

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining

Sebagaimana ditunjukkan oleh (Fayyad et al. 1996 dalam Suyanto, 2017:1) menyatakan bahwa information mining merupakan kemajuan ilmiah menuju ukuran pengungkapan informasi dalam kumpulan data atau Knowledge Discovery In Database yang diringkas sebagai KDD.

KDD (Knowledge Discovery In Database) mengelola metode rekonsiliasi dan wahyu logis, pemahaman dan representasi contoh dalam berbagai informasi. Rangkaian siklus ini memiliki tahapan-tahapan yang menyertainya (Tan, 2006):Pembersihan data yaitu untuk membuang data yang tidak konsisten dan noise

- a. Integrasi data yaitu penggabungan data dari beberapa sumber
- b. Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk di mining
- c. Aplikasi teknik data mining yaitu proses ekstraksi pola dari data yang ada
- d. Evaluasi pola yang ditemukan yaitu proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan
- e. Presentasi pengetahuan yaitu dengan teknik visualisasi

Tahap ini penting untuk tindakan pencarian informasi yang mencakup pemeriksaan apakah contoh atau data yang ditemukan bertentangan dengan kenyataan saat ini atau spekulasi yang ada sebelumnya. Kemajuan terakhir dari KDD adalah memperkenalkan informasi dalam struktur yang mudah diperoleh klien (Tan, 2006).

2. Clustering

Menurut Tan (2006) Pengelompokan adalah siklus untuk mengelompokkan informasi menjadi beberapa kelompok atau kumpulan sehingga informasi dalam satu tandan memiliki tingkat kemiripan yang paling ekstrim dan informasi antar tandan memiliki kedekatan dasar. Dengan menggunakan pengelompokan ini, kita dapat mengatur daerah tebal, menemukan desain - contoh sirkulasi umum, dan menemukan hubungan menarik antara kredit informasi. Dalam penambangan informasi, upaya dipusatkan di sekitar menemukan teknik untuk mengelompokkan kumpulan data yang sangat besar secara memadai dan produktif.

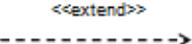
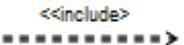
Sebagian dari persyaratan pengelompokan dalam penambangan informasi mencakup kemampuan beradaptasi, kemampuan untuk menangani berbagai jenis karakteristik yang dilengkapi untuk menjaga dimensi tinggi, menjaga informasi yang bergejolak, dan dapat diinterpretasikan tanpa masalah. Alasan informasi pengelompokan ini adalah untuk membatasi pekerjaan target yang ditetapkan dalam sistem pengelompokan, yang sebagian besar mencoba untuk membatasi varietas di dalam kelompok. Selanjutnya, batasi varietas antar tandan. Berbicara secara

komprehensif, ada beberapa teknik untuk pengelompokan informasi. Keputusan teknik pengelompokan bergantung pada jenis informasi dan alasan pengelompokan yang sebenarnya.

3. Pengertian Unified Modeling Language (UML)

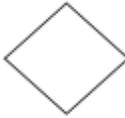
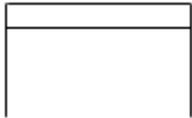
Menurut Sri Dharwiyanti (2003), Unified Modeling Language (UML) adalah bahasa yang bergantung pada ilustrasi/gambar untuk membayangkan, menunjukkan, membangun, dan merekam kerangka kerja peningkatan produk yang bergantung pada OO (Object Oriented).

Tabel 2. 1 Simbol Usecase Diagram

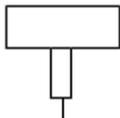
GAMBAR	NAMA	KETERANGAN
	ACTOR	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri, jadi walaupun simbol dari actor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama actor.
GAMBAR	NAMA	KETERANGAN
	USECASE	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama usecase.
	ASOSIASI/ASSOSIATION	Komunikasi antara actor dan usecase yang berpartisipasi pada usecase atau usecase memiliki interaksi dengan actor.
	EKSTENSI/EXTEND	Relasi usecase tambahan ke sebuah usecase dimana usecase yang ditambahkan dapat berdiri sendiri walau tanpa usecase tambahan memiliki nama depan yang sama dengan usecase yang ditambahkan.
	GENERALISASI/GENERALIZATION	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah usecase dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
	MENGGUNAKAN INCLUDE	Relasi usecase tambahan ke sebuah usecase dimana usecase yang ditambahkan

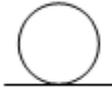
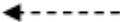
		memerlukan usecase ini untuk menjalankan fungsional atau sebagai syarat dijalankan usecase ini.
--	--	---

Tabel 2. 2 Simbol Activity Diagram

GAMBAR	NAMA	KETERANGAN
	<i>STATUS AWAL/INITIAL</i>	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	<i>AKTIVITAS/ACTIVITY</i>	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	<i>PERCABANGAN/DECISION</i>	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	<i>PENGGABUNGAN/JOIN</i>	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	<i>STATUS AKHIR/FINAL</i>	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status baru.
	<i>SWIMLINE</i>	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

Tabel 2. 3 Simbol Sequence Diagram

GAMBAR	NAMA	KETERANGAN
	<i>ACTOR</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>LIFELINE</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
	<i>GENERAL</i>	Merepresentasikan entitas tunggal dalam sequence.

	<i>BOUNDARY</i>	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
GAMBAR	NAMA	KETERANGAN
	<i>CONTROL</i>	Elemen mengatur aliran dari informasi untuk sebuah skenario. Objek ini umumnya perLaris dan perLaris bisnis.
	<i>ENTITAS</i>	Elemen yang bertanggung jawab menyimpan atau informasi. Ini dapat berupa beans atau model object.
	<i>ACTIVATION</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>MESSAGE ENTRY</i>	Berfungsi untuk menggambarkan pesan/hubungan antar objek yang menunjukkan urutan kejadian
	<i>MESSAGE TO SELF</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi
	<i>MESSAGE RETURN</i>	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

Tabel 2. 4 Simbol Class Diagram

GAMBAR	NAMA	KETERANGAN
	<i>GENERALIZATION</i>	Hubungan dimana objek anak (descendent) berbagi perLaris dan struktur data dari objek yang ada di atasnya objek induk (ancestor).
	<i>NARY ASSOCIATION</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
	<i>CLASS</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.

GAMBAR	NAMA	KETERANGAN
	<i>COLLABORATION</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
	<i>REALIZATION</i>	Operasi yang benar-benar dilakukan oleh suatu objek
	<i>DEPENDENCY</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri.
	<i>ASSOCIATION</i>	Apa yang menghubungkan antara objek satu dengan obyek lainnya.

