

Analisis Prediksi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritme Cart (*Classification And Regression Trees*)

(Studi Kasus Data Alumni Fakultas Teknik
Universitas Bengkulu)

Muhammad Fando Rizalno¹, Asahar Johar², Funny F. Coastera³
^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu
Jl. WR. Supratman Kandang Limun Bengkulu 38371A INDONESIA
(telp: 0736-341022; fax: 0736-341022)

¹fandorizalnol1@gmail.com,
²ffaradyc@unib.ac.id,
³asahar.johar@unib.ac.id

Abstrak: Salah satu sumber daya besar yang dimiliki universitas adalah basis data, namun basis data yang besar ini hanya disimpan dalam gudang data. Padahal data yang terkumpul dan berukuran besar tersebut merupakan aset yang dapat dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga. Melihat kondisi tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Universitas Bengkulu untuk melihat parameter yang paling berpengaruh pada lama masa studi mahasiswa, data yang akan dimanfaatkan disini adalah data akademik dan data wisudawan. Metode yang digunakan untuk menganalisisnya adalah Decision tree dengan menggunakan algoritme CART. Melakukan prediksi dibutuhkan dataset yang teratur, Dataset yang digunakan masih mengandung missing values sehingga dalam penelitian ini tahap preprocessing data dilakukan. Tahap preprocessing menggunakan sistem data warehouse dengan penerapan ETL. Aplikasi ini dibuat dengan menggunakan Visual studio code bahasa pemrograman PHP framework laravel. Hasil dari penelitian ini berupa aplikasi pengelompokan data alumni lulus tepat waktu dan prediksi lama masa studi untuk mahasiswa baru. Pengujian software menggunakan metode Evaluasi Accuracy dan Black box. Hasil dari penelitian ini adalah nilai parameter yang berpengaruh pada kelulusan mahasiswa dan dapat digunakan untuk proses prediksi masa studi mahasiswa.

Kata Kunci: Data, Decision tree, CART, Data warehouse, ETL, Accuracy.

Abstract: One of the major resources that universities have is databases, but these large databases are only stored in data warehouses. Whereas the data collected and large in size is an asset that can be used for analysis, the result of which is in the form of valuable knowledge or information. Seeing these conditions, research is needed to explore the data owned by Bengkulu University to see the most influential parameters on the length of the student's study period, the data that will be used here are academic data and graduate data. The method used to analyze it is a Decision tree using the CART algorithm. Making predictions requires regular datasets. The datasets used still contain missing values, so that in this study the data preprocessing stage was carried out. The preprocessing stage uses a data warehouse system with the application of ETL. This application is made using Visual studio code programming language PHP framework Laravel. The results of this study are an application for grouping alumni data to graduate on time and predict the length of the study period for new students. Software testing using Accuracy and Black box Evaluation methods. The results of this study are the parameter values that affect student graduation and can be used to predict the student's study period.

Keywords: Data, Decision tree, CART, Data warehouse, ETL, Accuracy.

I. PENDAHULUAN

Setiap program studi pada perguruan tinggi dituntut memiliki keunggulan untuk bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Salah satu sumber daya besar yang dimiliki universitas adalah basis data, namun basis data yang besar ini hanya disimpan dalam gudang data. Padahal data yang terkumpul dan berukuran besar tersebut merupakan aset yang dapat dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Kelulusan mahasiswa merupakan salah satu aspek penilaian dalam proses akreditasi pada suatu perguruan tinggi. Sehingga jika mahasiswa lulusnya tepat waktu akan membantu penilaian akreditasi suatu perguruan tinggi. Berdasarkan data peserta wisuda Program Sarjana (S1) di Fakultas Teknik Universitas Bengkulu banyak alumni yang menempuh masa studi lebih dari 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) yang menempuh lama studi lebih dari yang dijadwalkan yaitu 8 semester.

Melihat kondisi tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Universitas Bengkulu untuk melihat parameter yang paling berpengaruh pada lama masa studi mahasiswa, data yang akan dimanfaatkan disini adalah data akademik dan data wisudawan. Metode yang digunakan untuk menganalisisnya adalah Decision tree dengan menggunakan algoritme CART. Apabila data parameter yang paling

berpengaruh pada kelulusan mahasiswa dapat diketahui maka data tersebut dapat digunakan untuk memprediksi lama masa studi mahasiswa.

