

BAB II

TINJAUAN PUSTAKA

2.1 Teori

Pada bagian teori ini, peneliti akan membahas makna, dan hal-hal yang terkait dengan judul penelitian ini dengan merinci beberapa prinsip dasar logika atau argumen yang menjadi landasan pemahaman bagi peneliti guna untuk mempermudah proses penelitian.

2.1.1 Teori Umum

Pada bagian teori umum ini, peneliti akan menjelaskan teori-teori yang menjadi landasan pemahaman bagi peneliti guna untuk mendukung dan memperkuat kesimpulan yang muncul pada penelitian ini.

2.1.1.1 KDD (*Knowledge Discovery in Database*)

Dalam menganalisis data untuk mencari informasi baru tentunya ada beberapa langkah yang harus di jalani seperti langkah pada *Knowledge Discovery in Database*, langkah ini telah digunakan pada penelitian dalam pemilihan jenis laptop dimana metode ini digunakan untuk memperoleh pengetahuan dari basis data yang ada, sehingga keputusan diambil berdasarkan pengetahuan yang diperoleh melalui tahapan yang membentuk dasar pengetahuan, dengan hasil penelitian terdapat tiga variabel yang sangat mempengaruhi pemilihan laptop, yaitu harga, prosesor dan masa pakai baterai (Baginda, 2023).

Langkah-langkah *Knowledge discovery in database* juga digunakan dalam penelitian dalam mendianogsa radang hati, dimana langkah tersebut antara lain:

1. Data preparation

Pada langkah ini data yang telah dipilih akan disaring, dimana data yang digunakan tidak semua dipakai hanya yang sesuai dengan analisis data yang akan dilakukan.

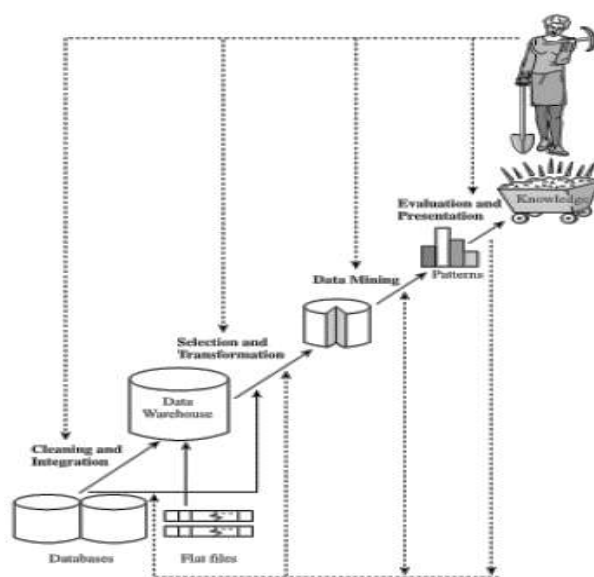
2. Algoritma *data mining*

Penggunaan algoritma *data mining* pada langkah ini guna untuk menggali data yang telah terintegrasi agar mempermudah proses identifikasi informasi bernilai.

3. Analisa data

Dari *data mining* dievaluasi guna melihat apakah knowledge domain dapat ditemukan dalam bentuk *rule* yang telah terekstrak dari jaringan (Amrin & Pahlevi, 2021)

