

## **BAB II**

### **KERANGKA TEORITIS**

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

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

Sebagai bidang ilmu yang relative baru, saat ini Data Mining menjadi salah satu pusat perhatian para akademis maupun praktisi. Menurut (Suntoro, 2019: 6 ) data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.

Data mining adalah proses menganalisa data dari yang berbeda dan menyimpulkannya menjadi informasi atau pengetahuan atau pola yang penting untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya Knowledge Discovery In Database (KDD) berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut menurut (Witten, 2016: 6):

- (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;

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 bentuk yang mudah dipahami.

##### **2. Bahasa Pemrograman**

###### **(a) PHP**

(Miftahul Jannah, 2016). PHP adalah singkatan dari “PHP: *Hypertext Preprocessor*”, yaitu bahasa pemrograman yang digunakan secara luas untuk penanganan pembuatan dan pengembangan sebuah situs web,

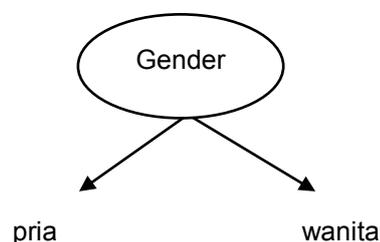
mulai dari halaman web yang sederhana sampai aplikasi kompleks yang membutuhkan koneksi ke database. *PHP* merupakan script yang menyatu dengan HTML dan berada pada server ( server side HTML embedded scripting). PHP (Hypertext Preprocessor), merupakan bahasa pemrograman web bersifat serverside, artinya bahasa berbentuk script yang disimpan dan dijalankan di computer server(WebServer) sedang hasilnya yang di kirimkan ke computer client (WebBrowser) dalam bentuk script HTML (Hypertext Markup Language).

(b) HTML

Menurut (Dieterici, 2018) menjelaskan bahwa “Hypertext Markup Language (HTML) adalah sebuah bahasa dasar bahasa markup untuk memformat konten halaman web yang digunakan untuk merancang halaman website statis. Menurut (Dieterici, 2018), “HTML adalah singkatan dari Hypertext Markup Language yang merupakan suatu kode semi pemrograman yang menjadi dasar terwujudnya web”.

### 3. Pengertian Algoritma C4.5

Menurut (Novita, 2016), Algoritma yang dapat digunakan untuk memprediksi atau mengklasifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain algoritma C4.5, yang merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J.Ross Quinlan. Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples merupakan data contoh yang digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang digunakan sebagai parameter dalam klasifikasi data.



**Gambar 2. 1 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

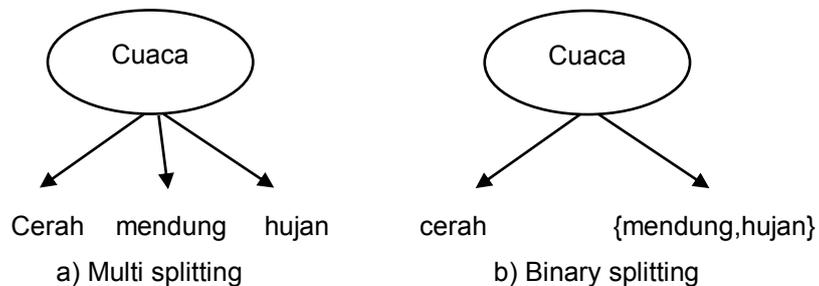
Fitur yang hanya mempunyai dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang. Contoh pemecahannya disajikan pada Gambar 2.1.

