

BAB II

TINJAUAN PUSTAKA

2.1. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sudah banyak di kenal 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 diskrit dari record yang baru. (Elisa, 2017)

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 pada cabang memiliki kelas yang sama.

Untuk menghitung nilai gain digunakan rumus seperti dalam persamaan berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (1)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi atribut A

$|S_i|$ = Jumlah kasus pada partisi ke-i

$|S|$ = Jumlah kasus dalam S

Untuk menghitung nilai Entropy, digunakan rumus seperti dalam persamaan berikut:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (2)$$

Keterangan:

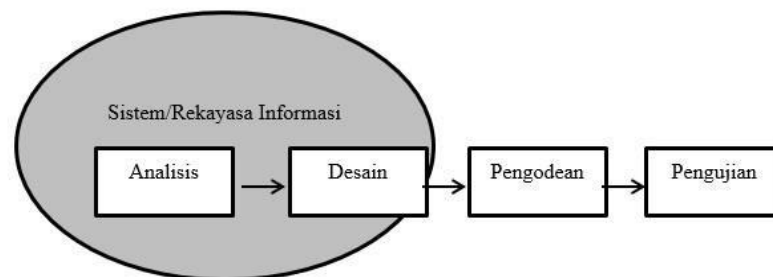
S = Himpunan kasus

n = Jumlah Partisi S

p_i = Proporsi S_i terhadap S

2.2. Metode Waterfall

Metode waterfall atau metode air terjun adalah metode yang menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau terurut dimulai dari analisis, desain, pengodean dan pengujian. (Yopi Handrianto, 2020). Berikut adalah gambar dari metode waterfall dalam SDLC:



Gambar 2.1 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 adalah proses beberapa tahapan Langkah pada rancangan pembuatan program perangkat lunak meliputi struktur data, arsitektur perangkat lunak, representasi antarmuka dan prosedur pengodean. Tahap ini mentranslasi kebutuhan perangkat lunak dari tahapan analisis kebutuhan ke representasi rancangan agar dapat diimplementasikan menjadi program pada tahap selanjutnya. Pada tahap ini, hasil dari desain perangkat lunak yang telah ada di dokumentasikan.

c. Pengkodean

Desain harus di implementasikan kedalam program perangkat lunak. Hasil dari tahap ini adalah program computer sesuai desain yang telah dibuat pada tahap desain. Atau tahapan penulis membuat program dengan Bahasa program seperti PHP, HTML, SCC dan lain-lain.

d. Pengujian

Tahapan ini penulis melakukan mengujian terhadap program yang telah dibuat untuk mengetahui kekurangan dari program tersebut.

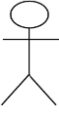




2.3. Usecase Diagram

Usecase Diagram adalah diagram yang menggambarkan actor, usecase dan relasinya sebagai suatu urutan Tindakan yang memberikan nilai terukur untuk actor. Sebuah usecase digambarkan sebagai elips horizontal dalam suatu Diagram UML Usecase. (Haviluddin, 2011). Usecase memiliki dua istilah:

1. System Usecase, interaksi dengan sistem.
2. Business usecase, interaksi bisnis dengan konsumen atau kejadian nyata

Berikut adalah penjelasan simbol-simbol usecase yang terdapat pada tabel 2.1.

Tabel 2.1 Simbol-Simbol Usecase



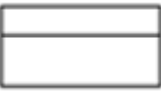




Simbol	Keterangan
 Aktor	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan usecase
Include 	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (dependent) akan memengaruhi elemen yang bergantung padanya elemen yang tidak mandiri (independent)
Association 	Apa yang menghubungkan antara objek satu dengan objek yang lainnya
Sistem 	Menspesifikasikan paket yang menampilkan sistem secara terbatas
Usecase 	Mendesripsikan urutan aksi-aksi yang di tampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor

2.4. Class Diagram

Class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi. Atribut merupakan variabel-variabel yang dimiliki oleh suatu kelas, sedangkan operasi atau metode adalah fungsi-fungsi yang dimiliki oleh suatu kelas (Rosa dan Shalahuddin, 2016).

Adapun simbol-simbol Classdiagram dapat dilihat pada Tabel 2.2.

Tabel 2.2. Simbol-Simbol Class Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<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>).
2		<i>N-Ary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
3		<i>Class</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
4		<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
5		<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek.
6		<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.
7		<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.

2.5. Basis Data

Basis data adalah suatu pengorganisasian sekumpulan data yang saling terkait sehingga memudahkan aktivitas untuk memperoleh informasi. Sebagai contoh, basis data akademis mengandung tabel-tabel yang berhubungan dengan data mahasiswa, data jurusan, data matakuliah, data pengambilan data matakuliah pada suatu semester, dan data nilai yang diperoleh mahasiswa (Kadir and Triwahyuni, 2014).

