

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Pengertian *Data Mining*

Menurut (Pramudiono, 2006, P.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 ekstraksi agar menjadi informasi yang baru dan dapat membantu dalam pengambilan keputusan. (Suntoro, 2019, P. 6). Ada beberapa tahapan dalam data mining, yaitu (Meilani & Susanti, 2014, P.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 performasi 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.

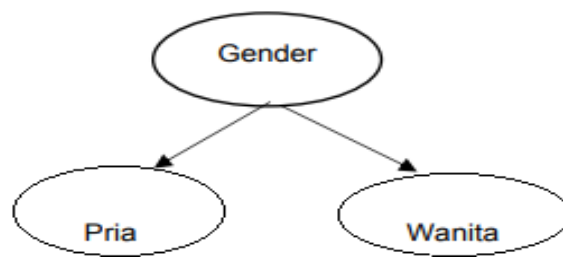
2. Pengertian Klasifikasi

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, P.34). 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, P.34) proses klasifikasi ada empat komponen (Gorunescu, 2011, P.21):
 - 1) *Class*, variabel dependen yang 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.

3. 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, P.12). Seperti dijelaskan pada gambar 2.1



Gambar 2.1 Syarat Pengujian Fitur Biner

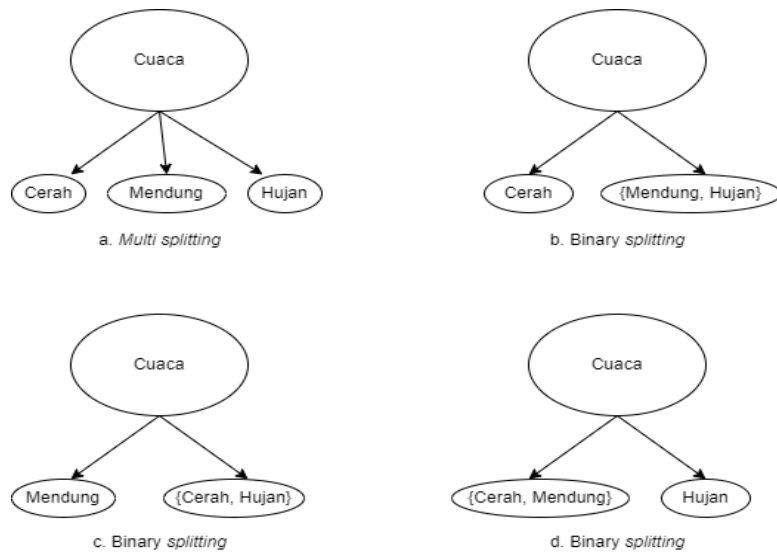
Yang menjadi hal penting dalam induksi *Decision Tree* adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 (tiga) kelompok penting dalam syarat pengujian node:

(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.

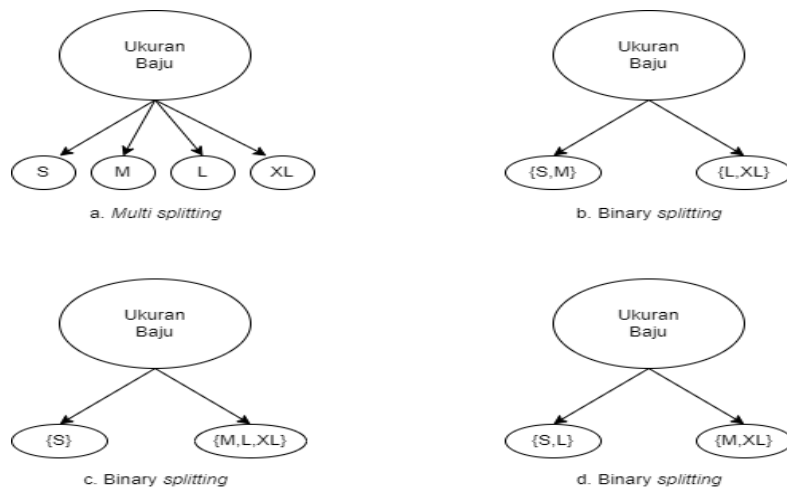
(b) Fitur bertipe Kategorikal

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.



