

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **II.1. Penelitian Terkait**

Adapun penelitian terkait yang akan digunakan sebagai sumber acuan yang relevan dan terkini yaitu:

Berdasarkan penelitian yang dilakukan oleh Koko Handoko (2016) yang berjudul “Penerapan Data Mining Dalam Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode *K-Means Clustering* (Studi Kasus Di Program Studi TKJ Akademi Komunitas Soloselatan)”, Koko Handoko menyimpulkan bahwa menentukan *centroid* (titik pusat) pada tahap awal algoritma *K-Means* sangat berpengaruh pada hasil *cluster* seperti pada hasil pengujian yang dilakukan dengan menggunakan 25 dataset dengan *centroid* yang berbeda menghasilkan hasil *cluster* yang berbeda pula.

Berdasarkan penelitian yang dilakukan oleh Fina Nasari (2016) yang berjudul “Penerapan Algoritma *K-Means Clustering* Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat”, Fina Nasari menyimpulkan Diare merupakan penyakit yang bertanggung jawab untuk sekitar seperempat dari 130.000 kematian tahunan diantara anak balita, terutama pada musim pancaroba seperti yang terjadi di hampir seluruh kawasan Indonesia tidak terkecuali di kabupaten langkat sumatera utara. Untuk melihat kawasan penyebarannya perlu dibuat sebuah pengelompokan kawasan penyebaran diare, agar diperoleh daerah

penyebaran diare dan pusat penyebarannya. Hasil pengelompokkan data pada iterasi pertama berbeda dengan kelompok awal, oleh karena itu lanjut pada tahap berikutnya yaitu menentukan kembali nilai *centroid* atau titik pusat data sementara.

Berdasarkan penelitian yang dilakukan oleh Ari Sulistiyo (2015) yang berjudul “Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode *K-Nearest Neighbor Classifier* (Studi Kasus : SMAN 16 Semarang)”, membahas Penentuan penempatan Siswa Kelas X dengan menggunakan beberapa kriteria yaitu Nilai Rapor, Nilai Ujian, Minat Siswa dan Rekomendasi BK. Dalam penentuan jurusan minat program studi minat siswa ini dimodelkan dengan metode *K-Nearest Neighbor Classifier*. Pemodelan ini berdasarkan kriteria-kriteria yang akan di-inputkan oleh pengguna. Langkah dalam perhitungan *K-Nearest Neighbor Classifier* ini, yaitu mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru dengan menentukan hubungan jarak bobot antar atribut Minat dan Rekomendasi pada data yang ada sehingga dapat dilakukan perhitungan yang kemudian dilakukan perhitungan jarak bobot selanjutnya menentukan akurasi terhadap data uji yang telah diklasifikasikan. Setelah didapat nilai akurasi, maka dapat ditentukan penentuan penjurusan siswa.

Berdasarkan penelitian yang dilakukan oleh Fitri Yunita yang berjudul “Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri)”,

Fitri Yunita menyimpulkan bahwa Setelah dilakukan pengelompokan data penerimaan mahasiswa menggunakan metode k-means Clustering terbentuk tiga cluster yaitu cluster satu dengan jumlah 195 items, cluster dua 271 dengan jumlah 271 items, dan cluster tiga dengan jumlah 50 items.

## **II.2. Tinjauan Pustaka**

### **II.2.1. Data Mining**

Data Mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan (Alfa Saleh : 2015; 3).

### **II.2.2 Clustering**

Salah satu teknik yang dikenal dalam data mining yaitu *clustering*. Pengertian *clustering* keilmuan dalam data mining adalah pengelompokan sejumlah data atau objek ke dalam *cluster* (group) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya (Fitri Yunita : 2018; 240).

### II.2.3. Algoritma K-Means

Pengertian dari *K-Means Clustering* adalah, K dimaksudkan sebagai konstanta jumlah cluster yang diinginkan, *Means* dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster*, sehingga *K-Means Clustering* adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervise (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Dasar algoritma *K-Means* adalah sebagai berikut :

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
2. Inisialisasi k sebagai centroid yang dapat dibangkitkan secara random.
3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut :

$$d(P, Q) = ||x - y|| = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya*.
5. Tentukan posisi *centroid* baru (k).

6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama (Marsono; 2019 : 33).

Adapun data yang akan di proses dalam tahap ini dapat dilihat pada Tabel dibawah ini. Tabel ini merupakan tabel yang bersikan data yang sudah dapat diolah dengan menggunakan metode *clustering k-means*.