4. Pengembangan System Development Life Cycle (SDLC)

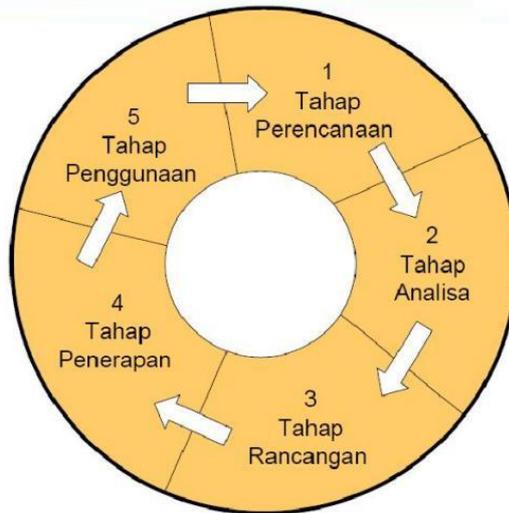
Raymond McLeod (2007) menyatakan bahwa pendekatan frameworks adalah sebuah teknik. Filsafat adalah metode yang disarankan untuk menyelesaikan sesuatu. Pendekatan kerangka kerja adalah prosedur mendasar dalam menangani segala jenis masalah. Siklus hidup kemajuan kerangka kerja (SDLC) adalah penggunaan cara kerangka kerja untuk menangani peningkatan kerangka kerja data.

Ada beberapa fase pekerjaan perbaikan yang harus dilakukan jika suatu usaha ingin memiliki kemungkinan pencapaian yang tinggi. Tahapan tersebut adalah: Perencanaan

- a. Analisis
- b. Desain
- c. Implementasi
- d. Penggunaan

Usaha dan aset yang diharapkan untuk menyelesaikan pekerjaan diatur dan kemudian diatur. Kerangka kerja saat ini juga diperiksa untuk memahami masalah dan menentukan prasyarat yang berguna dari kerangka kerja baru. Kerangka kerja baru ini kemudian direncanakan dan dilaksanakan. Setelah eksekusi, kerangka tersebut kemudian digunakan, di dunia yang sempurna untuk jangka waktu yang signifikan.

Karena pekerjaan di atas mengikuti contoh normal dan dilakukan dengan cara hierarkis, SDLC biasa sering disinggung sebagai pendekatan kaskade. Tindakan ini memiliki aliran satu arah yang mendorong hasil proyek.



Gambar 2. 1 Pola melingkar dari siklus hidup sistem

(Sumber : Raymond McLeod, 2007)

Ketika sebuah kerangka telah melampaui masa pakainya dan harus diganti, siklus hidup lain dimulai dengan tahap penataan. Tidak sulit untuk memahami bagaimana SDLC konvensional dapat dianggap sebagai pemanfaatan pendekatan kerangka kerja. Masalah ini akan ditandai dalam fase persiapan dan pemeriksaan. Pengaturan elektif diakui dan dinilai dalam tahap rencana. Kemudian, pada saat itu, pengaturan terbaik dilakukan dan digunakan. Selama tahap penggunaan, kritik dikumpulkan untuk memahami seberapa baik kerangka kerja dapat mengatasi masalah yang tidak benar-benar ditetapkan.

5. Bahasa Pemrograman

a. Pengertian PHP

Sesuai (Solichin, 2016:11) mengklarifikasi bahwa PHP adalah bahasa pemrograman online yang disusun oleh dan untuk insinyur web. PHP pertama kali dikembangkan oleh Rasmus Lerdorf, seorang insinyur produk dan individu dari Tim Apache dan dikirimkan pada akhir 1994.

b. Pengertian Hypertext Markup Language (HTML)

Sesuai (Solichin, 2016:10) menjelaskan bahwa "Hypertext Markup Language (HTML) adalah bahasa pemrograman web yang memberitahu browser internet bagaimana membuat dan menyajikan substansi pada halaman situs web. Pada akhirnya, HTML adalah pembentukan web. HTML dapat menampilkan item seperti teks, tabel, gabungan, gambar, rekaman.

c. Framework CodeIgniter

Sesuai Yenda Purbadian (2016: 18) mengatakan bahwa "CodeIgniter adalah struktur yang dibuat dengan menggunakan bahasa pemrograman PHP yang direncanakan untuk memudahkan para insinyur perangkat lunak web untuk membuat atau mengembangkan aplikasi elektronik".

6. Database Yang Digunakan

1. Database

Menurut Mustakini (2009: 46), basis informasi adalah berbagai informasi yang saling berhubungan satu sama lain, disimpan pada peralatan PC dan digunakan oleh pemrograman untuk mengontrol.

2. MySQL

Menurut Kustiyahningsih (2011: 145), My SQL adalah kumpulan data yang berisi satu atau berbagai tabel. Sebuah tabel terdiri dari berbagai baris dan setiap baris berisi satu atau berbagai tabel. Sebuah tabel terdiri dari berbagai kolom dan setiap baris berisi satu atau berbagai tabel. Tipe informasi My SQL, seperti yang ditunjukkan oleh Kustiyahningsih (2011: 147), "Tipe informasi My SQL adalah informasi yang terdapat dalam sebuah tabel sebagai kolom yang berisi sisi-sisi informasi. Nilai informasi dalam kolom memiliki jenisnya sendiri.

7. Web Server

Seperti yang ditunjukkan oleh Robert McCool (1995), Apache adalah nama server web yang bertanggung jawab untuk reaksi permintaan HTTP dan pencatatan data tertentu (penggunaan penting). Demikian juga, Apache juga dicirikan sebagai server web yang minimal, terukur, mengikuti pedoman konvensi HTTP, dan tentu saja sangat terkenal.

8. Intranet

Menurut Prakoso (2007:119), Intranet adalah kumpulan jaringan PC lokal yang menggunakan pemrograman web dan protokol TCP/IP atau HTTP. Dengan cara ini, jaringan intranet memiliki semua kantor yang diklaim oleh web, misalnya, email, File Transfer Protocol (FTP, dll. Jaringan intranet adalah jaringan web yang hanya diklaim oleh organisasi dan tidak dapat diperoleh ke dari luar.

9. Euclidean Distance (Jarak *Euclidean*)

Teknik Euclidean melihat jarak dasar gambar pengujian dengan dasar informasi gambar persiapan. Jarak Euclidean dari dua vektor x dan y ditentukan oleh situasi:

$$d(x, y) = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

Keterangan :

$d(x, y)$ =Tingkat Perbedaan (*Dissimilarity Degree*)

n =Jumlah vector

x_i = Vektor citra *input*

y_i = Vektor citra pembanding/*output*

Semakin kecil nilai $d(x, y)$, semakin komparatif kedua vektor tersebut dikoordinasikan/dipikirkan. Sebaliknya, semakin diperhatikan nilai $d(x, y)$, semakin unik koordinat kedua vektor tersebut (Budi Santosa, 2007).

10. Flowchart

Menurut Indrajani (2011:22), Flowchart adalah penggambaran grafis tentang sarana dan susunan strategi suatu program. Biasanya mempengaruhi tujuan masalah yang secara khusus harus direnungkan dan dinilai lebih lanjut.