Gambar 2.2 Syarat Pengujian Fitur Biner Bertipe Nominal

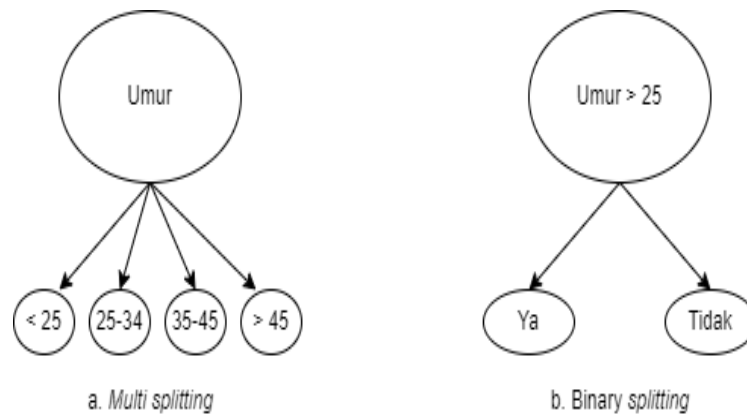
berbeda dalam fitur tersebut. Seperti yang dijelaskan pada gambar 2.2 dan gambar 2.3.



Gambar 2.3 Syarat Pengujian Fitur Biner 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.

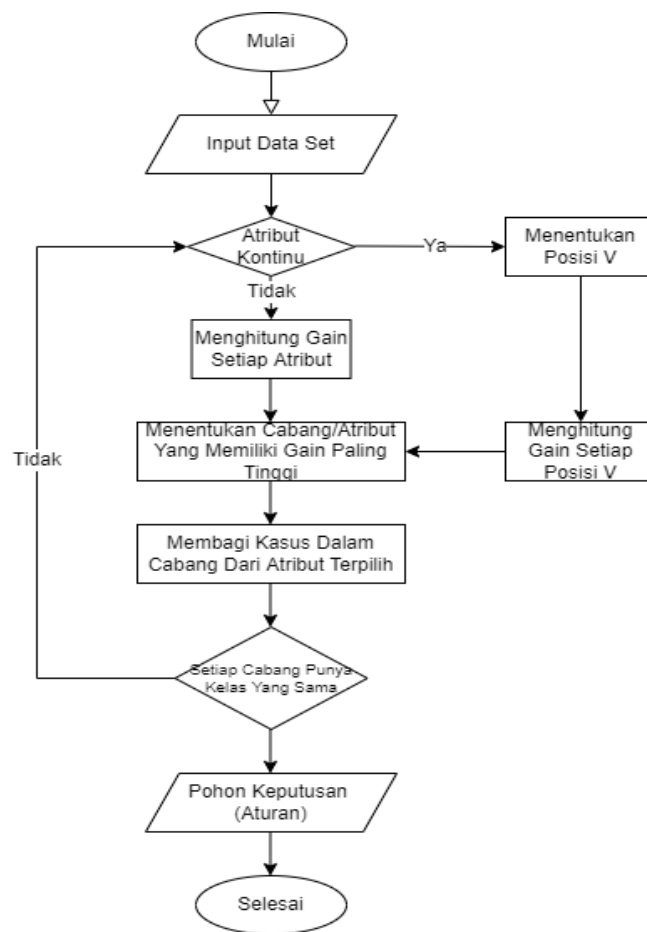


Gambar 2.4 Syarat Pengujian Fitur Biner Bertipe Numerik

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, P.12)

Berikut penjelasan dari gambar 2.5:

- 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.
- 7) Atribut kontinu.

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

5. Pengertian Metode *Prototype*

Menurut (Ogedebe & Jacob, 2012, P.14) 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 prototipe menggambarkan versi awal dari sistem untuk kelanjutan sistem sesungguhnya yang lebih besar.

Menurut (Ogedebe, 2012, P.17), 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.

