

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Terdapat beberapa referensi kajian pustaka yang digunakan dengan tema yang sama yaitu pemetaan pasar tradisional. Penelitian pertama berjudul “Pengelompokan Wilayah Persebaran Indekos Dengan Metode *K-Means* Dan Informasi Pengalokasian Akses Jalan Dengan Klasifikasi Menggunakan Analisis Data Spasial Pada Aplikasi E- Commerce Carikos” oleh (Latifah, 2017). dalam penelitian ini dibangun sebuah aplikasi CariKos yang dapat memudahkan antara pemilik indekos dan pencari indekos. Pada aplikasi yang dibangun dapat mengelompokkan wilayah indekos yang ada agar perhitungan menentukan jarak dari suatu cluster indekos ke jurusan kampus di ITS, atau ke tempat fasilitas daerah yang ada disekitar indekos tersebut menjadi lebih mudah. Uji coba dilakukan sebanyak tiga macam uji coba, yaitu uji coba *k-means*, uji coba analisis spasial, dan ujian fungsionalitas. Uji coba *k-means* dilakukan dengan menghitung sum squared error dan hasil yang didapat adalah sebesar 0,0003867. Hasil uji coba dari analisis spasial adalah menampilkan peta yang terdapat informasi akses jalan pada wilayah indekos..

Masih dengan topik yang sama yaitu terkait dengan Metode *K-means* dengan jurnal berjudul “Analisa Spasial *Clustering* Zonasi Rawan Bencana Tanah Longsor Wilayah Bogor Selatan Berbasis WebGIS” oleh (Suwardi, 2020). Tujuan dari penelitian ini adalah mengidentifikasi faktor yang berpengaruh terhadap peluang terjadinya longsor, zonasi wilayah berdasarkan tingkat kerawanan longsor, serta menganalisis hubungan antara faktor mekanika tanah dengan tingkat kerawanan. Analisis spasial dilakukan dengan sistem informasi geografis (SIG) dan menggunakan metode *clustering k-means*. Berdasarkan sebaran spasial, faktor curah hujan, jenis batuan, kemiringan lahan, dan penggunaan lahan merupakan faktor dominan yang dapat memicu kejadian longsor dengan peluang yang tinggi. Hasil penelitian ini dapat menunjukkan lokasi potensi longsor sebagai informasi awal untuk peringatan dini bahaya

longsor di wilayah penelitian, meskipun belum memperhitungkan kerugian materi. Peta Risiko Longsor ini juga dapat dijadikan dasar penataan ruang. Tingkat risiko longsor sangat dipengaruhi oleh kepadatan penduduk karena keterpaparan atau exposure sangat tergantung pada kepadatan dan tingkat kerawanan fisik longsor.

Penggunaan WEBGIS untuk mempermudah pengolahan informasi geografis juga diterapkan dalam sebuah penelitian dengan judul "Analisis Spasial *K-Means Clustering* Sebaran Keluhan Pelanggan PDAM Tirta Pakuan Berbasis Webgis" oleh (Randyka Kurniawan, 2019). tujuan dari penelitian ini untuk melakukan analisis sebaran keluhan pelanggan PDAM Tirta pakuan di Kota Bogor, lalu dapat menampilkan informasi tersebut menggunakan WebGIS agar lebih efektif dalam memberikan visualisasi data. Hasil dari penelitian ini adalah *clustering* sebaran keluhan pelanggan PDAM Tirta Pakuan Kota Bogor, telah berhasil mengelompokkan sebaran keluhan pelanggan PDAM Tirta Pakuan Kota Bogor dengan metode *K-Means clustering* berdasarkan kategori keluhan berbasis WebGIS.

Penggunaan sistem informasi geografis menggunakan metode *K-Means* juga digunakan dalam penelitian berjudul "Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode *K-Means*" oleh (Suryani, 2021). Pemetaan jalan rusak pada penelitian ini menggunakan openStreetMap dan untuk sistemnya menggunakan *website* dengan framework codeigniter. Hasil dari penelitian ini metode *K-Means* yang diterapkan pada *website* ini sudah tepat dengan tingkat presentase kecocokannya 100%.

Pemetaan menggunakan metode *K-Means Clustering* juga digunakan dalam penelitian dengan jurnal berjudul "Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karangbesuki Menggunakan Metode *K-Means Clustering*" oleh (Hasymi, 2021), Sistem yang dibangun pada penelitian ini yaitu sistem informasi geografis pemetaan warga kurang mampu menggunakan metode *clustering* serta dikombinasikan dengan QGIS 2.18. hasil dari penelitian

ini adalah memberikan informasi dan mengelompokan warga kurang mampu, sertalokasi warga kurang mampu di Kelurahan.

2.2 Pasar

Pasar adalah suatu area tempat barang diperjualbelikan dengan banyak penjual, disebut sebagai pusat perbelanjaan, pasar tradisional, toko, pusat perbelanjaan, alun-alun, pusat perdagangan, atau nama lain. Menurut (Undang-Undang Republik Indonesia Nomor 7 Tahun 2014 Tentang Perdagangan, 2014)Pasar Tradisional adalah pasar yang dibangun dan dikelola oleh Pemerintah, Pemerintah Daerah, Swasta, Badan Usaha Milik Negara dan Badan Usaha Milik Daerah termasuk kerjasama dengan swasta dengan tempat usaha berupa toko, kios, los dan tenda yang dimiliki/dikelola oleh pedagang kecil, menengah, swadaya masyarakat atau koperasi dengan usaha skala kecil, modal kecil dan dengan proses jual beli barang dagangan melalui tawar menawar.