Flowchart dipisahkan menjadi 5 macam flowchart, antara lain flowchart framework, flowchart laporan, flowchart skema, flowchart program, flowchart ukuran. Setiap jenis flowchart akan dijelaskan di bawah ini:

a. System Flowchart

Framework Flowchart dapat dicirikan sebagai grafik yang menunjukkan alur kerja umum dari framework. Grafik ini menggambarkan suksepsi metodologi dalam kerangka. Grafik aliran kerangka kerja menunjukkan apa yang dilakukan dalam kerangka kerja.

b. Document Flowchart

Flowchart arsip atau disebut juga structure flowchart atau desk work flowchart adalah diagram aliran yang menunjukkan perkembangan laporan dan struktur termasuk duplikatnya.

c. Schematic Flowchart

Diagram alir skematik adalah garis besar aliran yang seperti grafik aliran kerangka kerja, yang menggambarkan metodologi dalam kerangka kerja. Yang penting adalah, diagram aliran skema serta menggunakan gambar grafik aliran kerangka kerja, juga menggunakan gambar PC dan perangkat keras lain yang digunakan. Motivasi di balik penggunaan gambar-gambar ini adalah untuk bekerja dengan korespondensi dengan orang-orang yang ingin tahu tentang gambar grafik aliran. Penggunaan gambar-gambar ini mudah, namun merepotkan dan membosankan untuk menggambar.

d. Program Flowchart

Bagan alir program (program flowchart) adalah garis besar yang menggambarkan secara mendalam sarana interaksi program. Bagan alir program dibuat dari inferensi bagan alur kerangka kerja. Bagan alir program dapat terdiri dari dua jenis, yaitu bagan alur alasan program tertentu dan bagan alur program PC tertentu. Grafik aliran alasan program digunakan untuk menggambarkan setiap perkembangan dalam program PC dengan bijaksana. Kerangka instrumen dasar pemikiran program ini disiapkan oleh para ahli kerangka kerja. Gambar terlampir menunjukkan grafik aliran alasan program. Diagram alur program PC poin demi poin digunakan untuk menggambarkan arah program PC yang rumit. Diagram aliran ini disiapkan oleh pengembang.

e. Process Flowchart

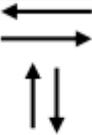
Bagan alir proses (process flowchart) merupakan bagan alir yang banyak digunakan di teknik industry. Bagan alir ini juga berguna bagi analisis sistem untuk

menggambarkan proses dalam suatu prosedur. Berikut ini merupakan notasi atau symbol-simbol yang digunakan dapat dibagi menjadi 3 (tiga) kelompok yaitu

1. Flow Direction Symbols (SimbolPenghubung/alur)

Simbol yang digunakan untuk menghubungkan antara symbol yang satu dengan yang lainnya. Simbol ini juga disebut connecting line, simbol tersebut adalah

Tabel 2. 5 Flow Direction Symbols

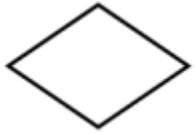
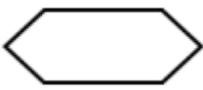
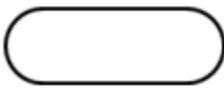
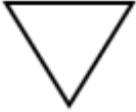
NO	Simbol	Nama	Keterangan
1		<i>Arus/Flow</i>	Untuk menyatakan jalannya arus suatu proses
2		<i>Communication link</i>	Untuk menyatakan bahwa adanya transisi suatu data atau informasi dari suatu lokasi ke lokasi lainnya
3		<i>Connector</i>	Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman / lembaran sama
4		<i>Offline</i>	Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman atau lembaran yang berbeda

2. Processing Symbols (Simbol Proses)

Simbol yang menunjukkan jenis operasi pengolahan dalam suatu proses prosedur. Simbol – simbol tersebut adalah :

Tabel 2. 6 Processing Symbols

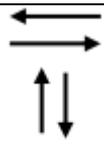
NO	Simbol	Nama	Keterangan
1		Proses	Sebuah fungsi pemrosesan yang dilaksanakan oleh

			komputer biasanya menghasilkan perubahan terhadap data atau informasi
2		<i>Symbol</i>	Untuk menyatakan suatu tindakan (proses) yang tidak dilakukan oleh <i>computer</i> (manual)
3		<i>Decision / Logika</i>	Untuk menunjukkan suatu kondisi tertentu, dgn dua kemungkinan, YA / TIDAK
4		<i>Predefine Process</i>	Untuk menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberi harga awal
5		<i>Terminal</i>	Untuk menyatakan permulaan atau akhir suatu program
6		<i>Offline Storage</i>	Untuk menunjukkan bahwa data dalam simbol ini akan disimpan ke suatu media tertentu
7		<i>Manual Input</i>	Untuk memasukkan data secara manual dengan menggunakan online <i>keyword</i>

3. Input / Output Symbols (Simbol Input – output)

Simbol yang menunjukkan jenis peralatan yang digunakan sebagai media input atau output. Simbol – simbol tersebut adalah

Tabel 2. 7 Input - Ouput

NO	Simbol	Nama	Keterangan
1		<i>Arus/Flow</i>	Untuk menyatakan jalannya arus suatu proses

2		<i>Communication link</i>	Untuk menyatakan bahwa adanya transisi suatu data atau informasi dari suatu lokasi ke lokasi lainnya
3		<i>Connector</i>	Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman / lembaran sama
4		<i>Offline</i>	Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman atau lembaran yang berbeda

B. Metode Algoritma K-Means

Teknik. Pada umumnya, K-Means telah menjadi salah satu perhitungan utama di bidang penambangan informasi (Wu dan Kumar, 2009).

Pada umumnya, tipe dasar K-Means ditemukan oleh berbagai ilmuwan dari berbagai disiplin ilmu. yang paling persuasif adalah Lloyd (1982), Forgey (1967), Friedman dan Rubin (1967), McQueen (1967). Perhitungan K-Means dibentuk menjadi setting yang lebih besar sebagai perhitungan slope climbing, seperti yang diperkenalkan oleh Gray dan Nuhoff (1998).

K-Means dapat diterapkan pada informasi yang dialamatkan dalam ruang r -dimensi suatu tempat. K-Means menggabungkan koleksi informasi r -dimensi, $= \{x_i | i = 1, \dots, \}$ dimana alamat informasi ke- i sebagai "fokus informasi". Seperti yang dijelaskan sebelumnya bahwa K-Means membagi X ke dalam tandan K , perhitungan K-Means menggabungkan semua informasi fokus di X sehingga setiap titik jatuh hanya pada salah satu bagian K . Yang penting adalah fakta bahwa hal itu ada dalam kelompok yang dilengkapi dengan cam dengan memberikan setiap poin ID grup. Fokus dengan ID kelompok yang sama berada dalam grup yang sama, sedangkan fokus dengan berbagai ID kelompok berada di berbagai grup. Untuk mengkomunikasikan ini, biasanya diselesaikan dengan vektor partisipasi tandan m dengan panjang N , di mana m , adalah ID grup titik. (Windarto, 2007:26-33)

Batasan yang harus dimasukkan ketika menggunakan perhitungan K-Means adalah bahwa nilai K yang digunakan umumnya didasarkan pada data yang baru diketahui tentang jumlah kelompok informasi yang muncul di X , jumlah tandan yang diperlukan untuk pelaksanaannya, atau jenisnya. banyak yang dicari dengan menyelidiki/melakukan tes. dengan beberapa kualitas K . Jumlah kualitas K dipilih, tidak

penting untuk melihat bagaimana K-Means mengemas koleksi informasi X. (Prasetyo, 2014: 190)

