001) untuk pertama kali mengaplikasikan algoritma *COP - K-Means*, mendemonstrasikan algoritma *K-Means clustering* dimodifikasi dengan

menerapkan *Pairwise Constraints* sebagai latar belakang ilmu pengetahuan untuk membatasi penugasan algoritma *K-Means*, mereka melakukan pengujian metode ini untuk mendeteksi jalur jalanan dari data *Global Positioning System* (GPS). Hasil pengujian penelitian ini menyatakan bahwa kinerja algoritma *COP - K-Means* secara umum lebih baik daripada algoritma konvensional *K-Means*. Namun algoritma *COP - K-Means* secara implisit menentukan kondisi yang harus dipenuhi berdasarkan urutan alokasi data dari proses *K-Means clustering*, yang menghasilkan hasil pengelompokan yang tidak stabil.

Okabe dan Yamada (2018) pada penelitiannya mengusulkan metode *Constrained Clustering* dengan mengkombinasikan algoritma baru *Constrained K-Means* dengan prinsip *boosting*. Dalam menjalankan tugasnya sebagai metode *Constrained Clustering* memiliki keunggulan dalam hal waktu komputasi yang lebih sedikit dibandingkan dengan algoritma tradisional *K-Means*.

Hasil eksperimen pada 12 kumpulan data dari tiga sumber data menunjukkan bahwa metode ini memiliki kinerja yang kompetitif dengan metode *Constrained Clustering* yang canggih untuk dataset yang besar dan membutuhkan waktu komputasi yang jauh lebih sedikit.

Terlepas dari keunggulan komputasi algoritma ini, namun secara umum dalam berkinerja algoritma ini membutuhkan beberapa *constraints* untuk mengoptimalkan kekuatan *boosting* karena metode ini melakukan pendekatan *Ensemble Learning*.

Penelitian lainnya yang dilakukan (Lei *et al.*, 2013) menyatakan bahwa hasil dari proses *clustering* sangat bergantung pada *initial center (centroid)* pada algoritma *K-Means* dan *Size - Constrained Clustering*. Oleh karena itu, mempertimbangkan dua konsep untuk mengoptimalkan *Size - Constrained Clustering*, yang pertama meningkatkan kecepatan konvergensi selanjutnya mengurangi RMSE.

Penelitian ini mengusulkan metode *Size - Constrained Clustering* sebagai pemilihan titik awal dengan menemukan titik padat di sekitar *centroid* yang telah terbentuk secara rekursif. *Root - Mean - Square Error* dan kecepatan konvergensi (jumlah iterasi) adalah dua metrik terpenting untuk pengelompokan menggunakan metode iteratif.

Pengujian dari penelitian ini dilakukan pada sekitar sepuluh ribu proposal penelitian *National Natural Science Foundation of China* dan hasilnya menunjukkan bahwa metode ini dapat mengurangi waktu iterasi hingga lebih dari 50% dan mendapatkan *root mean square error* yang lebih kecil, dengan demikian metode ini dapat dikombinasikan dengan skala *Size - Constrained Clustering* untuk mengatasi masalah *clustering* data dengan skala besar dalam *data mining*.

Selanjutnya terdapat dua bentuk pengembangan penelitian ini yang dapat dilakukan di masa akan datang. Pertama mengurangi kompleksitas waktu dari algoritma pemilihan titik *centroid* awal dengan menghilangkan perhitungan yang berlebihan, Kemudian, bereksperimen pada kumpulan data yang lebih kompleks untuk pengaplikasian permasalahan *real*.

Thota *et al.*, (2017) melakukan penelitian yang berjudul “*Cluster based Zoning of Crime Info*” dimana penelitian tersebut mengenai zonasi wilayah kejahatan yang terjadi di negara India, dimodelkan dengan pemanfaatan *data mining*, tindakan yang dilakukan antara lain mengumpulkan, menganalisis dan mempublikasikan data kejahatan menggunakan analisis *cluster*.

Penelitian ini mengusulkan algoritma *K-Means* yang berfungsi sebagai pembentuk pola wilayah yang rawan tindak kejahatan, dataset yang digunakan adalah kumpulan dari laporan info kejahatan lembaga *The National Crime Records Bureau* (NCRB).

Input dari data hasil *clustering* divisualisasikan dengan menggunakan peta khusus zona *cluster* yang berguna untuk menampilkan info kejahatan pada negara bagian, sehingga dapat membantu polisi dan penegak hukum dalam mengambil tindakan pencegahan untuk memerangi kejahatan dan merencanakan operasi yang strategi dalam melakukan tindakan investigasi.

Kelanjutan dari penelitian ini dapat dikembangkan dengan berbagai metode *data mining* pada kumpulan data *The National Crime Records Bureau* (NCRB) India dan menggunakan atribut kejahatan yang lebih banyak untuk mengidentifikasi *trending* kejahatan.

Penelitian yang dilakukan oleh (Alifi, Hayati, dan Supangkat., 2017) Berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia tentang standar pelayanan minimum untuk pendidikan dasar, pemerintah kota

Bandung memberlakukan sistem zonasi sekolah, penelitian ini mengusulkan rencana jalur sekolah untuk meminimalkan biaya transportasi sekolah, serta mempertimbangkan pembangunan dan perkiraan kapasitas daya tampung sekolah negeri.

Metode yang digunakan adalah *P - Median* sebagai model untuk jalur akses sekolah, pelaksanaan penelitian ini menggunakan perangkat lunak *ArcGIS* untuk menganalisis pemetaan akses jalur ke sekolah serta antara sekolah satu dengan yang lainnya. Dari hasil analisis pemetaan akses jalur yang diusulkan menurunkan impedansi sebesar 26,55% dibandingkan dengan akses jalur sebelumnya. Adapun hasil dari penelitian ini merekomendasi untuk mengurangi pendaftaran sekolah di 11 kecamatan, dan meningkatkan pendaftaran sekolah di dua kecamatan, serta membangun sekolah baru di 13 kecamatan.

Sitanggang, Radiatun, dan Nur Risal (2020) dengan judul penelitian “*Incremental Spatio Temporal Clustering Application on Hotspot Data*”.

Titik panas merupakan salah satu indikator kebakaran hutan dan lahan. Analisis data titik api perlu dilakukan sebagai langkah peringatan dini untuk mencegah terjadinya kebakaran hutan dan lahan. Penelitian ini bertujuan untuk membuat aplikasi *clustering* data *hotspot* berbasis *web* menggunakan modul *Incremental ST - DBSCAN*. Dataset yang digunakan pada penelitian ini adalah kumpulan data *hotspot* dari tahun 2014 sampai dengan tahun 2017.

*Software* yang dikembangkan menggunakan bahasa pemrograman *R* dan *framework Shiny*. Metodologi yang diterapkan adalah *Adaptive Software Development (ASD)*.

Fitur utama dari aplikasi ini adalah *incremental clustering* data *hotspot* serta visualisasi hasil *clustering*. Pengujian dilakukan dengan pendekatan *Black Box*, menunjukkan bahwa semua fitur *software* berfungsi dengan baik. Berdasarkan hasil evaluasi *usability*, kepuasan pengguna mencapai 78.5%, yang berarti *software* ini sangat mudah digunakan.

Selanjutnya penelitian yang dilakukan oleh Wang dan Gao (2019), menjelaskan algoritma *K-Means* banyak digunakan dalam proses *clustering* dimana bertugas untuk memproses data ke dalam kelas yang berbeda, karena kesederhanaan dan efisiensinya, serta telah menjadi alat penting dalam *data mining*.

Akan tetapi, algoritma *tradisional K-Means* hanya dapat mendeteksi *cluster* dengan pola lingkaran (bola) dan rentan terhadap *noise* atau titik terisolasi, yang dapat mempengaruhi hasil *clustering*.

Untuk mengatasi masalah tersebut, pada penelitian ini mengusulkan algoritma *K-Means* yang ditingkatkan berdasarkan uji *Kurtosis* dan metode *Monte Carlo*.

Algoritma *K-Means* ditingkatkan agar dapat beradaptasi dengan kumpulan data berbentuk kompleks, dan mengurangi dampak *outlier* data pada hasil *clustering*, serta membuat hasil algoritma *K-Means* lebih akurat.

Hasil penelitian menunjukkan bahwa algoritma yang diusulkan memiliki indeks eksternal yang lebih besar dari kinerja *metrics clustering*. Hal ini berarti akurasi hasil *clustering* meningkat secara signifikan.

Selanjutnya untuk pengembangan penelitian ini agar lebih memperhatikan *thresholds* yang ditetapkan dalam proses komputasi algoritma. Selain itu, untuk arah penelitian prospektif menjabarkan penggunaan nilai langsung *Kurtosis* sebagai indikator dinamis dalam proses iterasi algoritma.

Syaripudin *et al.*, (2019) mereka mendesain sistem rekayasa penentuan sekolah yang terdekat dengan menggunakan formula *Haversine* dengan tujuan menghitung jarak antara situs sekolah dan rumah siswa, dengan memanfaatkan layanan *location based service*. Untuk penunjuk posisi awal dengan menggunakan *Global Positioning System (GPS)*, hasil dari penelitian ini sangat baik dibandingkan hasil prediksi *default Google Distance* dari Google.

Akan tetapi, dalam memprediksi perhitungan jarak lokasi sekolah terdekat berdasarkan sistem zonasi tingkat keakuratannya tidak selalu optimal sebab sistem rekayasa ini mengacu pada tingkat akurasi sistem GPS di perangkat yang digunakan.

Selanjutnya penelitian yang dilakukan (Ratnasar *et al.*, 2020) yang terkait mengenai penerimaan siswa berbasis zonasi yaitu dengan menggunakan metode *Generate and Test* untuk menganjurkan rekomendasi jarak terpendek terbaik dan akurat, adapun dalam pelaksanaannya tanpa melibatkan aplikasi berbasis denah lainnya.

Hasil dari penelitian ini diuji dengan kalkulasi formula *Haversine*, dan dalam memprediksi pengukuran jarak menggunakan *google maps*. Namun, dalam berkinerja membutuhkan waktu yang cukup lama dan penggunaan memori yang besar yaitu 6,2 MB dan waktu rata-rata 0,0000216806982 Ms.