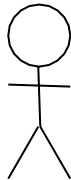
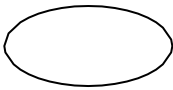
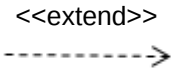
Manfaat lainnya dari penggunaan *prototyping* adalah :

- 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 prototipe 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.

6. *Unified Modeling Language (UML)*

Menurut (Dharwiyanti & Wahono, 2003, P.13) 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
	ACTOR	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri, jadi walaupun simbol dari actor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama actor.
	USE CASE	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama use case.
	ASOSIASI / ASSOCIATION	Komunikasi antara actor dan use case yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.
	EKSTENSI/ EXTEND	Relasi use case tambahan ke sebuah use case dimana use case yang ditambahkan dapat berdiri sendiri walau tanpa use case tambahan memiliki nama depan yang sama dengan use case yang ditambahkan.
	GENERALISASI / GENERALIZATION	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.

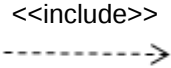
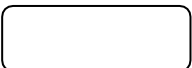
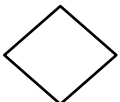
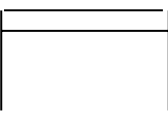
SIMBOL	NAMA	KETERANGAN
	MENGGUNAKAN INCLUDE	Relasi <i>use case</i> tambahan ke sebuah <i>use case</i> untuk menjalankan fungsional

Diagram use case merupakan suatu pemodelan yang bisa menggambarkan perilaku atau kebiasaan dari sistem yang ingin dibuat

(Sumber: Dharwiyanti & Wahono, 2003)

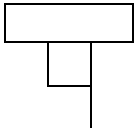
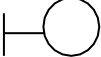
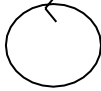
Tabel 2.2 Simbol Activity Diagram

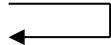
SIMBOL	NAMA	KETERANGAN
	STATUS AWAL / INITIAL	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	AKTIVITAS / ACTIVITY	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	PERCABANGAN / DECISION	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	PENGGABUNGAN/ JOIN	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	STATUS AKHIR/FINAL	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status baru.
	SWIMLINE	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

Runtutan proses dari suatu sistem digambarkan secara vertikal. *Activity* diagram merupakan pengembangan dari *Use Case* yang memiliki alur aktivitas.

(Sumber: Dharwiyanti & Wahono, 2003)

Tabel 2.3 Simbol Sequence Diagram

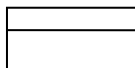
SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>Lifeline</i>	Menghubungkan objek selama <i>sequence</i> (<i>message</i> dikirim atau diterima).
	<i>General</i>	Merepresentasikan entitas tunggal dalam <i>sequence</i> .
	<i>Boundary</i>	Berupa tepi dari sistem, seperti <i>userinterface</i> 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 <i>beans</i> atau model <i>object</i> .
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah <i>sequence</i> yang menunjukkan sebuah objek mengirim atau menerima objek.

SIMBOL	NAMA	KETERANGAN
	<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.

Sequence diagram merupakan cara atau skenario untuk memvisualisasikan dan melakukan validasi berbagai skenario teknis secara runtime.

(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>).
	<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> .

SIMBOL	NAMA	KETERANGAN
←-----	<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 objek lainnya.

Class diagram adalah visual dari struktur sistem program pada jenis-jenis yang dibentuk. *Class diagram* merupakan alur jalannya database pada sebuah sistem. *System diagram* tool diagram design software.

(Sumber: Dharwiyanti & Wahono, 2003)

7. Pengertian Database

Menurut (Mustakini, 2009, P.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, P.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.

