

## BAB II KERANGKA TEORITIS

### A. Landasan Teori

#### 1. Data Mining

Perkembangan teknologi informasi telah memberikan kontribusi pada cepatnya pertumbuhan jumlah data yang dikumpulkan dan disimpan dalam basis data berukuran besar (big data), big data adalah istilah yang menggambarkan volume data yang besar, baik data yang terstruktur maupun data yang tidak terstruktur, big data memiliki potensi tinggi untuk mengumpulkan wawasan kunci dari informasi bisnis, big data dapat dianalisis untuk wawasan yang mengarah pada pengambilan keputusan dan strategi bisnis yang lebih baik.

Sebuah metode atau teknik diperlukan untuk dapat merubah data tersebut menjadi sebuah informasi berharga atau pengetahuan yang bermanfaat untuk mendukung pengambilan keputusan, suatu teknologi yang dapat digunakan untuk mewujudkannya adalah data mining.

Belakangan ini *data mining* telah diimplementasikan kedalam berbagai bidang diantaranya dalam bidang bisnis atau perdagangan, bidang pendidikan, dan telekomunikasi.

Data *mining* adalah bidang penelitian dan praktik yang berfokus pada penemuan pola-pola baru dalam data yang mengacu pada penggunaan *algoritma* dan komputer untuk menemukan pola baru dan menarik dalam data. (Stanton, 2013, p173).

a) Tipe Data Mining:

1. Tipe Data Numerik adalah tipe data yang diperoleh dengan cara pengukuran, dimana jarak dua titik pada skala sudah diketahui.
2. Tipe Data Kategorial adalah tipe data yang diperoleh dengan cara kategorisasi atau klasifikasi.
3. Tipe Data Rentang Waktu adalah tipe yang diperoleh dengan cara menunjukkan beberapa objek yang berbeda.

b) Tahap-Tahap Data Mining

Menurut (Tan, 2004, p124), serangkaian proses tahapan memiliki tujuh (7) tahapan yaitu:

1. Pembersihan Data (data cleaning)

Pembersihan data adalah proses untuk menghilangkan data-data yang tidak relevan. Data-data yang dibuang terkadang dibandingkan terlebih

dahulu dengan hipotesa yang telah dibuat. Sehingga pada proses selanjutnya dapat dengan mudah menemukan hasil yang diinginkan.

## 2. Integrasi data (data integration)

Integrasi data merupakan proses dalam menggabungkan data dari beberapa database kedalam satu database baru. Tidak sedikit data yang dibutuhkan diambil dari berbagai database atau teks file.

## 3. Seleksi data (data selection)

Data yang sudah ada di database seringkali tidak semuanya dibutuhkan, maka dari itu dibutuhkan penyeleksian data untuk data yang benar-benar dibutuhkan dalam proses selanjutnya.

## 4. Transformasi data (data transformation)

Data digabung atau diubah sesuai dengan proses yang digunakan dalam data mining. Karena beberapa format data mining membutuhkan format data yang khusus dalam pemrosesannya.

## 5. Proses mining

Adalah proses menggali data dari sebuah database atau kumpulan data untuk memperoleh informasi yang tersembunyi dari data yang diolah

## 6. Evaluasi Pola (pattern evaluation)

Dalam proses ini adalah hasil dari teknik data mining berupa pola-pola yang akan diujia pada hipotesa yang sudah dibuat sebelumnya. Sehingga akan memperoleh kesimpulan-kesimpulan yang mendekati hasil atau hipotesa untuk proses selanjutnya.

## 7. Presentasi pengetahuan (knowledge presentation)

Ini termasuk dalam langkah akhir dari data mining dalam tahap ini saatnya untuk mempresentasikan hasil yang telah dilakukan dengan mengimplementasikan analisis yang didapat. Sehingga akan memperoleh kesimpulan real.

## 2. Clustering

*Clustering* merupakan proses partisi satu *set* objek data kedalam himpunan bagian yang di sebut dengan *cluster*. Objek yang di dalam *cluster* memiliki kemiripan karakteristik antara satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering*. Oleh karena itu *clustering* sangat berguna dan bisa menemukan *group* atau kelompok yang tidak di kenal dalam data. *Clustering* banyak di gunakan dalam berbagai *aplikasi* seperti misalnya pada *business intelligence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). Di dalam *business intelligence clustering* bisa mengatur

banyak *customer* kedalam banyaknya kelompok. Contohnya mengelompokan *customer* ke dalam beberapa *cluster* dengan kesamaan karakteristik yang kuat. *Clustering* juga dikenal sebagai data segmentasi karena *clustering* mempratiasi banyak data *set* kedalam banyak *group* berdasarkan kesamaan. Selain itu *clustering* juga bisa sebagai *outlier detection*. (Edy Irwansyah Muhamad Faisal, 2015, p4).

### 3. Business Process Model and Notation (BPMN)

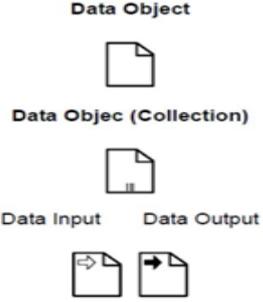
Menurut Yudho Yudhanto (2016: 1) *Business Process Modeling Notation* (BPMN) adalah notasi grafis yang menggambarkan logika dari langkah-langkah dalam proses bisnis. Notasi ini telah didesain secara khusus untuk mengkoordinasikan urutan proses dan pesan yang mengalir antara pelaku dalam kegiatan yang berbeda.

Tujuan utama dari usaha BPMN adalah menyediakan suatu notasi yang mudah dipahami oleh semua masyarakat terutama pegiat software. Dari analisis bisnis yang ada kemudian menciptakan draft permulaan dari proses-proses sampai dengan pengembangan teknis meliputi alur dan pekerjaan dalam bentuk model atau notasi. BPMN juga menciptakan suatu jembatan terstandarisasi untuk gap antara desain proses bisnis dan implementasi proses.

**Tabel 2. 1 Business Process Model and Notation (BPMN)**

Nama Simbol	Simbol	Keterangan
Event		sesuatu yang "terjadi" selama jalannya Proses atau Koreografi. Mempengaruhi aliran dari model dan bisanya memiliki penyebab (pemicu) atau dampak (hasil). <b>Event</b> digambarkan dalam lingkaran terbuka untuk membedakan fungsinya. Ada tiga jenis event, berdasarkan pengaruh aliran proses: Awal, Menengah, dan Akhir.

Nama Simbol	Simbol	Keterangan
Aktivitas	 Task  SubProcess	Sebuah istilah umum untuk suatu kegiatan yang memperlihatkan perusahaan melakukan Proses. Jenis Kegiatan yang merupakan bagian dari Proses sebuah Model Digambarkan bulat persegi panjang.
Gateway		Mengontrol perbedaan dan konvergensi dari urutan Arus dalam Proses. Dengan demikian, akan menentukan percabangan, forking, penggabungan, dan bergabung dengan jalur
Sequence Flow		Sebuah Arus Urutan digunakan untuk menunjukkan urutan kegiatan yang akan dilakukan dalam proses
Message Flow		Digunakan untuk menunjukan aliran Pesan antara dua pelaku yang telah dipersiapkan untuk mengirim dan menerima mereka.
Association		Digunakan untuk menghubungkan informasi dan Artefak dengan elemen BPMN grafis. Teks penjelasan dan Artefak lain dapat terkait dengan grafis elemen. Semua mata panah pada Asosiasi menunjukkan arah aliran (misalnya: data)

Nama Simbol	Simbol	Keterangan
Pool		<p>Representasi grafis dari pelaku/peserta kolaborasi. Hal ini juga bertindak sebagai "swimlane" dan wadah grafis untuk partisi satu set kegiatan dari Pools lain, biasanya dalam konteks situasi B2B. Pool A mungkin memiliki internal yang rinci, dalam bentuk proses yang akan dieksekusi.</p>
Lane		<p>Partisi sub-dalam Proses, terkadang dalam Pool, akan memperpanjang seluruh Proses baik secara vertikal ataupun horisontal. Jalur yang digunakan untuk mengatur dan mengategorikan Kegiatan.</p>
Data Object		<p>Memberikan informasi tentang kegiatan apa yang perlu diadakandan atau apa yang mereka hasilkan. Daata Object dapat mewaakili benda tunggal atau koleksi benda-benda. Data input dan Data Output memberikan informasi yang sama untuk Proses.</p>
Group		<p>Pengelompokan unsur-unsur grafis yang berada dalam kategori yang sama. Jenis pengelompokan tidak mempengaruhi Sequence Flow dalam Group. Nama</p>

Nama Simbol	Simbol	Keterangan
		Kategori muncul pada diagram sebagai label kelompok.