2.6. MySQL

MySQL adalah sistem manajemen database SQL yang bersifat Open Source dan paling populer saat ini. Sistem database MySQL mendukung beberapa fitur seperti multithreaded, multi-user, dan SQL database management system (DBMS) (Madcoms, 2016).

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 bersamaan (multi-threaded) (Raharjo, 2015).

2.7. HTML

HTML adalah bahasa markah yang digunakan untuk menyusun halaman web, kodenya berupa kumpulan tag-tag, dengan setiap tag ditandai dengan < dan diakhiri dengan >. Tag bisa mewakili elemen halaman web, misalnya menyatakan elemen gambar serta pasangan <a> dan menyatakan *hyperlink* (Kadir and Triwahyuni, 2014)

2.8. 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 adalah suatu Bahasa pemrograman berbasis kode-kode (script) yang

digunakan untuk mengolah suatu data dan mengirimkannya Kembali ke web browser menjadi kode HTML. (Astria Firman, 2016).

2.9. Penelitian Terkait

Berikut adalah tabel penelitian terkait yang bisa dilihat pada tabel 2.3 .

Tabel 3.3 Penelitian Terkait

Penulis	Tahun Terbit	Judul	Keterangan
Erlin Elisa	2017	Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT. Arupadhatu Adisesanti.	Banyaknya kecelakaan kerja yang terjadi tidak terlepas dari faktor Human Error, tentunya berdampak pada kinerja dan pekerjaan yang dilaksanakan, Metode yang digunakan dalam analisis ini adalah Algoritma C4.5 yang merupakan salah satu algoritma modern untuk melakukan Data Mining, Algoritma C4.5 disebut juga dengan pohon keputusan (decision tree) yaitu merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon, dan pada setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas, Konsep dari pohon keputusan ini adalah dengan mengumpulkan data selanjutnya dibuatkan decision tree yang kemudian akan dihasilkan rule-rule solusi permasalahan. dari hasil penelitian faktor-faktor yang menjadi penyebab terjadinya kecelakaan kerja kontruksi yang sering terjadi adalah Lingkungan Tempat Kerja, Rambu-Rambu Keselamatan dan Pekerja dan Cara kerja
Nurul Azwanti	2018	Analisa Algoritma C4.5 untuk Memprediksi Penjualan Motor Pada PT. Capella	Di Indonesia, motor yang paling banyak diminati adalah motor merk Honda. Wilayah penjualan sangat mempengaruhi terhadap penjualan motor, selain ke showroom langsung, di beberapa tempat ramai seperti mall bisa menjadi promosi dan penjualan

		Dinamik Nusantara Cabang Muka Kuning.	<p>motor karena mudah di jangkau oleh masyarakat.</p> <p>Namun jumlah data yang sangat banyak akan sangat sulit untuk di Analisa. Analisa di perlukan untuk melihat pola dari penjualan hingga dapat menghasilkan prediksi penjualan motor yang nantinya akan berguna untuk pendistribusian motor di beberapa wilayah.</p> <p>Maka dari itu penelitian ini menggunakan algoritma C4.5 untuk mendapatkan sebuah pohon keputusan untuk melihat pola prediksi perilaku konsumen dalam membeli motor.</p>
Nurul Azwanti	2018	Algoritma C4.5 untuk Memprediksi Mahasiswa yang Mengulang Mata Kuliah.	<p>Penelitian ini menggunakan algoritma Algoritma C4.5 dengan metode pohon keputusan dengan harapan dapat memberikan informasi rule prediksi untuk menggambarkan proses yang terkait dengan prediksi mahasiswa yang mengulang. Karakteristik data yang diklasifikasi dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun aturan sehingga dalam tahap pengujian dengan software WEKA dapat membantu dalam memprediksi mahasiswa yang mengulang mata kuliah. Hal ini meminimalisir kesalahan dalam memprediksi jumlah mahasiswa yang akan mengulang mata kuliah dan akan berdampak pada tidak tersedianya kelas dan pihak akademik harus mengatur ulang bagi mahasiswa yang tidak mendapatkan kelas</p>
Victor Saputra Ginting & Kusri & Emha Taufiq	2020	Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan	<p>Tidak dapat dipungkiri, SPP Sekolah merupakan salah satu faktor penting yang digunakan untuk mealokasi biaya pembangunan sekolah, biaya untuk guru, karyawan, dan lain-lain. Biaya SPP Sekolah ini umumnya diterapkan oleh sekolah swasta yang dibebankan kepada siswanya, berbeda dengan sekolah negeri, yang biaya pengelolaan sekolah masih ada bantuan dari biaya pemerintah.</p>

		Python.	Berdasarkan pemaparan diatas, perlu adanya sebuah penelitian untuk memprediksi keterlambatan pembayaran SPP Sekolah yang dilakukan murid. Data di ambil dari dataset penelitian terdahulu oleh (Muqorobin et al., 2019) sebanyak 30 data dengan variable penghasilan orang tua, tanggungan keluarga, Pendidikan ayah, umur ayah, Pendidikan ibu, umur ibu dan keterangan sebagai class kemudian di implementasikan kedalam sebuah pemograman dengan menggunakan python untuk menghasilkan keterangan hasil prediksi.
Asroni & Badrahini Masajeng Respati & Slamet Riyadi	2018	Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta.	Hal yang dapat dijadikan sebagai tolok ukurnya adalah jenis pekerjaan yang diperoleh alumni setelah lulus, tetapi masalahnya belum dilakukan pengolahan pada data alumni dengan menggunakan metode data mining. Klasifikasi adalah suatu teknik yang dilakukan untuk memprediksi class atau properti dari setiap instace data. Teknik ini dilakukan pada data baru dengan memanipulasi data yang ada dan telah diklasifikasikan menggunakan hasilnya untuk memberikan sejumlah aturan. Teknik klasifikasi yang di gunakan dalam memprediksi adalah Algoritma C4.5 Peneliti mendapatkan data dari database server BSI (biro sistem informasi) universitas Muhammadiyah Yogyakarta..

2.10. Kerangka Pemikiran

Salah satu hal terpenting dalam akademik adalah dengan memantau kelayakan mahasiswa untuk mengambil mata kuliah Tugas Akhir atau Skripsi. Dengan algoritma C45 maka dapat diambil keputusan dimana mahasiswa yang dapat atau layak untuk mengambil skripsi.

2.11. Hipotesa

2.11.1. Mencari Nilai Entropy Pada Seluruh Data

Untuk memulai dari perhitungan ini maka dicari nilai entropy pada seluruh data dengan perhitungan sebagai berikut:

1. Tentukan jumlah seluruh data

Jumlah kasus: 28

Jumlah kasus dengan keputusan “tidak”: 24

Jumlah kasus dengan keputusan “ya”: 4

2. Cari Entropy “Tidak” untuk seluruh data

$$= \frac{\text{Jumlah kasus "tidak"}}{\text{Jumlah Seluruh kasus}} * \text{Log}_2 \frac{\text{Jumlah kasus "tidak"}}{\text{Jumlah Seluruh kasus}}$$

$$= \frac{24}{28} * \text{Log}_2 \frac{24}{28} = 0,1906$$

3. Cari Entropy “Ya” untuk seluruh data

$$= \frac{\text{Jumlah kasus "ya"}}{\text{Jumlah Seluruh kasus}} * \text{Log}_2 \frac{\text{Jumlah kasus "ya"}}{\text{Jumlah Seluruh kasus}}$$

$$= \frac{4}{28} * \text{Log}_2 \frac{4}{28} = 0,401$$

4. Mencari total entropy untuk seluruh data

Entropy “tidak” + Entropy “ya”

$$= 0,1906 + 0,401$$

$$= 0,5917$$

2.11.2. Mencari Nilai Gain pada atribut jumlah SKS

Pada tahap selanjutnya dilakukan perhitungan untuk mencari Nilai Gain pada atribut jumlah SKS, dengan perhitungan sebagai berikut:

1. Entropy tidak untuk atribut jumlah SKS dengan Kriteria lebih dari 124:

$$= \frac{\text{Jumlah kasus tidak } (>124)}{\text{Jumlah Seluruh kasus } (>124)} * \text{Log}_2 \frac{\text{Jumlah kasus tidak } (>124)}{\text{Jumlah Seluruh kasus } (>124)}$$

$$= \frac{12}{14} * \text{Log}_2 \frac{12}{14} = 0,1906$$

2. Entropy tidak untuk atribut jumlah SKS dengan Kriteria kurang dari 124:

$$= \frac{\text{Jumlah kasus tidak } (<124)}{\text{Jumlah Seluruh kasus } (<124)} * \text{Log}_2 \frac{\text{Jumlah kasus tidak } (<124)}{\text{Jumlah Seluruh kasus } (<124)}$$

$$= \frac{12}{14} * \text{Log}_2 \frac{12}{14} = 0,1906$$

2.11.3. Mencari “Ya” untuk atribut jumlah SKS

Langkah terakhir dengan mencari “Ya” untuk atribut jumlah SKS dengan kriteria lebih dari 124, dengan perhitungan sebagai berikut:

$$= \frac{\text{Jumlah kasus ya } (>124)}{\text{Jumlah Seluruh ya } (>124)} * \text{Log}_2 \frac{\text{Jumlah kasus ya } (>124)}{\text{Jumlah Seluruh ya } (>124)}$$

$$= \frac{2}{14} * \text{Log}_2 \frac{2}{14} = 0,401$$