

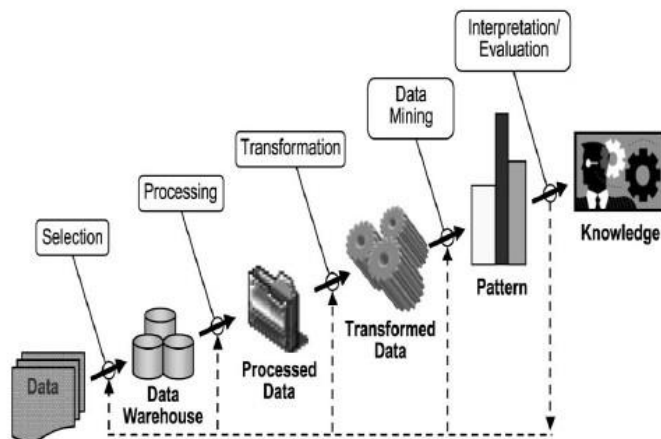
BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Pengertian Data Mining

Dalam proses menemukan informasi yang menarik terdapat sebuah data yang akan diolah dengan menggunakan metode tertentu. Menurut (Arhami and Nasir, 2020, p. 2-5) *data mining* merupakan proses untuk menganalisa sebuah pola pada data menjadi sebuah informasi yang berguna sehingga dapat membantu dalam membuat keputusan yang lebih baik. *Data mining* memiliki tujuan untuk menemukan sebuah pola yang memiliki arti dan maksud tertentu yang dapat digunakan untuk menggali dan menganalisis dari sebagian besar data yang ada.

Istilah *data mining* digunakan sebagai proses penguraian penemuan informasi dalam sebuah data. (Putra et al., 2023, p. 30) mengutarakan bahwa *data mining* memiliki hubungan di berbagai bidang keilmuan yang dapat membantu untuk mengidentifikasi pola ataupun hubungan tersembunyi dari sebuah data yang dapat mengungkapkan sebuah informasi berharga dari data. Proses yang terjadi seperti pengambilan informasi, analisa pola dan ringkasan pengetahuan yang memiliki pemahaman mengenai data serta menggiring pengukuran secara konstruktif dari area yang terlibat.

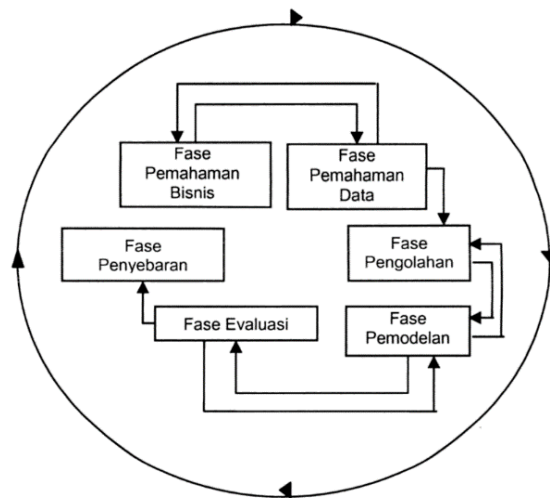


Gambar 2. 1 Tahapan Proses KDD

Data mining termasuk kedalam salah satu rangkaian *Knowledge Discovery in Database (KDD)*. Menurut (Amna et al., 2023, p. 10) *Knowledge Discovery in Database (KDD)* merupakan tahap penggalian dan menganalisa guna mendapatkan suatu informasi yang berguna dari sejumlah data. *Data mining* dapat digunakan untuk mendapatkan informasi pada bisnis dan analisis data. Gambar 2. 1 menunjukkan proses tahapan data mining menurut

Fayyad, Shapiro, dan Smith pada tahun 1996 dalam (Jollyta, Ramdhan and Zarlis, 2020, p. 48-49) sebagai berikut:

- (1) *selection*, tahapan awal dengan memilih atau menyeleksi beberapa data yang dibutuhkan dari sekumpulan data untuk diseleksi kembali pada tahap selanjutnya;
- (2) *preprocessing*, tahap ini dilakukan pembersihan data pada field yang dianggap tidak memadai untuk menjamin format data konsisten;
- (3) *transformation*, pada tahap ini data akan disesuaikan dengan teknik *data mining* yang digunakan agar kualitas data tepat dan sesuai untuk digunakan pada tahap selanjutnya;
- (4) *data mining*, tahap ini menggunakan sebuah teknik atau metode khusus untuk menganalisis data guna mendapatkan hasil yang berisi informasi penting yang tersembunyi dalam data;
- (5) *interpretation & evaluation*, tahap dimana menginterpretasikan pola menjadi suatu pengetahuan yang dapat digunakan untuk membantu mendukung pengambilan keputusan.



Gambar 2. 2 Proses Data Mining Pada CRISP-DM

Gambar 2. 2 menunjukkan proses data *mining* pada CRISP-DM atau *Cross-Industry Standard Process for Data Mining* yang digunakan sebagai strategi dalam pemecahan masalah pada bisnis maupun penelitian dengan menyediakan sebuah proses standar dari data *mining*. Menurut (Kusrini and Luthfi, 2009, p. 8) data *mining* memiliki enam fase siklus hidup dalam CRISP-DM, dimana pada fase siklus hidup tersebut memiliki sifat adaptif secara berurutan yang digambarkan dengan panah untuk menunjukkan hubungan

penting pada fase siklus hidup. Adapun fase CRISP-DM menurut Larose dalam (Kusrini and Luthfi, 2009, p. 9) sebagai berikut :

- (1) fase pemahaman bisnis (*Business Understanding Phase*); fase ini terdiri dari penentuan tujuan proyek dan kebutuhan secara spesifik secara keseluruhan, lalu menerjemahkan tujuan dan batasan menjadi sebuah acuan dari permasalahan dan pada akhirnya menyiapkan strategi awal untuk mencapai awal tujuan;
- (2) fase pemahaman bisnis (*Business Understanding Phase*); fase ini dimulai dari mengumpulkan data, lalu melakukan analisis pada data dan mengevaluasi kualitas pada data;
- (3) fase persiapan data (*Data Preparation Phase*); fase ini dimulai dari menyiapkan data, memilih variabel yang akan digunakan dan mengubahnya jika ada beberapa variabel yang diperlukan;
- (4) fase pemodelan (*Modeling Phase*); fase ini menerapkan teknik data *mining* yang sesuai dengan permasalahan yang diambil;
- (5) fase evaluasi (*Evaluation Phase*); fase ini dilakukan evaluasi kembali model yang dipilih agar permasalahan yang terdapat terpecahkan dengan baik;
- (6) fase penyebaran (*Deployment Phase*); fase ini menyebarkan model yang telah dihasilkan dari teknik data *mining* yang berupa laporan.

Data *mining* memiliki beberapa metode diantaranya klasifikasi, *clustering*, asosiasi, estimasi, prediksi dan regresi. Pada penelitian ini menggunakan metode *clustering*. (Arhami and Nasir, 2020, p. 149) mengutarakan bahwa *clustering* merupakan data atau nilai yang belum memiliki label pada kelasnya perlu diprediksi ke dalam sebuah kelas untuk mengetahui objek tersebut akan dimasukan pada kelas yang sesuai berdasarkan kesamaan pola atau karakteristik pada kelompoknya. Salah satu metode pada data mining yaitu pengelompokan data ke dalam kelompok berdasarkan dengan karakteristik tertentu.

Menurut (Putra *et al.*, 2023, p. 86) *clustering* termasuk pada *unsupervised learning* atau sebuah metode pembelajaran tanpa adanya pengawasan, dimaksudkan tidak memiliki ketergantungan pada data yang mempunyai label untuk dapat mengolahnya. *Clustering* digunakan dalam menganalisis data dengan melibatkan data yang sudah dikelompokkan berdasarkan karakteristik dari data tersebut, dapat pula menemukan sebuah pola terstruktur yang terdapat dalam data, sama halnya dengan mengidentifikasi sebuah pengelompokan alami (Putra *et al.*, 2023, p. 87). Pada penelitian ini dalam

memetakan penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan digunakan model *clustering* dengan pendekatan partisi *clustering*.