#### 4. Unified Modelling Language (UML)

*Unified Modelling Language* (UML) adalah sebuah "bahasa" pemodelan untuk sistem atau perangkat yang berparadigma "berorientasi obyek" yang telah menjadi standar dalam industri untuk *visualisasi*, merancang dan mendokumentasikan sistem perangkat lunak. UML menawarkan sebuah standar untuk merancang model sebuah sistem. Dengan menggunakan UML dapat dibuat model untuk semua jenis aplikasi perangkat lunak. Aplikasi tersebut dapat berjalan pada perangkat keras, sistem operasi dan jaringan apapun, serta ditulis dalam bahasa pemrograman apapun. Tetapi karena UML juga menggunakan *class* dan *operation* dalam konsep dasarnya, maka lebih cocok untuk penulisan perangkat lunak dalam bahasa-bahasa berorientasi objek seperti C++, Java, C# atau VB.NET. Walaupun demikian, UML tetap dapat digunakan untuk modeling aplikasi prosedural dalam VB atau C. Tahap awal adalah perencanaan (*planning*), yaitu menyangkut studi tentang kebutuhan pengguna (*user's spesification*), studi-studi kelayakan (*feacibility study*) serta penjadwalan pengembangan. Pada tahap ini *tool* UML menggunakan *use case* diagram untuk menangkap kebutuhan dan harapan pengguna. Tahap kedua adalah analisis (*analysis*), yaitu tahap untuk mengenali segala permasalahan yang muncul pada pengguna dengan merealisasikan *use case* diagram lebih lanjut, mengenali komponen sistem/perangkat lunak, obyek-obyek, dan hubungan antara obyek. Tahap ketiga adalah perancangan (*design*), yaitu mencari solusi permasalahan yang didapat dari tahap analisis. Dalam tahap ini dilakukan penambahan dan modifikasi kelas-kelas yang akan lebih mengefisisensikan dan mengefektifkan sistem/perangkat lunak yang akan dikembangkan. Tahap keempat adalah implementasi (*implementation*), yaitu mengimplementasikan perancangan sistem ke situasi nyata. Tahap kelima adalah pengujian (*testing*), yaitu melakukan pengujian untuk menentukan apakah sistem/perangkat lunak yang kita buat sudah sesuai dengan kebutuhan pengguna atau belum. Keuntungan menggunakan metodologi berorientasi obyek mulai dari analisis sampai implementasi menggunakan *tool* UML yang sama adalah proses *iterative* bisa berjalan dengan efisien serta lebih efektif ditinjau dari segi uang dan waktu (Nugroho, 2009, p7).

a) Notasi UML

Notasi UML merupakan simbol yang digunakan untuk pembuatan diagram. Beberapa notasi yang digunakan diantaranya *actor*, *use case*, *association*, *generalization*, *note*, *class*, *interface*, *interaction*, *realization*, *dependency*, dan *package*. Setiap notasi yang digunakan disesuaikan dengan diagram yang digunakan. Setiap diagram tentu akan menggunakan notasi yang berbeda).

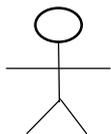
b) Diagram UML

Diagram UML terdiri dari *use case diagram*, *class diagram*, *state chart diagram*, *activity diagram*, *sequence diagram*, *collaboration diagram*, *component diagram*, dan *deployment diagram*.

1) Use Case Diagram

*Use Case Diagram* menjelaskan urutan kegiatan yang dilakukan *actor* dan *system* untuk mencapai suatu tujuan tertentu, walaupun menjelaskan kegiatan namun *use case* hanya menjelaskan apa yang dilakukan oleh *actor* dan *system*, bukan bagaimana *actor* dan *system* melakukan kegiatan tersebut. (Julius Hermawan, 2004, p16). Dimana manfaat dari *use case* itu sendiri antara lain memberikan kepastian pemahaman yang pas tentang pemetaan atau kebutuhan sebuah *system* serta dapat mengidentifikasi siapa yang sedang berinteraksi dengan *system* dan juga apa yang harus dilakukan untuk *system* tersebut.

**Tabel 2. 2 Use Case Diagram**

Notasi	Keterangan	Simbol
Actor	Pengguna sistem atau yang berinteraksi langsung dengan sistem, misalnya manusia, aplikasi atau objek lain	
Use Case	Digambarkan dengan lingkaran elips dengan nama use case nyata tertulis di tengah lingkaran	
Association	Digambarkan dengan sebuah garis yang berfungsi menghubungkan actor dengan use case	

2) Class Diagram

Sama seperti *class*, *class diagram* merupakan diagram yang selalu ada di permodelan *system* berorientasi obyek. *Class diagram* menunjukkan hubungan

antara *class* dalam *system* yang sedang di bangun dan bagaimana mereka saling berkolaborasi untuk mencapai suatu tujuan.

*Class diagram* umumnya tersusun dari elemen *class*, *Interface*, *Dependency*, *Generalization* dan *Association*. Relasi *dependency* menunjukkan bagaimana ketergantungan terjadi antara *class* yang ada. Relasi *Generalization* menunjukkan bagaimana suatu *class* menjadi super *class* dari *class* lainnya dan *class* yang lain menjadi sub-*class* dari *class* tersebut. Relasi *Association* menggambarkan navigasi antara *class*, berapa banyak obyek lain bisa berhubungan dengan satu obyek dan apakah suatu *class* menjadi bagian dari *class* lainnya. *Class diagram* digunakan untuk menggambarkan *desain statis* dari *system* yang sedang dibangun (Julius Hermawan, 2004, p27). Berdasarkan pengertian dari *class diagram*, dengan *class diagram* dapat dapat dibuat bangun secara terperinci dan jelas, dengan cara meperharikan kode spesifikasi apa saja yang dibutuhkan oleh program, hal ini mampu mengimplementasikan ke struktur yang dijelaskan.

**Tabel 2. 3 Class Diagram**

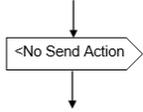
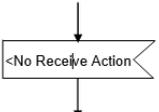
Notasi	Keterangan	Simbol
Class	Blok –blok pembangunprogram. Bagian atas classmenunjukkan nama dari class,bagian tengah mengindikasikanatribut dari class, dan bagianbawah mendefinisikan methoddari sebuah class	
Association	Menunjukkan relationship atauhubungan antar class	
Dependency	Menunjukkan ketergantunganantara satu class dengan classyang lain	
Generezation	Menunjukkan inheritance darisatu class ke beberapa class	

### 3) Activity Diagram

*Activity Diagram* merupakan state diagram khusus, dimana sebagian besar state adalah action dan sebagian transisi di trigger oleh selesainya state sebelumnya (*internal processing*). Oleh karena itu *Activity Diagram* tidak menggambarkan *behaviour* internal sebuah sistem (dan interaksi antar subsistem) secara eksak, tetapi lebih menggambarkan proses-proses dan jalur-jalur aktivitas dari level atas secara umum (Yuni Sugiarti, 2013, p58).

Berdasarkan pengertiannya *activity diagram* dapat menjelaskan urutan *activity* dalam suatu proses sehingga memudahkan memahami proses yang adadalam *system* secara keseluruhan, serta mengetahui aktivitas dari *actor* berdasarkan *use case* yang telah di buat sebelumnya.