2.2.1 Jenis- jenis Pasar

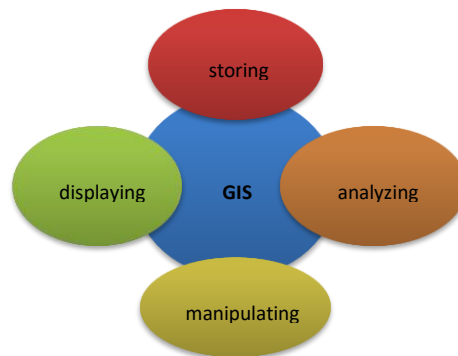
Terdapat berbagai jenis-jenis pasar, terdiri atas :

1. Jenis pasar Luas Jangkauan
 - a. Pasar Daerah membeli dan menjual barang di daerah produksinya. Dapat juga dikatakan bahwa pasar daerah melayani penawaran dan permintaan di suatu wilayah tertentu.
 - b. Pasar lokal mengacu pada pasar di mana komoditas diperjual belikan di kota yang sama. Dapat juga dikatakan bahwa pasar lokal melayani penawaran dan permintaan kota.
 - c. Pasar nasional adalah pasar di mana komoditas diperjualbelikan di negara-negara produsen. Dapat juga dikatakan bahwa pasar nasional meliputi permintaan dan penjualan dalam negeri.
 - d. Pasar internasional adalah pasar jual beli komoditi dari beberapa negara, bisa juga dikatakan pasar yang tersebar di seluruh dunia.
2. Berdasarkan jenis pasar barang yang dijual
 - a. Pasar barang konsumsi adalah pasar yang menjual barang yang dapat langsung digunakan untuk keperluan rumah tangga.

- b. Pasar produksi adalah pasar di mana faktor-faktor produksi diperjualbelikan dan pasar mesin, pasar tenaga kerja, dan pasar mata uang merupakan input untuk diperdagangkan.
3. Jenis pasar tergantung pada jam buka
 - a. Pasar harian adalah pasar untuk jual beli setiap hari. Pasar harian ini biasanya dapat ditemukan di kota-kota.
 - b. Pasar mingguan adalah pasar yang hanya buka seminggu sekali. Pasar mingguan ini ada di dalam negeri, biasanya pada hari-hari Jawa.
 - c. Pasar bulanan adalah pasar untuk jual beli sebulan sekali. Pasar Tahunan adalah pasar di mana penjualan dan pembelian dilakukan setahun sekali.
 - d. Pasar temporer adalah pasar yang diselenggarakan oleh organisasi/lembaga pada saat-saat tertentu atau hanya sesekali (tidak tetap), seperti pameran dan pameran.

2.3 Sistem Informasi Geografi

Dengan berkembangnya GIS itu sendiri, banyak definisi GIS yang muncul. Sistem informasi geografis adalah sistem informasi yang dirancang untuk memproses data referensi geografis atau koordinat geografis, atau dengan kata lain, SIG adalah sistem basis data yang dapat diakses untuk memproses data referensi geografis. Secara spasial dengan banyak karakteristik kinerja (Widana, 2019). Sistem Informasi Geografis tidak terlepas dari data spasial. Data spasial adalah data tentang posisi, objek, dan hubungan antar ruang yang ada di bumi. Data spasial adalah informasi yang memuat informasi tentang bumi, meliputi permukaan bumi, bawah tanah, air, lautan, dan atmosfer bawah (Geoinforadmin, 2017).



Gambar : 2. 1 Aspek- aspek dalam GIS

Konsep sistem informasi geografis dapat diringkas sebagai berikut:

1. Informasi geografis adalah informasi tentang letak permukaan bumi.
2. Teknologi informasi geografis meliputi *Global Positioning System* (GPS), remote sensing dan Sistem Informasi Geografis.
3. Sistem Informasi Geografis adalah sistem komputer dan perangkat lunak (software).
4. Sistem informasi geografis digunakan dalam berbagai aplikasi.
5. Komputasi geografis adalah ilmu di balik Teknologi Sistem Informasi Geografis.

Perkembangan SIG sangat erat kaitannya dengan perkembangan teknologi informasi dan pemrograman yang berkaitan langsung dengan pengolahan data spasial. Fungsi dasar dari GIS adalah sebagai berikut (Nirwansyah, 2017):

1. Input data untuk mengubah format data grafik menjadi data digital dalam format yang digunakan oleh SIG.
2. Manajemen data (manajemen), yang menyimpan data yang dimasukkan dan dapat dipanggil kembali jika diperlukan.
3. Mengelola dan menganalisis data yang ada untuk mendapatkan informasi yang lebih dalam dan lengkap dari SIG ini.
4. Output data untuk memperoleh informasi dari SIG hasil pengolahan SIG.