2. Algoritma K-Means

Dalam penelitian ini, pemetaan penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan diperlukannya sebuah algoritma. Menurut (Santosa, 2007, p. 42) K-Means merupakan sebuah teknik klastering yang umum dikenal dan paling sederhana, dengan K-Means dapat mengelompokkan suatu objek kedalam sebuah klaster. (Prasetyo, 2014, p. 189) mengutarakan bahwa algoritma K-Means sederhana untuk diimplementasikan, mudah dipahami, umum dalam penggunaannya dan relatif cepat. Dengan algoritma K-Means dapat dilakukan pembagian data ke dalam sebuah kelompok yang memiliki karakteristik yang serupa dikelompokkan ke dalam satu set yang serupa, sebaliknya jika memiliki karakteristik berbeda, maka akan dikelompokkan ke dalam set yang berbeda pula, hal ini dilakukan untuk meminimalkan nilai variatif pada suatu kelompok (Putra *et al.*, 2023, p. 87-88).

Menurut (Arhami and Nasir, 2020, p. 148-149) berikut ini langkah-langkah dasar pada algoritma K-means yaitu:

- (1) tentukan nilai k klaster sesuai dengan yang diinginkan;
- (2) pilih titik-titik atau sampel yang menjadi anggota klaster secara acak;
- (3) tentukan nilai centroid atau titik tengah dari klaster tersebut dengan rumus;

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} X_{ik}$$

- (4) hitung *square error* untuk setiap klaster C_k yang merupakan jumlah kuadrat dari jarak Euclidean antara tiap sampel dalam C_k dan titik tengahnya (*centroid*), error dikenal juga dengan nama *within cluster variation (WCV)*, yaitu;

$$e_k^2 = \sum_{i=1}^{n_k} (X_{ik} - M_k)^2$$

- (5) selanjutnya jumlah dari keseluruhan error dari *k-cluster* juga dihitung dengan rumus;

$$E_k^2 = \sum_{k=1}^k e_k^2$$

- (6) kelompokkan kembali semua sampel berdasarkan jarak minimum dari masing-masing pusat M_1, M_2, \dots, M_k sehingga diperoleh distribusi baru dari sampel sesuai kelasternya, untuk memperoleh distribusi baru dari sampel baru tersebut dapat dilakukan dengan menghitung jarak masing-masing titik pusat dengan keseluruhan sampel $d(M_1, x_1) \dots d(M_k, x_k)$; perhitungan jarak dari masing-masing titik tersebut dapat menggunakan beberapa metode, contohnya *Euclidean Distance*, dimana jarak antara dua titik dalam satu, dua, tiga atau sampai dimensi n dapat dihitung sebagai berikut;

$$d(p, q) = \sqrt{((p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2)}$$

- (7) tuliskan hasil anggota klaster baru sesuai dengan hasil yang diperoleh pada langkah ke-5 dan ulangi langkah ke-3 sampai beberapa iterasi yang nantinya akan ditemukan nilai total *square error* turun secara signifikan.

Menurut (Arhami and Nasir, 2020, p. 149-152), berikut ini perhitungan K-means yang dapat dipaparkan dengan data contoh sebagai berikut;

x_k	p	q
x_1	0	2
x_2	0	0
x_3	1.5	0
x_4	5	0
x_5	5	2

- (a) misal pada kasus seperti ini, nilai K yang diambil yaitu $K = 2$;
 (b) klaster diambil secara acak yaitu sebagai berikut;

$$C_1 = \{x_1, x_2, x_4\}$$

$$C_2 = \{x_3, x_5\}$$

- (c) pada langkah ke-2 dapat ditentukan titik tengah dari kedua klaster tersebut, yaitu;

$$M_1 = \left\{ \left(\frac{0 + 0 + 5}{3} \right), \left(\frac{2 + 0 + 0}{3} \right) \right\} = \{1.66, 0.66\}$$

$$M_2 = \left\{ \left(\frac{1.5 + 5}{2} \right), \left(\frac{0 + 2}{2} \right) \right\} = \{3.25, 1.00\}$$

- (d) *square error* diperoleh;

$$e_1^2 = [(0 - 1.66)^2 + (2 - 1.66)^2] + [(0 - 1.66)^2 + (0 - 1.66)^2] + [(5 - 1.66)^2 + (0 - 1.66)^2] = 19.36$$

$$e_2^2 = [(1.5 - 3.25)^2 + (0 - 1)^2] + [(5 - 3.25)^2 + (2 - 1)^2] = 8.12$$

sehingga diperoleh *square error* yaitu;

$$E^2 = e_1^2 + e_2^2 = 19.36 + 8.12 = 27.48$$

- (e) redistribusi sampel ke dalam kelas yang sesuai dengan hasil perhitungan jarak yang dimiliki oleh tiap sampel, dimana jarak minimum dari titik pusat kluster yang ada akan menjadi kluster baru bagi sampel tersebut; berikut hasil yang diperoleh;

$$d(M_1, x_1) = \sqrt{((1.66 - 0)^2 + (0.66 - 2)^2)} = 2.14$$

$$d(M_2, x_1) = \sqrt{((3.25 - 0)^2 + (1 - 2)^2)} = 3.40$$

$$d(M_1, x_2) = \sqrt{((1.66 - 0)^2 + (0.66 - 0)^2)} = 1.79$$

$$d(M_2, x_2) = \sqrt{((3.25 - 0)^2 + (1 - 2)^2)} = 3.40$$

$$d(M_1, x_3) = \sqrt{((1.66 - 1.5)^2 + (0.66 - 0)^2)} = 0.83$$

$$d(M_2, x_3) = \sqrt{((3.25 - 1.5)^2 + (1 - 0)^2)} = 2.01$$

$$d(M_1, x_4) = \sqrt{((1.66 - 5)^2 + (0.66 - 0)^2)} = 3.41$$

$$d(M_2, x_4) = \sqrt{((3.25 - 5)^2 + (1 - 0)^2)} = 2.01$$

$$d(M_1, x_5) = \sqrt{((1.66 - 5)^2 + (0.66 - 2)^2)} = 3.60$$

$$d(M_2, x_5) = \sqrt{((3.25 - 5)^2 + (1 - 2)^2)} = 2.01$$

berikut hasil perhitungan yang telah dilakukan dan jarak minimum dari titik tengah kluster ke sampel menjadi kelas sampel;

X_k	$d(M_1, x_k)$	$d(M_2, x_k)$	Kluster
X_1	2.14	3.40	C_1
X_2	1.79	3.40	C_1
X_3	0.83	2.01	C_1
X_4	3.41	2.01	C_2
X_5	3.60	2.01	C_2

sehingga kluster baru dapat dituliskan $C_1 = \{x_1, x_2, x_3\}$, $C_2 = \{x_4, x_5\}$, pada kedua kluster dihitung kembali titik tengah dan total *square error*-nya untuk memastikan perlu dilanjutkan atau tidak ke iterasi berikutnya; titik tengah dari kedua kluster dapat dihitung sebagai berikut:

$$M_1 = \left\{ \left(\frac{0 + 0 + 1.5}{3} \right), \left(\frac{2 + 0 + 0}{3} \right) \right\} = \{0.5, 0.66\}$$

$$M_2 = \left\{ \left(\frac{5 + 5}{2} \right), \left(\frac{0 + 2}{2} \right) \right\} = \{5, 1.00\}$$

dan *square error*-nya yaitu; $e_1^2 = 4.17$ dan $e_2^2 = 2$ sehingga total yang diperoleh adalah;

$$E^2 = e_1^2 + e_2^2 = 4.17 + 2 = 6.17$$

jika dilihat dari total *square error*-nya yang turun begitu signifikan dari 27.48 ke 6.17, maka untuk kasus ini pada iterasi pertama sudah cukup untuk mengambil kesimpulan bahwa kluster yang dihasilkan adalah $C_1 = \{x_1, x_2, x_3\}$, $C_2 = \{x_4, x_5\}$ karena jika analisis dilanjutkan berdasarkan dengan titik tengah kluster baru maka kluster yang akan diperoleh juga akan sama seperti kluster yang dihasilkan pada iterasi pertama.

