

BAB II

LANDASAN TEORI

2.1. Data Mining

Menurut Kusriani dan Luthfi, E. T. (2009) Dalam Buku “Algoritma Data Mining”, Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan di dalam database. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Menurut Fajrin, A. A., Maulana, A. (2018:29) Data Mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan Data Mining adalah kenyataan bahwa Data Mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, Data Mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- a. Jumlah data yang sangat besar
- b. Dimensi data yang tinggi
- c. Data yang heterogen dan berbeda sifat

Pengelompokan Data Mining dibagi menjadi beberapa kelompok yaitu :

- a. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan record lengkap yang menyediakan nilai variable target sebagai nilai prediksi.

c. Prediksi

Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

d. Klasifikasi

Dalam klasifikasi terdapat target variable kategori, missal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.

e. Pengklasteran

Merupakan pengelompokan record, pengamatan, atau memperhatikan dari membentuk kelas objek-objek yang memiliki kemiripan.

f. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.2. Metode ECLAT

Menurut Wijaya, Malik, Nurmaini (2020:368). Algoritma ECLAT adalah salah satu algoritma di dalam metode Association rules yang memiliki waktu pencarian frequent itemset lebih cepat di banding dengan metode Apriori dengan teknik pendekatan BDF

(Breath deep First) dengan melakukan scan pada database sebanyak 3(tiga) kali yang mentransformasikan table yang dapat mengurangi proses input dan output.

Metode ECLAT merupakan algoritma yang diterapkan untuk menemukan itemset yang sering muncul pada basis data. Algoritma ini pada dasarnya melakukan pencarian secara Depth First Search (DFS) dengan tata letak vertical dan jika database berbentuk horizontal terlebih dahulu menjadi vertical. Algoritma ECLAT di bagi menjadi 3 tahapan yaitu :

1. Inisialisasi yaitu konstruksi perhitungan global untuk frequent 2-itemset.
2. Tahap Transformasi, yaitu membagi Frequent 2-itemset dan menjadwalkan partisi tersebut di prosessor, yang selanjutnya melakukan transformasi vertical pada database.
3. Tahap Asynchronous, adalah pembentukan frequent k-itemset sesuai kebutuhan atau tujuan yang ingin di capai.

2.3. Metode FP-Growth

Menurut Fajrin, A. A., Maulana, A. (2018:30) Frequent Pattern Growth (FP-Growth) adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data. Algoritma FP-Growth merupakan pengembangan dari algoritma Apriori. Sehingga kekurangan dari algoritma Apriori diperbaiki oleh algoritma FP-Growth .

Menurut Wijaya dkk (2020:367) FP-Growth merupakan salah satu algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (Frequent

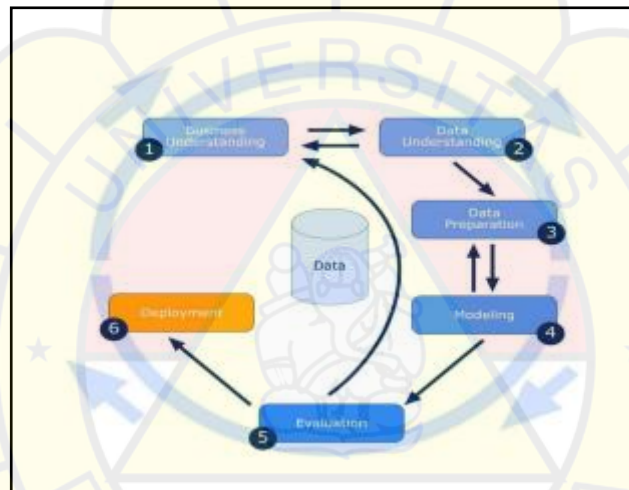
itemset) dalam sebuah kumpulan data. Karakteristik algoritma FP-Growth adalah struktur dari data digunakan adalah tree atau disebut dengan FP-Tree. Dengan menggunakan FP-Tree, algoritma ini dapat langsung mengekstrak frequent itemset dari FP-Tree.

