

BAB II

KERANGKA TEORITIS

A. Landasan Teori

1. Pengertian Data Mining

Menurut (Pramudiono, 2006, Hal 12), menyampaikan bahwa data mining adalah proses analisa yang dilakukan secara otomatis pada data yang kompleks dan berjumlah besar untuk memperoleh sebuah pola atau kecenderungan yang umumnya tidak disadari.

Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu di ekstrasi agar menjadi informasi yang baru dan dapat membantu dalam pengambilan keputusan. (Suntoro, 2019, Hal 6).

Ada beberapa tahapan dalam data mining, yaitu (Meilani & Susanti, 2014, Hal 7):

a) Pembersihan Data (*Data Cleaning*)

Pembersihan data merupakan langkah menghilangkan noisedan data yang tidak konsisten atau data yang tidak relevan. pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian – isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu ada juga atribut – atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data – data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah kompleksitasnya.

b) Integrasi Data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut – atribut yang mengidentifikasi entitas – entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi.

Tahapan proses *Data Mining* ada 3(tiga) macam, yaitu:

- 1) Pengumpulan data (*data collection*)
- 2) Transformasi data (*data transformation*)
- 3) Analisis data (*data analysis*).

Proses tersebut diawali dengan *preprocessing* yang terdiri atas pengumpulan data untuk menghasilkan data mentah (*raw data*) yang dibutuhkan oleh data mining, yang kemudian dilanjutkan dengan transformasi data untuk mengubah data mentah menjadi format yang dapat diproses oleh data mining, misalnya melalui filtrasi atau agregasi. Hasil transformasi data akan digunakan oleh analisis data untuk membangkitkan pengetahuan dengan menggunakan teknik seperti *Machine learning*. *Machine learning* adalah pengambilan pengetahuan dari suatu data. Penerapan metode *machine learning* dalam beberapa tahun terakhir telah berkembang dimana-mana dalam kehidupan sehari-hari, secara khusus algoritma ini dapat menciptakan sebuah output untuk sebuah input yang belum pernah ada sebelumnya tanpa bantuan dari manusia. Algoritma *machine learning* yang belajar dari pasangan *input / output* disebut *supervised learning*. *Supervised learning* yang dipahami dengan baik dan kinerjanya mudah diukur. Jika aplikasi diformulasikan sebagai masalah penyebab perceraian dan dapat membuat dataset yang mencakup hasil yang diinginkan.

2. Pengertian Klasifikasi

Klasifikasi adalah sebuah proses untuk menemukan sekumpulan model yang dijelaskan kelas-kelas data, sehingga model tersebut bisa digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Maka jika ingin mendapatkan model, kita harus melakukan analisis terhadap data latih. Sedangkan data uji digunakan untuk mengetahui tingkat akurasi dan model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai dari suatu objek data. Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kamber, 2006). Klasifikasi data terdiri dari dua langkah proses yaitu :

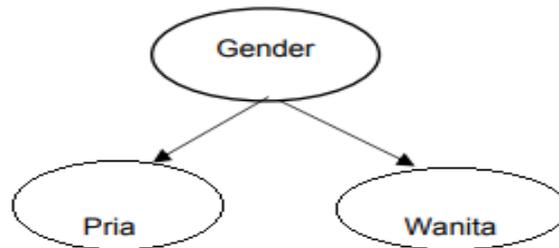
- (a) Pertama adalah *learning* (fase *training*), algoritma klasifikasi dibuat untuk menganalisis data training lalu dipresentasikan dalam bentuk rule klasifikasi.
- (b) Kedua adalah data tes digunakan untuk memperkirakan akurasi dari *rule*

klasifikasi (kamber,2006) proses klasifikasi ada empat komponen (gorunescu,2011):

1. *Class*, variabel dependenyang berupa kategorikal yang merepresentasikan label yang terdapat pada objek.
2. *Predictor*, variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
3. *Training Dataset*, data set yang berisi nilai dari kedua komponen yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
4. *Testing Dataset*, data baru yang akan diklasifikasikan oleh model yang telah dibuat dan klasifikasi dievaluasi.

B. Pengertian 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). Seperti dijelaskan pada gambar 2.1



Gambar 2. 1 Syarat pengujian fitur biner

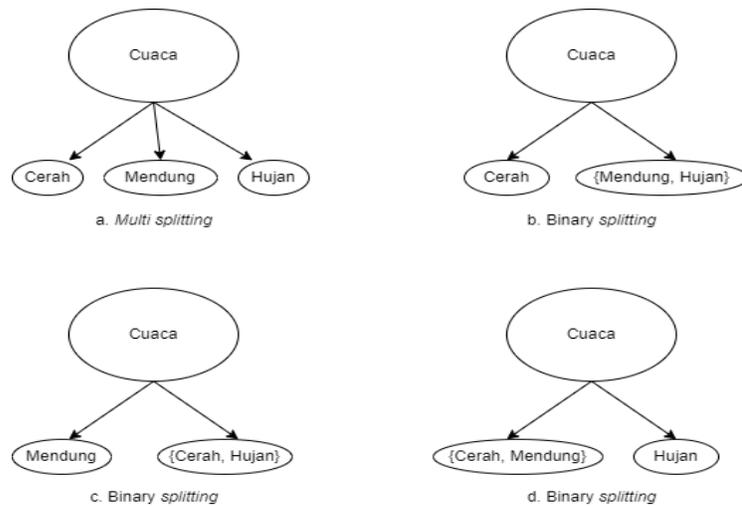
Yang menjadi hal penting dalam induksi *Decision Tree* adalah bagaimana menyatakan syarat pengujian pada node Ada 3 (tiga) kelompok penting dalam syarat pengujian node:

a. Fitur biner

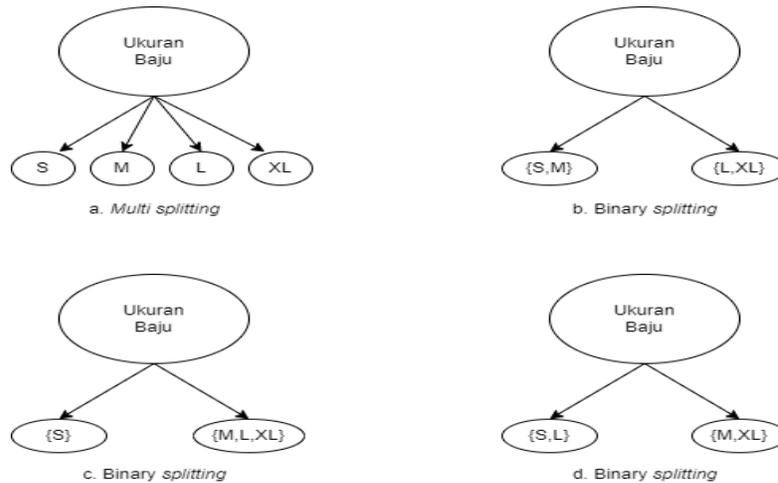
Fitur yang hanya mempunyai dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun *internal*) hanya punya 2 (dua) pilihan cabang. Contoh pemecahannya disajikan pada Gambar 2.2

b. Fitur bertipe kategorikal

Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Contohnya adalah fitur 'cuaca' mempunyai 3 (tiga) nilai berbeda dan ini bisa mempunyai banyak kombinasi syarat pengujian pemecahan. Secara umum ada 2 (dua), yaitu pemecahan biner (*binary splitting*) dan pemecahan dua biner (*multi splitting*). Kombinasinya disajikan seperti pada gambar. Untuk pemecahan yang hanya membolehkan pemecahan biner, seperti algoritma cart maka akan memberikan kemungkinan jumlah kombinasi pemecahan sebanyak 2^{k-1} , dimana k adalah jumlah nilai berbeda dalam fitur tersebut. Seperti yang dijelaskan pada gambar 2.3 dan gambar 2.4.



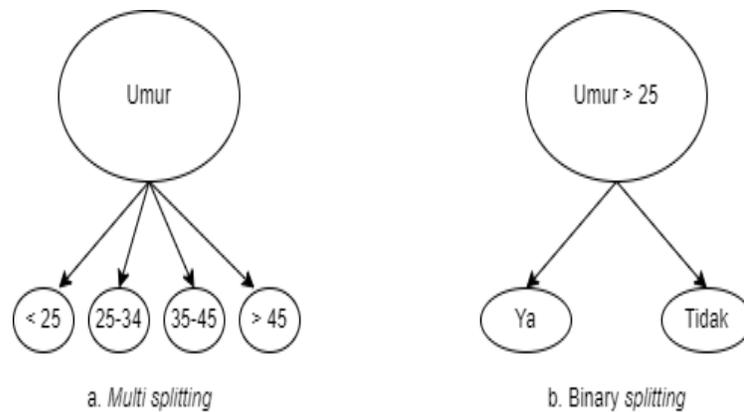
Gambar 2. 2 Syarat pengujian fitur bertipe nominal



Gambar 2. 3 Syarat pengujian fitur bertipe ordinal

c. Fitur bertipe numerik

Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun *internal*) dinyatakan dengan pengujian perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk $vi \leq A < vi+1$, 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. Seperti yang ada di gambar 2.4.



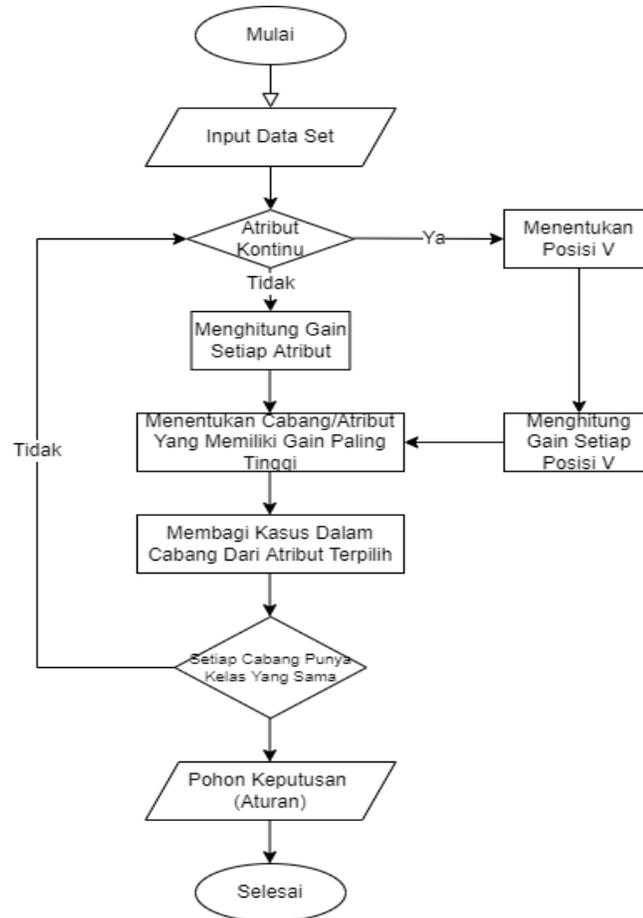
Gambar 2. 4 Syarat pengujian fitur bertipe numerik