3. Sistem Pendukung Keputusan (SPK)

Menurut Wijoyo, H. dalam (Yuswardi *et al.*, 2022, p. 1) menyatakan bahwa pengambilan keputusan dilakukan dengan beberapa perhitungan dimana termasuk mengumpulkan informasi, menentukan tujuan, mengidentifikasi alternatif, menetapkan kriteria dan memilih pilihan terbaik yang tersedia. Menurut (Yuswardi *et al.*, 2022, p. 3) mengutarakan bahwa sistem pendukung keputusan merupakan suatu proses berpikir dan memutuskan berbagai pilihan dari pemilihan alternatif untuk meramalkan keadaan masa depan yang dibutuhkannya sebuah pemahaman teori atau metode mengenai pendukung keputusan. Metode K-Means digunakan pada penelitian ini sebagai sistem pendukung keputusan.

Sistem pendukung keputusan dapat meningkatkan efektivitas dan efisiensi para pengambil keputusan dalam membantu pengambilan keputusan (Setyaningsih, 2015, p. 12). Dalam pengambilan keputusan terdapat tahapan yang harus dilalui agar dapat menghasilkan sebuah keputusan terbaik. Adapun tahapan dalam pengambilan keputusan menurut (Hutahaeen *et al.*, 2023, p. 4-5):

- (1) *intelligence phase*, proses dalam menelusuri dan mengenali permasalahan;
- (2) *design phase*, proses pengembangan dalam mencari solusi atau alternatif yang mungkin dapat diambil dengan adanya verifikasi dan validasi untuk mengetahui tingkat akurasi model yang digunakan;
- (3) *choice phase*, dilakukan sebuah pemilihan dari solusi alternatif yang telah ada dengan memperhatikan kriteria yang tepat dengan tujuan utamanya;
- (4) *implementation phase*, penerapan dilakukan sesuai dengan rancangan yang telah dibuat sebelumnya.

Selain tahapan pada sistem pendukung keputusan terdapat pula komponen utama pada sistem pendukung keputusan. Menurut (Hutahaeen *et al.*, 2023, p. 5) dalam sistem pendukung keputusan terdapat tiga komponen utama yang tersusun dalam menentukan kemampuan sistem pendukung

keputusan. Berikut merupakan tiga komponen utama dalam sistem pendukung keputusan menurut Carter dalam (Setiyaningsih, 2015, p. 16-22):

- (1) subsistem data (*data subsystem*), merupakan komponen yang terdapat pada komponen sistem pendukung keputusan dalam menyediakan data yang dibutuhkan;
- (2) subsistem model (*model subsystem*), merupakan sebuah model yang digunakan dalam sistem pendukung keputusan untuk mengolah data sehingga menghasilkan suatu hasil yang dapat digunakan dalam pengambilan keputusan;
- (3) subsistem dialog (*user system interface*), pengimplementasian sistem pendukung keputusan yang menjadikan pengguna dapat berkomunikasi dengan sistem yang telah dirancang secara interaktif.

4. Pengertian Data

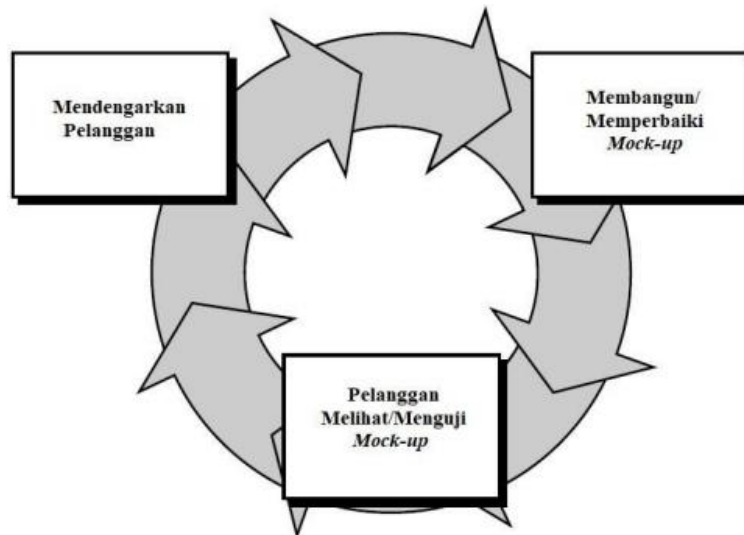
Komponen yang paling penting dalam sistem informasi yaitu data. Menurut (Yusuf and Daris, 2018, p. 16) data merupakan pemaparan atau keterangan yang belum memiliki arti sepenuhnya dari sebuah objek yang dapat berupa angka, suara, gambar, karakter maupun simbol. Pada penelitian ini menggunakan data yang berupa angka dari fakta kejadian yang sebenarnya terjadi. Data pada penelitian ini menggunakan file excel dalam pelaksanaannya. File excel merupakan buku kerja atau kertas kerja yang dapat diubah menjadi format data lain.

5. Pengertian System Development Life Cycle (SDLC)

Dalam hal ini penelitian yang dilakukan menggunakan sebuah model pengembangan agar proses yang terjadi dalam penelitian ini lebih terstruktur. Menurut (A.S. and Shalahuddin, 2013, p. 26) *System Development Life Cycle* (SDLC) merupakan suatu pengembangan pada sistem yang sudah ada sebelumnya dengan memiliki tahapan-tahapan yang harus dilalui agar terciptanya sebuah sistem yang berkualitas. Pada *System Development Life Cycle* (SDLC) terdapat beberapa metode, salah satu metode yang digunakan pada penelitian ini yaitu metode *prototyping*. Menurut (Khadir, 2014, p. 29) *prototyping* merupakan metode dalam pengembangan sebuah sistem yang dilakukannya pendekatan untuk menghasilkan spesifikasi yang terperinci sebelum pengguna melakukan evaluasi pada sistem tersebut.

Pada penelitian ini, *prototyping* digunakan untuk menjabarkan kebutuhan pengguna secara rinci. Menurut (A.S. and Shalahuddin, 2013, p 30-31) model *prototyping* dibutuhkan untuk memberi kejelasan mengenai spesifikasi kebutuhan yang diinginkan pada sebuah sistem dengan awal mula apa saja

yang dibutuhkan oleh user mengenai sistem yang dibutuhkan, *prototyping* menyediakan sebuah gambaran tampilan dari alur kerja sebuah sistem yang pada nantinya akan di evaluasi oleh pengguna hingga ditemukan spesifikasi yang sesuai dengan keinginan pengguna. Berikut ini model *prototyping* menurut (A.S. and Shalahuddin, 2013, p. 32) yang terdiri dari 3 tahapan ditunjukkan pada gambar 2. 3, yaitu;



Gambar 2. 3 Model *Prototyping*

Sumber: (A.S. and Shalahuddin, 2013, p. 32)

- (1) mendengarkan pelanggan untuk melakukan pengumpulan kebutuhan dari sebuah sistem bagaimana sistem yang sedang berjalan dengan melakukan analisis kebutuhan dari pengguna;
- (2) dilakukannya perancangan dan membangun *prototyping* sesuai dengan kebutuhan pengguna yang telah dijabarkan sebelumnya;
- (3) dilakukannya pengujian *prototyping* oleh pengguna kemudian pengguna memberikan evaluasi kekurangan dan kebutuhan apa saja yang dibutuhkan oleh pengguna sehingga pada saat pengujian masih terdapat kekurangan dan kebutuhan maka kembali kepada tahap mendengarkan pelanggan untuk memperbaiki *prototyping* yang sudah tersedia;

prototyping sebuah sistem yang dapat menyediakan dan mendemonstrasikan fungsi-fungsi yang terdapat serta mungkin dapat dilakukan pengujian desain sistem tersebut dapat dikatakan sebagai *mock-up*.