Untuk dapat mewujudkan beberapa fungsi di atas, SIG memiliki beberapa komponen yang saling terkait untuk menyajikan informasi spasial, yaitu (Nirwansyah, 2017) :

1. Hardware

Terdiri dari beberapa macam perangkat keras, seperti komputer, GPS, printer, plotter, scanner, digitizer, dan lain- lain. Sebagai media pengolahan GIS nantinya.

2. Software

Sekumpulan program yang digunakan untuk mempermudah dalam pembuatan dan pengolahan data, menyimpan, editing, layout dan juga analisis keruangan.

3. Brainware

Brainware bisa dikatakan juga dengan sumberdaya manusia atau manusia yang mengoperasikan hardware dan software untuk mengolah data terkait dengan GIS.

4. Data Spasial

Data atau informasi spasial merupakan data realitas yang akan diolah menjadi suatu sistem informasi berbasis keruangan.

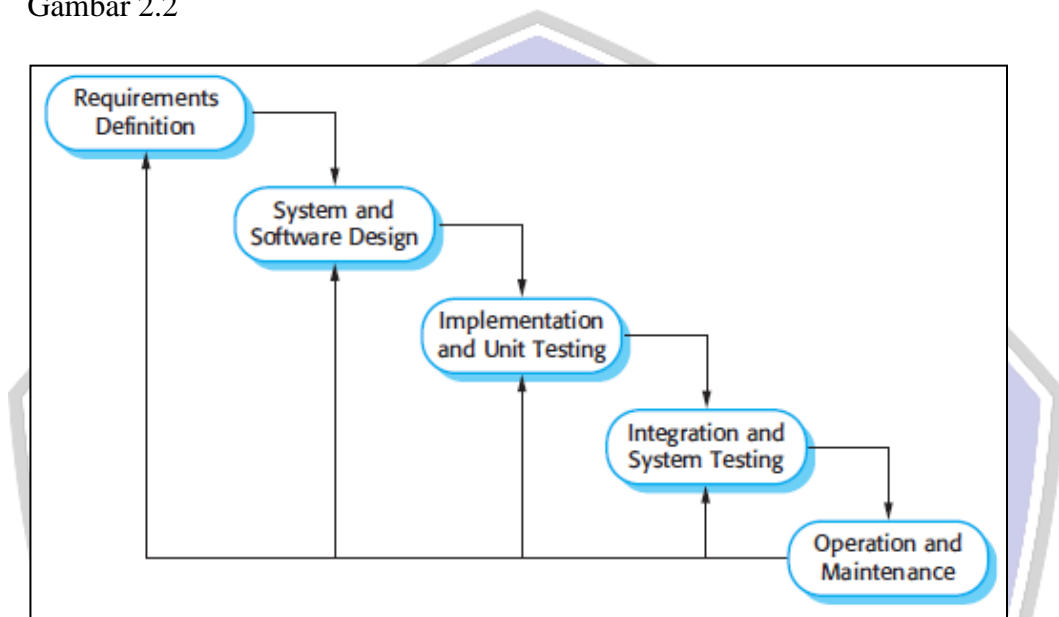
5. Metode

Metode digunakan sebagai analisis GIS untuk pengembang dapat memperoleh informasi yang relevan dengan pemangku kepentingan.

2.4 Waterfall SDLC

Pemilihan model pengembangan perangkat lunak didasarkan pada hasil analisis masalah yang ditemukan. Gunakan model *waterfall* ketika persyaratan yang ada sudah jelas dan dapat dipahami, sehingga pengembangan dapat dilakukan dalam beberapa langkah..

Model *waterfall* diselesaikan dengan menyelesaikan satu tahap kemudian dilanjutkan ke tahap berikutnya. Selama proses implementasi akan dihasilkan dokumen-dokumen yang telah disetujui pada setiap tahapan, sehingga apabila tahapan sebelumnya tidak disetujui maka tidak akan dilakukan tahapan selanjutnya, jika diperlukan modifikasi akan dilakukan setelah semua tahapan selesai. (Schach, 2011). Tahapan dalam model waterfall ini digambarkan pada Gambar 2.2



Gambar : 2. 2 SDLC Waterfall

Berikut merupakan penjelasan tahapan dalam pengembangan perangkat lunak model *waterfall* (Sommerville, 2016):

1. *Requirement Definition* Ini adalah proses analisis mendefinisikan persyaratan perangkat lunak, yang berasal dari negosiasi dengan peserta sistem. Pada tahap ini, batasan dan persyaratan didefinisikan, dan batasan dan persyaratan ini telah didefinisikan dan dijelaskan secara rinci dalam lembar spesifikasi.
2. *System and Software Design* merupakan proses perancangan perangkat lunak sesuai dengan metode pendekatan yang digunakan, bisa berupa pendekatan terstruktur ataupun pendekatan berorientasi objek. Untuk selanjutnya dilakukan perancangan perangkat lunak berupa perancangan database, perancangan pseudocode dan perancangan antarmuka.
3. *Implementation and Unit Testing* Merupakan proses pengembangan perangkat lunak yang sebelumnya dilakukan dalam bentuk bahasa

pemrograman. Hal ini juga dilakukan untuk melihat apakah setiap bagian dari program memenuhi spesifikasinya.