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 .

Gambar 2.5 merupakan flowchart dan penjelasan algoritma C4.5 :



Gambar 2. 5 Flowchart Algoritma C4.5

(Sumber: Larose, 2005)

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

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

1. Pengertian 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 Kustiyaningsih (2011:114), “PHP (atau resminya PHP: *Hypertext Preprocessor*) adalah skrip bersifat *server – side* yang ditambahkan ke dalam HTML”. Teoritis terkait dengan objek permasalahan.

2. Pengertian Metode *Prototype*

Menurut (Ogedebe & Jacob, 2012) menyampaikan bahwa *prototyping* merupakan metode pengembangan perangkat lunak, yang berupa model fisik kerja sistem dan berfungsi sebagai versi awal dari sistem. Dengan metode *prototyping* ini akan dihasilkan prototipe sistem sebagai perantara pengembang dan pengguna agar dapat berinteraksi dalam proses kegiatan pengembangan sistem informasi. Agar proses pembuatan *prototype* ini berhasil dengan baik adalah dengan mendefinisikan kebutuhan awal. *Prototype* akan dihilangkan atau ditambahkan pada bagiannya sehingga sesuai dengan perencanaan dan analisis yang dilakukan oleh pengembang sampai dengan uji coba dilakukan secara simultan seiring dengan proses pengembangan.

Ada 4 (empat) metodologi *prototyping* yang paling utama yaitu :

- 1) ***Illustrative***, menghasilkan contoh laporan dan tampilan layar.
- 2) ***Simulated***, mensimulasikan beberapa alur kerja sistem tetapi tidak menggunakan data *real*.
- 3) ***Functional***, mensimulasikan beberapa alur sistem yang sebenarnya dan menggunakan data *real*.
- 4) ***Evolutionary***, menghasilkan model yang menjadi bagian dari operasional sistem.

Dibuatnya sebuah *prototyping* bagi pengembang sistem bertujuan untuk mengumpulkan informasi dari pengguna sehingga pengguna dapat berinteraksi dengan model prototipe yang dikembangkan, sebab *prototype* menggambarkan versi awal dari sistem untuk kelanjutan sistem sesungguhnya yang lebih besar.

Menurut(Ogedebe, 2012), menegaskan telah ditemukan bahwa dalam analisis dan desain sistem, terutama untuk proses transaksi, dimana dialog yang ditampilkan lebih mudah dipahami. Semakin besar interaksi antara komputer dan pengguna, besar pula manfaat yang diperoleh ketika proses pengembangan sistem informasi akan lebih cepat dan membuat pengguna akan lebih interaktif dalam proses pengembangannya.

Prototyping dapat diterapkan pada pengembangan sistem kecil maupun besar dengan harapan agar proses pengembangan dapat berjalan dengan baik, tertata serta dapat selesai tepat waktu. Keterlibatan pengguna secara penuh ketika prototipe terbentuk akan menguntungkan seluruh pihak yang terlibat, bagi pimpinan, pengguna sendiri serta pengembang sistem.

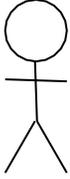
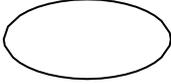
Manfaat lainnya dari penggunaan *prototyping* adalah :

- (a) Mewujudkan sistem sesungguhnya dalam sebuah replika sistem yang akan berjalan, menampung masukan dari pengguna untuk kesempurnaan sistem.
- (b) Pengguna akan lebih siap menerima setiap perubahan sistem yang berkembang sesuai dengan berjalannya *prototype* sampai dengan hasil akhir pengembangan yang akan berjalan nantinya.
- (c) Prototype dapat ditambah maupun dikurangi sesuai berjalannya proses pengembangan. Kemajuan tahap demi tahap dapat diikuti langsung oleh pengguna.
- (d) Penghematan sumber daya dan waktu dalam menghasilkan produk yang lebih baik dan tepat guna bagi pengguna.

3. *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.