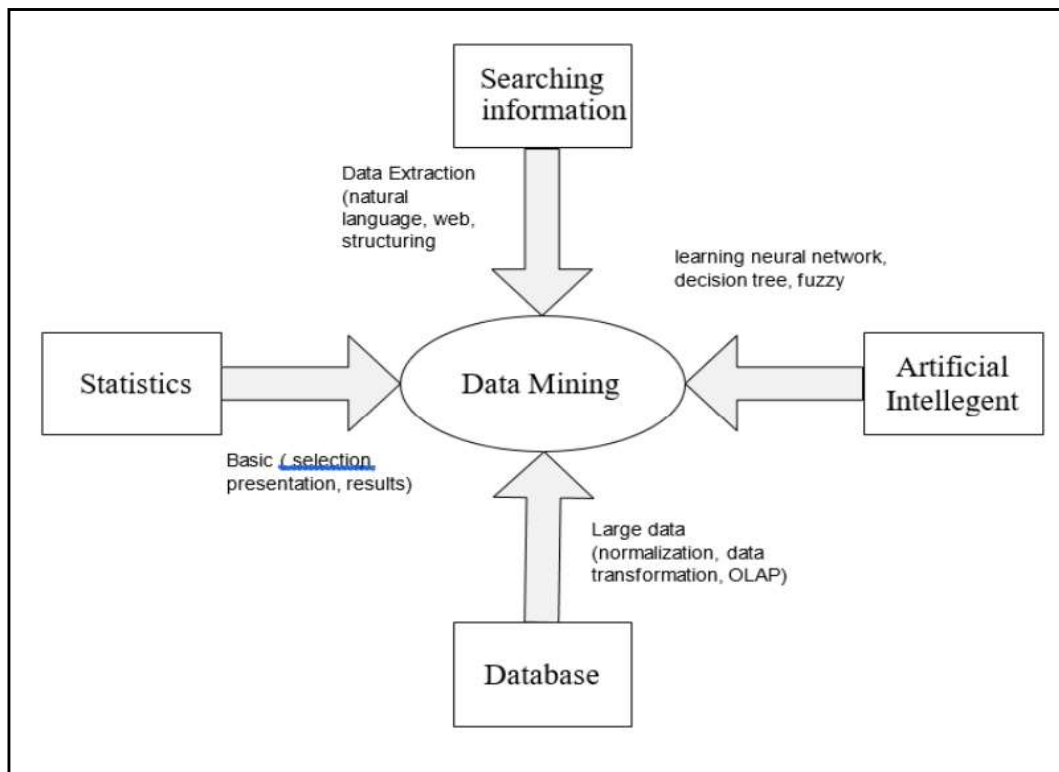
Gambar di bawah ini dapat menunjukkan apa itu langkah-langkah *Knowledge Discovery in Database*.



Gambar 2.1 Proses *Knowledge Discovery in Database* (KDD)

2.1.1.2 Data Mining

Menambang data dari *database* yang selama ini, dapat menghasilkan informasi baru seperti penelitian analisa peramalan trend mata uang kripto sehingga dapat membantu orang yang tertarik dalam investasi dibidang ini dapat mengetahui trend bitcoin saat ini, jadi dapat dikatakan, *Data Mining* merupakan tahap perolehan data atau biasa lebih dikenal proses penambang data yang bertujuan untuk mencari pola-pola yang menarik dengan maksud untuk mendapatkan pengetahuan dari suatu dataset yang memiliki volume data yang sangat besar, seperti yang ditemukan dalam Big Data (Julianto et al., 2022).



Sumber : (Darmawan, 2018).

Gambar 2.2 Bidang Ilmu *Datamining*

2.1.1.3 Pengelompokan *Data Mining*

Datamining memiliki peran yang penting dalam memperoleh informasi-informasi yang penting serta meningkatkan pemahaman bagi pengguna. Proses *data mining* melibatkan serangkaian tahapan umum, termasuk deskripsi, prediksi, estimasi, klasifikasi, pengklusteran, dan asosiasi (Wahono & Riana, 2020).

1. Deskripsi

Deskripsi mempunyai tujuan untuk mengidentifikasi pola yang berulang dalam data dan mengubahnya menjadi aturan serta kriteria yang dapat dipahami dengan mudah oleh para ahli di bidang aplikasinya. Aturan yang dihasilkan harus mudah dimengerti dan dipahami agar dapat meningkatkan pengetahuan dalam sistem. Tugas deskriptif adalah komponen penting dalam *data mining*, terutama dalam teknik *postprocessing*, yang digunakan untuk memvalidasi dan menjelaskan hasil dari proses *data mining*. *Postprocessing* adalah langkah yang diperlukan untuk memastikan bahwa hanya hasil yang valid dan bermanfaat yang disampaikan kepada pihak yang bersangkutan.