4. *Integration and System Testing* adalah proses mengintegrasikan dan menguji beberapa unit perangkat lunak yang direalisasikan ke dalam satu sistem. Jika persyaratan perangkat lunak terpenuhi, akhirnya ditentukan bahwa.
5. *Operation and Maintenance* adalah proses perangkat lunak setelah pengguna menggunakannya. Tidak ada bug yang ditemukan selama pengembangan.

2.5 Pemodelan Sistem

2.5.1 *Sequence Diagram*

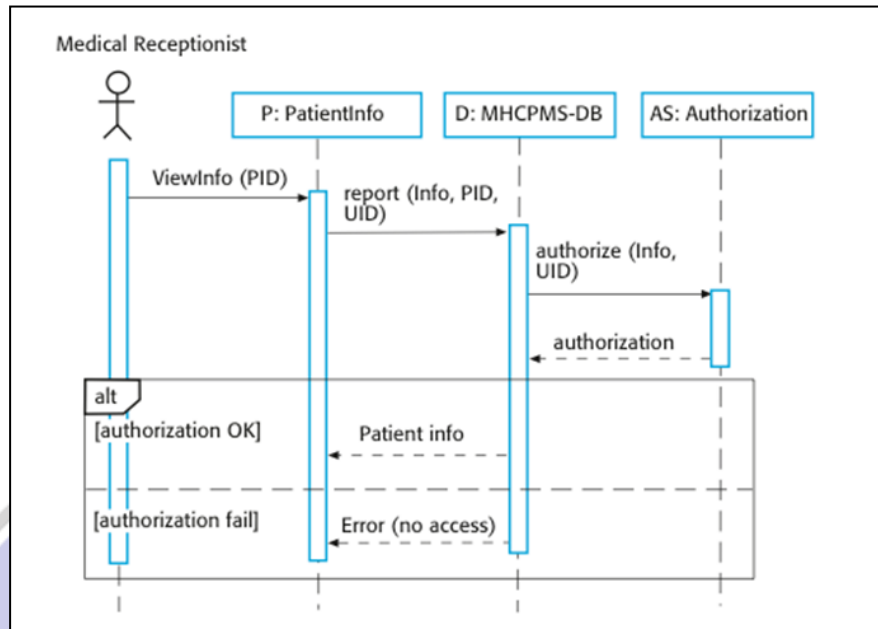
Sequence Diagram menggambarkan interaksi antara sistem dan objek sekitarnya (termasuk pengguna, layar, dll.) berdasarkan pesan yang dibuat dari waktu ke waktu. (Yasin, 2012). Digunakan sebagai gambaran skenario atau urutan langkah-langkah yang terjadi sebagai respon atau pemicu respon sehingga dapat dihasilkan beberapa output. *Sequence diagram* terdiri dari beberapa komponen yaitu:

1. Object merupakan komponen berbentuk kotak dari kelas atau objek. Sebagai gambaran bagaimana objek berinteraksi dengan sistem.
2. Activation merupakan sebuah komponen persegi panjang yang mewakili waktu yang dibutuhkan untuk menyelesaikan tugas objek. Semakin lama waktu aktivasi, semakin lama waktu yang dibutuhkan.
3. Aktors merupakan komponen berbentuk stick figure. Merepresentasikan user yang nantinya akan berinteraksi langsung dengan sistem.
4. Lifeline merupakan komponen dengan garis putus-putus. Berisi bidang dengan nama objek. Fungsinya adalah aktivitas objek.

2.5.2 *Class Diagram*

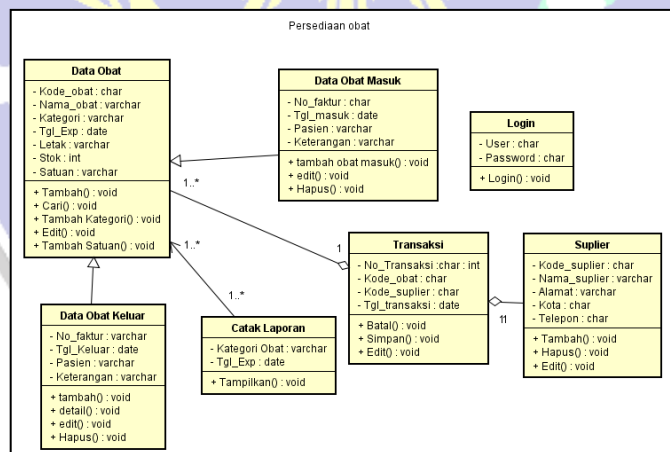
Class Diagram adalah spesifikasi yang membuat objek ketika dipakai dan merupakan inti dari desain dan pengembangan berorientasi objek. (Yasin, 2012). Menggambarkan struktur dan deskripsi kelas, paket, dan objek, serta hubungan di antara mereka, seperti pewarisan, asosiasi, dll. Keadaan sistem (atribut/atribut) dan operasinya dijelaskan oleh kelas. Kelas itu sendiri terdiri dari tiga bagian,

yaitu nama, atribut dan metode. Pada saat yang sama, properti dan metode memiliki properti saat dipanggil, sebagai berikut:



Gambar : 2. 3 Contoh *Sequence Diagram*

1. *Private*, tidak bisa dipanggil diluar class yang bersangkutan
2. *Protected*, hanya bisa dipanggil oleh class yang bersangkutan dan struktur class yang mewarisinya
3. *Public*, bisa dipanggil didalam maupun diluar class



Gambar : 2. 4 Contoh *Class Diagram*

Sebuah class merupakan representasi dari sebuah interface. Antar kelas bisa saling berinteraksi satu sama lain. Hubungan antar class bisa dijabarkan sebagai:

1. Asosiasi, yang berarti sebuah class memiliki atribut dari class lain.

2. Agregasi, yang berarti sebuah class terdiri atas class lain.
3. Pewarisan, yang berarti sebuah class diturunkan dari class lain dan mewarisi semua atribut dan method dari kelas tersebut.
4. Dinamis, yang berarti sebuah class terhubung atau mem-passing suatu pesan ke class lain.

2.6 Framework Laravel

Kerangka kerja Laravel adalah kerangka kerja aplikasi web yang menyederhanakan proses pengembangan dengan menyederhanakan tugas-tugas yang berulang dari aplikasi web (Armel, 2014) . Laravel dapat mengakomodir beberapa task, seperti penyajian web, pengelolaan basis data dan HTML. Ketika framework PHP lain membutuhkan konfigurasi yang cukup rumit sebelum memulai project, Laravel hanya memerlukan beberapa baris kode hingga siap dalam proses pengerjaan. Laravel adalah *end-user-centric*, yang berarti berfokus pada kejelasan dan kesederhanaan penulisan dan penelusuran, dan memiliki fitur aplikasi Web yang berfungsi seperti yang diharapkan. (Yasin, 2012). Beberapa keuntungan yang bisa didapatkan ketika menggunakan framework Laravel, yaitu:

1. *Website* menjadi lebih mudah untuk dikembangkan (*scalable*),
2. Menggunakan ruang nama dan layar untuk membantu mengatur dan mengelola sumber daya situs web Anda. ,
3. Pengembangan lebih cepat dan menghemat waktu karena dapat digunakan bersama dengan berbagai komponen kerangka pengembangan situs web lainnya..

Selain itu Laravel juga memiliki template layout yang ringan yang dapat ditambahkan css, gambar dan teks dengan struktur kode yang lengkap. Dengan library Object oriented yang sangat banyak, salah satunya dapat memudahkan dalam penggunaan fitur autentikasi. Laravel mendukung perancangan dengan ECB (*Entity Control Boundary*), yang berarti memisahkan antara logika dengan tampilan, dengan dokumentasi yang lengkap. Pembagian modul pada Laravel yang dibagi menjadi beberapa modul individu membuat aplikasi menjadi lebih informatif, modular dan responsive.

2.7 Pengujian Sistem

2.7.1 Pengujian *White-Box*

Pengujian *white-box* juga dikenal sebagai pengujian struktural, adalah metode pengujian yang berasal dari struktur perangkat lunak dan pengetahuan implementasi. Pengujian struktural berfokus pada unit terkecil dari program yang terkait dengan objek. Penguji dapat menganalisis kode program untuk mendefinisikan kasus uji. Analisis kode digunakan untuk memastikan setidaknya terjadi satu kali eksekusi untuk setiap komponen yang diuji.

Pengujian *white-box* kali ini menggunakan metode basic path testing, yang bertujuan untuk menganalisis jalur independent yang melalui sebuah komponen atau program (Sommerville, 2016). Setiap jalur dieksekusi, setidaknya semua komponen dieksekusi. Semua kalimat diuji kebenaran dan kepalsuannya. Pengujian dilakukan dengan menguji metode-metode yang berhubungan dengan objek. Jumlah jalur melalui program biasanya sebanding dengan ukurannya. Jalur pengujian disebut juga *flowgraph*. Graph terdiri dari node yang merepresentasikan logika dari program, dan edge yang merepresentasikan aliran kontrol. Jumlah jalur independent pada program dapat dihitung dengan kompleksitas cyclomatic yang dijabarkan pada persamaan :

$$CC(G) = \text{Jumlah (edge)} - \text{Jumlah (node)} + 2 \quad (2.1)$$

Setelah dilakukan perhitungan maka bisa dilakukan perancangan kasus uji, jumlah minimum kasus uji yang dibutuhkan adalah sama dengan jumlah kompleksitas cyclomatic.

2.8 Metode *Haversine*

Algoritma *Haversine* Formula adalah sebuah persamaan yang penting dalam bidang navigasi, untuk mencari jarak busur antara dua titik pada bola. Nantinya rumus haversine ini akan menghasilkan jarak terpendek antara dua titik, misalnya pada bola yang diambil dari garis bujur (*longitude*) dan garis lintang (*latitude*).