Prediksi disini merupakan proses untuk meramalkan suatu variabel di masa mendatang dengan berdasarkan pertimbangan data pada masa lampau. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin dengan kejadian.(Sinaga, 2018). Penelitian ini dirasa perlu karena jika parameter yang berpengaruh pada lama masa studi mahasiswa dapat dilihat dan diprediksi lebih dini, maka pihak jurusan dapat melakukan tindakan-tindakan yang dirasa perlu supaya mahasiswa dapat lulus tepat waktu sekaligus meningkatkan kualitas jurusan itu sendiri. Dan pihak jurusan dapat mengatur kuota masuk mahasiswa dengan mempertimbangkan berbagai parameter yang berpengaruh dalam kelulusan mahasiswa..

II. LANDASAN TEORI

A. Data Warehouse

Data warehouse adalah sebuah sistem yang mengambil dan menggabungkan data secara periodik dari sistem sumber data ke penyimpanan data bentuk dimensional atau normal. Data warehouse merupakan penyimpanan data yang berorientasi objek, terintegrasi, mempunyai variant waktu, dan menyimpan data dalam bentuk nonvolatile sebagai pendukung manajemen dalam proses pengambilan keputusan

Sebuah data warehouse merupakan penyimpanan data tetap sebagai implementasi fisik dari pendukung keputusan model data. Data warehouse juga biasanya dilihat sebagai arsitektur, pembangunan dan penyatuan data dari bermacam macam sumber data yang berbeda untuk

mendukung struktur dan atau query tertentu, laporan analisis, dan pembuatan keputusan.

Extract, transform, dan load (ETL) merupakan sebuah sistem yang dapat membaca data dari suatu data store, merubah bentuk data, dan menyimpan ke data store yang lain. Data store yang dibaca ETL disebut data source, sedangkan data store yang disimpan ETL disebut target. Proses pengubahan data digunakan agar data sesuai dengan format dan kriteria, atau sebagai validasi data dari source system. Proses ETL tidak hanya menyimpan data ke data warehouse, tetapi juga digunakan untuk berbagai proses pemindahan data. Kebanyakan ETL mempunyai mekanisme untuk membersihkan data dari source system sebelum disimpan ke warehouse. Pembersihan data merupakan proses identifikasi dan koreksi data yang kotor. Proses pembersihan ini menerapkan aturan-aturan tertentu yang mendefinisikan data bersih (kimball, 2004)..

B. *Analytical Network Process (ANP)*

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. (Han, 2006)

- Decision tree

Dalam decision tree tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil. Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan.

Ada beberapa macam algoritme decision tree diantaranya ID3, C4.5, dan CART. Beberapa isu utama dalam decision tree yang menjadi perhatian yaitu seberapa detail dalam mengembangkan decision tree, bagaimana mengatasi atribut yang bernilai continues, memilih ukuran yang cocok untuk penentuan atribut, menangani data training yang mempunyai data yang atributnya tidak mempunyai nilai, memperbaiki efisiensi perhitungan.

- Algoritme Cart

Pembentukan cabang pada algoritme CART (Classification and Regression Tress) berbeda dengan ID3 dan C4.5 yang berbasis pada perhitungan entropy. Algoritme CART menggunakan perhitungan IndexGini untuk pembentukan cabang, Sedangkan untuk pembentukan node pada algoritme CART digunakan perhitungan GiniGain.

$$\text{IndexGini} = \left| 1 - \sum_{i=1}^n (p_i^2) \right|$$

$$\text{GiniGain} = \text{Gini}(A,S) - \sum_{i=1}^n \left(\frac{s_i}{s} \times \text{Gini}(s_i) \right)$$

Keterangan:

P_i = jumlah kasus per cabang dibagi total kasus

S_i = jumlah kasus per cabang

S = jumlah total kasus

Gini(a,s) = index gini total kasus

Gini(s_i) = index gini per cabang kasus

- Akurasi

Accurasy pada klasifikasi merupakan persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Rumus Accuracy dapat didefinisikan pada rumus berikut ini:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / P$$

TP = jumlah record/tuple positif yang dilabelkan secara benar oleh model algoritme klasifikasi

TN = jumlah record/tuple negatif yang dilabelkan secara salah oleh model algoritme klasifikasi

P = total semua record yang dievaluasi. (Kamber, 2012)

C. Database

Pada penelitian yang akan dilakukan, acuan dari beberapa penelitian terdahulu menjadi sangat penting dalam melakukan sebuah penelitian dengan tujuan untuk mengetahui hubungan antara penelitian yang akan dilakukan dengan penelitian terdahulu, sehingga dengan menambahkan acuan tersebut dapat menghindari adanya suatu duplikasi dalam penelitian yang akan dilakukan.