2. Prediksi

Prediksi memiliki kemiripan dengan klasifikasi, tetapi perbedaannya terletak pada pengelompokan data berdasarkan perilaku atau suatu nilai yang diperkirakan di masa depan atau waktu yang akan datang. Contoh tugas prediksi yaitu memproyeksikan penurunan jumlah pelanggan suatu toko dalam waktu yang akan datang dan meramalkan harga saham untuk tiga bulan ke depan.

3. Estimasi

Estimasi memiliki kemiripan dengan prediksi, namun fokusnya lebih pada variabel target yang bersifat numerik daripada kategori. Dalam pemodelan estimasi, model dibangun menggunakan data lengkap yang mengandung nilai variabel target sebagai nilai prediksi. Pada tahap berikutnya, estimasi nilai variabel target dilakukan berdasarkan nilai variabel prediksi. Sebagai contoh, dalam konteks rumah sakit, estimasi tekanan darah sistolik pada pasien dapat dilakukan dengan mempertimbangkan faktor seperti usia pasien, jenis kelamin, berat badan, dan tingkat natrium dalam darah. Hubungan antara tekanan darah sistolik dan variabel prediksi selama proses pembelajaran akan menghasilkan model estimasi.

4. Klasifikasi

Klasifikasi adalah tahapan di mana suatu model atau fungsi digunakan untuk mengidentifikasi dan memisahkan data menjadi berbagai kelas. Dalam klasifikasi, karakteristik dari objek dianalisis, dan objek tersebut ditempatkan dalam salah satu kelas yang telah ditentukan sebelumnya.

5. Pengklusteran

Pengklusteran/pengelompokan, atau lebih umum disebut *clustering*, adalah suatu proses menggabungkan data ke dalam kelompok tanpa mempertimbangkan kelas-kelas tertentu. Kluster merupakan kumpulan dari catatan yang memiliki kesamaan satu sama lain, tetapi berbeda dengan catatan dalam kelompok lain. Tujuan dari pengelompokan adalah menghasilkan kelompok objek yang serupa di dalam kelompok-kelompok

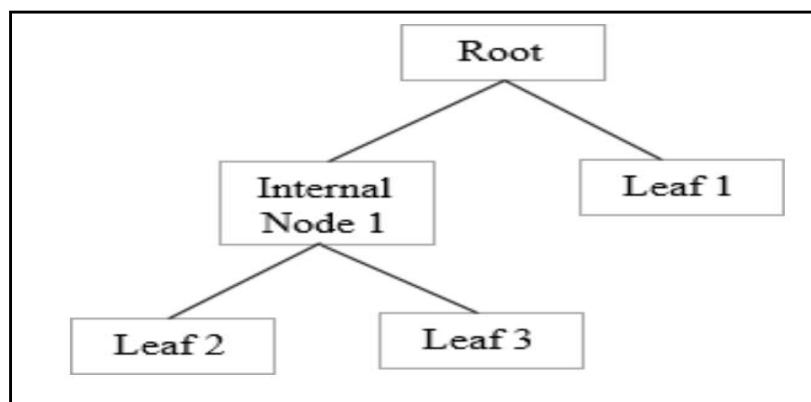
tertentu. Kualitas analisis pengelompokan meningkat ketika kesamaan objek dalam suatu kelompok semakin besar dan perbedaan antar kelompok semakin signifikan.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah mengidentifikasi atribut yang sering muncul bersama dalam waktu tertentu. Dalam konteks bisnis, ini sering dikenal sebagai analisis keranjang belanja (*market basket analysis*). Tujuan tugas asosiasi adalah untuk mengungkap pola dan aturan yang menjelaskan hubungan antara dua atau lebih atribut.

2.1.1.4 *Decision Tree*

Pohon keputusan disebutkan sebagai metode yang sangat didukung dan dikenal dengan baik dalam kelompok klasifikasi dan prediksi. Metode keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturan dapat dengan mudah dipahami dalam bahasa alami. Dan mereka juga dapat diungkapkan dalam bentuk basis data seperti bahasa *Structure Query* untuk mencari catatan dalam kategori tertentu (Darmawan, 2018).



Sumber : (Arif & Yulianto, 2022)

Gambar 2.3 Konsep dasar *Decision Tree*

2.1.1.5 Algoritma C4.5

Salah satu algoritma yang digunakan untuk mengubah informasi besar menjadi pohon keputusan (*decision tree*) yang menggambarkan aturan adalah C4.5. Terdapat beberapa perbaikan yang telah diterapkan pada C4.5, seperti kemampuan untuk mengatasi data yang hilang (*missing value*), data berkelanjutan (data kontinu), dan teknik pruning. Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan sebagai pengembangan dari algoritma ID3, dan algoritma ini digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling terkenal dalam klasifikasi. Pohon keputusan terdiri dari simpul-simpul, dan simpul pertama disebut akar, yang tidak memiliki simpul masukan (Adriansa et al., 2022).

$$Gains(S, A) = Entropi(S) - \sum_{i=1}^n \left\{ \frac{|S_i|}{|S|} \right\} *$$

Rumus 2.1 Rumus *Gain*

Di mana:

S: Himpunan kasus

A: Atribut

n: Jumlah partisi atribut A

|S_i|: Jumlah kasus partisi ke-i

|S|: Jumlah Kasus S

$$Entropi(S) = \sum_{i=1}^n p_i * \log_2 p_i$$

Rumus 2.2 Rumus Perhitungan *Entropy*

Di mana:

S: Perwakilan Himpunan

n: Perwakilan jumlah dari partisi yang terdapat pada S p_i adalah bagian dari S_i ke S

2.1.2 Teori Khusus

Pada bagian ini, peneliti akan membahas makna, dan hal-hal yang terkait dengan judul penelitian ini dengan merinci beberapa prinsip dasar logika atau argumen yang berkaitan langsung dengan variabel yang dipilih dan diambil oleh peneliti secara khusus membahas makna dari variabel-variabel tersebut.

2.1.2.1 Penjualan

Penjualan merupakan suatu sistem kerja yang terdiri dari berbagai komponen yang saling terkait dan berinteraksi untuk merumuskan rencana strategi di sektor usaha dengan tujuan memenuhi kebutuhan konsumen dan pencapaian tujuan tertentu (Surahmat, 2023). Pada umumnya penjualan merupakan kegiatan bisnis yang melibatkan proses menjual produk atau layanan kepada pelanggan atau konsumen, dengan tujuan untuk memperoleh pendapatan atau laba.

2.1.2.2 *Waikato Environment for Knowledge Analysis (WEKA)*

Waikato Environment for Knowledge Analysis, yang disingkat sebagai *WEKA*, adalah kumpulan alat praktis untuk mesin learning yang dikembangkan di Universitas Waikato, Selandia Baru, yang memiliki beragam aplikasi dan keahlian. *WEKA* memiliki sejumlah keunggulan, terutama dalam hal klasifikasi, yang sangat efektif dalam menyelesaikan permasalahan penambangan data di dunia nyata. *WEKA* juga mudah digunakan dalam berbagai tahap, serta menyediakan alat-alat untuk pra-pemrosesan data, klasifikasi, regresi, pengelompokan, aturan asosiasi,