Dengan mengamsusikan bahwa bumi berbentuk sebuah lingkaran atau bulat sempurna dengan jari-jari R 6.367, 45 km dan lokasi dari 2 titik dikoordinat bola (lintang dan bujur) masing masing dengan long1, lat1 dan long2, lat2 maka rumus dapat dituliskan dengan persamaan sebagai berikut:

$$\Delta lat = lat2 - lat1$$

$$\Delta long = long2 - long1$$

$$A = \sin^2(\Delta lat/2) + \cos(lat1) \cdot \cos(lat2) \cdot \sin^2(\Delta long/2)$$

$$C = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$D = R \cdot c$$

$$D = \text{acos}(\sin(lat1) \cdot \sin(lat2) + \cos(lat1) \cdot \cos(lat2) \cdot \cos(long2 - long1)) \cdot R$$

Keterangan :

lat1 : Lintang lokasi awal

lat2 : Lintang lokasi tujuan

long1 : Bujur lokasi awal

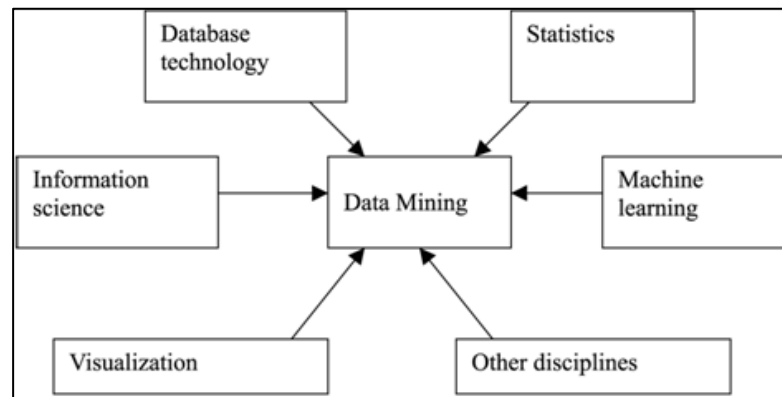
long2 : Bujur lokasi tujuan

R : Radian bumi (6371km)

D : Jarak dalam km

2.9 Data Mining

Data mining, juga dikenal sebagai *Knowledge Discovery in Database* (KDD), adalah aktivitas yang mengumpulkan data historis dan menggunakannya untuk menemukan hukum, pola, atau hubungan dalam kumpulan data yang besar. Hasil dari data mining ini dapat digunakan untuk memperbaiki solusi di masa mendatang. Oleh karena itu, istilah "pengenalan pola" jarang digunakan saat ini karena merupakan bagian dari data mining (Ediyanto, 2019).



Gambar : 2. 5 Data mining merupakan irisan dari berbagai disiplin

(Suharti, 2017)

Concept/Class Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Prediction, Cluster Analysis(Outlier analysis, dan Evolution analysis) ini semua merupakan fungsionalitas dari data mining

Dapat disimpulkan bahwa data mining merupakan kombinasi antara teori dan metode heuristik, yang berfokus pada keseluruhan proses penemuan pola/pengetahuan, termasuk pembersihan data, pelatihan, dan visualisasi hasil.

2.10 *Unsupervised Learning*

Pengenalan pola adalah bidang pembelajaran mesin, yang dapat didefinisikan sebagai "proses mengekstraksi data mentah dan mengambil tindakan berdasarkan klasifikasinya". Jadi ini adalah seperangkat aturan pembelajaran yang diawasi, dan ada beberapa definisi lain untuk pengenalan pola(Achsanisa, 2017).

Pengenalan pola adalah bagian dari penambangan data. Oleh karena itu, pengenalan pola adalah topik yang mempelajari bagaimana kita membagi objek ke dalam kategori yang berbeda dan bagaimana menemukan tren dalam data. Yang pertama terkait dengan klasifikasi kasus, dan yang kedua terkait dengan regresi.

Unsupervised learning adalah algoritma yang digunakan untuk menarik kesimpulan dari kumpulan data. Metode ini hanya memeriksa data berdasarkan kedekatan, atau data yang biasa disebut dengan *cluster*. Metode pembelajaran

tanpa pengawasan yang paling umum adalah analisis kelompok, yang digunakan saat menganalisis data untuk menemukan pola tersembunyi. Atau cluster dalam data (Rezika & Davita, 2020).

Dalam unsupervised learning kita harus mengelompokkan data ke dalam sejumlah kelas. Hal ini terutama karena data yang ada tidak ditandai. Label menunjukkan di mana data dikelompokkan. Untuk menyelesaikan tugas ini, kita dapat menggunakan metode pembelajaran tanpa pengawasan, termasuk grouping and self-organizing card (SOM) dalam kelompok ini (Ediyanto, 2019).

2.11 *Clustering*

Clustering adalah teknik terkenal yang telah banyak digunakan dalam data mining. *Clustering* adalah metode pengumpulan data. *Clustering* adalah proses pengelompokan data menjadi beberapa cluster atau kelompok sehingga data dalam cluster paling mirip, dan data antar cluster paling sedikit. (Irwansyah, 2014).

Menurut Han dan Kamber, 2012, syarat yang harus dipenuhi oleh suatu algoritma *clustering* adalah:

1. Skalabilitas

Metode *clustering* harus mampu menangani data dalam jumlah besar. Saat ini, sejumlah besar data sangat umum di berbagai bidang seperti database. Tidak hanya berisi ratusan objek, database yang sangat besar berisi jutaan objek.