Ratna dan indra (2018) menerapkan Algoritme C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. Penelitian ini menggunakan algoritme C4.5 untuk menentukan pohon keputusannya dengan menggunakan atribut asal daerah, ipk, jenis kelamin. Atribut ini didapat

dari data repositori lulusan dimana datanya banyak dan terdapat sedikit miss value pada datanya. Pada penelitian ini menghasilkan nilai akurasi sebesar 60,52% untuk data semuanya. (Putri, 2018)

Selanjutnya penelitian lain membandingkan klasifikasi algoritme C5.0 dengan CART pada data sosial kepala keluarga masyarakat desa teluk baru kecamatan muara ancalong oleh remi, memi, dan surya (2019). Hasil rata-rata tingkat akurasi ketepatan klasifikasi algoritme C5.0 sebesar 79,17% sedangkan metode CART sebesar 84,63%. Sehingga dapat dikatakan bahwa metode CART merupakan metode yang lebih baik dalam pengklasifikasian data rata-rata pendapatan masyarakat Desa Teluk Baru Kecamatan Muara Ancalong tahun 2019 dibandingkan dengan metode algoritme C5.0. (Pratiwi, 2019)

Mardiani (2019) melakukan penerapan klasifikasi dengan algoritme CART untuk prediksi kuliah bagi mahasiswa baru. Data yang digunakan yaitu 365 data mahasiswa jurusan sistem informasi yang didapat dari data akademik dalam bentuk database yang kemudian melalui proses cleansing untuk mencari atribut yang dapat digunakan. Proses tersebut didapatlah data mahasiswa dengan atribut yaitu NPM, IPK, grade masuk, gelombang masuk USM. Didapatlah IPK yang menjadi akar utama dari pohon kuputusan sehingga disimpulkan bahwa ipk yang paling mempengaruhi lama kelulusan mahasiswa sistem informasi. (mardiani, 2019).

III. METODOLOGI PENELITIAN

A. Jenis Penelitian

Dalam penelitian ini akan dibangun sistem untuk memprediksi lama masa studi mahasiswa dari berbagai aspek yang mempengaruhinya berdasarkan data alumni terdahulu, dan mengelompokan data lama masa studi alumni

berdasarkan berbagai aspek yang mempengaruhinya. Sistem ini diharapkan dapat membantu mahasiswa untuk mengetahui kemungkinan tingkatan kelulusannya, dan membantu fakultas untuk menentukan atau memilih calon mahasiswa barunya.

Sistem yang akan dibangun adalah sistem pengelompokan data menggunakan Data Warehouse dan melakukan proses mining dengan metode data mining Decision Tree algoritme CART. Data yang diperlukan diambil dari database mahasiswa dan database alumni. Tidak semua data wisuda akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses mining memakan waktu lama dan tingkat hubungannya pun rendah..

B. Teknik Pengumpulan Data

- Studi Pustaka

Dalam metode ini, penulis mempelajari buku-buku yang bersangkutan sebagai dasar dan acuan dalam memahami Data Warehouse dan Data mining. Selain itu, juga memanfaatkan internet sebagai sumber data guna menambah referensi terkait dalam membangun aplikasi berbasis android.

- Observasi

Metode pengumpulan data dengan cara mengamati dan menganalisa guna mendapatkan data-data sebagai penunjang perancangan Data Warehouse dan Data mining.

- Implementasi

Apabila tahapan analisa sudah selesai maka kebutuhan dan perancangan sistem maka akan masuk ke tahap selanjutnya yaitu tahap implementasi. Dimana tahapan implementasi ini merupakan tahapan nyata yang akan dilaksanakan

setelah penelitian. Pada tahap ini pun penulis akan memaksimalkan pengerjaan dari penelitian ini.

C. Metode Pengujian Sistem

Pengujian yang dilakukan pada sistem yang dibangun menggunakan metode black box testing. Black box testing adalah suatu kondisi pengujian yang dibangun berdasarkan fungsional program. Pengujian ini membutuhkan informasi tentang data inputan dan mengamati output, tetapi tidak tahu bagaimana sistem bekerja. Pengujian lebih difokuskan pada fungsionalitas program daripada spesifikasinya. Keuntungan dari pengujian ini adalah, pengujian ini mencocokkan apa yang diharapkan dilakukan oleh program, dan ini secara alami serta dimengerti oleh semua orang.