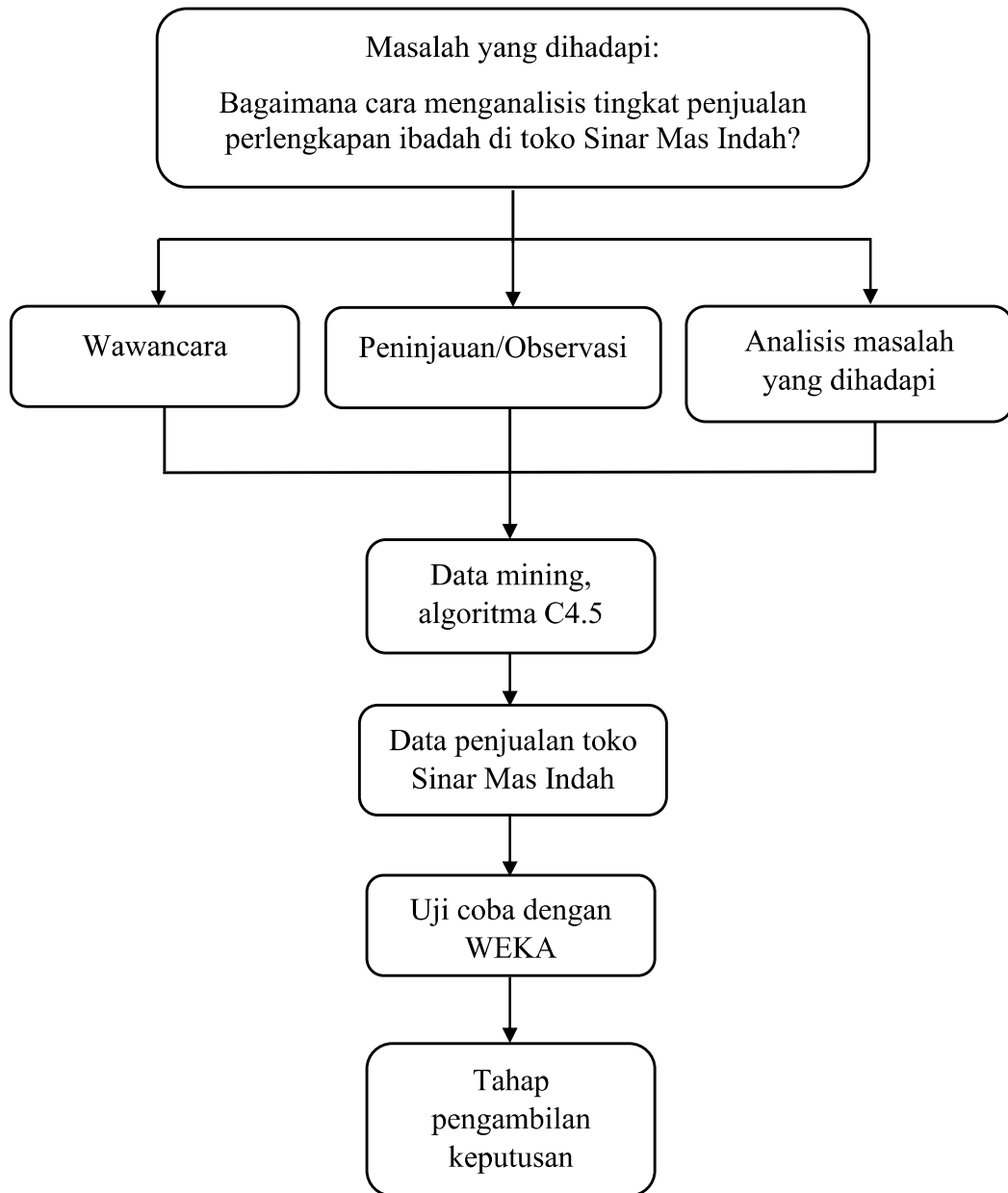
dan visualisasi. Aplikasi ini mendukung beberapa metode dalam penambangan data, terutama pohon keputusan, yang melibatkan node keputusan yang menghubungkan berbagai bagian mulai dari simpul akar hingga simpul daun (akhir). Ini merupakan bagian menarik dalam metode klasifikasi, di mana indikasi pada simpul keputusan diuji, dan setiap tindakan mengarah ke bagian yang sesuai. Setiap bagian kemudian di arahkan ke simpul lain atau simpul akhir untuk menghasilkan keputusan akhir (Baginda, 2023). Berikut adalah tampilan utama dari aplikasi *weka* ;



Gambar 2.4 Tampilan *WEKA*

2.2 Kerangka Pemikiran

Penelitian ini mengembangkan kerangka pemikiran yang menciptakan suatu gambaran yang akan digunakan oleh peneliti dalam proses penelitian untuk memperoleh hasil yang dapat dijadikan dasar untuk pengambilan keputusan. Di bawah ini merupakan struktur kerangka pikir yang peneliti lampirkan:



Gambar 2.5 Kerangka Pemikiran Peneliti

2.3 Hipotesis Penelitian

Hipotesis penelitian adalah suatu pernyataan atau dugaan yang diajukan oleh peneliti sebagai dasar untuk diuji atau diselidiki dalam rangkaian penelitian. Hipotesis menyatakan hubungan antara dua atau lebih variabel atau menyajikan jawaban yang diharapkan terhadap pertanyaan penelitian. Hipotesis digunakan sebagai kerangka acuan untuk mengarahkan penelitian dan mencari bukti empiris yang mendukung atau menolak dugaan tersebut. Hipotesis dapat berfungsi sebagai landasan untuk menyusun metode penelitian dan mengarahkan analisis data untuk mencapai kesimpulan yang lebih mendalam tentang topik penelitian.

