



Penerapan Metode Algoritma C4.5 Untuk Prediksi Mahasiswa Non Aktif

Irmayansyah^{1*}, Ersisya Lastrini²

¹Sistem Informasi/Universitas Binaniga Indonesia

Email: irma@stikombinaniaga.ac.id

²Sistem Informasi/ Universitas Binaniga Indonesia

Email: lastrisya2316@gmail.com

ABSTRACT

The problem of non-active students is something to be aware of because it can affect the quality of education and result in a decrease in campus financial income. If the problem of inactive students can be predicted faster, then the management can prevent and anticipate early. To solve the problem, c4.5 algorithm is applied to the prediction of non-active students in order to produce patterns based on classification results. By using IPS, Attendance, Annual Income, cost sources, and payment status. This is done to monitor students who are potentially non- active so as to anticipating for the decline of active students. In this study, feasibility test has been conducted, with a feasibility value of 87.50%, and also has been conducted accuracy test using confussion matrix formula with 81% accuracy result.

Keywords: Prediction; Non-Active Student; C4.5 Algorithm; Classification; feasibility.

ABSTRAK

Permasalahan mahasiswa non – aktif menjadi sesuatu yang harus di perhatikan karena hal ini dapat mempengaruhi kualitas pendidikan serta mengakibatkan penurunan pendapatan keuangan kampus. Jika masalah mahasiswa non-aktif dapat di prediksi lebih cepat, maka pihak manajemen dapat mencegah dan mengantisipasi lebih awal .Untuk mengatasi masalah tersebut maka diterapkan algoritma c4.5 untuk prediksi mahasiswa non - aktif agar menghasilkan pola berdasarkan dengan hasil klasifikasi. Dengan menggunakan variabel IPS, Absensi, Penghasilan per tahun, sumber biaya, dan status pembayaran. Hal ini dilakukan untuk memonitoring mahasiswa – mahasiswa yang berpotensi non – aktif sehingga dapat mengantisipasi penurunan mahasiswa yang aktif. Pada penelitian ini sudah dilakukan uji kelayakan, dengan nilai kelayakan sebesar 87, 50 %, dan juga sudah dilakukan uji akurasi dengan menggunakan rumus confussion matrix dengan hasil akurasi 81%.

Keywords: Prediksi; Mahasiswa Non – Aktif; Algoritma C4.5; Klasifikasi; Kelayakan.

A. PENDAHULUAN

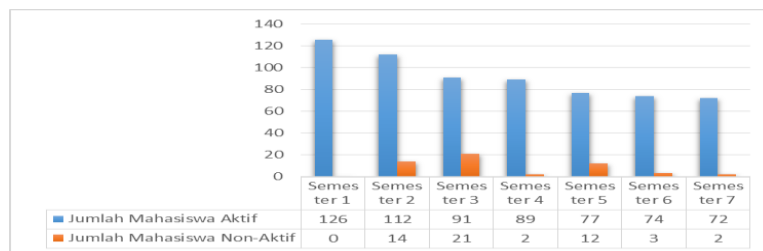
1. Latar Belakang

Status mahasiswa merupakan salah satu elemen dari akreditasi sebuah Perguruan Tinggi, pada saat ini Fakultas Informatika dan Teknologi di Universitas Binaniaga Indonesia pada saat ini untuk memprediksikan mahasiswa non – aktif berdasarkan pada mahasiswa yang sudah

melakukan perwalian dan mahasiswa yang mengikuti Ujian Akhir Semester. Hal tersebut dapat dilihat dari hasil wawancara yang telah dilakukan dengan Kepala Bagian Administrasi Akademik dan Kepala Bagian Administrasi Umum, dengan pertanyaan “Bagaimana proses yang dilakukan untuk pendataan keaktifan mahasiswa dalam proses perkuliahan?”

Berdasarkan hasil wawancara dengan Bagian Administrasi Akademik bahwa Universitas Binaniaga Indonesia pada Fakultas Informatika dan Teknologi dalam proses mendata mahasiswa aktif dilakukan dengan hasil melihat data mahasiswa yang sudah melakukan perwalian, dan selama ini belum dilakukan prediksi untuk mahasiswa non-aktif.