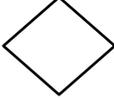
Tabel 2. 1 Simbol *use case diagram*

SIMBOL	NAMA	KETERANGAN
	<p><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 <i>actor</i> adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama <i>actor</i>.</p>
	<p><i>USE CASE</i></p>	<p>Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau <i>actor</i> biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama <i>use case</i>.</p>
	<p><i>ASOSIASI/ ASSOCIATION</i></p>	<p>Komunikasi antara <i>actor</i> dan <i>use case</i> yang berpartisipasi pada <i>use case</i> atau <i>use case</i> memiliki interaksi dengan <i>actor</i>.</p>

SIMBOL	NAMA	KETERANGAN
<<extend>> 	EKSTENSI/ EXTEND	Relasi <i>use case</i> tambahan ke <i>sebuah use case</i> dimana <i>use case</i> yang ditambahkan dapat berdiri sendiri walau tanpa <i>use case</i> tambahan memiliki nama depan yang sama dengan <i>use case</i> yang ditambahkan.
	GENERALISASI / GENERALIZATION	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah <i>use case</i> dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
<<include>> 	MENGGUNAKAN INCLUDE	Relasi <i>use case</i> tambahan ke <i>sebuah use case</i> untuk menjalankan fungsional

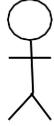
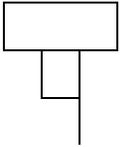
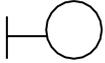
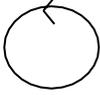
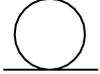
(Sumber: Dharwiyanti & Wahono, 2003)

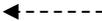
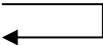
Tabel 2. 2 Simbol *activity diagram*

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: Dharwiyanti & Wahono, 2003)

Tabel 2. 3 Simbol *sequence diagram*

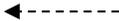
SIMBOL	NAMA	KETERANGAN
	<p><i>Actor</i></p>	<p>Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.</p>
	<p><i>Lifeline</i></p>	<p>Menghubungkan objek selama <i>sequence</i> (<i>message</i> dikirim atau diterima).</p>
	<p><i>General</i></p>	<p>Merepresentasikan entitas tunggal dalam <i>sequence</i>.</p>
	<p><i>Boundary</i></p>	<p>Berupa tepi dari sistem, seperti <i>user interface</i> dan alat yang berinteraksi dengan yang lain.</p>
	<p><i>Control</i></p>	<p>Elemen mengatur aliran dari informasi untuk sebuah skenario. Objek ini umumnya perilaku dan perilaku bisnis.</p>
	<p><i>Entitas</i></p>	<p>Elemen yang bertanggung jawab menyimpan atau informasi. Ini dapat berupa <i>beans</i> atau model <i>object</i>.</p>

SIMBOL	NAMA	KETERANGAN
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah <i>sequence</i> yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan/hubungan antar objek yang menunjukkan urutan kejadian
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman <i>message</i> 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 (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (<i>ancestor</i>).

SIMBOL	NAMA	KETERANGAN
	<i>Nary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari dua 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 <i>actor</i> .
	<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (<i>independent</i>) 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)

4. Pengertian Database

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

Menurut (Shalahuddin, 2015:43) menyatakan bahwa “basis data (*database*) merupakan salah satu bagian rekayasa perangkat lunak yang terkomputerisasi dan bertujuan utama sebagai media penyimpanan informasi agar dapat diakses dengan mudah dan cepat”.

Adapun kegunaan *database* yaitu sebagai alat untuk penyimpanan data yang saling berhubungan secara logikal dan deskripsi agar dapat diakses dengan lebih mudah dan cepat, dan terdapat tujuan utamanya yaitu untuk memelihara data yang sudah diolah dan membuat informasi tersedia pada saat dibutuhkan.

C. Tinjauan Pustaka

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) **(Angga riadi, Sundaru retno andani, Retno rafiqa dewi, 2021) Penerapan Data Mining Algoritma C4.5 menentukan Kelayakan Penerima Bantuan Covid 19** Dampak dari Pandemi COVID 19 dapat dilihat dari terpuruknya ekonomi masyarakat, khususnya pada masyarakat nagori Sitalasai tempat dilakukannya penelitian ini. Data dari Dinas Kelurahan menjelaskan bahwa hampir seluruh masyarakat merasakan sulitnya menjalankan roda perekonomian. Dalam mengatasi permasalahan ini pemerintah membuat program pemberian bantuan khusus bagi masyarakat yang layak. Untuk menentukankelayakan penerimaan bantuan Covid 19 ini dibutuhkan model atau aturan agar tepat pada sasaran. Saat ini pihak kelurahan hanya memiliki data yang bersumber dari Dinas Sosial dan tidak semua masyarakat masuk kedalamdaftar tersebut sehingga banyak masyarakat yang seharusnya layak menerima bantuan tidak masuk kedalam daftar penerima bantuan. Data mining dengan algoritma C4.5 merupakan salah satu metode didalam ilmu komputer untuk memecahkan permasalahan klasifikasi. Dengan algoritma C4.5 variabel untuk mengukur faktor-faktor

kelayakan penerima bantuan ini dikorelasikan untuk menghasilkan aturan. Hasil yang akan dicapai pada penelitian ini adalah aturan pohon keputusan kelayakan penerima bantuan Covid 19. Dengan metode pemecahan permasalahan ini menggunakan data mining algoritma C4.5 dapat mengklasifikasikan kelayakan penerima bantuan Covid19 sehingga data yang diolah valid dan tepat sasaran.