Penelitian lainnya yang mengacu untuk mengidentifikasi sistem penentuan zonasi penerimaan siswa baru yaitu dengan menggunakan algoritma *K-Means*.

Hasil yang diperoleh pada penelitian ini menyatakan bahwa dengan menggunakan algoritma *K-Means* berdasarkan distribusi data siswa relatif baik terhadap acuan parameter jarak antara domisili siswa dan sekolah, serta terbentuknya penetapan area penerimaan siswa berbasis zonasi secara non lingkaran.

Akan tetapi, hasil dari penelitian ini memperlihatkan bahwa setiap sekolah memiliki anggota *cluster* yang tidak seimbang atas sebaran siswa antara *cluster* satu dengan *cluster* lainnya, dalam hal ini sekolah sebagai titik *centroid cluster*.

Algoritma *K-Nearest Neighbour* dan *K-Means* pada prinsipnya menggunakan metode pembelajaran *Supervised* dan *Unsupervised*. Namun, dalam melaksanakan tugasnya terhadap dataset yang besar tidak berkinerja dengan optimal.

Selain itu, mereka berasumsi bahwa menganggap distribusi kelas data telah seimbang padahal pada kenyataannya banyak himpunan data besar yang sangat tidak seimbang. Yaitu mayoritas data berada pada satu kelas sedangkan pada data minoritas berada pada kelas lain.

Berdasarkan peninjauan penelitian sebelumnya mengenai sistem penentuan zonasi sekolah pada penerimaan siswa, sebagian besar hanya berfokus pada dimensi jarak antara domisili atau tempat tinggal siswa dengan jarak tempuh sekolah tanpa memperhatikan daya tampung antara sekolah satu dengan sekolah lainnya sehingga masih menyebabkan permasalahan pada akseptasi siswa melalui jalur zonasi.

Dalam penelitian ini ditawarkan desain metode yang terintegrasi untuk sistem penentuan wilayah area zonasi penerimaan siswa sekolah terhadap data yang tidak seimbang. Dengan memanfaatkan metode pendekatan *clustering* menggunakan algoritma *Constrained K-Means* dalam teknik pemerataan data.

Tujuan pendistribusian alokasi data harus optimal dan tidak menyebabkan tumpang tindih pada kondisi data tertentu, seperti halnya keadaan lokasi sekolah yang saling berapit satu sama lain sedemikian rupa. Hingga sekolah yang jaraknya berapit tidak dikelompokkan sebagai satu *cluster* yang sama demikian juga dengan anggota *cluster*.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana merancang penentuan sistem zonasi sekolah yang optimal dan efisien dengan menggunakan algoritma *Constrained K-Means*?
2. Bagaimana membuat *clustering* zona sekolah yang tidak tumpang tindih untuk sekolah yang berada dalam satu radius (kondisi letak sekolah saling berdekatan satu sama lain)?
3. Bagaimana menyeimbangkan daya tampung sebaran siswa di setiap zona sekolah (zona non lingkaran seperti saat ini)?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah maka tujuan yang akan dicapai dari penelitian ini adalah:

1. Merancang penentuan sistem zonasi sekolah yang optimal dan efisien dengan menggunakan algoritma *Constrained K-Means*.
2. Membuat *clustering* zona tiap sekolah berdasarkan daya tampung yang tidak tumpang tindih pada sekolah yang berada dalam satu radius (kondisi letak sekolah saling berdekatan satu sama lain).
3. Menyeimbangkan daya tampung sebaran siswa di setiap zona sekolah (zona non lingkaran seperti saat ini).

#### 1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Bagi peneliti, pengaplikasian dari sejumlah konsep dan pengetahuan yang berguna untuk menambah wawasan pengetahuan ataupun kemampuan/*skill* mengenai proses pengenalan pola metode *data mining* untuk *clustering* zona sekolah.
2. Bagi masyarakat, dapat digunakan sebagai referensi bagi pembaca dalam menambah wawasan dan ilmu pengetahuan dalam hal teknologi industri 4.0, khususnya dalam bidang *data mining*.
3. Bagi Pemerintah, dapat digunakan sebagai referensi dalam mengembangkan sistem pendidikan di Indonesia terkhusus pada instansi lembaga bidang pendidikan untuk proses Penerimaan Peserta Didik Baru, di era revolusi industri 4.0 sebagai dasar untuk membangun sarana pendidikan dalam acuan menetapkan area zonasi sekolah yang optimal dan akurat berdasarkan pusat *cluster* (sekolah terdekat).
4. Bagi institusi Pendidikan Magister Jurusan Teknik Elektro Konsentrasi Informatika, dapat digunakan sebagai referensi ilmiah dalam penelitian untuk pengembangan sistem pada topik *data mining* terkhusus pada metode *clustering* dengan objek atau kasus yang berbeda.

#### 1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini antara lain:

1. Data yang digunakan yaitu data dari alamat calon peserta didik yang akan melanjutkan pendidikan sekolah dari jenjang sekolah menengah pertama ke jenjang sekolah negeri menengah atas.
2. Parameter atribut pendukung yang diambil dari data siswa dan sekolah adalah jarak distribusi, jumlah calon peserta didik.
3. Kategori yang digunakan dalam penentuan sekolah pada setiap calon peserta didik adalah jarak terdekat antara domisili calon peserta didik dengan jarak sekolah.

4. Sistem yang dibuat hanya mengklaster data sebaran domisili calon peserta didik berdasarkan area penentuan zonasi sekolah yang terdekat, tidak untuk menentukan akses jalan berdasarkan medan topografi.

## 1.6 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini adalah:

### **Bab I Pendahuluan**

Pada bab I ini berisi penjelasan tentang latar belakang yang menjabarkan alasan dilakukannya penelitian terkait sistem pendidikan Indonesia terkhusus pada skema Penerimaan Peserta Didik Baru (PPDB) jalur zonasi sekolah, berdasarkan peluang penelitian dan uraian penelitian awal tentang *data mining* dengan metode *clustering* yang dilakukan, terkait rumusan masalah, tujuan, manfaat, ruang lingkup serta sistematika penulisan penelitian dibahas pada bagian ini.

### **Bab II Tinjauan Pustaka**

Pada bab II ini berisi penjelasan tentang landasan teori yang digunakan dalam penelitian meliputi sistem zonasi sekolah, *data mining*, *clustering*, algoritma *Constrained K-Means*, dan beberapa landasan teori lainnya.

Diuraikan pula tentang tinjauan pustaka yang merupakan penjelasan tentang hasil-hasil penelitian sebelumnya yang berkaitan dengan penelitian yang dilakukan. Landasan teori merupakan suatu penjelasan tentang sumber acuan terbaru dari pustaka primer seperti buku, artikel, jurnal, prosiding dan tulisan asli lainnya untuk mengetahui perkembangan penelitian yang relevan dengan judul atau tema penelitian yang dilakukan dan juga sebagai arahan dalam memecahkan masalah yang diteliti.

Pada bab ini juga diuraikan tentang kerangka pemikiran yang merupakan penjelasan tentang kerangka berpikir untuk memecahkan masalah yang sedang diteliti, termasuk menguraikan objek penelitian serta *state of the art* dari beberapa penelitian terkait.

### **Bab III Metodologi Penelitian**

Pada Bab III ini merupakan penjelasan tentang tahapan penelitian dimulai bagaimana metode penelitian, penentuan masalah, penentuan pendekatan komputasi, juga penjelasan bagaimana pengembangan dan penerapan sistem objek penelitian.

Selain itu, menjelaskan proses validasi hasil penerapan sistem, metode pengumpulan data, metode analisis data, penerapan pada masalah penelitian, konstruksi sistem serta pengujian sistem.

### **Bab IV Hasil dan Pembahasan**

Pada bab IV berisi penjabaran hasil dan pembahasan penelitian berdasarkan teknik implikasi sistem yang digunakan serta penelitian yang telah dilakukan.

Pada bagian hasil ini membahas tentang data kuantitatif yang telah dikumpulkan berdasarkan ketetapan pada metodologi. Pembahasan menjabarkan proses pengolahan data serta interpretasinya, baik berupa deskriptif maupun penarikan inferensinya.

Implikasi penelitian merupakan uraian penjabaran mengenai keberlanjutan penelitian yang relevan dengan aspek sistem, maupun pengembangan pada aspek penelitian. Hasil analisis kinerja sistem *clustering* dan olah *data mining* yang telah dirancang dirangkum dalam bentuk tabel, dan gambar *map box*.

### **Bab V Kesimpulan dan Saran**

Pada bab V berisi kesimpulan terhadap hasil yang didapatkan dalam penelitian ini, yang merujuk pada rumusan masalah dan saran pengembangan dari penelitian ini untuk menyempurnakan kekurangan-kekurangan yang belum tercapai pada penelitian ini, sehingga kedepannya penelitian yang dilakukan dapat dikembangkan dan bisa memperoleh hasil yang jauh lebih baik.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Landasan Teori

##### 2.1.1 Sistem zonasi penerimaan peserta didik baru



**Gambar 1.** Ilustrasi sistem zonasi sekolah

Penerimaan Peserta Didik Baru (PPDB) mengacu pada Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 tentang Penerimaan Peserta Didik Baru (PPDB) pada semua jenjang sekolah mulai dari taman kanak-kanak, sekolah dasar, sekolah menengah pertama, maupun sekolah menengah atas dan sekolah menengah kejuruan. Beberapa hal yang perlu diperhatikan mengenai sistem zonasi dalam Penerimaan Peserta Didik Baru (PPDB) diantaranya:

- a. Sekolah yang diselenggarakan oleh pemerintah daerah (pemda) wajib menerima calon peserta didik yang berdomisili pada radius zona terdekat dari sekolah dengan kuota paling sedikit 90% dari total jumlah keseluruhan peserta didik yang diterima.
- b. Domisili calon peserta didik yang termasuk dalam zonasi sekolah didasarkan pada alamat pada kartu keluarga (KK) yang diterbitkan paling lambat 6 (enam) bulan sebelum pelaksanaan PPDB.