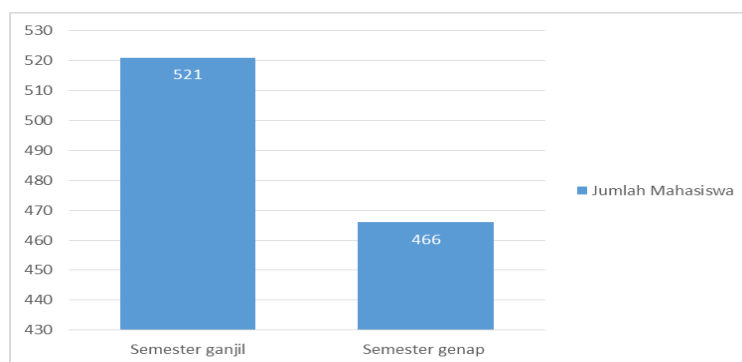
Selain wawancara diperoleh data mahasiswa non – aktif dari Bagian Administrasi Akademik pada tahun angkatan 2015/2016 menunjukkan perbandingan antara jumlah mahasiswa awal yang masuk dengan jumlah mahasiswa yang sudah melanjutkan tingkat semester terdapat selisih jumlah mahasiswa yang tidak melanjutkan disemester selanjutnya.



Gambar 1. Data Angkatan 2015/2016

Berdasarkan Gambar 1 bahwa pada Angkatan Tahun Akademik 2015/2016 mahasiswa masuk pada semester 1 dengan jumlah sebanyak 126 orang, namun pada saat semester 7 mahasiswa yang aktif hanya tersisa sebanyak 72 orang, yang diartikan bahwa mahasiswa non – aktif sebanyak 54 orang.

Namun berdasarkan hasil wawancara dengan Bagian Administrasi Umum bahwa selama ini sudah melakukan prediksi mahasiswa non-aktif dengan melihat mahasiswa mengikuti Ujian Akhir Semester, jika mahasiswa tersebut mengikuti Ujian Akhir Semester maka mahasiswa tersebut dianggap sebagai mahasiswa aktif, jika pada saat Ujian Akhir Semester tidak menghadiri namun mahasiswa tersebut mengikuti Ujian Susulan maka mahasiswa tersebut dianggap sebagai mahasiswa aktif, dan Bagian Administrasi Umum juga melihat dari tunggakan biaya pembayaran, semakin banyak mahasiswa tidak melakukan pembayaran maka diperkirakan bahwa mahasiswa tersebut akan menjadi mahasiswa non-aktif, dengan nilai akurasi prediksi 80%. Selain wawancara juga diperoleh data mahasiswa aktif pada tahun akademik 2019/2020 dari Bagian Administrasi Umum pada Gambar 1.2



Gambar 2. Data Mahasiswa Aktif

Bahwa berdasarkan hasil wawancara dan data yang diperoleh dari Bagian Administrasi Akademik Data Mahasiswa Aktif untuk setiap semester terjadi penurunan mahasiswa aktif rata-rata kurang lebih 50%, dan berdasarkan hasil wawancara dan data yang diperoleh dari Bagian Administrasi Umum untuk tahun akademik 2019/2020 untuk seluruh angkatan

berpengaruh terhadap pendapatan kampus yang mengalami penurunan pendapatan keuangan hingga 10%

2. Permasalahan

Berdasarkan permasalahan diatas maka dapat diidentifikasi sebagai berikut :

- a. Belum akurat dalam melakukan prediksi mahasiswa yang aktif dan non – aktif
- b. Belum efektifnya proses pendataan keaktifan mahasiswa dalam proses perkuliahan.

3. Tujuan

Adapun tujuan dari penelitian ini adalah :

- a. Mendapatkan rekomendasi mahasiswa yang akan non – aktif.
- b. Meningkatkan efektifitas dalam proses pendataan keaktifan mahasiswa dalam proses perkuliahan.
- c. Mengembangkan aplikasi menggunakan pemodelan Algoritma C4.5 untuk memprediksikan mahasiswa non – aktif.
- d. Mengukur tingkat keakuratan penerapan algoritma c4.5 untuk memprediksi mahasiswa non – aktif.

4. Tinjauan Pustaka

a. Data Mining

Menurut Tan (2006) dalam Eko Prasetyo (2013), menyatakan bahwa data mining adalah suatu proses pengambilan informasi dari sekumpulan data yang besar. Data mining juga dapat diartikan sebagai penggalian informasi baru yang didapat dari beberapa data besar untuk membantu pengambilan keputusan. Istilah data mining disebut juga knowledge discovery. Knowledge Discovery in Databases (KDD) adalah metode yang digunakan dalam datamining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD (Turban, dkk 2001).