**Tabel 1 Data Clustering K-Means**

No.	Produk	Jumlah Suplay	Produksi
1	Gofress Peppermint Flavor	66,693	402,047
2	Gofress Orange Flavor	31,136	182,848
3	Gofress Strawberry Flavor	49,23	259,581
4	Gofress Lemon Flavor	50,577	289,494
5	Gofress Manggo Flavor	51,083	281,392
6	Gofress Grape Flavor	65,597	464,498
7	Ice N Cool Herbal Mint	72,552	387,168
8	Ice N Cool Herbal Liquorice	162,619	964,001

9	Wake Up 12 Strips Coffe Extract	113,609	706,419
10	Herbafress Strips	61,33	329,557
11	Mentos Pure Fresh Chewing Gum	48,902	290,954

Dari data yang telah ada pada Tabel 1 maka dapat diketahui Algoritma Metode *k-means* sebagai berikut:

1. Jumlah *Cluster* 3
2. Penentuan pusat awal *cluster* dengan data random.

**Tabel 2 Pusat awal *cluster***

<b>No.</b>	<b>Pusat <i>Cluster</i></b>	<b>Jumlah Suplay</b>	<b>Produksi</b>
1	Ice N Cool Herbal Liquorice	162,619	964,001
2	Ice N Cool Herbal Mint	72,552	387,168
3	Gofress Orange Flavor	31,136	182,848

3. Dalam penelitian ini digunakan metode hard k-menas untuk mengalokasikan setiap data ke dalam suatu cluster yang memiliki jarak paling dekat dengan titik pusat setiap cluster. Untuk mengetahui cluster mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap cluster. Sebagai contoh, akan dihitung jarak dari data penyuplaian dari setiap produk pertama ke pusat cluster pertama.

$$D(1,1) = \sqrt{(66,693 - 162,619)^2 + (402,047 - 964,001)^2} = 570,082534$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data produk pertama dengan pusat cluster pertama adalah 570,082534.

Untuk cluster kedua dengan data pertama:

$$D(1,2) = \sqrt{(66,693 - 72,552)^2 + (402,047 - 387,168)^2} = 15,991$$

Untuk cluster ketiga dengan data pertama:

$$D(1,3) = \sqrt{(66,693 - 31,136)^2 + (402,047 - 182,848)^2} = 222,064 \quad (\text{Marsono; 2019 :}$$

34).

#### II.2.4. SQL Server

*Microsoft SQL Server* adalah produk database dari *Microsoft*. *Microsoft SQL Server* dapat disertakan dengan produk database server lain, seperti Oracle, IBM DB2, MySQL, dan sebagainya. Sejatinya, sebagai database server, *Microsoft SQL Server* harus diinstal pada sebuah komputer server. Dalam hal ini, komputer server tersebut menggunakan sistem operasi *Windows Server*. Namun untuk

keperluan development (pengembangan sistem aplikasi), kita bisa juga menginstal *SQL Server* edisi tertentu pada komputer desktop atau Pengertian *QR Code*. *QR code* adalah suatu jenis kode matriks atau kode batang dua dimensi yang dikembangkan oleh Denso Wave, sebuah divisi *Denso Corporation* yang merupakan sebuah perusahaan Jepang dan dipublikasikan pada tahun 1994 dengan fungsionalitas utama yaitu dapat mudah dibaca oleh pemindai, QR merupakan singkatan dari *quick response*, yang sesuai dengan tujuannya adalah untuk menyampaikan informasi dengan cepat dan mendapatkan respon yang cepat pula. Berbeda dengan kode batang, yang hanya menyimpan informasi secara horizontal, kode QR mampu menyimpan informasi secara horizontal dan vertikal, oleh karena itu secara otomatis Kode QR dapat menampung informasi yang lebih banyak dari pada kode batang. (M. Abu Jihad Plaza R;2015;106)

#### **II.2.5. Visual Studio**

Visual Studio merupakan suatu perangkat lunak yang dapat digunakan untuk pengembangan berbagai macam aplikasi yang memiliki berbagai macam tipe antara lain aplikasi *desktop* (*Windows Form*, *Command Line Console*, *Aplikasi Web*, *Windows Mobile (Paket PC)*). Visual studio ini dapat digunakan untuk membuat aplikasi yang berbasis desktop yang merupakan *platform windows*, namun juga dapat dijalankan dalam bentuk *Microsoft Intermediate Language* diatas *Net Framework*. Selain itu Visual Studio juga dapat digunakan untuk membuat aplikasi yang dapat dijalankan diatas *Windows Mobile* yang berjalan diatas *.Net Compact Framework*.