- c. Radius zona terdekat dalam sistem zonasi ditetapkan oleh pemda sesuai dengan kondisi di daerah tersebut dengan memperhatikan ketersediaan anak usia sekolah di daerah tersebut; dan jumlah ketersediaan daya tampung sekolah.
- d. Penetapan radius zona pada sistem zonasi ditentukan oleh pemda dengan melibatkan musyawarah atau kelompok kerja kepala sekolah.
- e. Bagi sekolah yang berada di daerah perbatasan provinsi/kabupaten/kota, ketentuan persentase penerimaan siswa dan radius zona terdekat dapat ditetapkan melalui kesepakatan tertulis antara pemerintah daerah yang saling berbatasan.
- f. Calon siswa di luar zonasi dapat diterima melalui beberapa cara yakni:
  - i. Melalui jalur prestasi dengan kuota paling banyak 5% (lima persen) dari total jumlah keseluruhan peserta didik yang diterima.
  - ii. Alasan perpindahan domisili orang tua/wali atau alasan terjadi bencana alam/sosial dengan paling banyak 5% (lima persen) dari total keseluruhan siswa yang diterima.
- g. Sistem zonasi menjadi prioritas utama atau terpenting dalam PPDB jenjang SMP dan SMA. Setelah seleksi zonasi baru kemudian dipertimbangkan hasil seleksi ujian tingkat SD atau hasil ujian nasional SMP untuk tingkat SMA.
- h. Untuk jenjang SD, sistem zonasi menjadi pertimbangan seleksi tahap kedua setelah faktor minimum usia masuk sekolah sudah terpenuhi. Sedangkan bagi SMK sama sekali tidak terikat mengikuti sistem zonasi.

### **2.1.2 Data mining**

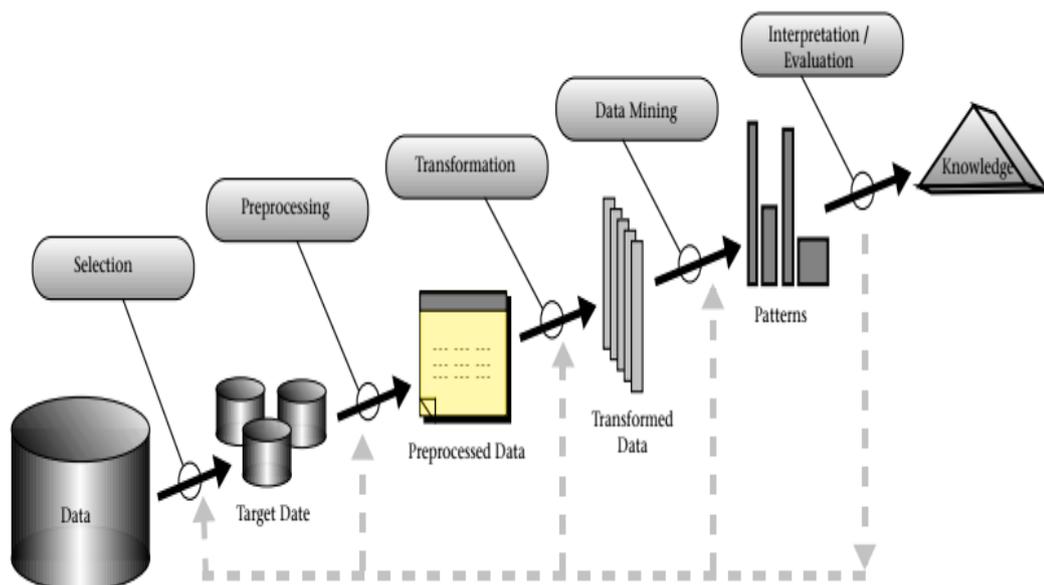
Menurut beberapa pakar ahli mendefinisikan *data mining* diantaranya sebagai berikut:

- a. Olson dan Delen (2008) mendefinisikan *data mining* sebagai analisis data eksplorasi. Data yang dihasilkan dari *cash register*, *scanning*, dan *database* pada perusahaan bisnis tertentu untuk dieksplorasi, dianalisis, dikurangi, dan digunakan kembali.

- b. Tan *et al.*, (2019) *data mining* adalah sebuah proses untuk menemukan informasi yang berguna dari repositori data yang besar secara otomatis. Teknik *data mining* digunakan untuk mencari serta menemukan pola baru pada kumpulan data besar yang berguna. Teknik *data mining* ini memiliki kemampuan untuk memprediksi hasil pengamatan di masa depan. Seperti jumlah yang akan dibayarkan dan dibelanjakan oleh pelanggan secara online atau di toko fisik.
- c. Wolfgang Ertel (2021) *Data mining* adalah suatu proses memperoleh pengetahuan dari data untuk menyajikannya, dan menerapkannya. Metode yang digunakan biasanya didasarkan pada statistik atau *machine learning* dan harus dapat diaplikasikan pada jumlah data yang besar (*big data*) dengan upaya biaya yang wajar.

Model *data mining* terdiri dari seperangkat persamaan, aturan, atau fungsi transfer yang kompleks yang dapat digunakan untuk mengidentifikasi pola yang bermanfaat.

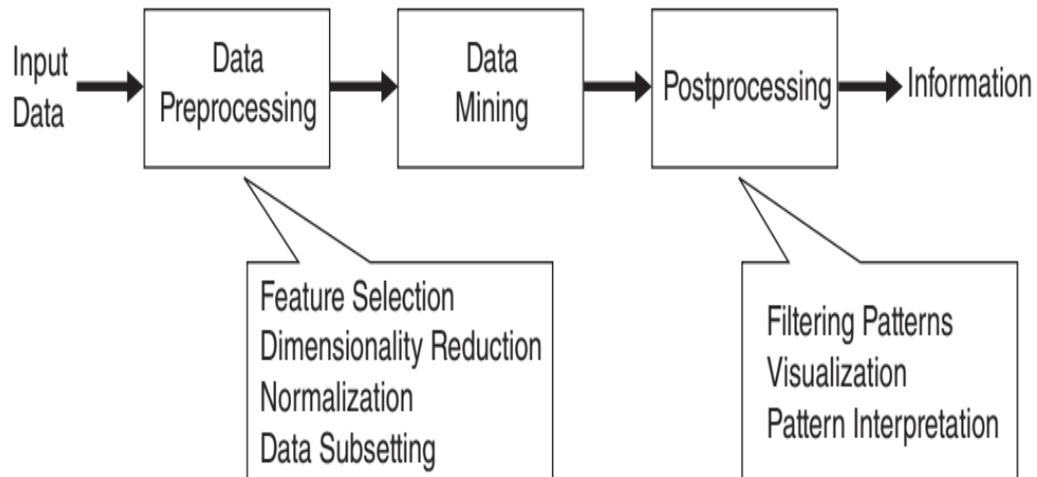
*Data mining* adalah proses pencarian informasi secara otomatis di penyimpanan data besar. Istilah yang umum digunakan yaitu *knowledge discovery in databases (KDD)*, *knowledge extraction*, *pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, dan *business intelligence*.



**Gambar 2.** Langkah-langkah menyusun *knowledge discovery in databases (KDD)*

*Data mining* merupakan bagian integral dari *Knowledge Discovery in Databases* (KDD), proses keseluruhan mengubah data mentah menjadi informasi yang berguna, seperti yang ditunjukkan pada Gambar 2.

Proses ini terdiri dari serangkaian langkah dari *pre-processing* data hingga *postprocessing* hasil *data mining*.

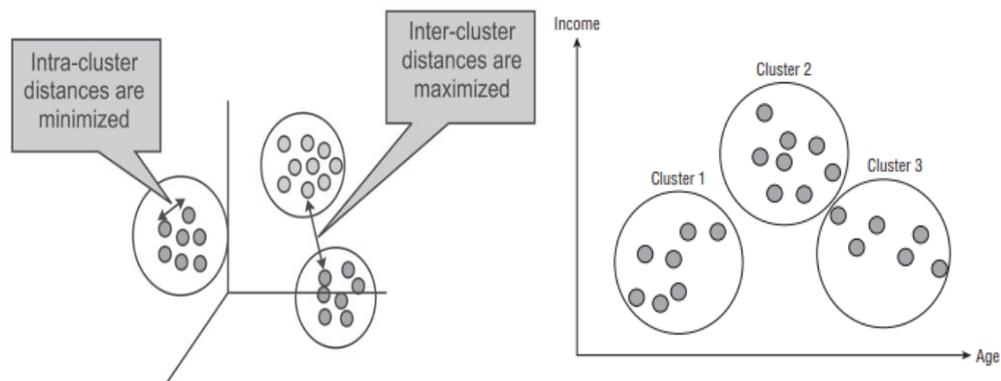


**Gambar 3.** Proses *knowledge discovery in databases* (KDD)

*Data mining* merupakan proses pencarian pengetahuan dari suatu data berukuran besar melalui metode statistik, *machine learning*, dan *artificial algorithm*. Hal yang paling utama dari suatu proses *data mining* adalah *feature selection* dan proses pengenalan pola dari suatu sistem *database*.

### 2.1.3 Clustering

*Clustering* dalam *data mining* didefinisikan sebagai proses segmentasi pengelompokan satu set data objek yang memiliki kesamaan ke dalam kelas atau *cluster*. Dengan kata lain, analisis *cluster* melakukan partisi terhadap kumpulan data ke dalam kelas-kelas sedemikian rupa sehingga kumpulan data tersebut memperlihatkan kesamaan yang tinggi satu sama lain di dalam suatu *cluster* (*intra-cluster*), akan tetapi memiliki ketidaksamaan yang tinggi dibandingkan dengan objek-objek dalam *cluster* lain (*antar-cluster*). Seperti yang ditunjukkan pada Gambar 4.