KDD melibatkan teknik penghubungan dan penemuan ilmiah, pemahaman dan penggambaran dari pola data. Rangkaian tahapan tersebut meliputi tahapan sebagai berikut (Tan, 2006):

- 1) Pembersihan data yaitu untuk menghilangkan data yang tidak tetap atau adanya kesalahan
- 2) Integrasi data yaitu menghubungkan atau mengombinasikan beberapa data.
- 3) Transformasi data yaitu merubah bentuk data menjadi bentuk data yang sesuai untuk melakukan proses mining.
- 4) Aplikasi teknik data mining yaitu proses pemisahan pola dari data yang ada.
- 5) Evaluasi pola yang ditemukan yaitu melakukan penilaian terhadap pola, agar menjadi pengetahuan yang dapat dipergunakan untuk mendukung pengambilan sebuah keputusan.
- 6) Presentasi pengetahuan yaitu menyampaikan pengetahuan melalui teknik penggambaran atau visualisasi.

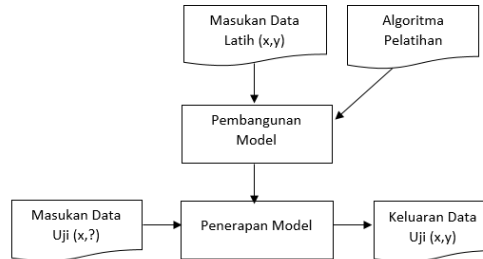
b. Pengertian Klasifikasi

Klasifikasi adalah cara untuk melihat perilaku dan atribut kelompok tertentu. Teknik ini dapat mengklasifikasikan data baru dengan mengolah data yang sudah ada yang telah diklasifikasikan dan menggunakan hasilnya untuk memberikan banyak aturan. Aturan ini digunakan untuk data baru yang akan diklasifikasikan.

Teknik ini menggunakan induksi yang diawasi, yang menggunakan serangkaian pengujian catatan klasifikasi untuk menentukan kategori lain (Kusnawi, 2007). Klasifikasi adalah proses mengevaluasi objek data untuk yang dimasukkan ke dalam kelas tertentu dari beberapa kelas yang telah tersedia. Terdapat 2 proses utama yang dilakukan untuk teknik klasifikasi, yaitu

- 1) Pembuatan model sebagai prototipe untuk disimpan sebagai memori dan
- 2) Tujuan dari model ini adalah agar memperkenalkan / mengklasifikasikan / memprediksi data lain untuk mengetahui kelas data tersebut berada dalam model yang disimpan. (Eko Prasetyo, 2012).

Sesuatu model dalam klasifikasi mempunyai makna yang sama dengan kotak hitam, terdapat model dalam kotak hitam yang menerima masukan, setelah itu bisa memikirkan masukan serta membagikan jawaban selaku hasil dari pemikiran tersebut. Proses kerja dari klasifikasi ditunjukkan pada Gambar 3 Angka ini membagikan banyak informasi latih(x, y) selaku informasi pembangunan model. Setelah itu pakai model buat memprediksi jenis bersumber pada informasi uji (x,?) Buat mengenali jenis sesungguhnya y.



Gambar 3. Proses Pekerjaan Klasifikasi

Klasifikasi dipergunakan untuk membuat model yang dapat memetakan setiap variabel ke setiap target, kemudian untuk menetapkan nilai target ke setiap variabel yang baru diperoleh dapat menggunakan model yang telah dibuat. Dalam mengembangkan model selama proses pelatihan, diperlukan suatu algoritma untuk membangunnya yang disebut dengan algoritma pelatihan. Banyak algoritma pelatihan yang telah dikembangkan oleh peneliti yaitu C4.5, Nearest Neighbor, Bayesian Classification, Neural Network, dll.

c. Metode Algoritma C4.5

Metode Algoritma C4.5 merupakan metode yang banyak digunakan untuk klasifikasi data yang memiliki atribut numerik dan kategorikal. Hasil proses klasifikasi berupa rules yang dapat dipergunakan untuk memprediksi nilai atribut tipe diskrit record baru. Algoritma C4.5 adalah pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal mengatasi data yang hilang, mampu mengatasi data kontinyu dan pemangkasan. (Larose, 2005).

