BAB II

LANDASAN TEORI

2.1 Tinjauan Terhadap Penelitian Yang Terkait

Berikut ini adalah beberapa hasil penelitian yang terkait dan menjadi referensi pada penelitian ini :

1. Penelitian yang dilakukan oleh Wiwit Agus Triyanto (2015). Dengan judul "Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk" Permasalahan yang diangkat dalam penelitian ini adalah perusahaan kesulitan dalam menentukan keputusan-keputusan mengenai target pasar, penempatan produk, bauran pemasaran, serta tingkat biaya pemasaran yang diperlukan. Untuk mengatasi masaalah tersebut. maka penelitian ini akan menggunakan algoritma K-Medoids untuk pengelompokan data penjualan, sehingga akan ditemukan informasi yang dapat digunakan untuk penentuan strategi pemasaran produk yang tepat Informasi yang bisa diperoleh dari hasil proses clustering tersebut adalah pada transaksi penjualan yang semula terdiri dari 1360 record transaksi, dapat dikelompokkan menjadi 5 cluster: dengan (1). Cluster pertama terdiri dari 909 record transaksi berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 1 - 2; dan jumlah barang yang tidak dibeli sekitar 301 - 302, (2).Cluster kedua terdiri dari 166 record transaksi berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 3 - 7; dan jumlah barang yang tidak dibeli sekitar 296 - 300, (3). Cluster ketiga terdiri dari 66 record transaksi berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 8 - 14; dan jumlah barang yang tidak dibeli sekitar 289 –

295, (4). Cluster keempat terdiri dari 132 record transaksi berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 15 -48; dan jumlah barang yang tidak dibeli sekitar 255 –288, (5). Cluster kelima terdiri dari 87 record transaksi berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 49 -303; dan jumlah barang yang tidak dibeli sekitar 0 –254. dari kelima klaster tersebut maka Strategi pemasaran produk dapat dilakukan dengan melakukan promosi pada cluster kelima yang memiliki kombinasi jumlah barang dibeli yang paling tinggi.

2. Penelitian yang dilakukan oleh Rima Dias Ramadhani (2014). Dengan judul "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro" Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data mahasiswa Universitas Dian Nuswantoro dengan memanfaatkan proses data mining dengan menggunakan teknik Clustering. Metode yang digunakan adalah CRISP-DM dengan melalui proses business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Algoritma yang digunakan untuk pembentukan cluster adalah algoritma K-Means. Implementasi menggunakan RapidMiner 5.3 digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan adalah kota asal, program studi dan IPK mahasiswa. Cluster mahasiswa yang terbentuk adalah tiga cluster, dengan cluster pertama 804 mahasiswa, cluster kedua 2792 mahasiswa dan cluster ketiga sejumlah 223 mahasiswa. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan

keputusan untuk menentukan strategi promosi berdasarkan cluster yang terbentuk oleh pihak admisi UDINUS.

2.2 Data Mining

Menurut (Efori Buulolo, 2020) dalam e-book yang berjudul "Data mining untuk Perguruan Tinggi". Data Mining atau kadang disebut juga Knowledge Discovery in Database (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar. Output dalam data mining dapat dipergunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki keputusan di masa yang akan datang.

Data mining digunakan pada proses pencarian suatu pola tertentu atau informasi pada suatu data menggunakan suatu teknik tertentu. Penggunaan data mining biasanya ditujukan untuk melakukan ekstraksi informasi penting pada suatu dataset yang besar, dimana data mining akan menghasilkan suatu keputusan akhir setelah dilakukan pengolahan data pada kumpulan data yang berjumlah besar.

Data Mining mempunyai beberapa model proses yang digunakan untuk mengarahkan pelaksanaan data mining, model proses yang digunakan dalam penelitian ini adalah *CRISP-DM* (*Cross Industry Standard for Data mining*)

CRISP-DM (Cross Industry Standard for Data mining) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data di dalamnya, Tujuan dari metode CRISP-DM ini yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan. CRISP-DM memiliki

tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan dalam penelitian. Pada CRISP-DM ini terdapat 6 tahap yaitu :

1. Business Understanding

Business Understanding adalah proses menentukan tujuan bisnis, memahami situasi dan kondisi pada saat penelitian dan menetapkan sebuah tujuan dari penelitian yang dilakukan ke dalam permasalahan yang diselesaikan dengan data mining.

2. Data Understanding

Data understanding adalah tahap persiapan, melakukan pengecekan terhadap data yang digunakan, mengumpulkan data awal serta melakukan identifikasi pada kualitas data. Dalam data understanding, data yang digunakan akan melalui proses deskripsi dari setiap fiturnya.