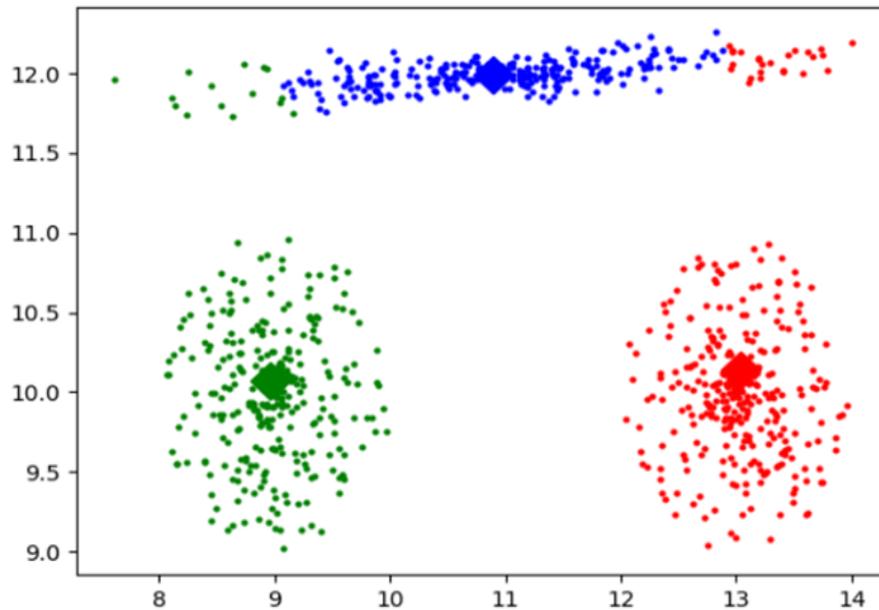
**Gambar 4.** Karakteristik *cluster*

Tidak seperti dengan metode klasifikasi, teknik *clustering* secara otomatis membagi dalam beberapa kelompok data sebelum label kelas diketahui. Namun, pada metode *clustering* juga dapat digunakan untuk kumpulan data dengan penunjukan kelas yang diketahui, dengan demikian metode *clustering* bersifat bersifat tanpa arahan (*unsupervised learning*).

#### 2.1.4 Algoritma *constrained k-means clustering*

Pada tahun 1967 McQueen menciptakan istilah *K-Means* untuk pertama kalinya sebagai algoritma *clustering*. Sampai saat ini telah menduduki peringkat terpopuler dari algoritma *clustering* lainnya. Alasannya adalah terletak pada kesederhanaan algoritma, fleksibilitas, implementasi yang mudah, serta digunakan di berbagai bidang. Hal ini didasarkan pada prinsip pola kerja algoritma *K-Means* pada konsep penentuan pusat awal (*centroid*) yang digunakan dalam mendefinisikan *cluster*, yaitu dengan membagi kumpulan data menjadi kelas-kelas (partisi) dari data yang diberikan ke dalam  $k$  *cluster* yang akan dibentuk, kemudian digunakan untuk menentukan jarak dari setiap titik data ke  $k$  *centroid* yang berbeda.

*K-Means* merupakan algoritma partisi terpopuler untuk *clustering* data karena diasumsikan efisien, namun algoritma *K-Means* juga memiliki beberapa kelemahan. Misalnya, rentan terhadap titik *noise* dan titik terisolasi. Pada saat yang sama, karakteristik *Euclidean Distance* menunjukkan bahwa algoritma *K-Means* hanya berlaku untuk kumpulan data dengan pola bentuk lingkaran dengan ukuran dan kepadatan yang sama. Gambar 5 menunjukkan keterbatasan algoritma *K-Means* tradisional.

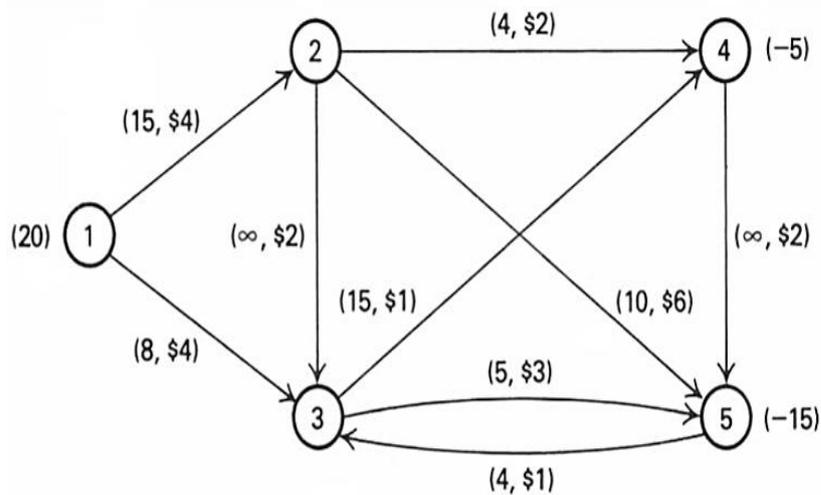


**Gambar 5.** Keterbatasan algoritma *k-means* tradisional

Kekurangan lain dari algoritma *K-Means* yaitu ketidakseimbangan (*imbalance*) pada *cluster* yang terbentuk. Dalam penerapan *K-Means* untuk kasus *clustering* yang tidak diketahui hasil bentuk dari pola *clusternya*, terlebih dahulu harus mengidentifikasi area yang tingkat kepadatannya objeknya yang tinggi terhadap dataset yang ada, kemudian mendistribusikannya ke setiap *cluster* yang telah terbentuk sehingga setiap *cluster* memiliki populasi atau kepadatan yang sama, dengan demikian menghasilkan optimasi pada hasil *clustering* yang terbentuk (*balanced clustering*).

Kelemahan dari algoritma *K-Means* lainnya adalah sering menyatukan data objek pada satu atau lebih *cluster* kosong atau mengumpulkan titik data yang sangat sedikit. Dalam beberapa kasus, ada batasan ukuran *cluster* yang mengharuskan setiap ukuran *cluster* berada dalam kisaran tertentu.

Solusi untuk masalah ini pertama kali disajikan dalam sebuah artikel penelitian yang dilakukan (Bradley *et al.*, 2000). Mereka mengusulkan penambahan *constraints* pada masalah optimasi *clustering*. Ini berarti bahwa setiap *cluster* harus memiliki jumlah poin minimum, berdasarkan *K-Means clustering*. Mereka mengubah langkah penugasan *cluster* menjadi masalah *Minimum Cost Flow* (MCF) dan menyelesaikannya dengan optimasi *programmang linier*.



**Gambar 6.** Model minimum cost flow (MCF)

Dapat dilihat pada Gambar 6 di atas menunjukkan bahwa terdapat lima simpul dengan busur berarah (panah). Setiap simpul memiliki nilai permintaan (negatif) atau penawaran (positif), dan busur memiliki nilai aliran dan biaya. Misalnya, busur 2–4 memiliki aliran 4 dan biaya \$2. Demikian pula, simpul satu memasok 20 unit, dan simpul 4 membutuhkan lima unit.

Algoritma *clustering K-Means* telah digunakan di banyak bidang, akan tetapi kelemahan dari algoritma ini adalah dapat berkumpul pada *cluster* kosong atau lebih dari satu *cluster* yang berisi sejumlah data subjek (misalnya satu subjek).

Bradley *et al.*, (2000) mengusulkan algoritma *Constrained K-Means* yang menambahkan batasan  $k$  ke masalah optimasi *clustering* yang mengharuskan subjek  $\tau_h$  untuk dimasukkan dalam  $k$  *cluster*.

Pada kumpulan dataset  $D = \{x_i\}_m$  dari  $m$  subjek di  $R^n$  dan sejumlah  $k$  *cluster* yang diinginkan, permasalahan algoritma *K-Means* tradisional terletak pada pencarian titik pusat *cluster*  $c_1, c_2, \dots, c_k$  di  $R^n$  sehingga jumlah kuadrat jarak antara setiap subjek  $x_i$  dengan pusat *cluster* terdekatnya  $c_h$  diminimalkan.

$$\min_{c_1, c_2, \dots, c_k} \sum_{i=1}^m \min_{h=1, \dots, k} (\|x_i - c_h\|^2) \quad (1)$$

Persamaan di atas ekuivalen dengan masalah berikut dimana *min - operation* dalam penjumlahan dihilangkan dengan memasukkan variabel indikator  $T_{i,h}$ .

$$\min_{C,T} \sum_{i=1}^m \sum_{h=1}^K T_{i,h} (\|x_i - c_h\|^2)$$

Subject to

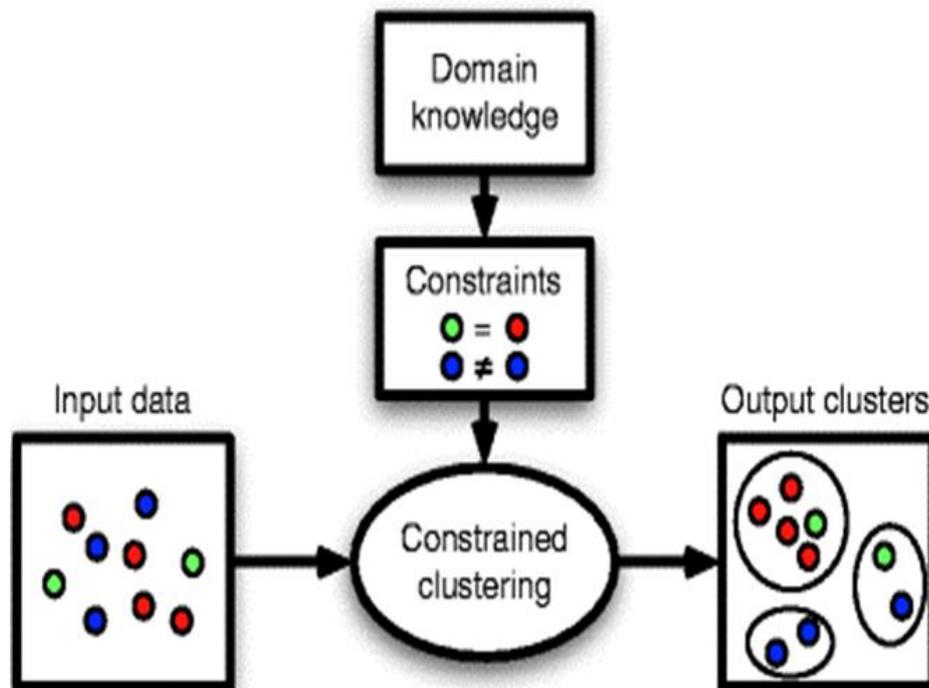
$$\sum_{h=1}^k T_{i,h} = 1; i = 1, \dots, m$$

$$T_{i,h} \geq 0, i = 1, \dots, m; h = 1, \dots, k \quad (2)$$

Pada persamaan di atas dapat dilihat bahwa  $T_{i,h} = 1$ , subjek  $x_i$  paling dekat dengan pusat *cluster*  $c_h$  sedangkan 0 (nol) sebaliknya.