**Tabel 2. 4 Activity Diagram**

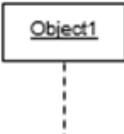
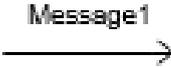
Notasi	Keterangan	Symbol
Initial State	Titik awal untuk suatu aktivitas	
Final State	Titik akhir untuk mengakhiri aktivitas	
Activity	Menandakan sebuah aktivitas	
Decision	Pilihan untuk mengambil keputusan	
Fork / Join	Menunjukkan kegiatan menggabungkan dua panel activity menjadi satu atau satu panel activity menjadi dua	
<i>Send</i>	<i>Tanda pengiriman</i>	
Receive	Tanda penerimaan	
Note	Catatan khusus untuk aktivitas	
Control Flow	Arus aktivitas	

#### 4) Sequence Diagram

Sequence diagram menjelaskan secara detail urutan proses yang dilakukan dalam *system* untuk mencapai tujuan dari *use case* interaksi yang terjadi antara *class*, operasi apa saja yang terlibat, urutan antar operasi, dan informasi yang

diperlukan oleh masing-masing operasi. Pembuatan Sequence Diagram merupakan aktifitas yang paling kritical dari proses *desain* karena artifak ini yang menjadi pendoma dalam proses *pemrograman* nantinya dan berisi aliran *control* dari program. (Julius Hermawan, 2004, p25). Sequence Diagram bisa digunakan untuk menggambarkan *scenario* atau rangkaian langkah yang di lakukan berdasarkan respon terhadap sebuah *event* untuk menghasilkan output tertentu, dan perubahan apa saja yang terjadi secara internal serta output apa yang di hasilkan.

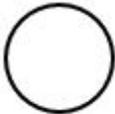
**Tabel 2. 5 Sequence Diagram**

Notasi	Keterangan	Simbol
Object	Instance dari sebuah class yang dituliskan tersusun secara horizontal diikuti lifeline	
Activation	Indikasi dari sebuah objek yang melakukan suatu aksi	
Lifeline	Indikasi keberadaan sebuah objek dalam basis waktu	
Message	Indikasi untuk komunikasi antar objek	
Self – Message	Komunikasi kembali ke dalam itu sendiri	

#### 5) Component Diagram

Component diagram menurut Munawar (2005, p119), mempresentasikan Duniariil item yaitu component software. Component software adalah bagian fisik 22 dari sebuah sistem karena menetap di komputer. Component diagram mengandung component, iterface dan relationship.

Tabel 2. 6 Component Diagram

Notasi	Simbol	Deskripsi
Package		<p>Package merupakan sebuah simbol yang dipakai untuk tempat komponen.</p>
Component		<p>Component system adalah simbol yang menjelaskan perangkat keras atau objek dalam sistem tersebut.</p>
Dependency		<p>Simbol yang menjelaskan sebuah keterkaitan antara komponen, satu komponen dengan yang lain. Arah panah dalam simbol tersebut diarahkan pada komponen yang dipakai.</p>
Interface		<p>Hal ini mirip dengan bahasa pemrograman berorientasi objek (PBO), dimana simbol ini dipakai untuk antar muka dengan fungsi supaya tidak langsung mengakses objek.</p>
Link		<p>Simbol link ini dipakai untuk mengarahkan relasi antar komponen, jika suatu komponen memiliki relasi atau keterkaitan dengan komponen lainnya maka dipakailah simbol link ini.</p>

6) Deployment Diagram

Menurut (Irmayani, Susyatih, 2017) “Deployment Diagram Menggambarkan hubungan antara software dan hardware terhadap sistem dan apa saja output yang dihasilkan. Software yang digunakan antara lain xampp sebagai server, phpmyadmin sebagai database dan NetBeans IDE 8.1 sebagai editor bahasa pemrograman java”.

**Tabel 2. 7 Deployment Diagram**

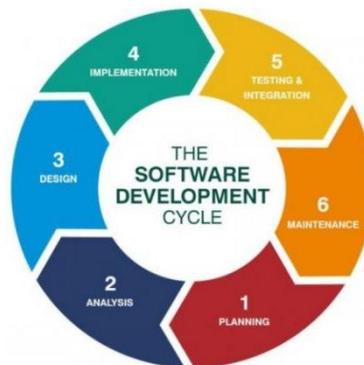
Notasi	Simbol	Deskripsi
Package		Package adalah sebuah bungkusan dari satu atau lebih node.
Node		Node biasanya mengacu pada hardware, perangkat lunak yang tidak dibuat sendiri (software). Jika didalam node disertai komponen untuk mengkonsistenkan rancangan maka komponen yang diikutsertakan harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen.
Dependency		Dependency adalah kebergantungan antar node, arah panah mengarah pada node yang dipakai.
Link		Link merupakan relasi antar node

## 5. Mysql

MySQL merupakan RDBMS (atau server database) yang mengelola database dengan cepat menampung dalam jumlah sangat besar dan dapat di akses oleh banyak user (Raharjo, 2011, p21).

## 6. System Development Life Cycle (SDLC)

Pendekatan sistem merupakan metodologi. Dimana menurut (Mcleod, 2007, p188) metodologi secara definisi ialah suatu jalan atau cara yang direkomendasikan dalam melakukan sesuatu, dan pendekatan sistem ialah metodologi dasar untuk pemecahan berbagai macam permasalahan, dan siklus hidup pengembangan sistem adalah suatu alplikasi dari pendekatan sistem untuk pengembangan suatu sistem informasi”.



(Mcleod Raymond, Schell George, 2007, p188)

**Gambar 2. 1 SDLC Konsep Mcleod Raymond, Schell George**

Secara tahap SDLC traditional ada tampak seperti diatas dimana semua tahap-tahap itu memiliki langkah-langkah perencanaan yang berbeda-beda, dimana setiap tahap akan selalu terkait dengan tahap-tahap yang lain, setiap langkah yang ada pada konsep pengembahan sistem daengan menggunakan *System Development Life Cycle* (SLDC) memang sangat praktis dan mudah dalam menyusun setiap langkah demi langkah akan tetapi kita harus mengetahui secara tepat konsep dan tujuan setiap langkah tersebut agar penyusunan dapat sesuai dengan yang diharapkan (Mcleod Raymond, Schell George, 2007, p188).

## 7. Structure Query Language (SQL)

Menurut (Prasetio, 2012, p182) “SQL (dibaca “ess-que-el”) merupakan kependekan dari Structured Query Language. SQL digunakan untuk berkomunikasi dengan sebuah Database. Sesuai ANSI, SQL merupakan bahasa standar untuk sistem manajemen database relasional”. Dapat disimpulkan bahwa SQL (Structured Query Language) merupakan struktur bahasa yang digunakan untuk membangun basis data dan digunakan untuk berkomunikasi dengan database.

## 8. Web Server

Pada umumnya web server berperan sebagai server yang memberikan layanan kepada komponen yang meminta informasi berkaitan dengan web, dalam web yang telah dirancang dalam internet.

Menurut (Kustiyahningsih dan Devie, 2011, p8)” web server adalah komputer yang digunakan untuk menyimpan dokumen-dokumen web, komputer ini melayani permintaan dokumen web dari kliennya”.

## 9. Bahasa Pemrograman

Programming language (bahasa pemrograman) merupakan suatu sintak untuk mendefinisikan program komputer, bahasa ini memungkinkan seorang programmer dapat membuat suatu program aplikasi (Dipraja, 2014, p26).

Menurut tingkat kedekatannya dengan mesin komputer, bahasa pemrograman terdiri dari:

1. Bahasa Mesin, yaitu memberikan perintah kepada komputer dengan memakai kode bahasa biner, contohnya 01100101100110;
2. Bahasa Tingkat Rendah, atau dikenal dengan istilah bahasa rakitan (bah.Ingggris Assembly), yaitu memberikan perintah kepada komputer dengan memakai kode-kode singkat (kode mnemonic), contohnya kode\_mesin|MOV, SUB, CMP, JMP, JGE, JL, LOOP;
3. Bahasa Tingkat Menengah, yaitu bahasa komputer yang memakai campuran instruksi dalam kata-kata bahasa manusia (lihat contoh Bahasa Tingkat Tinggi di bawah) dan instruksi yang bersifat simbolik, contohnya {, }, ?, <<, >>, &&, ||;
4. Bahasa Tingkat Tinggi, yaitu bahasa komputer yang memakai instruksi berasal dari unsur kata-kata bahasa manusia, contohnya begin, end, if, for, while, and, or, dsb.Komputer dapat mengerti bahasa manusia itu diperlukan program compiler atau interpreter.