Hal penting dalam induksi pohon keputusan adalah bagaimana menyatakan kondisi pengujian pada node. Ada 3 grup penting dalam persyaratan pengujian node:

- 1) Fitur biner
 Fitur dengan hanya dua nilai berbeda disebut fitur biner. Saat fungsi ini menjadi node (node root atau node internal), pengujian hanya memerlukan dua opsi cabang
- 2) Fitur bertipe kategorikal
 Fitur yang nilainya diklasifikasikan (nominal atau terurut) dapat memiliki beberapa nilai yang berbeda. Contohnya adalah bahwa fungsi "cuaca" memiliki 3 nilai berbeda dan dapat memiliki beberapa kombinasi untuk menyelesaikan kondisi pengujian. Biasanya ada dua yaitu binary split dan multiple split.
- 3) Fitur bertipe numeric
 Untuk fitur numerik, kondisi pengujian dari node (node root atau node internal) diwakili oleh tes perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau beberapa tes untuk hasil dengan banyak nilai. dari $v_{i-1} < v_i < v_{i+1}$, untuk $i = 1, 2, \dots, k$. Dalam kasus solusi biner, algoritme memeriksa semua kemungkinan posisi solusi v dan memilih posisi v terbaik. Untuk multi-mode, algoritme harus memeriksa semua kemungkinan rentang nilai berkelanjutan

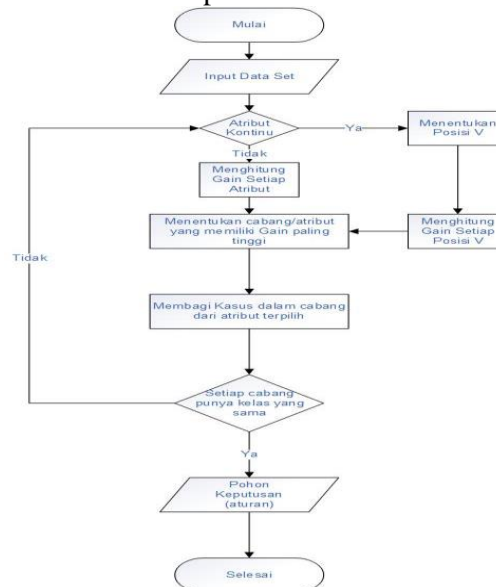
Dalam algoritma C4.5, kriteria yang paling banyak digunakan untuk memilih karakteristik pembagi adalah rasio gain, yang diwakili oleh persamaan berikut:

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

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.

Dapat dilihat pada Gambar 4 terdapat flowchart dari Algoritma C4.5 dan beserta penjelasannya :

- 1) Memasukan Data Set yang telah disediakan. Data Set adalah kumpulan objek dan atributnya.
- 2) Pada proses cek keputusan dilihat apakah data set tersebut termasuk Atribut Kontinu atau bukan, Atribut Kontinu adalah atribut yang mempunyai jangkauan real.
- 3) Jika ya, itu akan menentukan posisi V dan kemudian menghitung keuntungan dari setiap V. Jika tidak, hitung gain setiap atribut.
- 4) Setelah itu menentukan cabang yang memiliki gain paling tinggi.
- 5) Setelah itu membagi kasus dalam cabang dari atribut yang terpilih.
- 6) Jika setiap cabang memiliki kelas yang sama, pohon keputusan akan dibuat. Jika tidak, maka cabang atau atribut kembali ke pemeriksaan atribut kontinu.



Gambar 4. Flowchart Algoritma C4.5

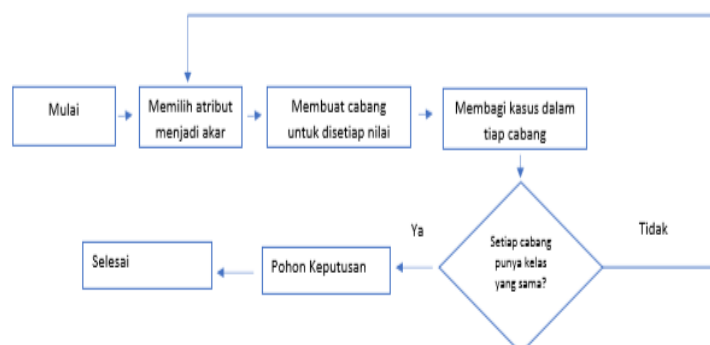
B. METODE

1. Model Konseptual

Metode konseptual yang digunakan adalah Algoritma C4.5, dalam memprediksi mahasiswa non – aktif, ada beberapa tahapan guna untuk memastikan upaya mencapai hasil yang maksimal dan sesuai dengan kebutuhan. Proses tersebut digambarkan dalam diagram alur proses metode Algoritma C4.5.