2. Kemampuan analisa beragam bentuk data

Algoritma *clustering* diterapkan pada berbagai bentuk data, seperti data nominal, data terurut, dan data agregat.

3. Menemukan cluster dengan bentuk yang tidak terduga

Banyak algoritma *clustering* menggunakan Euclidean Get atau Manhattan metode. Hasilnya bulat. Hasil pengelompokan mungkin aneh dan berbeda satu

sama lain. Oleh karena itu, harus dimungkinkan untuk menganalisis semua jenis pengelompokan dalam algoritma pengelompokan.

4. Kemampuan untuk dapat menangani noise

Data tidak selalu dalam kondisi baik. Terkadang data mungkin rusak, salah paham, atau hilang. Karena sistem ini membutuhkan algoritma *clustering* untuk mengolah data, maka satu set data dapat berisi banyak dimensi atau atribut. Untuk itu diperlukan suatu algoritma *clustering* yang dapat menangani data multidimensi.

5. Sensitifitas terhadap perubahan input

Perubahan atau penambahan data pada hari pemasukan dapat menyebabkan perubahan pada cluster yang ada. Jika menggunakan algoritma *clustering* dengan tingkat sensitivitas yang lebih rendah, maka digunakan algoritma *clustering* dengan tingkat sensitivitas yang lebih rendah. bahkan dapat menyebabkan perubahan yang nyata Perubahan.

6. Mampu melakukan *clustering* untuk data dimensi tinggi

kelompok data dapat berisi beberapa dimensi atau atribut. Untuk tujuan ini, diperlukan algoritma pengelompokan yang dapat menangani sejumlah besar data multi-dimensi.

7. Interpretasi dan kegunaan

Hasil *clustering* harus dapat diinterpretasi dan bermanfaat.

2.12 *K-Means Clustering*

Dua jenis data *clustering* yang sering digunakan dalam proses pengelompokan data yaitu "*hierarchical*" dan "*non-hierarchical*", dan *K-Means* merupakan salah satu metode *clustering* data non-hierarchical atau *partitioned clustering*. *K-Means* merupakan salah satu algoritma klastering dengan metode partisi (*partitioning method*) yang berbasis titik pusat (*centroid*) selain algoritma *k-Medoids* yang berbasis obyek (Irwansyah, 2014).

Algoritma KMeans mendefinisikan center of gravity cluster sebagai pusat cluster itu sendiri, dan algoritma KMeans menentukan center of gravity atau pusat cluster sebagai pusat cluster. Gunakan algoritma KMeans, jika diberikan sekumpulan data $X = \{x_1, x_2, \dots, x_n\}$ dimana $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ adalah system dalam ruang real R^n , maka algoritma *K-Means* akan menyusun partisi X dalam sejumlah k cluster (a priori)(Irwansyah, 2014). Jadi setiap cluster memiliki centroid yang merupakan nilai means dari dalam klaster.

Tahap awal dari algoritma *K-Means* adalah memilih k objek secara acak sebagai pusat centroid dalam data. Kemudian, gunakan jarak Euclidean untuk menghitung jarak antara benda dan pusat centroid. Algoritma *K-Means* secara berulang meningkatkan nilai, yang dihitung dari titik pusat cluster, dan objek berikutnya di setiap cluster ditempatkan di setiap cluster di grup terdekat. Ketika semua data dimasukkan ke grup berikutnya, titik pusat baru ditentukan. Ulangi proses penentuan titik pusat dan memasukkan data ke dalam kelompok sampai nilai rata-rata semua kelompok yang terbentuk tidak berubah lagi.

Dalam menentukan nilai centroid untuk awal iterasi, nilai awal dilakukan secara acak, maka digunakan rumus sebagai berikut :

$$v_{ij} = \frac{1}{N_i} + \sum_{k=0}^{N_1} x_{kj} \quad (2.2)$$

dimana :

v_{ij} : adalah centroid cluster ke-1 untuk variabel ke-j

N_1 : adalah jumlah data yang menjadi anggota cluster ke-i

i, k : adalah indeks dari cluster

j : adalah indeks dari variabel

x_{kj} : adalah nilai data ke-k yang ada didalam cluster untuk variabel ke-j

Untuk menghitung jarak semua data ke titik centroid dapat menggunakan teori jarak euclidean, sebagai berikut :

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + (X_{ki} - X_{kj})^2} \quad (2.3)$$

Dimana :

$D(i, j)$: Jarak data ke i ke pusat cluster j

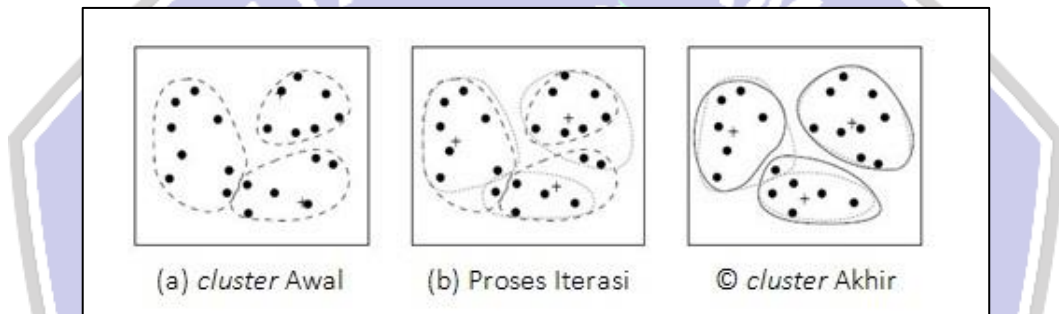
X_{ki} : Data ke i pada atribut data ke k