Dalam K-Means, setiap kelompok tandan K dialamatkan oleh satu titik di . Kumpulan delegasi grup dikomunikasikan sebagai $= \{c_j \mid j = 1, \dots, K\}$. Berbagai grup delegasi K juga disebut sebagai grup atau centroid grup (atau hanya centroid). Koleksi informasi dalam X dikumpulkan tergantung pada gagasan tentang kedekatan atau komparabilitas. Meskipun ide yang dirujuk untuk informasi yang dikumpulkan dalam satu kelompok adalah informasi yang sebanding, jumlah yang digunakan untuk mengukurnya adalah disparitas. Artinya, informasi dengan perbedaan kecil/dekat (jarak) terikat untuk berpartisipasi dalam satu tandan. Pengukuran yang biasanya digunakan untuk disparitas adalah Euclidean. (Prasetyo, 2013: 190)

Prasetyo mengungkapkan, ketika informasi telah menentukan perbedaan untuk setiap centroid, maka pada titik itu, disparitas terkecil kemudian dipilih sebagai kelompok yang akan diikuti sebagai pergerakan informasi dalam kelompok dalam suatu penekanan. Perpindahan suatu informasi dalam kelompok yang diikuti dapat dikomunikasikan dengan nilai partisipasi a_{ij} yang bernilai 0 atau 1. Nilai 0 jika bukan merupakan individu dari suatu kelompok dan 1 jika merupakan individu dari sekelompok. Karena K-Means secara ketat mengelompokkan informasi hanya dalam satu kelompok, maka, pada saat itu, dari nilai informasi di semua tandan, hanya satu yang bernilai 1, sedangkan yang lain bernilai 0 seperti yang dinyatakan oleh kondisi berikut:

$$a_{ij} = \begin{cases} 1 & \text{arg min } \{(X_i, c_j)\} \\ 0 & \text{lainnya} \end{cases}$$

(X_i, c_j) menyatakan ketidakmiripan (jarak) dari data ke- i ke cluster C_j

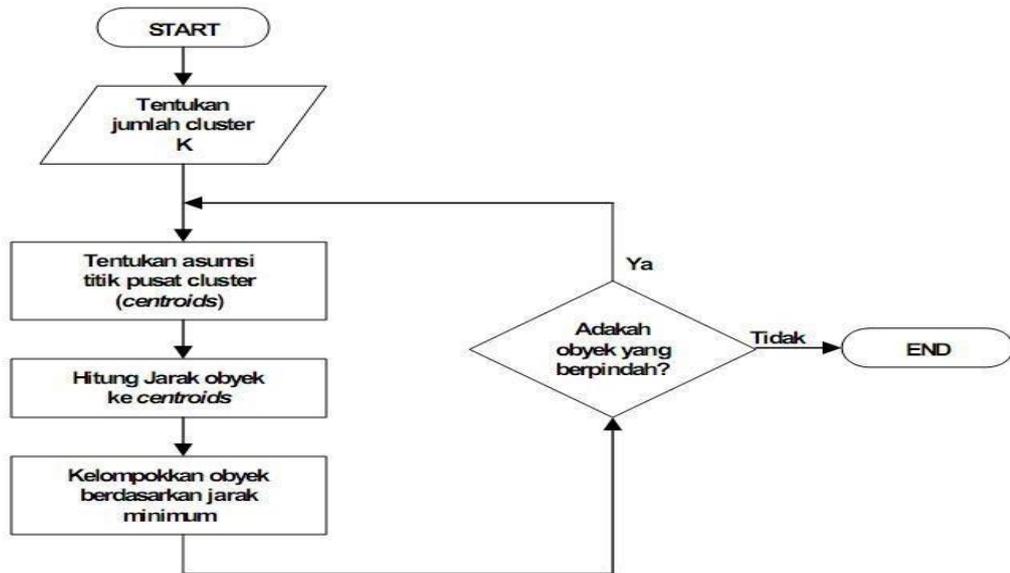
Sementara relokasi centroid untuk mendapatkan titik centroid C didapatkan dengan menghitung rata – rata setiap fitur dari semua data yang tergabung dalam setiap cluster. Rata- rata sebuah fitur dari semua data dalam sebuah cluster dinyatakan oleh persamaan berikut :

$$C_j = \frac{1}{N_k} \sum_{l=1}^{N_k} x_{jl}$$

N_k adalah jumlah data yang tergabung dalam sebuah cluster. Jika diperhatikan dari langkahnya yang selalu memilih cluster terdekat, maka sebenarnya K-Means berusaha untuk meminimalkan fungsi objektif/fungsi biaya non-negatif, seperti dinyatakan oleh persamaan berikut

$$J = \sum_{i=1}^N \sum_{l=1}^K a_{il} (x_i, C_l)^2$$

Dengan kata lain, K-Means berusaha untuk meminimalkan total jarak kuadrat (*squared distance*) di antara setiap titik x_i dan representasi cluster c_j terdekat. Berikut adalah taha



Gambar 2. 2 Tahapan Algoritma K-Means
(Sumber:Prasetyo,2013)

Keterangan :

1. Inisialisasi adalah tentukan nilai K sebagai jumlah cluster yang diinginkan dan metrik ketidakmiripan (jarak) yang diinginkan. Jika perlu, tetapkan ambang batas perubahan fungsi objektif dan ambang batas perubahan fungsi objektif dan ambang batas perubahan posisi centroid.
2. Pilih K data dari set data X sebagai centroid.
3. Alokasikan semua data ke centroid terdekat dengan metrik jarak yang sudah ditetapkan (memperbarui cluster ID setiap data).
4. Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu (a) perubahan fungsi objektif sudah di bawah ambang batas yang diinginkan; atau (b) tidak ada data yang berpindah cluster; atau (c) perubahan posisi centroid sudah di bawah ambang batas yang ditetapkan.

Perhitungan K-Means menggabungkan koleksi informasi X dalam kemajuan berulang. Dua kemajuan utama adalah (1) penjaminan ulang ID grup dan semua fokus informasi di X, dan (2) menyegarkan penggambaran kelompok (centroid) mengingat fokus informasi di setiap kelompok. (Prasetyo, 2013).

Fungsi perhitungannya sebagai berikut: pertama, penggambaran kelompok dipasang dengan sembarangan memilih informasi K pada . Kemudian, secara iteratif memainkan dua tahap yang menyertainya sampai kondisi bersamaan tercapai. (Prasetyo, 2013)

Tahap 1 Tugas data. Setiap informasi dialokasikan ke centroid terdekat dengan memutuskan hubungan tanpa jaminan. Hasilnya adalah informasi yang dibagi. Tahap 2 Relokasi "menandakan". Setiap penggambaran tandan dipindahkan ke tengah dengan rata-rata juggling angka dari semua informasi yang diturunkan ke dalamnya. Alasan untuk perkembangan ini tergantung pada persepsi bahwa dalam memberikan sekelompok fokus, penggambaran terbaik mutlak untuk set (sejauh membatasi jumlah kuadrat jarak Euclidean antara setiap titik informasi. Ini juga mengapa teknik ini digunakan. sering disebut rumpun mean atau group centroid, sesuai dengan nama yang dimiliki (Prasetyo, 2013).