(2) (Weni Ratna Sari Oktapia ningse, Sumarno, Zulaini Masruro Nasution, 2022)

Klasifikasi Algoritma C4.5 untuk penentuan penerima program Indonesia pintar pada Mis Al-Khoirot. Tujuan penelitian adalah untuk membantu pihak sekolah dalam menyeleksi data siswa sebagai penerima PIP (program Indonesia pintar) menjadi lebih objektif dan praktis serta dapat membantu dalam meningkatkan ketepatan sasaran dari penerima bantuan dana PIP tersebut. Pada penelitian ini menggunakan teknik Data mining dengan menggunakan algoritma C4.5 Sumber data penelitian yang digunakan diperoleh dari observasi dan wawancara kepada pihak sekolah MIS Al-khoirot Tambun Nabolon Pematang Siantar. Adapun variabel penelitian yang digunakan adalah pekerjaan orang tua, penghasilan orang tua, pemegang KKS, pemegang SKTM, pada penelitian ini alternatif yang digunakan sebagai sampel adalah data siswa hasil dari penelitian ini didapatkan bahwa atribut yang paling dominan adalah pemegang SKTM dengan gain sebesar 0,833764907.

(3) (puja Restu Adinda, 2022) Pohon Keputusan C4.5 Algoritma Untuk

Klasifikasi Program Bantuan Belajar. Di era pandemi, hampir semua orang berjuang untuk hidupnya. Mahasiswa adalah contohnya. Mereka kesulitan membayar uang sekolah untuk melanjutkan studi. Berdasarkan situasi problematis ini, Universitas Teknokrat Indonesia menghibahkan mahasiswa yang memiliki prestasi akademik baik dengan program bantuan biaya kuliah. Banyaknya fitur yang digunakan untuk menentukan hibah membuat sulitnya mengambil keputusan dalam waktu singkat atau bahkan membutuhkan waktu yang sangat lama. Untuk memudahkan manajemen dalam menentukan siapa mahasiswa yang berhak mendapatkan hibah, maka diperlukan model klasifikasi. Tujuan dari penelitian ini adalah untuk mengklasifikasikan penerima hibah dengan menggunakan algoritma pohon keputusan C4.5. Algoritma C4.5 digunakan karena dapat langsung menampilkan pola data dalam bentuk pohon keputusan yang dapat menentukan apakah seorang calon mahasiswa dapat diterima sebagai penerima beasiswa atau tidak.

- (4) (Firamon Syakti,pahmi Ladianto,2022) Implementasi Data Mining Untuk Bantuan Dana Bansos Dengan Menggunakan Algoritma C4.5 Di pemerintahan Kabupaten EmpatLawang.** Pemerintah Kabupaten Empat Lawang dalam memajukan kesejahteraan masyarakatnya mempunyai program berupa bantuan social. Pemerintah Kabupaten Empat Lawang dalam menyalurkan bantuan social berupa pembedahan rumah yang tak layak huni. Karena pada Kabupaten Empat Lawang masih banyak masyarakat dibawah garis kelayakan, tak terkecuali tempat tinggal atau rumah yang tak layak huni. Sedangkan proses untuk mendapatkan bedah rumah tersebut diawali dari pemerintah paling rendah yaitu Kepala desa mendata warganya yang mempunyai rumah yang tidak layak huni serta ekonomi pendapatan perhar dalam memenuhi kebutuhan hidup, kemudian data tersebut diajukan ke Pemerintah Kabupaten Empat Lawang. Tim Verifikasi yang ditunjuk Pemerintah Kabupaten Empat Lawang melakukan seleksi untuk menentukan siapa yang akan memperoleh dana bantuan social tersebut. Pada penelitian ini untuk mengumpulkan data yang diperlukan penulis akan menggunakan data yang telah dikumpulkan oleh pihak Pemerintah Kabupaten Empat Lawang mengenai objek yang akan menerima bantuan, jika sudah melakukan observasi, penulis akan memasukkan data tersebut pada kategori dan regulasi sesuai ketentuan yang ditetapkan diperaturan Menteri PUPR Nomor 13/PRT /M / 2016 . Berikutnya untuk menyaring kembali data yang sudah diperoleh penulis akan menggunakan System. System yang akan dibangun berupa system berbasis web yang akan dipadukan dengan Algoritma C4.5
- (5) (Nabilah iftah nella,Nanang Yudi Setiawan,Dian Eka Ratnawati,2022) Klasifikasi penerima Bantuan program keluarga Harapan menggunakan Algoritma Decision Tree C4.5.** Bantuan sosial merupakan pemberian bantuan berupa uang atau barang yang didistribusikan kepada individu atau kelompok tertentu. Sebagai bentuk tanggung jawab dalam mengentaskan kemiskinan pemerintah mengalokasikan biaya untuk menangani hal tersebut salah satunya adalah Program Keluarga Harapan (PKH). PKH adalah program pemberian bantuan tunai kepada rumah tangga miskin. Salah satu desa yang telah mendapatkan bantuan PKH dari pemerintah adalah desa Mlirip, Kabupaten Mojokerto. Dengan adanya bantuan PKH warga dengan ekonomi rendah merasa sangat terbantu. Tetapi sangat disayangkan, hingga saat ini PKH dianggap masih belum optimal karena banyak bantuan yang tidak tepat sasaran dan kantor kepala desa mlirip belum memiliki acuan dalam memilih keluarga yang benar-benar

layak. Klasifikasi dilakukan untuk mengetahui kelayakan penerima bantuan PKH agar tidak salah sasaran..