Sebagian besar bahasa pemrograman digolongkan sebagai Bahasa Tingkat Tinggi, hanya bahasa C yang digolongkan sebagai Bahasa Tingkat Menengah dan Assembly yang merupakan Bahasa Tingkat Rendah

Contoh Bahasa Pemrograman:

### a. Hypertext Preprocessor (PHP)

PHP sering dipakai para programmer untuk membuat situs web yang bersifat dinamis karena gratis dan berguna dalam merancang aplikasi web. PHP (PHP: Hypertext Preprocessor) adalah suatu bahasa pemrograman yang digunakan untuk menerjemahkan baris kode program menjadi kode mesin yang dapat dimengerti oleh komputer yang berbasis server-side yang dapat ditambahkan ke dalam HTML. (Supono dan Putratama, 2016, p3),

hypertext preprocessor (PHP) merupakan bahasa pemrograman untuk membuat/mengembangkan aplikasi berbasis web dan bersifat open source dan ditanamkan ke dalam script HTML.

b. Hypertext Markup Language (HTML)

Proses tampilnya sebuah halaman website di browser melibatkan HTML. HyperText Markup Language (HTML) tergolong dalam salah satu format yang digunakan dalam pembuatan dokumen yang terbaca oleh web.

Menurut (Prasetio, 2010, p4) mengemukakan bahwa HTML merupakan “bahasa pemrograman yang digunakan untuk mendesain sebuah halaman web”.

Berdasarkan teori dari para ahli di atas, maka hypertext markup language (HTML) merupakan bahasa pemrograman yang dikenal oleh browser untuk menampilkan informasi lebih menarik di halaman web melalui web browser

## 10. Intranet

Pengertian intranet adalah sebuah jaringan di dalam sebuah organisasi yang menggunakan teknologi internet (seperti web browser dan server, TCP / IP protokol jaringan, penerbitan dokumen HTML hypermedia dan database, dan lainnya) untuk menyediakan lingkungan internet dalam perusahaan untuk berbagi informasi, komunikasi, kolaborasi, dan dukungan dari proses bisnis (O'Brien dan Marakas, 2011, p229).

## B. Metode K-Means

K-means merupakan algoritma klasterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya. Menurut (Suyanto, 2017, p262) Algoritma k-means bekerja dengan empat langkah, yaitu:

1. Himpunan data yang akan diklasterisasi, dipilih sejumlah k objek secara acak sebagai *centroid* awal.
2. Setiap objek yang bukan *centroid* dimasukkan ke *cluster* terdekat berdasarkan ukuran jarak tertentu.
3. Setiap *centroid* diperbarui berdasarkan rata-rata dari objek yang ada di dalam setiap *cluster*.
4. Langkah kedua dan ketiga dilakukan secara diulang-ulang (diiterasi) sampai semua *centroid* stabil dalam arti semua *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid* yang dihasilkan pada iterasi sebelumnya.

a) Berikut ini adalah langkah-langkah algoritma K-Means:

1. Penentuan *cluster* awal

Dalam menentukan n buah pusat *cluster* awal dilakukan pembangkitan bilangan *random* yang merepresentasikan urutan data *input*, pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan *random* pusat awal dari data.

2. Perhitungan jarak dengan pusat *cluster*

Untuk mengukur jarak antara data pusat dengan *cluster* digunakan *euclidian distance*, algoritma perhitungan jarak data dengan pusat *cluster*:

- a. Pilih nilai data dan nilai pusat *cluster*
- b. Hitung *euclidian distance* data dengan tiap pusat *cluster*

$$(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \dots (1)$$

Penjelasan:

$x_i$ : Data kriteria

$\mu_j$ : *Centroid* pada *cluster* ke j

3. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Adapun cara pengelompokan data tersebut adalah:

- a. Pilih nilai jarak tiap pusat *cluster* dengan data
- b. Cari nilai jarak terkecil.
- c. Kelompokkan data dengan pusat *cluster* yang memiliki jarak terkecil.

4. Penentuan pusat *cluster* baru untuk mendapatkan pusat *cluster* baru

bisa dihitung dari rata-rata nilai anggota *cluster* dan pusat *cluster*, pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum *konvergen* proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah *konvergen* (pusat *cluster* baru sama dengan pusat *cluster* lama). Algoritma penentuan pusat *cluster* :

- a. Cari jumlah anggota tiap *cluster*
- b. Hitung pusat baru dengan rumus

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{x_j \in S_j} x_j \dots (2)$$

Penjelasan:

$\mu_j(t+1)$ : *Centroid* baru pada iterasi ke 1

$N_{sj}$ : Banyak data pada *cluster*  $s_j$

Hasil dari operasi *clustering* yang terbentuk selanjutnya akan di evaluasi menggunakan *Davies bouldin index* yang dihitung dengan persamaan:

$$DBI = \frac{1}{K} \sum_{i=1}^K \frac{R_i}{\min_{j=1, \dots, k, i \neq j} R_{ij}} = \frac{\sum_{i=1}^K \text{var}(C_i)}{\sum_{i=1}^K \|c_i - c_j\|}$$

Dimana:

$C_i = \text{Cluster } i$  dan cluster  $c_i$  adalah *centroid* dari *cluster*  $i$

5. Kelebihan dan Kekurangan Metode K-Means:

Sebagai fungsi penambangan data, analisis *cluster* dapat digunakan sebagai alat yang berdiri sendiri untuk memperoleh wawasan ke dalam distribusi data.

Adapun keuntungan lain dari metode ini (Han, dkk, 2012, p445) antara lain:

1. K-means juga disebut *segmentasi* data di beberapa aplikasi karena pengelompokan mempartisi *set* data besar ke dalam grup sesuai dengan kemiripannya.
2. K-means bisa juga digunakan untuk deteksi *outlier* (nilai yang "jauh" dari mana pun *cluster*).
3. K-means mempartisi sekumpulan objek data (atau pengamatan) ke dalam himpunan bagian, sehingga banyak digunakan dalam banyak aplikasi seperti 22 *intelijen* bisnis, pengenalan pola gambar, pencarian web, biologi, dan keamanan.

Selain itu, metode *clustering* memiliki beberapa kekurangan (Suyanto, 2017, p262) antara lain:

1. K-means tidak dapat menjamin *konvergen* pada *optimum* global.
2. K-means sering terjebak pada *optimum* lokal, dimana *centroid* akhir yang dihasilkan tidak benar-benar menjadi pusat *cluster* yang sesungguhnya.
3. Keluaran dari k-means bergantung pada *centroid* awal yang ditentukan secara acak.

b) Contoh Perhitungan K-Means

Perhitungan dari algoritma k-means berdasarkan dataset pada Tabel 2.8 dibawah ini (Abdurrahman, 2016, p2).

**Tabel 2. 8 Dataset**

Data	Jumlah Transaksi	Total Belanja
1	4	2.255
2	2	1.788
3	2	1.255
4	4	3.322

Data	Jumlah Transaksi	Total Belanja
5	4	3.012
6	3	1.261

Langkah pertama, ditentukan terlebih dahulu jumlah *cluster*, misalnya  $k=3$  dan ditentukan *centroid* (pusat *cluster*) awalnya adalah  $C1=(2,1)$ ;  $C2=(4,3)$ ,  $C3=(3,2)$ .

Dengan menggunakan Tabel 2.8, maka dihitung jarak *Euclidean* setiap item (Jumlah Transaksi dan Total Belanja) dengan *centroid* awal, maka diperoleh:

Maka data-1, diperoleh:

$$C1 = \sqrt{(4-2)^2 + (2.255-1)^2} = 2.36$$

$$C2 = \sqrt{(4-4)^2 + (2.255-3)^2} = 0.745$$

$$C3 = \sqrt{(4-3)^2 + (2.255-2)^2} = 1.745$$

dan seterusnya, sehingga diperoleh jarak antara data dengan *centroid* awal seperti pada Tabel 2.9

**Tabel 2. 9 Jarak Euclidean**

Data ke-i	C1	C2	C3
1	2.361149085	0.745	1.032000484
2	0.788	2.338577	1.022225024
3	0.255	2.654247	1.247006415
4	3.064585453	0.322	1.657613948
5	2.836925096	0.012	1.422724148
6	1.033499395	2.006021	0.739

Dari Tabel 2.9 dipilih jarak *Euclidean* terkecil, sehingga diperoleh *matriks* penempatan *cluster* seperti tampak pada Tabel 2.10 berikut ini.