1. *Data mining* menggunakan *decision tree* atau pohon keputusan dengan algoritma C4.5 diduga dapat memperoleh tingkat penjualan perlengkapan ibadah di toko Sinar Mas Indah.
2. Dengan menggunakan algoritma C.45 toko Sinar Mas Indah diduga akan mampu mengetahui perlengkapan apa yang sering dibeli oleh pembeli selama ini.

2.4 Penelitian Terdahulu

Penelitian terdahulu, atau yang juga disebut sebagai penelitian sebelumnya, merujuk pada studi atau penyelidikan yang telah dilakukan sebelumnya mengenai suatu topik atau masalah tertentu. Penelitian ini dapat mencakup temuan, metode, atau hasil penelitian sebelumnya yang relevan dengan subjek yang sedang diteliti. Dikarenakan penelitian ini menggunakan algoritma C4.5, jadi peneliti akan melampirkan beberapa penelitian terdahulu yang membahas tentang C4.5:

1. Sebuah penelitian dari karya (Iftitah & Setyadi, 2023) yang berjudul “Penerapan Algoritma C.45 Untuk Analisis Pengadaan Peralatan dan Mesin Kantor” menyatakan bahwa penerapan algoritma C4.5 pada data KIB peralatan dan mesin kantor memungkinkan prediksi usulan peralatan dan mesin kantor di BKAD Kabupaten Banyumas mencapai tingkat akurasi 100% dengan tingkat kesalahan klasifikasi 0%. Hasil penelitian ini menunjukkan bahwa nilai presisi dan recall juga mencapai 100%, yang berarti bahwa prediksi yang diberikan adalah sepenuhnya akurat dan seluruh klasifikasi yang dilakukan juga benar.
2. Sebuah penelitian dari karya (Yusuf Maulana et al., 2022) yang berjudul “Penerapan Data Mining dengan Algoritma C4.5 Dalam Memprediksi Penjualan Tempe” menyimpulkan bahwa faktor utama yang memengaruhi penjualan tempe, baik yang tinggi maupun rendah, dapat diselesaikan dengan menggunakan teknik *data mining* menggunakan Algoritma C4.5. Dalam proses ini, diperoleh sembilan aturan yang memiliki tingkat akurasi sekitar 70%. Hasil analisis menunjukkan bahwa faktor yang paling berpengaruh adalah jumlah (C1) dengan nilai gain sebesar 0,433764175. Jadi dapat disimpulkan bahwa *data mining* dengan algoritma C4.5 dapat membantu proses memprediksi penjualan tempe di AC Mandiri
3. Sebuah penelitian dari karya (Sulastri, 2022) yang berjudul “Identifikasi Tingkat Penjualan Produk Herbal HWI Menggunakan Algoritma C4.5” Penerapan metode Pohon Keputusan sangat bermanfaat bagi stokis produk herbal HWI di Benai. Hal ini membantu mereka dalam mengidentifikasi

produk yang diminati dan tidak diminati oleh pelanggan. Dengan demikian, stokis HWI di Benai dapat meningkatkan pasokan produk yang diminati dan dalam prosesnya, meningkatkan penjualan serta menjaga pelanggan setia. Penyebab utama kekurangan produk terlaris oleh stokis herbal HWI di Benai dapat dilihat dari faktor-faktor seperti status pembelian, jumlah pesanan yang diterima, dan status pelanggan.