FP-Tree merupakan struktur penyimpanan data yang dimampatkan. FP-tree dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam FP-tree. Karena dalam setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki item yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki item yang sama, maka proses pampatan dengan struktur data FP-tree semakin efektif. Kelebihan dari FP-tree adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien. Metode FP-Growth dapat dibagi menjadi 3 tahapan yaitu:

1. Tahap pembangkitan conditional pattern base, merupakan sub database yang berisi prefix path dan suffix pattern yang di dapat dari dari FP-tree yang telah di bentuk.
2. Tahap pembangkitan conditional FP-Tree, support count dari setiap item pada setiap conditional pattern base di jumlahkan, kemudian setiap item yang memiliki jumlah support lebih besar atau sama dengan nilai minimum support yang akan di bangkitkan.
3. Tahap pencarian frequent itemset, jika conditional FP-tree adalah lintasan tunggal (single path), maka bias di dapat frequent pattern dari kombinasi item untuk setiap conditional FP-Tree, dan jikan bukan lintasan tunggal, maka lakukan pembangkitan secara rekursif.

2.4. CRISP-DM

Menurut Hasanah, dkk (2021:104) CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah. Metodologi ini terdiri dari enam tahapan yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Proses metodologi ini terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut.



Gambar 2.1 Metodologi CRISP-DM (Hasanah,2021)

1. *Businnes Understanding* (Pemahaman Bisnis)

Beberapa hal yang dilakukan pada tahap ini seperti memahami kebutuhan serta tujuan dari sudut pandang bisnis selanjutnya mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data mining.

2. *Data Understanding* (Pemahaman Data)

Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, serta mengevaluasi kualitas data.

3. *Data Preparation* (Persiapan Data)

Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang akan dilakukan mencakup melakukan pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), record dan atribut-atribut, dan juga melakukan transformasi terhadap data (*Data Transformation*) untuk dijadikan masukan dalam tahap pemodelan.

4. *Modelling* (Pemodelan)

Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik data mining, alat bantu data mining, dan algoritma data mining yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap data preparation.

5. *Evaluation* (Pengujian)

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama.

6. *Deployment* (Penyebaran)

Tahap deployment atau rencana penggunaan model adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk Deployment dimulai selama Business Understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional. Pada akhirnya, rencana sistem Deployment mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data. Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

2.5. UML(*Unified Modelling Language*)

Menurut Muslihudin, M., & Oktafianto Dalam Buku Berjudul “Analisis dan Perancangan Sistem Informasi Menggunakan Model Terstruktur dan UML”. UML singkatan dari *Unified Modeling Language* yang berarti bahasa pemodelan standar. UML memiliki sintaks dan semantik. Ketika kita membuat model menggunakan konsep UML ada aturan-aturan yang harus diikuti. UML bukan hanya sekedar diagram, tetapi juga menceritakan konteks dalam membuat suatu sistem yang akan dibuat.

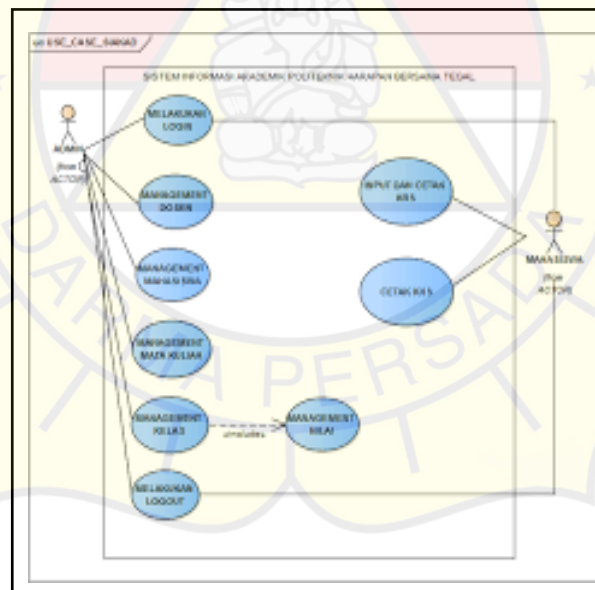
Menurut Prihandoyo, M. T.(2018:127) “*Unified Modeling Language* merupakan salah satu metode pemodelan visual yang digunakan dalam perancangan dan pembuatan sebuah software yang berorientasikan pada objek UML merupakan sebuah standar penulisan atau semacamnya blue print dimana didalamnya termasuk sebuah bisnis proses,

penulisan kelas-kelas dalam sebuah bahasa yang spesifik”. Terdapat beberapa jenis-jenis diagram di dalam UML sebagai berikut:

2.5.1. Use Case Diagram

Menurut Prihandoyo, M. T.(2018:127) “ *Use Case* Merupakan gambaran dari fungsionalitas yang di harapkan dari sebuah sistem dan merepresentasikan sebuah interaksi antara aktor dan sistem. Didalam *use case* terdapat aktor yang merupakan pekerjaan di sistem”. Diagram *Use Case* bersifat statis dan sangat penting untuk mengorganisasi dan memodelkan perilaku sistem yang dibutuhkan serta diharapkan pengguna.