Visual Studio terbagi menjadi beberapa tipe diantaranya :

1. Visual Studio *Express Edition* yang bisa digunakan secara gratis tanpa memberikan royalti kepada *Microsoft Inc.*
2. Visual Studio *Standard Edition.*
3. Visual Studio *Professional Edition.*
4. Visual Studio *Ultimate Edition.*

**Tabel II.3. Jendela Aplikasi Visual Studio**

No	Bagian	Keterangan
1	<i>Menu Bar</i>	Menampilkan suatu baris menu yang mempunyai 11 menu utama, masing-masing memiliki sub menu dan peritnah lengkap dengan <i>shortcut key</i> .
2	<i>Toolbar Standar</i>	suatu baris menu yang mempunyai fungsi yang sama pada setiap <i>tool</i> standard pada umumnya. Seperti fungsi untuk menyimpan, meng- <i>copy</i> , menambah <i>project</i> baru, mengatur tampilan program dan masih banyak yang lainnya.
3	<i>Form Design</i>	suatu lembar <i>form</i> yang berfungsi untuk merancang tampilan aplikasi secara visual dengan menempatkan kontrol-kontrol yang diperlukan.
4	<i>Toolbox</i>	suatu jendela yang berfungsi untuk menampung komponen-komponen <i>standard</i> .
5	<i>Solution Explorer</i>	Menampilkan suatu jendela yang berfungsi untuk menampilkan objek yang akan anda gunakan untuk membuat aplikasi seperti <i>form</i> , <i>class</i> , dan objek lainnya.
6	<i>Properties Explorer</i>	Menampilkan suatu jendela yang berfungsi untuk

		mengatur nilai properties dari masing-masing komponen yang akan digunakan.
7	<i>Error List</i>	suatu jendela yang digunakan untuk menampilkan setiap kesalahan dari pembuatan kode program suatu aplikasi.

(Sumber : Rolly Yesputra ; 2017:19-22)

### II.2.6. *Unified Modeling Language (UML)*


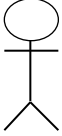

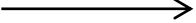
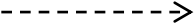
Hasil pemodelan pada OOAD terdokumentasikan dalam bentuk *Unified Modeling Language (UML)*. UML adalah satu alat bantu yang sangat handal didunia pengembangan sistem yang berorientasi objek. Hal ini disebabkan karena UML menyediakan bahasa pemodelan visual yang memungkinkan bagi pengembang sistem untuk membuat cetak biru atas visi mereka dalam membentuk yang baku, mudah dimengerti serta dilengkapi dengan mekanisme yang efektif untuk berbagi (*sharing*) dan mengkomunikasikan rancangan mereka dengan yang lain (Munawar ; 2018 : 49).

Alat bantu yang digunakan dalam perancangan berorientasi objek berbasis UML adalah sebagai berikut :

#### 1. *Use case Diagram*

*Use case diagram* merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* mendeskripsikan tipikal interaksi antara (pengguna) sebuah *system* dengan sistemnya sendiri melalui sebuah cerita bagaimana sebuah sistem yang dipakai (Munawar ; 2018 : 89).

Tabel II.2. Simbol *Use Case Diagram*

Gambar	Keterangan
	<p><i>Use Case</i> menggambarkan fungsionalitas yang disediakan sistem sebagai unit-unit yang bertukar pesan antar unit dengan aktor, biasanya dinyatakan dengan menggunakan kata kerja di awal nama <i>Use Case</i>.</p>
	<p>Aktor adalah <i>abstraction</i> dari orang atau sistem yang lain yang mengaktifkan fungsi dari target sistem. Untuk mengidentifikasi aktor, harus ditentukan pembagian tenaga kerja dan tugas-tugas yang berkaitan dengan peran pada konteks target sistem. Orang atau sistem bisa muncul dalam beberapa peran. Perlu dicatat bahwa aktor berinteraksi dengan <i>Use Case</i>, tetapi tidak memiliki <i>control</i> terhadap <i>Use Case</i>.</p>
	<p>Asosiasi antara aktor dan <i>Use Case</i>, digambarkan dengan garis tanpa panah yang mengindikasikan siapa atau apa yang meminta interaksi secara langsung dan bukannya mengindikasikan aliran data.</p>
	<p>Asosiasi antara aktor dan <i>Use Case</i> yang menggunakan panah terbuka untuk mengindikasikan bila aktor berinteraksi secara pasif dengan sistem.</p>
	<p><i>Include</i>, merupakan di dalam <i>Use Case</i> lain (<i>required</i>) atau pemanggilan <i>Use Case</i> oleh <i>Use Case</i> lain, contohnya adalah pemanggilan sebuah fungsi program.</p>

**Tabel II.2. Simbol Use Case Diagram(sambungan)**

←-----	<i>Extend</i> , merupakan perluasan dari <i>Use Case</i> lain jika kondisi atau syarat terpenuhi.
--------	---

(Sumber : Munawar ; 2018 : 893)

## 2. *Class Diagram* (Diagram Kelas)

Merupakan hubungan antar kelas dan penjelasan detail tiap-tiap kelas di dalam model desain dari suatu sistem, juga memperlihatkan aturan-aturan dan tanggung jawab entitas yang menentukan perilaku sistem.