Perhitungan K-Means tiba pada keadaan bersatu ketika redistribusi fokus informasi (dan selanjutnya area centroid) tidak berubah lagi. Interaksi dari satu penekanan ke penekanan lain sampai tercapai kondisi yang bersamaan juga dapat diperoleh. Dalam kondisi yang semakin bergabung, sangat mungkin terlihat bahwa nilai kapasitas target akan berkurang. Mengatasi masalah optima terdekat dapat diselesaikan dengan menjalankan perhitungan beberapa kali dengan berbagai centroid pengantar dan kemudian memilih hasil terbaik. (Prasetyo, 2013)

Pemilihan nilai K yang optimal juga menjadi hal sulit dilakukan. Jika ada informasi mengenai set data, seperti jumlah partisi yang secara alami menggambarkan set data, maka informasi tersebut dapat digunakan untuk memilih nilai K yang optimal. Jika tidak, maka harus menggunakan beberapa kriteria lain untuk memilih K, kemudian menyelesaikan masalah pemilihan model tersebut. Solusi yang naif adalah dengan mencoba beberapa nilai K berbeda dan memilih clustering yang nilai fungsi objektifnya minimal. Sayangnya nilai yang diberikan oleh fungsi objektif tidak cukup informatif untuk digunakan sebagai harapan penyelesaian masalah ini. Misalnya biaya solusi optimal terhadap peningkatan K menurun sampai menjadi nol ketika jumlah cluster sama dengan jumlah titik data berbeda. Hal tersebut menjadi lebih sulit jika menggunakan fungsi objektif untuk (a) secara langsung membandingkan solusi dengan jumlah cluster berbeda, dan (b) mencari nilai K yang optimal.

Oleh karena itu jika K yang dibutuhkan tidak diketahui, salah satu cara yang biasa digunakan adalah menjalankan K-Means dengan nilai K berbeda dan kemudian menggunakan beberapa kriteria yang lain yang cocok untuk memilih satu hasil yang terbaik. Misalnya X-Means menambahkan istilah kompleksitas (yang meningkat terhadap K) pada fungsi objektif yang asli dan kemudian mengidentifikasi K yang meminimalkan biaya penyesuaian (Pelleg dan Moore, 2000). Alternatif yang lain adalah secara progresif meningkatkan jumlah cluster dengan gabungan kriteria pemberhentian yang cocok. Bisecting K-Means melakukan hal tersebut dengan meletakkan semua data dalam cluster tunggal dan kemudian secara rekursif memecah cluster paling padat menjadi dua cluster menggunakan 2-means (Steinbach, Karypis, dan Kumar, 2000).

Pemilihan K titik data sebagai centroid awal juga memengaruhi hasil clustering.

Sifat ini menjadi karakteristik alami K-Means yang dapat mengakibatkan hasil cluster yang didapat pada percobaan berbeda mendapatkan hasil berbeda. Kondisi seperti ini dikenal dengan solusi yang *local optima*, yang artinya algoritma K-Means sangat sensitif terhadap lokasi awal centroid. Dengan kata lain, inisialisasi set representasi cluster C yang berbeda dapat mengakibatkan hasil cluster yang berbeda, bahkan pada set data X yang sama. Inisialisasi yang jelek dapat mengakibatkan hasil cluster yang jelek juga.

Contoh Kasus

Data-data yang digunakan pada penelitian ini adalah data transaksi penjualan produk pasta pada Gia Restaurant Jakarta, dari bulan Januari 2017 sampai Desember 2017. Penelitian ini memfokuskan hanya pada produk pasta saja. Untuk memudahkan pemrosesan data maka nama produk diubah menjadi kode produk dan menambahkan atribut rata-rata perbulan yang di ambil dari data jumlah transaksi selama satu tahun dibagi 12 bulan. Inisialisasi data transaksi penjualan bisa dilihat pada tabel 2.7.

Tabel 2. 8 Inisialisasi Data Transaksi Penjualan Pasta

Nama Produk	Jumlah Transaksi	Rata- Rata
Bottarga	3.804,00	317,00
Granchio	1.146,00	95,50
Ricci	1.692,00	141,00
Gamberi	350,00	29,17
Caserecce	1.039,00	86,58
Ragu Danatra	773,00	64,42
Salsiccia	790,00	65,83
Tacchino	373,00	31,08
Gnocchi	405,00	33,75
Agnolotti	1.215,00	101,25
Fettuccine	1.231,00	102,58
Scoglio	1.638,00	136,50
Pici	36,00	3,00
Gnocchi Sorrentina	423,00	35,25
Ravioli	631,00	52,58
Foglie d orate	284,00	3,67
Agnoloti plin	319,00	26,58
Fettuccine Verdi	679,00	56,58
Cavatelli	315,00	26,25
Linguine Aragosta	132,00	11,00
Paccheri Ai Ricci	109,00	9,08

Sumber: Hasil Pengolahan Data (2017)

Tabel 2. 9 Centroid Awal Setiap Cluster

Titik Pusat Awal (Centroid)	Produk	Kode Produk	Jumlah Transaksi (X)	Rata-rata (Y)
1	Agnoloti Plin	17	319,00	26,58
2	Agnolotti	10	1,22	101,25

Sumber: Hasil Pengolahan Data (2017)

$$d(x_1, c_2) = \sqrt{\sum_{j=1}^n (x_{1j} - c_{2j})^2} = \sqrt{(3.804 - 1.215)^2 + (317 - 101,25)^2}$$

$$= \sqrt{6.702.921 + 46548,06} = 2597,97$$

Jarak data ke-1 pada *cluster 2* adalah : Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data ke-1 dengan pusat cluster kedua adalah 2597,97.

Berdasarkan hasil kedua perhitungan di atas dapat disimpulkan bahwa jarak data kedua yang paling dekat dengan cluster 1, sehingga data kedua di masukan ke dalam cluster dua. Hasil perhitungan selengkapnya untuk 21 data produk pasta iterasi pertama dapat dilihat pada tabel 2.9

Tabel 2. 10 Hasil Perhitungan Data Ke Setiap Cluster

Kode produk	Jarak Centroid		Jarak Terdekat	Cluster yang diikuti
	1	2		
1	3.497,08	2.597,97	2.597,97	2
2	829,87	69,24	69,24	2
3	1.377,76	478,65	478,65	2
4	31,11	868,00	31,11	1
5	722,50	176,61	176,61	2
6	455,57	443,53	443,53	2
7	472,63	426,47	426,47	2
8	54,19	844,92	54,19	1
9	86,30	812,81	86,30	1
10	899,11	0,00	0,00	2
11	915,16	16,06	16,06	2
12	1.323,57	424,47	424,47	2
13	283,98	1.183,09	283,98	1

14	104,36	794,75	104,36	1
15	313,08	586,02	313,08	1
16	35,12	934,23	35,12	1
17	0,00	899,11	0,00	1
18	361,25	537,86	361,25	1
19	4,01	903,12	4,01	1
20	187,65	1.086,75	187,65	1

Sumber: Hasil Pengolahan Data (2017)

Langkah keempat, menghitung kembali centroid yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut.