Persamaan (2) secara matematis setara dengan (1) dan dinyatakan sebagai masalah *integer linear programming* karena kemungkinan  $T_{i,h}$  adalah 0 atau 1.

*Constrained K-Means* merupakan algoritma modifikasi dari formula tradisional *K-Means*, dalam menjalankan fungsinya melakukan pendekatan pada *Linear Programming Algorithm* (LPA), dengan mewajibkan setiap *cluster* memiliki subjek yang telah ditentukan.



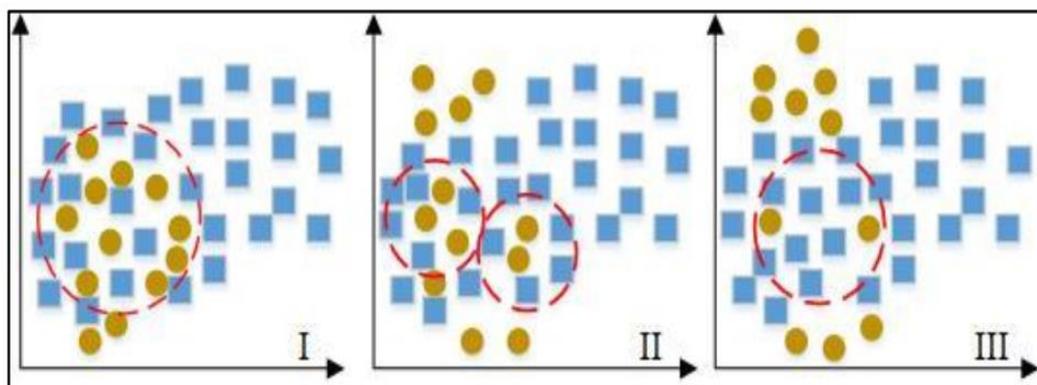
**Gambar 7.** Ilustrasi proses *constrained clustering*

Langkah-langkah pada *Constrained K-Means* dengan pendekatan *Linear Programming Algorithm* (LPA) dapat digambarkan secara singkat sebagai:

- a. Memilih pusat *cluster* awal ( $k$ ) baik secara *random sampling* (acak) atau dari solusi *Constrained K-Means*.
- b. Gunakan prosedur *Linier Pemrogramming Algorithm* (LPA) untuk mencari *clustering* yang optimal dengan *constraint* yang mengharuskan setiap *cluster* minimal memiliki 2 subjek.
- c. Perbarui pusat *cluster* dari hasil prosedur *Linear Programming Algorithm* (LPA).
- d. Jika masih terdapat perubahan dari keanggotaan *cluster* maka ulangi langkah b dan c.
- e. Ulangi langkah a - e untuk sejumlah dataset objek awal. Solusi *cluster* dengan nilai objektif minimal akan menjadi solusi akhir.

### 2.1.5 *Imbalanced data*

*Imbalanced data* atau ketidakseimbangan data adalah suatu keadaan pada kumpulan data jika jumlah data pada kelas tertentu (kelas mayoritas) lebih banyak daripada jumlah data pada kelas lain (kelas minoritas). Keadaan ini merupakan salah satu hal yang penting untuk diatasi dalam kasus *data mining* karena dapat mempengaruhi hasil komputasi. Terdapat 3 sifat kompleksitas pada kasus *imbalanced data* yang ditemukan oleh peneliti yaitu *overlap*, *small disjunct*, dan *outliner*. Gambar 8 menunjukkan ilustrasi sifat kompleksitas *imbalanced data*.



**Gambar 8.** Ilustrasi sifat kompleksitas *imbalanced data*: *overlap* (I), *small disjunct* (II), *outlier* (III)

Uraian penjelasan sifat kompleksitas *imbalanced* data sebagai berikut:

- a. *Overlap*: Keadaan kumpulan data yang banyak antara kelas. Jika suatu kumpulan data mengalami *overlapping*, maka *discriminative rule* data tersebut akan sulit untuk diproses sehingga berdampak pada keadaan semakin besarnya data pada kelas minoritas yang menghasilkan jumlah data yang lebih minim.
- b. *Small disjunct*: Kondisi beberapa data dalam *sub-cluster* memiliki jarak yang terlalu dekat antara kedua kelas. Dengan Adanya *sub-cluster* yang saling berdekatan dapat menambah kompleksitas data karena pada umumnya jumlah data pada masing-masing *sub-cluster* tersebut tidak seimbang.
- c. *Outlier*: Kondisi ketika kondisi data memiliki nilai ekstrim atau berbeda secara signifikan dari kelompok mayoritas.

### 2.1.6 Sistem informasi geografis

Sistem Informasi Geografis (SIG) atau *Geographic Information System* (GIS) adalah sebuah sistem yang terintegrasi menangkap, menyimpan, memanipulasi, menganalisa, mengatur dan menampilkan seluruh jenis data geografis.

Istilah sistem informasi geografis didefinisikan sebagai sistem berbasis komputer yang mampu memproses data bereferensi geografis terkait dengan pemasukan data, pengelolaan data, manipulasi dan analisis, serta pengembangan dan pencetakan produk.

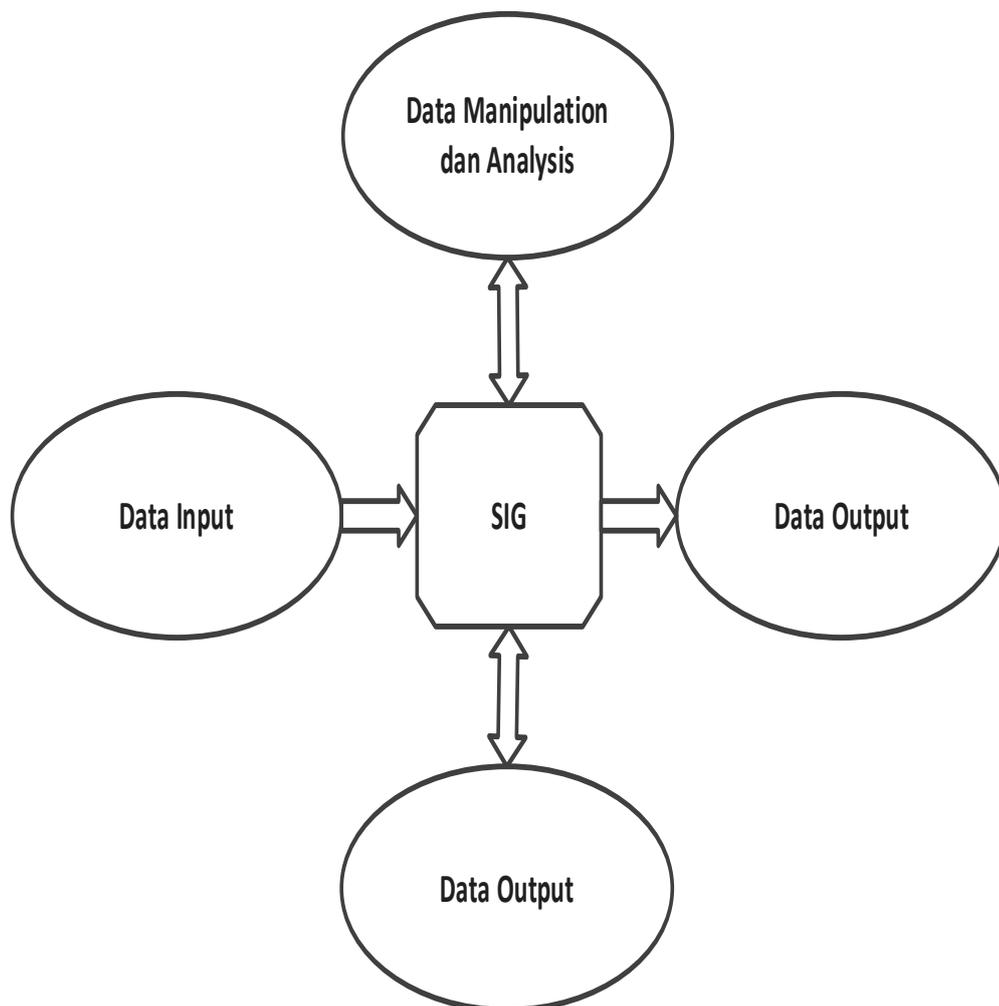
Berdasarkan pengertian yang telah diuraikan, dapat disimpulkan bahwa sistem informasi geografis adalah sistem komputer yang mengolah informasi geospasial, sistem ini terdiri dari tahapan input dan memproses hingga output.

Komponen-komponen pendukung SIG terdiri dari lima komponen yang bekerja secara terintegrasi yaitu perangkat keras (*hardware*), perangkat lunak (*software*), data, manusia, dan metode.

Untuk membuat suatu perancangan sistem pengambilan keputusan yang berkaitan dengan informasi geospasial diperlukan analisis data yang bereferensi geografis.

Analisis ini harus didukung oleh konsep ilmiah dan sejumlah data yang baik. Data atau informasi yang berkaitan dengan permasalahan akan dipecahkan dengan cara disortir dan diolah melalui pemrosesan yang akurat. Untuk keperluan tersebut SIG menyediakan sejumlah subsistem antara lain data *input*, data *output*, data *management*, serta data *manipulation* dan *analysis*.

Untuk melihat ilustrasi kerja dalam subsistem, sistem informasi geografis, dapat dilihat dalam Gambar 9.