1. Webserver (Apache)

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

(a) *Web server* merupakan mesin dimana tempat aplikasi atau *software* beroperasi dalam mendistribusikan *web page* ke *user*, tentu saja sesuai

dengan permintaan *user*;

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

B. Tinjauan 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. **(Nurhadi Wijaya, Marselina, Mujiatia Feliati 2020) Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi data status huni rumah rehabilitasi pasca erupsi merapi Erupsi** gunung Merapi berikut lahar hujan di Tahun 2010 berdampak pada kerusakan infrastruktur berikut ribuan hunian rumah di Kabupaten Sleman D.I. Yogyakarta dan Kabupaten Magelang Jawa Tengah. Melalui Peraturan Kepala BNPB No.5 Tahun 2011, rehabilitasi dan rekonstruksi perumahan yang terdampak erupsi Merapi, dilakukan dengan skema program Rehabilitasi dan Rekonstruksi Masyarakat dan Permukiman Berbasis Komunitas. Skema tersebut telah membangun rumah hunian sebanyak 2.516-unit. Berdasarkan *Key Performance Indikator* (KPI) oleh *The World Bank*, status huni rumah merupakan indikator keberhasilan kinerja skema program ini. Pelaksanaan program rehabilitasi rumah pasca erupsi Merapi didokumentasikan dan terekam ke dalam basis data. Dibidang *data mining*, basis data merupakan aset yang dapat digunakan sebagai bahan pengenalan dan penemuan pola-pola data yang dapat dipelajari dan diteliti guna menyelesaikan permasalahan baik pengelompokan data maupun klasifikasi data. Pada penelitian ini dilakukan penerapan algoritma decision tree C.45 untuk mengklasifikasi data status huni rumah rehabilitasi pasca erupsi gunung Merapi. Hasil klasifikasi penelitian diperoleh angka nilai tingkat akurasi klasifikasi mencapai 91.34%, dengan demikian terjawab bahwa algoritma *decision tree* C.45 dapat diterapkan untuk mengklasifikasi data status huni rumah rehabilitasi pasca erupsi gunung Merapi.

2. **(Bertha V.D.S Pinto, Emerensye S.Y. Pandie, Tiwuk Widyastuti 2023) Penerapan Algoritma C4.5 untuk menentukan prioritas penerima Bantuan Sosial Rumah Tidak Layak Huni (RTLH)** Bantuan sosial Rehabilitasi Rumah Tidak Layak Huni (RTLH) adalah bantuan yang diberikan kepada pemilik atau penghuni rumah yang kondisinya tidak memenuhi standar keamanan, kelayakan, atau kualitas hunian yang layak. Pada penelitian ini peneliti mengimplementasikan algoritma C4.5 dengan menggunakan *Rapid Miner* 10.1 untuk mengklasifikasikan kelayakan penerima bantuan sosial berdasarkan kriteria-kriteria yang ditentukan oleh Kementerian Sosial Republik Indonesia. Data dalam penelitian ini diperoleh melalui hasil wawancara berupa data penerima bantuan sosial RTLH pada tahun 2018-2021 sebanyak 2000 data mentah, setelah di *preprocessing* diperoleh data sebanyak 1850 data.
3. **(Puja Restu Adinda, 2022) Pohon Keputusan C4.5 Algoritma Untuk Klasifikasi Program Bantuan Belajar.** Di era pandemi, hampir semua orang berjuang untuk hidupnya. Mahawarga 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 Empat Lawang.** Pemerintah Kabupaten Empat Lawang dalam memajukan kesejahteraan masyarakatnya mempunyai program berupa bantuan sosial. Pemerintah Kabupaten Empat Lawang dalam menyalurkan bantuan sosial 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 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. (Fristi Riandari And Sarjon Defit, 2021) Penerapan Algoritma C4.5 untuk Seleksi Penerima Beasiswa.