Tahapan dalam pembuatan pohon keputusan adalah sebagai berikut (Eko Prasetyo, 2013):

- a. Memilih atribut menjadi akar.
- b. Membuat cabang untuk disetiap nilai.
- c. Membagi kasus dalam setiap cabang.
- d. Ulangi proses tersebut untuk setiap cabang hingga semua kasus pada cabang tersebut memiliki kelas yang sama



Gambar 5. Gambar Alur Proses Algoritma C4.5

2. Subjek Ujicoba

Subjek ujicoba yang dilibatkan harus diidentifikasi karakteristiknya secara jelas dan lengkap, tetapi terbatas dalam kaitannya dengan produk yang dikembangkan. Subjek pengguna yang terlibat pada penelitian ini yaitu terdiri dari 5 orang dosen wali, 1 orang bagian BAAK dan 2 orang ketua program studi pada Fakultas Informatika dan Komputer di Universitas Binaniaga Indonesia, dan subjek ahli yang terlibat pada penelitian ini adalah 2 orang dosen ahli sistem metode.

3. Instrumen Penelitian

a. Instrumen Untuk Ahli

Instrumen yang digunakan untuk ahli sistem adalah berupa kuesioner tertutup. (Sugiyono, 2019, p.406) menyatakan bahwa “Instrumen penelitian adalah alat ukur seperti tes, kuesioner, pedoman wawancara dan pedoman observasi yang digunakan peneliti untuk mengumpulkan data dalam suatu penelitian”. Dalam penelitian ini ahli sistem adalah dosen yang paham mengenai sistem. Instrumen yang dipakai adalah pengujian black box. Pengujian black box yaitu menguji perangkat lunak dari segi spesifikasi fungsional tanpa menguji desain dan kode program (Rosa A.S dan M. Shalahudin, 2011). Kategori – kategori kesalahan yang diuji oleh pengujian black box adalah fungsi – fungsi yang salah salah atau hilang, kesalahan interface, kesalahan dalam struktur data atau akses database eksternal, kesalahan performa, kesalahan inisialisasi dan terminasi (Lila, 2018). Black Box Testing berfokus pada spesifikasi fungsional dari perangkat lunak. Tester dapat mendefinisikan kumpulan kondisi input dan melakukan pengetesan pada spesifikasi fungsional program. Black Box Testing bukanlah solusi alternatif dari White Box Testing tapi lebih merupakan pelengkap untuk menguji hal-hal yang tidak dicakup oleh White Box Testing (Mustaqbal et al., 2015).

Menurut (Al Bahra, 2006, p.379) Black Box Testing mengarahkan untuk mendapatkan hal seperti berikut :

- 1) Fungsi yang salah atau hilang
- 2) Kesalahan antarmuka
- 3) Kesalahan kinerja.
- 4) Kesalahan inisialisasi dan terminasi.

b. Instrumen Untuk Pengguna

Tabel 1. Instrumen Untuk Pengguna

| No. | Pernyataan | Tidak Setuju / Setuju | | | | | | |
|-----|--|-----------------------|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | Secara keseluruhan, saya puas dengan kemudahan penggunaan aplikasi ini | | | | | | | |
| 2 | Aplikasi mudah digunakan | | | | | | | |
| 3 | Saya secara efektif dapat menyelesaikan tugas-tugas dan scenario menggunakan aplikasi ini | | | | | | | |
| 4 | Saya bisa menyelesaikan tugas-tugas dan scenario menggunakan aplikasi ini | | | | | | | |
| 5 | Saya dengan efisien dapat menyelesaikan tugas-tugas dan scenario menggunakan aplikasi ini | | | | | | | |
| 6 | Saya merasa nyaman menggunakan aplikasi ini | | | | | | | |
| 7 | Mudah untuk belajar menggunakan aplikasi ini | | | | | | | |
| 8 | Saya percaya saya bisa menjadi produktif dengan cepat menggunakan aplikasi ini | | | | | | | |
| 9 | Aplikasi ini memberikan pesan kesalahan yang jelas memberitahu saya bagaimana untuk memperbaiki masalah | | | | | | | |
| 10 | Setiap kali saya melakukan kesalahan dengan menggunakan aplikasi, saya bisa pulih dengan mudah dan cepat | | | | | | | |
| 11 | Informasi (seperti online pesan bantuan pada layer, dan dokumentasi lainnya) disediakan dengan jelas oleh aplikasi ini | | | | | | | |