Langkah kelima, setelah didapatkan centroid yang baru dari setiap cluster, Lariskan kembali dari langkah ketiga hingga centroid dari setiap cluster tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

Dalam penelitian ini, iterasi clustering data produk terjadi sebanyak empat kali iterasi. Pada iterasi keempat ini, centroid dari setiap cluster sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

C. Pengelompokkan Data Produk Herbal

Pengelompokkan adalah proses, cara, perbuatan mengelompokkan. Data adalah catatan atas kumpulan fakta. Data merupakan bentuk jamak dari datum, berasal dari bahasa Latin yang berarti "sesuatu yang diberikan". Dalam penggunaan sehari-hari data berarti suatu pernyataan yang diterima secara apa adanya.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Herdianto, 2013 : 8).

menurut Rangkuti (2004), persediaan merupakan suatu aktiva yang meliputi barang-barang milik perusahaan dengan maksud untuk dijual dalam suatu periode usaha tertentu atau persediaan barang-barang yang masih dalam pengerjaan atau proses produksi, ataupun persediaan bahan baku yang menunggu penggunaannya dalam suatu proses produksi.

Jadi, Pengelompokan Data Obat Herbal Untuk meprediksi produk herbal adalah suatu proses pengelompokkan dimana data obat herbal bisa digunakan sebagai acuan dalam menentukan produk herbal yang Laris dan tidak Laris dalam melakukan proses prediksi produk herbal.

D. Tinjauan Pustaka

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan, **Penerapan Algoritma K-MEANS Untuk Prediksi Produk Herbal Yang Laris.** Kesamaan metode yang digunakan yaitu Algoritma K-MEANS. Banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain pada tabel dibawah ini

1. Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. (Elly Muningsih¹ dan Sri Kiswati², AMIK BSI Yogyakarta,2015).

Berdasarkan hasil yang dicapai terkait dengan penelitian penentuan stok barang di online shop Ragam Jogja maka dapat disimpulkan bahwa:

Manajemen stok yang dilakukan secara tidak akurat dan asal-asalan akan menyebabkan biaya simpan yang tinggi dan tidak ekonomis, karena bisa terjadi kekosongan atau kelebihan produk tertentu. Hal ini tentu akan sangat merugikan semua peLaris usaha seperti halnya online shop. Penelitian ini bertujuan untuk mengelompokkan produk yang dijual pada online shop Ragam Jogja menjadi beberapa cluster untuk mengetahui produk mana yang paling diminati sehingga jumlah stok harus banyak, produk diminati untuk jumlah stok sedang dan produk kurang diminati untuk jumlah stok sedikit. Metode yang digunakan pada penelitian ini adalah metode K-Means yang merupakan salah satu metode terbaik dan paling populer dalam algoritma clustering dimana K-Means mencari partisi yang optimal dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Variabel yang digunakan adalah kode produk, jumlah transaksi, volume penjualan dan rata-rata penjualan. Penelitian juga didukung dengan software Rapidminer untuk pengolahan data dengan metode K-Means. Hasil akhir dari penelitian adalah berupa suatu program aplikasi yang dapat mengelompokkan produk menjadi kategori jumlah stok banyak, sedang dan sedikit berdasarkan transaksi penjualan.

2. Pemanfaatan Metode K-Means Dalam Penentuan Persediaan Barang.(Santoso Setiawan, Teknik Komputer; AMIK BSI Jakarta.2018).

Penjualan produk yang paling banyak terjual dapat diketahui dengan menggunakan KMeans, dengan mengelompokkan data produk, jumlah transaksi, dan rata-rata perbulan. Untuk rata-rata diambil dari data jumlah transaksi selama 1 tahun dibagi 12 bulan, lalu pilih 2 data secara acak untuk menentukan centroid awal pada setiap cluster, setelah data pada setiap cluster tidak berubah-ubah,maka dapat diketahui data yang paling banyak membutuhkan stok, yaitu data produk ke 1, 2, 3, 10, 11 dan 12. Dari hasil penelitian yang telah di dilakukan di dapat data yang paling sedikit terjual dimana untuk produk ke 4, 5, 6, 7, 8, 9, 13, 14, 15, 16, 17, 18, 19, 20,

dan 21. Sebagai pilihan produk yang dapat dibuatkan promosi oleh pihak perusahaan untuk meningkatkan penjualan.

3. DATA MINING DENGAN METODE CLUSTERING UNTUK PENGOLAHAN INFORMASI PERSEDIAAN OBAT PADA KLINIK SRIKANDI MEDIKA BERBASIS WEB.(Elmayati, STMIK Musirawas Lubuklinggau.2017).

Dari hasil penelitian yang penulis Lariskan di Klinik Srikandi Medika maka dapat ditarik kesimpulan bahwa hasil dari metode clustering data mining dapat digunakan untuk mengetahui kelompok obat yang paling laris dan tidak laris pada Klinik Srikandi Medika, sehingga apabila akan dilakukan pengadaan persediaan obat, petugas dapat melihat daftar obat dengan penjualan terlaris.

4. Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai.(Taslim1, Fajrizal2, Program Studi Teknik Informatika, Universitas Lancang Kuning, Pekanbaru.2016).

Dari hasil clusterisasi data obat dapat disimpulkan bahwa banyak diantara obat yang ada masuk dalam kelompok cluster kurang. Kecendrungan obat yang tergolong kurang ini adalah ada beberapa bulan yang tidak ada permintaan sama sekali terhadap obat tersebut. Sedangkan kelompok obat yang termasuk tinggi rata rata permintaan obatnya setiap bulan diatas 300 buah, sedangkan obat yang masuk dalam kelompok cluster sangat tinggi rata rata permintaan setiap bulannya adalah diatas 2000 buah. Dari hasil analisa cluster diatas mungkin perlu dilakukan lagi penelitian lanjutan agar clusterisasi data obat dapat dilakukan secara lebih valid dengan menetapkan nilai centroid terbaik.

5. Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means Clustering Sales Data at Outdoor Equipment Stores Using KMeans Method.(Fintri Indriyani1, Eni Irfiani2, Universitas Bina Sarana Informatika Jakarta.2019).

Penerapan metode K-Means dalam pengelompokan data penjualan pada Toko Genta Corp dapat menghasilkan rekomendasi barang yang laris, Kurang laris dan cukup laris. Sehingga data dijadikan rujukan bagi manajemen untuk mengatur stok barang agar toko tidak mengecewakan pelanggan karena barang yang ingin di beli tidak tersedia. Untuk pengembangan lebih lanjut dan agar pihak toko mudah dalam penggunaan metode ini maka perlu dibuat sebuah aplikasi dengan acuan metode K-Means

6. Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means.(Ricsa Andrian1 , Septian Fendy2 , Aryo Nugroho3.2019)

Berdasarkan dari hasil penelitian, metode K-Means dapat diterapkan untuk melakukan klasterisasi dengan mengubah data menjadi 3 kelompok kemudian kelompok tersebut dapat digunakan untuk mengetahui produk yang diminati, sedang dan kurang diminati. Dari hasil klasterisasi, perusahaan dapat melakukan

pengendalian persediaan berdasarkan kelompok produk seperti penjelasan hasil dan pembahasan di atas yaitu produk klaster 0 dapat digunakan sebagai persediaan dengan jumlah sedikit, sedangkan kelompok produk pada klaster 1 digunakan sebagai persediaan dengan jumlah banyak, dan kelompok produk klaster 2 dapat digunakan sebagai persediaan dengan jumlah sedang.