Program beasiswa merupakan salah satu teknik promosi yang digunakan oleh banyak universitas, dan pemberian beasiswa yang tepat tentu akan menjadi daya tarik bagi banyak orang. STMIK Pelita Nusantara adalah salah satu universitas yang menyelenggarakan program beasiswa. Dalam kondisi ekonomi yang sulit saat ini, program beasiswa menjadi target banyak calon mahasiswa yang ingin melanjutkan pendidikan ke jenjang yang lebih tinggi. Namun, tidak tersedianya alat untuk memproses data dalam jumlah besar membuat penentuan penerima beasiswa menjadi kurang efektif dan memakan banyak waktu. Situasi ini semakin diperparah dengan kenyataan bahwa beberapa mahasiswa masih belum mampu mempertahankan beasiswa yang mereka terima. Dalam penelitian ini, sebuah model klasifikasi diusulkan dengan menggunakan pendekatan algoritma C4.5 melalui pemanfaatan data masa lalu untuk memfasilitasi proses pengambilan keputusan dalam program beasiswa. Proses klasifikasi ini menghasilkan pohon keputusan yang dapat digunakan sebagai alat pengambilan keputusan. Beasiswa diberikan berdasarkan beberapa kriteria: potensi akademik, potensi vokasional, pendapatan orang tua, jumlah tanggungan, dan status pekerjaan. Berdasarkan hasil pengolahan data mahasiswa yang mendaftar beasiswa pada tahun 2020 dengan kriteria yang telah ditentukan, diperoleh akar tertinggi. Akar tersebut terdiri dari simpul 1 untuk potensi akademik, simpul 1.1 untuk potensi vokasional, dan simpul 1.2 untuk pendapatan orang tua. Pohon keputusan yang dihasilkan diharapkan dapat membantu membuat keputusan dengan cepat dan tepat sasaran.

6. (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 di kecamatan 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 berdasarkan "miskin" dan "tidak miskin".

7. (Lukman Bachtiar, Mahradianur 2023) Analisis Data Mining Menggunakan Metode Algoritma Untuk Menentukan Penerima Bantuan Langsung Tunai

Bantuan Langsung Tunai (BLT) adalah bantuan yang dibuat oleh pemerintah untuk orang atau keluarga yang kurang mampu dan rentan akan kemiskinan dengan mereka yang memenuhi persyaratan yang ada pada program. Tujuan penelitian ini untuk memberikan masukan tentang menentukan apakah keluarga tersebut masih berhak menerima atau tidak berhak menerima bantuan BLT, dimana masih ada keluarga yang berhak menerima dan memenuhi syarat dari program bantuan pemerintahan ini. Sumber data yang diperoleh dapat dari Kantor Desa Regei Lestari. Metode yang dipakai pada penelitian merupakan teknik data mining dengan algoritma C4.5 yang diterapkan pada software RapidMiner. Atribut yang dipakai untuk menentukan keluarga yang berhak menerima dan tidak berhak menerima pada program bantuan BLT yaitu Penghasilan, Status, Pekerjaan, Umur. Hasil dari klasifikasi data menggunakan metode algoritma C4.5 dan pengujian dengan RapidMiner adalah Jika Pekerjaan = buruh tani maka hasilnya adalah menerima dengan nilai 16 item dari menerima dan 0 yang tidak menerima, maka didapatkan faktor yang menentukan yang berhak menerima dan yang tidak berhak menerima bantuan langsung tunai (BLT).

8. (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 menentukan kelayakan 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 ke dalam daftar 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.

9. (Rizal Amegia Saputra, Sri Wasiyanti, Denny Pribadi 2021) Masalah kemiskinan diberbagai belahan dunia khususnya indonesia menjadi persoalan serius yang menjadi pusat perhatian pemerintah saat ini. Berbagai cara telah dilakukan oleh pemerintah untuk menekan angka kemiskinan meningkat, salah satunya memberikan bantuan kepada masyarakat miskin melalui perogram Bantuan Pangan Non Tunai (BPNT). Namun pemangku kepentingan memiliki permasalahan yang terjadi dilapangan yaitu sulitnya mengklasifikasikan data yang menerima bantuan sehingganya mengakibatkan kurang tepatnya sasaran dalam memberikan bantuan. Dalam penelitian ini menerapkan algoritma C4.5 dengan split kriteria information gain, algoritma C4.5 adalah salahsatu algoritma terbaik pada algoritma klasifikasi, dan dataset yang didapat pada tahun 2018 pada Desa Nagrak Utara Sukabumi sebanyak 130 data. Model pohon keputusan yang dihasilkan dari penerapan algoritma C4.5 yaitu menghasilkan akurasi sebesar 91.54% dengan nilai AUC 0.986, dengan demikian algoritma C4.5 untuk klasifikasi penentuan penerimaan BPNT itu termasuk algoritma Sangat Baik, karena range AUC diantara 0.90-1.00.