| | | | | | | | | |
|----|---|--|--|--|--|--|--|--|
| 12 | Mudah untuk menemukan informasi yang saya butuhkan | | | | | | | |
| 13 | Informasi yang disediakan aplikasi ini mudah dimengerti | | | | | | | |
| 14 | Informasi efektif dalam membantu menyelesaikan tugas-tugas dan skenario | | | | | | | |
| 15 | Organisasi informasi pada layer aplikasi jelas | | | | | | | |
| 16 | Antarmuka aplikasi ini menyenangkan | | | | | | | |
| 17 | Saya suka menggunakan antarmuka aplikasi ini | | | | | | | |
| 18 | Aplikasi ini memiliki semua fungsi dan kemampuan yang saya harapkan | | | | | | | |
| 19 | Secara keseluruhan, saya puas dengan aplikasi ini. | | | | | | | |

4. Teknik Analisa Data

a. Uji Produk

Dalam penelitian ini, metode analisis data dengan menggunakan presentase kelayakan. Adapun rumus yang digunakan adalah sebagai berikut:

$$\text{Presentasi kelayakan (\%)} = \frac{\text{Skor yang diobservasi}}{\text{Skor yang diharapkan}} \times 100\%$$

Hasil persentase digunakan untuk memberikan jawaban atas kelayakan aspek yang diteliti. Menurut Arikunto (2009, p.44), kualifikasi dibagi menjadi lima kategori. Rasio ini memperhitungkan kisaran persentase. Nilai harapan tertinggi adalah 100%, dan nilai minimum adalah 0%. Tabel 2 mencantumkan ruang lingkup kategori kualifikasi menurut Arikunto (2009, p.44).

Tabel 2 Kategori Kelayakan Menurut Arikunto

| Persentase Pencapaian | Interpretasi |
|-----------------------|--------------------|
| <21% | Sangat Tidak Layak |
| 21% - 40% | Tidak Layak |
| 41% - 60% | Cukup Layak |
| 61% - 80% | Layak |
| 81% - 100% | Sangat Layak |

(Sumber : Arikunto, 2009, p.44)

b. Uji Hasil

Confusion matrix adalah alat yang digunakan untuk mengevaluasi model klasifikasi untuk memperkirakan objek benar dan salah (F. Gorunescu, 2011). Matriks prediksi yang akan dibandingkan dengan kategori input, dengan kata lain memuat informasi tentang nilai aktual dan prediksi dari klasifikasi tersebut, berikut table confusion matrix :

Tabel 3. Confusion Matrix

| Classification | Predicted Class | |
|----------------|-----------------------|-----------------------|
| | Class = Yes | Class = No |
| Class = Yes | a (true positive-TP) | b (false negative-FN) |
| Class = No | c (false positive-FP) | d (true negative-TN) |

Akurasi adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus. Rumus untuk menghitung tingkat akurasi pada matrik adalah :

Keterangan :

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{A + D}{A + B + C + D}$$

- A = jika hasil prediksi positif dan data sebenarnya positif
- B = jika hasil prediksi negatif dan data sebenarnya positif
- C = jika hasil prediksi positif dan data sebenarnya negatif
- D = jika hasil prediksi negatif dan data sebenarnya negatif

C. HASIL DAN PEMBAHASAN

1. Hasil

a. Analisis Kebutuhan dan Hasil Analisis Kebutuhan

Tahap analisis kebutuhan dilakukan dengan cara pengumpulan dokumen dan wawancara . Adapun beberapa dokumen yang dikumpulkan adalah Data mahasiswa dan Status Pembayaran Mahasiswa Angkatan 2015/2016, yang diperoleh dari Biro Administrasi Akademik dan data status pembayaran mahasiswa diperoleh dari Biro Administrasi Umum.. Berdasarkan pada analisis dokumen yang dilakukan terdapat penurunan jumlah mahasiswa aktif disetiap semesternya, faktor yang mengakibatkan hal tersebut adalah keterlambatan mahasiswa melakukan pembayaran, dan keterlambatan mahasiswa melakukan perwalian. Untuk Wawancara dilakukan dalam rangka mengetahui proses prediksi yang berlangsung dan dilakukan kepada Kepala Biro Administrasi Keuangan Sumber Daya dan Kehumasan dan Kepala Biro Administrasi Akademik. Berdasarkan pada hasil wawancara diketahui proses bisnis lama. Pada proses bisnis lama tersebut terdapat kelemahan yaitu belum tepatnya dalam penanganan mahasiswa non – aktif, tidak terkoordinasi dengan baik dalam memonitoring mahasiswa non – aktif, dan penanganan hanya berdasarkan mahasiswa mengikuti Ujian Akhir Semester dan melakukan perwalian saja. Berdasarkan hasil analisis terhadap pengumpulan kebutuhan yang telah dilakukan diperlukan sebuah sistem informasi yang dapat digunakan untuk memprediksi mahasiswa non-aktif agar mempermudah dosen wali dalam memonitoring mahasiswa. Untuk dapat melakukan prediksi di tetapkan metode algoritma C.45 sebagai metode komputansi yang akan di terapkan pada sistem informasi prediksi mahasiswa non aktif. Dengan tahap-tahapan sebagai berikut :

1) Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data mahasiswa tahun akademik 2016/2017 sebanyak 172. Data mahasiswa dapat di lihat pada Tabel 4.

Tabel 4 Data Mahasiswa

| NPM | Nama | Tempat Lahir | Tanggal lahir | Agama | Jenis Kelamin | Warga Negara |
|----------|---------------|--------------|---------------|---------|---------------|--------------|
| 14150002 | Agatha Apri | Jakarta | 20/04/1995 | Kristen | Perempuan | Indonesia |
| 14150005 | Aldi Rivaldi | Bogor | 21/10/1995 | Islam | Laki - Laki | Indonesia |
| 14150009 | Anrico Rizki | Bogor | 10/05/1997 | Islam | Laki - Laki | Indonesia |
| 14150010 | Ardian H | Depok | 10/04/1997 | Islam | Laki - Laki | Indonesia |
| 14150012 | Asep Rifa'i | Bogor | 20/08/1992 | Islam | Laki - Laki | Indonesia |
| 14150015 | Della Lintang | Bogor | 11/12/1996 | Islam | Perempuan | Indonesia |
| 14150016 | Deni Lukman | Bogor | 09/06/1995 | Islam | Laki - Laki | Indonesia |
| 14150018 | Dinda Damar | Bogor | 29/09/1997 | Islam | Perempuan | Indonesia |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |
| 14157062 | Saeful Bahri | Bogor | 02/06/1992 | Islam | Laki - Laki | Indonesia |

Setelah data mahasiswa terkumpul, langkah selanjutnya adalah melakukan data selection dan penentuan variable

2) Data Selection

Data yang terkumpul pada tabel 4 memiliki beberapa atribut yang tidak terpakai dalam proses pengolahan data. Atribut yang tidak memiliki kontribusi terhadap tujuan penelitian, tidak digunakan dalam penelitian ini. Tabel 5 dibawah ini merupakan hasil dari proses data selection. Seleksi yang dilakukan adalah dengan menghapus field – field yang tidak digunakan, menghapus isi field yang kosong, da menambahkan data yang memiliki kontribusi terhadap tujuan penelitian seperti data Indeks Prestasi Semesr 7, Absensi semester 7, Sumber biaya, Penghasilan Per Tahun, Status Pembayaran dan field keterangan. Setiap mahasiswa

Tabel 5 Data Mahasiswa Setelah Dilakukan Data Selection

| No | NPM | Nama | IPS | Absensi | Sumber Biaya | Penghasilan Per tahun | Status Pembayaran | Ket |
|----|----------|--------------|------|---------|--------------|-----------------------|-------------------|-------------|
| 1 | 14150002 | Agatha | 2.67 | 12 | Orang Tua | < 24 jt | Lunas | Non - Aktif |
| 2 | 14150005 | Aldi Rivaldi | 0.83 | 6 | Sendiri | < 24 jt | Belum Lunas | Non - Aktif |