(6)(Rika Nur Adiha, Sundari Retno Andani, Widodo Saputra, 2022) APPLICATION OF DATA MINING IN DETERMINING SOCIAL ASSISTANCE RECIPIENTS WITH C4.5 ALGORITHM. *The Gunung Maligas District Office is a government agency tasked with running a government program, namely the Social Assistance Receipt program, to run the social assistance program, many residents complain that they do not receive assistance, while some residents who are considered capable actually get assistance, where each aid program is have different criteria in determining the recipient. Due to the large number of existing aid programs with different criteria in determining the acceptance of the aid program, of course, local government staff will have difficulty in conducting the selection process. So we need a system that is able to help local government staff to more easily determine the recipients of the social assistance. Based on the historical data of beneficiaries, recommendations for the classification of beneficiaries can be made that will assist government staff. Classification can be done using the C4.5 algorithm. In this study, it has parameters, namely, occupation, income, housing conditions and number of dependents. By applying the C4.5 data mining algorithm, it is hoped that it will make it easier and faster for government staff to determine the recipients of social assistance at the Gunung Maligas District Office.*

(7)(Fristi Riandari and Sarjon Defit, 2021) THE APPLICATION OF C4.5 ALGORITHM FOR SELECTING SCHOLARSHIP RECIPIENTS. *The scholarship program is one of the promotional techniques used by many universities, and the right scholarship award will certainly be an attraction for many people. STMIK Pelita Nusantara is one of the universities that organizes a scholarship program. In the current difficult economic conditions, the scholarship program is the target of many prospective students who want to continue their education in higher education. However, the absence of tools to process large amounts of data make determining scholarship recipients less effective and time-consuming. This situation is seen by the fact that some students are still unable to maintain the scholarships they receive. In the research, a classification model was proposed using the C4.5 algorithm approach by utilizing past data to facilitate the decision making of the scholarship program. This classification process produced a decision tree that could be used as a decision-making tool. Scholarships were awarded based on several criteria: academic potential, vocational potential, parents' income,*

number of dependents, and employment status. Based on the data processing results of students who apply for scholarships in 2020 with predetermined criteria, the highest root is obtained. It consists of node 1 for academic potential, node 1.1 for vocational potential, and node 1.2 for parental income. The resulting decision tree model is expected to help to make decisions quickly and on target.

(8)(Carissa Elma Purnomo,Rikendry,2021) penerapan metode algoritma C4.5 untuk klasifikasi warga miskin pada desa mengandung sari.Desa mengandung sari,Kecamatan sekampung udik,kabupaten lampung timur merupakan desa yang berasal dari nama “mehganung wo gannung”pemekaran dari desa toba.pada tanggal 01 januari 1961 oleh kepala kampung tuba yang bernama abdul gani gelar kerio pendetta mego,beserta kepala dusun minak braji,mereka mulai merintis atau membuka wilayah ini dengan nama desa”mengandung sari”.Desa mengandung sari adalah salah satu desa yang berada dikecamatan sekampung udik desa lampung ini memiliki 6 susun didalamnya desa mengandung sari diwajibkan untuk mengikuti dalam program dari kementerian sosial yang diberikan sebulan sekali dan lain-lain program ini diperuntukan bagi keluarga miskin sebagai upaya pemerintah dalam menanggulangi kemiskinan masalah baru muncul yaitu bagaimana satuan lingkungan setempat (SLS) menetapkan kepala keluarga yang benar-benar berhak menerima bantuan tersebut dengan adanya algoritma c4.5,data warga desa mengandung sari dapat diklasifikasikan berdesarkan “miskin” dan “tidak miskin” .

(9)(Mohammad Ahmad Maidanul Abrori, Abdul Syukur, Affandy, Moch Arief Soeleman, 2022) Improving C4.5 Algorithm Accuracy With Adaptive Boosting Method For Predicting Students in Obtaining Education Funding. *The level of accuracy in determining the prediction of the provision of educational funding assistance is very important for the education agency. The large number of data on prospective beneficiaries can be processed into information that can be used as decision support in determining eligibility for education funding assistance. The data processing is included in the field of data mining. One method that can be applied in predicting the feasibility of receiving aid funds is classification. There are several classification algorithms, one of which is a decision tree. The famous decision tree algorithm is C4.5. The C4.5 algorithm can be applied in classifying prospective recipients of educational aid funds. This study uses datasets from student data of SMK Al Fattah Kertosono. The purpose of this study is to increase the accuracy of the C4.5 algorithm by applying adaboost in classifying*

students who deserve education funding and not, by comparing the results before and after applying adaboost. Validation in this study uses cross validation.

