

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Pengertian Data Mining

Mengutip (Tan, 2006 dalam Eko Prasetyo, 2013:2) menyatakan bahwa data mining adalah proses memperoleh informasi bermanfaat yang berasal pada arsip database yang besar.

KDD yaitu teknik keterhubungan dan proses ilmu pengetahuan, pandangan serta penggambaran yang berasal pada pola beberapa *information*. Berikut merupakan tahapan pada compositions tersebut tahap (Tan,2006) :

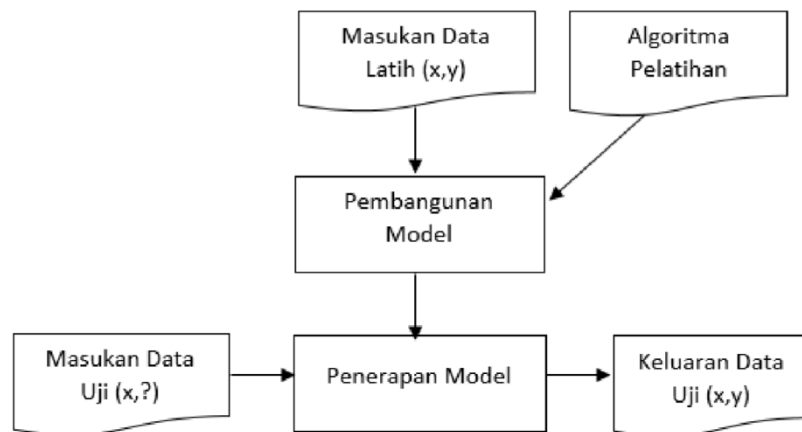
- a. Pembersihan data yaitu untuk menghapus data yang tidak serasi.
- b. Integrasi data yaitu data yang digabungkan pada beberapa sumber.
- c. Transformasi data yaitu perubahan data agar sesuai dengan mining.
- d. Aplikasi teknik data mining yaitu meringkas pola berdasarkan data sebelumnya.
- e. Evaluasi pola yang ditemukan yaitu kegiatan untuk mengambil pandangan seseorang terhadap pola yang sudah dibuat guna mengambil keputusan.
- f. Presentasi pengetahuan yaitu menyampaikan mengenai hal-hal yang telah selesai dilakukan.

Tahap ini yaitu bagian dari *expositions* dengan mencari pengetahuan dengan memeriksa pola yang didapatkan apakah berbeda dengan fakta. Kemudian langkah terakhir KDD yaitu menyampaikan pengetahuan yang dapat dimengerti oleh pengguna (Tan, 2006).

2. Pengertian Klasifikasi

Klasifikasi yaitu pekerjaan dalam menilai objek untuk menggolongkan informasi termasuk ke bagian kelas yang sudah pasti dari beberapa kelas yang ada. Pengklasifikasi dapat didefinisikan secara rinci yaitu kegiatan pelatihan pada fungsi target "x" pada salah satu dari beberapa label kelas yang ada "y". Kegiatan pelatihan tersebut akan memperoleh pola untuk disimpan pada memori (Eko Prasetyo, 2013:45).

Pola yang didapatkan berguna untuk memperkirakan label kelas pada data yang baru dengan menerapkan algoritma.



Gambar 2. 1 Proses Pekerjaan Klasifikasi

(Sumber : Eko Prasetyo, 2013)

Berikut merupakan algoritma pelatihan yang sudah dikembangkan oleh para peneliti yaitu C4.5, Nearest Neighbor, Bayesian Classification, Neural Network, dan sebagainya (Eko Prasetyo, 2013:46).

3. Pengertian Pengembangan SDLC

Pendekatan sistem adalah metodologi. Metodologi adalah praktik yang disarankan. Siklus Hidup Pengembangan Sistem (SDLC) yaitu aplikasi yang berasal dari pendekatan sistem untuk pengembangan sistem informasi (Raymond McLeod, 2007, p.200).

Berikut adalah tahapan SDLC :

- a. Rencana
- b. Analis
- c. Design
- d. Penerapan
- e. Menggunakan

Proyek dan sumber daya yang diperlukan dalam melaksanakan pekerjaan direncanakan dan dirakit. Pada sistem sebelumnya juga dilakukan analisis guna mengetahui permasalahan apa saja yang menjadi kendala selama ini, serta menetapkan kebutuhan-kebutuhan pada sistem baru yang akan dikembangkan. Langkah berikutnya itu merancang terhadap sistem baru yang akan digunakan, kemudian penerapan algoritma, kemudian sistem akan dilakukan uji coba penggunaan untuk mengetahui apakah sistem sudah sesuai dengan yang diharapkan, dan pengembangan sistem baru lebih baik dapat digunakan dalam jangka waktu yang lama.



Gambar 2. 2 Pola Melingkar dari Siklus Hidup Sistem

(Sumber: Raymond McLeod, 2007, p.200)

Gambar 2.2 menggambarkan sifat melingkar dari siklus hidup. Ketika sebuah sistem telah melewati masa manfaatnya dengan ditemukan berbagai kekurangan di dalamnya maka perlu diperbarui.

Masalah akan diidentifikasi selama fase desain dan analisis. Selama fase desain, alternatif diidentifikasi dan dievaluasi. Kemudian solusi terbaik diambil dan digunakan. Selama fase menggunakan sistem, hasil uji coba dikumpulkan untuk mengetahui bagaimana sistem baru dapat memecahkan masalah yang dihadapi.

4. Pengertian Metode Prototype

Menurut (Ogedebe, 2012) bahwa *prototyping* adalah metode pengembangan perangkat lunak. Dengan metode *prototyping* dapat memperoleh sebuah sistem *prototype* yang berfungsi menjadi perantara antara pengguna dan pengembang. Kemungkinan yang terjadi pada *prototype* yaitu adanya perubahan pada komponen-komponen tertentu berdasarkan perencanaan serta analisis dari pengembang hingga dilakukan pengujian

Ada metode utama pembuatan prototipe antara lain :

1. Illustrative, membuat tampilan layar.
2. Simulated, uji coba alur kerja sistem bukan pada data real.
3. Functional, uji coba alur kerja sistem berdasarkan data real.
4. Evolutionary, pola yang dihasilkan dan menjadi operasional sistem.

Prototyping dapat dibuat pada sistem yang besar maupun sistem kecil yang akan dikembangkan sesuai dengan yang diinginkan yaitu sistematis

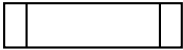
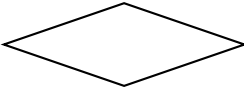




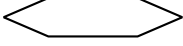
dan selesai tepat pada waktunya. Manfaat lain menggunakan prototipe adalah:

1. Menjalankan sistem sebenarnya dengan mengumpulkan masukan dari pengguna demi menghasilkan sistem yang sempurna.
2. Prototype akan mengalami perubahan sampai hasil akhir pengembangan, dan pengguna menjadi lebih siap menerima hal tersebut.
3. Prototype akan mengalami penambahan ataupun pengurangan namun sesuai dengan proses pengembangan yang diketahui oleh pengguna.
4. Menggunakan sumber daya dan waktu semaksimal mungkin guna menghasilkan produk yang diinginkan oleh pengguna.

5. Pengertian Flowchart

Sebuah diagram alur adalah gambar mengenai urutan-urutan pada program. Urutan tersebut menggunakan berbagai simbol tertentu. (Yuslena Sari, 2017:54).

Tabel 2. 1 Simbol Flowchart

SIMBOL	FUNGSI
	
	
	
	
	
	
	

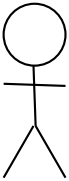
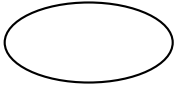
	
	




(Sumber : Sari, 2017)

6. Pengertian Unified Modeling Language (UML)

Menurut (Sri Dharwiyanti, 2003) menjelaskan bahwa Unified Modeling Language (UML) adalah "bahasa" yang digunakan untuk memvisualisasikan, desain dan mendokumentasikan sistem perangkat lunak.


Tabel 2. 2 Simbol Usecase Diagram

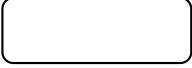
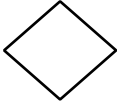
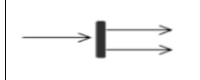

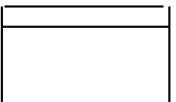
SIMBOL	NAMA	KETERANGAN
	<i>ACTOR</i>	
	<i>USECASE</i>	

	ASOSIASI/ ASSOCIATION	
<<extend>>	EKSTENSI/ EXTEND	
	GENERALISASI / GENERALIZATION	
<<include>> 	MENGGUINAKAN INCLUDE	

(Sumber : Sri, 2003)



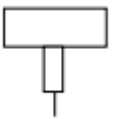
Tabel 2. 3 Simbol Activity Diagram

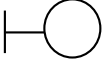
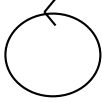
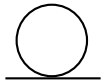



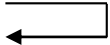
SIMBOL	NAMA	KETERANGAN
	STATUS AWAL/INITIAL	

	<i>AKTIVITAS/ ACTIVITY</i>	
	<i>PERCABANGAN/ DECISION</i>	
	<i>PENGGABUNGAN/ JOIN</i>	
	<i>STATUS AKHIR/FINAL</i>	
	<i>SWIMLINE</i>	

(Sumber : Sri, 2003)

Tabel 2. 4 Simbol Sequence Diagram



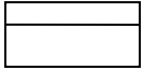

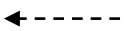
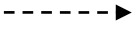

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	
	Lifeline	
	General	

	<i>Boundary</i>	
	<i>Control</i>	
	<i>Entitas</i>	
	<i>Activation</i>	
	<i>Message Entry</i>	
	<i>Message to Self</i>	
	<i>Message Return</i>	

(Sumber : Sri, 2003)


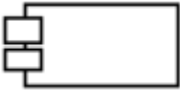
Tabel 2. 5 Simbol Class Diagram



SIMBOL	NAMA	KETERANGAN
---------------	-------------	-------------------

	<i>Generalization</i>	
	<i>Nary Association</i>	
	<i>Class</i>	
	<i>Collaboration</i>	
	<i>Realization</i>	
	<i>Dependency</i>	
	Association	

(Sumber : Sri, 2003)

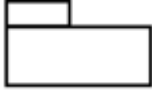
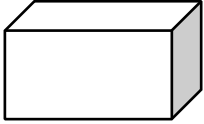

Tabel 2. 6 Simbol Component Diagram

SIMBOL	NAMA	KETERANGAN
	Package	
	Komponen	

	Kebergantungan/ dependency	
	Link	

(Sumber : Sri, 2003)

Tabel 2. 7 Simbol Deployment Diagram

SIMBOL	NAMA	KETERANGAN
	Package	
	Node	
	Kebergantungan/ dependency	

(Sumber : Sri, 2003)

7. Pengertian Database

Menurut (Sutabri, 2012:47) database adalah kumpulan data terkait dan di simpan pada perangkat keras komputer atau di sebut hardware dan di manipulasi oleh perangkat lunak atau software. Data harus disimpan dalam database untuk memberikan informasi tambahan.

8. Pengertian Webserver (Apache)

Menurut (Fadhla Binti Junus, 2020:19) menjelaskan bahwa tugas utama dari webserver adalah menerima (response) permintaan (request) dari client (browser). Request tersebut merupakan konten statis berupa halaman

HTML yang dikirimkan oleh browser melalui Hypertext Transfer Protocol (HTTP). Webserver menanggapi request dengan mengirimkan kembali jawaban ke browser melalui HTTP.

9. Pemrograman

a. Pengertian PHP

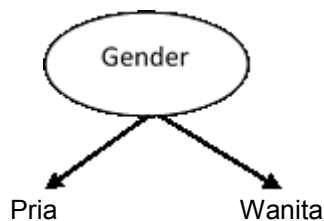
Menurut (Ahmad Solichin, 2016:11) menjelaskan bahwa PHP adalah bahasa pemrograman berbasis website yang di tulis oleh pengembang web untuk mengembangkan sebuah website.

b. Pengertian HTML

Menurut (Ahmad Solichin, 2016:10) menjelaskan bahwa "*Hypertext Markup Language (HTML)* adalah bahasa pemrograman web yang memberi tahu browser web cara menampilkan konten di halaman web."

B. Metode Algoritma C4.5

Algoritma C4.5 merupakan salah satu Algoritma klasifikasi, penggunaannya dapat menggunakan data atribut kategorial maupun angka. Algoritma c4.5 akan menghasilkan aturan yang dapat dipakai dalam memprediksi dari record baru. Algoritma C4.5 adalah perkembangan dari algoritma ID3, dimana pengembangan dilakukan pada kemampuan untuk memperbaiki *missing* data, bisa mengatasi data kontinu dan *pruning* (Larose, 2005).



Gambar 2. 3 Syarat pengujian fitur biner

Yang menjadi hal penting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node:

a. Fitur biner

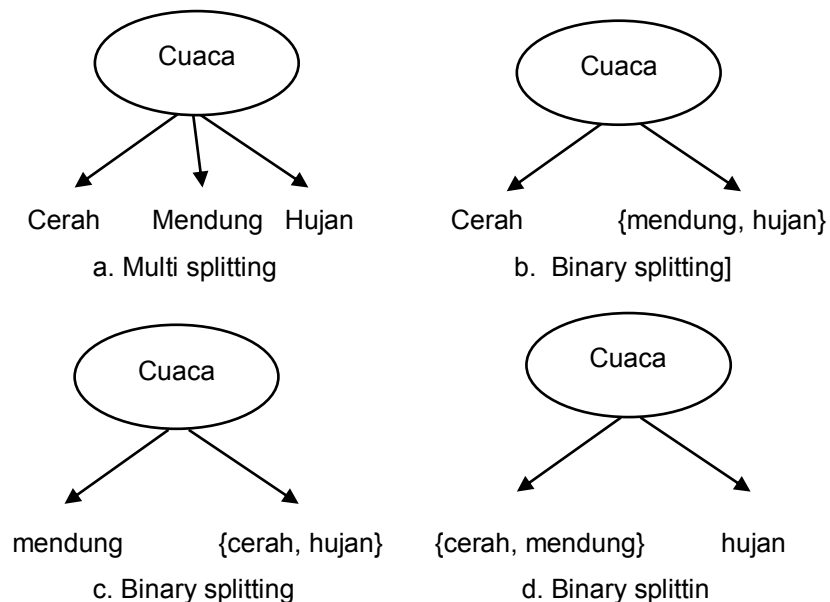
Karakteristik yang hanya memiliki dua nilai yang berbeda disebut karakteristik biner. Kondisi pengujian ketika fungsi ini menjadi simpul (root atau internal) hanya memiliki dua opsi cabang. Contoh pemecahannya disajikan pada Gambar 2.3.

b. Fitur bertipe kategorikal

Untuk Fitur dengan nilai kategorikal (nominal atau ordinal), mereka dapat mengasumsikan nilai yang berbeda. Contohnya adalah fitur 'cuaca' memiliki 3 nilai berbeda dan dapat memiliki banyak kombinasi kodisi pengujian pemecahan. Secara umum ada 2, yaitu pembagian biner (*binary splitting*) dan (*multi splitting*). Kombinasi tersebut disajikan seperti pada Gambar 2.4. Untuk solusi yang hanya memungkinkan pemecahan biner, seperti algoritma CART, maka akan memberikan sejumlah kombinasi pemecahan sebanyak 2^{k-1} , di mana k adalah jumlah nilai lain yang khas dalam fitur. Contoh pemecahannya disajikan pada Gambar 2.4 dan Gambar 2.5.

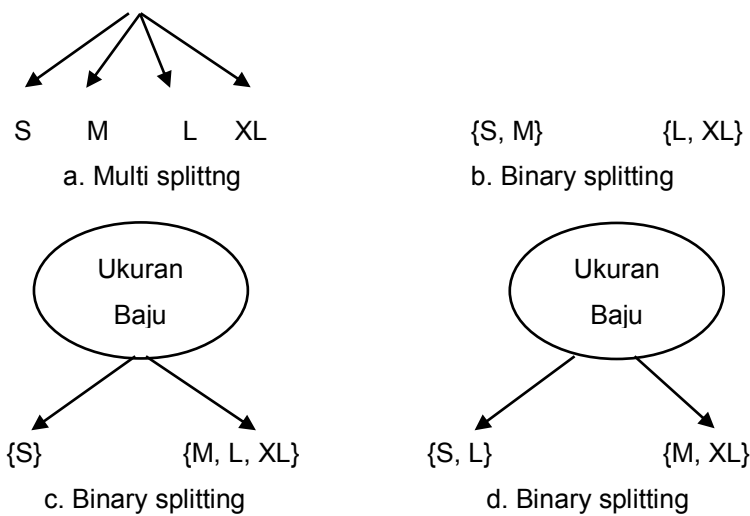
c. Fitur bertipe numerik

Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan tes perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk $v_i \leq A < v_{i+1}$, 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 multimetode, maka algoritma harus memeriksa semua kemungkinan jangkauan nilai *kontinyu*. Contoh pemecahan pada fitur numerik disajikan pada Gambar 2.6.

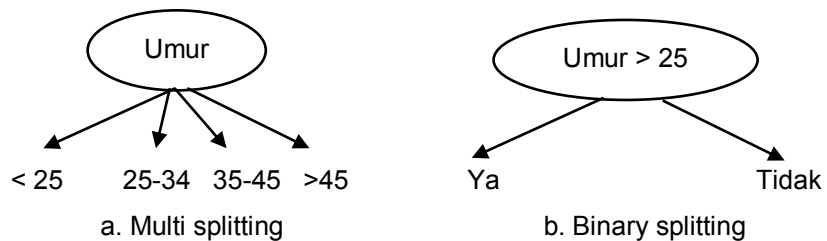


Gambar 2. 4 Syarat pengujian fitur bertipe nominal





Gambar 2. 5 Syarat pengujian fitur bertipe ordinal



Gambar 2. 6 Syarat pengujian fitur bertipe numerik

Kriteria sebagai pemecah dalam algoritma C4.5 adalah rasio gain, yang dirumuskan oleh persamaan berikut:

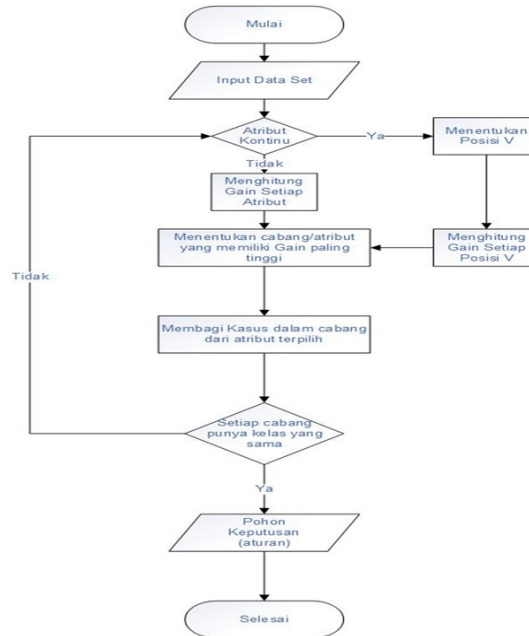
$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)}$$

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 .

Berikut ini adalah algoritma C4.5 menggunakan flowcart yang disajikan pada Gambar 2.7 dan berserta penjelasannya :

- Memasukan Data Set yang telah disediakan. Data Set adalah kumpulan objek dan atributnya.
- Pada proses cek keputusan dilihat apakah data set tersebut termasuk Atribut Kontinu atau bukan, Atribut Kontinu adalah atribut yang mempunyai jangkauan 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 cabang dari atribut yang terpilih.

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



Gambar 2. 7 Flowchart Algoritma C4.5

Sumber : (Larose, 2005)

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. Yaitu:

1. **“SISTEM PREDIKSI DAN EVALUASI PRESTASI AKADEMIK MAHASISWA DI PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN DATA MINING”**, Program Studi Teknik Informatika Politeknik Negeri Pontianak, Indonesia. Oleh Freska Rolansa, Yunita Yunita, Suheri Suheri (2020).

Dari hasil penelitian diperoleh bahwa sistem prediksi dan evaluasi prestasi akademik mahasiswa di Program Studi Teknik Informatika Politeknik Negeri Pontianak menggunakan data mining dengan algoritma regresi linear dapat memberikan prediksi nilai IP pada semester 5 sebagai data uji (angkatan 2017) berdasarkan data training pada angkatan 2016. Evaluasi yang dilakukan menggunakan algoritma C45 memiliki tingkat akurasi yang tinggi yakni 94,44% dan diperoleh IP pada semester 6 paling menentukan dalam memperoleh kategori 5. Dari evaluasi yang dilakukan

juga didapat 41 mahasiswa dari hasil klasifikasi tergolong kategori 5, 58 mahasiswa tergolong kategori 4, 6 mahasiswa tergolong kategori 3 dan hanya 2 mahasiswa tergolong klasifikasi 2, dan tidak ada mahasiswa yang tergolong kategori 1.

2. **“IMPLEMENTASI ALGORITMA C4.5 UNTUK MENENTUKAN PENERIMA BEASISWA DI STT HARAPAN MEDAN”**, Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan Medan. Oleh Rismayanti (2016).

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan. Hasil analisis data siswa STT Harapan Medan, terdapat 4 atribut yang digunakan sebagai atribut prediktif: Semester, Indeks Prestasi Kumulatif, Pendapatan Orang Tua (PO) dan Jumlah Tanggungan Orang Tua (JTO). Penetapan penerima beasiswa berujung pada penetapan bahwa mahasiswa yang “diterima” sebagai penerima beasiswa adalah mahasiswa dengan IPK Cum Laude dan IPK sangat baik dengan jumlah mahasiswa yang banyak, tanggungan dan orang tua “ditolak” karena beasiswa tersedia . adalah siswa dengan IPK yang baik dan IPK yang sangat baik ketika ada cukup tanggungan orang tua.

3. **“PENERAPAN METODE DECISION TREE ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA BERPRESTASI”**, Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan Medan. Oleh Nadiya Hijriana dan Riadhul Muttaqin, (2016).

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan. Hasil rangking siswa berprestasi 2012 yang diperoleh tidak hanya dapat digunakan sebagai tolak ukur untuk pengolahan data selanjutnya, seperti pengolahan data siswa beasiswa, tetapi juga dapat digunakan sebagai alat bagi tenaga pengajar untuk memantau hasil, salah satunya adalah memantau siswa yang mengikuti program tersebut. Mengklasifikasikan mahasiswa yang berpeluang lulus dengan IPK dengan predikat “Kurang” selanjutnya dapat menurunkan kualitas lulusan suatu universitas..

4. **“IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PRESTASI SISWA”**, Program Studi Teknik Informatika Fakultas Teknik Universitas Nusantara PGRI Kediri, (2017).

Berdasarkan hasil penelitian yang telah dilakukan dapat ditarik kesimpulan. Penelitian ini menggunakan data siswa kelas X, XI dan XII tahun pelajaran 2015/2016. Dari data siswa tersebut dijadikan sebagai data latih, dijadikan acuan dalam perhitungan algoritma C4.5. Data tes yang digunakan adalah data kelas X tahun pelajaran 2016/2017 dengan jumlah siswa sebanyak 153 siswa, dari hasil tes di atas akurasi = 83,66%.

5. **“PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN PENERIMA BEASISWA (Studi Kasus: SDN Karawang Kulon III)”**, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang, (2018).

Pengujian algoritma C4.5 diukur terhadap nilai akurasi, presisi dan tingkat kesalahan. Uji kinerja dengan menerapkan information gain dengan 7 kriteria yaitu: JTO, nilai, jarak, hunian, JPO, penerima KIP, penerima KPS menghasilkan nilai koreksi 93,58%, akurasi 95,5%, tingkat error 6,14% karena memiliki nilai akurasi prediksi benar tertinggi dan persentase terendah dari total data prediksi buruk. Hasil ini merupakan model terbaik dibandingkan skenario lain untuk memprediksi mahasiswa doctoral di SDN Karawang Kulon III, karena semua algoritma klasifikasi mencoba melatih model dengan akurasi tinggi. (Error Rate yang rendah).
6. **“MEMPREDIKSI PRESTASI SISWA DENGAN PENERAPAN ALGORITMA C4.5 DI SEKOLAH DASAR NEGERI 1 RAWA LAUT”**, Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, (2019).

Penelitian ini membuat prediksi keunggulan siswa di SD Negeri 1 Rawa Laut Bandar Lampung studi kasus, dengan kriteria status ekonomi, disiplin dan kehadiran, mungkin membantu sekolah membuat keputusan tentang pemberian beasiswa. Prediksi di atas berupa nilai siswa dan pencatatan data yang belum terkomputerisasi, sehingga dapat terjadi redundansi dan manipulasi data prestasi akademik..
7. **“Penerapan Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar (Studi Kasus : SD Sariputra Jambi)”**, Sistem Informasi, Teknologi Informasi dan Sistem Komputer, (2019).

Penelitian prediksi gaya belajar siswa ini menggunakan data siswa SD Sariputra tahun ajaran 2018 dan 2019 sebanyak 155 data dan menggunakan 11 atribut untuk membantu dalam proses peramalan yaitu minat, keterampilan, metode pembelajaran yang disukai, pengetahuan materi, dan atribut lainnya. Hasil prediksi gaya belajar Ada 3 jenis gaya belajar yaitu, Visual, Auditorial dan Kinestetik.
8. **“ALGORITMA C45 UNTUK MENENTUKAN MAHASISWA PENERIMA BEASISWA (STUDI KASUS : PPS IAIN RADEN INTAN BANDAR LAMPUNG)”**, PPS IAIN Radin Intan Bandar Lampung, (2015).

Dari sampel yang diambil yaitu sebanyak 40 (empat puluh mahasiswa) calon penerima beasiswa pada Program Pascasarjana IAIN Raden Intan Lampung dapat dilihat pada pohon keputusan yaitu sebanyak 18 (delapan belas) mahasiswa yang tidak kompeten untuk mendapatkan beasiswa

karena mempunyai IPK < 3,00, kemudian sebanyak 8 (delapan) mahasiswa yang tidak kompeten untuk mendapatkan beasiswa karena mempunyai masa kerja < 5 tahun dan pekerja Non PNS, sehingga dihasilkan sebanyak 14 (empat belas) mahasiswa yang kompeten untuk mendapatkan beasiswa karena telah memenuhi kriteria penerima beasiswa dari segi IPK, pekerjaan dan masa kerja yang telah ditentukan sebelumnya.

9. “Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar”, STMIK Hang Tuah Pekanbaru, (2016).

Berdasarkan hasil dari analisis data menggunakan metode algoritma C4.5 menghasilkan Decision Tree untuk memprediksi prestasi akademik berdasarkan faktor sosial ekonomi, motivasi, peran dosen, disiplin dan hasil belajar masa lalu yaitu : (1) variabel hasil belajar masa lalu adalah variabel yang menentukan potensi seseorang sukses atau gagal dalam prestasi akademik. Hal ini ditunjukkan bahwa Hasil Belajar menjadi node yang terpilih/pertama. (2) Variabel Peran Dosen menjadi variabel 2 untuk menentukan keberhasilan Prestasi akademik (3). Variabel Disiplin menjadi variabel 3 menentukan keberhasilan Prestasi Akademik . (4) Hasil Akurasi klasifikasi menggunakan metode Area Under Curve (AUC) diperoleh hasil nilai 65%.

10. “PREDIKSI PRESTASI SISWA SMP NURUL JADID MENGGUNAKAN ALGORITMA C4.5”, Sekolah Tinggi Teknologi Nurul Jadid, (2016).

Penelitian ini melihat uji coba dataset kehadiran dan hasil nilai raport siswa sebanyak 1006 sebagai sample data. Atribut Sikap memiliki pengaruh yang cukup besar dalam hal prestasi di lembaga SMP Nurul jadid ini ditunjukkan bahwa setiap siswa yang bersikap baik dan rajin masuk sekolah ada kemungkinan besar akan menjadi siswa tersebut berprestasi. Kehadiran juga memiliki pengaruh setelah attribute sikap, hal ini dibuktikan dengan Decision Tree yang terbentuk sikap yang baik belum tentu berprestasi.

Tabel 2. 8 Tinjauan Pustaka

No.	PENELITI	JUDUL	PERMASALAHAN	JURNAL	KONTRIBUSI / KEKURANGAN
1.	Rolansa, Yunita, Dan Suheri	“sistem prediksi dan evaluasi prestasi	Dibutuhkannya sebuah data mining untuk menggali informasi	Jurnal Pendidikan Informatika dan Sains 9	Kontribusi terhadap penelitian ini adalah

		akademik mahasiswa di program studi teknik informatika menggunakan data mining”	seputar prestasi mahasiswa	(1), 2020, 75-85	pengambilan metode Algoritma C4.5
2.	Rismayanti	“IMPLEMENTASI ALGORITMA C4.5 UNTUK MENENTUKAN PENERIMA BEASISWA DI STT HARAPAN MEDAN”	Dalam menentukan mahasiswa yang mendapatkan Beasiswa Peningkatan Prestasi Akademik (PPA) yang diberikan kepada mahasiswa berprestasi dan bagi mahasiswa yang kurang mampu dirasa belum tepat	Jurnal Media Infotama Vol. 12 No. 2, September 2016	Kontribusi terhadap penelitian ini adalah hasil decision tree dan rule yang didapat dari penelitian tersebut. Namun pada penelitian ini tidak terdapat cara perhitungan Algoritma C4.5 dan belum adanya perhitungan tingkat akurasi.
3.	Hijriana dan Muttaqin	“PENERAPAN METODE DECISION TREE ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA BERPRESTASI”	Belum adanya pengolahan data dengan baik dan benar	AI Ulum Sains dan Teknologi Vol.2 No.1 Nopember 2016	Kontribusi terhadap penelitian ini adalah pengambilan metode Algoritma C4.5
4.	Widodo dan	“ALGORITMA C4.5 UNTUK	Belum akuratnya prediksi	Artikel Skripsi	Kontribusi terhadap

	Ramadhani	MEMPREDIKSI MAHASISWA YANG MENGULANG MATA KULIAH (STUDI KASUS DI AMIK LABUHAN BATU)	mahasiswa yang berpotensi mengulang mata kuliah	Universitas Nusantara PGRI Kediri, 2017	penelitian ini adalah pengambilan variabel kedisiplinan dan ekonomi. Namun pada penelitian ini tidak terdapat cara perhitungan Algoritma C4.5
5.	Jaman dan Astuti	“Penerapan Algoritma C4.5 untuk penentuan penerima beasiswa (sdn karawang kulon III)”	Belum tepatnya dalam menentukan beasiswa untuk siswa yang kurang mampu	Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi, ISSN: 2503-054X Vol. 3 No: 1, April 2018	Kontribusi terhadap penelitian ini adalah pengambilan atribut nilai. Namun pada penelitian ini tidak terdapat cara perhitungan Algoritma C4.5
6.	Sari	“MEMPREDIKSI PRESTASI SISWA DENGAN PENERAPAN ALGORITMA C4.5 DI SEKOLAH DASAR NEGERI 1 RAWA LAUT”	Terkadang siswa yang rajin tidak berprestasi untuk itu dibutuhkan kajian lebih mendalam lagi tentang hal ini. Ada siswa yang memiliki nilai raport dibawah rata-rata sehingga mengakibatkan siswa tersebut tidak naik kelas, untuk mencari pola mengapa ada	Prosiding Seminar Bisnis & Teknologi ISSN : 2407-6171, Desember 2014	Kontribusi penelitian ini adalah pengambilan perhitungan algoritma C4.5 dan pengambilan atribut kehadiran

