

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining

Menurut Tan (2006) didalam Eko Presetyo (2013), menyatakan bahwa data mining adalah proses mendapatkan informasi yang berguna dari gudang database besar. Data mining juga dapat diartikan sebagai ekstraksi informasi yang baru diambil dalam sejumlah data besar yang dapat membantu dalam pengambilan keputusan. Istilah data mining disebut Knowledge Discovery. Knowledge Discovery in Database (KDD) adalah penerapan metode ilmiah pada data mining. Dalam hal ini, data mining merupakan salah satu tahapan dalam proses KDD (Turban, dkk 2001)

KDD berhubungan dengan mengintegrasikan teknologi dan penemuan ilmiah, interpretasi dan mode data yang divisualisasikan. Rangkaian proses ini memiliki tahapan sebagai berikut (Tan,2006) :

- a. Pembersihan data adalah untuk menghilangkan data dan kebisingan yang tidak konsisten.
- b. Integrasi data yaitu menggabungkan data dari berbagai sumber.
- c. Transformasi data yaitu mengubah data menjadi bentuk yang cocok untuk penambangan.
- d. Aplikasi teknik data mining yaitu proses ekstraksi pola dari data yang ada.
- e. Evaluasi pola yang ditemukan yaitu proses menafsirkan pola sebagai pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan.
- f. Presentasi pengetahuan yaitu dengan teknologi visualisasi

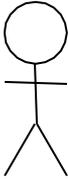
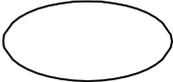
Tahap ini merupakan bagian dari proses pencarian pengetahuan, yang melibatkan pengecekan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau asumsi sebelumnya. Langkah terakhir KDD adalah menyajikan pengetahuan dalam bentuk yang mudah dipahami (Tan, 2006).

2. Pengertian Unified Modeling Language (UML)

Menurut Dharwiyanti "Pengantar *Unified Modeling Language*, 2003:2), menjelaskan bahwa *Unified Modeling Language* (UML) adalah sebuah bahasa yang berdasarkan grafik/gambar untuk memvisualisasi, menspesifikasikan, membangun, dan pendokumentasian dari sebuah sistem.

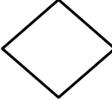
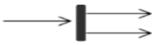
- 1) pengembangan software berbasis OO (*Object Oriented*). Berikut ini simbol UML,
- 2) diantaranya adalah :

Tabel 2.1. Simbol Use case Diagram

GAMBAR	NAMA	KETERANGAN
	<i>ACTOR</i>	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.
	<i>USE CASE</i>	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.
	<i>ASOSIASI/ ASSOCIATION</i>	Komunikasi antara actor dan use case, yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.
<i><<extend>></i>	<i>EKSTENSI/ EXTEND</i>	Relasi usecase tambahan ke sebuah usecase dimana use case yang ditambahkan dapat berdiri sendiri walau tanpa use case tambahan memiliki nama depan yang sama dengan use case yang ditambahkan.
	<i>GENERALISASI / GENERALIZATION</i>	Hubungan generalisasi dan spesialisasi (umum-khusus) dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
<i><<include>></i>	<i>MENGGUNAKAN INCLUDE</i>	Relasi use case tambahan ke sebuah use case dimana use case yang ditambahkan memerlukan use case ini untuk menjalankan fungsional atau sebagai syarat dijalankan use case ini.

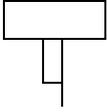
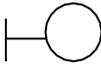
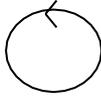
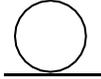
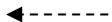
Sumber (Dharwiyanti Pengantar UML 2003:2)

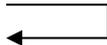
Tabel 2.2 Simbol Activity Diagram

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

Sumber (Dharwiyanti Pengantar UML 2003:2)

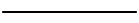
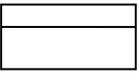
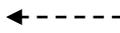
Tabel 2.3. Simbol Sequence Diagram

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>Lifeline</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
	<i>General</i>	Merepresentasikan entitas tunggal dalam sequence.
	<i>Boundary</i>	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
	<i>Control</i>	Elemen mengatur aliran dari informasi untuk sebuah skenario. Objek ini umumnya perilaku dan perilaku bisnis.
	<i>Entitas</i>	Elemen yang bertanggung jawab menyimpan atau informasi. Ini dapat berupa beans atau model object.
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan/hubungan antar objek yang menunjukkan urutan kejadian
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri,

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

Sumber (Dharwiyanti Pengantar UML 2003:2)

Tabel 2.4. Simbol Class Diagram

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

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

Sumber (Dharwiyanti Pengantar UML 2003:2)

3. Pengertian Klasifikasi

Menurut Han dan Kamber (2006:285) *Classification* adalah sebuah model dalam data mining dimana, *classifier* dikonstruksi untuk memprediksi categorical label, seperti “aman” atau “beresiko” untuk data aplikasi peminjaman uang; “ya” atau “tidak” untuk data marketing; atau “treatment A”, “treatment B” atau “treatment C” untuk data medis, kategori tersebut dapat direpresentasikan dengan nilai yang sesuai dengan kebutuhannya, dimana pengaturan nilai tersebut tidak memiliki arti tertentu.

4. Pengembangan Sistem SDLC

Pendekatan sistem adalah metodologi. Metodologi adalah cara yang direkomendasikan untuk melakukan sesuatu. Metode sistem adalah metode dasar untuk menyelesaikan berbagai masalah. System Development Life Cycle (SDLC) adalah penerapan metode sistem dalam pengembangan sistem informasi (Raymond McLeod 2007: 199).

Jika sebuah proyek memiliki probabilitas keberhasilan yang tinggi, diperlukan beberapa tahapan pekerjaan pengembangan. Tahapan tersebut adalah::

- a. Perencanaan
- b. Analisis
- c. Desain
- d. Implementasi
- e. Penggunaan

Proyek dan sumber daya yang dibutuhkan untuk melakukan pekerjaan direncanakan kemudian disatukan. Sistem yang ada juga dianalisis untuk memahami

masalah dan menentukan persyaratan fungsional dari sistem yang baru. Sistem baru ini kemudian dirancang dan diimplementasikan. Setelah implementasi, sistem kemudian digunakan, idealnya untuk jangka waktu yang lama.

Karena pekerjaan-pekerjaan di atas mengikuti satu pola yang teratur dan dilaksanakan dengan cara dari atas ke bawah, SDLC tradisional sering kali disebut sebagai pendekatan air terjun (*waterfall approach*). Aktivitas ini memiliki aliran satu arah menuju ke penyelesaian proyek

Gambar 2.1 berikut mengilustrasikan sifat melingkar dari siklus hidup. Ketika sebuah sistem telah melampaui masa manfaatnya dan harus diganti, satu siklus hidup baru akan dimulai dengan diawali oleh tahap perencanaan. Mudah bagi kita untuk melihat bagaimana SDLC tradisional dapat dikatakan sebagai suatu aplikasi dari pendekatan sistem. Masalah akan didefinisikan dalam tahap-tahapan perencanaan dan analisis. Solusi-solusi alternatif diidentifikasi dan dievaluasi dalam tahap desain. Lalu, solusi yang terbaik diimplementasikan dan digunakan. Selama tahap penggunaan, umpan balik dikumpulkan untuk melihat seberapa baik sistem mampu memecahkan masalah yang telah ditentukan. (RaymondMcLeod, 2007:188).



**Gambar 2. 1 Pola Melingkar dari Siklus Hidup Sistem
(Raymond MCLeod, Jr, 2007)**

5. Bahasa Pemrograman

a. Pengertian PHP

Menurut Anhar (2010:3), menjelaskan bahwa PHP singkatan dari *Hypertext Preprocessor* yaitu bahasa pemrograman web *server-side* yang bersifat *open source* atau dikhususkan untuk pengembangan web dan dapat ditanamkan pada sebuah skripsi HTML berbasis kode-kode (*script*) yang digunakan untuk mengolah suatu data dan mengirimkannya kembali ke web browser menjadi kode HTML.

b. Pengertian *Hypertext Markup Language* (HTML)

Menurut Simarmata (2010):52), menjelaskan bahwa HTML adalah bahasa *markup* untuk menyebarkan informasi pada web. bahasa pemrograman standar yang digunakan untuk membuat sebuah halaman web, yang kemudian dapat diakses untuk menampilkan berbagai informasi di dalam sebuah penjelajah *web* internet (*browser*). informasi yang terstruktur di dalam dokumen atau sekumpulan dokumen. Walaupun HTML tidak dengan mudah dapat dipahami kebanyakan orang, ketika diterbitkan penggunaannya menjadi jelas.

c. Pengertian My SQL

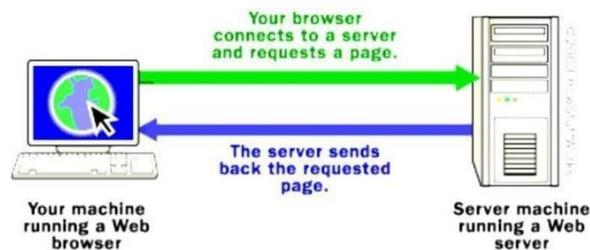
Menurut Kustiyahningsih (2011:145-147), menjelaskan bahwa “My SQL adalah sebuah basis data yang mengandung satu atau jumlah tabel. Tabel terdiri atas sejumlah baris dan setiap baris mengandung satu atau sejumlah tabel. Tabel terdiri atas sejumlah baris dan setiap baris mengandung satu atau sejumlah tabel”. Tipe data My SQL adalah data yang terdapat dalam sebuah tabel berupa *field-field* yang berisi nilai dari data tersebut. Nilai data dalam *field* memiliki tipe sendiri-sendiri.

d. Pengertian Database

Menurut Mustakini (2009:46), menjelaskan bahwa database adalah kumpulan dari data yang saling berhubungan satu dengan yang lainnya, tersimpan di perangkat keras komputer dan digunakan perangkat lunak untuk memanipulasi.

6. Web Server (Apache)

Menurut Kurniawan (2008: 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).



Gambar 2.2 Cara Kerja Web Server

(Sumber : Rulianto Kurniawan, 2008)

Keterangan :

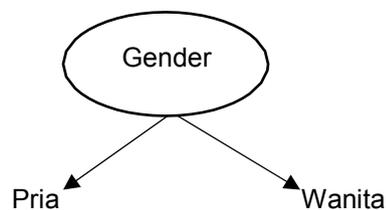
1. Cara kerja Web Server Web server merupakan mesin dimana tempat aplikasi atau software beroperasi dalam mendistribusikan web *page* ke *user*, tentu saja sesuai dengan permintaan *user*.
2. Hubungan antara web server dan browser internet merupakan gabungan atau jaringan komputer yang ada di seluruh 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).

7. Intranet

Menurut O'Brien dan Marakas (2011), menjelaskan bahwa intranet adalah suatu jaringan didalam sebuah organisasi dengan menggunakan teknologi internet (seperti web browser dan server, TCP/IP protokol jaringan, penerbitan dokumen HTML, hypermedia dan database, dan lain-lain), untuk menyediakan lingkungan internet dalam suatu perusahaan untuk berbagi informasi, komunikasi, kolaborasi, serta dukungan dari proses bisnis.

a. Metode Algoritma C4.5

Algoritma C4.5 adalah algoritma terkenal yang digunakan untuk mengklasifikasikan data dengan atribut numerik dan kategorial. Hasil proses klasifikasi dalam bentuk rule dapat digunakan untuk memprediksi nilai atribut tipe diskrit record baru. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3 yang pengembangannya didasarkan pada penanggulangan data yang hilang, dapat menangani data kontinyu dan pemangkasan. (Elisa, 2017).



Gambar 2. 1 Syarat pengujian fitur biner

Isi penting dari induksi pohon keputusan adalah bagaimana menyatakan kondisi pengujian pada node. Ada 3 kelompok persyaratan pengujian node yang penting:

1) Fitur biner

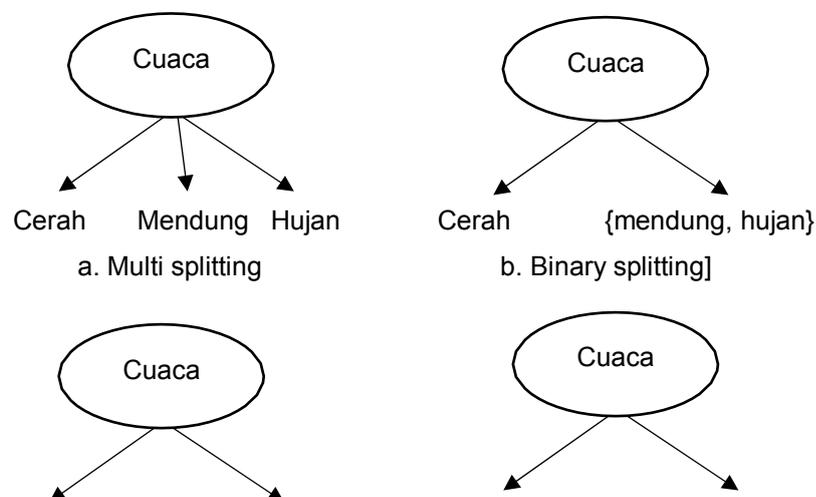
Fitur dengan hanya dua nilai berbeda disebut fitur biner. Saat fungsi ini menjadi node (node root atau node internal), pengujian hanya memerlukan dua opsi cabang. Gambar 2.3 menunjukkan contoh solusi.

2) Fitur bertipe kategorial

Untuk fitur yang nilainya merupakan tipe klasifikasi (nominal atau ordinal), fitur tersebut dapat memiliki beberapa nilai berbeda. Contohnya adalah bahwa fungsi "cuaca" memiliki 3 nilai berbeda, dan dapat memiliki beberapa kombinasi untuk menyelesaikan kondisi pengujian. Biasanya ada dua yaitu binary split dan multiple split. Kombinasi tersebut ditunjukkan pada Gambar 2.3. Untuk solusi yang hanya mengizinkan solusi biner, seperti algoritme CART, mungkin terdapat kombinasi solusi $2^{(k-1)}$, di mana k adalah jumlah nilai yang berbeda dalam fitur tersebut. Contoh solusi ditunjukkan pada Gambar 2.3 dan Gambar 2.4.

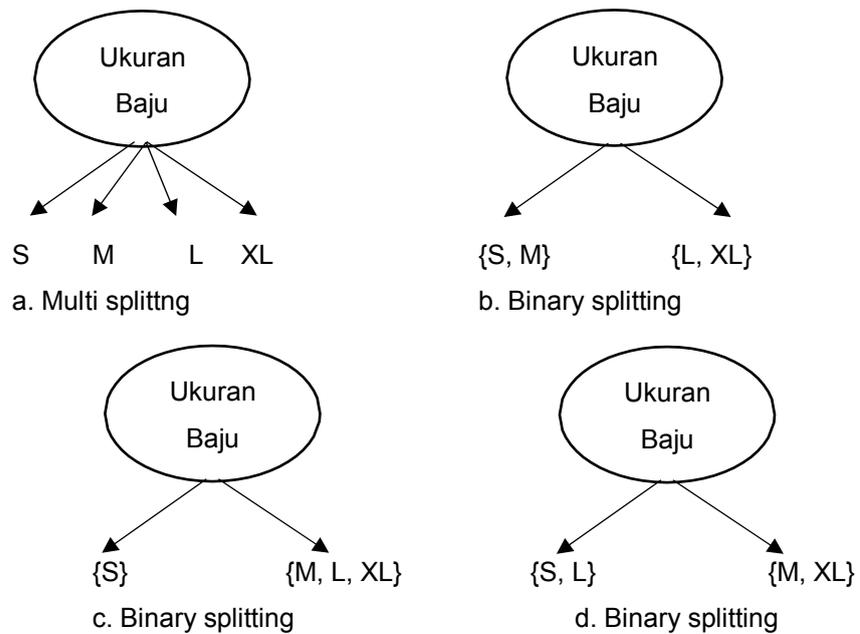
3) Fitur bertipe numerik

Untuk fitur numerik, kondisi pengujian node (root atau internal) diwakili oleh uji perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau untuk beberapa pengujian dengan beberapa nilai dalam bentuk hasil $v_i \leq A < v_{(i+1)}$, untuk $i = 1, 2, \dots, k$. Dalam kasus solusi biner, algoritme memeriksa semua kemungkinan posisi solusi v dan memilih posisi v terbaik. Untuk multi-mode, algoritme harus memeriksa semua kemungkinan rentang nilai kontinu. Gambar 2.4 menunjukkan contoh pemecahan fitur numerik.

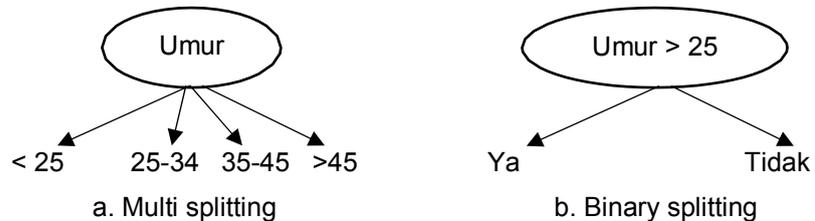


mendung {cerah, hujan} {cerah, mendung} hujan
 c. Binary splitting d. Binary splitting

Gambar 2. 2 Syarat pengujian fitur bertipe nominal



Gambar 2. 3 Syarat pengujian fitur bertipe ordinal

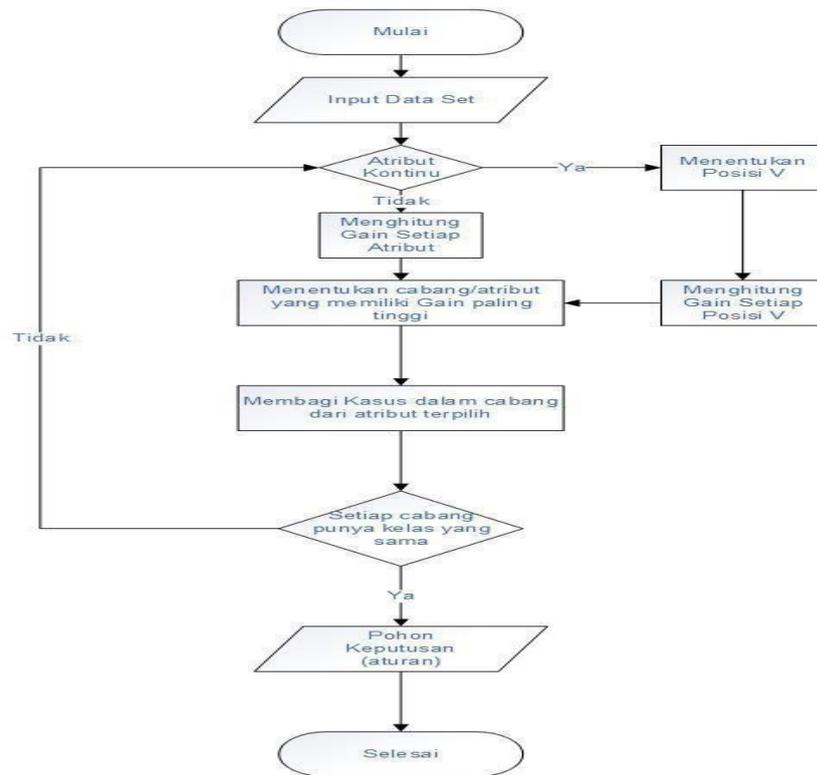


Gambar 2. 4 Syarat pengujian fitur bertipe numerik

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

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

Rumus di atas merepresentasikan rasio gain pada fitur 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. (Prasetyo, 2014). Adapun gambaran algoritma C4.5 menggunakan flowcart yang dapat dilihat pada gambar 2.6.



Gambar 2. 5 Flowchart Algoritma C4.5
(Larose, 2005)

Berdasarkan flowchart gambar 2.7 alur algoritma C4.5 dapat diuraikan sebagai berikut :

- Memasukan data set yang telah disediakan. Data set adalah kumpulan objek dan atributnya.
- Pada proses cek keputusan dilihat apakah data set tersebut termasuk kedalam atribut kontinu atau bukan, atribut kontinu adalah atribut yang mempunyai jangkuan real.
- Jika iya, maka akan menentukan posisi V, setelah itu menghitung *Gain* setiap V, jika tidak maka menghitung *Gain* setiap atribut.
- Setelah itu menentukan cabang/atribut yang memiliki *Gain* paling tinggi.
- Setelah itu membagi kasus dalam dalam cabang dari atribut terpilih.
- Jika setiap cabang mempunyai kelas yang sama, maka menghasilkan pohon keputusan. jika tidak maka cabang atau atribut tersebut kembali melakukan cek atribut kontinu.

Menurut (Ogedebe, 2012), prototyping adalah metode pengembangan perangkat lunak, yang merupakan model fisik kerja sistem dan merupakan

versi awal dari sistem. Dengan menggunakan metode prototipe ini, prototipe sistem akan diproduksi sebagai perantara antara pengembang dan pengguna sehingga dapat berinteraksi dalam proses kegiatan pengembangan sistem informasi. Oleh karena itu, proses keberhasilan pembuatan prototipe ini adalah dengan menentukan kebutuhan awal. Prototipe akan dihapus atau ditambahkan ke bagian-bagian agar sejalan dengan perencanaan dan analisis yang dilakukan oleh pengembang hingga proses uji coba dan pengembangan dilakukan secara bersamaan.

Ada 4 metode prototipe utama, yaitu:

1. Illustrative, menghasilkan sample laporan dan tangkapan layar.
2. Simulated, mensimulasikan beberapa alur kerja sistem, tetapi tidak menggunakan data aktual.
3. Functional, mensimulasikan beberapa proses sistem yang sebenarnya dan menggunakan data aktual.
4. Evolutionary, Menghasilkan model yang merupakan bagian dari sistem operasi.

Tujuan pembuatan prototipe untuk pengembang sistem adalah untuk mengumpulkan informasi dari pengguna sehingga pengguna dapat berinteraksi dengan model prototipe yang dikembangkan, karena prototipe menggambarkan versi awal sistem untuk kelangsungan sistem sebenarnya yang lebih besar.

(Ogedebe, 2012), menegaskan bahwa dalam analisis dan perancangan sistem, khususnya untuk pemrosesan transaksi, dimana dialog yang ditampilkan lebih mudah dipahami. Semakin besar interaksi antara komputer dan pengguna, semakin cepat proses pengembangan sistem informasi, dan semakin banyak interaksi antar pengguna dalam proses pembangunan, maka semakin besar pula manfaat yang didapat.

Prototipe tersebut dapat diaplikasikan pada pengembangan sistem skala kecil dan besar, diharapkan proses pembangunan dapat berjalan dengan baik, teratur dan selesai tepat waktu. Setelah prototipe terbentuk, partisipasi penuh pengguna akan menguntungkan semua pihak yang terlibat, termasuk pimpinan, pengguna itu sendiri, dan pengembang sistem.

Manfaat lain menggunakan prototipe adalah:

1. Menerapkan sistem aktual dalam salinan sistem yang akan dijalankan, dan mengakomodasi masukan pengguna untuk melengkapi sistem.
2. Pengguna akan lebih bersedia menerima setiap perubahan sistem yang dikembangkan berdasarkan jalannya prototipe hingga hasil pengembangan akhir akan dijalankan nanti.

3. Prototipe dapat ditambahkan atau dikurangi sesuai dengan proses pengembangan. Pengguna dapat melacak kemajuan langkah demi langkah.
4. Menghemat sumber daya dan waktu, serta menghasilkan produk yang lebih baik dan lebih efisien bagi pengguna.

B. Pengertian Rekomendasi

Rekomendasi adalah suatu proses komunikasi atas produk atau jasa tertentu yang berguna untuk memberikan informasi secara personal. Contoh rekomendasi yang paling umum digunakan menurut Kotler dan Keller adalah word of mouth communication (WOM) atau rekomendasi dari mulut ke mulut. Rekomendasi secara sederhana bisa disebut sebagai saran yang menganjurkan, membenarkan, atau menguatkan mengenai sesuatu atau seseorang (Kotler dan Keller, 2007). Rekomendasi perhatian bahwa orang yang disebut dapat dipercaya, baik (biasa dinyatakan dengan surat) penyuguhan, 2 saran yang menganjurkan (membenarkan, menguatkan). Ambil contoh, rekomendasi pengangkatan karyawan kontrak menjadi karyawan tetap selalu berdasarkan data dan informasi terbaru yang didasarkan pengamatan audit saat bekerja Begitupun Rekomendasi siswa berprestasi.

Siswa berpotensi mengikuti lomba adalah siswa yang menguasai pengetahuan atau keterampilan yang dikembangkan oleh mata pelajaran, lazimnya ditunjukkan dengan nilai tes atau angka nilai yang diberikan guru, kemampuan yang sungguh-sungguh ada atau dapat diamati (actual ability) dan yang dapat diukur langsung dengan tes tertentu. Menurut Sumadi Suryabrata (2006: 297), prestasi dapat puladi definisikan sebagai berikut : “nilai merupakan perumusan terakhir yang dapat diberikan oleh guru mengenai kemajuan/prestasi belajar siswa selama masa tertentu”. Jadi Rekomendasi siswa berpotensi mengikuti lomba adalah suatu saran yang dianjurkan bahwa adanya siswa yang berpotensi di kelas tersebut.

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. Aplikasi Data mining menggunakan metode Decision Tree untuk menampilkan laporan hasil nilai akhir mahasiswa (Studi kasus di Fakultas Teknik Umsurabaya) (Novianti, Aziz Program Studi D3 Teknik Komputer UmSurabaya, Surabaya 60113)
Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan Aplikasi data mining ini dapat digunakan untuk menampilkan informasi yang berguna tentang nilai akhir mahasiswa dengan teknik data mining. Informasi yang ditampilkan berupa hubungan antara nilai akhir mahasiswa dengan nilai Aktifitas, Tugas, Ujian Tengah Semester (UTS) dan ujian akhir semester (UAS). Hal ini terlihat pada pohon keputusan (*decision tree*) dan rancangan aplikasi berbasis pohon keputusan. Hasil evaluasi dan validasi dengan Cross Validation mengajukan nilai akurasi pada Algoritma C4.5 sebesar 100%. Data untuk mahasiswa yang diproses data mining meliputi data proses masuk data asal sekolah, data kota mahasiswa, dan data program studi. Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi tingkat kelulusan khususnya faktor dalam daya induk mahasiswa.
2. Implementasi Decision Tree Algoritma C4.5 untuk memprediksi kesuksesan pendidikan karakter (Abdilah, Setyanto) Magister Teknik Informatika Universitas Amikom Yogyakarta.
Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Hasil Accuracy sebesar 60,91%, menunjukkan bahwa algoritma C4.5 layak digunakan untuk melakukan prediksi tingkat kesuksesan pada pendidikan karakter. Namun masih butuh banyak kajian mendalam terutama dalam menentukan atribut-atribut yang memang benar-benar mempengaruhi kesuksesan pendidikan karakter sehingga diharapkan akan dapat *accuracy* yang lebih baik, serta terjadi efisiensi dalam pengelompokan atribut agar tidak terlalu banyak atribut – atribut, dimana dalam penelitian ini dari 12 regular atribut didapat atribut turunan hingga 64 atribut. Yang dapat membuktikan beberapa asumsi terkait pendidikan khususnya pendidikan karakter, diantaranya: Bidang Bahasa dan Sosial berperan berkorelasi positif dengan latar belakang pendidikan agama serta mempunyai orang tua yang menjadi seorang pendidik/guru berkorelasi positif dengan pendidikan karakter, dan kemampuan untuk membaca Al-Qur'an berkorelasi positif dengan pendidikan karakter
3. Algoritma C4.5 Untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan rapid miner (Ardiansyah, Walim , Pascasarjana Magister Ilmu Komputer / Sekolah Tinggi Manajemen Informatika dan Komputer (STIMIK))
Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang

digunakan dalam perhitungan, yaitu pada siswa SMP kelas VIII yang digunakan dalam perhitungan alternatif tertinggi penentuan siswa yang akan mengikuti lomba cerdas cermat. Metode yang diusulkan untuk proses seperti yang digunakan algoritma C4.5 dengan keriterian yang digunakan sebagai berikut: Nama Siswa, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Nilai Mtematika, Nilai keaktifan, Perolehan Skor IQ dan nilai bimbingan belajar. 1. Klasifikasi proses seleksi calon peserta lomba siswa SMP dapat mengklasifikasikan siswa dalam tahapan lolos atau tidaknya dalam seleksi. 2. Dari 33 data siswa yang digunakan menunjukkan tingkat akurasi dengan algoritma C4.5 sebesar 81,81%.

4. ALGORITMA C4.5 UNTUK PENILAIAN KINERJA KARYAWAN (Julianto, Yunitarini, dan Sophan Universitas Trunojoyo Madura)

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Berdasarkan evaluasi yang dilakukan dapat diketahui bahwa proses pembentukan pohon menggunakan teknik pruning memiliki kecepatan yang lebih tinggi karena penyederhanaan pohon, tetapi tidak selalu memiliki akurasi yang lebih besar Perbedaan pohon keputusan yang dihasilkan disebabkan oleh perbedaan jumlah data training yang digunakan pada masing-masing partisi. 3. Pohon keputusan Partisi A menggunakan teknik pruning dengan jumlah data training lebih besar daripada data testing memiliki akurasi tertinggi dibandingkan dengan pohon keputusan yang lain, yaitu mencapai 90 %. 4. Dengan adanya system ini, pengolahan data meliputi penempatan dan kinerja karyawan dapat dilakukan dengan mudah dan cepat.

5. SISTEM REKOMENDASI PENJURUSAN SEKOLAH MENENGAH KEJURUAN DENGAN ALGORITMA C4.5 (Indra Mukti Prabowo dan Subiyanto Fakultas Teknik)

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Berdasarkan confusion matrix, sistem berhasil memberikan rekomendasi jurusan yang sesuai dengan data jurusan siswa akademis tiga besar berprestasi pada data eksperimen sebanyak 40 data siswa. Jumlah data yang hasil rekomendasi jurusan dari sistem tidak sesuai dengan data jurusan siswa berprestasi adalah 8 data. Dengan demikian tingkat akurasi dari sistem rekomendasi jurusan SMK sebesar 83,33% dan error rate 16,67% Hasil rekomendasi jurusan yang tepat 100% terjadi pada data pengujian dengan jurusan AK, TKJ, PM, MM, PBS , F, TKR, BB, dan AP. Data pengujian dengan jurusan TPMI memiliki nilai akurasi rekomendasi 67,33% yaitu dua dari tiga memiliki hasil rekomendasi yang sesuai dengan data kenyataan siswa, yaitu TPMI. Satu data lain direkomendasikan jurusan TSM. Data dengan jurusan TSM sebanyak sembilan data dapat direkomendasikan sesuai dengan jurusan siswa sebanyak delapan data, sedangkan satu data direkomendasikan oleh sistem sebagai jurusan TKJ. Nilai akurasi sistem dalam uji data dengan jurusan TSM adalah 88,9%. Keseluruhan data pengujian dengan jurusan

TKPI dan NKPI tidak dapat direkomendasikan sesuai dengan data kenyataan jurusan siswa. Tiga data dengan jurusan TKPI direkomendasikan sistem sebagai jurusan TKR, TPMI, dan TKJ. Tiga data dengan jurusan NKPI direkomendasikan oleh sistem sebagai jurusan TPMI, MM, dan TKJ. Secara ringkas Artikel ini menyajikan pentingnya pengembangan sistem rekomendasi pemilihan jurusan SMK yang dapat membantu siswa yang mengalami kesulitan atau hambatan dalam memilih jurusan pada sekolah lanjutan (SMK). Sudah disajikan hasil analisis dari penerapan algoritma C4.5 untuk pengembangan sistem rekomendasi jurusan SMK memiliki tingkat akurasi sebesar 83,33%. Oleh karena itu, decision tree yang dibangun dengan algoritma C4.5 memiliki potensi yang besar untuk menyelesaikan masalah siswa dalam memilih jurusan SMK.

6. Penerapan Algoritma C4.5 dalam Menentukan Kandidat Siswa/Siswi SMK dalam Mengikuti Lomba Kompetensi Siswa (Putra, Fardiansah Universitas Sumatera Utara)
Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Berdasarkan hasil pengujian, maka dapat ditarik kesimpulan sebagai berikut : 1. Berdasarkan analisa yang dilakukan, untuk proses penerapan algoritma C4.5 maka terlebih dahulu untuk menentukan kriteria yang diperlukan untuk proses pemilihan kandidat siswa yang nanti akan dipilih untuk mengikuti Lomba Kompetensi Siswa. 2. Untuk membuat suatu pola keputusan yang dapat dijadikan landasan, maka harus melakukan proses penambahan data pemilihan kandidat siswa pada Sekolah Nurul Amaliyah Tanjung Morawa untuk perhitungan dengan menggunakan rumus C4.5. 3. Proses pengimplementasian algoritma C4.5 untuk mendapatkan suatu pohon keputusan dapat diuji dengan menggunakan aplikasi rapid miner berdasarkan data yang sudah ada maupun data baru.
7. KLASIFIKASI SISWA YANG AKAN MENGIKUTI LOMBA OLIMPIADE SAINS NASIONAL (OSN) MENGGUNAKAN ALGORITMA C4.5 (FAUZI, RAHMAD (2017)Universitas Muhammadiyah Gresik)

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan Hasil yang didapatkan dari penelitian ini dapat disimpulkan sebagai berikut: 1. Sistem ini dapat menghasilkan informasi untuk rekomendasi siswa yang akan terpilih mengikuti lomba Olimpiade Sains Nasional menggunakan metode Decision Tree C4.5. 2. Pemilihan data training dalam pengujian ini menggunakan trial and error. Dari tiga kali percobaan, pohon keputusan terbaik yang digunakan untuk mengklasifikasi status siswa adalah pohon keputusan pada percobaan ke-2. 3. Komposisi data training akan mempengaruhi tingkat akurasi yang didapatkan. Dimana setiap percobaan yang telah dilakukan akan mendapatkan akurasi yang berbeda-beda dan membentuk pohon keputusan yang terbaik. Nilai akurasi, laju error, sensitivitas, dan spesifisitas dari hasil uji coba yang telah dilakukan sebanyak tiga kali percobaan, percobaan ke-1 dengan jumlah data 50, jumlah data yang

tepat 36 dan jumlah data tidak tepat 14. Kemudian didapatkan akurasi sebesar 72%, laju error 28%, sensitivitas 93,33% spesifitas 62,86%, merupakan hasil pengujian percobaan ke-2 dengan jumlah data 43, jumlah data yang tepat 37 dan jumlah data tidak tepat 6. Kemudian didapatkan akurasi sebesar 86,05% , laju error 13,95%, sensitivitas 93,75% spesifitas 81,48%. merupakan hasil pengujian percobaan ke-3 dengan jumlah data 27, jumlah data yang tepat 16 dan jumlah data tidak tepat 11, Kemudian didapatkan akurasi sebesar 59,26% laju error 40,74%, sensitivitas 90% spesifitas 41,18%.

8. Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa (Riandari Agustina Simanggusong, STMIK Pelita Nusantara Medan)

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan yaitu: berdasarkan hasil penelitian yang telah dilakukan, didapatkan informasi bahwa variabel yang paling berpengaruh terhadap kepuasan mahasiswa adalah Responsiveness, Reability dan Empathy, data mining dengan algoritma C4.5 model decision tree dapat mengklasifikasi variabel – variabel dalam mengukur tingkat keputusan mahasiswa sebagai dasar dalam membuat sebuah keputusan dalam menentukan tindakan apa yang harus dilakukan untuk dapat meningkatkan kepuasan mahasiswa, hasil implementasi di uji menggunakan salahsatu aplikasi data mining algoritma C4.5 yaitu rapid miner studio dan didapatkan hasil klasifikasi yang sama output atau hasil pada penelitian ini dibagi menjadi 2 kategori yaitu puas dan tidak puas berdasarkan output yang dihasilkan maka digunakan data mining 33 dengan teknik klasifikasi karena pada teknik ini terdapat variabel variabel kategori yang akan menghasilkan informasi, 1. Dapat slection yaitu menggabungkan data dari beberapa sumber, 2.data cleaning yaitu untuk membuat data yang tidak konsistem dan nois, data tranformation yaitu mentranformasikan data menjadi bentuk yang sesuai untuk di mining, data mining yaitu proses esensial dimana metode yang intelegent digunakan untuk mengekstrak pola data, kemudian patren evaluation untuk mengidentifikasi pola yang benar benar menarik yang mewakili pengetahuan berdasarkan atau beberapa tindakan yang menari, graphicial user interfaces (GUI) yaitu untuk end User.

9. SISTEM INFORMASI SEKOLAH MENGGUNAKAN ALGORITMA C4.5 UNTUK MENGIDENTIFIKASI FAKTOR FAKTOR PENURUNAN MUTU PEMBELAJARAN SISWA SMK (Supangat, Ayub Budi Anggara TeknikInformatika. Fakultas Teknik, Universitas 17 Agustus 1945) Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan dari sistem informasi sekolah menggunakan algoritma C4.5 untuk mengidentifikasi faktor – faktor penurunan mutu pembelajaran siswa SMK dari penggunaan dari algoritma C4.5 dalam mengambil sebuah keputusan, data training berjumlah 250 data dimana sudah diisi oelh alumni SMK Muhmmadiyah 1 taman, kemudian data dilakukan perhitungan menggunakan algoritma C4.5 sehingga menunculkan sebuah rule – rule

keputusan juga dapat dilihat pohon keputusan juga. Kemudian untuk menguji akurasi algoritma C4.5 telah mendapatkan sebanyak 150 data uji dari pengisian yang dilakukan oleh siswa kelas X(sepuluh) yang sudah melakukan uji akhir semester dikarenakan dalam sistem terdapat pertanyaan nilai rata rata UAS untuk pelajaran umum dan produktif, data uji dibagi menjadi 3 (tiga) yaitu tingkat akurasi dengan 50 data, tingkat akurasi dengan 100 data dan tingkat akurasi dengan 150 data dan menghasilkan tingkat akurasi yang berbeda jika 50 data uji tingkat akurasi prediksi sebesar 56% dan laju error sebesar 44%, jika 100 data uji menghasilkan tingkat akurasi sebesar 64% dan laju error sebesar 36% kemudian jika 150 data uji maka menghasilkan tingkat akurasi sebesar 71,33% dan laju error sebesar 28,67% dalam mengambil keputusan algoritma C4.5 dapat menghasilkan tingkat akurasi atau ketepatan yang baik antara 56% - 71%.

10. ANALISIS DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA BERDASARKAN DATA NILAI AKADEMIK (Selvia Lorena Br Ginting¹ , Wendi Zarman² , Ida Hamidah³ ^{1,2,3}Jurusan Teknik Komputer, Fakultas Teknik dan Ilmu Komputer, UNIKOM Bandung) Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil Kesimpulan Aplikasi untuk memprediksi masa studi mahasiswa berhasil dirancang dan diimplementasikan menggunakan algoritma C4.5. Aplikasi tersebut dapat memprediksi masa studi mahasiswa. Berdasarkan training dan pengujian kemudian dilakukan analisis maka dapat diambil beberapa kesimpulan bahwa jumlah data training mempengaruhi persentase kecocokan atau keakurasian. Persentase kecocokan hasil pengujian yang didapatkan dari data training 112 lebih kecil di bandingkan dengan data training 70 untuk berapapun kombinasi atribut. Demikian pula untuk semua konfigurasi atribut, persentase kecocokan hasil pengujian dengan data training 112 lebih kecil di bandingkan dengan data training 70. Jumlah data testing adalah 35, jumlah data yang sama dengan data training sebanyak 16 dan persentase kecocokan dengan data training adalah $16/35 \times 100\% = 47,71\%$, terdapat gain untuk masing-masing mata kuliah, dimana hasil gain dari masing-masing mata kuliah tersebut diantaranya adalah: 1. Algoritma Pemrograman 1 = 0,9183 2. Fisika 1 = 0,2516.

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
1.	Triuli Novianti , Abdul Aziz	Aplikasi data mining menggunakan metode desicion tree untuk menampilkan laporan hasil akhir mahasiswa (studi kasus di fakultas teknik umsurabaya)	Nilai IPK mahasiswa pada fakultas teknik umumnya rendah dan tidak lulus beberapa matakuliah oleh karena itu prodi memanfaatkan data induk mahasiswa dan data nilai mahasiswa, dapat diketahui hasil nilai akhir mahasiswa melalui teknik data mining	Jurnal Ilmiah NERO Vol. 1 No. 3 2015	Kontribusi penelitian ini adalah konsep output yang menggunakan target atribut sebagai keputusan dengan nilai “lulus” dan “tidak lulus” dalam 1 matakuliah. Kelemahan penelitian ini yaitu tidak terdapat perhitungan C4.5, dan melakukan uji coba menggunakan Aplikasi Rapid Miner.
2.	M.A.Abdilah, Arief Setyanto	Implementasi Desicion Tree Algoritma C4.5 untuk memprediksi kesuksesan pendidikan karakter	Proses studi literatur terkait penelitian sebelumnya mengenai implementasi data mining Decision Tree C4.5 serta pengumpulan data di Unires .	Vol.XV Nomor 2 Juli 2020 – Jurnal Teknologi Informasi ISSN: 1907- 2430	Kontribusi penelitian ini adalah pengambilan uji hasil menggunakan confusion matrix. Kelemahan pada penelitian ini yaitu melakukan proses perhitungan menggunakan Aplikasi Rapid Miner dan Cross Validation
3.	Dian Ardiansyah, Walim	Algoritma C4.5 untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat siswa SMP dengan menggunakan Aplikasi Data mining	Pada siswa kelas VIII yang digunakan untuk perhitungan alternatif tertinggi penentuan siswa yang akan mengikuti lomba cerdas cermat.	Jurnal Inkofar * Volume 1 No.2, Desember 2018 * ISSN: 2615- 364	Kontribusi penelitian ini adalah permasalahan atribut dari nilai Bahasa Indonesia, bahasa Inggris, IPA, IPS dan MTK. Kelemahan pada penelitian ini yaitu melakukan proses perhitungan Aplikasi Rapid Miner.

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
4.	Windy Julianto, Rika Yunitarini	Algoritma C4.5 untuk penilaian kinerja karyawan	Proses seleksi dan kinerja karyawan pada kasus ini proses seleksi menggunakan metode C4.5, Metode ini memberikan output yang mencocokkan inputan dengan pohon keputusan yang dibuat.	Jurnal Inkofar *Volume 1 No.2, Desember 2018 * ISSN: 2615-364	Kontribusi penelitian ini adalah hasil rule yang didapat dari penelitian tersebut. Kelemahan pada penelitian ini yaitu SPK dapat dikembangkan dengan lebih dinamis dengan mengurangi dan menambah atribut lain
5.	Indra Mukti Prabowo dan Subiyanto	Sistem Rekomendasi penjurusan sekolah menengah kejuruan dengan Algoritma C4.5	Proses penentuan variabel data dilakukan yang digunakan dalam pembentukan decision tree. Data jurusan dipilih menjadi variabel keputusan dari pembentukan decision tree.	JURNALKEPE NDIDIKAN, Volume 1, Nomor 1, Juni 2017	Kontribusi penelitian ini adalah hasil menggunakan confusion matrix dan pengambilan metode Algoritma C4.5 dan cara perhitungan C4.5
6.	Fardiansah Putra	Penerapan Algoritma C4.5 dalam menentukan kandidat siswa/siswi SMK dalam mengikuti lomba kompetensi siswa	Pemilihan siswa yang dilakukan adalah untuk seleksi tahap awal perwakilan siswa yang akan mengikuti lomba kompetensi siswa, siswa yang dipilih adalah siswa kelas XII jurusan akuntansi pada sekolah amaliyah Tanjung Morawa	Sains, Informatics Journal, and Economy (SAIKO), Vol. 1 No. 2 Juli, 2018, ISSN : 2016	Kontribusi penelitian ini adalah pengambilan metode Algoritma C4.5 dan cara perhitungan C4.5 Kelemahan dari penelitian ini adalah tidak terdapat nilai akurasi

