

BAB 2 KERANGKA TEORITIS

A. Landasan Teori

1. Pengertian Data Mining

Data mining adalah kumpulan teknik untuk penemuan otomatis yang efisien yang sebelumnya tidak diketahui, valid, pola baru, berguna dan dapat dimengerti dalam database besar. Pola harus dapat ditindaklanjuti sehingga mereka dapat.

digunakan dalam pengambilan keputusan perusahaan. (Sumathi and Esakkirajan, 2007:17)

- (a) Data mining adalah proses penemuan otomatis dari pola yang sebelumnya tidak diketahui secara besar-besarnya volume data;
- (b) Volume data yang besar ini biasanya merupakan data historis dari sebuah organisasi yang dikenal sebagai data gudang;
- (c) Data mining berkaitan dengan volume data yang besar dalam kasus big data;
- (d) Data mining memungkinkan bisnis untuk menentukan pola historis untuk memprediksi perilaku masa depan;

Meskipun data mining dimungkinkan dengan jumlah data yang lebih kecil, semakin besar data semakin baik akurasi dalam prediksi.

2. Pengertian Klasifikasi

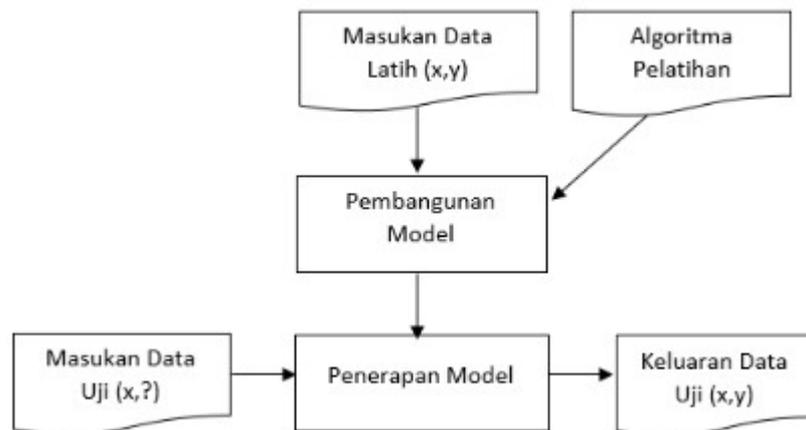
Klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi.

Teknik ini menggunakan supervised induction, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas-kelas tambahan (Kusnawi k, 2007). Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar

diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya (Eko Prasetyo, 2012).

Klasifikasi dapat didefinisikan secara detail sebagai waktu pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target / yang memetakan setiap vektor (set fitur) x ke depan satu dari sejumlah tabel kelas y yang tersedia pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori (Eko Prasetyo, 2013).

Model dalam dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukkan pada Gambar 2.1. Pada gambar tersebut disediakan sejumlah data latih (x,y) untuk digunakan sebagai data pembangun model. Model tersebut kemudian dipakai untuk memprediksi kelas dari data uji $(x,?)$ sehingga diketahui kelas y yang sesungguhnya.

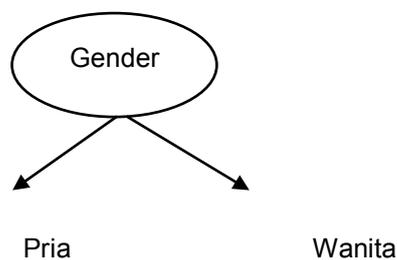


Gambar 2.1 Proses Pekerjaan Klasifikasi

Klasifikasi digunakan untuk pembuatan model yang dapat melakukan pemetaan dari setiap himpunan variabel kesetiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan variabel yang baru didapat. Dalam pembangunan model selama proses pelatihan tersebut diperlukan adanya suatu algoritma untuk membangunnya yang disebut dengan Algoritma pelatihan. Ada banyak algoritma pelatihan yang sudah dikembangkan oleh para peneliti yaitu C4.5, Nearest Neighbor, Bayesian Classification, Neural Network, dll.

3. Metode Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal, bisa mengatasi *missing* data, bisa mengatasi data kontinu dan *pruning* (Larose, 2005).



Gambar 2.2 Syarat pengujian fitur biner

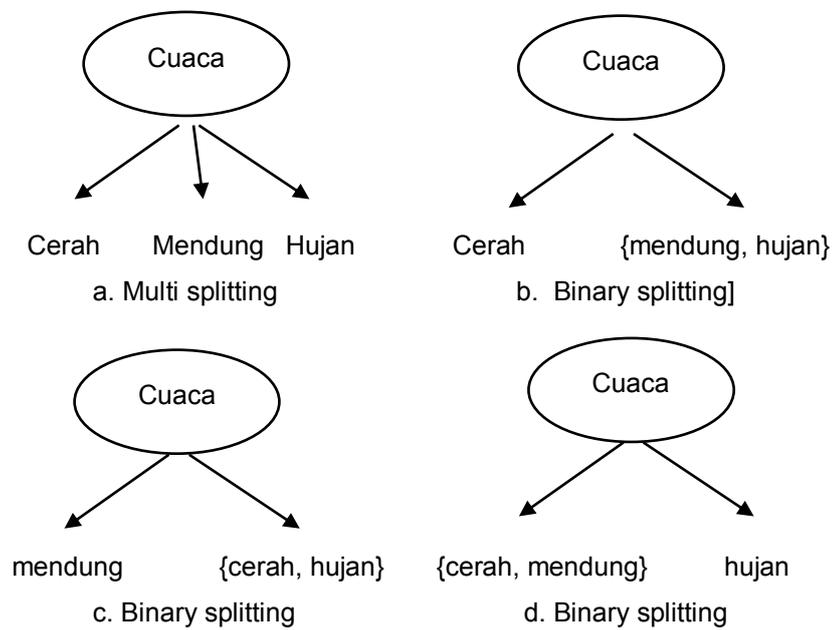
Yang menjadi hal penting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node:

(a) Fitur biner

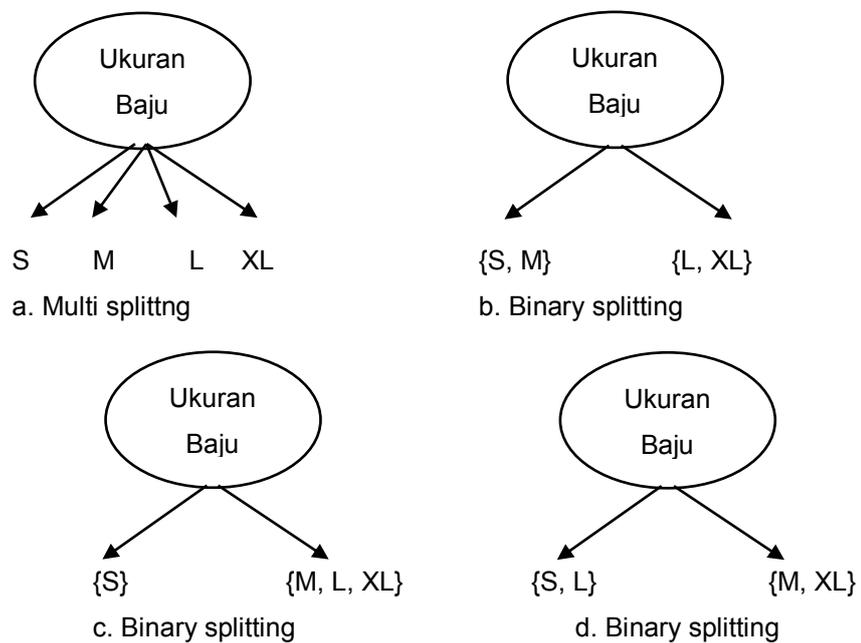
Fitur yang hanya mempunyai dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang. Contoh pemecahannya disajikan pada Gambar 2.3

(b) Fitur bertipe kategorikal

Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Contohnya adalah fitur 'cuaca' mempunyai 3 nilai berbeda dan ini bisa mempunyai banyak kombinasi syarat pengujian pemecahan. Secara umum ada 2, yaitu pemecahan biner (*binary splitting*) dan (*multi splitting*). Kombinasinya disajikan seperti pada Gambar. Untuk pemecahan yang hanya membolehkan pemecahan biner, seperti algoritma CART, maka akan memberikan kemungkinan jumlah kombinasi pemecahan sebanyak 2^{k-1} , dimana k adalah jumlah nilai berbeda dalam fitur tersebut. Contoh pemecahannya disajikan pada Gambar 2.4 dan Gambar 2.5.



Gambar 2.3 Syarat pengujian fitur bertipe nominal

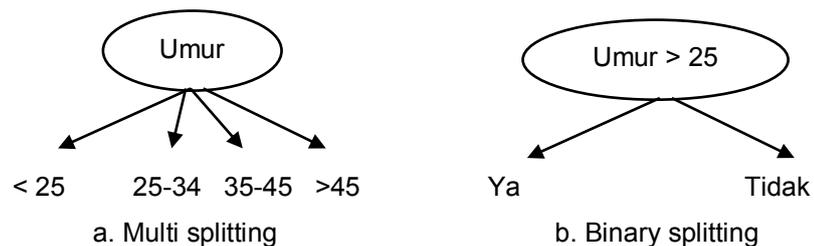


Gambar 2.4 Syarat pengujian fitur bertipe ordinal

(c) Fitur bertipe numerik

Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan pengujian perbandingan ($A < v$) atau ($A \geq$

v) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk $v_i \leq A < v_{(i+1)}$, untuk $i = 1, 2, \dots, k$. Untuk kasus pemecahan biner, maka algoritma akan memeriksa semua kemungkinan posisi pemecahan v dan memilih posisi v terbaik. Untuk cara multi, maka algoritma harus memeriksa semua kemungkinan jangkauan nilai kontinyu. Contoh pemecahan pada fitur numerik disajikan pada gambar 2.5.



Gambar 2.5. Syarat pengujian fitur bertipe numerik

Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C4.5 adalah rasio gain, yang diformulasikan oleh persamaan berikut:

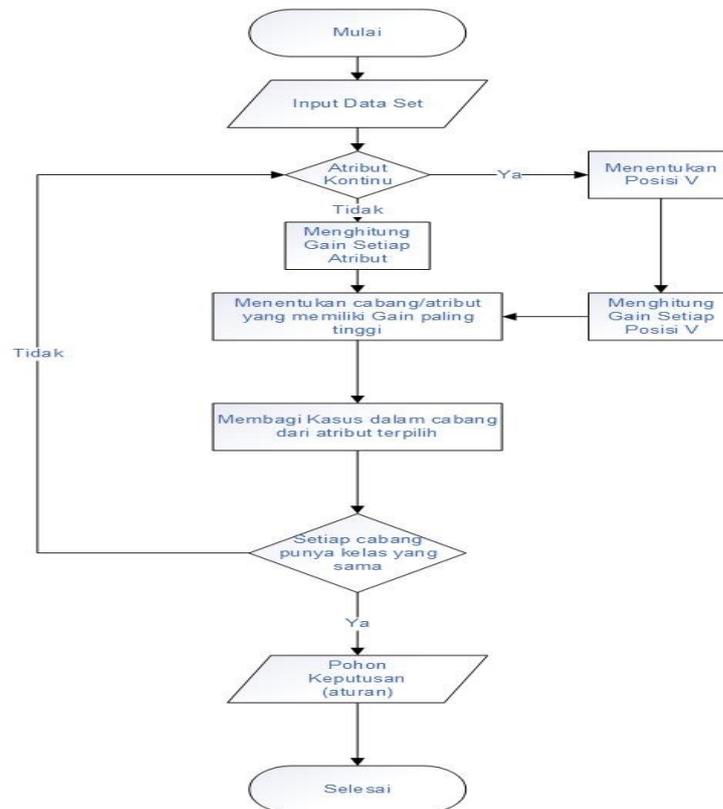
$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)}$$

Persamaan diatas menyatakan nilai rasio gain pada fitur ke- 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 .

Berikut ini adalah algoritma C4.5 menggunakan flowcart yang disajikan pada Gambar 2.6 dan berserta penjelasannya :

- (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 Atribut Kontinu atau bukan, Atribut Kontinu adalah atribut yang mempunyai jangkauan real.
- (3) Jika Iya maka akan menentukan posisi V , setelah itu menghitung gain setiap V . Jika tidak maka menghitung gain setiap atribut.
- (4) Setelah itu menentukan cabang/atribut yang memiliki gain paling tinggi.
- (5) Setelah itu membagi kasus dalam cabang dari atribut yang terpilih.

(6) Jika setiap cabang mempunyai kelas yang sama, maka menghasilkan pohon keputusan. Jika tidak maka cabang atau atribut tersebut kembali melakukan cek atribut kontinu.



Gambar 2.6 Flowchart Algoritma C4.5

Sumber : (Larose, 2005)

4. Pengembangan Sistem SDLC

Pendekatan sistem merupakan sebuah metodologi. Metodologi adalah satu cara yang direkomendasikan dalam melakukan sesuatu. Pendekatan sistem adalah metodologi dasar dalam memecahkan segala jenis masalah. Siklus hidup pengembangan sistem (System Development Life Cycle – SDLC) adalah aplikasi dari pendekatan sistem bagi pengembangan suatu sistem informasi (Raymon Mcleod , Jr, 2007)

Terdapat beberapa tahapan pekerjaan pengembangan yang perlu dilakukan jika suatu proyek ingin memiliki kemungkinan berhasil yang besar. Tahapan-tahapan tersebut adalah:

- (a) Analisis
- (b) Desain
- (c) Implementasi
- (d) Penggunaan

Proyek dan sumber daya yang dibutuhkan untuk melakukan pekerjaan direncanakan kemudian disatukan. Sistem yang ada juga dianalisis untuk memahami masalah dan menentukan persyaratan fungsional dari sistem yang baru. Sistem baru ini kemudian dirancang dan diimplementasikan. Setelah implementasi, sistem kemudian digunakan, idealnya untuk jangka waktu yang lama.

Karena pekerjaan-pekerjaan di atas mengikuti satu pola yang teratur dan dilaksanakan dengan cara dari atas ke bawah, SDLC tradisional sering kali disebut sebagai pendekatan air terjun (waterfall approach). Aktivitas ini memiliki aliran satu arah menuju ke penyelesaian proyek.



Gambar 2.7 Pola Melingkar dari Siklus Hidup Sistem

(Raymon Mcleod , Jr, 2007)

Gambar 2.7 mengilustrasikan sifat melingkar dari siklus hidup. Ketika sebuah sistem telah melampaui masa manfaatnya dan harus diganti, satu siklus hidup baru akan dimulai dengan diawali oleh tahap perencanaan.

Mudah bagi kita untuk melihat bagaimana SDLC tradisional dapat dikatakan sebagai suatu aplikasi dari pendekatan sistem. Masalah akan didefinisikan dalam tahap-tahapan perencanaan dan analisis. Solusi-solusi alternatif diidentifikasi dan dievaluasi dalam tahap desain. Lalu, solusi yang terbaik diimplementasikan dan digunakan. Selama tahap penggunaan, umpan balik dikumpulkan untuk melihat seberapa baik sistem mampu memecahkan masalah yang telah ditentukan.

5. PHP

Menurut (Kustyahningsih, 2011) PHP atau kependekan dari Hypertext Preprocessor adalah salah satu bahasa pemrograman open source yang sangat cocok atau dikhususkan untuk pengembangan web dan dapat ditanamkan pada sebuah skripsi HTML. Bahasa PHP dapat dikatakan menggambarkan beberapa bahasa pemrograman seperti C, Java, dan Perl serta mudah untuk dipelajari. PHP merupakan bahasa scripting server – side, dimana pemrosesan datanya dilakukan pada sisi server. Sederhananya, serverlah yang akan menerjemahkan skrip program, baru kemudian hasilnya akan dikirim kepada client yang melakukan permintaan. Adapun pengertian lain PHP adalah akronim dari Hypertext Preprocessor, yaitu suatu bahasa pemrograman berbasis kode – kode (script) yang digunakan untuk mengolah suatu data dan mengirimkannya kembali ke web browser menjadi kode HTML”. “PHP (atau resminya PHP: Hypertext Preprocessor) adalah skrip bersifat server – side yang ditambahkan ke dalam HTML”. Teoritis terkait dengan objek permasalahan

6. Business Process Model and Notation (BPMN)

Menurut (Pane, Lase, & Mali, 2020, hal. 196) menjelaskan bahwa Business Process Model and Notation (BPMN) merupakan sebuah representasi grafis dalam model proses bisnis untuk menentukan proses bisnis; tujuan BPMN untuk menyediakan pengguna teknis dan bisnis dengan markup intuitif kepada pelanggan bisnis manajemen proses bisnis, mengekspresikan kualitas dari proses yang kompleks; Dengan simbol – simbol sebagai berikut :

SIMBOL	NAMA	KETERANGAN
	; Start	; Merupakan awal dimulainya suatu aktivitas;
	; Intermediate	; Terjadi setelah proses sudah dimulai dan sebelum suatu proses berakhir;
	End	; Merupakan akhir dari segala aktivitas;



Pool

- Digunakan sebagai tempat grafis pada partisi satu set proses dari pool lain;



Line

- Digunakan untuk mengidentifikasi aktor yang terlibat;



Abstract

- Aktivitas yang dilakukan;



User Task

- Aktivitas yang dilakukan oleh user dengan menggunakan perangkat lunak;



Script Task

- Ketika task atau aktivitas dimulai maka mesin akan menjalankan script, begitu pula sebaliknya;



Manual Task

- Aktivitas yang dijalankan tanpa bantuan mesin atau aplikasi apapun;



Business Rule Task

- Aktivitas akan memungkinkan mengirimkan data ke dan menerima data dari business rule engine;



Service Task

- Aktivitas yang dijalankan akan otomatis oleh aplikasi;

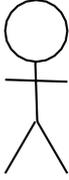
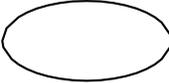
SIMBOL	NAMA	KETERANGAN
	Exclusive Gateway	Kondisi dimana hanya ada satu pilihan di dalamnya;
	Parallel Gateway	Kondisi dimana adanya beberapa pilihan yang dilalui;
	Inclusive Gateway	Kondisi dimana adanya satu kondisi atau lebih bisa dilalui;
	Sequence Flow	Penghubung untuk task berikutnya yang terdapat pada satu line;
	Message Flow	Digunakan untuk menyampaikan pesan dari dua pool;
	Association Flow	Digunakan untuk menghubungkan elemen dengan artifact.

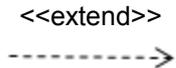
7. Unified Modeling Language (UML)

Menurut (Rosa and Shalahuddin, 2016, p. 155), Unified Modeling Language (UML) adalah bahasa pemodelan yang digunakan untuk mendefinisikan atau menggambarkan sistem perangkat lunak berdasarkan objek dalam sistem. UML tidak menentukan metode yang akan digunakan saat mengembangkan sistem, hanya notasi standar yang biasa digunakan untuk pemodelan objek.

Menurut (Rosa and Shalahuddin, 2016, p. 155), Use Case Diagram adalah pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Use Case mendeskripsikan sebuah interaksi antara satu atau lebih actor dengan sistem informasi yang akan dibuat. Berikut adalah simbol dari use case diagram.

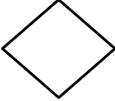
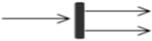
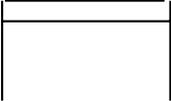
Simbol use case diagram

SIMBOL	NAMA	KETERANGAN
	<i>ACTOR</i>	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri, jadi walaupun simbol dari actor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama actor.
	<i>USE CASE</i>	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama use case.

	ASOSIASI / <i>ASSOCIATION</i>	Komunikasi antara actor dan use case yang partisipasi pada use case atau use case memiliki interaksi dengan actor.
	EKSTENSI / <i>EXTEND</i>	Relasi use case tambahan ke sebuah use case dimana use case yang ditambahkan dapat berdiri sendiri walau tanpa use case tambahan memiliki nama depan yang sama dengan use case yang ditambahkan.
	GENERALISASI / <i>GENERALIZATION</i>	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
	MENGGUNAKAN <i>INCLUDE</i>	Relasi use case tambahan ke sebuah use case untuk menjalankan fungsional

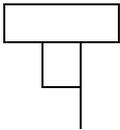
(Sumber: Dharwanto & Wahono, 2003)

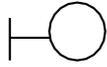
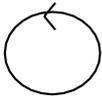
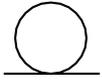
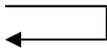
SIMBOL	NAMA	KETERANGAN
	STATUS <i>AWAL/INITIAL</i>	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	AKTIVITAS/ <i>ACTIVITY</i>	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengankata kerja.

	Simbol Activity Diagram <i>PERCABANGAN/ DECISION</i>	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	<i>PENGGABUNGAN/ JOIN</i>	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	<i>STATUS AKHIR/FINAL</i>	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status baru.
	<i>SWIMLINE</i>	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

Sumber: Dharwiyanti & Wahono, 2003)

Simbol Sequence Diagram

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>Lifeline</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
	<i>General</i>	Merepresentasikan entitas tunggal dalam sequence.

SIMBOL	NAMA	KETERANGAN
	<i>Boundary</i>	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
	<i>Control</i>	Elemen mengatur aliran dari informasi untuk sebuah skenario. Objek ini umumnya perilaku dan perilaku bisnis.
	<i>Entitas</i>	Elemen yang bertanggung jawab menyimpan atau informasi. Ini dapat berupa beans atau model object.
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan/hubungan antar objek yang menunjukkan urutan kejadian
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

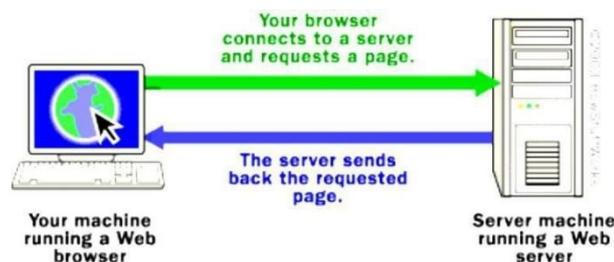
(Sumber: Dharwiyanti & Wahono, 2003)

8. Database

Menurut (Mustakini, 2009, hal. 49) database adalah kumpulan dari data yang saling berhubungan satu dengan yang lainnya, tersimpan di perangkat keras komputer dan digunakan perangkat lunak untuk memanipulasi.

9. Webserver (Apache)

Menurut (Kurniawan, 2008, hal. 2), menjelaskan bahwa *webserver* adalah sebuah perangkat lunak server yang berfungsi menerima permintaan HTTP atau HTTPS dari klien yang dikenal dengan web browser dan mengirimkan kembali hasilnya dalam bentuk halaman-halaman web yang umumnya berbentuk halaman HTML; web server yang terkenal diantaranya adalah apache dan microsoft internet information service (IIS); dengan ilustrasi cara kerja sebagai berikut :



- (1) Web server merupakan mesin dimana tempat aplikasi atau software beroperasi dalam mendistribusikan web page ke user, tentu saja sesuai dengan permintaan user;
- (2) Hubungan antara web server dan browser internet merupakan gabungan atau jaringan komputer yang ada di seluruh dunia. Setelah terhubung secara fisik, protocol TCP/IP (networking protocol) yang memungkinkan semua komputer dapat berkomunikasi satu dengan yang lainnya; pada saat browser meminta data web page ke server maka instruksi permintaan data oleh browser tersebut di kemas di dalam TCP yg merupakan protocol transport dan dikirim ke alamat yg dalam hal ini merupakan protocol berikutnya yaitu Hyper Text Transfer Protocol (HTTP).

B. 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. **(Dewi,2020) Analisa Algoritma C4.5 untuk prediksi penjualan obat pertanian di toko dewi sri** Pada Toko Dewi Sri ini belum dilakukan prediksi penjualan baik secara manual maupun menggunakan sistem. Hal ini mengakibatkan sering terjadi kesalahan dalam menentukan jumlah stok obat untuk periode berikutnya. Pestisida merupakan zat atau bahan kimia yang digunakan untuk membunuh hama pada tanaman yang dapat menghambat pertumbuhan tanaman sehingga berdampak pada hasil panen yang kurang baik. Toko Dewi Sri merupakan salah satu toko obat pertanian yang ada di Desa Beru Kecamatan Sarirejo yang menyediakan berbagai macam obat-obat pertanian yang dibutuhkan oleh para petani di Desa Beru Kecamatan Sarirejo. Adapun tujuan dari penelitian ini adalah untuk memprediksi penjualan obat pertanian. Dimana metode ini bekerja dengan membentuk suatu pohon keputusan. Berdasarkan hasil perhitungan yang telah dilakukan didapatkan hasil Akurasi sebesar 75%.
2. **(Esthyningtias,2019)Penerapan Algoritma C4.5 untuk menentukan persediaan obat studi kasus di RsBethesda yogyakarta** Dalam pelayanan Farmasi sebuah Rumah Sakit, persediaan obat adalah salah satu faktor penting. Jumlah persediaan atau stok harus dipertimbangkan, dengan tujuan untuk menjaga stok tersedia stabil. Stok tersedia berarti bahwa di warehouse tidak ada kelebihan stok atau kekurangan stok. Dalam penelitian ini yang dimaksud adalah warehouse Farmasi Rawat Jalan Lantai I Rumah Sakit Bethesda. Masalahnya adalah bahwa setiap item obat memiliki pengeluaran yang berbeda, sangat tergantung pada peresepan dokter. Maka perhitungan diperlukan untuk menentukan pola obat apa yang perlu disediakan (reorder), yang tidak perlu disediakan atau bahkan perlu dikurangi jumlahnya di warehouse. Algoritma C4.5 adalah sekelompok algoritma yang menggunakan pohon keputusan. Pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Algoritma ini digunakan untuk menganalisis obat yang perlu direorder agar ketersediaan obat tetap terjaga. Penerapan algoritma C4.5 dapat digunakan untuk memprediksi ketersediaan stok di Farmasi Rawat Jalan Bawah Lantai 1 Rumah Sakit Bethesda. Hasil penghitungan nilai entropy dalam algoritma C4.5 menggunakan Decision tree RapidMiner dapat

menghasilkan pola apakah stok di Farmasi Rawat Jalan Bawah Lantai 1 Rumah Sakit Bethesda perlu ditambah, cukup atau perlu dikurangi.

3. **(Anestiviya,2021) Analisis pola menggunakan metode C4.5 untuk peminatan jurusan siswa berdasarkan kurikulum** Peminatan jurusan merupakan suatu keputusan yang dilakukan oleh peserta didik baru untuk memilih jurusan “IPA” dan “IPS” yang ada pada siswa. Pemilihan peminatan dilakukan atas dasar minat calon peserta didik. Ketepatan dalam menentukan peminatan dapat menentukan keberhasilan belajar siswa. Sebaliknya, kesempatan yang sangat baik bagi siswa akan hilang karena kurang tepatan dalam penentuan peminatan. Kurikulum yang digunakan pada sekolah SMA Negeri 1 Natar saat ini adalah kurikulum 2013. Maka dari itu pihak sekolah berharap dengan adanya penjurusan yang tepat sejak awal akan memberikan akhir proses pembelajaran yang berhasil sehingga pihak sekolah membutuhkan variabel dasar penentuan keberhasilan proses kegiatan belajar mengajar (KBM) yaitu nilai Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, serta minat dari calon siswa. Penggunaan data mining dengan teknik klasifikasi melalui metode C4.5 diharapkan dapat memberikan analisis pola peminatan jurusan siswa yang sesuai sehingga proses KBM dapat berjalan secara lancar dan berhasil. Dari hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa.
4. **(Purnamasari,2021)Klasifikasi siswa berdasarkan mata pelajaran lintas minat menggunakan metode C4.5** Lintas minat adalah program untuk memperluas dan mengembangkan minat, bakat dan kemampuan peserta didik yang mereka miliki dengan memilih kelompok mata pelajaran. Kurikulum 2013 dirancang untuk memberikan kesempatan kepada siswa belajar berdasarkan minat siswa. Selain memilih mata pelajaran dalam suatu peminatan tertentu, siswa juga diberi kesempatan untuk mengambil mata pelajaran lintas minat. SMA Negeri 1 Anyer salah satu sekolah yang telah menerapkan program lintas minat. Dalam proses penentuan kelas lintas minat disekolah tersebut masih mengalami kendala yaitu tidak terspesifikasinya siswa yang memiliki minat pada mata pelajaran tertentu, dan pada proses pemilihan lintas minat ditentukan oleh pihak sekolah. Penelitian ini bertujuan untuk mengklasifikasi siswa berdasarkan minat dan bakat siswa pada mata pelajaran tertentu. Metode yang digunakan yaitu Decision Tree dan algoritma C4.5. Pada

penelitian ini didapat nilai akurasi sebesar 82,82%. Penelitian menghasilkan sebuah sistem penentuan kelas lintas Minat. Model klasifikasi ini dapat membantu siswa dalam menentukan lintas minat dan dapat digunakan sebagai alternatif referensi bagi guru BK untuk dapat mengelompokkan siswa berdasarkan minat dan bakat siswa.

5. **(Zakir,2020)Penerapan data mining untuk klasifikasi data penjualan makanan terlaris dengan Algoritma C4.5** Pada penelitian ini diambil permasalahan terhadap data penjualan pada Rumah burger yang mengalami kesulitan mendapatkan informasi strategis seperti tingkat penjualan per periode atau makanan terlaris. Berdasarkan permasalahan dibuatlah analisa data mining pada data penjualan di rumah burger. penerapan data mining menggunakan metode Algoritma C4.5 yang dapat membantu user untuk analisis dan memprediksi produk makanan terlaris atau tidak terlaris dan ada perekomendasi ialah dengan memastikan tingkatan confidence tertinggi yang dicapai setiap variabel pada semua produk, pemenuhan dan prioritas variabel untuk menganalisis dan memprediksi.
6. **(Gunawan,2021)Klasifikasi rasa susu almond berdasarkan minat pelanggan menggunakan Algoritma C4.5 di home made mami nia ice pematang siantar** Minat adalah timbulnya perasaan suka terhadap suatu objek yang ditandai dengan adanya rasa senang atau tertarik terhadap sesuatu tanpa ada paksaan. Susu almond adalah minuman susu nabati yang dibuat dari kacang almond. Sari kacang almond memiliki warna yang lebih pucat dan tekstur yang lebih kental dibandingkan susu biasanya. Adapun manfaat susu almond dalam dunia kesehatan yaitu untuk meningkatkan produksi ASI, mencegah tekanan darah tinggi, memperkuat kekebalan tubuh, melindungi kesehatan tulang, menjaga kesehatan jantung bayi, penangkal radikal bebas, memperlancar pencernaan, membuat kulit cerah setelah melahirkan. Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD). Oleh sebab itu penulis memberikan solusi klasifikasi dalam Data Mining dilakukan dengan menggunakan Algoritma C4.5. Dengan Algoritma C4.5 akan didapatkan sebuah Pohon Keputusan yang mudah dipahami dan mudah dimengerti. Dengan demikian dapat membantu pemilik menentukan perencanaan produksi susu almond yang paling banyak diminati serta tanpa khawatir terjadi kelebihan barang atau kekurangan bahan.
7. **(Renhorhan,2018)Penerapan Algoritma C4.5 untuk menentukan stok dan target permintaan material yang paling dibutuhkan gudang logistik pada Pt PLN (persero) area kebun jeruk** mengklasifikasikan data stok dan target

permintaan material yang user/ pengawas pekerjaan yang nantinya akan dibuatkan klasifikasi data set dengan jumlah kasus yang paling dibutuhkan. Pada proses penelitian yang sudah dilakukan dari tahap ini yaitu perlu adanya peringkasan data atau proses pengubahan data mentah menjadi data yang mudah dikelola. Data mentah akan dimasukkan pada kategorinya masing-masing kemudian data juga dipilih, sehingga data tersebut sesuai untuk proses data mining yang telah dilakukan pada Penerapan Klasifikasi Algoritma C4.5 dengan pengolahan datanya menggunakan metode Decision Tree dan mempermudah jalannya di sistem aplikasi RapidMiner. metode Decision Tree berguna sekali dalam penelitian ini yaitu untuk menentukan data stok material/barang dari target permintaan material gudang yang paling sering diminta dan dibutuhkan oleh user. Sehingga dapat disimpulkan bahwa permasalahan dalam menentukan data stok material/barang bisa diselesaikan menggunakan teknik data mining, yaitu dengan Algoritma C4.5 dan mendapatkan tingkat akurasi yang dihasilkan oleh sistem dengan metode decision tree pada aplikasi RapidMiner adalah 100%. sehingga dapat disimpulkan bahwa penelitian ini dapat membantu user/pengawas gudang dalam memilih dan menyiapkan stok material yang dibutuhkan sesuai standar dengan sigap dan cukup memadai.

8. **(Pritalia,2018)Penerapan Algoritma C4.5 untuk penentuan ketersediaan barang E-commerce** Dalam sebuah perusahaan e-commerce, stok barang merupakan salah satu faktor penting yang harus diperhatikan oleh perusahaan. Jumlah barang masuk maupun keluar perlu diperhatikan, tujuannya untuk menjaga stok yang tersedia dalam gudang menjadi stabil. Stok stabil yang dimaksud dalam gudang adalah barang tidak kelebihan stok maupun barang tidak kehabisan stok. Permasalahan yang dihadapi pada penjualan barang adalah setiap barang memiliki minat beli yang berbeda-beda. Barang dengan minat beli yang tinggi akan cepat habis dibandingkan dengan barang yang memiliki minat beli rendah. Untuk itu, stok barang perlu ditambah pada waktu stok barang menipis. Algoritma C4.5 merupakan kelompok algoritma dengan menggunakan pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Semakin kaya informasi atau pengetahuan yang dikandung oleh data training, maka akurasi akan semakin meningkat. Algoritma ini digunakan untuk menganalisis waktu pembelian stok barang yang sudah menipis dengan mengklasifikasi barang mana yang sudah waktunya di tambah stok maupun belum, sehingga ketersediaan barang tetap stabil dan terjaga. Hasil dari analisis menggunakan

algoritma C4.5 adalah untuk menentukan waktu penentuan ketersediaan barang memiliki tingkat keakuratan sebesar 98.9%.

9. **(Hutagaol,2021)Penerapan Algoritma C4.5 untuk menentukan kualitas telur ayam australia terbaik**

Telur Ayam Australia merupakan salah satu bahan makanan hewani yang dikonsumsi selain daging, ikan dan susu. Telur ayam Australia mengandung gizi yang tinggi, ketersediaan yang berkelanjutan, dan harga yang relatif lebih dibandingkan dengan telur lainnya sehingga menjadikan telur ayam Australia sangat diminati oleh para konsumen. Namun, telur ayam Australia mudah mengalami kerusakan dan penurunan kualitas akibat masuknya bakteri ke dalam telur. CV. Bintang Ternak pemilihan telur yang berkualitas masih dilakukan secara manual oleh karyawan, sehingga perlu adanya metode terbaru dalam penentuan telur yang berkualitas. CV. Bintang Ternak memiliki sampel telur ayam Australia 90 butir. Sampel yang digunakan yaitu kulit telur ayam Australia, dalam menentukan kualitas telur ayam Australia dapat dilihat dari kondisi kulit telurnya. Perencanaan dalam penyelesaian permasalahan tersebut akan menggunakan metode algoritma C4.5. Dengan adanya penerapan metode algoritma C4.5 yang terdapat pada aplikasi Rapidminer5 maka ada beberapa keuntungan yang didapat diperoleh yaitu dalam mengambil suatu keputusan lebih sederhana, mempersingkat pekerja karyawan, lebih fleksibel dalam pemilihan atribut kelas, keakuratan dalam memprediksi lebih efektif dan efisien dan mudah dipahami dengan adanya kesimpulan yang dihasilkan oleh aplikasi Rapidminer5 yang dapat dilihat dari pohon keputusan atau rule yang diperoleh dengan menggunakan 90 sampel data uji menghasilkan akurasi sebesar 90%.

10. **(Fikri,2020) Penerapan data mining untuk prediksi penjualan alat medis menggunakan Algoritma C4.5 Pt Murni Indah Sentosa**

Alat Medis merupakan perangkat apapun atau peralatan yang bertujuan untuk keperluan para dokter-dokter yang bertugas di rumah sakit maupun klinik. PT. Murni Indah Sentosa perusahaan yang bergerak di bidang penjualan alat medis. Dalam bisnisnya, perusahaan ini sering kali dihadapkan dengan permasalahan menentukan persediaan stok alat medis dan tidak dapat mengatasi prediksi penjualan pada periode berikutnya. Sehingga dengan cara seperti ini perusahaan sering kali mengalami kesulitan dalam menentukan persediaan stok alat medis, jadi aplikasi pada sistem yang lama sangat sulit untuk membantu manager dalam menentukan prediksi penjualan untuk periode berikutnya. Tidak jarang stok alat medis yang diminta merupakan stok yang peminatnya kurang dan terjadi penumpukan stok di gudang. Untuk

menyelesaikan masalah tersebut dibutuhkan suatu teknologi sistem informasi yang akan membantu manajemen dalam mengelola perusahaan agar setiap keputusan yang diambil ditunjang dengan informasi aktual dan dapat dipercaya. Salah satu teknik yang ada pada data mining adalah klasifikasi dengan algoritma C4.5. Pada penelitian ini akan dibahas teknik klasifikasi yang diterapkan untuk menganalisa data penjualan terutama hal yang berkaitan dengan transaksi penjualan alat medis. Kesimpulan yang diperoleh yaitu dengan menggunakan data mining algoritma C4.5 dapat membantu perusahaan dalam menentukan prediksi penjualan alat medis. Sehingga aplikasi data mining yang dibangun dapat langsung digunakan oleh perusahaan dan dapat dirasakan manfaatnya.

Tabel 2.1 Tinjauan Studi

N O	TAHUN	JUDUL	SUMBER	KONTRIBUSI
1.	(,Kiki Rosita Dewi,2020)	Analisa Algoritma C4.5 Untuk Prediksi Penjualan Obat Pertanian Di Toko Dewi Sri	<u>Analisa Algoritma C4.5 Untuk Prediksi Penjualan Obat Pertanian Di Toko Dewi Sri - Search (bing.com)</u>	Kontribusi terhadap penelitian ini adalah Pengambilan variabel jenis obat untuk dijadikan sebagai variabel penelitian.
2.	(Gunartatik Esthiningtyas, 2019)	Penerapan Algoritma C4.5 Untuk Menentukan Persediaan Obat Studi Kasus: Di Rs	<u>https://www.bing.com/ck/a?!&&p=621929e99668930ec068e22cbc30adfbe78cf06a7bcd2d264987360a</u>	Kontribusi terhadap penelitian ini adalah untuk menjaga stok tersedia stabil. tidak ada kelebihan stok atau kekurangan stok.

NO	TAHUN	JUDUL	SUMBER	KONTRIBUSI
		Bethesda Yogyakarta		
3	(Vista Anestiviya,2021)	Anlisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum Studi Kasus : SMAN 1 NATAR	https://www.bing.com/ck/a?!&&p=f8f4f91d33ad5421c75019d1e79fb72897203e6f4f8d1c22a7148180cf3e69JmltdHM9MTY1Nzc	Kontribusi terhadap penelitian ini adalah untuk penggunaan perhitungan menggunakan algoritma c4.5 untuk dijadikan dalam metode dalam penelitian.
4	(Suherman,2021)	Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas minat Menggunakan Metode Decision Tree C4.5	<u>Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas Minat Menggunakan Metode Decision Tree C4.5 Marlia Purnamasari - Academia.edu</u>	Kontribusi terhadap penelitian ini adalah untuk Penggunaan perhitungan menggunakan algoritma c4.5 untuk dijadikan dalam metode dalam penelitian.

NO	TAHUN	JUDUL	SUMBER	KONTRIBUSI
5	(Ahmad Zakir,2020)	Penerapan Data Mining Untuk Klasifikasi Data Pejualan Makanan Terlaris Dengan Algoritma C4.5	https://www.bing.com/ck/a?!&&p=aeb130e2ed6320efc852a1ec12647295cd9dc2c66205e5b6ae230b8e	Kontribusi terhadap penelitian ini adalah Pengambilan variabel data penjualan untuk dijadikan sebagai variabel penelitian.
6	(Indra Gunawan,2021)	Klasifikasi Rasa Susu Almond Berdasarkan Minat Pelanggan Menggunakan Algoritma C4.5 Di Home Made Mami Nia Ice Pematang Siantar	Klasifikasi Rasa Susu Almond Berdasarkan Minat Pelanggan Menggunakan Algoritma C4.5 di Home Made MamiNia Ice Pematangsiantar Request PDF (researchgate.net)	Kontribusi terhadap penelitian ini adalah mengenai hasil rule yang didapat dari penelitian tersebut
7	(Bahrun Said Benhoran,2018)	Penerapan Algoritma C4.5	<u>PENERAPAN KLASIFIKASI ALGORITMA C4.5 UNTUK MENENTUKAN</u>	Kontribusi terhadap penelitian ini adalah Pengambilan output ditambah dan dikurang

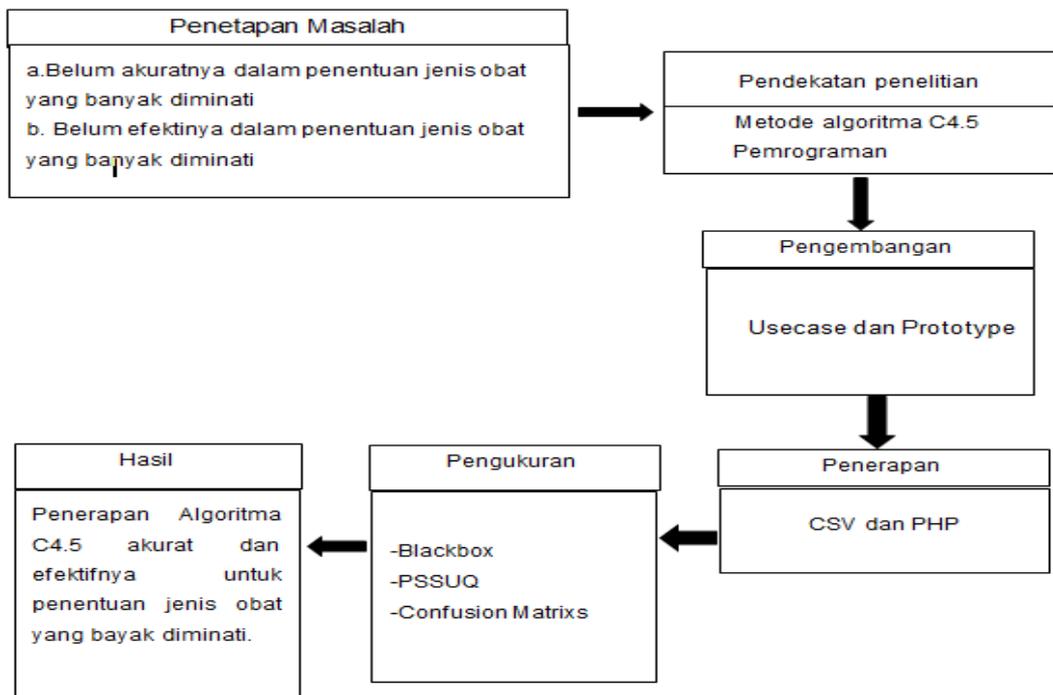
NO	TAHUN	JUDUL	SUMBER	KONTRIBUSI
		Untuk Menentukan Data Stok Dan Target Permintaan Material yang Paling Dibutuhkan Gudang Logistik Pada PT PLN (Persero) Area Kebun Jeruk	<u>DATA STOK DAN TARGET PERMINTAAN MATERIAL YANG PALING DIBUTUHKAN GUDANG LOGISTIK PADA PT PLN (Persero) AREA KEBON JERUK (nusamandiri.ac.id)</u>	untuk dijadikan sebagai output penelitian
8	(GL Pritalia ,2018)	Penerapan Algoritma C4.5 Untuk Penentuan Ketersediaan Barang E-commerce	<u>(PDF) Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce (researchgate.net)</u>	Kontribusi terhadap penelitian adalah ma ini digunakan untuk menganalisis waktu pembelian stok barang yang sudah menipis dengan mengklasifikasi barang mana yang sudah waktunya di tambah stok maupun belum
9	(Prodi Sistem	Penerap	https://www.bing.com/sea	Kontribusi terhadap

NO	TAHUN	JUDUL	SUMBER	KONTRIBUSI
	Informasi STMIK Royal,2021)	an Metode Algoritma C4.5 Untuk Menentukan Kualitas Telur Ayam Australia Terbaik	rch?q=PENERAPAN+METODE+ALGORITMA+C4.5+UNTUK+MENENTUKAN+KUALITAS+TELUR+AYAM+AUSTRALIA+TERBAIK&cvid=2475eddabe534439b3949db6edbf046&aqs=edge..69i57j69i6	penelitian ini adalah dihasilkan oleh aplikasi rapidminer5 yang dapat dilihat dari pohon keputusan atau rule yang diperoleh dengan menggunakan 90 sampel data uji menghasilkan akurasi sebesar 90%.
10	(Wiwi Veriana,2020)	Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT Murni Indah Sentosa	<u>Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT Murni Indah Sentosa - Search (bing.com)</u>	Kontribusi terhadap penelitian ini adalah dengan menggunakan data mining algoritma C4.5 dapat membantu perusahaan dalam menentukan prediksi penjualan alat medis

Metode dan hasil dari 10 jurnal tersebut menggunakan metode algoritma C 4.5 yang dapat digunakan untuk klasifikasi. Sedangkan untuk persamaan dengan penelitian ini yaitu menggunakan metode algoritma C4.5. Dan untuk perbedaannya dengan penelitian ini yaitu metode algoritma C4.5 dapat digunakan untuk Penentuan Jenis Obat Bebas diapotik.

C. Kerangka Pemikiran

Kerangka pemikiran ini dibuat mewakili konsep pemecahan masalah penelitian yang meliputi masalah yang akan di teliti, metode penelitian, dan hasil dari penelitian, berikut dijelaskan berdasarkan gambar 2.8



Gambar 2.8 Kerangka Pemikiran

Keterangan :

1. Penetapan masalah yang ada adalah belum akuratnya dalam penentuan jenis obat bebas diapotik dan belum efektifnya dalam penentuan jenis obat bebas diapotik.
2. Pendekatan penelitian ini menggunakan metode algoritma c4.5 dan Pemrograman sebagai dasar penelitian
3. Aplikasi dan komponen didalamnya yaitu ada usecase dan prototype aplikasi
4. Di dalam penerapan menggunakan CSV dan PHP
5. kemudian dilakukan uji hasil yaitu menguji ketepatan hasil perhitungan algoritma C4.5 apakah sudah sesuai atau belum dengan menggunakan confusion matrix, pengujian sistem kepada ahli sistem menggunakan blackbox, dan uji kebergunaan kepada pengguna menggunakan PSSUQ
6. Lalu hasil dari Penerapan Algoritma C4.5 akurat dan efektifnya untuk penentuan jenis obat yang bayak diminati.

D. Hipotesis

Berdasarkan permasalahan yang dihadapi yaitu belum akurat dan efektifnya pada saat melakukan proses prediksi untuk penentuan jenis obat yang banyak diminati. maka perlu adanya suatu cara untuk mengatasi permasalahan tersebut. Dalam teori Algoritma C4.5 ada beberapa metode yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan pengalaman di masa lalu, diantaranya metode Algoritma C4.5 yang akan menghasilkan hasil prediksi yang telah ditentukan dari data penjualan perbulan nantinya hasil prediksi tersebut bisa digunakan untuk memprediksi jenis obat bebas diapotik. Berdasarkan dari pengungkapan penulis, maka dapat ditetapkan hipotesisnya yaitu metode Algoritma C4.5 diduga dapat memprediksi untuk penentuan jenis obat bebas diapotik.