3. Data Preparation

Data preparation merupakan proses yang dilakukan setelah data telah dikumpulkan. Pada tahap ini, data akan melalui proses identifikasi, pemilihan data, pembersihan data dan transformasi data.

4. Modeling

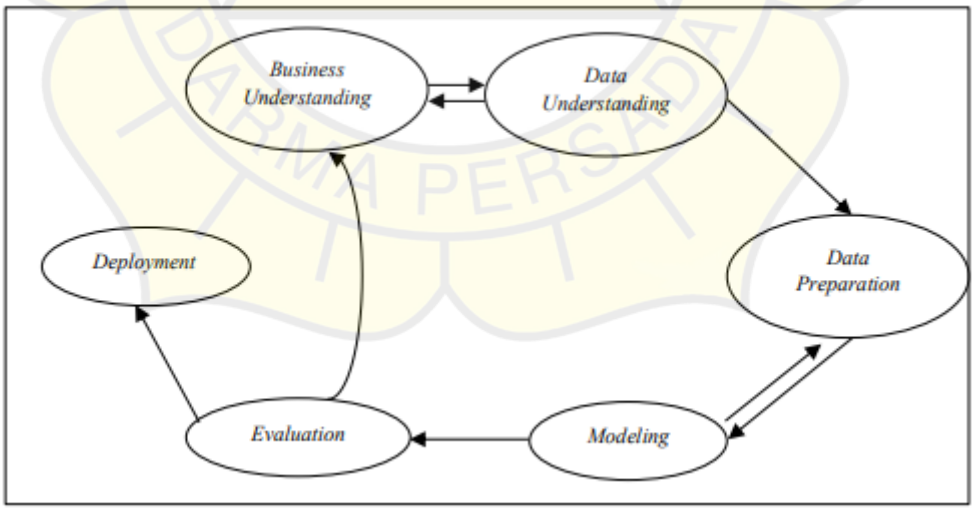
Modeling merupakan tahap implementasi algoritma yang akan digunakan untuk melakukan pencarian, identifikasi, serta menghasilkan pola yang akan digunakan pada data penelitian.

5. Evaluation

Evaluation adalah suatu proses untuk melakukan pengukuran hasil evaluasi dari model yang telah diimplementasikan sebelumnya di tahap modelling. Hasil evaluasi tersebut menggambarkan proses dari data mining yang telah dilakukan dan mengukur model yang paling baik untuk digunakan.

6. Deployment

Deployment adalah proses pembuatan aplikas beserta laporan atau artikel jurnal menggunakan hasil penelitian.



Gambar 2. 1 Metode *CRISP -DM*

2.3 Kegunaan Data Mining

Secara garis besar kegunaan *data mining* dibagi menjadi dua yaitu : deskriptif dan prediktif. Secara deskriptif artinya data mining berarti menemukan pola-pola yang digunakan untuk menjelaskan karakteristik data. Sedangkan secara prediktif data mining berarti dapat digunakan untuk menemukan model-model pengetahuan yang digunakan untuk melakukan prediksi.

Data mining berdasarkan fungsionalitasnya dapat dikelompokkan menjadi enam bagian yaitu (Suyanto, 2017) :

1. Klasifikasi (*classification*)

   Diterapkan pada data baru untuk mengelompokkan jenis objek. Klasifikasi termasuk pada model *supervised*. Pada persoalan klasifikasi kita memiliki sampel data dan memprediksi beberapa *class* yang ada berdasarkan sampel yang ada. Hanya satu atribut di antara banyak atribut yang disebut dengan atribut predator. Klasifikasi ini juga umum, digunakan untuk pemodelan bisnis dan lainnya. Misal klasifikasi menentukan penyakit tertentu atau menentukan *customer* berdasarkan model pembayarannya.

2. Klastering (*Clustering*)

   Berbeda dengan klasifikasi, klatering termasuk model *unsupervised*. Klastering mengelompokkan data yang tidak diketahui labelnya. Klasteringyang diorganisasi ke dalam struktur hierarkikal akan mendefinisikan taksonomi dari data. Penerapan metode klastering yang tepat akan menghasilkan *clustering* yang berkualitas. Suatu *cluster* dikarakterisasi oleh *centroid,* atribut *histogram* dan *clustering* model

hierarchial tree.

3. Regresi (*Regression*)

   Merupakan suatu fungsi yang digunakan untuk memodelkan data untuk meminimalkan hasil kesalahan prediksi. Umumnya regresi dilakukan dengan data yang bersifat *time series*.