IV. ANALISIS DAN PERANCANGAN

A. Analisis Sistem

Dalam penelitian ini akan dicari hubungan tingkat kelulusan dengan berbagai aspek data wisudawan. Tidak semua data wisuda akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses mining memakan waktu lama dan tingkat hubungannya pun rendah. Serta atribut yang digunakan harus memiliki data yang banyak karena pada data mining semakin banyak data maka tingkat akurasi akan semakin tinggi. Adapun yang akan diproses mining meliputi:

- IPK

Hubungan IPK terhadap tingkat kelulusan bermanfaat untuk mengetahui apakah dari jenis IPK yang tinggi banyak mempengaruhi masa studi

yang cepat atau IPK yang rendah lebih mempengaruhi tingkat kelulusan mahasiswa.

Pada IPK akan dibagi menjadi 3 kelompok yaitu untuk 3,50 sampai 4,00 termasuk golongan tinggi, untuk 3,00 sampai 3,49 termasuk golongan sedang dan 2,99 ke bawah termasuk golongan rendah.

- Jenis Kelamin

Hubungan tingkat kelulusan dengan jenis kelamin bermanfaat untuk mengetahui jenis kelamin yang lebih banyak menghasilkan lulusan tepat waktu. Jenis kelamin mungkin tidak terlalu logis untuk dijadikan atribut namun jenis kelamin dapat menjadi atribut pendukung untuk meningkatkan akurasi dari prediksi nantinya.

Dalam jenis kelamin pastinya akan di kelompokkan untuk laki-laki yaitu L dan untuk perempuan yaitu P.

- Program studi

Dari atribut program studi dapat diketahui hubungan tingkat kelulusan dan program studi untuk mengetahui tingkat kelulusan pada tiap program studi yang ada sehingga dapat diketahui apakah pembelajaran tiap program studi mempengaruhi masa studi mahasiswanya.

Pada Program Studi merupakan program studi yang terdapat pada Fakultas Teknik UNIB yang terbagi menjadi 4 yaitu Teknik Informatika, Teknik Mesin, Teknik Sipil, Teknik Elektro.

- Jalur Masuk

Hubungan tingkat kelulusan dengan Jalur Masuk bermanfaat untuk mengetahui jalur masuk mana yang mempunyai tingkat kelulusan tinggi ataupun rendah. Sehingga dapat diketahui jalur masuk mana yang dapat menghasilkan banyak mahasiswa lulus tepat waktu dan pihak prodi dapat mengatur kuota tiap jalur masuk yang ada.

Pada Jalur Masuk terbagi menjadi 4 yaitu SNMPTN, SBMPTN, SPMU, dan untuk jalur yang lain dari pada itu dimasukkan ke Jalur Lainnya.

- Domisili

Hubungan Domisili dengan tingkat kelulusan bermanfaat untuk mengetahui pengaruh daerah asal yang mempunyai tingkat kelulusan tinggi ataupun rendah. Dan dapat diketahui apakah daerah asal bukan Bengkulu dapat menghasilkan mahasiswa lulus tepat waktu.

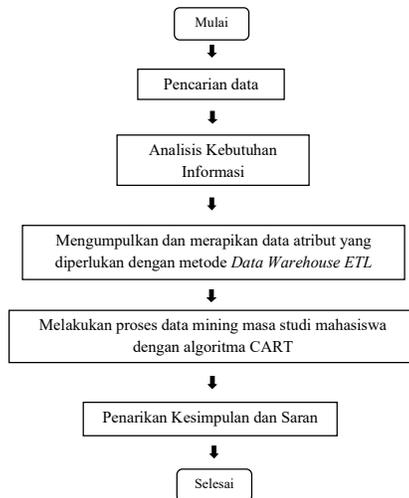
Atribut Domisili terbagi 2 yaitu Bengkulu untuk mahasiswa yang berasal dari daerah pada provinsi Bengkulu, dan Luar kota untuk mahasiswa yang tidak berasal di provinsi Bengkulu.