7. Penerapan Metode K-Means Clustering untuk Menentukan Persediaan Stok Barang Pada Toko Pensmart Jambi.(Nengsi Anggraini¹, Jasmir², Pareza Alam Jusia³,STIKOM Dinamika Bangsa, Jambi.)

Adapun beberapa yang dapat diambil dari penelitian ini yaitu, antara lain :

1. Dengan menerapkan metode k-means clustering pada pengolahan data dari ketiga jenis data yang penulis gunakan sebagai bahan penelitian ini dan penulis gabungkan menjadi 1 jenis data menghasilkan sebuah hasil (output) yang berbeda-beda pula seperti yang telah dijelaskan sebelumnya.
2. Dari hasil (output) yang telah diperoleh dari proses penerapan metode k-means clustering, maka hasil dari analisis dan evaluasi yang di dapat berupa sebuah kelompok data yang telah ter-cluster dari masing-masing kelompok, dan hasil (output) tersebut dapat dijadikan sebuah informasi yang berguna bagi pemilik Toko Pensmart jambi, serta dapat menjadi acuan untuk pemilik toko dapat menerapkan metode k-means clustering dalam membantu proses pengolahan stok barang mereka selanjutnya.

8. Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru.(Gustientiedinaa, M.Hasmil Adiyaa, Yenny Desnelita Teknik Informatika, Sekolah Tinggi Ilmu Komputer Pelita Indonesia, Pekanbaru.2019).

Dari hasil clusterisasi pada data obat – obatan dapat ditarik kesimpulan bahwa kelompok obat yang termasuk pemakaian sedikit rata rata permintaan obat setiap tahunnya kurang dari 18000 buah, dan obat yang termasuk pemakaian sedang rata rata permintaan obat setiap tahunnya diantara 18000–70000 buah, sedangkan obat yang masuk kedalam kelompok obat yang pemakaian tinggi rata – rata permintaan obat setiap tahunnya diatas 70000 buah.

9. SISTEM PENGELOMPOKAN DAN REKOMENDASI PENGADAAN BARANG DENGAN MENGGUNAKAN METODE K-MEANS.(BUDI WAHYU APRIANTO, UNIVERSITAS NUSANTARA PGRI KEDIRI.2018).

Kurang tepatnya sistem pengadaan barang yang dilakukan pada sebuah toko dapat mengakibatkan menurunnya angka penjualan di dalam toko tersebut, karena pemilihan barang yang kurang tepat akan berdampak pada kurangnya model ataupun jumlah barang yang ada dalam persediaan dan hal tersebut akan berpengaruh besar pada angka penjualan toko yang akan menurun dan berakibat merugikan pemilik toko tersebut. Permasalahan yang dihadapi pemilik toko adalah (1) kurang efektifnya sistem pengadaan barang yang ada. (2) dibuatnya sebuah

aplikasi yang dapat mengelompokkan barang dan memberi data yang tepat sebagai bahan acuan untuk pengadaan barang agar pengadaan barang lebih efektif sehingga dapat menaikkan angka penjualan di dalam toko. Metode yang digunakan dalam sistem ini adalah metode K-Means Clustering dimana data yang dipakai adalah data penjualan. Sistem akan mengelompokkan data penjualan berdasarkan kelas tertentu, lalu sistem akan merekomendasikan barang dengan acuan dari kelas-kelas yang hasil dari Clustering. Hasil pada sistem yang telah dibuat adalah (1) Sistem dapat mengelompokkan barang. (2) Membuat sistem yang dapat merekomendasi barang apa yang akan diadakan berdasarkan kelas.

10. KLASTERISASI DATA PADA TOKO OBAT MENGGUNAKAN METODE K-MEANS UNTUK ANALISA KETERSEDIAAN BARANG.(Yusuf1), Lukman Junaedi, Universitas Narotama Surabaya.2020).

Produk yang paling banyak terjual dapat diketahui dengan menggunakan K-Means, dengan mengelompokkan data penjualan, jumlah transaksi, dan rata-rata. Untuk rata-rata diambil dari data jumlah transaksi selama 2 bulan, Dari hasil klasterisasi data obat bahwa dapat disimpulkan masih banyak di antara obat yang dalam kelompok tergolong kurang Laris. Kecendrungan data obat yang kurang Laris 69 obat tersebut adalah beberapa hari tidak ada permintaan sama sekali pada obat tersebut, sedangkan obat yang termasuk Laris rata 35 buah dari hasil analisa cluster diatas mungkin perlu dilakukan lagi penelitian lanjutan agar clustering data obat dapat dilakukan secara lebih valid dengan nilai centroid yang terbaik.

Tabel 2. 11 Tinjauan Pustaka

No	Nama Peneliti	Jurnal	Judul	Permasalahan	Kontribusi
1.	Elly Muningsih 1 dan Sri Kiswati 2	Jurnal Bianglala Informatika Vol 3 No 1 Maret 2015 – lppm3.bsi.ac.id/jurnal	Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang	Permasalahan yang terjadi pada online shop tersebut disebabkan karena online shop mengalami kesulitan dalam menentukan stok minimum tiap barang yang harus dipenuhi berdasarkan minat konsumen.	Dengan menggunakan metode K-Means yang menghasilkan sistem penentuan stok barang

2.	Santoso Setiawan*	Jurnal Penelitian Ilmu Komputer, System Embedde d & Logic p-ISSN: 2303-3304,e-ISSN: 2620-3553 6 (1): 41 - 48 (Maret 2018)	Pemanfaatan Metode K-Means Dalam Penentuan Persediaan Barang	Manajemen stok yang tidak akurat akan menyebabkan biaya penyimpanan yang tinggi dan tidak ekonomis, karena kemungkinan terjadinya kekosongan atau kelebihan produk tertentu.	Dengan Pemanfaatan metode k-means dapat menghasilkan sistem penentuan persediaan barang
3.	Elmayati	Jurnal Pelita Informatika, Volume 6, Nomor 2, Oktober 2017 ISSN 2301-9425 (Media Cetak) Hal: 159-164	DATA MINING DENGAN METODE CLUSTERING UNTUK PENGOLAHAN INFORMASI PERSEDIAAN OBAT PADA KLINIK SRIKANDI BERBASIS WEB	persediaan obat pada Klinik Srikandi Medika hanya berdasarkan pada pengalaman sebelumnya, yaitu hanya dengan memeriksa persediaan obat yang hampir habis	Dengan menggunakan metode k-means dapat menghasilkan sistem pada persediaan obat
4.	Taslim, Fajrizal	Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 7,	Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai	Ketersediaan obat	Clusterisasi dapat digunakan untuk menganalisa pemakaian obat untuk mengajukan permintaan obat pada periode yang akan datang.