6. Pengertian *Unified Modeling Language* (UML)

Pada sebuah penelitian membutuhkan sebuah teknik pemodelan visual dalam menggambarkan sebuah sistem. (Hasanah and Untari, 2020, p. 64) menjelaskan bahwa *Unified Modelling Language* (UML) adalah bahasa yang digunakan untuk mendefinisikan kemampuan sebuah sistem, desain, analisis serta gambaran mengenai arsitektur pada sebuah sistem. Menurut (Suhendar and Gunadi, 2002, p. 34) *Unified Modelling Language* (UML) merupakan bahasa pemodelan dari sebuah sistem untuk menggambarkan, membangun dan mendokumentasikan sistem yang besar dan rumit yang terdapat notasi untuk membantu dalam menjelaskan secara visual mengenai komponen yang terdapat pada pemodelan.

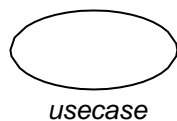
Pada *Unified Modelling Language* (UML) terdapat beberapa diagram yang dapat memodelkan sebuah sistem. Menurut (Henderi, 2007, p. 4), *Unified Modelling Language* (UML) merupakan pemodelan untuk memvisualisasikan, merancang ataupun mendokumentasikan sebuah sistem perangkat lunak pada proses bisnis guna memperluas tingkat fungsionalitas. Adapun diagram-diagram yang terdapat pada *Unified Modelling Language* (UML) sebagai berikut:

(a) *usecase* diagram;

usecase diagram mendeskripsikan sebuah pemodelan perlakuan pada sebuah sistem yang akan dibuat dimana perlakuan tersebut berupa cara pengguna dan sistem berinteraksi, sedangkan *usecase* naratif menjelaskan setiap prosedur interaksi yang dilakukan secara detail dan tersusun berupa tekstual, berikut simbol yang terdapat pada *usecase*;



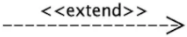

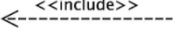
; menggambarkan sebuah interaksi seseorang dengan sistem yang biasanya dinyatakan menggunakan kata benda di awal frasa nama *actor*;



; menggambarkan mengenai sesuatu kegiatan yang dikerjakan oleh sistem yang dinyatakan dengan menggunakan kata kerja;


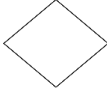
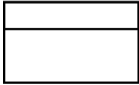

asosiasi/ association ;

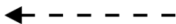
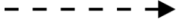

menggambarkan sebuah komunikasi antara *actor* dan *usecase*;

	;	<p>menggambarkan sebuah hubungan usecase tambahan ke sebuah usecase dimana usecase yang ditambahkan dapat berdiri sendiri walau tanpa usecase tambahan;</p>
 <p style="margin-left: 20px;">generalisasi / <i>generalization</i></p>	;	<p>menggambarkan tubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah usecase;</p>
	;	<p>menggambarkan hubungan usecase dengan usecase yang pada prosesnya harus dilaksanakan;</p>

(b) class diagram;



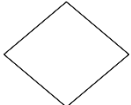
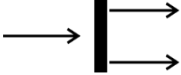


class diagram menjelaskan mengenai aliran data pada sebuah program yang dimana menggambarkan struktur sistem dengan mendefinisikan *class object* yang menyusun sistem dan hubungan yang terdapat pada *class object* tersebut, berikut merupakan simbol pada *class diagram*;

	;	<p><i>generalization</i> menggambarkan suatu hubungan dimana <i>descendent</i> berbagi perilaku dan struktur data dari objek yang ada di atasnya <i>ancestor</i>;</p>
	;	<p><i>nary association</i> merupakan upaya untuk menghindari asosiasi lebih dari 2 objek;</p>
	;	<p><i>class</i> menggambarkan himpunan dari objek-objek yang berbagi atribut serta operasi yang sama;</p>
	;	<p><i>collaboration</i> mendeskripsikan mengenai urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu <i>actor</i>;</p>

	; <i>realization</i> merupakan operasi yang benar-benar dilakukan oleh suatu objek;
	; <i>dependency</i> menggambarkan suatu hubungan perubahan yang terjadi pada suatu elemen mandiri yang akan mempengaruhi elemen yang bergantung pada elemen yang tidak mandiri;
	; <i>association</i> menghubungkan antara objek satu dengan objek lainnya;

(c) *activity* diagram;

activity diagram didefinisikan sebagai alur dari aktivitas yang terjadi pada sebuah sistem yang menggambarkan kegiatan yang akan dilakukan pada sistem atau program tersebut, berikut merupakan simbol pada *activity* diagram;

	; status awal atau <i>initial</i> menunjukkan awal dari suatu proses;
	; <i>activities</i> menggambarkan langkah kegiatan yang dilakukan pada sistem;
	; <i>decision</i> menunjukkan percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu yang dapat dipilih;
	; <i>join</i> menunjukkan penggabungan lebih dari satu aktivitas;
	; <i>swimlane</i> digunakan untuk memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas;
	; status akhir atau <i>final</i> digunakan pada akhir kegiatan yang terjadi dalam sebuah sistem;

(d) *sequence diagram*;

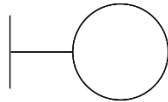
sequence diagram memberi gambaran bagaimana interaksi antar *object* yang terlibat yang diilustrasikan dengan pesan terkirim maupun diterima dalam sekuensinya, adapun simbol pada *sequence diagram* sebagai berikut;



actor mempresentasikan entitas yang berada ; diluar sistem dan berinteraksi diluar sistem;



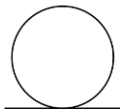
lifeline menghubungkan objek selama ; *sequence* (message dikirim atau diterima).



boundary menggambarkan tampilan form ; dan alat yang berinteraksi dengan yang lain;



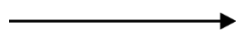
control menggambarkan aliran informasi ; yang menghubungkan antara *boundary* dengan tabel;



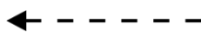
entity menggambarkan hubungan kegiatan ; dimana bertanggung jawab menyimpan informasi;



activation merupakan titik objek mulai ; berpartisipasi yang menunjukkan sebuah objek dalam mengirim atau menerima objek;



message entry menggambarkan hubungan ; antar objek yang menunjukkan urutan kejadian;