**Tabel 2. 10 Matriks Penempatan Cluster**

Data ke-i	C1	C2	C3
1		*	
2	*		
3	*		
4		*	
5		*	
6			*

Ditentukan *centroid* baru yakni  $(C1=(C11,C12), C2=(C21,C22), C3=(C31,C32))$  dengan melihat *matriks* penempatan *cluster* dan menggunakan data pada Tabel 2.7

$$C11 = \text{Average}(2,2) = 2$$

$$C12 = \text{Average}(1.788, 1.255) = 1.5215$$

$$C21 = \text{Average}(4,4,4) = 4$$

$$C22 = \text{Average}(2.255, 3.322, 3.012) = 2.863$$

$$C31 = \text{Average}(3) = 3$$

$$C32 = \text{Average}(1.261) = 1.261$$

Sehingga *centroid* baru tersebut adalah:

$$C1 = (C11, C12) = (2, 1.5215)$$

$$C2 = (C21, C22) = (4, 2.863)$$

$$C3 = (C31, C32) = (3, 1.261)$$

Dengan menggunakan *centroid* baru, ditentukan kembali jarak setiap item terhadap *centroid* baru tersebut, sehingga diperoleh data jarak terhadap *centroid* baru seperti terlihat pada Tabel 2.9

**Tabel 2. 11 Jarak Euclidean**

Data ke-i	C1	C2	C3
1	2.130263423	0.608	1.409977305
2	0.2665	2.2706	1.130366755
3	0.2665	2.566255	1.000018
4	2.691059317	0.459	2.290790475
5	2.494311578	0.149	2.016432741
6	1.033373238	1.888493	0

Dari Tabel 2.11 dipilih jarak *Euclidean* terkecil sehingga diperoleh *matriks* penempatan *cluster* seperti tampak pada Tabel 2.12

**Tabel 2. 12 Matriks Penempatan Cluster**

Data ke-i	C1	C2	C3
1		*	
2	*		
3	*		
4		*	
5		*	
6			*

Dengan membandingkan Tabel 2.10 dan Tabel 2.12 ternyata tidak terjadi perpindahan *centroid*, sehingga iterasi *clustering* dihentikan.

c) *Silhouette Coefficient*

*Silhouette Coefficient* merupakan metode yang digunakan untuk melihat kualitas dan kekuatan dari *cluster*. Metode *Silhouette Coefficient* merupakan gabungan dari dua metode yaitu metode kohesi yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster*, dan metode separasi yang berfungsi untuk mengukur seberapa jauh sebuah *cluster* terpisah dengan *cluster* lain. Tahapan perhitungan *Silhouette Coefficient*.

1. Hitung rata-rata jarak dari suatu objek, misalkan  $i$  dengan semua objek lain yang berada dalam satu cluster dengan menggunakan rumus dibawah ini :

$$a_i = \frac{1}{|A|-1} \sum_{j \in A, i \neq j} d(i, j)$$

Keterangan:

$|A|$  = Banyaknya data dalam *cluster* A

$i, j$  = indeks dari kolom

$d(i, j)$  = jarak antara dokumen  $i$  dengan dokumen ke  $j$ .

2. Hitung rata-rata jarak dari dokumen  $i$  tersebut dengan semua dokumen di cluster lain menggunakan rumus berikut :

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)$$

Keterangan:  $d(i, C)$  = jarak rata-rata objek  $i$  dengan semua objek pada cluster lain.

3. Hitung nilai *Silhouette Coefficient*-nya dengan rumus berikut :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Kriteria subjektif pengukuran pengelompokkan berdasarkan *Silhouette Coefficient* menurut (Kaufman dan Rousseeuw, 1990, p17);

**Tabel 2. 13 Keterangan Nilai Silhouette Coefficient**

NO	Rentang Nilai Silhouette Coefficient (SC)	Keterangan
1	$0,7 < SC \leq 1$	Strong Structure
2	$0,5 < SC \leq 0,7$	Medium Structure
3	$0,25 < SC \leq 0,5$	Weak Structure
4	$SC \leq 0,25$	No Structure

## **C. Pemetaan wilayah layanan data internet**

### **1. Internet**

Internet adalah komunikasi jaringan komunikasi global yang menghubungkan seluruh komputer di dunia meskipun berbeda sistem operasi dan mesin (Hermawan, 2013, p68).

### **2. Layanan**

Layanan adalah suatu tindakan sukarela dari satu pihak ke pihak lain dengan tujuan hanya sekedar membantu, atau adanya permintaan kepada pihak lain untuk memenuhi kebutuhannya secara sukarela, pelayanan adalah aspek yang tidak bisa disepelekan dalam persaingan bisnis manapun, karena dengan pelayanan konsumen akan menilai kemudian menimbang apakah selanjutnya dia akan *loyal* kepada pemberi layanan tersebut, sehingga tidak jarang para pebisnis memaksimalkan layanannya untuk menarik konsumen sebesar-besarnya, maka dari itu bila ingin menarik konsumen dengan sebanyak-banyaknya harus mengetahui arti dari layanan itu sendiri. Pengertian layanan atau pelayanan secara umum, menurut Purwadarminta adalah menyediakan segala apa yang dibutuhkan orang lain, sedangkan menurut Tjiptono definisi layanan adalah kegiatan yang dilakukan perusahaan kepada pelanggan yang telah membeli produknya, sedangkan menurut Barata bahwa suatu pelayanan akan terbentuk karena adanya proses pemberian layanan tertentu dari pihak penyedia layanan kepada pihak yang dilayani. Ada beberapa ciri pelayanan yang baik yang dapat memberikan kepuasan kepada pelanggan, (a) Memiliki karyawan yang profesional khususnya yang berhadapan langsung dengan pelanggan, (b) Tersedianya sarana dan prasarana yang baik yang dapat menunjang kelancaran produk ke pelanggan secara cepat dan tepat, (c) Tersedianya ragam produk yang diinginkan, (d) Bertanggung jawab kepada setiap pelanggan dari awal hingga selesai, (e) Mampu melayani secara cepat dan tepat, tentunya jika dibandingkan dengan pihak pesaing, (f) Mampu berkomunikasi dengan jelas, menyenangkan dan mampu menangkap keinginan dan kebutuhan pelanggan, (g) Memberikan jaminan kerahasiaan setiap transaksi, terutama dalam hal keuangan, (h) Memiliki

pengetahuan dan kemampuan yang baik tentang produk yang dijual dan pengetahuan umum lainnya, (i) Mampu memberikan kepercayaan kepada pelanggan, sehingga pelanggan merasa yakin dengan apa yang telah dilakukan perusahaan (Purwadarminto, 1996, p245).

#### **D. Tinjauan Pustaka**

Dalam penelitian ini, permasalahan yang akan diangkat adalah mengenai Penerapan Metode K-means untuk Pemetaan Wilayah Layanan Telekomunikasi, berdasarkan permasalahan tersebut, diperoleh 10 (Sepuluh) penelitian lain yang berhubungan dengan penelitian ini. Berikut ini penelitian mengenai Penerapan Metode K-means yang menjadi tinjauan studi dalam penelitian ini, diantaranya:

1. Penelitian yang dilakukan oleh Suhandio Handoko pada tahun 2020 dengan judul, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING" Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Perkembangan industri telekomunikasi saat ini sangat pesat karena telekomunikasi sudah menjadi kebutuhan utama bagi masyarakat sehingga banyak perusahaan yang bergerak di industry telekomunikasi. Banyaknya *industry* Telekomunikasi menuntut para pengembang untuk menemukan strategi atau suatu pola yang dapat meningkatkan penjualan dan pemasaran produk, salah satu strateginya adalah dengan memanfaatkan data transaksi. Paket data merupakan produk dibidang telekomunikasi. Proses *clustering* saat ini masih di lakukan secara manual sehingga membutuhkan waktu, *proses* perhitungan dan ketelitian yang tinggi. Pada penelitian ini dibuat aplikasi berbasis website dengan tujuan untuk mempermudah *clustering* data sehingga dapat digunakan sebagai referensi dalam perencanaan promosi produk telkomsel ke berbagai daerah. Metode yang digunakan untuk mengatasi permasalahan tersebut yaitu metode *clustering* dengan menggunakan Algoritma K-Means. Algoritma K-Means merupakan algoritma pengelompokkan sejumlah data menjadi menjadi kelompok-kelompok data tertentu. Pada penelitian ini data penjualan dikelompokkan menjadi 3 yaitu data penjualan rendah, data penjualan sedang dan data penjualan tinggi. Pengujian clustering dengan algoritma K-Means pada aplikasi terhadap data transaksi penjualan paket telkomsel diperoleh persentase kesesuaian yaitu 100% dibandingkan dengan *clustering* manual.
2. Penelitian yang dilakukan oleh Tri Ginanjar Laksana pada tahun 2017 dengan judul, " ANALISA POLA PEMILIHAN PROVIDER TELEKOMUNIKASI