X_{kj} : Titik pusat ke j pada atribut ke k

Sedangkan rumus pengukuran jarak metode euclidean distance sebagai berikut :

$$j(v_1 + v_2) = \sqrt{\sum_{k=1}^n (v_1(k) + v_2(k))^2} \quad (2.4)$$

v_1 dan v_2 merupakan dua vektor yang nanti jaraknya akan dihitung dan N menyatakan panjang vektor itu sendiri

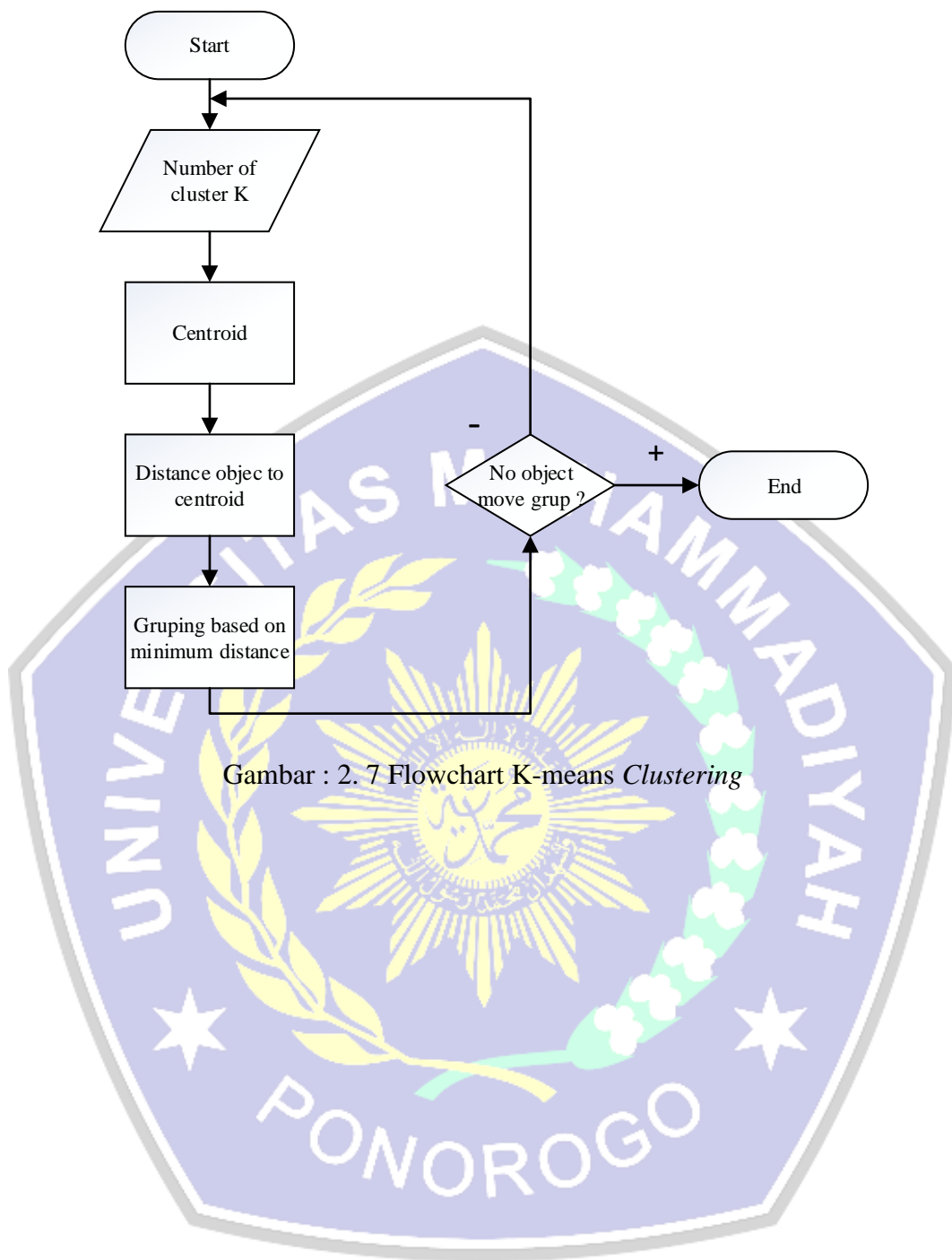


Gambar : 2. 6 Proses *Clustering* Obyek Menggunakan metode *K-Means*

(Irwansyah, 2014)

Langkah-langkah *clustering* menggunakan *K-Means* yaitu :

1. Pilih jumlah cluster K .
2. Tentukan titik awal cluster K secara acak.
3. Alokasi semua data/objek ke cluster terdekat.
4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang.
5. Tugaskan lagi setiap objek memakai pusat cluster yang baru.



Gambar : 2. 7 Flowchart K-means *Clustering*