10.(Winda Lidiyarti, Heru Satria Tambunan, Hendry Qurimawan 2022) Penerapan Data Mining dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko dengan Algoritma C4.5 Kasus Kantor Kelurahan Martoba Sosial Bantuan sosial, yang biasa dikenal sebagai Bansos, adalah program pemerintah yang dibebankan pada Anggaran Pendapatan dan Belanja Daerah. Berdasarkan ketentuan umum Pasal 15 Peraturan Menteri Nomor 32 Tahun 2011 tentang Pemberian Subsidi dan Bantuan Sosial dalam Anggaran Pendapatan dan Belanja Daerah, definisi bantuan sosial berlaku bagi individu, keluarga, kelompok masyarakat yang tidak berkelanjutan, dan bersifat selektif untuk mencegah kemungkinan risiko sosial. Pemerintah Kota Pematangsiantar menyediakan berbagai jenis program bantuan, salah satunya adalah Bantuan Sosial PEMKO yang disalurkan melalui Kantor Kelurahan Martoba.

Dalam proses seleksi untuk menentukan penerima Bantuan Sosial PEMKO di Kantor Kelurahan Martoba, masih belum sepenuhnya memanfaatkan teknologi informasi untuk mendukung kinerja pegawai. Oleh karena itu, terdapat hambatan dan prosesnya memakan waktu lama. Maka dari itu, dibutuhkan sebuah sistem yang dapat membantu pegawai dengan mudah menentukan penerima bantuan. Penerapan Data Mining adalah serangkaian proses untuk menambahkan nilai dengan cara semi-manual dalam bentuk pengetahuan yang sebelumnya tidak diketahui dari data itu sendiri. Dalam penelitian ini, digunakan parameter seperti jumlah tanggungan, pekerjaan, pendapatan, dan status rumah. Dengan menerapkan Algoritma Data Mining C4.5, diharapkan dapat mempermudah dan mempercepat pegawai dalam menentukan penerima Bantuan Sosial PEMKO di Kantor Kelurahan Martoba.

Tabel 2.5 Tinjauan Pustaka

No	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
1.	Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi data status huni rumah rehabilitasi pasca erupsi merapi Erupsi	Nurhadi Wijaya, Marselina, Mujiatna, Feliati 2020	Jurnal PROSIDING SEMINAR NASIONAL ISSN :2567 -2397 https://prosiding.respati.ac.id/index.php/PSN/article/view/309/301	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian
2.	Penerapan Algoritma C4.5 untuk menentukan prioritas penerima Bantuan Sosial Rumah Tidak Layak Huni (RTLH)	Bertha V.D.S Pinto, Emerensye S.Y. Pandie, Tiwuk Widyastuti 2023	Jurnal Inovatif Wira Wacana Vol 2 No 1 (2023) https://ojs.unkriswina.ac.id/index.php/inovatif/article/view/626	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian
3	Pohon Keputusan C4.5 Algoritma Untuk Klasifikasi Program Bantuan Belajar.	puja Restu Adinda, 2022	Portalddata.org Volume 2 (9), 2022 http://portalddata.org/index.php/portalddata/article/view/215/204	. Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