- Asal sekolah

Dari atribut asal sekolah dicari hubungan tingkat kelulusan dengan asal sekolah dengan harapan dapat mengetahui tingkat keberhasilan mahasiswa dari berbagai jenis sekolah tertentu. Sehingga dapat menjadi bahan acuan prodi untuk mengatur kuota undangan masuk untuk tiap jenis sekolah.

Pada Asal Sekolah terdapat 3 kelompok yaitu SMA, SMK, dan untuk kelompok yang lainnya akan masuk ke sekolah lainnya.

B. Alur Sistem



Gambar 1. Diagram Alur Sistem

- Proses Data Warehouse ETL

1.Import data

Proses import data adalah proses load data dari data akademik dan data wisuda ke data warehouse. Untuk awalnya semua data akan dimasukkan terlebih dahulu.

2.Selection

Selection data adalah proses menyeleksi atribut apa saja yang diperlukan untuk sistem.

3.Transformasi

Transformasi data merupakan proses memindahkan data yang telah di seleksi sebelumnya ke database yang baru, sehingga data tersebut dapat di satukan dengan menggunakan NPM sebagai penghubungnya.

4. Cleaning

Cleaning data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Dalam tahap ini semua data yang digunakan akan

dibersihkan dari record data yang tidak mempunyai atribut lengkap.

- Proses Mining

Pada proses mining data yang telah di rapikan dengan proses data warehouse ETL akan digunakan. Data akan di proses dengan data mining metode desission tree algoritme CART.

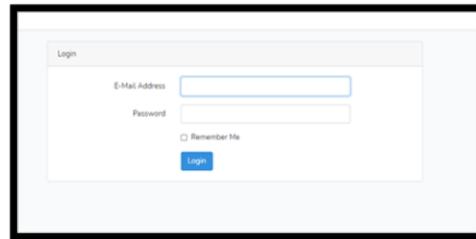
V. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

Setelah melakukan perancangan sistem, maka selanjutnya dilakukan implementasi sistem. Berikut merupakan halaman sistem web untuk prediksi tingkat kelulusan.

1) Halaman Login

Halaman Login merupakan halaman yang digunakan untuk menampilkan nama pengguna dan kata sandi untuk dimasukkan untuk masuk ke halaman utama. Pada Gambar 5.12 menunjukkan rancangan antarmuka halaman login sistem.

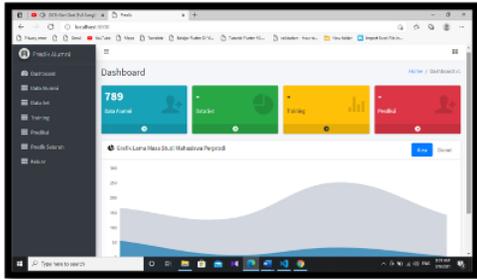


Gambar 2. Halaman Login

Pada gambar 2 halaman login, admin menginputkan email dan kata sandi untuk masuk ke halaman utama pada sistem, jika nama pengguna dan kata sandi yang diinputkan benar maka akan masuk ke dalam halaman utama sistem dan jika salah maka akan kembali lagi ke halaman login..

2) Halaman Utama

Halaman Utama adalah halaman yang digunakan admin untuk melihat Grafik dari data yang telah dikelompokkan. Berikut adalah halaman utama yang akan ditunjukkan pada Gambar 3.

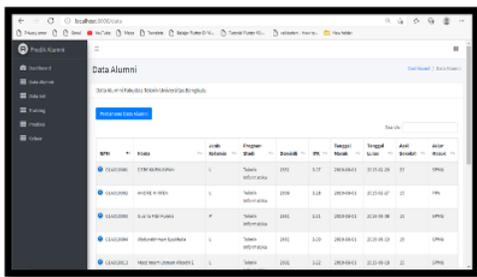


Gambar 3. Halaman Utama

Pada gambar 3 ini, halaman utama dari sistem dimana terdapat grafik dari data yang telah dikelompokkan. Seperti grafik perbandingan data mahasiswa yang lulus tepat dan tidak tepat.

3) Halaman Data Alumni

Halaman Data alumni adalah halaman yang digunakan admin untuk melihat data alumni yang ada. Berikut adalah halaman data alumni yang akan ditunjukkan pada Gambar 4

The screenshot shows a 'Data Alumni' page with a table of alumni records. The table has columns for 'ID Alumni', 'Nama', 'Jenis Kelamin', 'Program Studi', 'Status', 'Rata-rata', 'Tingkat', 'Tahun Masuk', and 'Tahun Keluar'. There are five rows of data, each with a blue circular icon to the left of the 'ID Alumni' column.