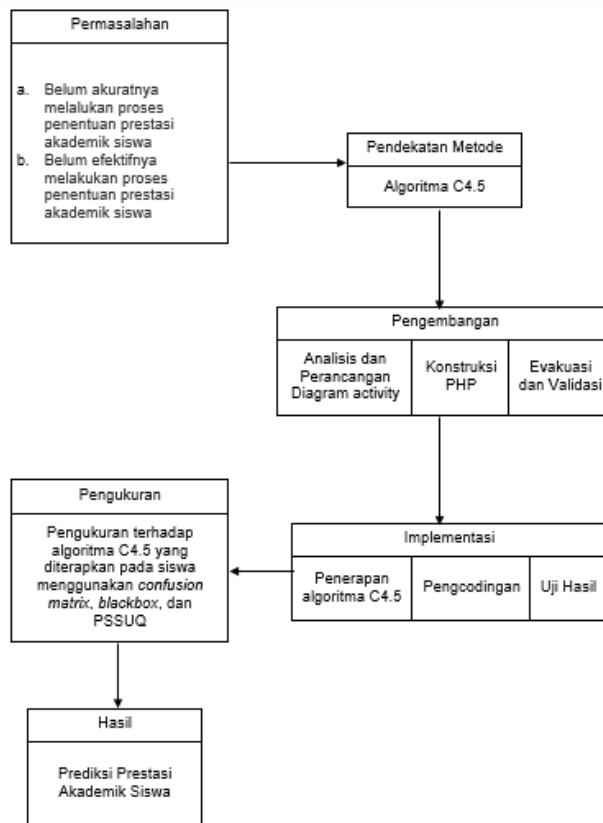
			siswa yang tidak naik kelas		
7.	Oscario, Jasmir, dan Novianto	“Penerapan Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar (Studi Kasus : SD Sariputra Jambi)”	Belum tepatnya gaya belajar bagi siswa sekolah dasar untuk mencapai prestasi belajar	Processor: Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer R –VOL. 14, No. 2, Oktober 2019	Kontribusi terhadap penelitian ini adalah decision tree
8.	Rahman	“ALGORITMA C45 UNTUK MENENTUKAN MAHASISWA PENERIMA BEASISWA (STUDI KASUS : PPS IAIN RADEN INTAN BANDAR LAMPUNG)”	Dengan Jumlah pendaftar beasiswa sangat banyak dan ada banyak petunjuk kriteria, sehingga diperlukan suatu sistem yang dapat membantu menentukan siapa saja yang berhak mendapatkan beasiswa perguruan tinggi	Jurnal TIM Darmajaya Vol. 01 No. 02 Oktober 2015	Kontribusi terhadap penelitian ini adalah pengambilan metode Algoritma C4.5. Kelemahan pada penelitian ini yaitu belum adanya perhitungan nilai akurasi.
9.	Sabna dan Muhardi	“Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen,	Seiring pergantian semester, nilai akademik mahasiswa semakin menurun. Belum akuratnya prediksi yang ada, dibutuhkan	Jurnal CoreIT, Vol.2, No.2, Desember 2016 ISSN: 2460-738X	Kontribusi terhadap penelitian ini adalah pengambilan variabel ekonomi dan kedisiplinan.

		Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar”	sebuah data mining untuk mengetahui informasi penyebabnya		
10.	Malik dan Faid	“PREDIKSI PRESTASI SISWA SMP NURUL JADID MENGGUNAKAN ALGORITMA C4.5”	Mencari pola mengapa siswa yang rajin tidak berprestasi mengapa hal seperti ini bisa terjadi untuk itu dibutuhkan kajian lebih mendalam lagi tentang hal ini	Nusantara Journal of Computers and its Applications Volume 2 – No.4, Desember 2016	Kontribusi terhadap penelitian ini adalah atribut kehadiran dan sikap

Berdasarkan uraian diatas, maka terdapat kekurangan yang menjadi salah satu dasar penelitian ini yaitu terdapat jurnal yang hanya menggunakan aplikasi yang sudah tersedia, dan beberapa jurnal tidak melakukan perhitungan tingkat akurasi. Sedangkan pada penelitian ini yaitu melakukan pengembangan aplikasi serta menghitung tingkat akurasinya.

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 2.8 dapat dijelaskan sebagai berikut :

- a. Permasalahan terkait yaitu belum akurat dalam penentuan prestasi akademik siswa dan belum efektifnya proses dalam penentuan prestasi akademik siswa.
- b. Pendekatan penelitian menggunakan Algoritma C4.5.
- c. Pengembangan yaitu tahap melakukan analisis dan perancangan gambar untuk diagram activity maupun class diagram, kemudian menentukan bahasa pemrograman dengan konstruksi PHP dan melakukan evaluasi serta validasi
- d. Implementasi yaitu tahap menerapkan metode C4.5 ke dalam aplikasi yang akan dibuat dan pembuatan coding lalu melakukan uji hasil dari sistem informasi tersebut.
- e. Pengukuran yaitu menguji ketepatan hasil perhitungan algoritma C4.5 dengan menggunakan confusion matrix, pengujian sistem kepada ahli sistem menggunakan blackbox, dan uji kebergunaan kepada pengguna menggunakan PSSUQ.
- f. Hasil yaitu sistem informasi menampilkan hasil prediksi prestasi akademik siswa.

E. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu belum akurat dan belum efektif pada saat melakukan proses penentuan prestasi akademik siswa, maka sangat memerlukan cara salam mengatasi permasalahan tersebut. Dalam teori data mining ada beberapa metode yang dapat melakukan pengklasifikasian untuk memperkirakan suatu hal di masa depan dari pengalaman di masa lalu, yaitu menggunakan metode C4.5. Teori klasifikasi yang dipakai yaitu C4.5 yang akan menghasilkan hasil prediksi yang telah ditentukan nilai probabilitas akhir dari setiap kelas atau label dan melihat hasil akhir yang terbesar dari setiap kelas nantinya hasil prediksi tersebut bisa digunakan untuk membantu wali kelas dalam memprediksi prestasi siswa. Berdasarkan hal tersebut, maka dapat ditetapkan hipotesis metode C4.5 diduga dapat untuk membantu wali kelas dalam memprediksi prestasi akademik siswa.