Contoh Diagram Use Case :

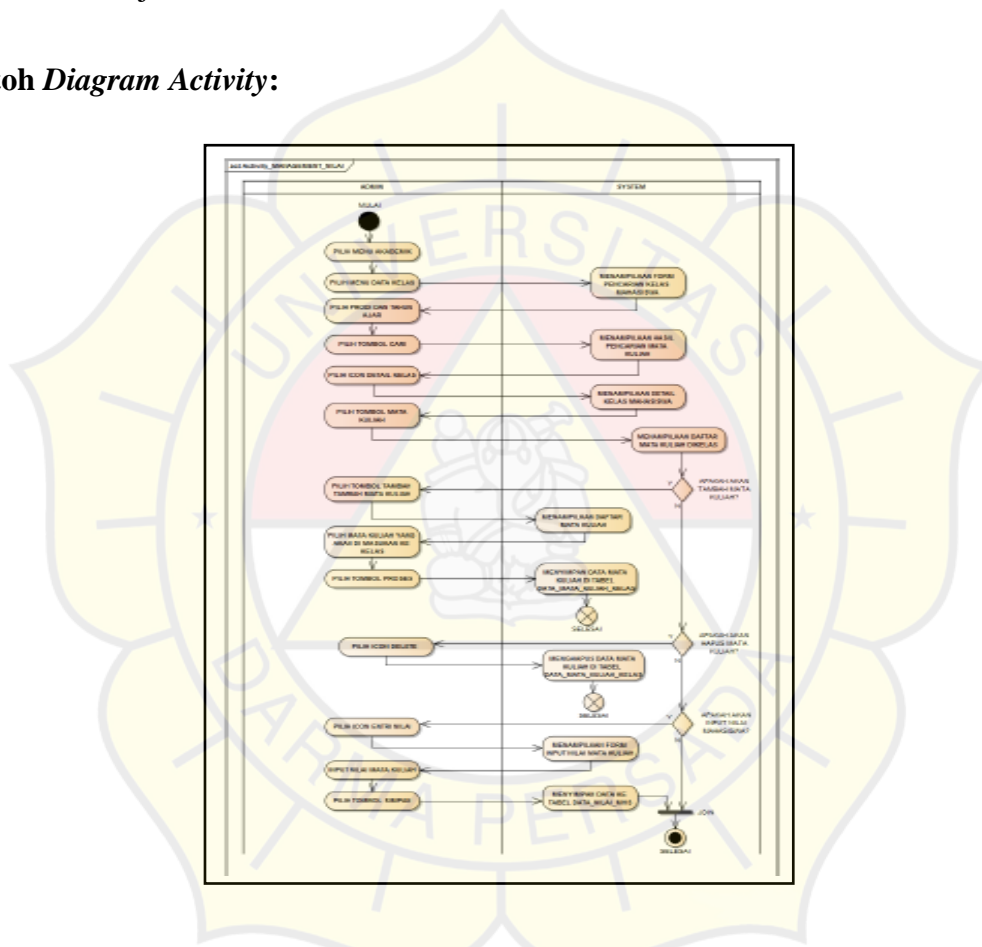


Gambar 2.2 Diagram Use Case (M Teguh Prihandoyo, 2018)

2.5.2. Activity Diagram

Menurut Prihandoyo, M. T.(2018:127) “Activity Diagram Merupakan gambaran air dari aktivitas-aktivitas didalam sistem yang berjalan”. Activity diagram ini bersifat dinamis dalam pemodelan fungsi-fungsi suatu sistem dan memberi tekanan pada aliran kendali antar objek.