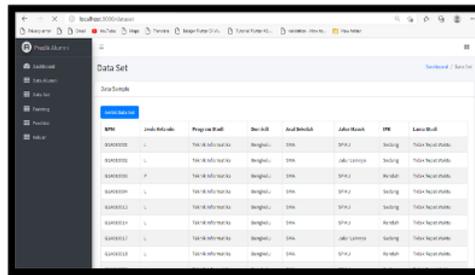
Gambar 4. Halaman Data Alumni

Pada gambar 4 ini, merupakan tampilan data alumni, dimana disini terdapa data awal dari alumni yang akan diolah. Terdapat tombol ambil data yang berfungsi untuk mengambil terpilih data dari database akademik dan database wisudawan, lalu disatukan dalam satu tabel. Data disini masih

beupa data mentah yang diambil langsung dari data base akademik dan wisudawan.

4) Halaman Data Set

Halaman data set adalah halaman yang digunakan admin untuk melihat data set yang ada. Berikut adalah halaman data set yang akan ditunjukkan pada Gambar 5

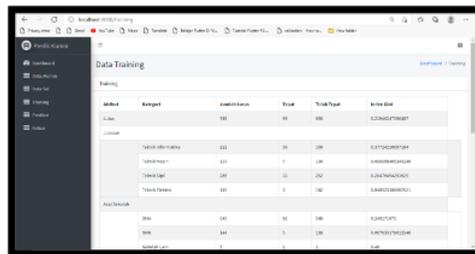
The screenshot shows a 'Data Set' page with a table of data set records. The table has columns for 'ID Data Set', 'Jenis Kelamin', 'Program Studi', 'Dini-Id', 'Aval-Berkas', 'Jenis-Berkas', 'IPK', and 'Lama-Studi'. There are ten rows of data, each with a blue circular icon to the left of the 'ID Data Set' column.

Gambar 5. Halaman Data Set

Pada gambar 5 ini, merupakan tampilan data set, dimana disini terdapa data set yang akan digunakan pada proses mining. Terdapat tombol ambil data yang berfungsi mengambil data pada tabel alumni sebelumnya yang kemudian akan dikelompokkan tiap atributnya untuk mempermudah proses mining. Data yang dikelompokkan per poinnya menjadi data non nominal dan menentukan lama studi dari data tanggal masuk dan keluar.

5) Halaman Data Training

Halaman Data Training adalah halaman yang digunakan admin untuk melihat data training yang ada. Berikut adalah halaman data training yang akan ditunjukkan pada Gambar 6.

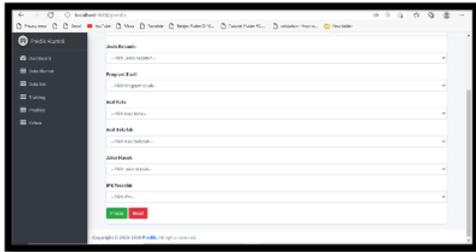
The screenshot shows a 'Data Training' page with a table of training data. The table has columns for 'Label', 'Berkas', 'Jumlah-Berkas', 'Berkas', 'Tingkat-Tujuan', and 'Berkas-Berkas'. There are two main sections: 'Label' and 'Berkas-Berkas', each with a table of data.

Gambar 6. Halaman Data Training

Pada gambar 6 ini, merupakan tampilan data training, dimana disini terdapa data training yang akan digunakan untuk proses mining. Data di sini juga dapat digunakan sebagai informasi dari total perhitungan jumlah mahasiswa lulus tepat dan tidak dari berbagai atribut.

6) Halaman Prediksi

Halaman Prediksi adalah halaman yang digunakan admin untuk mencoba prediksi lama masa studi mahasiswa. Berikut adalah halaman prediksi yang akan ditunjukkan pada Gambar 7.

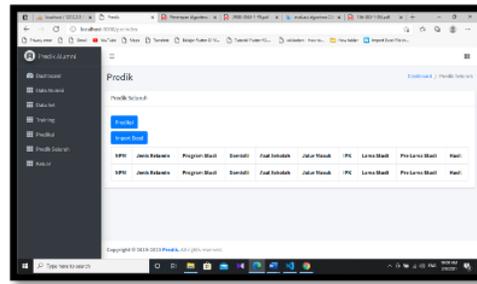


Gambar 7. Halaman Prediksi

Pada gambar 7 ini, merupakan tampilan halaman prediksi, dimana disini terdapat pertanyaan atau kuisisioner yang perlu diisi supaya dapat diprediksi lama masa studi mahasiswa. Setelah menekan tombol proses maka akan terlihat lama masa studi dari mahasiswa yang datanya telah diinputkan

7) Halaman Prediksi seluruh

Halaman Prediksi Seluruh adalah halaman yang digunakan admin untuk mencoba prediksi lama masa studi mahasiswa. Berikut adalah halaman prediksi seluruh yang akan ditunjukkan pada Gambar 8.



Gambar 8. Halaman Prediksi Seluruh

Pada gambar 8, merupakan tampilan halaman prediksi, dimana disini berguna untuk memprediksi dengan jumlah yang banyak sehingga terdapatlah tombol import excel untuk memasukan datanya. Setelah menekan tombol proses maka akan terlihat lama masa studi dari mahasiswa yang datanya telah diinputkan. Di dini terdapat lama studi dan prediksi lama studi yang digunakan untuk melihat akurasi dari algoritme yang ada.

B. Pengujian perhitungan metode *Analytical Network Process (ANP)*

Di dalam penelitian, untuk menguji performa algoritme digunakanlah pencarian accuracy. Accuracy pada klasifikasi merupakan persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Rumus Accuracy dapat didefinisikan pada rumus berikut ini:

$$\text{Accuracy} = (TP+TN)/P$$

TP = jumlah record/tuple positif yang dilabelkan secara benar oleh model algoritme klasifikasi

TN = jumlah record/tuple negatif yang dilabelkan secara salah oleh model algoritme klasifikasi

P = total semua record yang dievaluasi.

Berikut adalah tabel akurasi dari algoritme cart dengan menggunakan data set alumni fakultas teknik:

Data set yang digunakan	Benar (TP)	Salah (TN)	Total data evaluasi yang digunakan (P)	Akurasi
100% (789)	711	78	789	80,22%
90% (710)	707	82	789	79,21%
70% (550)	692	97	789	75,41%
50% (394)	668	121	789	69,32%

Gambar 9. Tabel akurasi per data set

Dari tabel diatas dapat di lihat jika menggunakan data set 100% menghasilkan akurasi 80.22% sedangkan untuk data set 50% menghasilkan akurasi 69.32%. maka dapat di simpulkan bahwa semakin banyak data set maka akurasi akan semakin tinggi.

Berikut adalah tabel akurasi dari algoritme cart dengan menggunakan data set alumni fakultas teknik per parameter:

Data		Total	Benar	Salah	Akurasi	Rata-rata
IPK	Rendah	112	112	0	100%	88,52%
	Sedang	562	519	43	84,69%	
	tinggi	115	104	11	80,86%	
Jurusan	Teknik Informatika	222	182	40	63,96%	82,48%
	Teknik Mesin	137	130	7	89,78%	
	Teknik Sipil	285	255	30	78,94%	
Jalur Masuk	Teknik Elektro	145	143	2	97,24%	82,61%
	SNMPTN	433	379	54	75,05%	
	SBMPTN	89	80	9	79,77%	
Jenis Kelamin	SPMU	185	176	9	90,27%	74,71%
	Jalur Lainnya	82	76	6	85,36%	
	Laki-laki	571	534	37	87,04%	
Asal Sekolah	Perempuan	218	177	41	62,36%	77,10%
	SMA	640	566	74	76,87%	
	SMK	144	140	4	94,44%	
Domisili	Sekolah Lainnya	5	4	1	60%	78,91%
	Bengkulu	762	686	76	80,05%	
	Luar Kota	27	24	3	77,77%	

Gambar 10. Tabel akurasi per atribut

Dari tabel diatas dapat di lihat parameter yang menghasilkan akurasi yang tinggi adalah IPK Rendah dengan nilai 100%, sedangkan terendah

adalah asal sekolah lainya dengan nilai 60%. Dan setelah di rata-rata kan persentasi tertinggi adalah atribut IPK dengan persentase 88.52% dan yang terendah adalah jenis kelamin dengan persentase 74.71%. Maka dari itu dari segi persentase keseluruhan dari yang tertinggi berpengaruh adalah IPK – Jalur Masuk – Jurusan - Domisili – Asal Sekolah – Jenis Kelamin. Sesuai saat dilihat dengan pohon keputusan yang paling berpengaruh adalah IPK sedangkan untuk tidak terlalu berpengaruh pada pohon keputusan adalah domisili dan untuk persentase adalah jenis kelamin..