**Gambar 9.** Ilustrasi subsistem SIG

## 2.2 Penelitian Terkait

Beberapa penelitian terkait mengenai *Constrained K-Means clustering*, sistem zonasi dan sistem penentuan sistem zonasi sekolah dengan beberapa metode berbeda seperti pada beberapa penelitian:

1. Penelitian yang dilakukan oleh (Bradley *et al.*, 2000) mengembangkan algoritma *K-Means clustering* dengan melakukan usulan pendekatan terhadap metode *Linear Programming Algorithm* (LPA), mengingat dimana nilai “*k*” dalam algoritma tradisional *K-Means* sesuai dengan dengan jumlah pusat (*means*) *cluster* di sekitar *cluster* yang akan terbentuk dengan demikian tidak ada *cluster* yang terbentuk dari jumlah nilai “*k*”. Dengan metode yang diusulkan oleh (Bradley *et al.*, 2000) mengharuskan setiap *cluster* memiliki setidaknya 2 atau lebih subjek di setiap *cluster* yang terbentuk serta menghasilkan optimasi *cluster* yang lebih baik terhadap data yang tidak seimbang.
2. Pada penelitian lainnya yang pernah dilakukan oleh (Wagstaff *et al.*, 2001) untuk pertama kali mengaplikasikan algoritma *COP - K-Means*, mendemonstrasikan algoritma *K-Means clustering* dimodifikasi dengan menerapkan *Pairwise Constraints* sebagai latar belakang ilmu pengetahuan untuk membatasi penugasan algoritma *K-Means*, mereka melakukan pengujian metode ini untuk mendeteksi jalur jalanan dari data *Global Positioning System* (GPS). Hasil pengujian penelitian ini menyatakan bahwa kinerja algoritma *COP - K-Means* secara umum lebih baik daripada algoritma konvensional *K-Means*. Namun algoritma *COP - K-Means* secara implisit menentukan kondisi yang harus dipenuhi berdasarkan urutan alokasi data dari proses *K-Means clustering*, yang menghasilkan hasil pengelompokan yang tidak stabil.
3. Okabe dan Yamada (2018) pada penelitiannya mengusulkan metode *Constrained Clustering* dengan mengkombinasikan algoritma baru *Constrained K-Means* dengan prinsip *boosting*. Dalam menjalankan tugasnya sebagai metode *Constrained Clustering* memiliki keunggulan dalam hal waktu komputasi yang lebih sedikit dibandingkan dengan algoritma tradisional *K-Means*.

Hasil eksperimen pada 12 kumpulan data dari tiga sumber data menunjukkan bahwa metode ini memiliki kinerja yang kompetitif dengan metode *Constrained Clustering* yang canggih untuk dataset yang besar dan membutuhkan waktu komputasi yang jauh lebih sedikit. Terlepas dari keunggulan komputasi algoritma ini, namun secara umum dalam berkinerja algoritma ini membutuhkan beberapa *constraints* untuk mengoptimalkan kekuatan *boosting* karena metode ini melakukan pendekatan *Ensemble Learning*.

4. Penelitian lainnya yang dilakukan (Lei *et al.*, 2013) menyatakan bahwa hasil dari proses *clustering* sangat bergantung pada *initial center (centroid)* pada algoritma *K-Means* dan *Size - Constrained Clustering*. Oleh karena itu, mempertimbangkan dua konsep untuk mengoptimalkan *Size - Constrained Clustering*, yang pertama meningkatkan kecepatan konvergensi selanjutnya mengurangi RMSE. Penelitian ini mengusulkan metode *Size - Constrained Clustering* sebagai pemilihan titik awal dengan menemukan titik padat di sekitar *centroid* yang telah terbentuk secara rekursif. *Root - Mean - Square Error* dan kecepatan konvergensi (jumlah iterasi) adalah dua metrik terpenting untuk pengelompokan menggunakan metode iteratif. Pengujian dari penelitian ini dilakukan pada sekitar sepuluh ribu proposal penelitian *National Natural Science Foundation of China* dan hasilnya menunjukkan bahwa metode ini dapat mengurangi waktu iterasi hingga lebih dari 50% dan mendapatkan *root mean square error* yang lebih kecil, dengan demikian metode ini dapat dikombinasikan dengan skala *Size - Constrained Clustering* untuk mengatasi masalah *clustering* data dengan skala besar dalam *data mining*. Selanjutnya terdapat dua bentuk pengembangan penelitian ini yang dapat dilakukan di masa akan datang. Pertama mengurangi kompleksitas waktu dari algoritma pemilihan titik *centroid* awal dengan menghilangkan perhitungan yang berlebihan, Kemudian, bereksperimen pada kumpulan data yang lebih kompleks untuk pengaplikasian permasalahan *real*.

5. Thota *et al.*, (2017) melakukan penelitian yang berjudul “*Cluster based Zoning of Crime Info*” dimana penelitian tersebut mengenai zonasi wilayah kejahatan yang terjadi di negara India, dimodelkan dengan pemanfaatan *data mining*, tindakan yang dilakukan antara lain mengumpulkan, menganalisis dan mempublikasikan data kejahatan menggunakan analisis *cluster*. Penelitian ini mengusulkan algoritma *K-Means* yang berfungsi sebagai pembentuk pola wilayah yang rawan tindak kejahatan, dataset yang digunakan adalah kumpulan dari laporan info kejahatan lembaga *The National Crime Records Bureau* (NCRB). Input dari data hasil *clustering* divisualisasikan dengan menggunakan peta khusus zona *cluster* yang berguna untuk menampilkan info kejahatan pada negara bagian, sehingga dapat membantu polisi dan penegak hukum dalam mengambil tindakan pencegahan untuk memerangi kejahatan dan merencanakan operasi yang strategi dalam melakukan tindakan investigasi. Kelanjutan dari penelitian ini dapat dikembangkan dengan berbagai metode *data mining* pada kumpulan data *The National Crime Records Bureau* (NCRB) India dan menggunakan atribut kejahatan yang lebih banyak untuk mengidentifikasi *trending* kejahatan.
6. Penelitian yang dilakukan oleh (Alifi, Hayati, dan Supangkat., 2017) Berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia tentang standar pelayanan minimum untuk pendidikan dasar, pemerintah kota Bandung memberlakukan sistem zonasi sekolah, penelitian ini mengusulkan rencana jalur sekolah untuk meminimalkan biaya transportasi sekolah, serta mempertimbangkan pembangunan dan perkiraan kapasitas daya tampung sekolah negeri. Metode yang digunakan adalah *P - Median* sebagai model untuk jalur akses sekolah, pelaksanaan penelitian ini menggunakan perangkat lunak *ArcGIS* untuk menganalisis pemetaan akses jalur ke sekolah serta antara sekolah satu dengan yang lainnya. Dari hasil analisis pemetaan akses jalur yang diusulkan menurunkan impedansi sebesar 26,55% dibandingkan dengan akses jalur sebelumnya. Hasil dari penelitian ini merekomendasi untuk mengurangi pendaftaran sekolah di 11

kecamatan, dan meningkatkan pendaftaran sekolah di dua kecamatan, serta membangun sekolah baru di 13 kecamatan.

7. Sitanggang, Radiatun, dan Nur Risal (2020) dengan judul penelitian “*Incremental Spatio Temporal Clustering Application on Hotspot Data*”. Titik panas merupakan salah satu indikator kebakaran hutan dan lahan. Analisis data titik api perlu dilakukan sebagai langkah peringatan dini untuk mencegah terjadinya kebakaran hutan dan lahan. Penelitian ini bertujuan untuk membuat aplikasi *clustering* data *hotspot* berbasis *web* menggunakan modul *Incremental ST - DBSCAN*. Dataset yang digunakan pada penelitian ini adalah kumpulan data *hotspot* dari tahun 2014 sampai dengan tahun 2017. *Software* yang dikembangkan menggunakan bahasa pemrograman *R* dan *framework Shiny*. Metodologi yang diterapkan adalah *Adaptive Software Development (ASD)*. Fitur utama dari aplikasi ini adalah *incremental clustering* data *hotspot* serta visualisasi hasil *clustering*. Pengujian dilakukan dengan pendekatan *Black Box*, menunjukkan bahwa semua fitur *software* berfungsi dengan baik. Berdasarkan hasil evaluasi *usability*, kepuasan pengguna mencapai 78.5%, yang berarti *software* ini sangat mudah digunakan.
8. Selanjutnya penelitian yang dilakukan oleh Wang dan Gao (2019), menjelaskan algoritma *K-Means* banyak digunakan dalam proses *clustering* dimana bertugas untuk memproses data ke dalam kelas yang berbeda, karena kesederhanaan dan efisiensinya, serta telah menjadi alat penting dalam *data mining*. Akan tetapi, algoritma *tradisional K-Means* hanya dapat mendeteksi *cluster* dengan pola lingkaran (bola) dan rentan terhadap *noise* atau titik terisolasi, yang dapat mempengaruhi hasil *clustering*. Untuk mengatasi masalah tersebut, pada penelitian ini mengusulkan algoritma *K-Means* yang ditingkatkan berdasarkan uji *Kurtosis* dan metode *Monte Carlo*. Algoritma *K-Means* ditingkatkan agar dapat beradaptasi dengan kumpulan data berbentuk kompleks, dan mengurangi dampak *outlier* data pada hasil *clustering*, serta membuat hasil algoritma *K-Means* lebih akurat. Hasil penelitian menunjukkan bahwa algoritma yang diusulkan memiliki indeks eksternal yang lebih besar dari kinerja *metrics clustering*.

Hal ini berarti akurasi hasil *clustering* meningkat secara signifikan. Selanjutnya untuk pengembangan penelitian ini agar lebih memperhatikan *thresholds* yang ditetapkan dalam proses komputasi algoritma. Selain itu, untuk arah penelitian prospektif menjabarkan penggunaan nilai langsung *Kurtosis* sebagai indikator dinamis dalam proses iterasi algoritma.