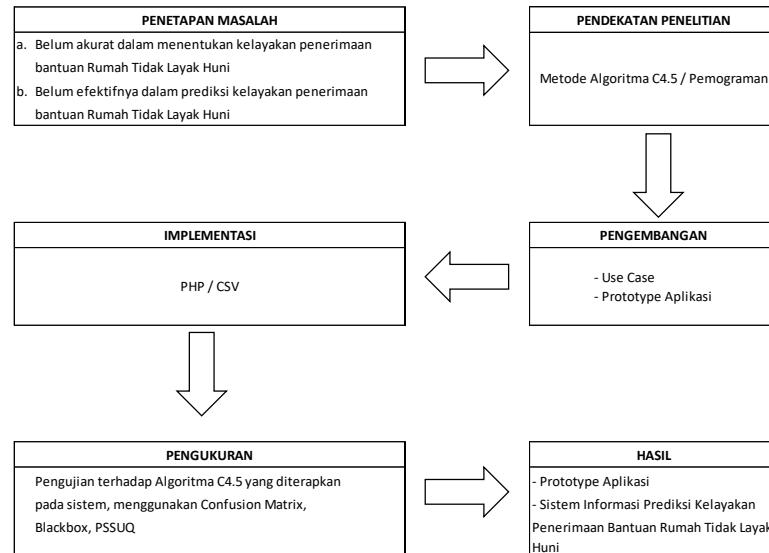
No	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
4	Implementasi Data Mining Untuk Bantuan Dana Bansos Dengan Menggunakan Algoritma C4.5 Di pemerintahan Kabupaten EmpatLawang.	Firamon Syakti,pahmi Ladioanto,2022	JurnalJupiter,Vol.14No .2BulanOktober,Tahun 2022,Hal.287-296 https://jurnal.polsri.ac.id/index.php/jupiter/article/view/5075	. Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
5	The Application Of C4,5 Algorithm For Selecting Scholarship Recipients	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 Algoritma C4.5 dijadikan dalam metode dalam penelitian
6	Penerapan metode C 4.5 untuk klasifikasi warga miskin pada desa mengandung sari	Carissa Elma Purnomo,Rikendry	Jurnal teknologi dan sistem informasi(JTSI) Vol.2,No.3,september 2021,14-25 E-ISSN:2746-3699 file:///C:/Users/LENOVO/Downloads/936-2807-1-PB.pdf	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian
7	Analisis Data Mining Menggunakan Metode Algoritma Untuk Menentukan Penerima Bantuan Langsung Tunai	Lukman Bachtiar,Mahradianur	Jurnal Informatika Fakultas Teknik Informatika dan Bina Sarana Informatika Vol 10, No 1 (2023) https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/15115	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian
8	Penerapan Data Mining Algoritma C4.5 menentukan Kelayakan Penerima Bantuan Covid 19	Angga riadi, Sundaru retno andani, Retnora fiqa dewi	JUKI : Jurnal Komputer dan Informatika Volume 3 Nomor 2 November 2021 e-ISSN : 2722-4368 https://www.ioinformatika.org/index.php/JUKI/article/view/60	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.

No	JUDUL	PENYUSUN	SUMBER	KONTRIBUSI
9	Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (BPNT)	Rizal Amegia Saputra, Sri Wasiyanti, Denny Pribadi	IJUBI : Indonesia Journal Of Bussiness Intelligence Vol 4, No 1 https://ejournal.almaata.ac.id/index.php/IJUBI/article/view/1757	Penggunaan perhitungan menggunakan Algoritma C4.5 dijadikan dalam metode dalam penelitian.
10	Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4.5 Kasus Kantor Kelurahan Martoba	Winda Lidysari, Heru Satria Tambunan, Hendry Qurniawan	KESATRIA Journal Penerapan Sistem Informasi (Komputer dan Manajemen) Vol 3, No 1 https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/97	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. Pada penelitian ini yaitu melakukan pengembangan aplikasi serta menghitung tingkat akurasi. 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.

C. 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 rumah tidak layak huni.

(2) Pendekatan Penelitian

Pendekatan penelitian ini yaitu, untuk menentukan kelayakan penerimaan rumah tidak layak huni. 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 rumah tidak layak huni.

D. Hipotesis

Algoritma C4.5 adalah algoritma klasifikasi pohon keputusan dan dapat mengimplementasikan metode algoritma c4.5 yang digunakan dengan memasukan data sampel dan dimenghitung secara bertahap berdasarkan kelas. Berdasarkan permasalahan yang dihadapi yaitu belum efektif dalam menentukan kelayakan penerimaan rumah tidak layak huni 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 rumah tidak layak huni. 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, p.21) maka hipotesis dari penelitian ini dapat ditetapkan bahwa algoritma C4.5 diduga efektif dalam memprediksi penerimaan rumah tidak layak huni menggunakan algoritma C4.5