VI. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijabarkan sebelumnya, maka kesimpulan yang dapat diambil adalah penelitian ini telah berhasil membangun data mining untuk memprediksi masa studi mahasiswa serta dapat mengimplementasikan metode decision tree dengan algoritme CART pada aplikasi predisi masa studi mahasiswa Fakultas Teknik. Pada hasil pohon keputusan algoritme CART dapat dilihat tingkatan parameter yang dimulai dari IPK – Jurusan - Jalur Masuk – Jenis Kelamin – Asal Sekolah – Domisili. Tingkatan ini dilihat dari kemunculan cabang pada tiap node yang ada sehingga dapat disimpulkan berdasarkan pohon keputusan bahwa parameter yang paling berpengaruh adalah IPK dan yang paling tidak berpengaruh adalah domisili. Dengan menggunakan algoritme CART parameter yang menghasilkan persentase akurasi yang tinggi adalah IPK Rendah dengan nilai 100%, sedangkan terendah adalah asal sekolah lainya dengan nilai 60%. Dan setelah di rata-rata kan persentasi tertinggi adalah atribut IPK dengan persentase

88.52% dan yang terendah adalah jenis kelamin dengan persentase 74.71. Maka dari itu dari segi persentase keseluruhan dari yang tertinggi berpengaruh adalah IPK – Jalur Masuk – Jurusan - Domisili – Asal Sekolah – Jenis Kelamin. Maka dapat disimpulkan dari ke dua proses yang berpengaruh adalah IPK dan yang paling tidak berpengaruh adalah domisili dan jenis kelamin..

B. *Saran*

Berdasarkan Analisa, implementasi, dan pengujian sistem ada beberapa saran bagi penulis dimasa mendatang yakni Sistem dapat dikembangkan lebih lanjut dengan menggunakan metode yang berbeda seperti KNN dan menggunakan algoritme C5.0, atau mengkombinasikan satu metode dengan metode yang lainnya sehingga mendapatkan hasil yang lebih sesuai dengan yang diharapkan dan dapat dibandingkan dengan hasil dari algoritme CART.

REFERENSI

- [1] Azuri, D. F., Zulhanif, & Pontoh, R. S. (2016). Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means. ISBN 978-602-72216-1-1 , 78-83.
- [2] Az-zahara, f. (2016). Multidimensional Database. yogyakarta: dictio.
- [3] Han, J. a. (2006). Data Mining Concepts and Techniques Second Edition. San Francisco.: Morgan Kauffman.
- [4] Kamber, J. H. (2012). Data Mining:Concept and Techniques. Burlington: Elsevier.
- [5] kimball, R. (2004). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting. new york: wiley publishing.
- [6] Kusriani. (2017). Strategi Perancangan dan Pengelolaan Basis Data. yogyakarta: Universitas Amikom.
- [7] mardiani. (2019). penerapan klasifikasi dengan algoritme CART untuk prediksi kuliah bagi mahasiswa baru. . penerapan klasifikasi dengan algoritme CART untuk prediksi kuliah bagi mahasiswa baru. .
- [8] Pratiwi, R. (2019). PERBANDINGAN KLASIFIKASI ALGORITME C5.0 DENGAN CLASSIFICATION AND REGRESSION TREE. PERBANDINGAN KLASIFIKASI ALGORITME C5.0 DENGAN CLASSIFICATION AND REGRESSION TREE.
- [9] Putri, R. P. (2018). Penerapan Algoritme C4.5 pada Aplikasi Prediksi . Penerapan Algoritme C4.5 pada Aplikasi Prediksi .
- [10] Rosa, A., & Shalahuddin, M. (2018). Rekayasa Perangkat Lunak Edisi Revisi. Bandung: Informatika.
- [11] Silberschatz, a. (2003). Operating System Concepts. 1009-1010.: The British Journal of Psychiatry.
- [12] Sinaga, E. (2018). PERANCANGAN APLIKASI PREDIKSI JUMLAH KELULUSAN. PERANCANGAN APLIKASI PREDIKSI JUMLAH KELULUSAN, 6.
- [13] Zaki, A. (2008). 36 Menit Belajar Komputer: Php Dan Mysql. Jakarta: Media Komputindo.