9. Syaripudin *et al.*, (2019) mereka mendesain sistem rekayasa penentuan sekolah yang terdekat dengan menggunakan formula *Haversine* dengan tujuan menghitung jarak antara situs sekolah dan rumah siswa, dengan memanfaatkan layanan *location based service*. Untuk penunjuk posisi awal dengan menggunakan *Global Positioning System* (GPS), hasil dari penelitian ini sangat baik dibandingkan hasil prediksi *default Google Distance* dari Google. Akan tetapi, dalam memprediksi perhitungan jarak lokasi sekolah terdekat berdasarkan sistem zonasi tingkat keakuratannya tidak selalu optimal sebab sistem rekayasa ini mengacu pada tingkat akurasi sistem GPS di perangkat yang digunakan.
10. Selanjutnya penelitian yang dilakukan (Ratnasar *et al.*, 2020) yang terkait mengenai penerimaan siswa berbasis zonasi yaitu dengan menggunakan metode *Generate and Test* untuk menganjurkan rekomendasi jarak terpendek terbaik dan akurat, adapun dalam pelaksanaannya tanpa melibatkan aplikasi berbasis denah lainnya. Hasil dari penelitian ini diuji dengan kalkulasi formula *Haversine*, dan dalam memprediksi pengukuran jarak menggunakan *google maps*. Namun, dalam berkinerja membutuhkan waktu yang cukup lama dan penggunaan memori yang besar yaitu 6,2 MB dan waktu rata-rata 0,0000216806982 Ms.
11. Penelitian lainnya yang mengacu untuk mengidentifikasi sistem penentuan zonasi penerimaan siswa baru yaitu dengan menggunakan algoritma *K-Means*. Hasil yang diperoleh pada penelitian ini menyatakan bahwa dengan menggunakan algoritma *K-Means* berdasarkan distribusi data siswa relatif baik terhadap acuan parameter jarak antara domisili siswa dan sekolah, serta terbentuknya penetapan area penerimaan siswa berbasis zonasi secara non lingkaran.

### 2.3 State of The Art

**Tabel 1.** *State of the art* penelitian

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
1.	<i>Constrained K-Means Clustering</i>	(Bradley, Bennett, and Demiriz 2000)	<i>Microsoft Research</i>	2000	1. Melakukan usulan pendekatan terhadap metode <i>Linear Programming Algorithm</i> (LPA) terhadap <i>K-Means</i> tradisional 2. Optimasi <i>cluster</i>	Hasil pengujian pada dataset real numerik menyatakan bahwa <i>Constrained K-Means</i> mampu mengatasi permasalahan <i>cluster</i> kosong terhadap keterbatasan <i>K-Means</i> tradisional yang meringkas sejumlah data	1. Mengharuskan setiap <i>cluster</i> memiliki setidaknya 2 atau lebih subjek di setiap <i>cluster</i> serta menghasilkan optimasi <i>cluster</i> yang lebih baik terhadap data yang tidak seimbang 2. Mengatasi permasalahan <i>cluster</i> kosong atau <i>cluster</i> yang memiliki objek yang sangat sedikit
2.	<i>Constrained K-Means Clustering with Background Knowledge</i>	(Wagstaff <i>et al.</i> , 2001)	<i>International Conference on Machine Learning ICML</i>	2001	1. <i>COP - K-Means</i> 2. Melakukan usulan pendekatan <i>Pairwise Constraints</i>	Hasil pengujian tersebut menyatakan bahwa kinerja algoritma <i>COP - K-Means</i> secara umum lebih baik daripada algoritma konvensional <i>K-Means</i>	1. Melakukan pengujian metode ini untuk mendeteksi jalur jalanan dari data <i>Global Positioning System</i> (GPS) 2. Algoritma <i>COP - K-Means</i> secara implisit menentukan kondisi yang harus dipenuhi berdasarkan urutan alokasi data dari proses <i>K-Means clustering</i> , yang menghasilkan hasil <i>clustering</i> yang tidak stabil

Lanjutan tabel 1.

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
3.	<i>Clustering Using Boosted Constrained K-Means Algorithm</i>	(Okabe and Yamada 2018)	Frontiers in Robotics and AI	2018	1. Mengusulkan metode <i>Constrained Clustering</i> dengan mengkombinasikan algoritma baru <i>Constrained K-Means</i> dengan prinsip <i>boosting</i>	Hasil eksperimen pada 12 kumpulan data dari tiga sumber data menunjukkan bahwa metode ini memiliki kinerja yang kompetitif dengan metode <i>Constrained clustering</i> yang canggih untuk sebagian besar kumpulan data dan membutuhkan waktu komputasi yang jauh lebih sedikit	1. Terlepas dari keunggulan komputasi algoritma ini, namun secara umum dalam berkinerja algoritma ini membutuhkan beberapa <i>constraints</i> untuk mengoptimalkan kekuatan <i>boosting</i> karena metode ini melakukan pendekatan <i>Ensemble Learning</i>
4.	<i>Size-Constrained Clustering Using an Initial Points Selection Method</i>	(Lei et al., 2013)	<i>International Conference on Machine Learning ICML</i>	2013	1. <i>Size-Constrained Clustering</i> 2. <i>Root-Mean -Square Error (RMSE)</i>	Hasilnya menunjukkan bahwa metode dapat mengurangi waktu iterasi hingga lebih dari 50% dan mendapatkan <i>root mean square error</i> yang lebih kecil, dengan demikian metode ini dapat dikombinasikan dengan skala <i>Size-Constrained Clustering</i> untuk mengatasi masalah <i>clustering</i> skala besar dalam <i>data mining</i>	1. Pengujian dari penelitian ini dilakukan pada sekitar sepuluh ribu proposal penelitian <i>National Natural Science Foundation of China</i>

Lanjutan tabel 1.

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
5.	<i>Cluster Based Zoning of Crime Info</i>	(Thota <i>et al.</i> , 2017)	2017 2nd International Conference on Anti-Cyber Crimes, ICACC 2017	2017	1. Dalam penelitian ini untuk menganalisis data mengusulkan algoritma <i>K-Means</i>	Dari input data hasil <i>clustering</i> divisualisasikan dengan menggunakan peta khusus dengan zona <i>cluster</i> yang berguna menampilkan info kejahatan pada negara bagian, sehingga membantu polisi dan penegak hukum mengambil tindakan pencegahan untuk memerangi kejahatan dan merencanakan operasi strategi untuk investigasi lanjutan	1. Dataset yang digunakan adalah kumpulan dari laporan info kejahatan lembaga <i>The National Crime Records Bureau (NCRB)</i> 2. Untuk kelanjutan dari penelitian ini dapat dikembangkan dengan berbagai metode <i>data mining</i> pada kumpulan data <i>The National Crime Records Bureau (NCRB)</i> India dan menggunakan atribut kejahatan yang lebih banyak untuk mengidentifikasi tren dan pola kejahatan
6.	<i>Optimization of School Network Using Location-Allocation Analysis: Case study: Bandung, Indonesia</i>	(Alifi, Hayati, and Supangkat 2017)	TENSYMP 2017 - IEEE International Symposium on Technologies for Smart Cities	2017	1. Metode yang digunakan adalah <i>P - Median</i>	Dari hasil analisis pemetaan akses jalur yang diusulkan menurunkan impedansi sebesar 26,55% dibandingkan dengan akses jalur sebelumnya. Adapun hasil dari penelitian ini merekomendasi untuk mengurangi pendaftaran sekolah di 11 kecamatan, dan meningkatkan pendaftaran sekolah di dua kecamatan, serta membangun sekolah baru di 13 kecamatan	1. Penelitian ini mengusulkan rencana jalur sekolah untuk meminimalkan biaya transportasi sekolah, serta mempertimbangkan pembangunan dan perkiraan kapasitas daya tampung sekolah negeri 2. Pelaksanaan penelitian ini menggunakan perangkat lunak <i>ArcGIS</i> untuk menganalisis pemetaan akses jalur ke sekolah serta antara sekolah satu dengan yang lainnya

Lanjutan tabel 1.

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
7.	<i>Incremental Spatio Temporal Clustering Application on Hotspot Data</i>	(Sitanggang, Radiatun, and Risal 2020)	IOP Conference Series: Earth and Environmental Science	2020	<ol style="list-style-type: none"> <li>1. Modul <i>Incremental ST-DBSCAN</i>.</li> <li>2. Dalam studi ini, metodologi yang diterapkan adalah <i>Adaptive Software Development (ASD)</i>.</li> </ol>	Berdasarkan hasil evaluasi <i>usability</i> , kepuasan pengguna mencapai 78.5%, yang berarti aplikasi ini sangat mudah digunakan.	<ol style="list-style-type: none"> <li>1. Data set yang digunakan pada penelitian ini adalah kumpulan data <i>hotspot</i> dari tahun 2014 sampai dengan tahun 2017</li> <li>2. Aplikasi <i>clustering</i> data <i>hotspot</i> berbasis web</li> <li>3. Fitur utama dari aplikasi ini adalah <i>incremental clustering</i> data <i>hotspot</i> dan visualisasi hasil <i>clustering</i></li> <li>4. Pengujian dilakukan dengan pendekatan <i>Black Box</i>.</li> </ol>
8.	<i>An Improved K-Means Algorithm Based on Kurtosis Test</i>	(Wang and Gao 2019)	Journal of Physics: Conference Series	2019	<ol style="list-style-type: none"> <li>1. Mengusulkan algoritma <i>K-Means</i> yang ditingkatkan berdasarkan uji <i>Kurtosis</i> dan metode <i>Monte Carlo</i>.</li> </ol>	Hasil eksperimen menunjukkan bahwa algoritma yang diusulkan memiliki indeks eksternal yang lebih besar dari kinerja dari <i>metrics clustering</i> . Ini berarti akurasi hasil <i>clustering</i> meningkat secara signifikan.	<ol style="list-style-type: none"> <li>1. Menyatakan kelemahan algoritma <i>tradisional K-Means</i> hanya dapat mendeteksi <i>cluster</i> dengan pola lingkaran (bola) dan rentan terhadap <i>noise</i> atau titik terisolasi, yang dapat mempengaruhi hasil <i>clustering</i></li> <li>2. Algoritma <i>K-Means</i> ditingkatkan berdasarkan uji <i>Kurtosis</i> dan metode <i>Monte Carlo</i> agar dapat beradaptasi dengan kumpulan data berbentuk kompleks, serta mengurangi dampak <i>outlier</i> data pada hasil <i>clustering</i></li> </ol>

Lanjutan tabel 1.