4. *Association Rule*

   Merupakan pemodelan kebergantungan. Fungsi asosiasi ini biasanya kita kenal dengan istilah "*market basket analysis*" yang merupakan fungsi untuk menemukan relasi atau korelasi antara himpunan item-item. Aturan asosiasi diartikan pada basket data yang digunakan untuk keperluan promosi, desain katalog untuk meningkatkan penjualan. Contoh penerapan asosiasi adalah ketika *customer* membeli pamper maka ada kemungkinan membeli bir.

5. *Anomaly Detection*

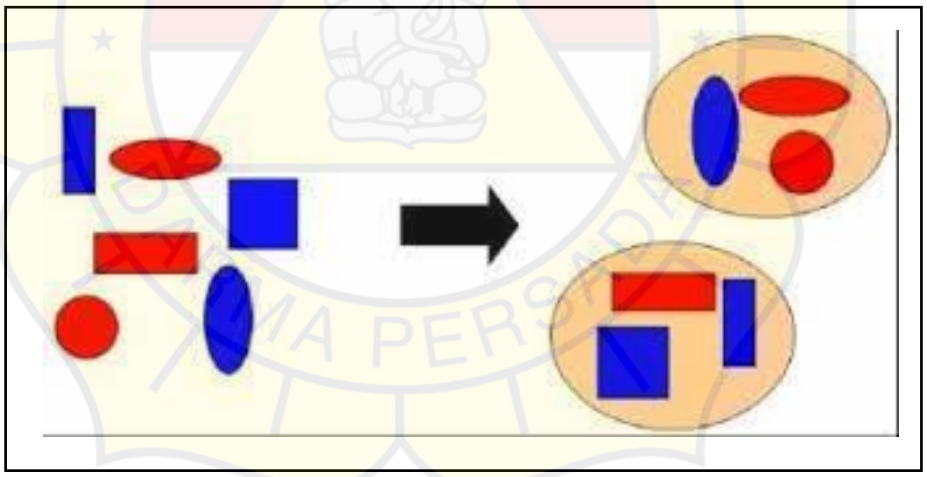
   Mengidentifikasi data yang tidak umum. Bisa berupa *outlier*, perubahan deviasi/bias yang penting dan perlu investigasi lebih lanjut.

6. *Summarization*

   Menyediakan representasi data yang lebih sederhana meliputi pelaporan, visualisasi data yang dipergunakan untuk menunjang informasi dan penguatan keputus

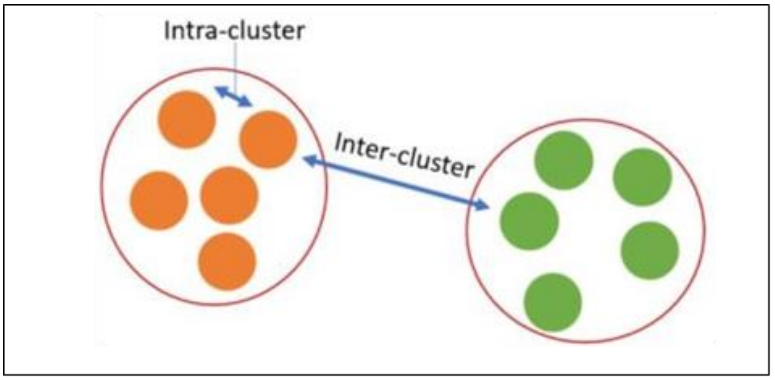
2.4 Teknik Clustering

Teknik *Clustering* merupakan klasifikasi tanpa pengawasan dan merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas atau pengklusteran merupakan teknik yang membagi data menjadi kelompok-kelompok (Kusuma and Andriani, 2020). Pembagian kclompok data tersebut berdasarkan kritcria tertentu. Tujuan dari tcknik *clustering* adalah mengelompokkan data yang mempunyai kriteria sama ke dalam satu wilayah atau kelompok yang sama dan mcmisahkan data yang mempunyai perbedaan kriteria. Data akan dikelompokkan dalam satu kelompok atau *cluster* mempunyai tingkat kemiripan yang maksimum dan setiap kelompok mempunyai tingkat kemiripan yang minimum.



Gambar 2. 2 Penggambaran Teknik *Clustering*

Pada gambar 2.4 Mengilustrasikan data yang dikelompokkan berdasarkan kriteria bentuknya. Cluster yang baik ketika jarak antara objek data yang ada pada *cluster* yang sama (*intra-cluster*) dapat diminimalkan dan jarak antara objek yang ada pada *cluster* yang berbeda (*inter-cluster*) dapat dimaksimalkan.



Gambar 2. 3 Perbedaan *intra-cluster* dengan *inter-cluster*

Berikut merupakan beberapa contoh implementasi teknik *clustering* dalam dunia nyata (Aggarwal, 2016):

1. *Data summarization* atau peringkasan data. *Data mining* berkaitan dengan mcringkas informasi dari data sehingga proses *clustering* sering kali merupakan langkah awal dari banyak algoritma *data mining*.