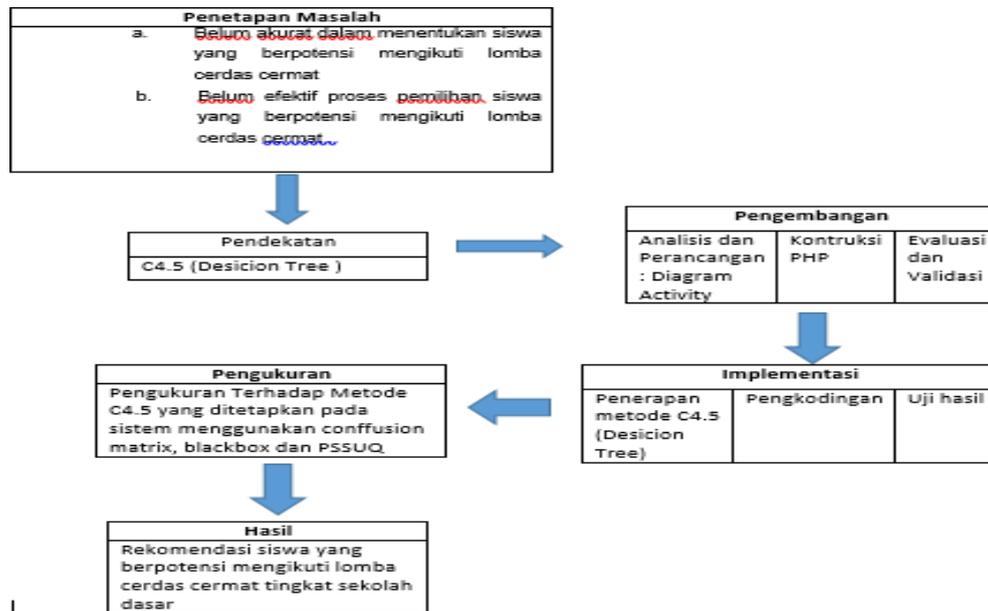
No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
7	Rahmad Fauzi	Klasifikasi Siswa yang akan mengikuti lomba olimpiade sains nasional (OSN) menggunakan Algoritma C4.5	Bagaimana Menentukan siswa yang akan mengikuti lomba Olimpiade Sains Nasional di SMPN 2 Geresik	2018	Kontribusi peneliti ini adalah konsep output yang diharapkan untuk merekomendasi siswa yang Akan terpilih mengikuti lomba. Kelemahan peneliti ini adalah tidak disampaikannya langkah langkah perhitungan C4.5
8	Fristi Riandari Agustina SImagunsong	Penerapan Algoritma C4.5 untuk mengukur tingkat kepuasan mahasiswa	Mengukur tingkat kepuasan mahasiswa untuk mengingat tingginya tingkat persaingan dalam dunia pendidikan pengetahuan dan teknologi	Vol.0.No.tahun(2019); buku terbit tahun 2019 ISBN; 978-623-90597-2-9	Kontribusi penelitian ini adalah menggunakan metode yang sama, kelemahan pada penelitian ini adalah tidak terdapat nilai uji coba,dan uji hasil menggunakan aplikasi
9	Selvia Lorena Br Giniting	Analisis dan penerapan metode Algoritma C4.5 dalam data mining untuk memprediksi masa studi mahasiswa berdasarkan data nilai akademik	Masih banyak mahasiswa program sarjana reguler jurusan teknik komputer yang menempuh lama studi lebih dari 8 semester	Yogyakarta, 15 November 2014, ISSN;1979-911X Prosidang Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014	Kontribusi penelitian ini adalah pengambilan metode algoritma C4.5, kelemahan penelitian ini adalah tidak terdapat cara perhitungan Algoritma C4.5

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
10.	Asmaul Husnah, Nasrullah	Sistem Informasi Sekolah menggunakan Algoritma C4.5 untuk mengidentifikasi faktor faktor penurunan mutu pembelajaran siswa SMK	Penguat atau pengembangan keterampilan yang dimiliki oleh siswa sebelum masuk ke dunia industri	Jurnal Teknik Informatika. Fakultas Teknik, Universitas 17 Agustus 1945.	Kontribusi penelitian ini adalah tampilan rule algoritma C4.5

Berdasarkan uraian, maka terdapat beberapa kekurangan yang menjadi salah satu dasar penelitian ini dilakukan yaitu keseluruhan jurnal hanya membahas tentang kesamaan dalam menentukan metode sedangkan pada penelitian ini rekomendasi siswa yang berpotensi mengikuti lomba tingkat sekolah dasar. Selain itu, variabel- variabel yang digunakan juga berbeda dengan penelitian sebelumnya dimana pada penelitian ini hanya menggunakan 9 variabel yaitu Nama Siswa, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai Ilmu Pengetahuan Alam (IPA), Nilai Ilmu Pengetahuan Sosial (IPS), Nilai Matematika, Nilai keaktifan (meliputi keaktifan mengerjakan soal dan menjawab pertanyaan ketika bimbingan belajar berlangsung), Perolehan skor IQ, dan Nilai Bimbingan Belajar. Output yang dihasilkan dari penelitian sebelumnya dalam mengukur evaluasi uji hasil juga kebanyakan penelitian ini mengukur evaluasi uji hasil menggunakan Aplikasi Rapid Miner.

D. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang digambarkan pada gambar berikut:



Gambar 2.8 Kerangka Pemikiran

Kerangka pemikiran pada gambar 2.4 dapat dijelaskan sebagai berikut :

1. Penetapan masalah mencakup fenomena yaitu Belum Akurat dalam menentukan siswa yang berpotensi mengikuti lomba, serta belum efektif proses pemilihan siswa yang berpotensi mengikuti lomba cerdas cermat
2. Pendekatan penelitian yaitu metode yang ditetapkan atau digunakan yaitu C4.5
3. Pengembangan yaitu tahap melakukan analisis dan perencanaan gambar untuk *diagram activity*, kemudian menentukan bahasa pemrograman dengan kontruksi PHP dan melakukan evaluasi serta validasi.
4. Implementasi yaitu tahap menerapkan metode C4.5 kedalam sistem aplikasi dan pembuatan coding lalu melakukan uji hasil dai sistem tersebut.
5. Penguuran yaitu melakukan pengujian terhadap metode C4.5 yang ditetapkan ke sistem menggunakan conffusion matrix, blackbox, dan PSSUQ.
6. Hasil yaitu sistem informasi menampilkan hasil rekomendasi siswa yang berpotensi mengikuti lomba cerdas cermat tingkat sekolah dasar.

E. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu efektif dan belum tepat pada saat melakukan proses pemilihan siswa yang berpotensi mengikuti lomba cerdas cermat tingkat sekolah dasar. Dalam teori data mining ada beberapa metode yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan dari pengalaman dimasa sebelumnya, diantaranya adalah metode algoritma C4.5. Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Berdasarkan hal tersebut maka dapat ditetapkan, hipotesis pada penelitian ini adalah penerapan Algoritma C4.5 diduga dapat merekomendasi siswa yang berpotensi mengikuti lomba cerdas cermat tingkat sekolah dasar.