Contoh Diagram Activity:



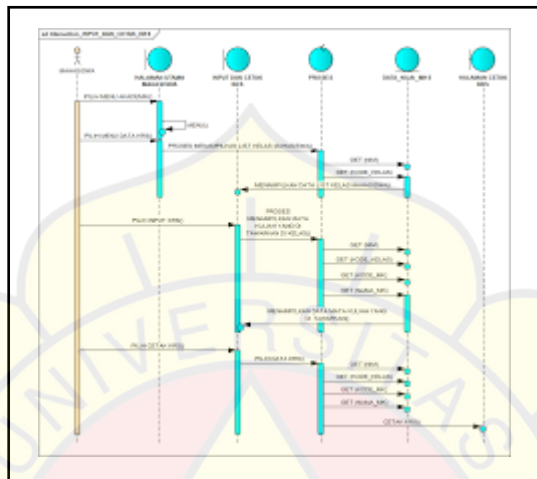
Gambar 2.3 Diagram Activity (M Teguh Prihandoyo, 2018)

2.5.3. Sequence Diagram

Menurut Prihandoyo, M. T.(2018:127) “Sequence Diagram adalah Menggambarkan interaksi antar objek didalam dan di sekitar sistem yang berupa message yang digambarkan terhadap waktu”. Sequence Diagram bias menggambarkan aktivitas

objek pada use case dengan mendeskripsikan waktu hidup objek serta pesan yang dikirimkan serta diterima antar objek.

Contoh Diagram Sequence:



Gambar 2.4 Diagram Sequence (M Teguh Prihandoyo, 2018)

2.6. Website

Menurut Sintinjak, dkk (2020) website adalah sering juga disebut web, dapat diartikan suatu kumpulan-kumpulan halaman yang menampilkan berbagai macam informasi teks, data, gambar diam ataupun bergerak, data animasi, suara, video maupun gabungan dari semuanya, baik itu yang bersifat statis maupun yang dinamis, yang dimana membentuk satu rangkaian bangunan yang saling berkaitan dimana masing-masing dihubungkan dengan jaringan halaman atau hyperlink.

2.7. PHP

Menurut Rahmasari, T.(2019:414) PHP merupakan singkatan dari “Hypertext Preprocessor”. PHP adalah sebuah bahasa scripting yang terpasang pada HTML. Sebagian besar sintaknya mirip dengan bahasa pemrogramn C, Java, ASP dan Perl

ditambah beberapa fungsi PHP yang Spesifik dan mudah dimengerti. PHP digunakan untuk membuat tampilan web menjadi lebih dinamis, dengan PHP anda bisa menampilkan atau menjalankan beberapa file dalam 1 file dengan cara di include dan require. PHP itu sendiri sudah dapat berinteraksi dengan beberapa database walaupun dengan kelengkapan yang berbeda yaitu seperti DBM, MySQL, Oracle .

2.8. Database

Menurut Ultariani, N. dkk (2020:221) Database adalah suatu aplikasi yang menyimpan sekumpulan data. Setiap database mempunyai API tertentu untuk membuat, mengakses, mengatur, mencari, menyalin data yang ada di dalamnya. Database yaitu kumpulan file-file yang berhubungan satu dengan yang lainnya, diatur sedemikian rupa sehingga dapat digunakan oleh beberapa program aplikasi database.

2.9. MySQL

Menurut Betha Sidik dalam buku yang “Pemrograman web dengan PHP (2012 : 333)” menyebutkan bahwa : “*MySQL* merupakan software database yang termasuk paling populer di lingkungan Linux, kepopuleran ini karena ditunjang karena performansi query dari database nya yang saat itu bisa dikatakan paling cepat dan jarang bermasalah”. *MySQL* merupakan aplikasi database server. *SQL* kepanjangan dari *Structured Query Language*. *SQL* merupakan bahasa terstruktur yang digunakan untuk mengolah database. *MySQL* dapat digunakan untuk membuat dan mengelola data di dalam database. Fungsi dari *MySQL* adalah untuk menambahkan mengubah, dan menghapus data di dalam database. *MySQL* berawal dari proyek yang dimulai oleh kedua orang developer, yakni Michael Widenius dan David Axmark di tahun 1994. Proyek ini didasari karena ingin

membuat suatu sistem database yang murah, meskipun ketika itu ada database yang power full yakni oracle, namun database ini bersifat komersil yang harganya mahal, dan begitu menguasai pasar.

Menurut Rahmasari, T.(2019:414) “Definisi MySQL merupakan software RDMS (Relational Database Management System) yang dapat mengelola database dengan sangat cepat, dapat menampung data dalam jumlah sangat besar, dapat diakses oleh banyak pengguna dan dapat melakukan suatu proses secara sinkron atau bersamaan”.

Menurut Sintinjak, dkk (2020) “MySQL adalah sebuah software database. MySQL merupakan tipe data relasional yang artinya MySQL menyimpan datanya dalam bentuk table-tabel yang saling berhubungan. Keuntungan menyimpan data di database adalah kemudahannya dalam penyimpanan dan menampilkan data karena dalam bentuk tabel”.