

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. *Data Mining*

Menurut (Deny Jollyta, William Ramdhan, Muhammad Zalis"Konsep Data Mining Dan Penerapan",2020:22) menyatakan bahwa *data mining* merupakan proses untuk menggali (*mining*) pengetahuan dan informasi baru dari data yang berjumlah banyak pada *data warehouse*,dengan menggunakan kecerdasan buatan (*Artificial Intelligence*),*statistic* dan matematika. *Data mining* merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pemakainya. Beberapa teknik dan sifat *data mining* menurut (Hermawati, 2013:14) adalah sebagai berikut :

(a) Klasterisasi

Adalah mempartisi data-set menjadi beberapa *sub-net* atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set property yang di share bersama, dengan tingkat similaritas yang tinggi dalam suatu kelompok yang rendah. Disebut juga dengan "*unsupervised learning*"

(b) Regresi

Adalah memprediksi nilai dari suatu variabel *kontinyu* yang diberikan berdasarkan nilai adari variabel yang lain, dengan mengasumsikan sebuah model ketergantungan *linier atau nonlinier*

(c) Klasifikasi

Adalah menentukan sebuah *record* data baru ke salah satu dari beberapa kategori (kelas) yang telah didefinisikan sebelumnya dan disebut juga dengan "*supervised learning*"

(d) Kaidah Asosiasi (*association rule*)

Adalah mendeteksi kumpulan atribut-atribut yang muncul bersamaan (*co-occur*) dalam frekuensi yang sering dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut

2. Klasifikasi

Menurut Han dan Kamber (2006:285) *Classification* adalah sebuah model dalam data mining dimana, *classifier* dikonstruksi untuk memprediksi categorical label, seperti "aman" atau "beresiko" untuk data aplikasi peminjaman uang; "ya" atau "tidak" untuk data marketing; atau "*treatment A*", "*treatment B*" atau "*treatment C*" untuk data medis, kategori tersebut dapat

direpresentasikan dengan nilai yang sesuai dengan kebutuhannya, dimana pengaturan nilai tersebut tidak memiliki arti tertentu.

3. Algoritma C4.5

Menurut Sukma, Halfis dan Hermawan (2019, p. 23) Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross quunlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 (dua) prinsip kerja, yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*). Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

Terdapat 4 langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

- (a) Memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
- (b) Membuat cabang untuk masing-masing nilai, artinya membuat cabang sesuai dengan jumlah nilai variabel gain tertinggi.
- (c) Membagi setiap kasus dalam cabang, berdasarkan perhitungan nilai gain tertinggi dan perhitungan dilakukan setelah perhitungan nilai gain tertinggi awal dan kemudian dilakukan proses perhitungan gain tertinggi kembali tanpa meyeritakan nilai variabel gain awal.
- (d) Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama, mengulangi semua proses perhitungan gain tertinggi untuk masing- masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

Algoritma C4.5 secara rekursif mengunjung setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Dari ketiga peneliti yang telah dilakukan tersebut, klasifikasi dengan Algoritma C4.5 digunakan oleh para peneliti sebagai solusi untuk mengambil keputusan yang diharapkan mampu membantu dalam pengambilan keputusan dengan lebih mudah dan cepat.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu:

- (1) Menyiapkan data *training*. Data *training* biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
- (2) Menentukan akar dari pohon . akar akan diambil dari atribut yang terpilih dengan cara menghitung nilai Gain dari masing- masing atribut, nilai Gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai Gain dari atribut, hitung dahulu nilai entropy adapun formulasi untuk menghitung nilai entropy yaitu :

$$Entropy (S) = \sum - \pi * \log_2 \pi \quad i=1$$

Kemudian hitung nilai Gain, nilai gain merupakan selisih antara nilai pretest dan posttest.

Berikut rumusnya :

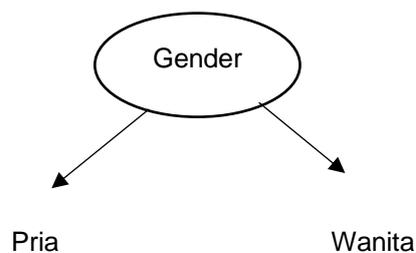
$$Gain (S, A) = Entropy (S) - \sum |s| * Entropy$$

Ulangi langkah ke-2 hingga semua semua tupel terpartisi.

Proses partisi pohon keputusan akan berhenti saat:

- (a) Semua tupel dalam node N mendapat kelas yang sama.
- (b) Tidak ada atribut di dalam tupel yang dipartisi lagi.
- (c) Tidak ada tupel di dalam cabang yang kosong.

Menurut (Elisa, 2017) Algoritma terkenal digunakan untuk mengklasifikasikan data. Hasil proses klasifikasi dalam bentuk *rule*.



Gambar 2. 1 Syarat pengujian fitur biner

Isi penting dari induksi pohon keputusan adalah bagaimana menyatakan kondisi pengujian pada node. Ada 3 kelompok persyaratan pengujian node yang penting:

1. Fitur biner

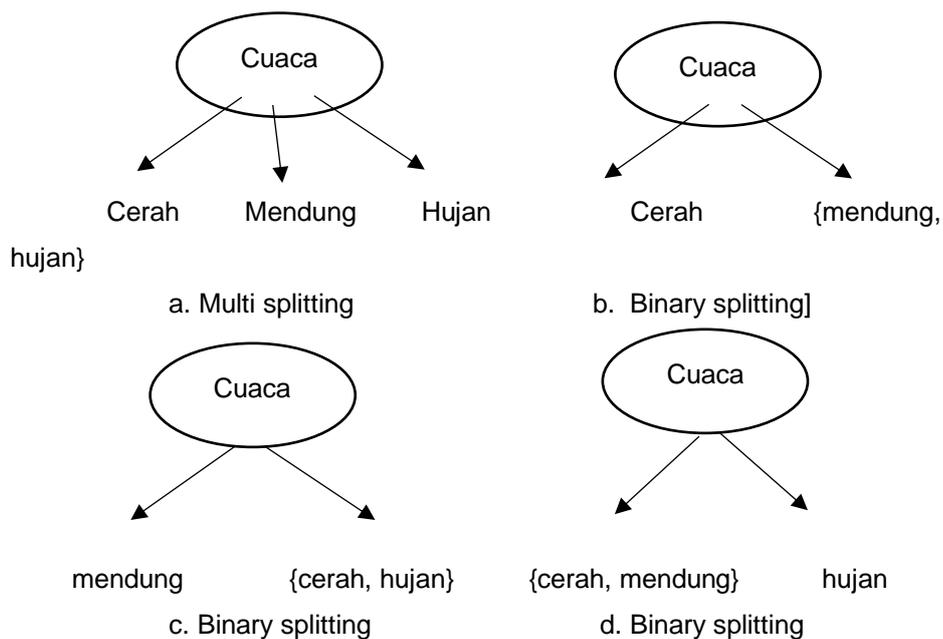
Fitur dengan Saat fungsi ini menjadi *node* (*node root* atau *node internal*), pengujian hanya memerlukan dua opsi cabang. Gambar 2.3 menunjukkan contoh solusi.

2. Fitur bertipe kategorial

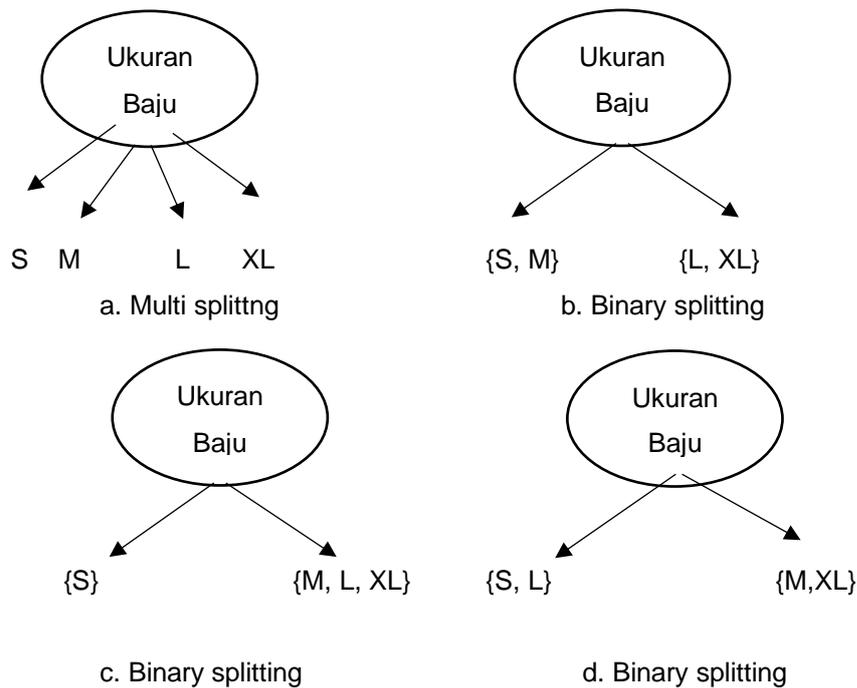
Untuk fitur yang nilainya merupakan tipe klasifikasi, fitur tersebut dapat memiliki beberapa nilai. Kombinasi tersebut ditunjukkan pada Gambar 2.3. Untuk solusi yang hanya mengizinkan solusi biner, seperti algoritme CART, mungkin terdapat kombinasi solusi $2^{(k-1)}$, di mana k adalah jumlah nilai yang berbeda dalam fitur tersebut. Contoh solusi ditunjukkan pada Gambar 2.3 dan Gambar 2.4.

3. Fitur bertipe numerik

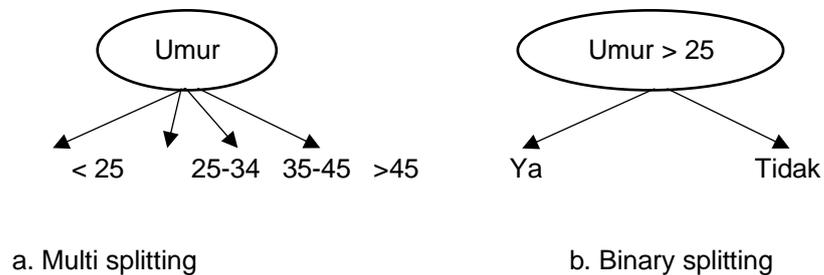
Untuk fitur numerik, kondisi pengujian node (root atau internal) diwakili oleh uji perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau untuk beberapa pengujian dengan beberapa nilai dalam bentuk hasil $v_i \leq A < v_{(i+1)}$, untuk $i = 1, 2, \dots, k$. Dalam kasus solusi biner, algoritme memeriksa semua kemungkinan posisi solusi v dan memilih posisi v terbaik. Gambar 2.4 menunjukkan contoh pemecahan fitur numerik.



Gambar 2. 2 Syarat pengujian fitur bertipe nominal



Gambar 2. 3 Syarat pengujian fitur bertipe ordinal



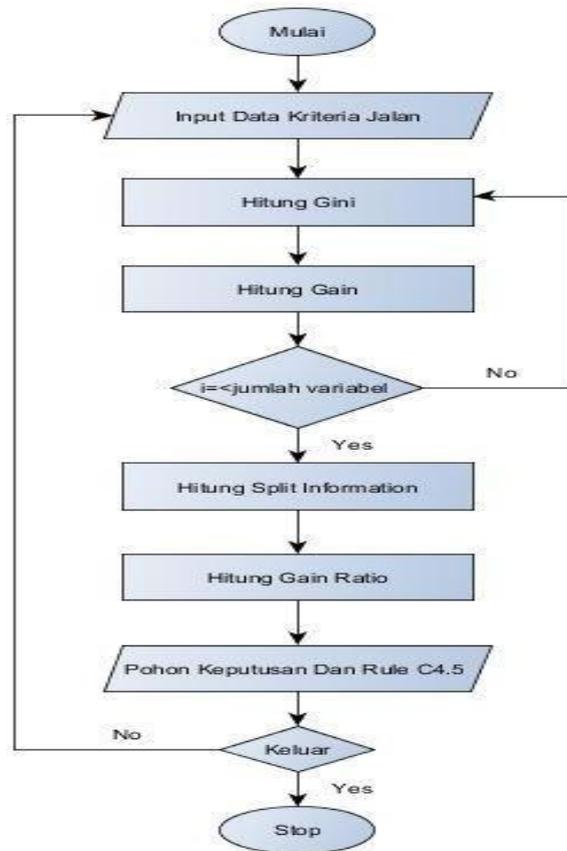
Gambar 2. 4 Syarat pengujian fitur bertipe numerik

Kriteria paling banyak yang diformulasikan oleh persamaan berikut:

$$RasioGain(s,j) = \frac{Gain(s,j)}{SplitInfo(s,j)}$$

Rumus di atas merepresentasikan rasio gain pada fitur j . $SplitInfo(s,j)$ didapat dari $SplitInfo(s,j) = -\sum_{i=1}^k p(v_i|s) \log_2 p(v_i|s)$ dimana k menyatakan jumlah pemecahan. (Prasetyo, 2014). Adapun gambaran algoritma C4.5 menggunakan flowchart yang dapat dilihat pada gambar 2.6.

Berikut adalah flowchart dari algoritma C4.5 :



Gambar 2. 5 Flowchart Algoritma C4.5

Berdasarkan flowchart gambar 2.6 alur algoritma C4.5 dapat diuraikan sebagai berikut :

- (1) Memasukan data set yang telah disediakan. Data set adalah kumpulan objek dan atributnya.
- (2) Pada proses cek keputusan dilihat apakah *data set* tersebut termasuk kedalam atribut kontinu atau bukan, atribut *kontinu* adalah atribut yang mempunyai jangkuan real.
- (3) Jika iya, maka akan menghitung, jika tidak maka menghitung *Gain* setiap atribut.
- (4) Setelah itu menentukan cabang/atribut yang memilih *Gain max* .

4. System Development Life Cycle (SDLC)

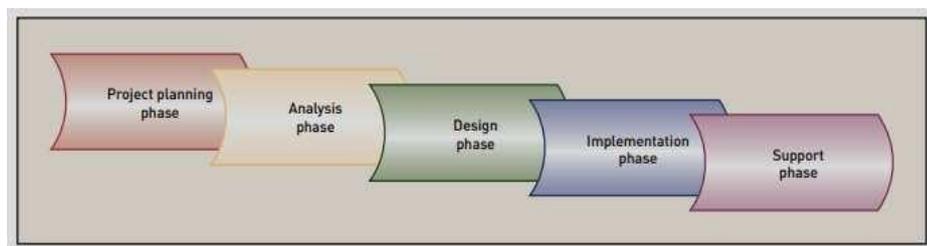
Menurut Simarmata (2010:39), SDLC mengacu pada model dan proses yang digunakan untuk mengembangkan sistem perangkat lunak dan

menguraikan proses, yaitu pengembang menerima perpindahan dari permasalahan ke solusi.

SDLC digunakan untuk membangun suatu sistem informasi agar dapat berjalan sesuai dengan apa yang diharapkan. SDLC (*Systems Development Life Cycle*, Siklus Hidup Pengembangan Sistem) atau *Systems Life Cycle* (Siklus Hidup Sistem), dalam rekayasa sistem dan rekayasa perangkat lunak, adalah proses pembuatan dan perubahan sistem serta model dan metodologi yang digunakan untuk mengembangkan sistem-sistem tersebut. Konsep ini umumnya merujuk pada sistem komputer atau informasi. SDLC juga merupakan pola yang diambil untuk mengembangkan sistem perangkat lunak, yang terdiri dari tahap-tahap: rencana (*planning*), analisis (*analysis*), desain (*design*), implementasi (*implementation*), uji coba (*testing*) dan pengelolaan (*maintenance*).

Dalam rekayasa perangkat lunak, konsep SDLC mendasari berbagai jenis metodologi pengembangan perangkat lunak. Metodologi-metodologi ini membentuk suatu kerangka kerja untuk perencanaan dan pengendalian pembuatan sistem informasi, yaitu proses pengembangan perangkat lunak. Terdapat 3 jenis metode siklus hidup sistem yang paling banyak digunakan, yakni: siklus hidup sistem tradisional (*traditional system life cycle*), siklus hidup menggunakan prototyping (*life cycle using prototyping*), dan siklus hidup sistem orientasi objek (*object-oriented system life cycle*).

Pada umumnya SDLC memiliki 5 (lima) fase utama yaitu fase perencanaan (*Project Planning*), fase analisis (*Analysis*), fase design (*Design*), fase implementasi (*Implementation*) dan fase dukungan (*Support*).



Gambar 2. 6 Fase – fase dalam SDLC

Sumber : Satzinger, Jackson, & Burd, (2010, p40)

Tabel 2. 1 Fase dan Tugas Dalam SDLC

Fase	Tugas
<i>Project Planning</i>	Untuk mengidentifikasi ruang lingkup sistem baru, memastikan bahwa proyek ini layak, dan mengembangkan jadwal, rencana sumber daya, dan anggaran dari proyek
<i>Analysis</i>	Untuk memahami dan merincikan kebutuhan bisnis dan persyaratan pengolahan sistem baru
<i>Design</i>	Untuk merancang sistem yang menghasilkan solusi berdasarkan persyaratan yang ditetapkan dan keputusan yang dibuat selama analisis
<i>Implementation</i>	Untuk membangun, menguji, dan memasang sistem informasi yang handal dengan pengguna dilatih siap untuk mendapatkan keuntungan seperti yang diharapkan dari penggunaan sistem
<i>Support</i>	Untuk menjaga sistem agar mampu berjalan secara produktif, baik pada awalnya dan selama bertahun-tahun hidup sistem

Sumber : Satzinger, Jackson, & Burd, (2010, p40)

5. Metode Prototype

Menurut (Ogedebe, 2012), *prototyping* adalah metode pengembangan perangkat lunak, yang merupakan model fisik kerja sistem dan merupakan versi awal dari sistem. Dengan menggunakan metode prototipe ini, prototipe sistem akan diproduksi sebagai perantara antara pengembang dan pengguna sehingga dapat berinteraksi dalam proses kegiatan pengembangan sistem informasi.

Tujuan pembuatan *prototipe* untuk pengembang sistem adalah untuk mencari informasi dari pengguna. (Ogedebe, 2012) Perancangan sistem, khususnya untuk pemrosesan pembayaran, semakin cepat proses pengembangan sistem informasi, dan semakin banyak interaksi antar

pengguna dalam proses pembangunan, maka semakin besar pula manfaat yang didapat.

Prototipe tersebut dapat diaplikasikan pada pengembangan sistem skala kecil dan besar, diharapkan proses pembangunan dapat berjalan dengan baik, teratur dan selesai tepat waktu.

Manfaat lain menggunakan *prototipe* adalah:

1. Menerapkan sistem aktual dalam salinan sistem yang akan dijalankan, dan mengakomodasi masukan user untuk melengkapi sistem.
2. User akan lebih bersedia dikembangkan berdasarkan jalannya prototipe hingga hasil pengembangan akhir akan dijalankan nanti.
3. Prototipe ditambahkan atau kurang sesuai dengan proses pengembangan. Pengguna dapat melacak kemajuan langkah demi langkah.
4. Menghemat sumber daya dan waktu, serta menghasilkan produk yang lebih baik dan lebih efisien bagi pengguna.

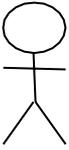
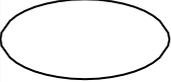
6. ***Unified Modeling Language (UML)***

Menurut Rosa dan Shalahuddin (2014:133), berpendapat bahwa UML (*Unified Modeling Language*) adalah “Salah standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan *requirement*, membuat analisa & desain, serta menggambarkan arsitektur dalam pemrograman berorientasi objek”. Dari beberapa penjelasan teori tersebut dapat disimpulkan bahwa UML (*Unified Modeling Language*) adalah bahasa yang sering digunakan untuk membangun sebuah sistem perangkat lunak dengan melakukan penganalisaan desain dan spesifikasi dalam pemrograman berorientasi objek.

Menurut Fowler, M. dalam (B. O. Lubis, 2016) UML (*Unified Modeling Language*) adalah “Keluarga notasi grafis yang didukung oleh meta-model tunggal, yang membantu pendeskripsian dan desain sistem perangkat lunak, khususnya sistem yang dibangun menggunakan pemrograman berorientasi objek (OO). Definisi ini merupakan definisi yang sederhana”. UML (*Unified Modeling Language*) dapat diartikan sebagai Bahasa visual untuk menggambarkan definisi-definisi tentang requirement, membuat analisis dan desain serta menggambar arsitektur dalam pemrograman berorientasi objek dengan menggunakan teks-teks pendukung.

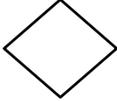
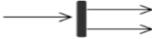
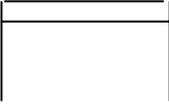
Berikut ini simbol UML, menurut Dharwiyanti Pengantar 2003:2 diantaranya adalah :

Tabel 2. 2 Simbol Use case Diagram

GAMBAR	NAMA	KETERANGAN
	<i>ACTOR</i>	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri.
	<i>USE CASE</i>	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor .
	<i>ASOSIASI/ ASSOCIATION</i>	Komunikasi antara actor dan use case, yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.
<<extend>>	<i>EKSTENSI/ EXTEND</i>	Relasi usecase tambahan ke sebuah usecase dimana use case yang ditambahkan dapat berdiri sendiri.
	<i>GENERALISASI / GENERALIZATION</i>	Hubungan generalisasi dan spesialisasi (umum-khusus) dua buah use case .

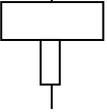
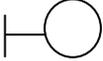
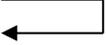
Sumber (Dharwiyanti Pengantar UML 2003:2)

Tabel 2. 3 Simbol Activity Diagram

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

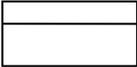
Sumber (Dharwiyanti Pengantar UML 2003:2)

Tabel 2. 4 Simbol Sequence Diagram

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Merepresentasikan entitas satu sistem dan berinteraksi dengan sistem lainnya.
	<i>Lifeline</i>	Penghubungan antara objek satu dengan objek yang lain
	<i>General</i>	Merepresentasikan entitas tunggal
	<i>Boundary</i>	Berupa tepi dari sistem
	<i>Control</i>	Elemen mengatur aliran dari informasi untuk sebuah skenario.
	<i>Entitas</i>	Elemen yang bertanggung jawab menyimpan atau informasi..
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan/hubungan antar objek .
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri.
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman message .