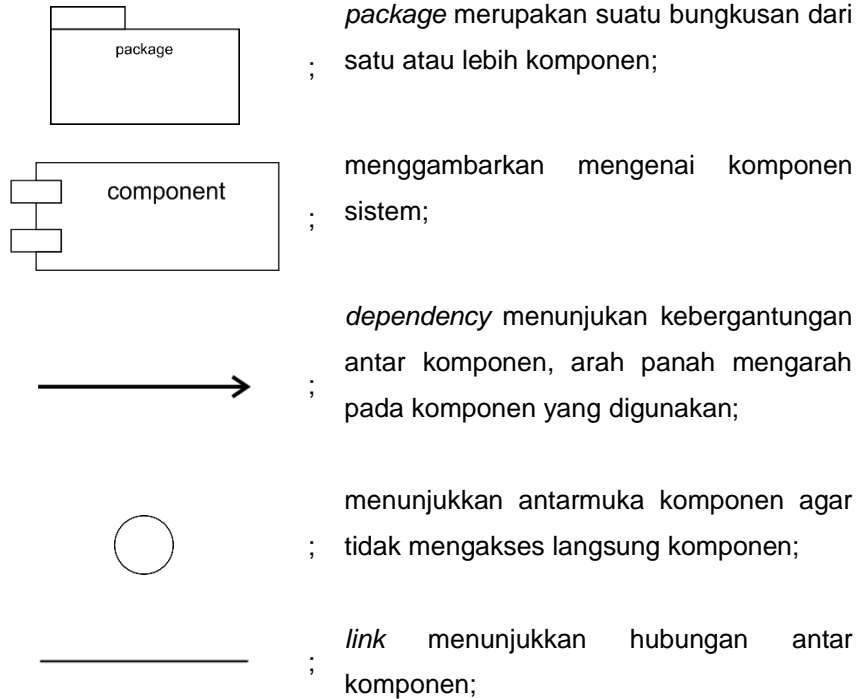
message to self menggambarkan hubungan ; objek itu sendiri, yang menunjukkan urutan kejadian;



message return menggambarkan hasil dari ; pengiriman *message* yang digambarkan dengan arah dari kanan ke kiri.

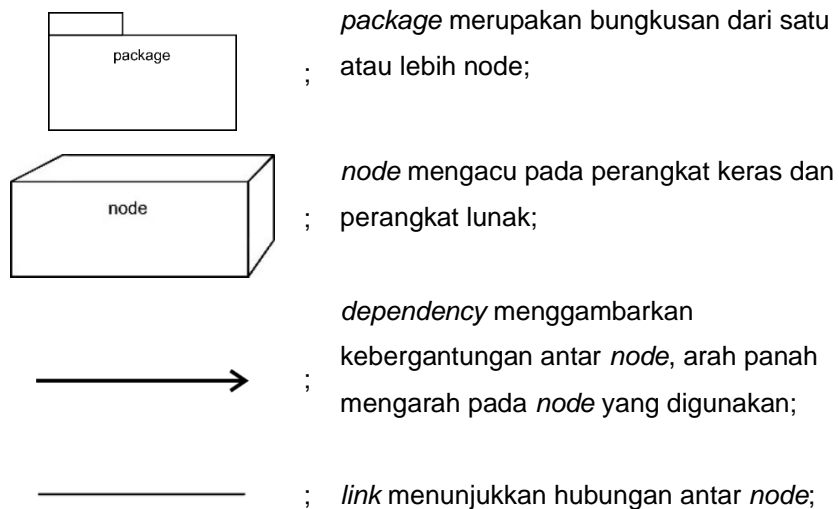
(e) *component diagram*;

component diagram menggambarkan mengenai keterkaitan antara komponen ataupun modul yang terdapat yang menunjukkan bagaimana kode pemrograman dibagi, berikut simbol mengenai *component diagram*;



(f) *deployment diagram*;

deployment diagram digunakan untuk menggambarkan mengenai *software* maupun *hardware* pada sistem yang menunjukkan konfigurasi komponen pada proses pemakaian aplikasi tersebut, adupun simbol yang terdapat pada *deployment diagram* sebagai berikut;



7. Pemrograman

Bahasa pemrograman yang tepat dibutuhkan dalam membangun sebuah aplikasi yang terdapat pada program komputer guna menghasilkan sebuah aplikasi yang sesuai dengan yang diharapkan. Menurut (Khadir, 2014, p. 192) bahasa pemrograman merupakan sekumpulan bahasa perintah yang dimengerti oleh komputer dalam hal berkomunikasi untuk menjalankan perintah tertentu yang pada akhirnya perintah tersebut dapat dijalankan oleh komputer. Pada penelitian ini digunakan bahasa pemrograman sebagai pengolah data untuk melakukan sebuah analisis.

Bahasa pemrograman yang digunakan pada penelitian ini yaitu Python. (Budiharto, 2018, p. 13) menjelaskan bahwa Python merupakan *tools* perangkat lunak yang dapat dilakukan untuk perhitungan numerik, visualisasi dan menganalisis dengan memiliki berbagai *library* pendukung. Bahasa pemrograman Python menunjang untuk pemrograman berorientasi objek dan pemrograman fungsional.

8. Pengujian Black Box dan White Box

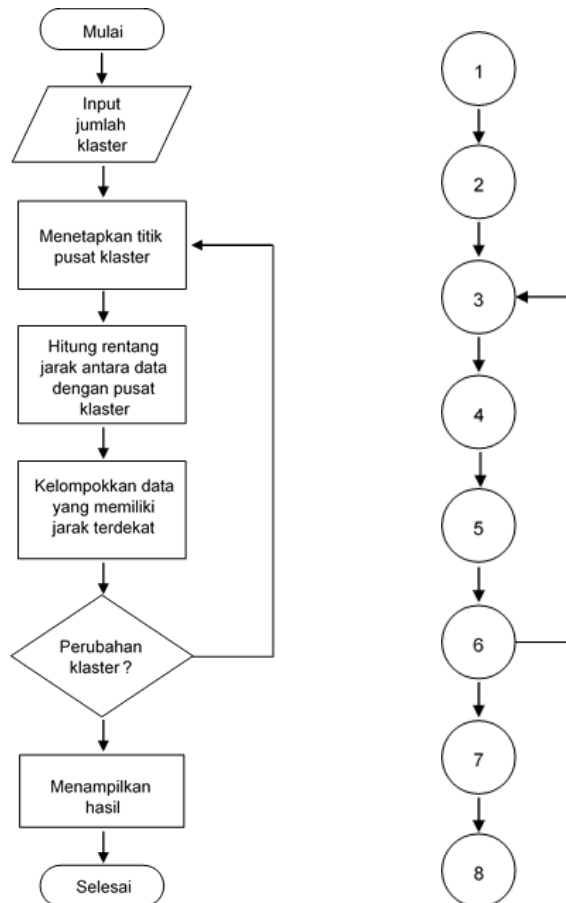
Terdapat dua teknik yang dibutuhkan dalam rancangan suatu program, dua teknik itu yaitu adalah *black-box testing* dan *white-box testing*. Pengujian *black box* dilakukan untuk mengetahui fungsi-fungsi yang terdapat pada program sesuai dengan spesifikasi yang dibutuhkan. *Black box testing* merupakan pengujian yang menganalisa hasil pemeriksaan melewati uji coba data dan mengetahui kegunaan yang terdapat pada perangkat lunak (Sita Eriana, Subariah and Farizy, 2022, p. 172). *Black box testing* dikerjakan dengan hanya menganalisa hasil pengecekan melewati uji coba dan menganalisa kegunaan perangkat lunak. Tabel 2. 1 menunjukkan contoh tabel pengujian tabel *black box* menurut (Sita Eriana, Subariah and Farizy, 2022, p. 173):

Tabel 2. 1 Contoh Tabel Pengujian *Black Box*

Proses Uji Coba	Hasil Yang Diharapkan	Hasil Uji Coba	Kesimpulan
Pencet Tombol "cari lokasi"	Timbul maps lokasi dengan warna berbeda beda	Warna maps pada prototype aplikasi berguna sesuai yang diinginkan	[X] Diterima

Menurut (Sita Eriana, Subariah and Farizy, 2022, p. 172) *white box testing* merupakan suatu pengujian pada aplikasi guna memeriksa rancangan aplikasi secara keseluruhan dan membagi kedalam beberapa bagian dalam menguji aplikasi tersebut. Pengujian *white box* dilakukan dengan cara mengecek

keseluruhan rancangan pada aplikasi dan dibagi menjadi beberapa bagian untuk saat pengujiannya. Adapun contoh pengujian *white box* ditunjukkan pada gambar 2. 4 dengan contoh perhitungan *cyclomatic complexity* sebagai berikut:



Gambar 2. 4 Contoh *Flowchart* dan *Flowgraph* Pengujian *White Box*

diketahui:

$E = 8$ dan $N = 8$;

ditanya: $V(G)$?

jawab:

$V(G) = E - N + 2$;

$V(G) = 8 - 8 + 2$;

$V(G) = 2$;

jadi, nilai $V(G)$ atau CC adalah 2, nilai $V(G)$ atau CC diperlukan untuk mengetahui jumlah independent path yang dapat dibuat pada tahap selanjutnya. Independent path dapat dilihat pada tabel dibawah ini :

Jalur 1 = 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8;

Jalur 2 = 1 – 2 – 3 – 4 – 5 – 6 – 3 – 4 – 5 – 6 – 7 – 8;

Tabel 2. 2 Contoh Tabel Pengujian *White Box*

No	Independent Path
1	1 – 2 – 3 – 4 – 5 – 6 – 7 – 8
2	1 – 2 – 3 – 4 – 5 – 6 – 3 – 4 – 5 – 6 – 7 – 8

Dari tabel 2. 2 dapat dilihat *independent path* bahwa *node* awal dan *node* akhir sudah saling terhubung dan semua *edge* dilewati. Kemudian dari *independent path* dapat diimplementasikan kedalam bentuk *test case* dimana semua alur yang terdapat akan dijalankan agar diketahui hasil dari pengujian apakah sudah sesuai atau tidak.

9. Tentang Penyakit Diabetes

Diabetes merupakan penyakit tidak menular yang ditandai oleh kadar glukosa dalam darah yang memiliki nilai lebih. Menurut (*World Health Organization, 2019*) diabetes merupakan suatu gangguan metabolisme yang ditandai dengan kadar gula darah yang tinggi yang mengganggu metabolisme karbohidat, protein dan lipid yang diakibatkan dari buruknya produksi insulin. Biasanya produkis insulin yang buruk disebabkan oleh gangguan pada sel-sel kelenjar pankreas ataupun kurang responnya tubuh terhadap insulin.

Berdasarkan keputusan menteri kesehatan republik Indonesia nomor HK.01.07/MENKES/1961/2022 tentang rumah sakit jejaring pengampuan pelayanan diabetes mellitus bahwa penyakit diabetes melitus merupakan salah satu penyakit yang mempunyai morbiditas dan angka kematian (mortalitas) yang masih tinggi, maka dari itu diperlukannya optimalisasi pelayanan di rumah sakit untuk meningkatkan kemampuan sumber daya manusia, sarana, prasarana, dan alat kesehatan, serta memperhatikan penatalaksanaan dan rujukan dalam pelayanan kesehatan yang diberikan.

10. Mengenai Pemberian Penyuluhan

Pada penelitian ini bertujuan untuk memberikan rekomendasi prioritas pemberian penyuluhan. Menurut (*Nurmala et al., 2018, p. 63*) penyuluhan merupakan suatu kegiatan pembelajaran dan instruksi di masyarakat dalam mendapatkan pengetahuan dan kemauan untuk melakukan sebuah perubahan perilaku hidup sehat guna mencapai kondisi hidup yang diinginkan dengan dilakukannya kegiatan penyuluhan dapat memberikan kesempatan pada masyarakat untuk memberi *feedback* dari materi yang disampaikan serta kegiatan penyuluhan dapat menurunkan angka kematian dan kesakitan di masyarakat karena telah menerapkan pola hidup sehat, baik secara fisik,

mental maupun sosial. Dengan itu rekomendasi prioritas pemberian penyuluhan pada penelitian ini dapat membantu tenaga kesehatan.

B. Tinjauan Studi

Dalam sebuah penelitian diperlukannya mencari sebuah informasi yang sudah pernah dilakukan pada penelitian sebelumnya. Penelitian rujukan mengenai persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan yang terdapat pada penelitian sebelumnya, demikian pula berdasarkan dengan kesamaan teori Algoritma K-Means. Penelitian tersebut sebagai sumbangan sumber ilmu pengetahuan bagi penelitian yang dilaksanakan. Berikut ini penelitian sebelumnya yang menjadi acuan pada penulisan penelitian ini, yaitu:

(1) Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Penyakit Diabetes Melitus Di Kecamatan Bagor (Niken *et al.*, 2021)

Pada penelitian ini dilakukannya perhitungan dengan algoritma K-Means pada data penderita diabetes melitus di kecamatan Bagor pada tahun 2019 dan 2020. Penelitian ini jumlah cluster yang akan dibentuk adalah 2 cluster dan penentuan Centroid awal dilakukan dengan mengambil data/nilai secara random yang kemudian hasil dan pembahasan penelitian tersebut yaitu pusat cluster 1 terdapat 6 Desa dengan pusat penyebaran penyakit diabetes tingkat tinggi yang harus mendapatkan perhatian lebih dari pemerintah dan pusat cluster 0 terdapat 15 Desa dengan jumlah penderita diabetes mellitus tingkat rendah sampai dengan menengah.

(2) A Novel Application of K-means Cluster Prediction Model for Diabetes Early Identification using Dimensionality Reduction Techniques (Vamshi *et al.*, 2023)

Pada penelitian ini menggunakan data PIMA Indian Diabetes. Akurasi prediksi diabetes berkisar pada rentang 0,9875 hingga 1,0. KCPM (model prediksi *cluster* K-means) dengan tahap preprocessing dan pengumpulan data, dilanjutkan dengan pengelompokan data dengan diberikan nilai K dan tahap terakhir penghapusan titik data yang tidak terdapat di *cluster*.

(3) Clusterization Of Diabetes Health Indicators With K-Means Cluster Algorithm (Perwira *et al.*, 2022)

Pada penelitian ini menggunakan 22 variabel indikator kesehatan diabetes dengan hasil akhir membentuk 3 *cluster*. *Cluster* 0 sebanyak 235 orang sebagai orang yang tidak terkena diabetes, *cluster* 1 sebanyak 1816 orang sebagai penderita pradiabetes dan *cluster* 2 sebanyak 298 orang sebagai orang penderita diabetes tipe 2.

(4) Grouping Diabetes Diagnosis Based on Age Range with K-means Algorithm (Mantik *et al.*, 2021)

Penelitian ini menggunakan algoritma K-Means dan data yang digunakan yaitu data rekam medis pasien Rumah Sakit Royal Prima tahun 2019. Hasil akhir yang diperoleh dari penelitian ini yaitu membentuk 2 *cluster* dimana *cluster* 1 merupakan kelompok pasien penderita diabetes tipe 1 dan *cluster* 2 merupakan kelompok pasien penderita diabetes tipe 2.

(5) Analisis Cluster Menggunakan Algoritma K-Means Cluster Untuk Culstering Jenis Penyakit Menular Pada Puskesmas Di Kecamatan Kota Tangerang (Rachma, Aden and Rusdiana, 2019)

Penelitian ini menggunakan algoritma K-Means untuk Clustering jenis penyakit pada Puskesmas perkecamatan di Kota Tangerang dengan menggunakan data sekunder berupa data bersumber BPS Kota berbentuk dalam dokumen yang peneliti jadikan sebagai populasi dan sampel. Hasil akhir pada penelitian ini berupa 4 klaster, dimana pada klaster 1 TB Paru mendapatkan 12 Puskesmas, klaster 2 Pneumonia mendapatkan 2 Puskesmas, klaster 3 DBD mendapatkan 5 Puskesmas, dan klaster 4 Diare mendapatkan 11 Puskesmas.