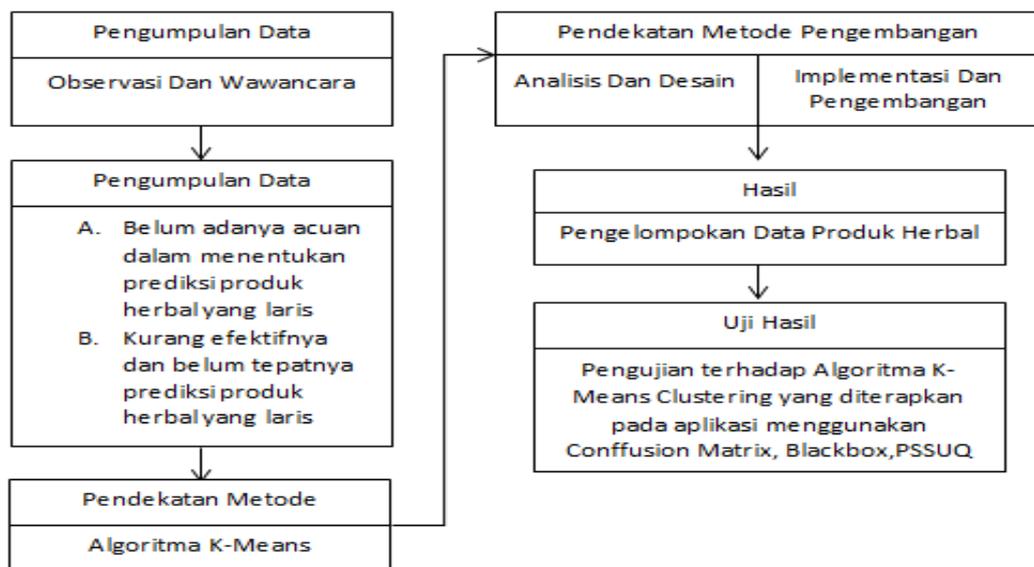
		Nomor 2,Novemb er 2016:108- 114			
5.	Fintri Indriyani1 , Eni Irfiani2	<i>JUITA: Jurnal Informatika e-ISSN: 2579-9801; Volume 7, Nomor 2, November 2019</i>	<i>Clustering</i> Data Penjualan pada Toko Perlengkapan <i>Outdoor</i> Menggunakan Metode K-Means	kurangnya persediaan stok perlengkapan standar keamanan yang dimiliki para pecinta	algoritma K-Means dapat membantu Genta Corp untuk pengelompokan data penjualannya agar dapat memaksimalkan stoknya.
6.	Ricsa Andrian1 , Septian Fendy2 , Aryo Nugroho	journal of Information Technology and Computer Science (JOINTCS) Vol. 4, No. 1, Januari 2019	Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means	menentukan jumlah persediaan barang yang dibutuhkan.	algoritma K-Means, menunjukkan bahwa kelompok 1 (rendah) merupakan kelompok produk yang kurang diminati, kelompok 2 (sedang) merupakan kelompok kategori sedang, kelompok 3 merupakan kelompok produk yang banyak diminati
7.	<i>Nengsi Anggraini¹, Jasmir², Pareza Alam</i>	<i>STIKOM Dinamika Bangsa, Jambi</i>	Penerapan Metode K-Means Clustering untuk Menentukan Persediaan Stok Barang Pada Toko Pensmart Jambi	proses dalam penambahan stok barang, dari sekian banyak jenis item barang . Tidak diketahui barang apa saja yang paling diminati dan kurang diminati	metode <i>K-Means Clustering</i> . ini juga dapat mengelompokkan item barang pada <i>cluster</i> masing-masing yang telah ditentukan.

				terkadang stok numpuk saat barang kurang Laris sedangkan stok yang kosong barang yang paling diminati	
8.	Gustientie dinaa , M.Hasmil Adiyaa , Yenny Desnelita b	JURNAL NASIONAL TEKNOLOGI DAN SISTEM INFORMASI- VOL. 05 NO. 01 (2019) 017-024	Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru	perencanaan dan pengendalian pasokan medis di rumah sakit	<i>Clustering</i> dalam data mining dapat digunakan untuk menganalisa pemakaian obat-obatan, perencanaan dan pengendalian obat-obatan di rumah sakit.
9.	Budi Wahyu Aprianto	Simki-Techsain Vol. 02 No. 03 Tahun 2018 ISSN : 2599-3011	SISTEM PENGELOMPOKAN DAN REKOMENDASI PENGADAAN BARANG DENGAN MENGGUNAKAN METODE K-MEANS	kurang efektifnya sistem pengadaan barang	metode K-Means Clustering dimana data dapat dipakai untuk mengelompokkan data penjualan berdasarkan kelas tertentu, lalu sistem akan merekomendasikan barang dengan acuan dari kelas-kelas yang hasil dari Clustering
10.	Yusuf1) , Lukman Junaedi2) ,	Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC) Vol. 13, No. 2. Agustus 2020	KLASTERISASI DATA PADA TOKO OBAT MENGGUNAKAN METODE K-MEANS UNTUK ANALISA KETERSEDIAAN BARANG	ketersediaan obat pada apotek	metode K-Means, dapat mengelompokkan data berdasarkan pola tertentu. Proses klasterisasi berhasil mendeteksi bahwa terdapat 69 jenis obat yang tidak ada permintaan sama sekali, dan 35 obat yang cepat Laris.

Dari 10 jurnal penelitian, jurnal yang menjadi acuan dalam penelitian ini adalah “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang”. Metode dan hasil dari 10 jurnal tersebut menggunakan metode K-Means yang dapat digunakan untuk mengklasterisasi. Sedangkan untuk persamaan dengan penelitian ini yaitu menggunakan metode K – Means. Dan untuk perbedaannya dengan penelitian ini yaitu pada atribut dalam mengelompokkan data obat herbal

E. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang dapat dilihat pada Gambar 2.3



Gambar 2. 3 Kerangka Pemikiran

Kerangka pemikiran pada gambar 2.3 dapat dijelaskan sebagai berikut :

1. Penelitian ini diawali dengan melakukan pengumpulan data mengenai objek permasalahan tersebut dengan cara observasi dan wawancara.
2. Kemudian dilakukan mengidentifikasi masalah mengenai memprediksi produk herbal yang Laris dan tidak Laris, lalu dapat diidentifikasi masalah diantaranya yaitu :
 - a. Belum adanya acuan dalam memprediksi produk herbal yang Laris.
 - b. Kurang efektif dan belum tepatnya prediksi produk herbal yang Laris
3. Dari permasalahan tersebut peneliti ingin menyelesaikan masalah dengan menggunakan pendekatan al Algoritma K-Means untuk mengelompokkan data obat herbal untuk membantu memprediksi produk herbal yang Laris.
4. Prosedur penelitian menggunakan pendekatan metode melalui analisis dan desain serta implementasi dan pengembangan
5. Penggunaan metode Algoritma K-Means diperoleh hasil pengelompokkan obat herbal berdasarkan kriteria-kriteria terpilih
6. Pengujian hasil kualitas *cluster* dan uji aplikasi.

F. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu, efektif dan belum akurat pada saat melakukan proses prediksi produk herbal yang laris. Dalam teori data mining ada beberapa metode yang dapat melakukan pengelompokan untuk memprediksi suatu peluang dimasa depan berdasarkan dari pengalaman dimasa sebelumnya, diantaranya adalah metode algoritma K-means adalah algoritma yang sudah banyak dikenal dan digunakan untuk proses pengelompokan. Hipotesis pada penelitian ini adalah penerapan Algoritma K-Means diduga dapat prediksi produk herbal yang laris.