2. *Customer segmentation* atau segmentasi konsumen yang sering kali digunakan untuk menganalisis perilaku umum dari kelompok konsumen yang serupa. Sebagai contoh pola perilaku pemilihan produk suatu konsumen yang ada pada suatu kelompok dapat digunakan untuk merekomendasikan produk tersebut kepada konsumen lain yang masih dalam satu kelompok.

3. *Social network analysis* atau analisis jaringan sosial yang digunakan untuk menganalisis perilaku manusia yang diperoleh dari mcnganalisis dinamika yang tcrjadi pada komunitas. Jaringan sosial merupakan simpul yang telah ter kluster secara erat melalui hubungan

yang saling tcrkait yang biasanyakelompok tcman yang sama atau komunitas.

4. Berhubungan dengan masalah *data mining* yang lain.

   Berikut merupakan metode yang umum digunakan pada teknik *clustering* antaraIain:

   1. *Partitioning mcthod* merupakan teknik yang digunakan untuk mengelompokkan data sesuai dengan jumlah cluster atau kelompok yang ditentukan. Dalam hal ini jumlah kelompok ditentukan di awal. Algoritma yang populer digunakan pada *partitioning method* adalah K-means, K- *medoids dan CLARA (Clustering Large Application)*.
   2. *Hierarchical method* mau metode hierarki merupakan alternatif dari *partitioning method* dan tidak memerlukan menentukan jumlah *cluster* atau kclompok di awal. Terdapat dua pendckatan yaitu *Divisive* dan *Agglotnerative*. Pendekatan *divisive* yang merupakan pendekatan *top-down*.Semua objek data dikumpulkan dalam satu *cluster* dan selanjutnya dipecah menjadi *cluster* yang lcbih kecil menggunakan iterasi berulang hingga kondisi selesai tercapai. Sedangkan pendekatan *agglomerative* merupakan pendekatan *bottom-up* di mana setiap kelompok data terpisah sejak awal dan digabungkan hingga kondisi selesai tercapai.

2.5 Pemrograman Aplikasi

2.5.1 Web

Menurut Rohi Abdulloh (2015), Web adalah sekumpulan halaman yang terdiri dari beberapa halaman yang berisi informasi dalam bentuk data digital baik berupa text, gambar, video, audio, dan animasi lainnya yang disediakan melalui jalur koneksi internet.

Menurut Agus Hariyanto (2015), Web dapat diartikan sebagai kumpulan halaman yang menampilkan informasi data teks, data gambar, data animasi, suara, video dan gabungan dari semuanya, baik yang bersifat statis maupun dinamis yang membentuk satu rangkaian bangunan yang saling terkait, dimana masing-masing dihubungkan dengan jaringan-jaringan halaman *(hyperlink)*.

Menurut (Mara Destiningrum, Qadhli Jafar Adrian, 2017) Web adalah Sebuah *software* yang berfungsi untuk menampilkan dokumen - dokumen pada suatu web yang membuat pengguna dapat mengakses internet melalui *software* yang terkoneksi dengan internet.

2.5.2 HTML

"*HyperText Markup Language* (HTML) adalah sebuah bahasa *markup* yang digunakan untuk membuat sebuah halaman web, menampilkan berbagai informasi di dalam sebuah Penjelajah web Internet dan *formating hypertext* yang sederhana ditulis kedalam berkas format ASCII agar dapat menghasilkan tampilan wujud yang terintegerasi. Dengan kata lain, berkas yang dibuat dalam perangkat lunak pengolah kata dan disimpan kedalam format ASCII normal sehingga menjadi *home page* dengan perintah perintah HTML. Bermula dari sebuah bahasa yang

sebelumnya banyak digunakan didunia penerbitan dan percetakan yang disebut dengan SGML(*Standard Generalized Markup Language*), HTML adalah sebuah standar yang digunakan secara luas untuk menampilkan halaman web. HTML saat ini merupakan standar Internet yan didefinisikan dan dikendalikan penggunaannya oleh *World Wide Web Consortium*(W3C). HTML dibuat oleh kolaborasi Caillau TIM dengan Berners-lee robert ketika mereka bekerja di CERN pada tahun 1989 (CERN adalah lembaga penelitian fisika energi tinggi di Jenewa)" (Ahmad dan Syarif, 2016).

2.5.3 CSS

CSS (*Cascading Style Sheet*) merupakan suatu bahasa program web yang digunakan untuk mengendalikan dan membangun berbagai komponen dalamn web yang digunakan untuk mengendalikan dan membangun berbagai komponen dalam web sehingga tampilan web akan lebihrapi, terstruktur dan seragam.

CSS saat ini dikembangkan oleh *World Wide Web Consortium* (W3C) dan menjadi bahasa standar dalam pembuatan web. CSS difungsikan sebagai penopang atau pendukung, dan pelengkap dari file html yang berperan dalam penataan kerangka dan *layout*.

CSS mampu di berbagai platform, maksudnya dapat dijalankan pada berbagai macam sistem operasi dan web browser. Secara umum, yang dilakukan oleh CSS adalah pengaturan layout, kerangka, teks, gambar, warna, tabel, spasi, dan lain sebagainya. Kita akan mengetahui masing-masing pengaturan tersebut pada pembahasan selanjutnya. (M.Taofik Chulkamdi, Sulis Purnomo, 2016).

2.5.4 Java Script

JavaScript merupakan bahasa *scripting* yang pada awalnya dikembangkan oleh Netscape. Dengan menggunakan javascript, dapat dibuat tampilan web yang lebih interaktif lagi. Agar user dapat menjalankan javascript, alat yang dibutuhkan hanyalah *browser* yang kemampuan javascript-nya telah diaktifkan. Penulisan javascript dapat dipadukan pada halaman HTML, dan dari sini dapat dilihat bahwa javascript merupakan bahasa pemrograman yang berjalan di sisi *client*. Karena proses dari javascript tergantung proses dari *browser* yang digunakan oleh *user*. Contoh penggunaan JavaScript adalah sebagai berikut:

```
<html>

<body>

<script language="javascript">

 document.write("Ini dihasilkan oleh javascript");

</script>

</body>

</html>
```

2.5.5 Boostrap

Bootstrap adalah framework yang kuat menyediakan set kelas CSS dan fungsi JavaScript untuk memudahkan proses pembangunan antarmuka halaman web. Mengaktifkan fitur *design* responsif dukungan untuk menampilkan desktop maupun mobile. Situs dikembangkan dapat bekerja dengan baik pada desktop maupun mobile. Developer tidak harus bekerja dengan CSS untuk membuat website terlihat menarik atau mendukung prinsip desain *responsive*, kecuali diperlukan . Twitter Bootstrap dapat diunduh secara gratis di website resminya.

yakni www.getbootsrap.com, setelah itu tinggal memanggil file CSS Bootstrap pada file project website yang akan menggunakan Bootstrap. Begitu selesai memanggil Bootstrap, maka secara otomatis akan mengubah tampilan website tanpa harus melakukan pengetikan sintak-sintak CSS seperti biasa dilakukan" (Alfan dan Bamban, 2016). Bootsap merupakan framework yang disediakan sediakan oleh twitter untuk membangun sebuah website. Untuk dapat menggunakan CSS serta fungsi java script yang terkandung pada bootstrap, terdapat template HTML yang disediakan oleh bootsrap yang bisa didapatkan melalui website resmi bootstrap . Cara meggunakannya adalah dengan mencopy paste kodingan dari template tersebut, kemudian edit kode- kode tersebut sesuai dengan kebutuhan. Sebagai batasan, ada beberapa kode yang tidak boleh diubah agar tetap selalu dapat menggunakannya, sepertimisalannya pemanggilan link yang terdapat pada tag head " ".

2.5.6 PHP

PHP atau *(hypertext preprocessor)*, merupakan bahasa pemrograman pada sisi *server* yang memperbolehkan *programmer* menyisipkan perintah- perintah perangkat lunak web *server* akan dieksekusi sebelum perintah itu dikirim oleh halaman ke *browser* yang me-*request*-nya. Sesuai dengan fungsinya php berjalan di sisi *server* maka php adalah bahasa pemrograman yang digunakan untuk membangun teknologi web *application*. oleh karena itu, salah satu tool yang harus tersedia sebelum memulai pemograman PHP adalah *server*(Jubilee, 2017). Dengan menggunakan bahasa pemrograman PHP sebuah *website* dapat menjadi *website* dinamis. Tidak seperti HTML yang bertipe statis. PHP dirancangan untuk dapat bekerja sama dengan *database server* dan dibuat sedemikian rupa sehingga pembuatan dokumen HTML yang dapat mengakses *database* menjadi begitu mudah. Tujuan dari bahasa *scripting* ini adalah untuk membuat aplikasi di mana aplikasi tersebut yang dibangun oleh PHP pada umumnya akan memberikan hasil pada web *browser*, tetapi prosesnya secara keseluruhan dijalankan di *server*.

2.5.7 Database

Menurut Mukhamad Masrur (2016), *Database* adalah sekumpulan file data yang satu sama lainnya saling berhubungan yang diorganisasi sedemikian rupa sehingga memudahkan untuk mendapatkan dan memproses data tersebut. Lingkungan sistem *database* menekankan pada data yang tidak tergantung (*independent*) pada aplikasi yang akan menggunakan data tersebut.

Penggunaan *database* pada komputer dilakukan dengan menggunakan *table-table*. Pada tabel-tabel tersebut masih dikelompokkan lagi menjadi beberapa bagian untuk membedakan data yang satu dengan data yang lain. Pada sebuah tabel

database harus memiliki setau kategori data yang digunakan sebagai kunci untuk membedakan data-data yang ada didalam satu tabel. Data kunci tersebut tidak boleh sama antara satu data dengan data lainnya. Data kunci sering disebut dengan *Primary Key*.

2.5.8 MySQL

MySql adalah suatu sistem manajemen database (*Relasional Database Management System*) bersifat terbuka (*Open Source*). Maksudnya terbuka adalah mysql boleh di-download oleh siapa saja baik versi kode program aslinya maupun versi binernya dan bisa digunakan secara gratis. Sejarah Mysql yang merupakan hasil buah pikiran dari Michael ”Monty” Widenius, David Axmark dan Allan Larson dimulai tahun 1995. mereka bertiga mendirikan perusahaan bernama Mysql di Swedia. (Lukman,2015:148). Dengan berbagai keunggulan yang dimiliki, membuat *software database* ini banyak digunakan oleh praktisi untuk membangun suatu *project*. Adanya fasilitas API (*Application Programming Interface*) yang dimiliki oleh Mysql, memungkinkan bermacam-macam aplikasi Komputer yang ditulis dengan berbagai bahasa pemograman dapat mengakses basis data MySQL.

2.6 Algoritma Sistem

2.6.1 Algoritma K-Means Clustering

Algoritma K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat *cluster* awal. Pada tahap ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya K-Means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap – tiap

kluster. Posisi pusat kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat kluster dan terakhir akan terbentuk posisi pusat *Master* yang baru (Danni dan Setiawan, 2016).

Algoritma klastering K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetanpkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat.

Langkah-langkah pengelompokan data adalah :

1. Pilih jumlah *cluster* k.
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random.
3. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap tiitk pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut (*Euclidean Distance*) dengan rumus seperti dibawah ini (Witanto, Ratnawati dan Anam, 2019) :

$$\boldsymbol{D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}} \qquad (1)$$

dimana:
$D\ (i,j)$ = Jarak data ke i ke pusat *clusterj*
X_{ki} = Data ke i pada atribut data ke k
X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi maka proses pengklasteran selesai.

2.6.2 Metode *Elbow*

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan cluster terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah cluster yang ditambah . Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik .

Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai cluster. Karena semakin besar jumlah cluster K maka nilai SSE akan semakin kecil. Rumus SSE pada K-Means

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_K} \left\| X_i - C_k \right\|_2^2$$

Setelah dilihat akan ada beberapa nilai K yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai K akan turun secara perlahan-lahan sampai hasil dari nilai K tersebut stabil. Misalnya nilai cluster K=2 ke K=3, kemudian dari K=3 ke K=4, terlihat penurunan drastis membentuk siku pada titik K=3 maka nilai cluster k yang ideal adalah K=3.

Algoritma Metode *Elbow* dalam menentukan nilai K pada K-Means :

1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil *sum of square error* dari tiapnilai K
5. Melihat hasil *sum of square error* dari nilai K yang turun secara drastis
6. Tetapkan nilai K yang beberbentuk siku
7. Selesai

2.7 Pemodelan UML

"*Unified Modeling Language* (UML) bukanlah suatu proses melainkan bahasa pemodelan secara grafis untuk menspesifikasikan, memvisualisasikan, membangun, dan mendokumentasikan seluruh artifak sistem perangkat lunak. Penggunaan model ini bertujuan untuk mengidentifikasikan bagian-bagian yang termasuk dalam lingkup sistem dibahas dan bagaimana hubungan antara sistem

suatu dengan subsistem maupun sistem lain di luarnya." *Unified Modeling Language* (UML)" adalah sebuah bahasa yang berdasarkan grafik atau gambar untuk memvisualisasi, menspesifikasikan dari sebuah sistem pengembangan software berbasis object oriented." Dari Pengertian diatas penulis menyimpulkan bahwa *Unified Modeling Language* (UML) merupakan bahasa pemodelan yang berbentuk grafis yangdigunakan untuk memvisualisasi, menspesifikasikan suatu sistem perangkat lunak (Yuhanar, 2018).

2.7.1 Use Case Diagram

Diagram *use case* merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat. Secara kasar, use case digunakan untuk mengetahui fungsi apa saja yang ada di dalam suatu sistem dan siapa saja yang berhak menggunakan fungsi-fungsi tersebut." (Yunahar, 2018). Berikut adalah simbol-simbol yang ada pada diagram use case:

Tabel 2. 1 Simbol *Use Case Diagram* (Yunahar, 2018)

Simbol	Deskripsi
Use Case	Fungsional yang disediakan sistem sebagai unit-unit yang saling bertukar antar unit atau aktor biasanya dinyatakan dengan menggunakan kata kerja di awal *frase* nama *use case*
Actor	Orang, proses, atau sistem lain yang beriteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi yang akan dibuat itu sendiri, jadi walaupun simbol dari aktor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal *frase* nama aktor
Asosiasi	Komunikasi antar aktor dan *use case* yang berpatisipasi pada *use case* atau *use case* memiliki interaksi dengan aktor
Extensi	Relasi *use case* tambahkan kesebuah *use case* dimana *use case* yang ditambahkan dapat berdiri sendiri walau tanpa *use case* tambahan
Generalisasi	Hubungan generalalisasi dan sepealisasi (umum-khusus) antara dua buah *use case* dimana fungsi yang satu adalah fungsi yang lebih umum dari yang lainnya
Include	Relasi *use case* tambahkan ke sebuah *use case* dimana *use case* yang ditambahkan memerlukan *use case* ini untuk menjalankan fungsinya atau sebagai syarat dijalankan *use case* ini

2.7.2 Activity Diagram

Activity Diagram menggambarkan *work flow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. Yang perlu diperhatikan disini adalah bahwa diagram aktivitas menggambarkan aktivitas sistem bukan apa yang dilakukan aktor, jadi aktivitas dapat dilakukan oleh sistem" (Yunahar, 2018). Simbol-simbol yang digunakan dalam activity diagram sebagai berikut :

Tabel 2. 2 Simbol *Activity Diagram* (Yunahar, 2018)

Simbol	Deskripsi
	Simbol *start* untuk menyatakan awal dari suatu proses
	Simbol *stop* untuk menyatakan akhir dari suatu proses
	Simol *desticion* digunakan untuk menyatakan kondisi dari suatu proses
	Simbol *action* menyatakan aksi yang dilakukan dalam suatu arsitektur sistem

2.7.3 Sequence Diagram

"*Sequence Diagram* adalah *tool* yang sangat populer dalam pengembangan sistem informasi secara *object-oriented* untuk menampilkan interaksi antar objek. Berdasarkan definisi tersebut, dapat disimpulkan bahwa *Sequence Diagram* adalah *tool* yang digunakan dalam pengembangan *system*" (Yunahar, 2018).

Tabel 2. 3 Simbol *Sequence Diagram* (Yunahar, 2018)

Nama Komponen	Keterangan	Simbol
Lifeline	Mengindikasikan keberadaan sebuah objek dalam basis waktu. Notasi untuk *lifeline* adalah garis putus-putus vertikal yang ditarik dari sebuah objek	
Activation	Dinotasikan sebagai sebuah kotak segi empat digambar pada sebuah lifeline mengindentikasikan sebuah objek yang akan melakukan sebuah aksi	
Message	Digambarkan dengan anak panah horizontal antara activation. Message mengindekasikan komunikasi antara objek-objek	
Object	Merupakan *instance* dari sebuah class dan dituliskan tersusun secara horizontal.	
Actor	*Actor* juga dapat berkomunikasi	

2.7.4 Deployment Diagram

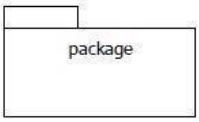
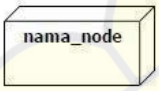
Menurut Sukamto dan Shalahudin (2015) mengemukakan bahwa "*Deployment Diagram* menunjukan konfigurasi komponen dalam *proses* eksekusi aplikasi". *Deployment Diagram* dapat digunakan untuk menggambarkan hal-hal sebagai berikut:

1. Sistem tambahan *(embedded system*) yang menggambarkan rancangan *device, node dan hardware*.
2. Sistem *client* atau *server*.
3. Sistem terdistribusi murni.
4. Rekayasa ulang aplikasi.

Menurut Hendini (2016) *Deployment Diagram* digunakan untuk menggambarkan detail bagaimana komponen disusun di infrastruktur sistem.

Dapat disimpulkan bahwa *Deployment Diagram* merupakan diagram yang digunakan untuk menggambarkan konfigurasi pada sistem.

Tabel 2. 4 Simbol *Deployment Diagram* (Rosa dan M.Shalahuddin, 2016)

Simbol	Deskripsi
Package 	package merupakan sebuah bungkusan dari satu atau lebih *node*
Node 	biasanya mengacu pada perangkat keras (*hardware*), perangkat lunak yang tidak dibuat sendiri (*software*), jika di dalam *node* disertakan komponen untuk mengkonsistenkan rancangan maka komponen yang diikutsertakan harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen
Kebergantungan / *dependency*	Kebergantungan antar *node*, arah panah mengarah pada *node* yang dipakai
Link	relasi antar *node*

2.8 Metodologi Pengembangan Sistem

Metodologi Agile adalah sekelompok metodologi pengembangan perangkat lunak yang didasarkan pada prinsip-prinsip yang sama atau pengembangan sistem jangka pendek yang memerlukan adaptasi cepat dari pengembang terhadap perubahan dalam bentuk apapun.

Agile development merupakan salah satu dari Metodologi pengembangan perangkat lunak yang digunakan dalam pengembangan perangkat lunak.

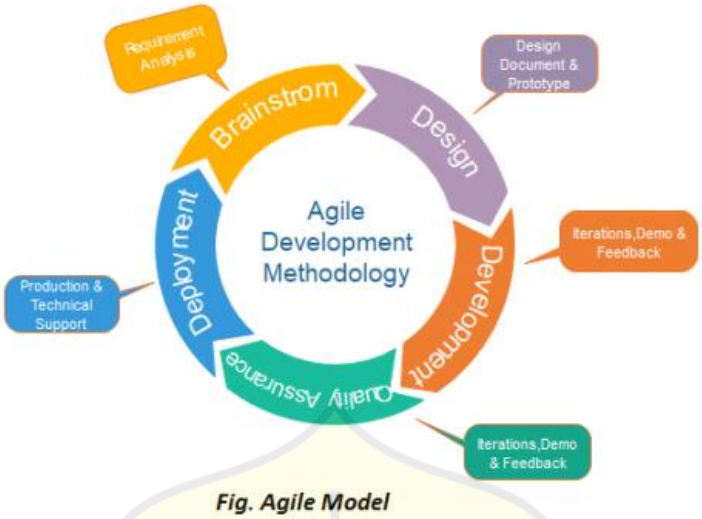


Fig. Agile Model

Gambar 2. 4 *Metodologi Agile*

2.8.1 Requirement

Dalam langkah ini merupakan analisa terhadap kebutuhan sistem. Pengumpulan data dalam tahap ini bisa melakukan sebuah penelitian, wawancara atau *study literatur*. Seorang *system* analisis akan menggali informasi sebanyak-banyaknya dari *user* sehingga akan tercipta sebuah sistem komputer yang bisa melakukan tugas-tugas yang diinginkan oleh *user* tersebut. Tahapan ini akan menghasilkan dokumen *user requirement* atau bisa dikatakan sebagai data yang berhubungan dengan keinginan *user* dalam pembuatan sistem. Dokumen inilah yang akan menjadi acuan *system analisis* untuk menterjemahkan kedalam bahasa pemrograman.

2.8.2 Design

Proses *design* akan menterjemahkan syarat kebutuhan kesebuah perancangan perangkat lunak yang dapat diperkirakan sebelum dibuat koding. Proses ini berfokus pada struktur data, arsitektur perangkat lunak, representasi

interface, dan *detail* (algoritma) prosedural. Tahapan ini akan menghasilkan dokumen yang disebut *software requirement*. Dokumen inilah yang akan digunakan *programmer* untuk melakukan aktivitas pembuatan sistemnya.

2.8.3 Development

Tahapan inilah yang merupakan tahapan secara nyata dalam mengerjakan suatu sistem. Dalam artian penggunaan *computer* akan dimaksimalkan dalam tahapan ini. Setelah pengkodean selesai maka akan dilakukan *testing* terhadap *system* yang telah dibuat tadi. Tujuan *testing* adalah menemukan kesalahan-kesalahan terhadap sistem tersebut dan kemudian bisa diperbaiki.

2.8.4 Testing

Disini Perangkat Lunak yang telah di buat di tes oleh bagian *control* kualitas agar *Bug* yang ditemukan bisa segera diperbaiki dan kuliatas perangkat lunak terjaga.

2.8.5 Deployment

yaitu proses yang dilakukan oleh penjamin kualitas untuk menguji kualitas sistem. Setelah sistem memenuhi syarat maka perangkat lunak siap *dideployment*.

2.8.6 Review

Tidak ada perangkat lunak yang 100% bebas dari *bug*, oleh karena itu sangatlah penting agar perangkat lunak dipelihara secara berkala.