BERDASARKAN PEKERJAAN DENGAN ALGORITMA KLUSTERING K – MEANS” Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Tujuan dari penelitian ini, diharapkan menghasilkan analisa pola pemilihan provider telekomunikasi berdasarkan hasil pengelompokan kemiripan nilai *cluster* dari data set primer yang diperoleh dari hasil kuesioner pertanyaan kepada 200 pengguna provider telekomunikasi. Selain hal tersebut, tujuan dari penelitian ini dapat digunakan sebagai pedoman penggambaran (analisa) distribusi pola pemilihan jenis *provider* telekomunikasi berdasarkan kondisi daerah tempat tinggal, jenis kelamin dan kepemilikan merk handphone. Metode yang digunakan dalam penelitian ini adalah algoritma *clustering* teknik k-means, dimana k-means memiliki kemampuan *cluster* yang mana tahap pertama, dilakukan dengan memilih nilai K secara acak, K adalah jumlah banyaknya *cluster* yang terbentuk, yaitu jenis *provider* yang telah dipilih oleh pengguna, yang didasarkan kepada *cluster* lingkungan tempat tinggal (daerah), jenis kelamin, dan *cluster merk handphone*. Kemudian algoritma k-means yang berisi data *cluster* yang memiliki tingkat kemiripan (*similarity*) yang sering muncul pada *cluster* yang sama pada tingkat kemiripan yang kurang pada *cluster* yang berbeda. Hasil dari penelitian ini berupa acuan penggambaran pola pemilihan pengelompokan berdasarkan 3 (Tiga) *cluster*, untuk memperoleh nilai *presentasi cluster yang optimal* maka dilakukan perhitungan *performance vektor* perhitungan *GiniCoofficient* dan *SumOfSquare*.

3. Penelitian yang dilakukan oleh BAGAS PRAYOGO HARYAJI pada tahun 2018 dengan judul, ” PENERAPAN ALGORITMA K-MEANS UNTUK MEMETAKAN GARIS KEMISKINAN MENURUT PROVINSI DI INDONESIA”. Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Indonesia memiliki jumlah penduduk yang cukup besar, Badan Pusat Statistik di tahun 2012 mencatat ada 250.000.000 lebih penduduk yang tersebar di seluruh provinsi Indonesia. Selain itu Indonesia juga memiliki beragam masalah sosial, salah satunya kemiskinan. Angka garis kemiskinan di Indonesia perlu diperbaiki. Teknik pemanfaatan data menjadi sebuah informasi baru disebut juga sebagai data mining. Salah satu metode data mining yang cukup populer yaitu clustering dengan menggunakan algoritma k-means. K-means dapat mengolah data tanpa diberitahu lebih dahulu label kelasnya. Penelitian ini menghasilkan tiga kelompok provinsi sesuai dengan angka pendapatan sangat rendah, rendah dan cukup. Pengolahan data angka garis kemiskinan di Indonesia menggunakan algoritma k-means mendapatkan hasil *davies bouldin index* sebesar 0,288. Hasil tersebut

dinilai cukup baik sebab semakin dekat hasil yang diperoleh dengan angka nol, maka kemiripan data anggota antar *cluster* semakin baik.

4. Penelitian yang dilakukan oleh FANI MULYANA NASUTION pada tahun 2019 dengan judul, "PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN KETAHANAN TANAMAN PANGAN KABUPATEN/KOTA DI PROVINSI SUMATERA UTARA (FANI MULYANA NASUTION)". Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Analisis *cluster* merupakan salah satu teknik *statistik* yang ditujukan untuk mengelompokkan objek atau variabel ke dalam beberapa kelompok tertentu dimana setiap objek atau variabel yang terbentuk memiliki sifat dan karakteristik yang berdekatan. Salah satu jenis *cluster* adalah algoritma K-Means *Clustering* yang merupakan *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini untuk mengetahui daerah berpotensi ketahanan pangan untuk menunjang terpenuhinya kebutuhan pangan masyarakat di Provinsi Sumatera Utara. Berdasarkan data hasil Badan Pusat Statistik Sumatera Utara yang mencatat semua hasil tanaman pangan Kabupaten dan Kota di Provinsi Sumatera Utara seperti: padi, jagung, kacang kedelai, ubi kayu dan ubi jalar yang bervariasi jumlahnya. Untuk itu diperlukan pengelompokan kabupaten dan kota untuk mengetahui daerah potensial ketahanan pangan paling tinggi ataupun paling rendah. Oleh karena itu dibutuhkan metode untuk memudahkan dalam pengelompokan daerah potensial ketahanan pangan. Dengan pendekatan pengklasteran K-means, pembagian kelompok daerah dapat dilakukan berdasarkan luas panen (Ha), produksi (ton), dan luas tanam (Ha). Penggunaan K-means *clustering* bertujuan agar memudahkan pengelompokan suatu daerah yang ketahanan pangannya paling tinggi, sedang, dan rendah.
5. Penelitian yang dilakukan oleh Amelia Ayu Anggraini pada tahun 2017 dengan judul, "PENGELOMPOKAN KECAMATAN MENGGUNAKAN METODE K-MEANS CLUSTER". Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Kabupaten Bondowoso sebagai salah satu Kabupaten di Propinsi Jawa Timur, memerlukan suatu kebijakan pengaturan ruang sebagai pengarah perkembangan wilayah di masa mendatang, Kabupaten Bondowoso memiliki banyak potensi, khususnya sektor pertanian, perkebunan, dan pariwisata, tetapi keberadaan potensi tersebut belum dapat dimanfaatkan secara optimal. K-Means *Clustering* merupakan salah satu metode data *clustering non hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster*. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan data

yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster* yang lain sehingga data yang berada dalam satu *cluster* memiliki tingkat *variasi* yang kecil. Dari hasil penentuan kelompok kecamatan dengan menggunakan Algoritma K-Means *Clustering* berhasil dilakukan dan pada *cluster* ke-1 (C1) memiliki anggota kecamatan “Sukosari,Binakal,Pakem,Taman Krocok dan Sempol”, pada *cluster* ke-2 (C2) terdiri kecamatan “Tlogosari,Sumber wringin,Tapen, Wonosari,Tenggarang,Bondowoso dan Tegalampel” sedangkan pada *cluster* ke-3 (C3) terdiri kecamatan “Maesan, Grujugan, Tamanan, Jambesari, Pujer, Curahdami, Wringin, Klabang, Botolinggo, Prajekan dan Cermee”. Proses *clustering* dapat menggunakan beberapa pusat *cluster* (*centroid*) yang diperoleh secara acak.

6. Penelitian yang dilakukan oleh Totok Suprawoto pada tahun 2016 dengan judul, “KLASIFIKASI DATA MAHASISWA MENGGUNAKAN METODE KMEANS UNTUK MENUNJANG PEMILIHAN STRATEGI PEMASARAN”. Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Analisis *cluster* merupakan teknik data mining yang bertujuan untuk mengidentifikasi suatu kelompok dari objek yang memiliki karakteristik yang sama, jumlah kelompok yang dapat diidentifikasi tergantung pada sejumlah data dan jenis dari objeknya, K-Means adalah salah satu metode *clustering* data yang dibagi kedalam bentuk satu atau lebih *cluster* yang memiliki karakteristik sama, *clustering* data mahasiswa menggunakan metode K-Means, terdiri dari nilai rerata ujian nasional (UN) dan indeks prestasi kumulatif (IPK) mahasiswa, penelitian ini menggunakan data mahasiswa angkatan 2014/2015, kemudian diperoleh kesimpulan bahwa kelompok mahasiswa dengan nilai rerata UN yang rendah memiliki pengaruh terhadap prestasi akademik mahasiswa yang rendah pada jenjang diploma-3( D-3) dan strata-1 (S-1), jika mahasiswa memiliki nilai UN yang tinggi maka prestasi akademik mahasiswa juga tinggi pada semua jenjang, dari hasil pengelompokan berdasarkan daerah asal sekolah IPK rata-rata yang tertinggi berasal dari propinsi Daerah Istimewa Yogyakarta (DIY) dan Jawa Tengah (Jateng).
7. Penelitian yang dilakukan oleh Evy Dwi Cahyati pada tahun 2017 dengan judul, “IMPLEMENTASI K-MEANS CLUSTERING UNTUK PEMETAAN DESA DAN KELURAHAN DI KABUPATEN BANGKALAN BERDASARKAN CONTRACEPTIVE PREVALENCE RATE DAN TINGKAT PENDIDIKAN”. Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Anak (DKBPP dan PA) Kabupaten Bangkalan melakukan pendataan keluarga setiap lima tahun dan pemutakhiran data tiap tahun, pendataan keluarga dilakukan di setiap desa dan kelurahan, data yang