(10) (Asrina Yuni Ikhsanti, Yuli Fauziah, Rifki Indra Perwira, 2021) **Implementation of the c4.5 decision tree learning algorithm for sentiment analysis in e-commerce application reviews on google play store.** Forknowing the level of accuracy of the C4.5 Decision Tree Learning Algorithm in sentiment analysis of reviews of e-commerce applications on the google play store. Design/method/approach: Using C4.5 Algorithm Decision Tree Learning.

Tabel 2. 5 Tinjauan studi

NO	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
1	Penerapan Data Mining Algoritma C4.5 menentukan Kelayakan Penerima Bantuan Covid 19	Angga riadi, Sundaru retno andani, Retno rafiqa dewi	JUKI : Jurnal Komputer dan Informatika Volume 3 Nomor 2 November 2021 e-ISSN : 2722-4368 https://www.ioinformatic.org/index.php/JUKI/article/view/60	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
2	Klasifikasi Algoritma C4.5 untuk penentuan penerima program Indonesia pintar	Weni Ratna oktapia Ningse, Sumarno, Zulaini Masruro Nasution, 2022	Journal of Machine Learning and Artificial Intelligence Vol. 1, No. 1, Maret 2022, pp. 65~76	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

NO	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
	pada MIS-AL-KHOIROT		file:///C:/Users/LENOVO/Downloads/65-76+JOMLAI+WE NI+RATNA+SA RI+OKTAPIA+NINGSE%20(1).pdf	
3	Pohon Keputusan C4.5 Algoritma Untuk Klasifikasi Program Bantuan Belajar.	puja Restu Adinda,2022	Portaldata.org Volume 2 (9), 2022 http://portaldata.org/index.php/portaldata/article/view/215/204	. Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
4	Implementasi Data Mining Untuk Bantuan Dana Bansos Dengan Menggunakan Algoritma C4.5 Di pemerintahan Kabupaten EmpatLawang.	Firamon Syakti,pahmi Ladianto,2022	JurnalJupiter,Vol .14No.2BulanOktober,Tahun2022,Hal.287-296 https://jurnal.polisri.ac.id/index.php/jupiter/article/view/5075	. Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

NO	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
5	Klasifikasi penerima bantuan program keluarga Harapan menggunakan Algoritma <i>Decission Tree</i> C4.5	Nabilah iftahnella,Nanang Yudi Setiawan,Di an Eka Ratnawati,2022	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 6, No. 3, Maret 2022, hlm. 1332-1339. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10801	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
6	<i>The Application Of C4,5 Algorithm For Selecting Scholarship Recipients</i>	Fristi Riandari and Sarjon Defit	ComTech: Computer, Mathematics and Engineering Applications, 13(1), 11-21. https://doi.org/10.21512/comtech.v13i1.7307	Penggunaan perhitungan menggunakan Alogritma C4.5 dijadikan dalam metode dalam penelitian
7	<i>Application Of Data Mining Social Determining Social Assistance</i>	Rika Nur Adiha, Sundari Retno Andani, Widodo Saputra	Journal of Artificial Intelligence and Engineering Applications 15th February	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian

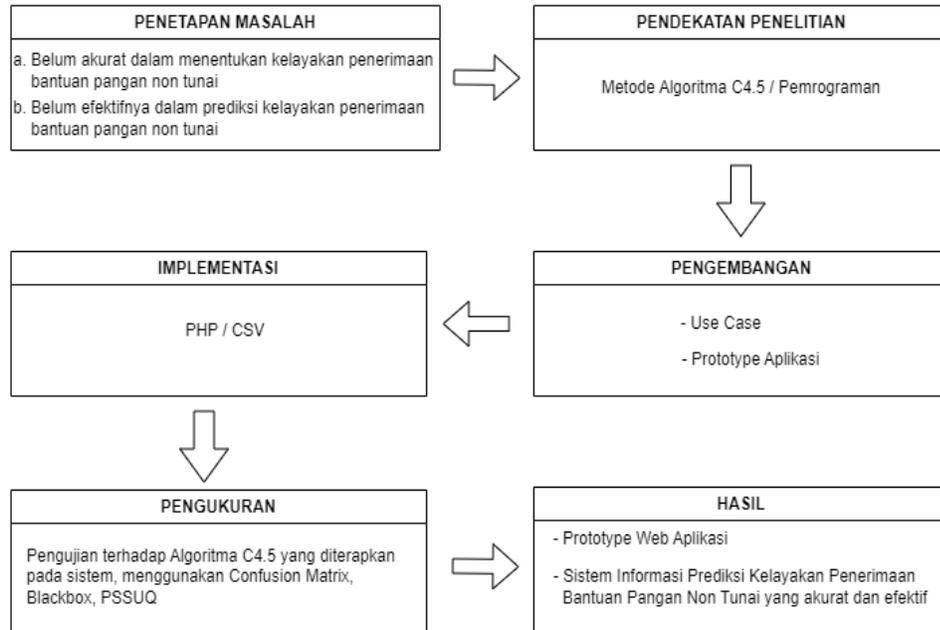
NO	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
	<i>Recipients With C4.5 Algorithm</i>		2022. Vol. 1. No. 2 https://ioinformat ic.org/	
8	Penerapan metode C 4.5 untuk klasifikasi warga miskin pada desa mengandung sari	Carissa Elma Purnomo, Ri kendry	Jurnal teknologi dan sistem informasi(JTSI) Vol.2,No.3,september 2021,14-25 E-ISSN:2746-3699 file:///C:/Users/L ENOVO/Downloads/936-2807-1-PB.pdf	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian
9	<i>Improving C4.5 Algorithm Accuracy With Adaptive Boosting Method For Predicting Students in Obtaining Education Funding.</i>	Mohammad Ahmad Maidanul Abrori, Abdul Syukur, Affandy, Moch Arief Soeleman	Available online at JDR Website: http://journal.unublitar.ac.id/jdr Journal Of Development Research, 6 (2), November 2022, Pages 137-140 DOI:	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian

NO	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
			https://doi.org/10.28926/jdr.v6i2.205	
10	<i>Implementation of the c4.5 decision tree learning algorithm for sentiment analysis in e-commerce application reviews on google play store</i>	Asrina Yuni Ikhsanti, Yuli Fauziah, Rifki Indra Perwira	Computing and Information Processing Letters Vol. 1., No. 1, November 2021, pp. 25-30 http://jurnal.upnyk.ac.id/index.php/cip/index	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian

Berdasarkan uraian rujukan penelitian yang sudah dilakukan sebelumnya dapat disimpulkan kekurangan yaitu terdapat jurnal yang hanya menggunakan aplikasi yang sudah tersedia, dan beberapa jurnal tidak melakukan perhitungan tingkat akurasi, terdapat juga kesamaan metode yang digunakan yaitu Algoritma C4.5, menggunakan confusion matrix sebagai uji hasilnya. Pada penelitian ini yaitu melakukan pengembangan aplikasi serta menghitung tingkat akurasinya. Selain itu pada penggunaan variabel pada penelitian kali ini menggunakan variabel Nama Penerima Bantuan, Penghasilan, Jumlah Tanggungan jenis kloset, Jenis Lantai, Jenis Dinding, Sumber Air, jenis kloset, Listrik, memiliki kendaraan roda 2, cara memperoleh air minum. Sehingga hasil prediksi dalam penelitian ini akan menghasilkan akurasi prediksi yang lebih baik.

D. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang dapat dilihat pada Gambar 2.6.



Gambar 2. 6 Kerangka pemikiran

Penelitian ini diawali dengan penetapan masalah terkait pengumpulan data yang didapatkan melalui hasil observasi dan wawancara yang dilakukan pada objek permasalahan.

(1) Penetapan Masalah

Pernyataan masalah yang ada adalah belum akurat dalam menentukan kelayakan penerimaan bantuan pangan non tunai

(2) Pendekatan Penelitian

Pendekatan penelitian ini yaitu, untuk menentukan kelayakan penerimaan bantuan pangan non tunai. Metode penelitian ini menggunakan algoritma C4.5 menjadi dasar analisis data hasil penelitian.

(3) Pengembangan

Pengembangan yang digunakan dalam pembangunan system menggunakan *Use Case* dan Aplikasi *Prototype*.

(4) Implementasi

Implementasi sebagai dasar tahap evaluasi dan validasi system *prototype* yang akan dibuat menggunakan PHP / CSV.

(5) Pengukuran

Pengukuran pengujian terhadap algoritma C4.5 yang diterapkan pada *system* menggunakan *Confusion Matrix* sebagai uji metode, *Blackbox* untuk uji ahli, dan PSSUQ untuk uji kebergunaan.

(6) Hasil

Hasil yang di targetkan dari penelitian ini yaitu menghasilkan *prototype* sistem informasi berbasis *website* khusus sebagai informasi penerimaan kelayakan bantuan pangan non tunai.

E. Hipotesis

Algoritma C4.5 adalah algoritma klasifikasi pohon keputusan dan dapat mengimplementasikan metode algoritma C4.5 yang digunakan dengan memasukan data sampel dan menghitung secara bertahap berdasarkan kelas. Berdasarkan permasalahan yang dihadapi yaitu belum efektif dalam menentukan kelayakan penerimaan bantuan pangan non tunai maka perlu adanya suatu cara untuk mengatasi atau setidaknya mengurangi permasalahan tersebut, salah satu keunggulan metode ini yaitu penentuan kelayakan penerimaan bantuan berdasarkan data label yang ditentukan dan yang akan mengetahui kelayakan seorang penerimaan bantuan pangan non tunai. ada beberapa metode yang dapat dilakukan yaitu mengklasifikasi untuk melakukan prediksi suatu peluang yaitu salah satunya dengan metode algoritma C4.5, Teori klasifikasi yang digunakan yaitu algoritma C4.5. Algoritma C4.5 juga sudah banyak dikenal dan digunakan oleh penelitian sebelumnya oleh (Weni, sumarno, & Zulaini, 2022) maka dapat diduga bahwa hipotesis pada penelitian ini adalah dalam memprediksi penerimaan bantuan pangan non tunai menggunakan algoritma C4.5.