(6) Pengelompokan Penyakit Berdasarkan Lingkungan Dengan Algoritma K-Means Pada Puskesmas Sungai Tarab 2 (Permata Sari, 2021)

Penelitian ini menggunakan algoritma K-Means untuk membantu pihak puskesmas dalam memberikan penyuluhan kepada masyarakat dengan menggunakan data pasien puskesmas sungai tarab 2 per oktober 2021 dan diambil 25 sampel data. Hasil akhir pada penelitian ini mengnialisasi sebanyak 2 klaster, dimana cluster 0 pengelompokkan penyakit berbasis lingkungan lingkungan seperti malaria, dbd dan dermatitis sebanyak 10 sedangkan cluster 1 tidak berbasis lingkungan seperti vertigo dan bronkitis sebanyak 5. Algoritma K-Means pada penelitian ini memiliki keakuratan sebesar 90% dari data sampel yang telah diolah.

(7) Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah (Sugianto, Rahayu and Gusman, 2020)

Penelitian ini menggunakan algoritma K-Means untuk membantu pihak puskesmas dalam memberikan penyuluhan kepada masyarakat dengan menggunakan data penyakit pasien Puskesmas Cigugur Tengah dari rentang tanggal 1 Januari 2017 sampai 31 Desember 2017 dengan jumlah 250 records. Hasil akhir pada penelitian ini menginisialisasi sebanyak 2 klaster dengan jumlah cluster 0 untuk mengelompokkan penyakit akut ada 241 item,

cluster 1 mengelompokkan penyakit tidak akut ada 9 item dengan total jumlah data sebanyak 250.

(8) Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering (Fajri and Purnamasari, 2022)

Penelitian ini menggunakan algoritma K-Means untuk mencari pola baru yaitu kecenderungan penyakit dengan usia dengan menggunakan data pasien di Puskesmas Muara Enim dari tahun 2020-2021. Hasil akhir pada penelitian ini dibagi kedalam 3 cluster dimana cluster 0 merupakan cluster dengan penyakit dan penderita tertinggi, cluster 1 merupakan cluster dengan penyakit dan penderita sedikit, sementara cluster 2 merupakan cluster dengan penyakit dan penderita sedang pada Puskesmas Muara Enim dari tahun 2020 hingga 2021.

(9) Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal (Kusnaldi, Gulo and Aripin, 2022)

Penelitian ini menggunakan algoritma K-Means metode K-Means untuk menentukan prioritas bantuan Uang Kuliah Tunggal. Dimana metode yang digunakan dalam normalisasi data yaitu Desimal Scaling dan metode pengelompokan data yaitu K-Means dengan hasil pengujian metode secara manual yaitu terdapat 3 cluster.

(10) Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means (Anggraini *et al.*, 2022)

Penelitian ini menggunakan algoritma K-Means untuk mencari pola baru yaitu kecenderungan penyakit dengan usia dengan menggunakan data data Rekam Medis Pasien Di Puskesmas Ujung Batu. Hasil akhir pada penelitian ini terbentuk sebanyak 2 *cluster*, dengan *cluster* 1 terdiri dari 710 data didominasi oleh desa Sukadamai dengan Jumlah pasien laki-laki sebanyak 247 data dan pasien perempuan 463 data dengan pasien penderita penyakit menular sebanyak 28 data yang di dominasi penyakit Acute upper respiratory infection dan unspecified. *Cluster* 2 terdiri dari 3165 data yang didominasi oleh desa Ngaso dengan jumlah pasien laki-laki sebanyak 1181 data dan pasien perempuan 1984 data, pasien penyakit menular sebanyak 218 data yang di dominasi penyakit Acute upper respiratory infections of multiple and unspecified sites. Dengan pengujian menggunakan algoritma K-Means menunjukkan pembentukan 2 *cluster* dengan nilai 0,646.

Tabel 2. 3 Tinjauan Studi

No	Peneliti	Judul	Jurnal	Kontribusi
1	Roro Niken Enggar Tiasti, Abdulloh Hamid, Dian Candra Rini Novitasari (2021)	Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Penyakit Diabetes Melitus Di Kecamatan Bagor	Jurnal Mahasiswa Matematika ALGEBRA Vol. 02, No. 01 (2021), pp. 69-84 e-ISSN 2774-8677 https://jurnalsaintek.uinsby.ac.id/mhs/index.php/algebra/article/view/77	Berkontribusi dalam algoritma K-Means pada klasterisasi penyebaran penyakit diabetes dan variabel pada penelitian ini
2	Vamshi Krishna B., Raguru Jaya K., Bhuvaneshwari A. P., Gururaj H. L., Vinayakumar Ravi, Meshari Almeshari and Yasser Alzamil (2023)	<i>A Novel Application of K-means Cluster Prediction Model for Diabetes Early Identification using Dimensionality Reduction Techniques</i>	The Open Bioinformatics Journal, 2023, Volume 16 DOI: 10.2174/18750362-16-230825-2023-18, 2023, 16, e187503622307310 https://openbioinformaticsjournal.com	Berkontribusi dalam algoritma K-Means
3	Yuda Perwira, Wira Apriani, Ahmad Zein, Sinta Lia Alfari (2022)	<i>Clusterization Of Diabetes Health Indicators With K-Means Cluster Algorithm</i>	Jurnal Infokum, Volume 10, No. 3, Agustus 2022 ISSN : 2302-9706 http://infor.seaninstitute.org/index.php/infokum/article/view/614	Berkontribusi dalam menghasilkan <i>cluster</i> dengan algoritma K-Means
4	Allwin M. Simarmata, Maria Albina Pratiwi Sianipar, Sukbinder Singh, Intan Is Mutiara Gulo, Jhon Boy Ricardo Purba (2021)	<i>Grouping Diabetes Diagnosis Based on Age Range with K-means Algorithm</i>	Jurnal Mantik, 5 No. 2 (2021) 1408-1412 https://iocscience.org/ejournal/index.php/mantik/article/view/1586	Berkontribusi sebagai acuan metode K-Means