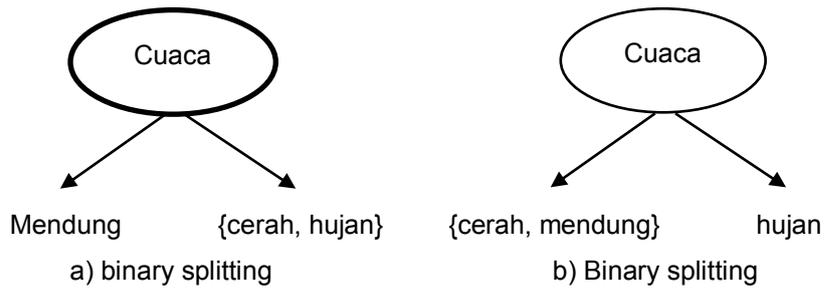
(b) Fitur bertipe kategorikal

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 2.2. 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 2.2 dan Gambar 2.3.

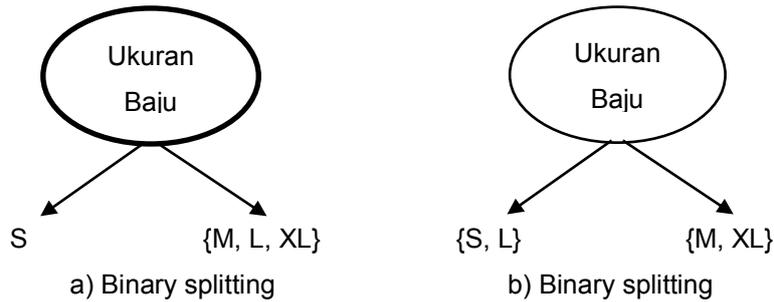
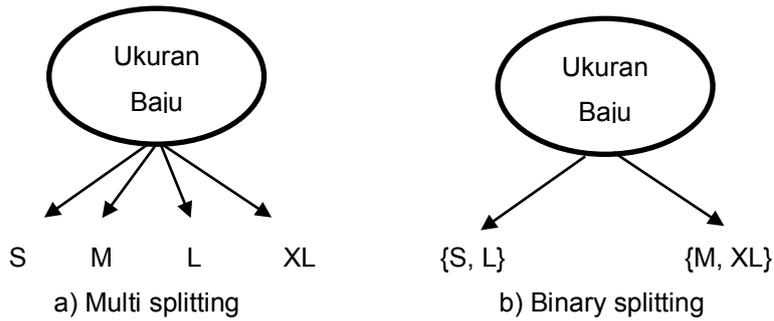
(c) Fitur bertipe numerik

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 2.4.

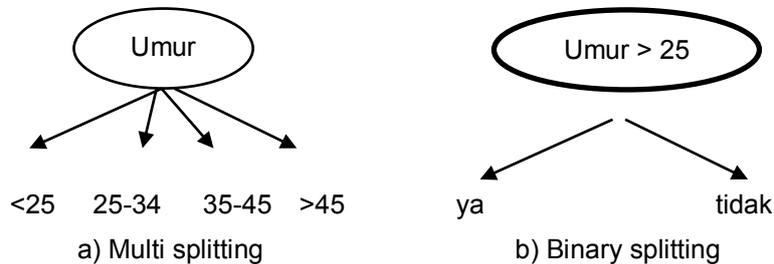




**Gambar 2. 2 Syarat pengujian fitur bertipe nominal**



**Gambar 2. 3 Syarat pengujian fitur bertipe ordinal**



**Gambar 2. 4 Syarat pengujian fitur bertipe numerik**

Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C 4.5 adalah rasio gain, yang diformulasikan oleh persamaan berikut:

$$RasioGain(s,j) = \frac{Gain(s,j)}{SplitInfo(s,j)} \quad (1)$$

Menurut (Prasetyo,2014), persamaan diatas menyatakan nilai rasio gain pada fitur ke-j. SplitInfo(s,j) dapat dari  $SplitInfo(s,j) = -\sum_{i=1}^k p(v_i|s) \log_2 p(v_i|s)$  dimana k menyatakan jumlah pemecahan.

