

## **BAB II**

### **KERANGKA TEORITIS**

#### **A. Landasan Teori**

##### **1. Pengertian Data Mining**

Mengutip (Tan, 2006 dalam Eko Prasetyo, 2013:2 ) data mining yaitu sebagai bentuk proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan sebuah informasi baru yang diambil dari potongan data besar yang membantu dalam pengambilan keputusan. Istilah data mining disebut juga knowledge discovery. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam hal ini data mining merupakan satu langkah dari proses KDD (Turban, dkk 2001). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola pada sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut (Tan, 2006) :

- (a) Pembersihan data yaitu untuk membuang data yang tidak konsisten dan noise;
- (b) Integrasi data yaitu penggabungan data dari beberapa sumber;
- (c) Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk di mining;
- (d) Aplikasi teknik data mining yaitu proses ekstraksi pola dari data yang ada;
- (e) Evaluasi pola yang ditemukan yaitu proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan;
- (f) Presentasi pengetahuan yaitu dengan teknik visualisasi
- (g) Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Langkah terakhir KDD adalah mempresentasikan pengetahuan dalam

##### **2. Pengertian Prediksi**

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. (Elsa Paskalis Krisda, 2019)

Analisis regresi dapat digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen atau prediktor dan sebuah variabel dependen atau respon (yang bernilai kontinyu). Dalam konteks data mining, variabel prediktor adalah atribut yang berkepentingan mendeskripsikan baris data (yakni yang menyusun vektor atribut). Secara umum, nilai dari variabel prediktor adalah diketahui. Variabel respon adalah apa yang akan diprediksikan. Ketika diberikan sebuah baris data yang dideskripsikan oleh variabel prediktor, maka akan diprediksikan nilai yang bersesuaian dengan variabel respon.

Analisis regresi adalah pilihan yang bagus manakala semua nilai dari variabel prediktor bernilai kontinyu juga. Banyak permasalahan dapat dipecahkan oleh Regresi Linier, dan akan lebih banyak lagi dapat dikerjakan dengan mengaplikasikan transformasi kepada variabel-variabel yang bersesuaian sehingga masalah non-linier dapat diubah menjadi masalah linier.

### 3. 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 (Raymond McLeod 2007: 199).

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

- (a) Perencanaan
- (b) Analisis
- (c) Desain
- (d) Implementasi
- (e) 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. 1 Pola Melingkar dari Siklus Hidup**

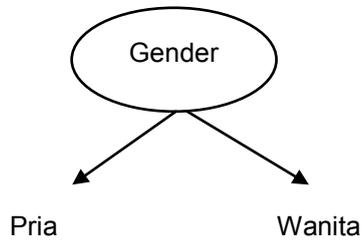
(Sumber: Raymond McLeod 2007: 199)

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.

#### 4. 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

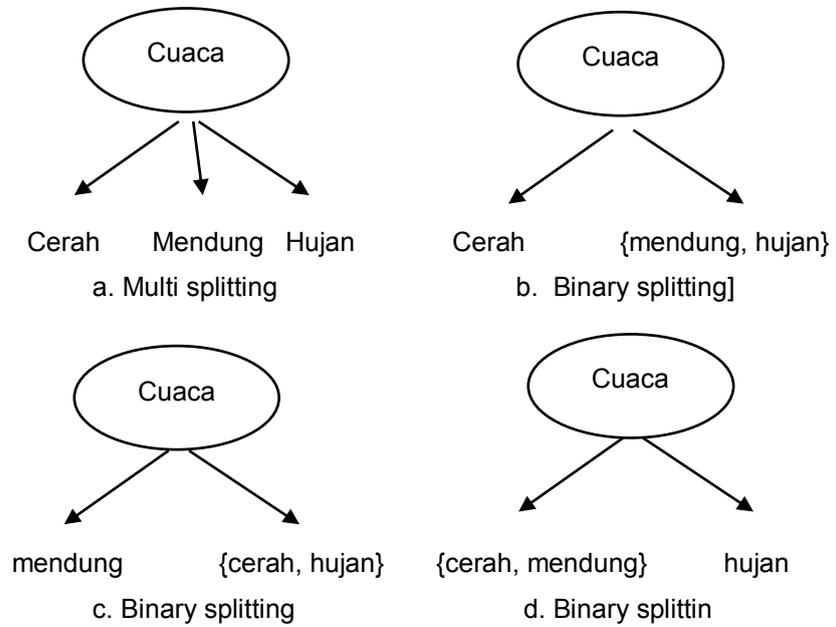
Fitur yang hanya mempunyai dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang. Contoh pemecahannya disajikan pada Gambar 2.2

(b) Fitur bertipe nominal

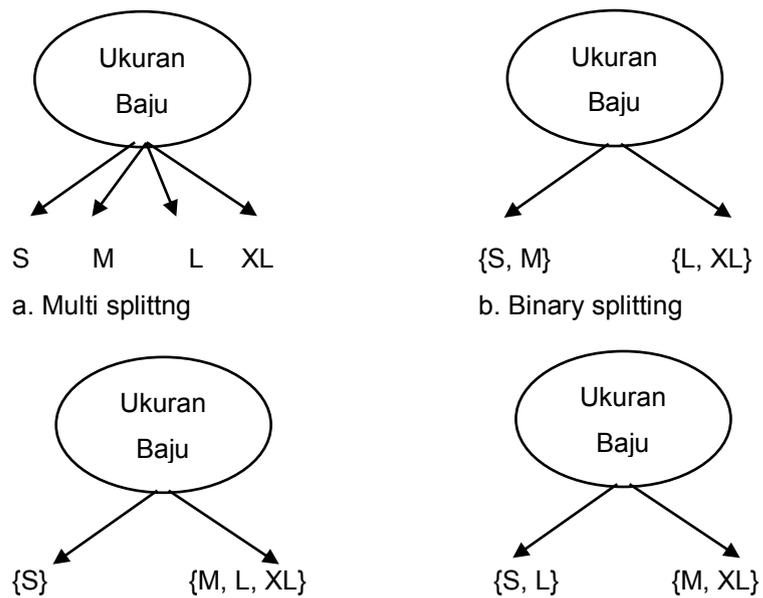
Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Contohnya adalah fitur 'cuaca' mempunyai 3 nilai berbeda dan ini bisa mempunyai banyak kombinasi syarat pengujian pemecahan. Secara umum ada 2, yaitu pemecahan biner (*binary splitting*) dan (*multi splitting*). Kombinasinya disajikan seperti pada Gambar 2.1. 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 2.1 dan Gambar 2.2

(c) Fitur bertipe numerik

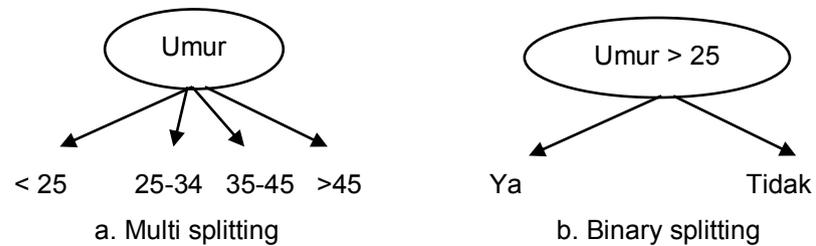
Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan pengujian perbandingan ( $A < v$ ) atau ( $A \geq v$ ) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk  $v_i \leq A < v_{i+1}$ , untuk  $i = 1, 2, \dots, k$ . Untuk 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. 3 Pengujian Bertipe Nominal**



**Gambar 2. 4 Syarat Pengujian Fitur Biner**



**Gambar 2. 5 Syarat Pengujian Fitur Bertipe Numerik**

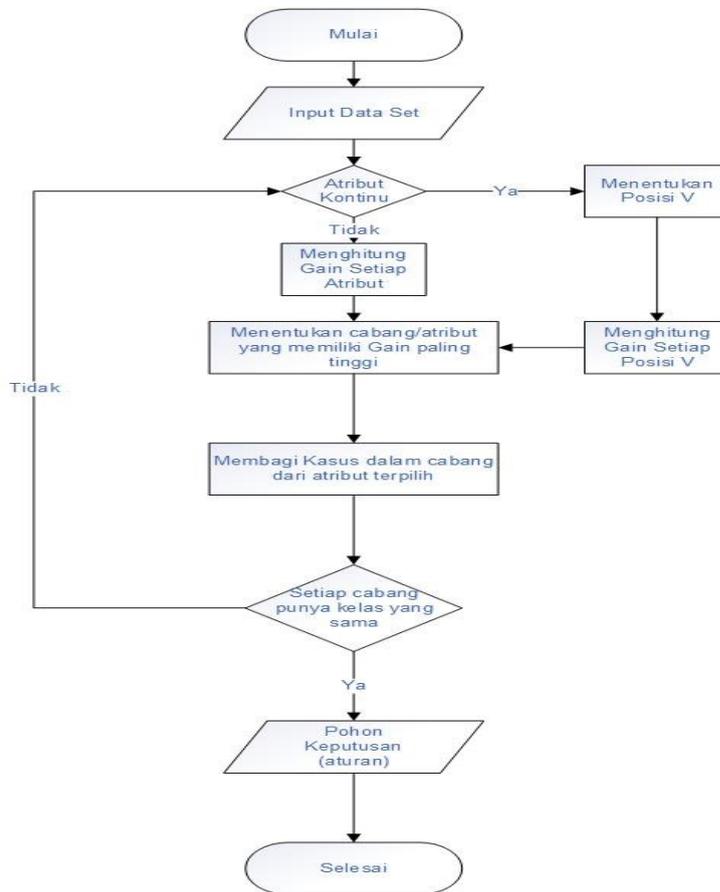
Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C4.5 adalah rasio gain, yang diformulasikan oleh persamaan berikut:

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

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. 5 dan beserta penjelasannya :

- (1) Pada proses cek keputusan dilihat apakah data set tersebut termasuk Atribut Kontinu atau bukan, Atribut Kontinu adalah atribut yang mempunyai jangkauan real;
- (2) Jika Iya maka akan menentukan posisi  $V$ , setelah itu menghitung gain setiap  $V$ . Jika tidak maka menghitung gain setiap atribut;
- (3) Setelah itu menentukan cabang/atribut yang memiliki gain paling tinggi;
- (4) Setelah itu membagi kasus dalam cabang dari atribut yang terpilih;
- (5) 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)**

## 5. PHP

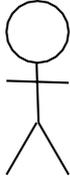
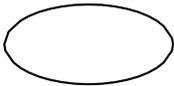
PHP atau kependekan dari Hypertext Preprocessor adalah salah satu bahasa pemrograman open source yang sangat cocok atau dikhususkan untuk pengembangan web dan dapat ditanamkan pada sebuah skripsi HTML. Bahasa PHP dapat dikatakan menggambarkan beberapa bahasa pemrograman seperti C, Java, dan Perl serta 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”. Menurut Kustiyarningsih (2011:114), “PHP (atau resminya PHP:

Hypertext Preprocessor) adalah skrip bersifat server – side yang ditambahkan ke dalam HTML”. Teoritis terkait dengan objek permasalahan.

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

Menurut (Dharwiyanti & Wahono, 2003) menjelaskan bahwa *Unified Modeling Language (UML)* adalah sebuah gambar yang digunakan untuk memvisualisasikan, merancang dan mendokumentasikan sistem perangkat lunak yang akan dikembangkan.

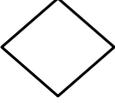
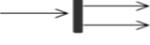
**Tabel 2. 1 Simbol UseCase Diagram**

SIMBOL	NAMA	KETERANGAN
	<p style="text-align: center;"><i>ACTOR</i></p>	<p>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 akan menggunakan kata benda diawal frase nama actor.</p>
	<p style="text-align: center;"><i>USE CASE</i></p>	<p>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.</p>

SIMBOL	NAMA	KETERANGAN
	<p style="text-align: center;"><i>ASOSIASI/ ASSOCIATION</i></p>	<p>Komunikasi antara actor dan use case yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.</p>
<p style="text-align: center;">&lt;&lt;extend&gt;&gt;  </p>	<p style="text-align: center;"><i>EKSTENSI/ EXTEND</i></p>	<p>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.</p>
	<p style="text-align: center;"><i>GENERALISASI / GENERALIZATION</i></p>	<p>Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.</p>
<p style="text-align: center;">&lt;&lt;include&gt;&gt;  </p>	<p style="text-align: center;"><i>MENGGUNAKAN INCLUDE</i></p>	<p>Relasi use case tambahan ke sebuah use case untuk menjalankan fungsional</p>

(Sumber: Dharwiyanti & Wahono, 2003)

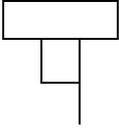
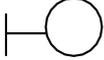
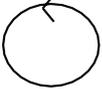
**Tabel 2. 2 Simbol Activity Diagram**

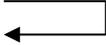
<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
	<i>STATUS AWAL/INITIAL</i>	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	<i>AKTIVITAS/ ACTIVITY</i>	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	<i>PERCABANGAN/ DECISION</i>	Asosiasi 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)

**Tabel 2. 3 Simbol Sequence Diagram**

<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.

SIMBOL	NAMA	KETERANGAN
	<i>Lifeline</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
	<i>General</i>	Merepresentasikan entitas tunggal dalam sequence.
	<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

SIMBOL	NAMA	KETERANGAN
		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)

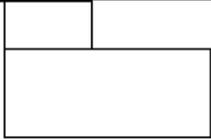
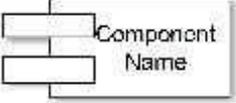
Tabel 2. 4 Simbol Class Diagram

SIMBOL	NAMA	KETERANGAN
	<i>Generalization</i>	Hubungan dimana objek anak (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor).
	<i>Nary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari dua (2) objek.
	<i>Class</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
	<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
	<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek

SIMBOL	NAMA	KETERANGAN
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri.
	<i>Association</i>	Apa yang menghubungkan antara objek satu dengan obyek lainnya.

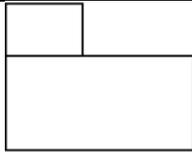
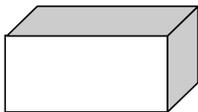
(Sumber: Dharwiyanti & Wahono, 2003)

**Tabel 2. 5 Simbol Component Diagram**

SIMBOL	NAMA	KETERANGAN
	<i>Package</i>	Merupakan sebuah bungkus dari satu atau lebih komponen
	<i>Component</i>	Komponen sistem
	Kebergantungan/ <i>dependency</i>	Kebergantungan antar node, arah panah mengarah pada node yang dipakai
	<i>Link</i>	Relasi antar node

(Sumber: Dharwiyanti & Wahono, 2003)

**Tabel 2. 6 Simbol Deployment Diagram**

SIMBOL	NAMA	KETERANGAN
	<i>Package</i>	Merupakan sebuah bungkus dari satu atau lebih mode
	<i>Node</i>	Biasanya mengacu pada perangkat keras (hardware), perangkat lunak yang tidak dibuat sendiri (software), jika didalam node disertakan komponen untuk mengkonsistenkan rancangan maka komponen yang diikut sertakan harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen.
	Kebergantungan/ <i>dependency</i>	Kebergantungan antar node, arah panah mengarah pada node yang

(Sumber: Dharwiyanti & Wahono, 2003)

### B. Teoritis terkait dengan objek permasalahan

Pemahaman Konveksi, dilansir dari LinkedIn, Konveksi ialah perusahaan yang beroperasi di sektor industri baju jadi. Konveksi memang biasa dipakai warga untuk mengatakan industri baju jadi.

Konveksi Menurut Pemerintah Dengan luas, konveksi masuk dalam kelompok Industri Tekstil serta Produk Tekstil dipersingkat industri TPT Indonesia. Industri ini terdiri atas beberapa macam industri yang membuat satu susunan dari hulu ke hilir. Serangkaian meliputi industri serat serta benang (fiber), pemintalan, penenunan serta perajutan, persetakan atau pengecapan dan industri baju jadi (garmen).

Ketidaksamaan Konveksi dengan Produsen Pakaian Yang lain Seringkali ada orang salah pengertian dengan konveksi. Konveksi pada umumnya mempunyai keunikan dibandingkan produsen pakaian lain. Diantaranya pesanan dengan minimum order (jumlahnya banyak) serta Ukuran size bugar (S,M,L,XL,dst).

Terdapat beberapa jenis bahan yang biasa digunakan untuk keperluan konveksi, diantaranya Drill, bahan ini biasa digunakan untuk pembuatan seragam kantor, lembaga maupun sekolah. Dapat juga dipakai untuk baju lapangan dan berbagai kebutuhan lain nya. Kemudian bahan CVC (Chief Value Cotton) bahan ini dibuat dri serat kombinasi dari cotton dan polyester. Biasa digunankan untuk keperluan seragam perusahaan, maupun kegiatan yang bersifat gathering. Kemudian **Cotton Combed** adalah kain katun yang paling halus dan telah mengalami proses penyisiran (**combed**) terlebih dahulu sebelum dipintal menjadi gulungan benang. Jenis kain ini terbuat dari 100% serat kapas sehingga lebih sejuk dan mudah menyerap keringat ketika digunakan. Digunakan untuk bahan baku pakain jenis kaos unisex. Kemudian jenis Milano. Merupakan jenis kain rajutan yang terbuat dari campuran katun, rambut, sutra, polyester, viskos, dan spandex dalam komposisinya. jenis kain ini sangat nyaman digunakan sepanjang hari. bahan ini juga dapat menyerap keringat dengan baik sehingga membuat tubuh tidak lembab. diperuntukan untuk pembuatan jersey sepakbola dan kegiatan olahraga lain nya. seperti funbike, motocross, dll. Lalu ada kain ripstop. Merupakan kain tenun yang sering terbuat dari benang nilon, dalam penenunan menggunakan teknik memperkuat khusus yang membuat mereka tahan terhadap robek. kain jenis ini memiliki kelebihan dalam rasio kekuatan -to-weight yang menguntungkan dan bahwa air tidak dapat dengan mudah menyebar. kain ini banyak digunakan untuk pembuatan seragam tentara, ransel, tas koper dll.

### C. Tinjauan Studi

Penelitian rujukan pada penelitian ini adalah sebagai berikut:

**(1) Analisis Persediaan Stok Barang Menggunakan Algoritma C 4.5 (Studi Kasus CV Harapan Raya)** (Darsono Nababan, Anastasia Venessa Tanlim, 2019)

Dalam sebuah perusahaan distributor, stok merupakan salah satu faktor penting yang perlu diperhatikan dipertimbangkan oleh perusahaan. Jumlah stock yang keluar masuk harus diperhatikan, tujuannya adalah untuk menjaga stok yang tersedia stabil. Stok yang stabil berarti di dalam gudang terdapat tidak ada kelebihan stok atau kekurangan stok. Masalahnya setiap item memiliki pembelian yang berbeda minat. Maka diperlukan perhitungan untuk dapat memprediksi item apa

saja yang perlu ditambahkan atau dikurangi di gudang. Algoritma C4.5 adalah sekelompok algoritma yang menggunakan keputusan pohon. Pohon keputusan adalah metode klasifikasi yang sangat kuat dan terkenal dan ramalan. Semakin kaya informasi atau pengetahuan yang dikandung oleh data pelatihan, semakin akurasi akan meningkat. Algoritma ini digunakan untuk menganalisis waktu pembelian barang yang telah habis dengan mengelompokkan barang-barang yang sudah ada dalam persediaan ditambah atau tidak, sehingga ketersediaan barang tetap stabil. Penerapan algoritma C 4.5 dapat digunakan untuk memprediksi ketersediaan stok di CV Harapan Raya. Hasil menghitung Nilai entropi pada algoritma C 4.5 dapat memprediksi jumlah stok yang disediakan oleh CV Harapan Raya.

**(2) Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-Commerce** , Oleh : G L Pritalia. . 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. Atribut dalam penelitian ini adalah Sku, Nat\_Inv, Fore\_3\_Month, Sales\_1\_Month, Pieces\_Past\_Due, dan Back Order. 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. Hasil dari analisis menggunakan algoritma C4.5 adalah untuk menentukan waktu penentuan ketersediaan barang memiliki tingkat keakuratan sebesar 98.9%.

**(3) Penerapan Algoritma C4.5 Untuk Menentukan Persediaan Obat (Studi Kasus Di RS Bethesda Yogyakarta)**, Oleh : Gunartatik Esthiningtyas, Putri Taqwa Prasetyaningrum. 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 persepsan 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 atau perlu dikurangi.

#### **(4) PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN ALAT MEDIS MENGGUNAKAN ALGORITMA C4.5 PT. MURNI INDAH SENTOSA,** Oleh: Abdul

Fikri, Wiwi Verina. Alat Medis merupakan perangkat apapun atau peralatan yang bertujuan untuk keperluan para dokter-dokter yang bertugas di rumah sakit maupun klinik. PT. Murti 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 priode 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.

#### **(5) Penerapan Algoritma C4.5 untuk Memprediksi Penjualan Motor Pada PT.Cappella Dinamik Nusantara Cabang Muka Kuning (Nurul Azwanti Program Studi Sistem Informatika, Fakultas Teknik, Universitas Putera Batam 2018),**

Oleh : Nurul Azwanti. Berdasarkan hasil penelitian yang dilakukan maka dapat disimpulkan bahwa Algoritma C4.5 dapat digunakan sebagai metode klasifikasi dalam memprediksi penjualan motor oleh PT. Capella Dinamik Nusantara Cabang Muka Kuning dengan memperhatikan nilai gain (penguatan) tertinggi dari empat atribut seperti Semester, Warna, Tahun Produksi, Segmen dan Harga. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi rule prediksi untuk menggambarkan proses yang terkait dengan prediksi penjualan motor. Dari pohon keputusan yang telah terbentuk, variabel tertinggi dalam memprediksi

penjualan adalah variabel penjualan via, artinya pendistribusian motor dari berbagai tempat sangat mempengaruhi terhadap penjualan motor. Harga dalam range Sangat Terjangkau Sekali (STS) yaitu harga dibawah Rp 17.699.000,00 masih mendominasi motor matic.

**(6) PENERAPAN ALGORITMA C4.5 UNTUK MENENTUKAN DATA STOK DAN TARGET PERMINTAAN MATERIAL YANG PALING DIBUTUHKAN GUDANG LOGISTIK PADA PT PLN (Persero) AREA KEBON JERUK,** Oleh: Bahrin Said Renhoran, Nova Nurhandayani, Laila Septiana. Setelah melakukan penelitian ini maka dengan adanya sistem peramalan persediaan barang dapat membantu mempermudah proses pelayanan pemilik dalam menyediakan barang untuk kedepannya. Variabel yang digunakan dalam penelitian ini meliputi: Nama Barang, Satuan, Kategori Material, Valuation Type, Jumlah Barang, Fisik Material, Stok Out/Target Permintaan. Dari hasil penilaian dapat disimpulkan kelebihan dan kekurangan pada proses penelitian diatas, sebagai berikut: Permasalahan dalam menentukan data stok material/barang dapat 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%. Dengan adanya penerapan *Data Mining* algoritma C4.5 diharapkan mampu memberikan solusi dalam menentukan permintaan barang yang paling dibutuhkan dengan permintaan yang banyak dari masing-masing user pada PT PLN (Persero) Area Kebon Jeruk.

**(7) Penerapan *Data Mining* untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5,** oleh : Juna Eska. Penjual harus lebih cermat menyediakan wallpaper mana yang akan banyak peminatnya, sehingga perlu perekomendasi jenis jenis wallpaper dengan menggunakan metode Klasifikasi dilakukan menggunakan *Data Mining* algoritma C4.5. data yang diperlukan yaitu data merk wallpaper terbaik, warna, motif, kualitas bahan, ukuran, dan harga. Setelah dilakukan perhitungan dengan menggunakan excel, selanjutnya dilakukan perhitungan menggunakan aplikasi rapid miner. Setelah diproses , maka di dapatkan *Decision Tree*. Keluaran yang dihasilkan dalam penelitian ini yaitu: Banyak atau Sedikit dari jumlah motif wallpaper. Algoritma C4.5 adalah algoritma klasifikasi data bertipe pohon keputusan. Pohon keputusan Algoritma C4.5 dibangun dengan beberapa tahap yang meliputi pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan membagi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dari penyelesaian pohon keputusan maka akan didapatkan beberapa rule.

**(8) Analisis Persediaan Stok Barang Menggunakan Algoritma C 4.5 (Studi Kasus CV Harapan Raya)**, Oleh: Darsono Nababan, Anastasia Venessa Tanlim. Dalam sebuah perusahaan distributor, stok merupakan salah satu faktor penting yang perlu diperhatikan dipertimbangkan oleh perusahaan. Atribut dalam penelitian ini adalah : Nama Barang, Bahan Barang, Jenis Barang. Jumlah stock yang keluar masuk harus diperhatikan, tujuannya adalah untuk menjaga stok yang tersedia stabil. Stok yang stabil berarti di dalam gudang terdapat tidak ada kelebihan stok atau kekurangan stok. Masalahnya setiap item memiliki pembelian yang berbeda minat. Maka diperlukan perhitungan untuk dapat memprediksi item apa saja yang perlu ditambahkan atau dikurangi di gudang. Algoritma C4.5 adalah sekelompok algoritma yang menggunakan keputusan pohon. Pohon keputusan adalah metode klasifikasi yang sangat kuat dan terkenal dan ramalan. Semakin kaya informasi atau pengetahuan yang dikandung oleh data pelatihan, semakin akurasi akan meningkat. Algoritma ini digunakan untuk menganalisis waktu pembelian barang yang telah habis dengan mengelompokkan barang-barang yang sudah ada dalam persediaan ditambah atau tidak, sehingga ketersediaan barang tetap stabil. Rule yang telah diperoleh dengan Algoritma C 4.5, diujikan dengan Software *Data Mining* yaitu Rapidminer. Proses ini akan menampilkan pola terbaik yang dapat diterapkan untuk menentukan kriteria unit. Penerapan algoritma 4.5 dapat digunakan untuk memprediksi ketersediaan stok di CV Harapan Raya.

**(9) Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)**, Oleh: Pandu Pratama Putra, Andi Supriadi Chan. Sumber daya perusahaan terbatas sehingga harus dikelola dengan efektif dan efisien agar menjamin kelangsungan perusahaan. Dibutuhkan suatu teknologi sistem informasi yang akan membantu manajemen dalam mengelola perusahaan agar setiap keputusan yang diambil ditunjang dengan informasi aktual dan dapat dipercaya. Penerapan algoritma C 4.5 bisa digunakan untuk memprediksi stock motor pada Saudara Motor. Atribut yang digunakan dalam penelitian ini adalah Type Motor, Warna Motor, Jenis Motor. Melalui tahapan-tahapan proses yang ada pada algoritma C 4.5 yaitu mempersiapkan data training, menghitung akar dari pohon, dan proses partisi hasil keputusan. Proses hasil keputusan akan berhenti saat semua record dalam simpul N mendapat kelas yang sama, tidak ada atribut di dalam record yang di partisi lagi, tidak ada record di dalam cabang yang kosong. Rule yang telah diperoleh dengan Algoritma C 4.5, diujikan dengan Software *Data Mining* yaitu Rapidminer. Proses ini akan menampilkan pola terbaik yang dapat diterapkan untuk menentukan kriteria unit. Hasil dari perhitungan nilai entropy pada algoritma C 4.5

dapat memprediksi jumlah stok motor yang disediakan dealer saudara motor ditambah atau dikurang. Proses penerapan *Data Mining* menggunakan algoritma C4.5 diperoleh pohon keputusan sehingga menghasilkan 12 aturan-aturan (rule) dalam menentukan prediksi persediaan stock motor pada dealer Saudara Motor.

**(10) PREDIKSI PENJUALAN PRODUK ROTI MENGGUNAKAN ALGORITMA C4.5 PADA PT. PRIMA TOP BOGA**, Oleh: M. Makmun Effendi, Diah Rahmawati.

Penjualan adalah kegiatan yang perlu diperhatikan dalam perusahaan, karena ujung dari sebuah produksi adalah penjualan barang. Atribut dalam penelitian ini yaitu : Nama barang, Bentuk, Ukuran, Topping, dan Warna. Ketersediaan stok barang tergantung dari hasil penjualan, semakin penjualan tinggi semakin besar untung yang diperoleh perusahaan. Untuk menjaga kestabilan penjualan pada perusahaan, banyak perusahaan yang menggunakan *Data Mining* untuk membantu menyelesaikan permasalahan produksi. Algoritma C4.5 adalah algoritma yang mempunyai fungsi untuk memprediksi atau meramalkan sebuah kejadian dimasa mendatang dengan menggunakan data sebelumnya. Mengukur tingkat akurasi barang laku dan kurang laku dengan aplikasi Rapid miner .Prediksi ini sangat menguntungkan bagi perusahaan dengan mengeluarkan hasil terjaga kestabilan stok untuk memenuhi produk penjualan yang laku dan kurang laku dipasaran. Dengan menggunakan pohon keputusan, algoritma C4.5 mampu memprediksi hasil penelitian dengan akurasi yang tinggi. Analisa data menggunakan algoritma ini memiliki tingkat akurasi sampai dengan 93%.

**Tabel 2. 7 Tinjauan Studi**

No	Peneliti	Judul	Sumber	Kontribusi
1	Darsono Nababan, Anastasia Venessa Tanlim (2019)	Analisis Persediaan Stok Barang Menggunakan Algoritma C 4.5 (Studi Kasus CV Harapan Raya)	<a href="https://ejournal.upbatam.ac.id/index.php/jif/article/view/839">https://ejournal.upbatam.ac.id/index.php/jif/article/view/839</a>	Pengambilan <i>output</i> ditambah dan dikurang untuk dijadikan sebagai <i>output</i> penelitian Rapid Miner untuk mempermudah memprediksi kompetensi siswa menggunakan metode C45.

2	G L Pritalia (2018)	Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-Commerce	<a href="https://ojs.uajy.ac.id/index.php/IJIS/article/view/1727">https://ojs.uajy.ac.id/index.php/IJIS/article/view/1727</a>	Pengambilan cara mengklasifikasi barang mana yang sudah waktunya di tambah stok maupun belum untuk di jadikan sebagai referensi penelitian.
3	Gunartatik Esthiningtyas, Putri Taqwa Prasetyaningrum (2020)	Penerapan Algoritma C4.5 Untuk Menentukan Persediaan Obat (Studi Kasus Di RS Bethesda Yogyakarta)	<a href="http://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/view/3">http://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/view/3</a>	Pengambilan <i>output</i> ditambah dan dikurang untuk dijadikan sebagai <i>output</i> penelitian.
4	Abdul Fikri, Wiwi Verina (2020)	Penerapan <i>Data Mining</i> Untuk Prediksi Penjualan Alat Medis Menggunakan ALGORITMA C4.5 PT. MURNI INDAH SENTOSA	<a href="https://e-journal.potensi-utama.ac.id/ojs/index.php/INFOSYS/article/view/1080">https://e-journal.potensi-utama.ac.id/ojs/index.php/INFOSYS/article/view/1080</a>	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
5	Nurul Azwanti (2018)	Penerapan Algoritma C4.5 untuk Memprediksi Penjualan Motor Pada PT.Cappela Dinamik Nusantara Cabang Muka Kuning	<a href="http://e-journals.unmul.ac.id/index.php/JIM/article/view/629">http://e-journals.unmul.ac.id/index.php/JIM/article/view/629</a>	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

6	Bahrin Said Renhoran, Nova Nurhandayani, Laila Septiana (2019)	Penerapan Algoritma C4.5 Untuk Menentukan Data Stok Dan Target Permintaan Material Yang Paling Dibutuhkan Gudang Logistik Pada PT PLN (Persero) AREA KEBON JERUK	<a href="https://ejournal.nusamandiri.ac.id/index.php/inti/article/view/1564">https://ejournal.nusamandiri.ac.id/index.php/inti/article/view/1564</a>	Pengambilan perhitungan menggunakan algoritma c4.5 dijadikan sebagai metode penelitian.
7	Juna Eksa (2016)	Prediksi untuk penjualan wallpaper terbanyak dengan metode klasifikasi menggunakan metode <i>Data Mining</i> algoritma c4.5. variabel yang dipakai adalah harga, ukuran, motif, bahan, warna	<a href="https://osf.io/preprints/inarxiv">https://osf.io &gt; preprints &gt; inarxiv &gt;</a>	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
8	Darsono Nababan, Anastasia Venessa Tanlim (2019)	Analisis Persediaan Stok Barang Menggunakan Algoritma C 4.5 (Studi Kasus CV Harapan Raya),	<a href="https://ejournal.upbatam.ac.id/index.php/jif/article/view/839">https://ejournal.upbatam.ac.id/index.php/jif/article/view/839</a>	Pengambilan perhitungan menggunakan algoritma c4.5 dan <i>output</i> ditambah dan dikurang dijadikan sebagai metode penelitian.

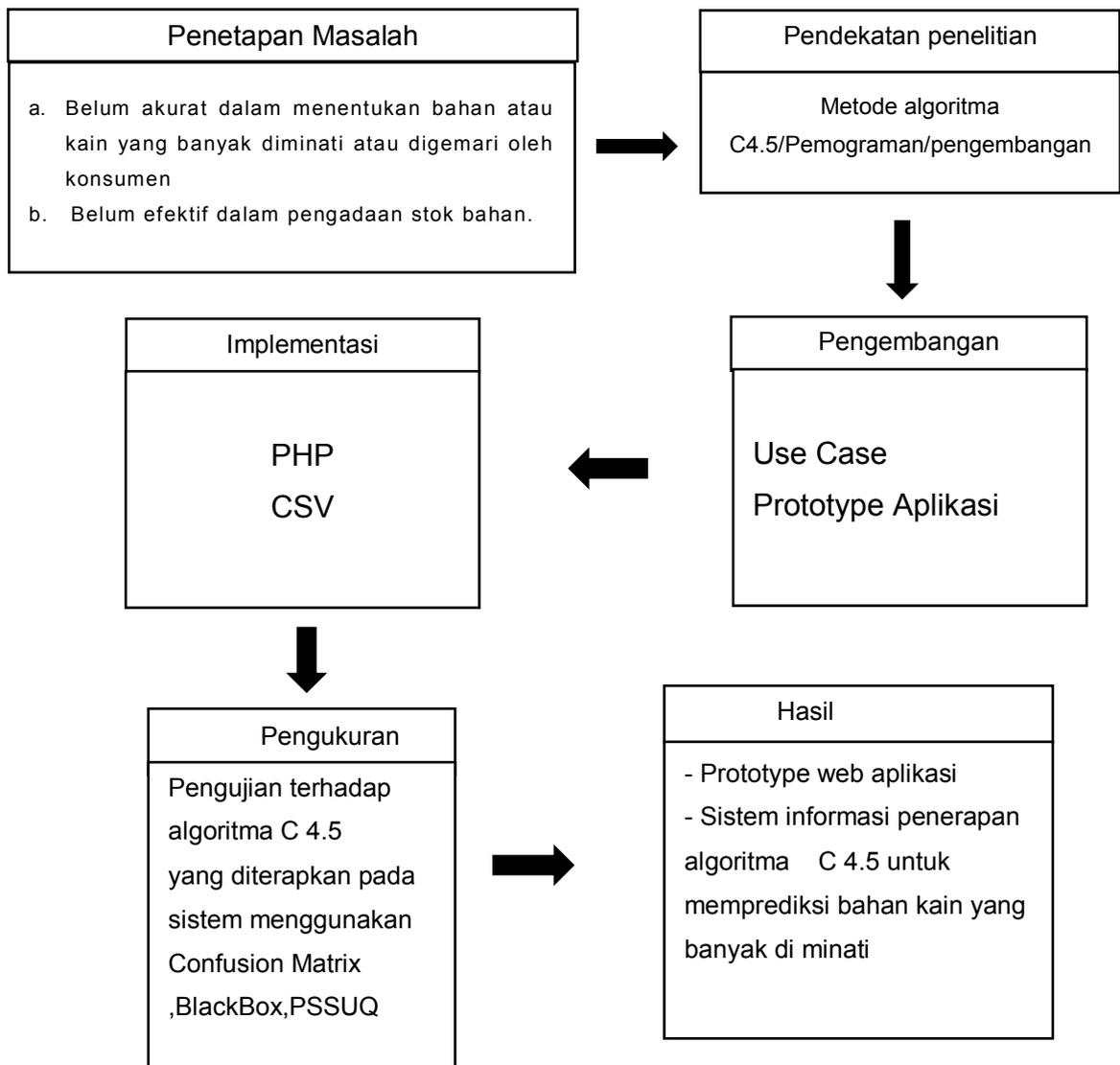
9	Pandu Pratama Putra, Supriadi (2019)	Andi Chan	Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Ea Motor)	<a href="http://ejournal.polbeng.ac.id/index.php/ISI/article/view/296/233">http://ejournal.polbeng.ac.id/index.php/ISI/article/view/296/233</a>	Pengambilan <i>output</i> ditambah dan dikurang untuk dijadikan sebagai <i>output</i> penelitian.
10	M. Effendi, Rahmawati (2018)	Makmun Diah	Prediksi Penjualan Produk Roti Menggunakan Algoritma C4.5 Pada PT. PRIMA TOP BOGA	<a href="https://jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/464">https://jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/464</a>	Penggunaan perhitungan menggunakan Rapid Miner dan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

Berdasarkan referensi jurnal pada table 2.7, terdapat penelitian yang berfokus pada stok barang dengan menggunakan metode ALGORITMA C.45. dari penelitian tersebut dapat diketahui bahwa metode C.45 dapat digunakan untuk mengklasifikasikan barang yang harus ditambah stok maupun yang belum memerlukan stok tambahan dengan menggunakan atribut Sku, Nat\_Inv,Fore\_3\_Month, Sales\_1\_Month, Pieces\_Past\_Due, dan Back Order.

Berdasarkan penelitian tersebut akan dilakukan penelitian menggunakan algoritma C.45 untuk prediksi bahan konveksi yang banyak diminati dengan menggunakan atribut jumlah order, jenis bahan, dan stok.

#### D. 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



Gambar 2. 7 Kerangka Pemikiran

Keterangan :

Berdasarkan Gambar 2.7 dapat diketahui

- (1) Pengambilan data pada penelitian ini dilakukan dengan cara Wawancara dan Obserasi.
- (2) Identifikasi Permasalahan dalam penelitian ini yaitu belum akurat dan efektif dalam memprediksi bahan konveksi paling banyak diminati
- (3) Pendekatan penelitian pada penelitian ini menggunakan metode algoritma C.4.5 sebagai dasar analisis data penelitian dan pengembangan yang digunakan dalam pembangunan system nantinya menggunakan metode *prototype*.
- (4) Tujuan penelitian ini yaitu, dapat membantu bagian analis dalam memprediksi bahan konveksi paling banyak diminati.
- (5) Aplikasi dan komponen didalamnya yaitu ada yang digunakan untuk pembangunan sistem prototype dalam penelitian pengembangan ini ada diagram UML dan ERD, desain database menggunakan MySQL, pemrograman menggunakan Bahasa pemrograman PHP. Sebaran angket sebagai proses pengujian system prototype dan implementasi sebagai dasar evaluasi dan validasi system prototype yang akan dibuat nantinya.
- (6) Hasil yang diharapkan dari penelitian ini yaitu sebuah Prototype aplikasi sistem informasi memprediksi bahan konveksi paling banyak diminati

## **E. Hipotesis**

Dalam teori data mining, algoritma C4.5 adalah salah satu algoritma yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan dari pengalaman dimasa sebelumnya. Algoritma C4.5 sudah banyak dikenal dan digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru di penelitian-penelitian sebelumnya dengan permasalahan yang serupa. Permasalahan yang dihadapi oleh peneliti yaitu belum efektif dan belum akurat proses memprediksi bahan konveksi yang banyak diminati. Berdasarkan hal tersebut maka dapat ditetapkan, hipotesis pada penelitian ini adalah penerapan Algoritma C4.5 diduga dapat memprediksi bahan konveksi yang banyak diminati.