4. Sebuah penelitian dari karya (Diansyah & Exprada, 2022) yang berjudul “Penerapan Algoritma C4.5 Untuk Memprediksi Penjualan Lele Pada Kolam Pancing Galatama” menyimpulkan bahwa Algoritma C4.5 dapat berfungsi sebagai metode klasifikasi untuk memproyeksikan penjualan lele di Kolam Pancing Galatama. Hal ini dilakukan dengan mempertimbangkan nilai gain tertinggi dari empat atribut utama, yakni berat, jenis, harga, dan jumlah terjual. Pohon Keputusan hasil dari penerapan Algoritma C4.5 menunjukkan bahwa variabel berat dan jenis memiliki gain tertinggi dalam memprediksi penjualan lele. Oleh karena itu, pengelola Kolam Pancing Galatama disarankan untuk memberikan perhatian lebih pada kedua variabel tersebut sebagai acuan utama dalam meningkatkan penjualan ikan lele.
5. Sebuah penelitian dari karya (Salsabilla & Sulastri, 2022) yang berjudul “Implementasi Algoritma C4.5 Untuk Klasifikasi Produk Laris Sepeda Motor Honda Pada Cv Cendana Motor Cepiring” menyimpulkan bahwa Berdasarkan pohon keputusan yang terbentuk, dapat disimpulkan bahwa variabel utama dalam meramalkan keberhasilan penjualan produk adalah

Nama Motor, yang mengindikasikan bahwa jenis motor memiliki pengaruh signifikan terhadap tingkat penjualan.

6. Sebuah penelitian karya dari (Wijaya & Mutatkin Bakti, 2023) yang berjudul “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Penjualan Alat-Alat Telekomunikasi Dan IT (Information Technology) Pada PT Julian Adiputra Utama” menyimpulkan bahwa Dengan menganalisis hasil prediksi menggunakan algoritma C4.5, diperoleh tingkat akurasi sebesar 72.31%.
7. Sebuah penelitian karya dari (Rusdiansyah et al., 2022) yang berjudul “Data Mining Implementation On SMUN Scholarship Recipient Candidates Using The C4.5 Algorithm” menyimpulkan bahwa klasifikasi untuk calon penerima beasiswa dapat mengelompokkan siswa pada tahap lulus atau tidak lulus dalam seleksi. Dari 33 data siswa yang digunakan, terlihat tingkat akurasi sebesar 81,81%. Dapat disimpulkan bahwa aplikasi algoritma C4.5 untuk rekomendasi penerima beasiswa sudah memiliki tingkat akurasi yang cukup baik, dan seleksi sudah memiliki aturan yang diterapkan melalui proses klasifikasi dengan *data mining*. Dengan demikian, aplikasi ini membantu badan mahasiswa dalam pemilihan calon penerima beasiswa untuk tahun berikutnya.
8. Sebuah penelitian karya (Tanyu et al., 2021) yang berjudul “Landslide Susceptibility Analyses Using Random Forest, C4.5, And C5.0 With Balanced And Unbalanced Datasets” yang menyimpulkan bahwa pada analisis yang dilakukan menggunakan C4.5 dengan semua 13 parameter

kondisioning (Skenario 3) baik untuk data citra raster maupun vektor (nilai akurasi dan kappa sebesar 99,8% dan 0,93 untuk data raster dan 96% dan 0,92 untuk data vektor, secara berturut-turut). Model RF menunjukkan efektivitas yang paling rendah dalam memprediksi longsor untuk ketiga skenario.

9. Sebuah penelitian karya (Elisa, 2022) yang berjudul “Penerapan Algoritma C4.5 Untuk Memprediksi Penjualan Barang pada PT Batam Bangun Prathama” dengan menggunakan algoritma C4.5 dan pengujian dengan aplikasi *WEKA* dapat mempermudah pelanggan dalam mengambil keputusan sebelum membeli barang di PT Batam Bangun Prathama mulai dari perbandingan harga, kualitas dan barang yang diberi garansi atau tidak.
10. Sebuah penelitian karya (Mulyanda et al., 2023) yang berjudul “Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas” membuktikan bahwa penggunaan algoritma C4.5 sangat tepat dalam memprediksi harga pasar mobil bekas, sehingga penjual mobil dapat meraup lebih banyak keuntungan.