#### 4. Pengertian Klasifikasi

Menurut (Prasetyo, 2013), 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. Klasifikasi bertujuan untuk mengklasifikasikan item data menjadi satu dari beberapa kelas standar. Sebagai contoh, suatu program email dapat mengklasifikasikan email yang sah dengan email spam. Beberapa algoritma klasifikasi antara lain pohon keputusan, nearest neighbor, naive bayes, neural networks dan support vector machines.

#### 5. Pengembangan Sistem SDLC

Pendekatan sistem merupakan sebuah metodologi. Metodologi adalah satu cara yang direkomendasikan dalam melakukan sesuatu. Pendekatan system adalah metodologi dasar dalam memecahkan segala jenis masalah. Siklus Hidup pengembangan system (System Development Life Cycle –SDLC) adalah aplikasi dari pendekatan system bagi pengembangan suatu system informasi. (Kristianto, 2019) 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 direncanakan dan diatur. Ini juga menganalisis sistem yang ada untuk memahami masalah dan menentukan persyaratan fungsional untuk sistem baru. Selanjutnya, sistem

baru ini akan dirancang dan diimplementasikan. Setelah implementasi, sistem ini idealnya digunakan untuk jangka waktu yang lama. SDLC tradisional sering disebut sebagai pendekatan air terjun karena pekerjaan di atas mengikuti pola yang teratur dan dilakukan dari atas ke bawah. Kegiatan ini memiliki aliran satu arah yang mengarah ke akhir proyek.



**Gambar 2. 5 Pola Melingkar Siklus Kehidupan**

(Sumber: Raymond McLeod 2007: 199)

## 6. Pengertian Unified Modeling Language (UML)

*Unified Modelling Language (UML)* adalah sebuah bahasa yang telah menjadi standar dalam industry untuk visualisasi, merancang dan mendokumentasikan sistem perangkat lunak. *UML* menawarkan sebuah standar untuk merancang model sebuah sistem. (Rosa & Shalahuddin, 2013)

**Tabel 2. 1 Simbol UseCase**

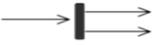
SIMBOL	NAMA	KETERANGAN
—	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.

<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
     	<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 sendiri, yang menunjukkan urutan kejadian yang terjadi

SIMBOL	NAMA	KETERANGAN
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

(Sumber : Habibi, Putra , & Putri, 2020).

**Tabel 2. 2 Simbol Diagram Activity**

SIMBOL	NAMA	KETERANGAN
	<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 : Habibi, Putra , & Putri, 2020).

**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.
	<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

SIMBOL	NAMA	KETERANGAN
	<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.

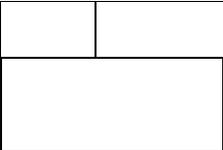
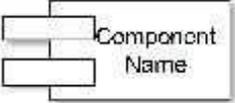
(Sumber : Habibi, Putra , & Putri, 2020).

**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 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
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung padanya elemen

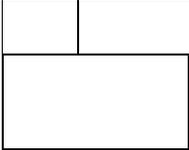
		yang tidak mandiri
	Association	Apa yang menghubungkan antara objek satu dengan obyek lainnya.

**Tabel 2. 5 Simbol Component Diagram**

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

(Sumber : Habibi, Putra , & Putri, 2020).

**Tabel 2. 6 Simbol Deployment Diagram**

SIMBOL	NAMA	KETERANGAN
	<i>Package</i>	Merupakan sebuah bungkus dari satu atau lebih mode
		Biasanya mengacu pada perangkat keras (hardware), perangkat lunak yang tidak dibuat sendiri (software), jika didalam node

	<i>Node</i>	disertakan komponen untuk mengkonsistenkan rancangan maka komponen yang diikuti 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 : Habibi, Putra , & Putri, 2020).

## 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) **Prediksi Penjualan Produk Roti Menggunakan Algoritma C4.5 Pada PT. Prima Top Boga** (M. Makmun Effendi, 2018)

Berdasarkan hasil penelitian mengenai “Prediksi penjualan Roti Menggunakan Algoritma C4.5 pada PT. Prima Top Boga” dengan akurasi mencapai 93.12% maka penulis dapat menyimpulkan bahwa, brownies dengan bentuk bulat kurang laku atau kurang diminati, sedangkan dengan bentuk persegi panjang lebih laku dan brownies dengan Topping Almond sangat laku dan diminati daripada topping cheese dan chocochip, oleh konsumen PT. Prima Top Boga.

### 2) **Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce** (Lukhayu Pritalia, 2018)

Berdasarkan hasil implementasi dan uji coba aplikasi, dapat disimpulkan bahwa: Data mining dengan algoritma C4.5 dapat diimplementasikan untuk penentuan ketersediaan barang e-commerce dengan dua kategori yaitu barang tersebut dalam kondisi tidak atau belum tersedia untuk pemesanan barang untuk pelanggan/backorder dan barang dalam kondisi tersedia untuk pemesanan barang untuk pelanggan. Aplikasi Weka berhasil menentukan ketersediaan barang e-commerce dengan presentase keakuratan sebesar 98% dari 5000 sampel data dan 23 atribut dengan kecepatan waktu menjalankan program selama 0.38 detik. Hasil penentuan ketersediaan barang e-commerce dari aplikasi penelitian ini dapat membantu bagian perusahaan untuk mengetahui status barang yang siap untuk

diperjualbelikan. Hal ini dapat menjadi rekomendasi pengambilan keputusan dalam menerima pesanan agar barang tersebut tetap terjaga ketersediaannya dan tidak mengalami backorder. Hal ini penting agar pelanggan tetap terjaga kepercayaannya dalam melakukan transaksi barang.

**3) Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor) (Putra & Chan, 2018)**

Berdasarkan penelitian dan pembahasan yang dilakukan, maka dapat disimpulkan beberapa hal. Hasil dari perhitungan nilai entropy telah memberikan nilai. Berdasarkan proses perhitungan yang telah dilakukan melalui beberapa langkah pada proses penerapan Data Mining menggunakan algoritma C4.5 maka diperoleh pohon keputusan tersebut diperoleh 12 aturan-aturan (rule) dalam menentukan prediksi jumlah persediaan stock motor pada dealer Saudara Motor.

**4) Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri (Rosita Dewi & Farouq Mauladi, 2020)**

Berdasarkan hasil perancangan, implementasi dan pengujian sistem prediksi Toko Dewi Sri ini dapat disimpulkan bahwa:

Sistem prediksi penjualan menggunakan algoritma C4.5 telah berhasil dibuat dan berfungsi dengan baik dengan fitur menu utama, fitur menu data admin, fitur menu tambah data admin, fitur menu data barang, fitur menu tambah data barang, fitur menu data penjualan, fitur menu tambah data penjualan, fitur data mining, fitur pohon keputusan, fitur prediksi, dan fitur akurasi. Analisis prediksi penjualan menggunakan algoritma C4.5 dapat dipergunakan untuk memprediksi penjualan obat pertanian di periode yang akan datang pada Toko Dewi Sri berdasarkan data penjualan dari bulan Oktober sampai bulan November tahun 2019.

**5) Penerapan Algoritma C4.5 Untuk Menentukan Data Stok Dan Target Permintaan Material Yang Paling Dibutuhkan Gudang Logistik PADA PT PLN (Persero) Area Kebon Jeruk. (Bahrun Said Renhoran, Nova Nurhandayani, 2018)**

Setelah melakukan penelitian ini maka dengan adanya sistem peramalan persediaan barang dapat membantu mempermudah proses pelayanan pemilik dalam menyediakan barang untuk kedepannya. 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.

**6) APLIKASI PREDIKSI PENJUALAN AC MENGGUNAKAN DECISION TREE DENGAN ALGORITMA C4.5** (Izyuddin & Wibisono, 2020)

Dalam penelitian ini menghasilkan sebuah aplikasi yang dapat digunakan untuk memprediksi penjualan produk AC terlaris. Dengan kemampuan melakukan prediksi penjualan AC terlaris ini maka manajemen Global Elektronik dapat secara lebih tepat dalam memperhitungkan ketersediaan stok produk AC, sehingga permintaan konsumen pada produk AC dapat terpenuhi, dengan demikian kepuasan konsumen dapat terjaga dan keuntungan yang diperoleh dapat lebih besar.

**7) IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI POLA PEMBELIAN SEPEDA MOTOR PADA SHOWROOM CV. VIVA MAS MOTORS DENGAN METODE ALGORITMA C4.5** (Purwadi, 2018)

Setelah dilakukan analisa dan implementasi pembahasan maka penelitian ini menyimpulkan beberapa hal yaitu Hasil dari analisa yang diperoleh dari data mining Teknik Klasifikasi dengan metode Decision Tree mengenai pola pembelian sepeda motor dapat membantu CV. VIVA MAS Motors mengetahui merk dan tahun sepeda motor yang laku dipasaran. Dan juga ditemukan pola yang saling berkaitan erat antara atribut merk, tahun dan harga sepeda motor terhadap penjualan sepeda motor. Hasil analisa yang didapat dengan menggunakan aplikasi Rapid Miner dalam memprediksi penjualan sepeda motor di CV. VIVA MAS Motors memperoleh hasil yang akurat bila dibandingkan dengan proses perhitungan manual. Dan menghasilkan rule atau aturan yang tepat sesuai dengan proses manualnya.

**8) ANALISIS PERSEDIAAN STOK BARANG MENGGUNAKAN ALGORITMA C 4.5 (STUDI KASUS CV HARAPAN RAYA)** (Darsono Nababan, 2019)

Berdasarkan pembahasan dan penelitian dapat kita simpulkan beberapa hal. Berdasarkan proses perhitungan yang telah dilakukan melalui beberapa langkah pada proses penerapan Data Mining menggunakan algoritma C4.5 maka diperoleh pohon keputusan tersebut diperoleh 9 aturan-aturan (rule) dalam menentukan prediksi jumlah persediaan stock barang pada CV Harapan Raya. Selanjutnya penulis menyarankan agar hasil analisis ini dapat di implementasikan dengan baik pada perusahaan.

**9) Penerapan Data Mining Untuk Klasifikasi Data Penjualan Makanan Terlaris Dengan Algoritma C45 (Zakir et al., 2020)**

dari penelitian ini dapat disimpulkan, dari pembuatan sistem klasifikasi makanan laris dan tidak laris dengan penerapan algoritma klasifikasi C4.5 yang dibuat dengan menggunakan PHP. Kesimpulan ini didapatkan berdasarkan perkembangan selama proses pembuatan sistem. Algoritma C4.5 dapat digunakan untuk memudahkan dalam pengambilan keputusan dengan memproyeksikan data-data yang ada ke dalam bentuk pohon keputusan, berdasarkan nilai entropy dan gain yang dimiliki masing-masing atribut data. Untuk hasil prediksi yang lebih akurat dibutuhkan data dalam jumlah besar, artinya semakin besar jumlah data yang digunakan maka semakin akurat hasil prediksi yang dihasilkan. Menghasilkan sistem yang dapat menentukan makanan laris dan tidak laris menggunakan algoritma C4.5 untuk klasifikasi penjualan makanan laris dan tidak laris.

**10) Klasifikasi Rasa Susu Almond Berdasarkan Minat Pelanggan Menggunakan Algoritma C4.5 di Home Made MamiNia Ice Pematangsiantar (Febriani et al., 2021)**

Berdasarkan hasil penelitian yang telah dilakukan, menunjukkan bahwa atribut Almond Milk Original (C1) memiliki pengaruh paling tinggi dalam klasifikasi minat tertinggi. Hasil yang diperoleh dari perhitungan manual dengan menggunakan Microsoft Excel 2010 dan hasil yang diberikan oleh Software RapidMiner Studio Versi 5.3 mendapatkan hasil berupa rule dan pohon keputusan yang sama. Dan dengan adanya penerapan data mining Algoritma C4.5 dalam melakukan klasifikasi rasa susu almond berdasarkan minat pelanggan dapat membantu owner MamiNia Ice Pematangsiantar dalam mengambil keputusan dalam menentukan rasa susu almond mana yang paling diminati pelanggannya.

**Tabel 2. 7 Tinjauan Studi**

No	Peneliti	Judul	Sumber	Kontribusi
1	Diah Rahmawati, M. Makmun Effend	Prediksi Penjualan Produk Roti Menggunakan Algoritma C4.5 Pada PT. Prima Top Boga	<a href="https://www.jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/464">https://www.jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/464</a>	Kontribusi terhadap penelitian ini adalah pengambilan metode Algoritma C4.5. Kelemahannya kurangnya variabel yang digunakan..

2	G L Pritalia	Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce ALGORITMA C4.5	<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>	Kontribusi terhadap penelitian ini adalah pengambilan metode Algoritma C4.5 dan cara perhitungan C4.5 dalam prediksi pengambilan variabel.
3	Pandu Pratama Putra, Andi Supriadi Chan	Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor	<a href="https://www.researchgate.net/publication/333143097_">https://www.researchgate.net/publication/333143097_</a>	Kontribusi terhadap penelitian ini adalah pengambilan metode Algoritma C4.5 dan mengklasifikasi data dengan akurat
4	Kiki Rosita Dewi, Kemal Farouq Mauladi, Masruroh	Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri	<a href="https://proceeding.unpkediri.ac.id/index.php/inotek/article/download/72/46/134">https://proceeding.unpkediri.ac.id/index.php/inotek/article/download/72/46/134</a>	Kontribusi: Pengambilan variabel jumlah anggota keluarga untuk dijadikan variabel yang akan digunakan.

5	Nova nurhandayani	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>	Kontribusi terhadap penelitian ini adalah Metode <i>decision tree</i> yang digunakan di penelitian ini akurat sehingga menjadi contoh untuk penelitian ini
6	Ade Izyuddin	APLIKASI PREDIKSI PENJUALAN AC MENGGUNAKAN DECISION TREE DENGAN ALGORITMA C4.5	<a href="https://ejournal.stmiklombok.ac.id/index.php/misi/article/view/208">https://ejournal.stmiklombok.ac.id/index.php/misi/article/view/208</a>	Kontribusi terhadap penelitian ini adalah Penentuan jumlah persediaan setiap model AC ditentukan secara subjektif, hanya didasarkan pada perkiraan manajemen saja, tanpa mempertimbangkan perbandingan antara penjualan dan persediaan barang.
7		IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI POLA PEMBELIAN SEPEDA MOTOR PADA SHOWROOM CV. VIVA MAS MOTORS DENGAN	<a href="https://jurnal.kaputama.ac.id/index.php/JSIK/article/view/106">https://jurnal.kaputama.ac.id/index.php/JSIK/article/view/106</a>	Kontribusi dari penelitian ini adalah Hasil analisa yang didapat dengan menggunakan aplikasi Rapid Miner dalam memprediksi penjualan sepeda motor di CV. VIVA MAS

		METODE ALGORITMA C4.		Motors memperoleh hasil yang akurat bila dibandingkan dengan proses perhitungan manual. Dan menghasilkan rull atau aturan yang tepat sesuai dengan proses manualnya.
8	Darsono Nababan, Anastasia Venessa Tanlim	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>	Kontribusi penelitian ini adalah pengambilan metode Algoritma C4.5.
9	Ahmad Zakir, Yermia Ndruru, Edrian Hadinata, Ihsan lubis	Penerapan Data Mining Untuk Klasifikasi Data Penjualan Makanan Terlaris Dengan Algoritma C45	<a href="http://jifti.upnjatim.ac.id/index.php/jifti/article/view/33">http://jifti.upnjatim.ac.id/index.php/jifti/article/view/33</a>	Konstruksi dari penelitian ini yaitu memastikan tingkatan confidence tertinggi yang dicapai setiap variabel pada semua produk, pemenuhan dan prioritas variabel untuk menganalisis dan memprediksi.
10	Febriani, Indra Gunawan, Rafiq Dewati	Klasifikasi Rasa Susu Almond Berdasarkan Minat Pelanggan	<a href="https://journal.y3a.org/index.php/satesi/article/view/33">https://journal.y3a.org/index.php/satesi/article/view/33</a>	Konstruksi dari penelitian ini adalah Algoritma C4.5 akan

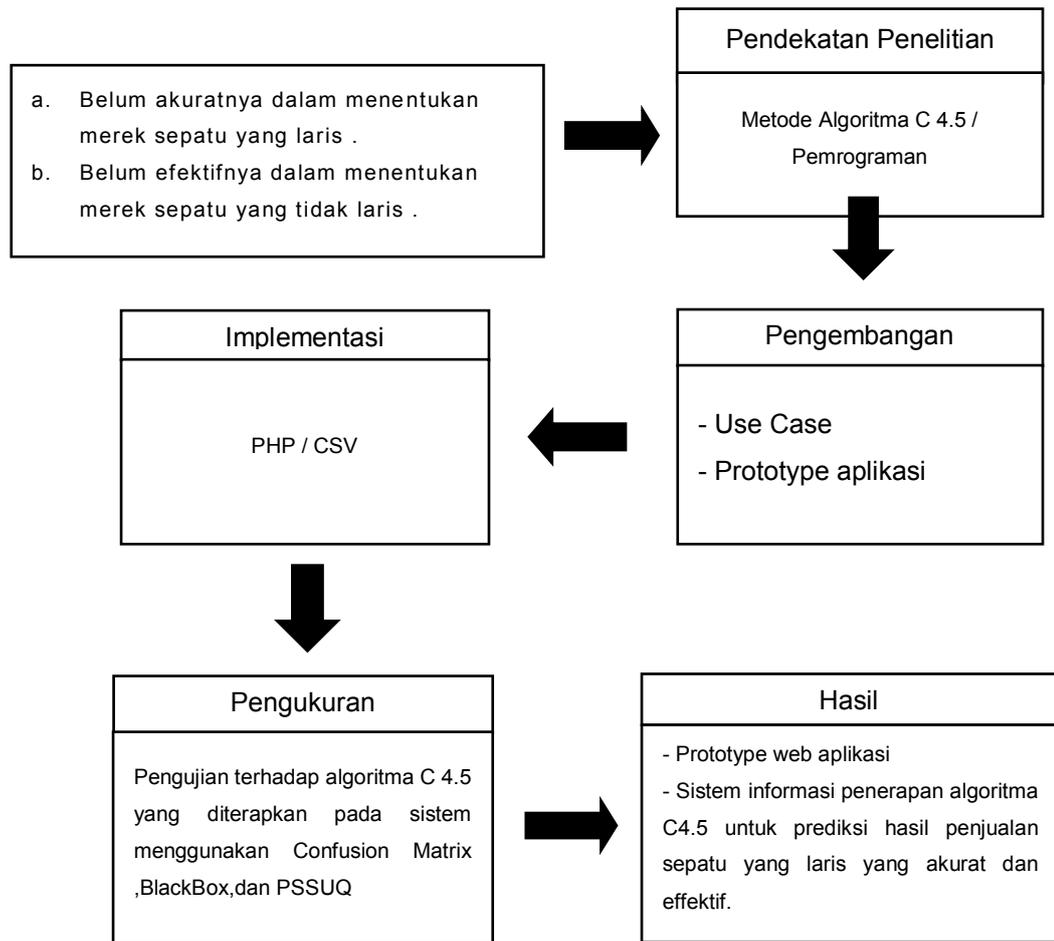
	Dedy Hartama , Muhammad Ridwan Lubis	Menggunakan Algoritma C4.5 di Home Made MamiNia Ice Pematangsiantar		didapatkan sebuah Pohon Keputusan yang mudah dipahami dan mudah dimengerti. Dengan demikian dapat membantu pemilik menentukan perencanaan produksi susu almond yang paling banyak diminati serta tanpa khawatir terjadi kelebihan barang atau kekurangan bahan.
--	--------------------------------------------	---------------------------------------------------------------------------------	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Berdasarkan uraian rujukan yang sudah dilakukan sebelumnya dapat disimpulkan kekurangan yaitu terdapat jurnal yang hanya menggunakan aplikasi yang sudah tersedia dan beberapa jurnal tidak dapat melakukan perhitungan akurasi, terdapat juga kesamaan metode yaitu algoritma C4.5, menggunakan *Confusion matrix* sebagai uji hasilnya. Pada penelitian ini yaitu melakukan pengembangan aplikasi serta menghitung tingkat akurasi. Selain itu pada penggunaan Variabel pada penelitian kali ini menggunakan Variabel data penjualan selama satu tahun, merek, kode, warna, ukuran, indent, stok, terjual, sisa, harga

### C. Kerangka Pemikiran

Berdasarkan dukungan landasan teoritis yang diperoleh dari eksplorasi teori yang dijadikan rujukan penelitian, maka dapat disusun kerangka pemikiran sebagai berikut:

Penetapan masalah
-------------------



**Gambar 2. 6 Kerangka Pemikiran**

Keterangan :

Penelitian ini diawali dengan munculnya permasalahan terkait melalui pengumpulan data dengan cara observasi & wawancara yang dilakukan pada objek permasalahan.

1. Penetapan Permasalahan dalam penelitian ini yaitu belum akuratnya dalam prediksi pembelian sepatu yang laris dan belum efektifnya dalam menentukan produk yang tidak laris.
2. Pendekatan penelitian ini yaitu, untuk prediksi pembelian merek sepatu yang laris. Metode penelitian pengembangan ini menggunakan metode algoritma C.4.5 sebagai dasar analisis data penelitian dan pengembangan yang digunakan dalam pembangunan system nantinya menggunakan metode prototype.
3. 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.

4. implementasi sebagai dasar evaluasi dan validasi system prototype yang akan dibuat nantinya.
5. Pengukuran pengujian terhadap algoritma C4.5 yang diterapkan pada sistem menggunakan Confusion Matrix, Blackbox dan PSSUQ
6. Hasil yang diharapkan dari penelitian ini yaitu menghasilkan sistem informasi prototype khusus sebagai informasi terkait produk yang laku

#### **D. Hipotesis**

Berdasarkan permasalahan yang dihadapi yaitu belum efektif dan belum akurat pada saat melakukan proses pemilihan penjualan sepatu yang laris, 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 di masa sebelumnya, Salah satunya adalah metode C4.5. C4.5 adalah program yang digunakan untuk menghasilkan peraturan taksonomi dengan menggunakan pohon keputusan dari sekumpulan data yang diberikan. Berdasarkan hasil penelitian mengenai “Prediksi penjualan Roti Menggunakan Algoritma C4.5 pada PT. Prima Top Boga” dengan akurasi mencapai 93.12% maka penulis dapat menyimpulkan bahwa, brownies dengan bentuk bulat kurang laku atau kurang diminati, sedangkan dengan bentuk persegi panjang lebih laku dan brownies dengan Topping Almond sangat laku dan diminati daripada topping *cheese* dan *chocochip*, oleh konsumen PT. Prima Top Boga oleh (Diah Rahmawati). Berdasarkan hal tersebut, maka dapat ditetapkan hipotesis pada penelitian ini yaitu metode C4.5 diduga dapat menentukan prediksi merek sepatu yang laris tidak laris.