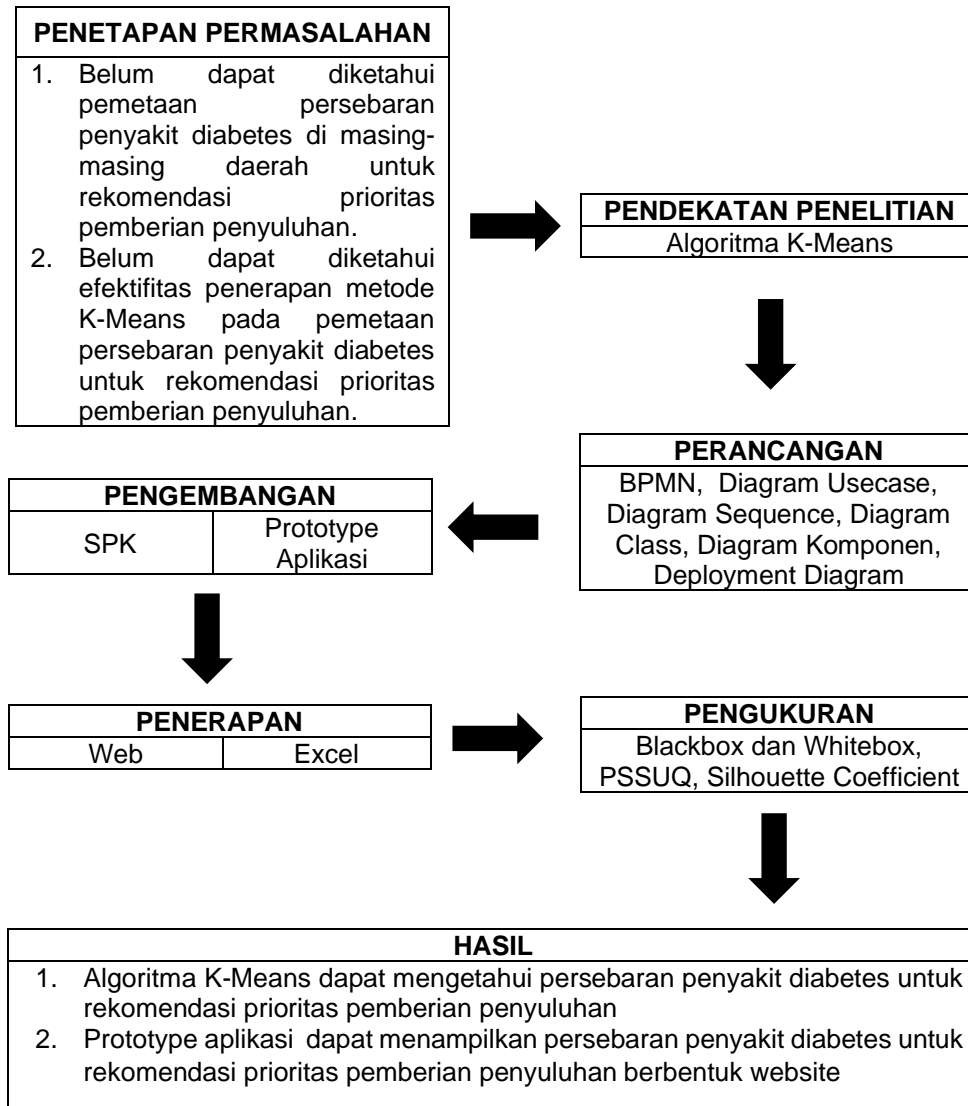
No	Peneliti	Judul	Jurnal	Kontribusi
5	Aninda Rachman, Aden, S.Si., M.Pd., dan Yulianti Rusdiana, S.Si., M.Sc (2019)	Analisis Cluster Menggunakan Algoritma K-Means Cluster Untuk Culstering Jenis Penyakit Menular Pada Puskesmas Di Kecamatan Kota Tangerang	p-ISSN 2621-7856 e-ISSN 2655-7312 JURNAL SAINTIKA UNPAM Vol. 2 No. 1 Juli 2019 (15-27) http://openjournal.unpam.ac.id/index.php/journal/article/view/2915	Berkontribusi dalam pengklasteran daerah rawan penyebaran penyakit
6	Dian Permata Sari (2019)	Pengelompokan Penyakit Berdasarkan Lingkungan Dengan Algoritma K-Means Pada Puskesmas Sungai Tarab 2	JOISIE Journal Of Information System And Informatics Engineering Vol. 5, No.2 Desember 2021, Hlm 75-81 p-ISSN: 2503-5304 e-ISSN: 2527-3116 https://www.ejournal.pelitaindonesia.ac.id/ojs32/index.php/JOISIE/article/view/1700	Berkontribusi pada teknik data mining <i>clustering</i> metode K-Means
7	Castaka Agus Sugianto, Ayu Hendrati Rahayu, Aditia Gusman (2020)	Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah	JOINT (Journal of Information Technology) (p)ISSN: 2527-9467/(e)ISSN: 2656-7539 Vol. 02 No 02 Agustus 2020, pp. 39-44 https://jurnal.stmik-amikbandung.ac.id/joint/article/view/30	Berkontribusi pada pengelompokan penyakit pasien dengan metode K-Means
8	Muhammad Bhakti Fajri, Susan Dian Purnamasari (2022)	Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering	Journal of Information Technology Ampera Vol. 3, No. 3, December 2022 e-ISSN: 2774-2121 https://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/584	Berkontribusi pada <i>cluster</i> dengan dataset kunjungan pasien dengan metode K-Means

No	Peneliti	Judul	Jurnal	Kontribusi
9	Muhammad Rafli Kusnaldi, Timotius Gulo, Soeb Aripin (2022)	Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal	Journal of Computer System and Informatics (JoSYC), Volume 3, No. 4, August 2022 https://ejournal.seminar-id.com/index.php/josyc	Berkontribusi dalam melakukan normalisasi data yaitu <i>decimal scalling</i> .
10	Rahayu Angraini, Elin Haerani, Jasril, lis Afrianty (2022)	Pengelompokkan Penyakit Pasien Menggunakan Algoritma K-Means	JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 6, Desember 2022 e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak) DOI 10.30865/jurikom.v9i6.5145 https://ejournal.stmik-budidarma.ac.id/index.php/jurikom/issue/view/176	Berkontribusi dalam perhitungan <i>Silhouette Coefficient</i> menggunakan metode K-Means

Berdasarkan tinjauan pustaka yang sudah dipaparkan pada tabel 2. 3 didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Kontribusi dalam rujukan penelitian ini memberikan pengetahuan ilmu mengenai permasalahan pada persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan dengan proses perhitungan K-Means. Penelitian dilakukan karena ada pengembangan dari penelitian sebelumnya. Penelitian ini dilakukan tahap pengembangan dari penelitian sebelumnya dengan menggunakan metode K-Means yang akan diuji dengan *silhouette coefficient* pada persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan dengan menggunakan variabel yang digunakan adalah nama pasien, faktor risiko penderita diabetes seperti merokok, kurang aktifitas fisik, pola makan, konsumsi alkohol, pemeriksaan gula, dan kelurahan.

C. Kerangka Berfikir

Kerangka pemikiran merupakan rancangan dari dasar pemikiran pada sebuah penelitian. Gambar 2. 5 merupakan kerangka pemikiran pada penelitian ini.



Gambar 2. 5 Kerangka Pemikiran

Penjelasan kerangka pemikiran pada gambar 2. 5 sebagai berikut:

- (1) permasalahan yang terjadi pada penelitian ini yaitu belum dapat diketahui penyakit diabetes di masing-masing daerah dan belum dapat diketahui efektifitas penerapan metode K-Means pada pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan;

- (2) pendekatan penelitian dilakukan untuk menemukan solusi dari permasalahan yang terjadi sehingga dilakukan dengan menggunakan algoritma K-Means dan pemrograman;
- (3) perancangan penelitian dilakukan untuk merancang aplikasi secara sistematis dengan menggunakan BPMN, *Diagram Usecase*, *Diagram Sequence*, *Diagram Class*, *Diagram Component* dan *Deployment Diagram*;
- (4) tahap pengembangan dilakukan dengan sistem pendukung keputusan dan kemudian membuat *prototyping* dengan menggunakan bahasa pemrograman *Python*;
- (5) penerapan penelitian dilakukan dengan menggunakan web dan file excel sebagai data yang digunakan;
- (6) pengujian yang dilakukan dengan menggunakan *black box* dan *white box*, PSSUQ, dan *silhouette coefficient*;
- (7) hasil dari penelitian & pengembangan ini bahwa algoritma K-Means dapat mengetahui persebaran penyakit diabetes di masing-masing daerah untuk rekomendasi prioritas pemberian penyuluhan.

D. Hipotesis

Algoritma K-Means merupakan algoritma pada data *mining* yang terdapat pada permasalahan pengelompokan melakukan klusterisasi data berdasarkan kesamaan pada karakteristiknya. Beberapa penelitian rujukan yang menggunakan algoritma K-Means menunjukkan tingkat keberhasilan penggunaan metode K-Means. Basis kerja pada metode K-Means yaitu melakukan suatu pengelompokan, maka bahwa pada penelitian ini juga berkaitan dengan pengelompokan data yang di duga dapat diketahui persebaran penyakit diabetes dan diketahui efektifitas penerapan metode K-Means pada persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan kepada pihak terkait agar dapat mengambil keputusan yang tepat mengenai pencegahan dan penanganan pada permasalahan tersebut. Dari pemahaman diatas maka dapat ditetapkan hipotesis penelitian ini adalah bahwa penerapan metode K-Means di duga dapat mengetahui persebaran penyakit diabetes dan diketahui efektifitas penerapan metode K-Means pada persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan.