

BAB II

TINJAUAN PUSTAKA

2.1 Algoritma C4.5

Elisa (2017) mengungkapkan Algoritma C4.5 merupakan salah satu solusi pemecahan masalah pada teknik klasifikasi. Hasil dari algoritma C4.5 berbentuk decision tree, serupa dengan teknik klasifikasi lainnya. Decision tree adalah struktur yang memecah data besar menjadi himpunan record yang lebih kecil melalui serangkaian aturan keputusan. Setiap tahap pemecahan menghasilkan himpunan anggota yang semakin serupa satu sama lain.

Algoritma C4.5 dikenal luas dan digunakan dalam klasifikasi data dengan atribut numerik dan kategorial. Aturan-aturan yang dihasilkan dari proses klasifikasi digunakan untuk memprediksi nilai atribut diskrit pada record baru.

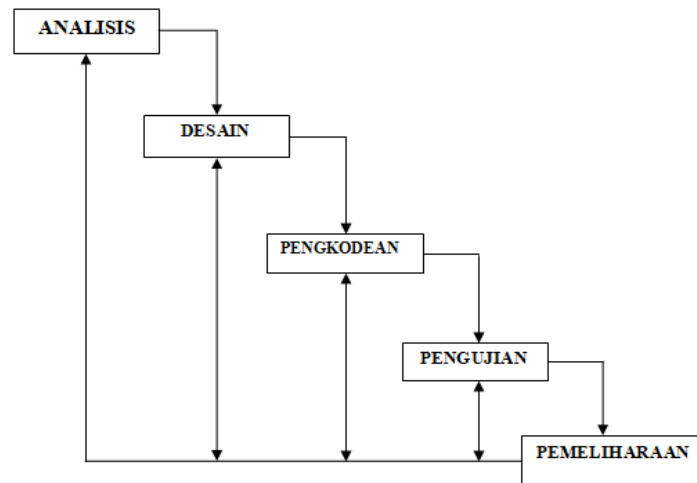
Secara umum algoritma C4.5 untuk membangun keputusan sebagai berikut:

1. Pilih atribut sebagai node akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus memiliki kelas yang sama.

2.2 Metode Waterfall

Menurut Rianto (2021) metode waterfall atau metode air terjun adalah metode ini menggambarkan pendekatan yang sistematis dan juga berurutan pada pengembangan perangkat lunak secara sekuensial atau terurut dimulai dari analisis, desain, pengodean dan pengujian.

Berikut adalah alur gambar dari metode waterfall menurut :



Gambar 2.1 Ilustrasi proses metode Waterfall

Tahapan dari metode waterfall:

a. Analisis

Melakukan analisis kebutuhan perangkat lunak, fungsi dan prose dari web yang di buat, pengidentifikasian kendala dalam pembuatan web, menganalisis keandalan, kelemahan dan teknologi yang di pakai.

b. Desain

Desain perangkat lunak merupakan rangkaian tahapan yang melibatkan beberapa langkah. Tahap dalam proses rancangan pembuatan program perangkat lunak mencakup struktur data, arsitektur perangkat lunak, representasi antarmuka, dan prosedur pengkodean. Proses ini menerjemahkan kebutuhan perangkat lunak dari tahap analisis kebutuhan menjadi representasi desain yang dapat diimplementasikan dalam bentuk program pada tahap berikutnya. Dalam tahap ini, hasil dari desain perangkat lunak didokumentasikan secara rinci.

c. Pengkodean

Desain harus di implementasikan kedalam program perangkat lunak. Keluaran dari tahap ini berupa program komputer sesuai dengan desain yang telah disusun pada tahap sebelumnya. Pada tahapan ini, penulis mengimplementasikan program menggunakan bahasa pemrograman seperti PHP, HTML, CSS, dan lainnya.

d. Pengujian

Pada tahap ini, penulis melakukan pengujian terhadap program yang telah dibuat guna mengidentifikasi potensi kekurangan dalam program tersebut.

e. Pemeliharaan

Tahap ini melakukan pengoperasian program dan melakukan pemeliharaan terhadap perangkat lunak dengan penyesuaian atau perubahan terhadap situasi sebenarnya.

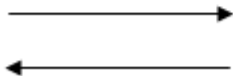
2.3 DFD (*Data Flow Diagram*)

Pengertian *Data Flow Diagram* (DFD) Menurut Krist Diagram yang menggunakan notasi simbol untuk menggambarkan arus data system (Krist, 1998). *Data Flow Diagram* (DFD) sering digunakan untuk menggambarkan suatu sistem yang sudah ada atau sistem yang baru akan dikembangkan secara logika dan menjelaskan arus data dari mulai pemasukan sampai dengan keluaran data tingkatan diagram arus data mulai diagram konteks yang menjelaskan secara umum suatu sistem atau batasan sistem dari level 0 dikembangkan menjadi level 1 sampai sistem tergambar secara rinci. Gambaran ini tidak tergantung pada perangkat keras, perangkat lunak, struktur data atau organisasi file.

Beberapa simbol yang digunakan dalam pembuatan data *flow diagram* ini meliputi :

- a. *External entity* (kesatuan luar)
- b. *Data flow* (arus data)
- c. *Process* (proses)
- d. *Data store* (penyimpanan data)

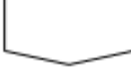
Tabel 2.1 Simbol - simbol DFD

Simbol	Nama	Fungsi
	Simbol entitas eksternal	Digunakan untuk menunjukan tempat asal <i>data</i> .
	Simbol proses	Digunakan untuk menunjukan tugas atau proses yang dilakukan baik secara manual atau otomatis.
	Simbol penyimpanan data	Digunakan untuk menunjukan gudang informasi atau <i>data</i> .
	Simbol arus data	Digunakan untuk menunjukan arus dari proses.

2.4 Flowchart

Flowchart adalah bagan yang menggambarkan urutan instruksi dan hubungan satu proses dengan proses lainnya menggunakan simbol-simbol tertentu, digunakan sebagai alat bantu komunikasi dan dokumentasi. Dalam analisis sistem, flowchart ini digunakan secara efektif untuk menelusuri alur suatu laporan atau form. Adapun simbol flowchart dapat dilihat pada tabel berikut :

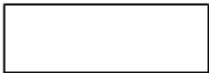
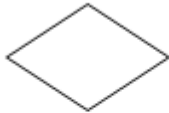
Tabel 2.2 Simbol – simbol Flowchart

Simbol	Nama
	Proses : digunakan untuk pengolahan aritmatika dan pemindahan data.
	Terminal : digunakan untuk menunjukkan awal dan akhir program.
	Preparation : digunakan untuk memberikan nilai awal pada satu variabel.
	Keputusan : digunakan untuk mewakili operasi perbandingan logika.
	Proses terdefinisi : digunakan untuk proses yang detailnya dijelaskan terpisah.
	Penghubung : digunakan untuk menunjukkan hubungan arus proses yang terputus masih dalam halaman yang sama.
	Penghubung halaman lain : digunakan untuk menunjukkan hubungan arus proses yang terputus masih dalam halaman yang sama.

2.5 Entity Relationship Diagram

Entity Relationship Diagram (ERD) merupakan teknik yang digunakan untuk memodelkan kebutuhan data dari suatu organisasi, biasanya oleh sistem analisis dalam tahap analisis persyaratan proyek pengembangan sistem. Sementara seolah-olah teknik diagram atau alat peraga memberikan dasar untuk desain database relasional yang mendasari sistem informasi yang dikembangkan. ERD bersama-sama dengan detail pendukung merupakan model data yang pada gilirannya digunakan sebagai spesifikasi untuk database. (Oleh, n.d.)

Tabel 2.3 Simbol dan Fungsi Entity Relationship Diagram

Simbol	Keterangan
	Entitas
	Atribut
	Hubungan
	Garis

2.5.1 Entitas

Objek dalam dunia nyata yang dapat dibedakan dengan objek lain. Entitas terdiri atas beberapa atribut mengidentifikasi atau membedakan yang satu dengan yang lainnya. Pada setiap entitas baru harus memiliki atribut unik atau yang disebut dengan *primary key*.

2.5.2 Atribut

Isi dari atribut mempunyai elemen yang dapat mengidentifikasi isi elemen satu dengan yang lain. Ada dua jenis atribut, yaitu :

1. *Identifier key* digunakan untuk menentukan suatu entity secara unik (*primary key*).
2. *Descriptor (nonkey attribute)* digunakan untuk mengspesifikasi karakteristik dari suatu entity yang tidak unik

2.5.3 Kardinalitas

Menyatakan jumlah himpunan relasi antar entitas. Pemetaan kardinalitas terdiri dari :

1. One-to-one, sebuah entitas A berhubungan dengan entitas B paling banyak.
2. One-to-many, sebuah entitas A berhubungan dengan entitas B lebih dari satu.
3. Many-to-many, sebuah entitas A berhubungan dengan entitas B lebih dari satu dan entitas B berhubungan dengan entitas A lebih dari satu juga.

2.5.4 Hierarchy Chart

Hierarchy Chart adalah gambaran atau pendefinisian dan mengilustrasikan organisasi sistem secara berjenjang kedalam modul-modul dan sub modul yang digunakan untuk menggambarkan struktur dari suatu sistem (Oleh, n.d.). Bagan terstruktur menjelaskan :

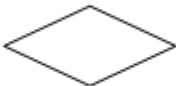
1. Komunikasi antara masukan dan keluaran
2. Pembagian suatu sistem menjadi modul

Bagan terstruktur digunakan dalam perancangan terstruktur dari suatu sistem informasi yang menunjukkan hubungan elemen data dan elemen kontrol, modul dan hubungan antar modul dalam suatu program.

Dengan adanya bagan terstruktur dapat diketahui :

1. Masukan dari sebuah modul.
2. Keluaran dari sebuah modul.
3. Apa yang diketahui oleh modul tersebut.

Tabel 2.4 Simbol dan Fungsi Hierarchy Chart

Simbol	Keterangan
	Module Mengembangkan suatu modul (kumpulan perintah atau instruksi program).
	Connection Menghubungkan suatu modul dengan modul lain.
	Loop Menyatakan perulangan selama kondisi terpenuhi dalam suatu modul.
	Decision Simbol ini menunjukkan suatu penyeleksian kondisi di dalam modul.
	Couple Menunjukkan suatu data atau elemen kontrol yang dikirimkan dari suatu modul kemodul lainnya.

2.6 Visual Studio Code

Menurut Schroeder & Cure (2021) Visual Studio Code (VS Code) adalah editor kode sumber lintas platform yang dikembangkan oleh Microsoft. Sejak perilisannya pada tahun 2015, VS Code telah menjadi salah satu alat pengembangan perangkat lunak yang paling populer di kalangan pengembang. Tinjauan pustaka ini bertujuan untuk memberikan pemahaman yang komprehensif tentang fitur, kegunaan, dan manfaat VS Code dalam proses pengembangan perangkat lunak.

2.7 RapidMiner

RapidMiner merupakan salah satu perangkat lunak data mining yang populer dan kuat. RapidMiner menyediakan lingkungan yang intuitif dan visual untuk melakukan berbagai tugas data mining, termasuk pemrosesan data, transformasi, visualisasi, pembelajaran mesin, dan evaluasi model. Dengan antarmuka yang user-friendly, RapidMiner memungkinkan pengguna dengan berbagai tingkat keahlian untuk melakukan analisis data yang kompleks tanpa harus memiliki latar belakang teknis yang mendalam. Menurut Elisa (2017).

2.8 Basis Data

Basis data adalah susunan data yang diatur sedemikian rupa sehingga data-data tersebut memiliki kaitan dan mempermudah akses untuk mendapatkan informasi. Sebagai contoh, dalam basis data akademis, terdapat tabel-tabel yang saling berhubungan, mencakup data mahasiswa, data jurusan, data matakuliah, data pengambilan matakuliah dalam suatu semester, dan data nilai yang diperoleh oleh mahasiswa. Kurniati (2015), *Buku Ajar Basis Data Untuk Informatika*.

2.9 MySQL

Menurut Solichin (2016), “Pemrograman Web dengan PHP dan MySQL” merujuk pada sistem manajemen basis data SQL yang bersifat Open Source dan saat ini sangat populer. Sistem basis data MySQL mendukung berbagai fitur, termasuk multithreaded, multi-user, dan manajemen sistem basis data SQL (DBMS). MySQL merupakan software RDBMS (atau server database) yang dapat mengelola database dengan sangat cepat, dapat menampung data dalam jumlah sangat besar, dapat diakses oleh banyak user (multi-user), dan dapat melakukan suatu proses secara sinkron atau berbarengan (multi-threaded).

2.10 HTML

Menurut Moshinsky (1959), "*HTML & CSS The Complete Reference, Fifth Edition*" adalah bahasa markup yang digunakan untuk menyusun halaman web, kode dalam bentuk sekumpulan tag, masing-masing ditandai dengan < dan >. Tag bisa mewakili elemen halaman web, misalnya menyatakan elemen gambar serta pasangan <a> dan menyatakan *hyperlink*.

2.11 PHP

PHP atau kependekan dari Hypertext Preprocessor adalah salah satu Bahasa pemrograman open source yang sangat cocok untuk pengembangan web dan dapat di tanamkan pada sebuah skripsi HTML. Bahasa PHP dapat dikatakan mendeskripsikan beberapa Bahasa pemrograman seperti C, Java, dan Perl serta mudah untuk di pelajari. Pemrosesan data pada PHP ini dilakukan pada sisi server karena PHP merupakan Bahasa Scripting server.

PHP merupakan bahasa pemrograman berbasis skrip yang digunakan untuk memproses data dan mengirimkannya kembali ke peramban web dalam bentuk kode HTML. (2016), "*Pemrograman Web dengan PHP dan MySQL Achmad*".

2.12 Penelitian Terkait

- Dalam konteks penelitian ini, penulis mengidentifikasi permasalahan keterbaruan dalam penelitian sebelumnya yang dilakukan oleh Fitriani (2018) dengan judul "Penerapan Algoritma C4.5 untuk Seleksi Penerimaan Siswa Baru pada SD Islam Terpadu Permata Bunda Demak ". Meskipun penelitian tersebut memberikan temuan yang signifikan dan berharga, kami menemukan bahwa tingkat keterbaruan penelitian tersebut kurang memadai. Informasi yang diberikan dalam publikasi tersebut tidak

cukup rinci dan terstruktur sehingga sulit untuk mereproduksi metode dan proses yang digunakan.

- Perbedaan utama antara penelitian ini dan penelitian yang dilakukan oleh Fitriani(2018) adalah Dalam penelitian tersebut, pengujian menggunakan Algoritma C4.5 menghasilkan Accuracy 99.05% sedangkan penelitian ini mendapatkan nilai akurasi sebesar 100%. Dalam penelitian ini, penulis memberikan penekanan yang lebih besar pada aspek keterbaruan dengan memastikan bahwa metode, data, dan langkah-langkah analisis yang digunakan dalam penelitian ini sepenuhnya didokumentasikan dan dapat direproduksi oleh pihak lain. Hal ini memungkinkan peneliti lain untuk mengulangi penelitian ini dengan mudah dan memvalidasi temuan. Dalam rangka mencapai tingkat keterbaruan yang tinggi, kami menggunakan perangkat lunak RapidMiner, yang menyediakan antarmuka visual dan intuitif untuk analisis data. Penggunaan RapidMiner membantu penulis dalam membangun dan mereproduksi proses analisis data dengan lebih efisien, menjadikannya alat yang efektif dalam mencapai keterbaruan yang baik dalam penelitian ini.

Dengan memperhatikan tantangan keterbaruan yang dihadapi dalam penelitian sebelumnya dan memfokuskan pada keterbaruan sebagai salah satu tujuan utama penelitian ini, penulis berharap bahwa penelitian ini dapat memberikan kontribusi yang signifikan dalam memperbaiki tingkat keterbaruan dalam analisis data dan eksperimen komputasi.

Tabel 2.5 Penelitian Terkait

Penulis dan Tahun Terbit	Judul	Keterangan
(Husein & Hutaeruk, 2022)	Penerapan Algoritma C4.5 Dalam Pemilihan Siswa Berprestasi di SMPN 10 Medan	Melakukan perhitungan untuk mengklasifikasikan hasil probabilitas dari sample yang digunakan. Apabila perhitungan probabilitas telah selesai, maka tahap selanjutnya dilakukan pengujian dengan menggunakan tools data mining yaitu RapidMiner untuk mengetahui hasil klasifikasi dan memvalidasi hasil perhitungan sebelumnya. Pengujian data ini berhasil mendapatkan akurasi yang cukup. Dengan akurasi yang didapat adalah 56,17%, precision 55,28% dan recall 31.25%. Sementara penelitian ini melakukan seleksi siswa
(Khotimah, 2022)	Teknik Data Mining menggunakan Algoritma Decision Tree (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi	Berdasarkan hasil penelitian implementasi data mining menggunakan algoritma decision tree (C4.5) diperoleh nilai akurasi sebesar 100%. Uji akurasi juga dilakukan dengan algoritma Naïve Bayes untuk memperoleh perbandingan tingkat akurasi. Berdasarkan uji akurasi dua algoritma diperoleh data tingkat akurasi sebesar 100% pada algoritma Decision Tree (C4.5) dan 90,16 % pada algoritma Naïve bayes. Hal ini dapat disimpulkan akurasi algoritma Decision Tree (C4.5) untuk prediksi calon mahasiswa penerima beasiswa jalur KIP lebih baik dibandingkan algoritma Naïve bayes. Sementara dalam penelitian ini dilakukan pengujian decision tree menggunakan aplikasi RapidMiner untuk melakukan seleksi terhadap siswa baru.
(Ikhtal & Irfan, 2021)	Menentukan Penjurusan Siswa Dengan Menggunakan Metode <i>Decision Tree</i> Algoritma	Dalam menentukan penulis penjurusan siswa selama ini belum diterapkannya proses klasifikasi jurusan dengan akurasi nya juga belum diketahui, penelitian ini bertujuan untuk mengklasifikasi penjurusan siswa dan

	C4.5 (Studi Kasus : SMA Negeri 2 Padang)	menerapkan metode data mining menggunakan algoritma decision tree C4.5. Klasifikasi ini menggunakan 6 atribut sebagai atribut kontrol yang merupakan nilai rata-rata rapor SMP yang meliputi mata pelajaran B. Indonesia, B. Inggris, Matematika dan IPA, serta rekomendasi guru BK SMP. Berdasarkan pengujian yang dilakukan terhadap siswa tahun masuk ajaran 2020/2021 diperoleh akurasi sebesar 68.42% dengan 304 sampel data dengan atribut rekomendasi guru BK SMP yang mempengaruhi keputusan penjurusan siswa. Demikian dapat disimpulkan bahwa menentukan penjurusan dengan menggunakan data mining algoritma decision tree C4.5 dapat mempercepat dan akurat dalam pengambilan keputusan penjurusan siswa SMA N 2 Padang. Sementara dalam penelitian ini dilakukan pengujian decision tree menggunakan aplikasi RapidMiner untuk melakukan seleksi terhadap siswa baru.
(Arfida, 2020)	Penerapan Metode Topsis Dalam Sistem Pendukung Keputusan Untuk Menentukan Pemenang Lomba Desa/Kelurahan	Penulis melakukan penilaian perlombaan dilakukan dengan membandingkan tingkat perkembangan 2 (dua) tahun terakhir berdasarkan data profil desa/kelurahan sesuai dengan indikator penilaian. Sistem pendukung keputusan lomba desa/kelurahan dapat mengatasi kelemahan dan kekurangan dari pelaksanaan lomba sebelumnya. Sistem yang baru ini dapat dilakukan dengan cepat, terbuka dan kompetitif meskipun jumlah data relatif banyak tetapi keakuratan perhitungan serta laporan dapat dicapai semaksimal mungkin. Sehingga efisiensi waktu dalam pengerjaan dan penyelesaian laporan akan menjadi lebih baik. Aplikasi pendukung keputusan untuk menentukan pemenang lomba Desa/Kelurahan menggunakan perhitungan metode Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Sementara

		penelitian ini melakukan seleksi siswa. baru menggunakan Algoritma C4.5 dimana tingkat akurasi lebih baik.
(Syuryanny Nurindah & Syahdan, 2022)	Penerapan Algoritma C4.5 Dalam Klasifikasi Penerimaan Peserta Didik Baru	Hasil analisis data dengan menggunakan metode Algoritma C4.5 menunjukkan bahwa atribut yang paling mempengaruhi dalam penerimaan peserta didik baru adalah nilai tes matematika yang memberikan kontribusi terbesar selanjutnya yaitu nilai tes bahasa inggris dan IPA serta yang memiliki kontribusi terkecil dalam pembentukan pohon keputusan yaitu AS (Asal Sekolah). Berdasarkan analisa hasil pengujian yang dilakukan dengan tingkat akurasi 90.76 %, presisi 94,20 % dan recall 92,85 %. Hal ini berarti algoritma klasifikasi C4.5 dapat diimplementasikan dengan baik dalam penerimaan peserta didik baru. Sementara dalam penelitian ini melakukan seleksi terhadap siswa baru.