Sumber (Dharwiyanti Pengantar UML 2003:2)

Tabel 2.5. Simbol Class Diagram

GAMBAR	NAMA	KETERANGAN
	<i>Generalization</i>	Hubungan dimana objek anak (descendent) berbagi perilaku
	<i>Nary Association</i>	Upaya untuk asosiasi dengan lebih dari 2 objek.
	<i>Class</i>	Himpunan objek-objek yang berbagi atribut serta operasi yang sama.
	<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem.
	<i>Realization</i>	Operasi yang dilakukan oleh suatu objek
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen .
	<i>Association</i>	Perhubungan antara objek satu dengan obyek lainnya.

Sumber (Dharwiyanti Pengantar UML 2003:2)

7. Bahasa Pemrograman

a. PHP

Menurut Supono & Putratama (2018: 1) mengemukakan bahwa “PHP (PHP: *hypertext preprocessor*) adalah suatu bahasa pemrograman yang digunakan untuk menterjemahkan basis kode program menjadi kode mesin yang dapat dimengerti oleh komputer yang bersifat *server-side* yang ditambahkan ke HTML”.

Kegunaan PHP adalah untuk membuat website dinamis bisa digunakan untuk menyimpan data ke dalam database, membuat halaman yang dapat berubah-ubah sesuai dengan input *user*, memproses *form*, membuat dan membuka file dalam suatu sistem, PHP dapat menangani formulir, seperti mengambil data-data tersebut dari file, menyimpan data tersebut dalam bentuk file, dapat mengirimkan data melalui email, dan mengembalikan data ke pengguna, Anda dapat menambahkan, menghapus, dan mengubah elemen dalam *database*.

b. *Hypertext Markup Language (HTML)*

Menurut (Wardana, 2016:3) *Hypertext markup language (HTML)* merupakan bahasa pemrograman dasar untuk mengelola website. Akan tetapi, HTML hanya terbatas pada pembuatan website statis (website yang tidak dapat berinteraksi aktif dengan user). Maka dari itu, HTML biasa dikombinasikan dengan bahasa pemrograman web lainnya.

Kegunaan HTML adalah untuk mengelola serangkaian data dan informasi sehingga suatu dokumen dapat diakses dan ditampilkan di Internet melalui layanan web, membuat halaman web, menampilkan berbagai informasi di dalam sebuah browser internet, membuat link menuju halaman web lain dengan kode tertentu (*hypertext*).

c. *My SQL*

Menurut Risnandar (2013:92) mendefinisikan bahwa “*My SQL* merupakan basis data yang bersifat open source sehingga banyak digunakan di dunia. Walaupun gratis, *My SQL* tetap berkualitas dan sudah cukup memberikan performa yang memadai”. *My SQL* merupakan *software database* open source yang sering digunakan untuk mengolah basis data yang menggunakan bahasa *SQL* (Subagia, 2018:67). Dapat ditarik kesimpulan bahwa *My SQL* merupakan aplikasi yang digunakan untuk mengolah basis data yang banyak digunakan untuk membangun aplikasi yang menggunakan *database*.

Kegunaan *My SQL* adalah untuk membuat dan mengelola *database* pada sisi server yang memuat berbagai informasi dengan menggunakan bahasa

SQL, memudahkan pengguna dalam mengakses data berisi informasi dalam bentuk *String* (teks), yang dapat diakses secara personal maupun publik dalam web.

d. Database

Menurut Ladjamudin (2013:129), *Database* adalah sekumpulan *data store* (bisa dalam jumlah yang sangat besar) yang tersimpan dalam *magnetic disk, optical disk, magnetic drum*, atau media penyimpanan sekunder lainnya. Dapat disimpulkan bahwa *Database* adalah sekumpulan file yang saling berhubungan yang menyimpan data dan tersimpan dalam sebuah media penyimpanan. Kegunaan dari *database* yaitu Mengelompokkan data dan informasi sehingga lebih mudah dimengerti, mencegah terjadinya duplikat data maupun inkonsistensi data, mempermudah proses penyimpanan, akses, pembaharuan, dan menghapus data. menjaga kualitas data dan informasi yang diakses sesuai dengan yang diinput, dan membantu proses penyimpanan data yang besar.

e. Web Server (Apache)

Menurut Kustiyahningsih dan Devie (2011:8) "*web server* adalah komputer yang digunakan untuk menyimpan dokumen-dokumen web, komputer ini melayani permintaan dokumen web dari kliennya". Dapat disimpulkan web server adalah komputer yang digunakan untuk menyimpan dokumen dengan mengakses dan menampilkan halaman web tersebut dari komputer client.

Kegunaan dari web server sendiri yaitu seperti pembahasan pengertian diatas. Web server merupakan *software* yang akan memberikan layanan kepada user berupa halaman website sesuai dengan permintaan *user*. Lebih lengkapnya apabila seorang pengunjung internet dan melakukan pencarian menggunakan browser, maka permintaan dari pencarian tersebut akan masuk kedalam *web server*. Setelah permintaan tersebut masuk kedalam server, selanjutnya permintaan akan diproses oleh TCP (*Tansmission Control Protocol*) selaku protokol transport. Saat data sudah ditemukan, selanjutnya server akan mengirimkan kembali kepada *web server* melalui HTTP atau HTTPS dalam bentuk tampilan berupa halaman *website*. Akan tetapi jika server tidak berhasil menemukan data sesuai dengan permintaan, maka akan muncul notifikasi halaman error 404 atau page not found. Proses yang rumit tersebut nyatanya dapat berlangsung sangat cepat saat kita mengaksesnya.

f. **Intranet**

Menurut Prakoso (2007:119), Intranet adalah sebuah kumpulan jaringan komputer lokal yang menggunakan perangkat lunak internet dan protokol TCP/IP atau HTTP. Oleh karena itu, sebuah jaringan intranet memiliki semua fasilitas yang dimiliki oleh internet seperti e-mail, *File Transfer Protocol* (FTP), dan lain sebagainya. Kegunaan dari intranet adalah Menghubungkan komputer dalam satu organisasi/perusahaan, Menghubungkan *resource* dalam satu jaringan agar dapat digunakan bersama, Mengamankan jaringan dalam organisasi atau perusahaan dari gangguan pihak luar, dan Mengamankan jaringan dalam organisasi atau perusahaan dari gangguan pihak luar.

B. **Prediksi Nilai Ketuntasan Kriteria Minimal (KKM)**

Prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan.

Menurut Prayitno (2009: 418) Ketuntasan Kriteria Minimal (KKM) adalah suatu acuan penentu seorang siswa memenuhi kriteria persyaratan penguasaan materi pelajaran tertentu secara minimal. Sedangkan menurut Lukmanul Hakiim (2009: 22) KKM merupakan batas minimal pencapaian kompetensi yang harus dikuasai oleh siswa pada setiap mata pelajaran. Kriteria yang ideal untuk setiap indikator adalah sebesar 75% dan sekolah diharapkan untuk terus meningkatkan KKM hingga mencapai nilai ideal (100%).

Maka dari itu prediksi nilai ketuntasan kriteria minimal (KKM) adalah suatu saran yang dianjurkan bahwa adanya siswa yang nilainya tidak tercapai dalam memenuhi nilai ketuntasan kriteria minimal (KKM) dalam pembelajarannya.

C. **Tinjauan Studi**

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan kesamaan metode yang digunakan yaitu Algoritma C4.5. Banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain :

- (1) Penelitian ini dilakukan oleh **Aprianto Tumanngor, Penda Sudiarto Hasugian, Dengan judul “Penerapan Data Mining Untuk Memprediksi**

Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4.5 Pada SDN 105351 Bakaran Batu”

Perkembangan teknologi yang sangat maju dapat membantu segala aspek kehidupan, salah satunya dalam dunia pendidikan. Dunia pendidikan tentunya sangat membutuhkan peran teknologi dalam mengelola kegiatan dengan memanfaatkan teknologi informasi. Permasalahan yang muncul adalah sistem pengolahan tumpukan data untuk mengetahui mampu atau tidaknya siswa mengikuti mata pelajaran di SDN 105351 Bakaran Batu. Proses implementasi aplikasi data mining ini hanya membahas tentang memprediksi tingkat kemampuan anak dalam mengikuti mata pelajaran. Data yang digunakan adalah data siswa kelas 5 SD 2 tahun terakhir (2018-2019) di SDN 105351 Bakaran Batu. Salah satu model klasifikasi adalah membuat pohon keputusan. Dengan mengetahui tingkat kemampuan anak dalam mengikuti mata pelajaran siswa diharapkan dapat memberikan solusi untuk kedepannya dalam mengolah data siswa dengan sistem yang terkomputerisasi khususnya data mengenai tingkat kemampuan anak dalam mengikuti mata pelajaran. Cara kerja sistem dibangun menggunakan metode C4.5 dan aplikasi *data mining* yang akan dirancang dengan *database VB Net* dan *Microsoft Access*.

- (2) Penelitian ini dilakukan oleh **Oscario Oscario, Jasmir Jasmir, Yudi Novianto, Dengan judul “Penerapan Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar”** *Data Mining* merupakan serangkaian proses yang digunakan untuk mendapatkan informasi yang berguna dari database yang besar. SD Sariputra merupakan sekolah dasar yang didirikan pada tahun 1971. Pada proses pembelajaran di SD Sariputra Jambi, masih banyak murid yang tidak fokus atau tidak memiliki keinginan untuk belajar. Dimana kemungkinan terbesar adalah ketidakcocokan gaya belajar dan metode belajar guru tersebut. Oleh karena itu, penulis mempelajari dan melakukan analisis data mining terhadap data murid dan kepribadiannya agar data tersebut dapat berubah menjadi informasi yang berharga bagi SD Sariputra. Informasi ini diharapkan dapat membantu meningkatkan proses belajar mengajar di SD Sariputra. Penulis menggunakan hasil data quisioner pada siswa-siswi kelas 3 & 4 sebanyak 155 data siswa-siswi yang telah penulis sajikan dalam *file ekstensi arff*. Dalam melakukan analisis ini, penulis menggunakan *tools WEKA*. Metode yang digunakan adalah *Decision Tree* dengan Algoritma C4.5 dengan 11 atribut. Analisis tersebut menghasilkan 23 *Rule*. *Rule* tersebut dihasilkan dari hasil akhir bentuk *Decision Tree*.

- (3) Penelitian ini dilakukan oleh **Suherman, Marlia Purnamasari, Fitriani Dwi Hastuti Dengan Judul “Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas Minat Menggunakan Metode Decision Tree C4.5”** Kurikulum 2013 dirancang untuk memberikan kesempatan kepada siswa untuk belajar sesuai dengan minat siswa. Selain memilih mata pelajaran minat tertentu, siswa juga diberi kesempatan untuk mengambil mata pelajaran lintas minat. SMA Negeri 1 Anyer merupakan salah satu sekolah yang telah melaksanakan program lintas minat. Dalam proses penentuan kelas lintas minat, sekolah masih mengalami kendala yaitu tidak adanya kekhususan siswa yang memiliki minat pada mata pelajaran tertentu, dan dalam proses pemilihan pemilihan umum yang ditentukan oleh sekolah. Penelitian ini bertujuan untuk mengelompokkan siswa berdasarkan minat dan bakat siswa pada mata pelajaran tertentu. Metode yang digunakan adalah *Decision Tree* dan algoritma C4.5. Penelitian ini memperoleh nilai akurasi sebesar 82,82%. Penelitian menghasilkan sistem penentuan kelas lintas negara. Model klasifikasi ini dapat membantu siswa dalam menentukan lintas minat dan dapat dijadikan sebagai alternatif referensi bagi guru BK untuk dapat mengelompokkan siswa berdasarkan minat dan bakat siswa.
- (4) Penelitian ini dilakukan oleh **Nisa, Sofiatun, Dengan judul “Korelasi antara kemampuan berpikir kritis dengan hasil belajar siswa pada materi pencemaran lingkungan : Studi penelitian di kelas X SMA Negeri 1 Cibinong Cianjur Selatan”** Proses pembelajaran adalah kegiatan terarah untuk mencapai tujuan pembelajaran. Salah satu cara mengukur aspek-aspek dalam proses pembelajaran adalah dengan adanya hasil belajar. Untuk menunjang hasil belajar siswa terdapat aspek-aspek yang penting untuk dimiliki salah satunya adalah aspek kemampuan berpikir kritis. Karena dengan berpikir kritis manusia dapat menyelesaikan berbagai persoalan yang ada mengikuti jaman yang kian berubah. Penelitian ini bertujuan untuk menjelaskan hubungan antara kemampuan berpikir kritis siswa dengan hasil belajar pada materi pencemaran lingkungan. Metode penelitian yang digunakan adalah penelitian korelasional dengan analisis korelasional koefisien dan analisis regresi. Hasil penelitian menunjukkan bahwa terdapat korelasi positif dan signifikan antara kemampuan berpikir kritis siswa dengan hasil belajar siswa pada materi pencemaran lingkungan di SMA Negeri 1 Cibinong Cianjur Selatan dengan nilai $r = 0,61$. Hubungan antara kemampuan berpikir kritis dengan hasil belajar tergolong pada kategori cukup. Kemampuan berpikir kritis siswa terhadap hasil belajar berpengaruh sebesar

35%. Kemampuan berpikir kritis siswa termasuk kedalam faktor internal yang dapat memengaruhi hasil belajar siswa.

- (5) Penelitian ini dilakukan oleh **Alvendo Wahyu Aranski, Koko Handoko** Dalam judul **“Data Mining Dalam Pengelompokan Nilai IQ Siswa”** Pendidikan merupakan salah satu aspek kehidupan yang harus dimiliki oleh setiap manusia. Saat-saat yang paling penting untuk pendidikan adalah selama sekolah. Selama sekolah, siswa akan mulai dari taman kanak-kanak sampai akhirnya lulus di tingkat sekolah menengah. Salah satu aspek yang sangat diperhatikan selama sekolah adalah tingkat kecerdasan. Tingkat kecerdasan sering dikaitkan dengan tingkat IQ. Menurut penelitian, tingkat IQ seseorang menjadi tolak ukur tingkat kecerdasan orang tersebut. Tingkat kecerdasan dapat mempengaruhi kehidupan siswa dalam menerima pelajaran dengan baik sehingga sekolah harus mengklasifikasikan siswa dengan baik. Pengelompokan siswa berdasarkan tingkat IQ bertujuan agar siswa dapat menerima metode pembelajaran yang lebih tepat dan efektif sehingga siswa dapat menerima pelajaran dengan baik. Di SMA Yos Sudarso Batam banyak siswa yang mendaftar dan pengelompokan siswa akan memakan banyak waktu dan tidak efisien jika dilakukan secara manual. Pengelompokan siswa dapat dilakukan dengan cepat dan tepat dengan memanfaatkan *data mining*. Salah satu metode dalam *data mining* adalah clustering dengan menggunakan K-means. Tahapan dari metode ini dimulai dengan pemilihan acak K, K berikut adalah jumlah cluster yang ingin dibentuk. Kemudian atur nilai K secara acak, sedangkan nilai tersebut menjadi pusat cluster atau biasa disebut dengan *centroid, mean* atau *“means”*. Menghitung jarak setiap data yang ada pada setiap *centroid* menggunakan rumus *Euclidian* untuk mencari jarak terdekat dari setiap data dengan *centroid* setelah perhitungan manual dijalankan pada *software data mining Rapid Miner*. Penggunaan metode K-means untuk pengelompokan merupakan salah satu metode yang tepat jika dilihat dari variabel yang akan digunakan yaitu nilai tingkat IQ siswa. Metode K-means ini akan membentuk cluster yang mengelompokkan siswa berdasarkan tingkat IQ. Dengan menerapkan data mining clustering dengan metode K-means diharapkan dapat membantu pihak sekolah dalam mengklasifikasikan siswa secara tepat sehingga memudahkan pihak sekolah dalam memastikan siswa mendapatkan metode pembelajaran yang tepat.
- (6) Penelitian ini dilakukan oleh **Muhamad Muhamad, Agus Perdana Windarto, Suhada Suhada** Dengan judul **“ Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out”** Siswa merupakan salah satu substansi

yang perlu diperhatikan dalam kaitannya dengan dunia pendidikan saat ini. Sulitnya mendapatkan siswa membuat pihak sekolah harus mengoptimalkan sistem pembelajaran dan sarana prasarana serta untuk menarik minat siswa baru dan juga agar siswa yang sudah bersekolah tidak putus sekolah atau putus sekolah. Salah satu faktor penyebab banyaknya siswa putus sekolah adalah karena kurangnya kebijakan dan tindakan dari lembaga pendidikan untuk mencegah siswanya putus sekolah. Tujuan dari penelitian ini adalah untuk mengklasifikasikan siswa yang berpotensi putus sekolah dan tidak berpotensi putus sekolah dengan algoritma C4.5 sebagai acuan dalam membuat kebijakan dan tindakan untuk mengurangi jumlah siswa putus sekolah. Hasil klasifikasi algoritma C4.5 dievaluasi dan divalidasi dengan *Rapid Miner Studio* untuk mengetahui keakuratan Algoritma C4.5 dalam mengklasifikasikan potensi putus sekolah.

- (7) Penelitian ini dilakukan oleh **Achmad Shoddiq Bayu Asmoro, Wahyu Sakti Gunawan Irianto, Utomo Pujianto** Dengan judul “Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan” Sistem manajemen *E-learning* merupakan bentuk kemajuan teknologi dalam bidang pendidikan dan telah banyak menghasilkan kumpulan data-data pendidikan yang salah satunya adalah data aktivitas pembelajaran siswa dalam sistem manajemen *E-learning*. Banyaknya data pendidikan yang belum tereksplorasi dengan baik dapat di manfaatkan dengan menggunakan teknik data mining. Pada penelitian ini akan dilakukan perbandingan 3 model data berbeda yaitu data awal tanpa *preprocessing* dan data yang di *preprocessing* menggunakan seleksi fitur *correlation-based feature selection* dan *Information Gain*. Data yang digunakan adalah data aktivitas pembelajaran siswa dalam sistem manajemen *E-learning*. Selanjutnya proses pengujian data dengan menggunakan 10 *folds cross validation* dengan metode C4.5 dan evaluasi data menggunakan *confusion matrix*. Hasil dari pengujian data menggunakan algoritma C4.5 yang dikombinasikan dengan seleksi fitur *correlation-based feature selection* menghasilkan nilai akurasi yang lebih tinggi dengan nilai akurasi sebesar 76.92%. Sementara itu hasil dari pengujian data awal tanpa seleksi fitur dan data yang di seleksi fitur menggunakan *information gain* memiliki nilai akurasi yang sama dengan nilai akurasi sebesar 76.19%. Hal ini dikarenakan data yang diproses menggunakan algoritma C4.5 tanpa *preprocessing* dan data yang telah di *preprocessing* menggunakan *information gain* sama-sama menghitung nilai gain untuk membuat model pohon keputusan, dan

menghasilkan model pohon keputusan yang sama. Sehingga hasil dari proses pengujian data memiliki nilai akurasi yang sama.

- (8) Penelitian ini dilakukan oleh **Arisna, Wayan Veni, Dengan judul “Penerapan Data Mining Untuk Analisa Pola Nilai Siswa Menggunakan Algoritma C4.5”** Berdasarkan peraturan kurikulum 2013, proses penjurusan siswa baru pada tingkat SMA mulai dilakukan saat siswa duduk dibangku kelas X(sepuluh) dimana dalam prosesnya, siswa diberi kesempatan memilih jurusan sesuai minat, baik itu jurusan IPA atau IPS, kemudian akan di klasifikasikan lagi keputusan jurusannya oleh pihak sekolah dengan mempertimbangkan nilai SMP dan nilai hasil test penjurusan. Oleh karena itu siswa kelas IX(sembilan) yang saat ini sedang menempuh jenjang pendidikan SMP sudah pasti akan dihadapkan dengan masalah pemilihan jurusan yang akan diambil. SMP Negeri 09 Bandar Lampung merupakan salah satu lembaga pendidikan yang memiliki data nilai siswa melimpah yang berpotensi untuk menghasilkan pengetahuan baru jika diolah. Penggunaan *data mining* menjadi salah satu cara untuk mengolah data nilai siswa yang melimpah sehingga menghasilkan informasi yang tepat dan berkualitas. Penulis memilih teknik klasifikasi untuk dijadikan sebagai solusi karena, teknik ini merupakan teknik yang paling tepat untuk digunakan dalam menganalisa pola nilai siswa. Berdasarkan hasil visualisasi decision tree yang terbentuk dari pengolahan data nilai siswa, teknik klasifikasi dengan penerapan algoritma C4.5 menggunakan *Tools WEKA* memiliki kinerja yang cukup baik dalam menganalisa pola nilai siswa. Hal tersebut dapat dilihat dari hasil perhitungan akurasi yang mencapai 0,6319 atau sebesar 63,19%. Kata Kunci: *Data Mining*, Klasifikasi, C4.5, WEKA, Pola Nilai.
- (9) Penelitian ini dilakukan oleh **M A Abdillah, Arief Setyanto, Sudarmawan Sudarmawan Dengan judul “Implementasi Decision Tree Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter”** Pencanangan kurikulum Pendidikan Karakter dalam sistem pendidikan di Indonesia, adalah sesuatu hal yang baru dan belum banyak dikaji dalam penelitian *Educational Data Mining* (EDM). Sebagian besar penelitian dalam ranah EDM masih menggunakan faktor kognitif dalam penilaiannya, berbeda dengan pendidikan karakter yang lebih berorientasi kepada pengajaran nilai-nilai karakter, serta mempertimbangkan latar belakang peserta didik. Oleh karena itu, diperlukan cara atau metode untuk mengidentifikasi calon peserta didik, serta memprediksi kesuksesannya dalam sistem pendidikan karakter. Algoritma C4.5 dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap

calon siswa dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon siswa baru, dalam penelitian ini peneliti menggunakan data mahasiswa Unires Yogyakarta sebagai objek penelitian. Dengan penelitian ini, juga diharapkan dapat diketahui tingkat akurasi *Decision Tree* Algoritma C.45 dalam mengukur pengaruh atribut-atribut latar belakang siswa tersebut terhadap kesuksesan pendidikan karakter, sehingga akan diketahui apakah *Decision Tree* Algoritma C.45 memenuhi aspek reliabilitas dan validitas sebagai alat ukur kesuksesan pendidikan karakter. Dari hasil pengukuran, diketahui bahwa kombinasi atribut Bidang Bahasa dan Sosial, Latar belakang pendidikan agama dan orang tua yang menjadi seorang pendidik/guru, serta kemampuan untuk membaca Al-Qur'an berkorelasi positif terhadap kesuksesan pendidikan karakter. Nilai *accuracy* sebesar 60,91%, menunjukkan bahwa algoritma *decision tree* C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter. Namun masih butuh banyak kajian mendalam terutama dalam menentukan atribut-atribut yang memang benar-benar mempengaruhi kesuksesan pendidikan karakter sehingga diharapkan akan didapat *accuracy* yang lebih baik, serta terjadi efisiensi dalam pengelompokan atribut.

- (10) Penelitian ini dilakukan oleh **Hanii Mustofa, Ikbal Andrian Prasetyo, Saman Supriadi, Ramadhan Hendra Prasetyo, Muhammad Yasri L, Fitriyani** Dengan judul **“Metode Decision Tree Dalam Pemilihan Gaya Belajar Pada Siswa Sekolah Dasar”** Jenis Gaya belajar memiliki peranan penting dalam proses kegiatan belajar mengajar. Dengan mengenal gaya belajar, seseorang akan dapat menentukan cara belajar yang lebih efektif. Jenis Gaya belajar sendiri dibedakan kedalam tiga kelompok, yaitu Jenis Gaya Belajar Visual, Jenis Gaya Belajar Auditorial, dan Jenis Gaya Belajar Kinesthetic. Siswa pada tingkat sekolah dasar memerlukan perhatian dan pengawasan dalam proses pembelajaran baik di sekolah ataupun di rumah. Jika kebiasaan dan gaya belajar belajar yang efektif sudah dilakukan siswa sejak sekolah dasar maka akan mempengaruhi prestasi belajar yang memuaskan di masa mendatang. Penelitian ini menggunakan dataset *“Data Set Of Learning Style Preference”* dan menggunakan algoritma *Decision Tree*. Hasil Penelitian Ini untuk mengidentifikasi masing-masing gaya belajar setiap siswa dan juga dapat bermanfaat untuk guru, siswa, orang tua ataupun pihak sekolah yang berhubungan serta melaksanakan proses belajar dan pembelajaran baik di sekolah ataupun dirumah untuk memilih strategi pembelajaran yang tepat serta dapat dijadikan acuan untuk mengembangkan gaya belajar siswa sesuai

dengan jenis pembelajarannya dan tentunya ini mendukung terciptanya proses belajar yang efektif dan efisien.

Tabel 2. 5 Penelitian Rujukan

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
1.	Aprianto Tumanggor, Pend a Sudiarto Hasugian (2021)	Penerapan Data Mining Untuk Memprediksi Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4.5 Pada SDN 105351 Bakaran Batu	Jurnal Nasional Komputasi Vol 4 No. 1 2021	Kontribusi penelitian ini adalah persamaan atribut dari nilai Bahasa Indonesia, ipa, i ps dan matematika Pada tahap pengujian menggunakan <i>black box</i> Kelemahan penelitian ini adalah tidak terdapat nilai akurasi
2.	Oscario Oscario, Jasmir Jasmir, Yudi Novianto (2019)	Penerapan Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar	Entropy (Si), 2019	Kontribusi penelitian ini adalah menggunakan pengambilan metode Algoritma C4.5 Kelemahan penelitian ini adalah tidak terdapat nilai akurasi

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
3.	Suherman,Marlia Purnamasari,Fitri ani Dwi Hastuti (2021)	Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas Minat Menggunakan Metode <i>Decision Tree</i> C4.5	JSil (Jurnal Sistem Informasi Vol. 8 No. 2 2021)	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan confussion matrix dan pengambilan metode Algoritma C4.5 Kelemahan penelitian ini adalah tidak terdapat cara perhitungan algoritma C4.5
4.	Nisa, Sofiatun (2020)	Korelasi antara kemampuan berpikir kritis dengan hasil belajar siswa pada materi pencemaran lingkungan : Studi penelitian di kelas X SMA Negeri 1 Cibinong Cianjur Selatan	Jurnal Digital Library (2020)	Kontribusi penelitian ini adalah pengambilan metode Algoritma C4.5 Kelemahan penelitian ini yaitu tidak terdapat nilai uji coba,dan uji hasil menggunakan aplikasi

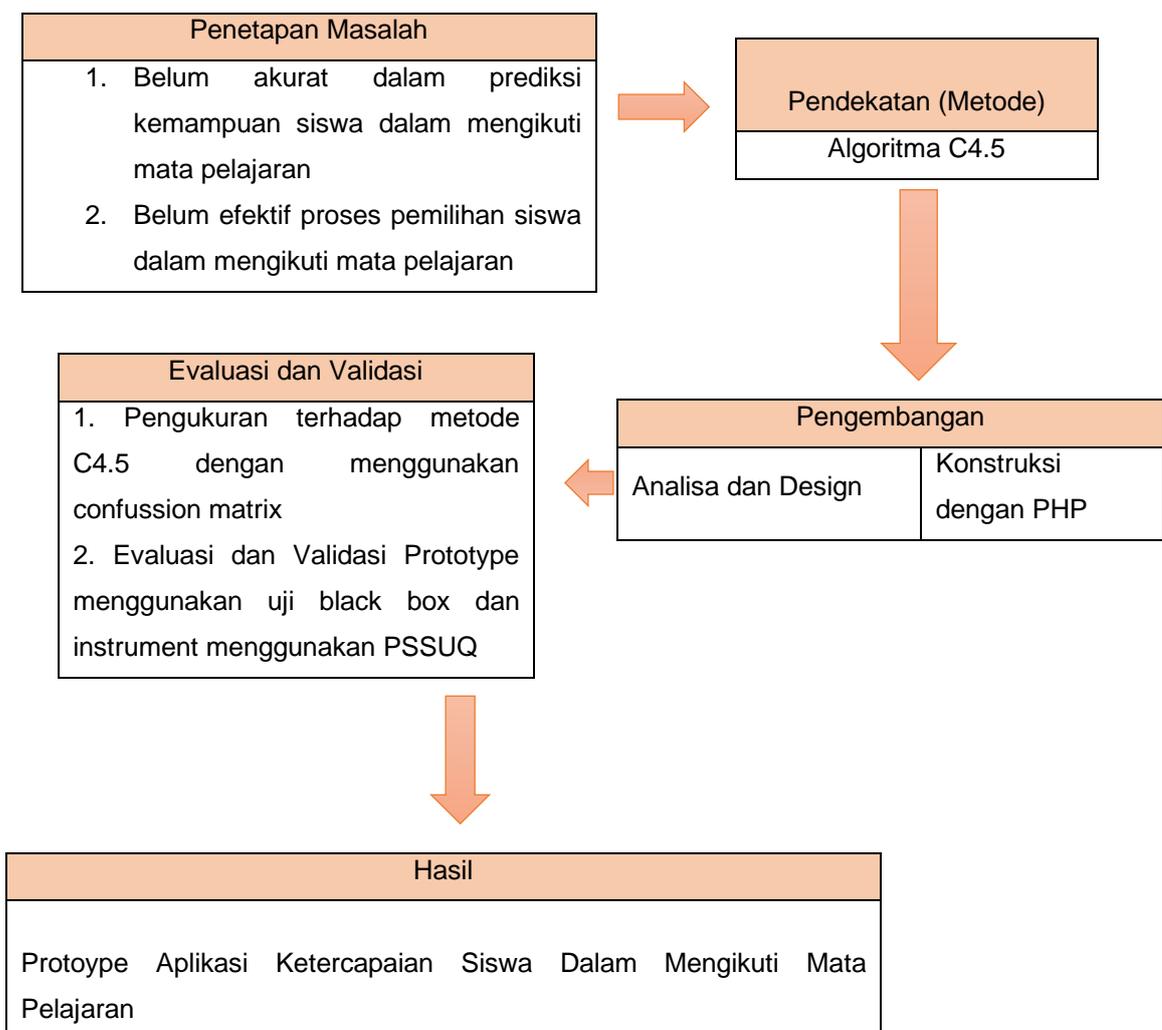
No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
5.	Alvendo Wahyu Aranski, Koko Handoko (2019)	Data Mining Dalam Pengelompokan Nilai IQ Siswa	Jurnal teknologi dan open source, 2019	Kontribusi penelitian ini adalah pengambilan metode menggunakan Algoritma C4.5 Kelemahan penelitian ini adalah belum adanya uji hasil menggunakan aplikasi
6.	Muhamad Muhamad, Agus Perdana Windarto, Suhada Suhada	Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out	KOMIK, Konferensi Nasional Teknologi Informasi dan Komputer, Vol. 3 No. 1 2019)	Kontribusi penelitian ini adalah pengambilan metode algoritma C4.5 dan cara perhitungan C4.5 Kelemahan penelitian ini adalah tidak terdapat nilai akurasi

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
7.	Achmad Shoddiq Bayu Asmoro, Wahyu Sakti Gunawan Irianto, Utomo Pujianto (2018)	Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan	Jurnal Edukasi dan Penelitian Informatika Vol. 4 No. 2 2018	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan confusion matrix dan pengambilan metode Algoritma C4.5 Kelemahan penelitian ini adalah tidak disampaikannya langkah langkah perhitungan C4.5
8	Arisna, Wayan Veni (2020)	Penerapan Data Mining Untuk Analisa Pola Nilai Siswa Menggunakan Algoritma C4.5	Jurnal Diss Teknorat Indonesia 2020	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan confusion matrix dan pengambilan metode menggunakan algoritma C4.5 Kelemahan penelitian ini adalah tidak disampaikannya langkah langkah perhitungan C4.5

No	Nama Peneliti	Judul Penelitian	Jurnal	Kontribusi
9.	M A Abdillah, Arief Setyanto, Sudarmawan Sudarmawan (2020)	Implementasi <i>Decision Tree</i> Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter	Jurnal Ilmiah Teknologi Informasi Vol. 15 No. 2 2020	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan <i>confussion</i> <i>matrix</i> dan pengambilan metode menggunakan algoritma C4.5 Kelemahan penelitian ini adalah belum adanya aplikasi untuk perhitungan
10	Hanii Mustofa, Ikbal Andrian Prasetyo, Saman Supriadi, Ramadhan Hendra Prasetyo, Muhammad Yasri L, Fitriyani (2021)	Metode <i>Decision Tree</i> Dalam Pemilihan Gaya Belajar Pada Siswa Sekolah Dasar	Jurnal Ilmiah Sains dan Teknologi Industri Vol. 5 No. 1 2021	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan <i>confussion</i> <i>matrix</i> dan pengambilan metode menggunakan algoritma C4.5

Berdasarkan uraian rujukan penelitian yang sudah dilakukan sebelumnya dapat disimpulkan terdapat kesamaan pada metode yang digunakan yaitu algoritma C4.5, Sedangkan perbedaannya terletak pada penggunaan variabel yang digunakan penelitian sebelumnya ada 4 variabel yaitu nilai Pendidikan Agama Islam (PAI), Bahasa Inggris, SBDP, PJOK sedangkan pada penelitian yang akan dilakukan variabel yang digunakan ada 7 variabel yaitu nilai Pendidikan Agama Islam (PAI), PPKN, Bahasa Indonesia, Matematika, Ilmu Pengetahuan Sosial (IPS), Ilmu Pengetahuan Alam (IPA), dan Bahasa Sunda. Selain itu juga pada penelitian sebelumnya belum diketahui tingkat akurasi dari metode algoritma C4.5 untuk prediksi ketercapaian kemampuan siswa dalam mengikuti mata pelajaran, hal tersebut akan dilakukan pada penelitian yang akan dilakukan.

D. Kerangka Pemikiran



Gambar 2. 7 Kerangka Pemikiran

Kerangka pemikiran pada gambar 2.7 dapat dijelaskan sebagai berikut :

1. Penetapan masalah mencakup fenomena yaitu belum akurat dalam menentukan siswa yang tercapai dalam mengikuti mata pelajaran
2. Pendekatan penelitian yaitu metode yang ditetapkan atau digunakan yaitu C4.5
3. Pengembangan yaitu tahap melakukan analisis dan desain kemudian menentukan bahasa pemrograman dengan konstruksi PHP
4. Evaluasi dan Validasi yaitu mengukur ketepatan hasil perhitungan metode C4.5 dengan menggunakan *confusion matrix*, pengukuran terhadap pengguna dengan menggunakan PSSUQ, dan pengukuran terhadap ahli dengan menggunakan *black box*
5. Hasil yaitu menampilkan prototype aplikasi ketercapaian siswa dalam mengikuti mata pelajaran

E. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu belum akurat dan belum efektif pada saat melakukan proses penentuan untuk prediksi hasil ketercapaian kemampuan siswa dalam mengikuti mata pelajaran di sekolah dasar, maka perlu adanya suatu cara untuk mengatasi permasalahan tersebut. Dalam teori data mining ada beberapa metode yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan dari pengalaman dimasa sebelumnya, diantaranya adalah metode algoritma C4.5. Metode Algoritma C4.5 akan menghasilkan hasil prediksi yang telah ditentukan nilai probabilitas akhir dari setiap kelas atau label dan melihat hasil akhir yang terbesar dari setiap kelas. Penelitian yang dilakukan oleh Aprianto Tumanggor, Penda Sudiarto Hasugian (2021) di dalam penelitiannya yang berjudul “PENERAPAN DATA MINING UNTUK MEMPREDIKSI TINGKAT KEMAMPUAN ANAK DALAM MENGIKUTI MATA PELAJARAN DENGAN METODE C4.5 PADA SDN 105351 BAKARAN BATU” menggunakan metode Algoritma C4.5. Maka dapat ditetapkan hipotesis pada penelitian ini penerapan Algoritma C4.5 diduga dapat memprediksi ketercapaian kemampuan siswa dalam mengikuti mata pelajaran di sekolah dasar.