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
9.	<i>Haversine Formula Implementation to Determine Bandung City School Zoning Using Android Based Location Based Service</i>	(Syaripudin <i>et al.</i> , 2019)	1st International Conference on Islam, Science and Technology, ICONISTECH 2019	2019	1. Formula <i>Haversine</i> 2. <i>Global Positioning System</i> (GPS)	Hasil dari penelitian ini sangat baik dibandingkan hasil prediksi <i>default Google Distance</i> dari Google	1. Menggunakan formula <i>haversine</i> dengan tujuan menghitung jarak antara situs sekolah dan rumah, dimana memanfaatkan layanan <i>Location based Service</i> 2. Namun dalam memprediksi perhitungan lokasi sekolah terdekat berdasarkan sistem zonasi tingkat keakuratannya tidak selalu optimal sebab sistem rekayasa ini mengacu pada tingkat akurasi sistem GPS di perangkat yang digunakan
10.	<i>Implementation of Generate and Test Algorithm for Junior High School Zoning System in Malang</i>	(Ratnasar <i>et al.</i> , 2020)	4th International Conference on Vocational Education and Training, ICOVET 2020	2020	1. Menggunakan metode <i>Generate and Test</i>	Hasil dari penelitian diuji dengan kalkulasi formula <i>Haversine</i> , dan dalam memprediksi pengukuran jarak menggunakan <i>google maps</i>	1. Sistem penerimaan siswa berbasis zonasi 2. Menggunakan metode <i>Generate and Test</i> untuk menganjurkan rekomendasi jarak terpendek terbaik dan akurat 3. Namun dalam berkinerja membutuhkan waktu yang cukup lama dan penggunaan memori yang besar yaitu adalah 6,2 MB penggunaan memori dan waktu rata-rata 0,0000216806982 Ms

Lanjutan tabel 1.

No	Judul	Penulis	Penerbit	Tahun	Metode	Hasil	Keterangan
11.	<i>School Zoning System Using K-Means Algorithm for High School Students in Makassar City</i>	(Febriana, Zainuddin, and Nurtanio 2019)	2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019	2019	1. Mengusulkan metode <i>K-Means clustering</i>	Hasil yang diperoleh pada penelitian ini menyatakan bahwa dengan menggunakan algoritma <i>K-Means</i> dengan berdasarkan distribusi data siswa relatif baik terhadap acuan parameter jarak antara domisili siswa dan sekolah, serta terbentuknya penetapan area penerimaan siswa berbasis zonasi secara non lingkaran	1. Mengidentifikasi sistem penentuan zonasi penerimaan siswa 2. Namun hasil dari penelitian ini memperlihatkan bahwa setiap sekolah memiliki anggota <i>cluster</i> sendiri yang tidak seimbang atas sebaran siswa antara <i>cluster</i> satu dengan <i>cluster</i> lainnya, dalam hal ini sekolah sebagai titik <i>centroid cluster</i>

## 2.4 Kerangka Pikir

Kerangka pikir menjelaskan alur penelitian yang akan dilakukan. Pada tahap pertama menjelaskan permasalahan yang ada dalam hal percepatan pemerataan kualitas pendidikan di Indonesia terkhusus pada aturan Penerimaan Peserta Didik Baru (PPDB) dengan menerapkan sistem zonasi.

Kebijakan pelaksanaan sistem zonasi ini tertuang pada Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) nomor 51 tahun 2018 tentang kebijakan pemerintah yang mengatur sistematika penerimaan peserta didik baru pada semua jenjang sekolah pelaksanaan sistem zonasi ini bertujuan untuk melaksanakan penerimaan peserta didik baru yang objektif, transparan, tanpa diskriminasi, dan berkeadilan dalam rangka meningkatkan akses layanan dan mutu pendidikan.

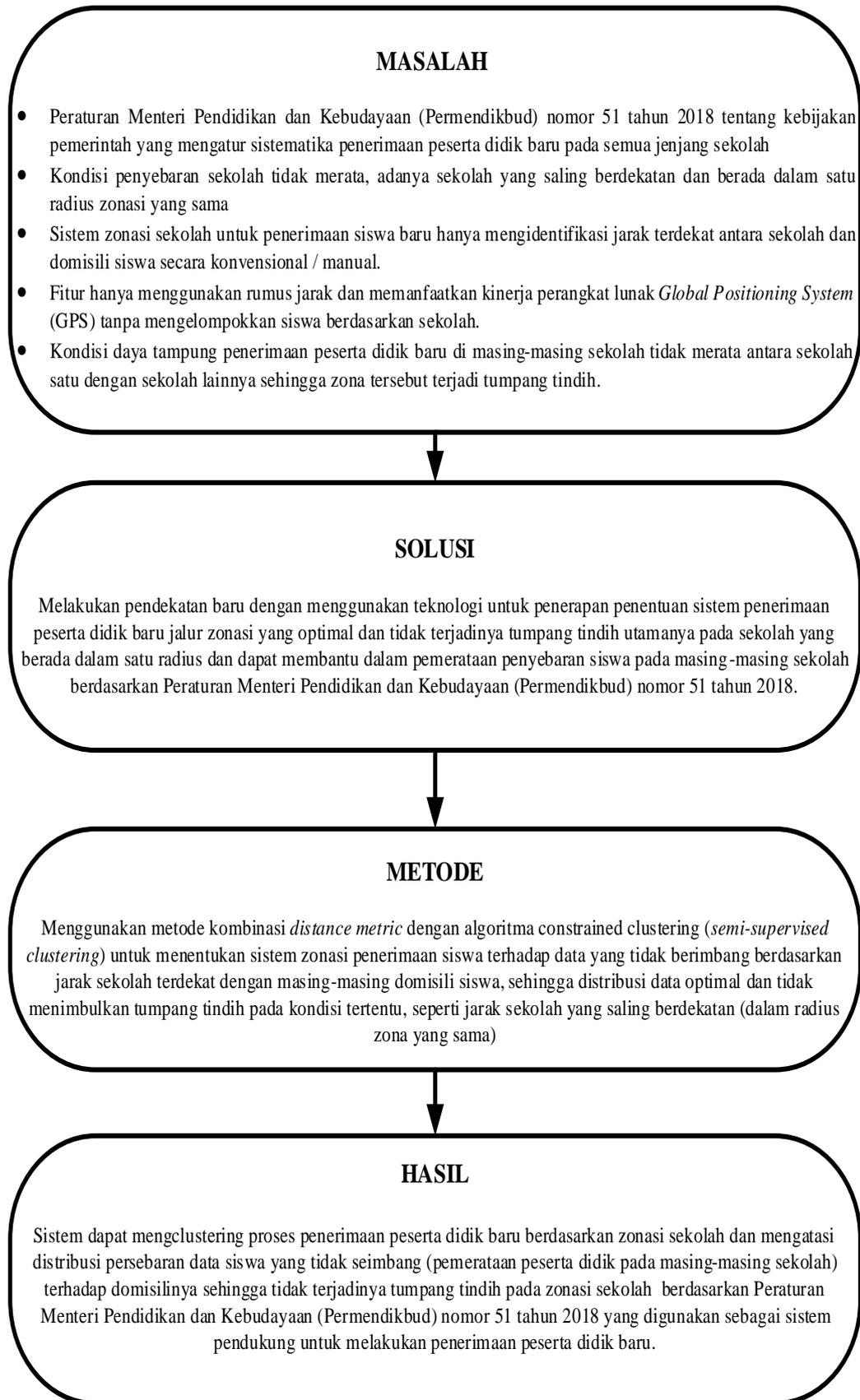
Sistem pelaksanaan penerimaan peserta didik berbasis zonasi masih dilakukan secara konvensional dan tentunya memiliki banyak kekurangan dari segi efisiensi dan efektifitas serta ditemukan banyak calon siswa yang pendaftarannya tidak diakomodir oleh pihak sekolah. Sementara itu, jarak domisili calon siswa dengan sekolah tujuan tidak saling berjauhan, permasalahan lainnya yaitu persebaran satuan pendidikan sekolah yang tidak merata, terbukti banyaknya sekolah yang letaknya saling berapit satu sama lain atau berada dalam satu radius yang mengakibatkan area tersebut akan tumpang tindih dalam hal kapasitas. Oleh karena itu, berdasarkan perkembangan teknologi khususnya dibidang *data mining*, dapat dimanfaatkan untuk mengatasi masalah tersebut yaitu dengan menggunakan metode *clustering*.

Oleh sebab itu, dilakukan perancangan sistem untuk mengatasi permasalahan tersebut dengan mengacu dari beberapa penelitian sebelumnya yang relevan, yaitu sistem hanya melakukan olah data berupa mencari jarak terdekat antara letak sekolah dengan domisili menggunakan algoritma *distance matrix* kemudian diuji dengan *Global Positioning System* (GPS) dan algoritma tradisional *K-Means* yang merupakan algoritma partisi yang banyak digunakan, namun dalam beberapa kasus ditemukan permasalahan pada batasan ukuran *cluster* yang mengharuskan setiap *cluster* berada dalam kisaran sama, dalam kasus demikian

*K-Means* sering menggabungkan satu atau lebih *cluster* yang bernilai kosong dengan demikian *K-Means* merangkum sejumlah data yang berukuran kecil.

Oleh karena itu, sistem yang dirancang menggunakan pendekatan algoritma *Constrained K-Means*.

*Constrained K-Means* merupakan algoritma modifikasi dari formula tradisional *K-Means*, dalam menjalankan fungsinya melakukan pendekatan pada *Linear Programming Algorithm* (LPA), dengan mewajibkan setiap *cluster* memiliki subjek. Implementasi hasil dari proses *clustering* terhadap banyaknya data anggota pada setiap *cluster center*/sekolah menghasilkan data yang seimbang dan tidak tumpang tindih di setiap *cluster center*/sekolah dan tidak ada *cluster* yang kosong atau bernilai 0. Adapun kerangka pikir penelitian dapat dilihat pada Gambar 10.



**Gambar 10.** Kerangka pikir penelitian