*Class diagram* merupakan diagram statis dari suatu aplikasi. *Class Diagram* tidak hanya digunakan untuk memvisualisasikan, menggambarkan, dan mendokumentasikan berbagai aspek sistem tetapi juga untuk membangun kode eksekusi (*executable code*) dari aplikasi perangkat lunak (Munawar ; 2018 : 101).

**Tabel II.3. Simbol Class Diagram**



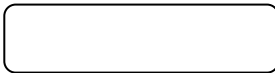
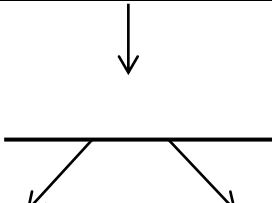
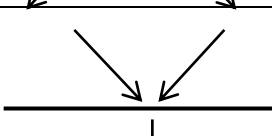
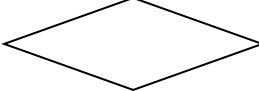

<b>Multiplicity</b>	<b>Penjelasan</b>
1	Satu dan hanya satu
0..*	Boleh tidak ada atau 1 atau lebih
1..*	1 atau lebih
0..1	Boleh tidak ada, maksimal 1
n..n	Batasan antara. Contoh 2..4 mempunyai arti minimal 2 maksimum 4

(Sumber :Munawar ; 2018 : 101)

### 3. Diagram Aktivitas (*Activity Diagram*)

*Activity Diagram* bagian penting dari UML yang menggambarkan aspek dinamis dari sistem. Logika prosedural, proses bisnis dan aliran kerja suatu bisnis bisa dengan mudah dideskripsikan dalam *activity diagram* (Munawar ; 2018 : 137).

**Tabel II.4. Simbol Diagram Aktivitas**

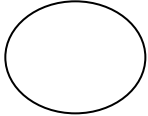
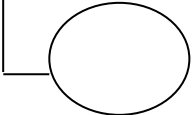
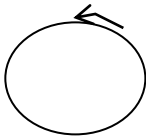

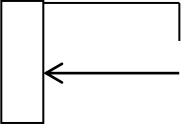
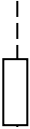
Gambar	Keterangan
	<i>Start point</i> , diletakkan pada pojok kiri atas dan merupakan awal aktifitas.
	<i>End point</i> , akhir aktifitas.
	<i>Activites</i> , menggambarkan suatu proses/kegiatan bisnis.
	<i>Fork</i> (Percabangan), digunakan untuk menunjukkan kegiatan yang dilakukan secara parallel atau untuk menggabungkan dua kegiatan paralel menjadi satu.
	<i>Join</i> (penggabungan) atau rake, digunakan untuk menunjukkan adanya dekomposisi.
	<i>Decision Points</i> , menggambarkan pilihan untuk pengambilan keputusan, <i>true</i> , <i>false</i> .
 New Swimline	<i>Swimlane</i> , pembagian <i>activity diagram</i> untuk menunjukkan siapa melakukan apa.

(Sumber : Munawar ; 2018 : 137)


#### 4. Diagram Urutan (*Sequence Diagram*)

*Sequence diagram* adalah salah satu *interaction diagram*. Karena *sequence diagram* mengacu kepada obyek, maka sbelum membuat diagram ini class diagram sudah harus teridentifikasi (Munawar ; 2018 : 186).

**Tabel II.5. Simbol Diagram Urutan**

Gambar	Keterangan
	<i>EntityClass</i> , merupakan bagian dari sistem yang berisi kumpulan kelas berupa entitas-entitas yang membentuk gambaran awal sistem dan menjadi landasan untuk menyusun basis data.
	<i>Boundary Class</i> , berisi kumpulan kelas yang menjadi <i>interface</i> atau interaksi antara satu atau lebih aktor dengan sistem, seperti tampilan <i>formentry</i> dan <i>form</i> cetak.
	<i>Control class</i> , suatu objek yang berisi logika aplikasi yang tidak memiliki tanggung jawab kepada entitas, contohnya adalah kalkulasi dan aturan bisnis yang melibatkan berbagai objek.
	<i>Message</i> , simbol mengirim pesan antar <i>class</i> .
	<i>Recursive</i> , menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri.
	<i>Activation</i> , mewakili sebuah eksekusi operasi dari objek, panjang kotak ini berbanding lurus dengan durasi aktivitas sebuah operasi.

**Tabel II.5. Simbol Diagram Urutan(sambungan)**

	<i>Lifeline</i> , garis titik-titik yang terhubung dengan objek, sepanjang <i>lifeline</i> terdapat <i>activation</i> .
---	---

(Sumber : Munawar ; 2018 : 186)