dihasilkan dari pendataan keluarga perlu dianalisis untuk mengetahui potensi dan kondisi di tiap desa dan kelurahan,tujuan penelitian yaitu untuk menerapkan algoritma K-Means untuk mengelompokkan dan memetakan desa dan kelurahan di Kabupaten Bangkalan, tahapan implementasi K-Means *clustering* untuk pemetaan desa dan kelurahan di Kabupaten Bangkalan berdasarkan *contraceptive prevalence rate* dan tingkat pendidikan dilakukan dalam lima tahap. Tahap pertama adalah pengumpulan data untuk memperoleh pemahaman teori dan pendataan keluarga,tahap kedua adalah pengolahan data, meliputi seleksi data, pembersihan data, dan *transformasi* data,tahap ketiga yaitu pengelompokan menggunakan algoritma KMeans,dengan cara penentuan pusat *cluster* awal yaitu *random* dan *partis*,tahap keempat merupakan tahapan dalam perancangan hingga pembangunan sistem,meliputi perancangan *database*,perancangan sistem menggunakan *system flowchart*,implementasi sistem berbasis *web*,pengujian sistem. Tahap kelima melakukan analisis hasil *cluster* yang dihasilkan oleh system,dari hasil penelitian didapatkan tiga *cluster* dengan masing-masing *Contraceptive Prevalence Rate* pada *cluster* 1 adalah 55.30%, *cluster* 2 adalah 55.07%, dan *cluster* 3 adalah 48.20%. Tingkat pendidikan mayoritas suami pada *cluster* 1 adalah tamat SMA, sedangkan pada istri adalah SD dan SMA,tingkat pendidikan mayoritas suami dan istri pada *cluster* 2 dan cluster 3 adalah tamat SD, persebaran wilayah pada *cluster* 1 dan *cluster* 3 tidak semuanya mengelompok, sedangkan pada *cluster* 2 persebaran semua wilayahnya mengelompok.

8. Penelitian yang dilakukan oleh Ramdani Budiman pada tahun 2019 dengan judul, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)" Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Setiap perguruan tinggi menginginkan peningkatan dalam penerimaan mahasiswa baru pada setiap tahunnya. Sehingga setiap perguruan tinggi memiliki strategi masing-masing dalam menarik minat para calon mahasiswanya, serta menentukan daerah mana saja yang menjadi lokasi potensial dalam melakukan promosi. Penelitian ini bermaksud untuk menentukan lokasi atau daerah mana saja yang berpotensi untuk mendatangkan mahasiswa baru di perguruan tinggi tersebut. Dengan menerapkan data mining dengan metode *clustering* mengelompokkan item objek penelitian berdasarkan kemiripan sifatnya, sehingga akan diketahui daerah mana saja yang berpotensi untuk mendatangkan mahasiswa baru. Penentuan lokasi promosi penerimaan mahasiswa baru dengan metode data mining akan memberikan dampak yang

bagus dan terarah dalam melakukan promosi, sehingga dapat meningkatkan jumlah mahasiswa baru setiap tahunnya

9. Penelitian yang dilakukan oleh BAYU RESI INDRAWAN pada tahun 2018 dengan judul, " Penerapan Algoritma K-Means clustering untuk menentukan strategi Promosi UNIVERSITAS ISLAM NEGRI SUNAN KALIJAGA YOGYAKARTA". Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Pada proses penerimaan mahasiswa baru Universitas UIN Yogyakarta akan menghasilkan data-data yang baru yang berlimpah berupa profil dari mahasiswa baru. Berdasarkan hal diatas informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa sehingga menghasilkan informasi yang bermanfaat bagi pihak universitas. Informasi yang didapatkan dapat membantu pihak universitas sebagai bahan pertimbangan dalam menentukan strategi promosi penerimaan mahasiswa baru di tahun depan. Penelitian ini bertujuan untuk mengelompokkan data mahasiswa kedalam sebuah cluster dengan menggunakan metode Data Mining algoritma K-Means *Clustering*. Data mahasiswa dikelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama akan berada dalam satu *cluster*. Atribut yang digunakan adalah kota asal, asal sekolah, dan IPK mahasiswa selama 2 semester awal dengan nilai diatas 2,75. *Cluster* yang terbentuk setelah proses K-Means Clustering adalah tiga cluster dengan cluster pertama berjumlah 508 data mahasiswa, *cluster* kedua 139 data mahasiswa, dan *cluster* ketiga 62 data mahasiswa. *Cluster* dengan nilai rata – rata IPK paling tinggi adalah *cluster* satu. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan cluster yang terbentuk oleh pihak Universitas UIN Yogyakarta.
10. Penelitian yang dilakukan oleh Putrama Alkhairi pada tahun 2019 dengan judul, " Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara". Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu Data mining ialah istilah yang telah digunakan untuk proses *otomatik* yang menggunakan matematika, kecerdasan buatan, statistik, dan *machine learning* untuk *mengidentifikasi* dan *mengekstrasi* informasi pengetahuan yang tersembunyi dalam database, *clustering* ialah suatu proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di suatu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain, K-means *clustering* ialah metode data clustering *non-hirarkii* yang mengelompokkan data dalam bentuk satu maupun lebih dari suatu *cluster*, algoritma k-means merupakan algoritma teknik *clustering*

yang berbasis suatu partisi, tujuan penelitian ini ialah untuk mengetahui pengukuran dari kinerja algoritma dan mengelompokkan daerah potensi karet produktif, pengukuran dilihat dari hasil cluster dengan cara menghitung nilai kemurnian (*purity measure*) masing – masing *cluster* yang di hasilkan, data penelitian ini menggunakan ialah data dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Diharapkan dengan adanya penelitian ini bisa berkontribusi dalam mengefektifkan dalam penggunaan lahan pertanian karet pada setiap daerah. Pada tinjauan studi penelitian dari ke-sepuluh judul tinjauan memiliki kesamaan yaitu menggunakan metode penelitian *K-Means* serta memiliki output yang sama yaitu pemetaan wilayah khususnya pada penelitian yang di lakukan Suhandio Handoko dan Tri Ginanjar Laksana memiliki subjek yang sama yaitu layanan data internet dan operator seluler, serta pada penelitian Bagas Prayogo Haryaji menggunakan *tool* yang sama yaitu rapidminer, untuk perbedaannya pada penelitian Suhandio Handoko hanya menggunakan satu operator seluler sedangkan pada penelitian ini menggunakan 3 operator seluler, serta pada penelitian yang lainnya memiliki perbedaan subjek penelitian.

**Tabel 2.14 Tinjauan Studi**

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
1	Suhandio Handoko - 2020	Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering	Belum Benarnya Pengelompokan Penjualan Paket Data Telkomsel	Jurnal Ilmiah Teknologi dan Rekayasa Volume 25 No. 1 April 2020	Memiliki Output yang sama
2	Tri Ginanjar Laksana (2017)	Analisa Pola Pemilihan Provider Telekomunikasi Berdasarkan Pekerjaan Dengan Algoritma Klustering K-Means	Sulitnya memilih provider yang tepat untuk setiap warga yang berbeda-beda kebutuhannya	Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Paper UNISBANK Ke 3 (Sendi_U3) 2017 ISBN: 9-789-7936-499-93	Memiliki Output yang sama

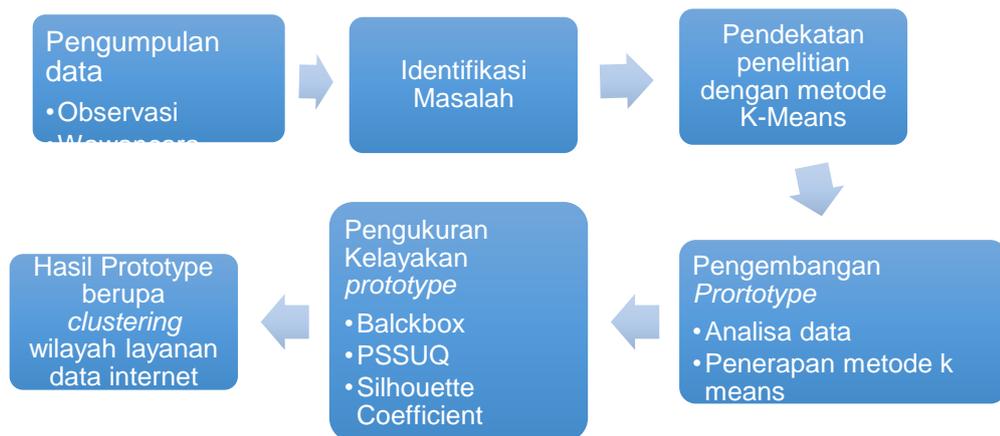
No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
3	BAGAS PRAYOGO HARYAJI (2018)	Penerapan Algoritma K-Means Untuk Memetakan Garis Kemiskinan Menurut Provinsi Di Indonesia	Memetakan tingkat kemiskinan di setiap provinsi di Indonesia	Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa Bekasi, Bagas Prayogo Haryaji, 311410430, 2018	Menggunakan Tools Rapidminer sebagai tool pengujian metode penelitian
4	Amelia Ayu Anggraini (2017)	Pengelompokan Kecamatan Menggunakan Metode K-Means Cluster	Kabupaten di Propinsi Jawa Timur, memerlukan suatu kebijakan pengaturan ruang sebagai pengarah perkembangan wilayah di masa mendatang.	Thesis (Undergraduate), Computer Science, Information, 10 Jul 2021	Memiliki artibut yang hampir mirip pada jumlah guru dan siswa
5	Totok Suprawoto (2016)	Klasifikasi Data Mahasiswa Menggunakan Metode Kmeans Untuk Menunjang Pemilihan Strategi Pemasaran	Pendataan keluarga dilakukan di setiap desa dan kelurahan. Data yang dihasilkan dari pendataan keluarga perlu dianalisis untuk mengetahui potensi dan kondisi di tiap desa dan kelurahan.	Jurnal Informatika dan Komputer (JIKO) – Vol. 1, No. 1, Februari 2016	Memiliki Output yang sama pemetaan wilayah
6	Evy Dwi Cahyati (2017)	Implementasi K-Means Clustering Untuk Pemetaan Desa Dan Kelurahan Di Kabupaten Bangkalan Berdasarkan Contraceptive Prevalence Rate Dan Tingkat Pendidikan	Belum Tepatnya pengelompokan wilayah	Seminar Nasional Matematika dan Aplikasinya, 21 Oktober 2017 Surabaya, Universitas Airlangg	Artibut Jumlah Penduduk
7	Ramdani Budiman (2019)	Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)	Kurang Tepatnya penentuan wilayah untuk promosi	Jurnal ProTekInfo Vol. 6 No 1, September 2019	Memiliki Output yang sama pemetaan wilayah

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
8	FANI MULYANA NASUTION, 2019)	Penerapan Metode K-Means Clustering Untuk Mengelompokkan Ketahanan Tanaman Pangan Kabupaten/Kota Di Provinsi Sumatera Utara	Menentukan ketahanan tanaman pangan di setiap wilayahnya	Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Sumatera Utara, Fani Mulyana Nasution-170823037, 2019	Memiliki Output yang sama pemetaan wilayah
9	BAYU RESI INDRAWAN (2018)	Penerapan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta	Membuat kelompok-kelompok wilayah untuk promosi	Thesis (Skripsi), Tehnik Informatika, 20 Mar 2019	Memiliki Output yang sama pemetaan wilayah
10	Putrama Alkhairi (2019)	Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif Di Sumatera Utara	Mengelompokkan Wilayah yang memiliki potensi pertanian karet	ISBN: 978-602-52720-1-1, SAINTEKS 2019 Januari 2019, Hal: 762 – 767	Memiliki Output yang sama pemetaan wilayah

Untuk tinjauan pustaka di atas pada judul penelitian Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering memiliki persamaan objek penelitian yaitu memetakan layanan data internet di setiap wilayahnya tepati dari hasil penelitiannya hanya berdasarkan *revenue*. Pada penelitian Analisa Pola Pemilihan Provider Telekomunikasi Berdasarkan Pekerjaan Dengan Algoritma Klustering K-Means juga membahas tentang pemetaan pemilihan provider telekomunikasi berdasarkan pekerjaan di setiap wilayah dimana objek penelitiannya sama dengan objek penelitian yang di teliti yaitu penyedia jasa telekomunikasi, dan pada penelitian Penerapan Algoritma K-Means Untuk Memetakan Garis Kemiskinan Menurut Provinsi Di Indonesia menggunakan *tool*/ penguji metode yang sama yaitu *Rapidminer*.

## E. Kerangka Pemikiran

Metode K-Means merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* yang bisa di jadikan metode pemecahan masalah pemetaan layanan telekomunikasi di setiap wilayah, kerangka pemikiran dibawah ini dibuat berdasarkan masalah-masalah pemetaan wilayah yang terjadi.



**Gambar 2. 2 Kerangka Pemikiran**

Berdasarkan gambar 2.2 tahap dalam kerangka pemikiran adalah:

1. Melakukan pengumpulan data yang di butuhkan.
  - a) Observasi : Melakukan *Observasi* pada setiap wilayah untuk mengetahui kualitas layanan data internet yang ada pada setiap wilayah.
  - b) Wawancara: Melakukan wawancara kepada provider telekomunikasi untuk mengetahui kriteria menentukan wilayah dengan layanan data yang baik serta kebutuhan fasilitas pendukung yang harus ada pada setiap wilayah agar mendapatkan layanan data yang maksimal.
2. Melakukan identifikasi masalah  
Dari hasil pengumpulan data di lakukan proses analisa terhadap data yang telah di kumpulkan untuk mengidentifikasi masalah yang terjadi serta mengetahui kebutuhan terhadap fasilitas pendukung layanan data internet yang seharusnya tersedia pada setiap wilayahnya
3. Melakukan pendekatan penelitian dengan metode *K-Means* terhadap data yang telah di kumpulkan, sehingga dapat menghasilkan *clustering* yang tepat sesuai dengan kebutuhan untuk menyediakan layanan data internet yang maksimal.
4. Melakukan pengembangan
  - a) Analisa terhadap data lalu merancang *prototype* yang sesuai dengan kebutuhan output yang di inginkan

- b) Penerapan metode algoritma K Means terhadap data yang telah di analisa sebelumnya
  - c) Memulai pengkodean terhadap *prototype* yang akan di kembangkan berdasarkan hasil analisa dan perancangan *prototype*
5. Melakukan pengukuran kelayakan pada *prototype* menggunakan dua cara yaitu:
- a) Blackbox dilakukan pengujian yang didasarkan pada detail aplikasi seperti tampilan aplikasi, fungsi-fungsi yang ada pada aplikasi, dan kesesuaian alur fungsi dengan bisnis proses yang diinginkan oleh customer
  - b) PSSUQ mengukur kepuasan yang dirasakan pengguna terhadap produk atau sistem. PSSUQ memberikan skor kepuasan secara keseluruhan dengan rata-rata sub-skala, yaitu kualitas sistem, kualitas informasi dan kualitas antarmuka. PSSUQ tersedia secara gratis (Garcia, 2013).
  - c) *Silhouette Coefficient* untuk mengetahui kemiripan data hasil *clustering*.
6. Mengeluarkan hasil Output berupa data hasil *clustering* wilayah layanan data internet yang di hasilkan oleh *prototype* sehingga dapat membantu penyedia layanan data internet meningkatkan layanannya pada setiap wilayah sehingga dapat memenuhi kebutuhan masyarakat terhadap layanan data internet.

## **F. Hipotesis Penelitian**

Metode K-Means diduga dapat menyelesaikan masalah dalam pemetaan layanan data internet dan fasilitas pendukungnya yang saat ini masih kurang tepat.