| | | | | | | | | |
|-----|----------|-----------------------|------|-----|-----------|---------|-------------|-------------|
| 3 | 14150009 | Anrico Rizki | 3.38 | 12 | Orang Tua | < 24 jt | Lunas | Aktif |
| 4 | 14150010 | Ardian Haryanto | 2.58 | 0 | Sendiri | < 24 jt | Lunas | Non - Aktif |
| 5 | 14150012 | Asep Rifa'i | 3.17 | 13 | Sendiri | < 24 jt | Lunas | Aktif |
| 6 | 14150015 | Della Lintang Melinda | 3.46 | 12 | Orang Tua | < 24 jt | Lunas | Aktif |
| 7 | 14150016 | Deni Lukman Hakim | 3.31 | 12 | Orang Tua | < 24 jt | Lunas | Aktif |
| 8 | 14150018 | Dinda Damar Sastrri | 2.00 | 14 | Sendiri | < 24 jt | Belum Lunas | Non - Aktif |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 70 | 15150046 | M. Ridho M | 1.56 | 14 | Sendiri | < 24 jt | Belum Lunas | Non - Aktif |

3) Variabel dalam Penelitian

Variabel yang digunakan dalam prediksi mahasiswa non – aktif ini adalah IPS semester 7, Absensi semester 7, sumber biaya, penghasilan per tahun, dan status pembayaran semester 7

Tabel 6 variable dan Kategori

| NAMA VARIABEL | Kategori | Nama Variabel | Kategori |
|-------------------|----------------|-----------------------|-------------------|
| IPS | <2.00 | Sumber Biaya | Orang Tua |
| | >=2.00 & <3.00 | | Wali |
| | >=3.00 & <3.50 | | Sendiri |
| | >=3.50 | | Perusahaan |
| Absensi | <8 | Penghasilan Per Tahun | Kurang dari 24 jt |
| | >=8 dan <16 | | 24 jt - 48 jt |
| | >=16 | | 48 jt - 60 jt |
| Status Pembayaran | Lunas | | Lebih dari 60 jt |
| | Belum Lunas | | |

4) Perhitungan Algoritma c4.5

Untuk membangun pohon keputusan dalam algoritma C4.5 secara umum perhitungannya sebagai berikut :

Melakukan perhitungan dari jumlah total kasus yang ada, jumlah total kasus yang aktif dan tidak aktif. Setelah itu lakukan perhitungan untuk mencari nilai Entropy dan Gain dari setiap atribut.

Menghitung nilai Entropy dari total kasus :

$$Entropy(total)$$

$$= \left(-\frac{47}{70} \times \log_2 \left(\frac{47}{70}\right)\right) + \left(-\frac{23}{70} \times \log_2 \left(\frac{23}{70}\right)\right) = 0,9135$$

Hitung Nilai Entropy IPS (indeks prestasi semester)

$$Entropy(<2.00)$$

$$= \left(-\frac{0}{6} \times \log_2 \left(\frac{0}{6}\right)\right) + \left(-\frac{6}{6} \times \log_2 \left(\frac{6}{6}\right)\right) = 0$$

$$Entropy(>=2.00 \ \& \ <3.00)$$

$$= \left(-\frac{6}{22} \times \log_2 \left(\frac{6}{22}\right)\right) + \left(-\frac{16}{22} \times \log_2 \left(\frac{16}{22}\right)\right) = 0,8454$$

$$Entropy(>=3.00 \ \& \ <3.50)$$

$$= \left(-\frac{25}{26} \times \log_2 \left(\frac{25}{26}\right)\right) + \left(-\frac{1}{26} \times \log_2 \left(\frac{1}{26}\right)\right) = 0,2352$$

$$Entropy(>=3.50)$$

$$= \left(-\frac{16}{16} \times \log_2 \left(\frac{16}{16}\right)\right) + \left(-\frac{0}{16} \times \log_2 \left(\frac{0}{16}\right)\right) = 0$$

Setelah selesai menghitung nilai entropy dari variable dan kategori (atribut) nilai indeks prestasi semester (IPS) selanjutnya dilakukan perhitungan untuk mencari nilai

gain dari masing-masing atribut. Temukan nilai keuntungan dari setiap atribut. Setelah mendapatkan nilai entropi dari tiap atribut, nilai gain dihitung dengan mengurangkan nilai total dikalikan jumlah kasus dan mengalikan entropi tiap kasus dengan total entropi. Perhitungan masing – masing nilai Gain dari empat atribut adalah sebagai berikut :

$$\text{Gain (Total, IPS)} = 0,9135 - \left(\left(\frac{6}{70} \right) \times 0 \right) + \left(\left(\frac{22}{70} \right) \times 0,8454 \right) + \left(\left(\frac{26}{70} \right) \times 0,2352 \right) + \left(\left(\frac{16}{70} \right) \times 0 \right) = 0,5604$$

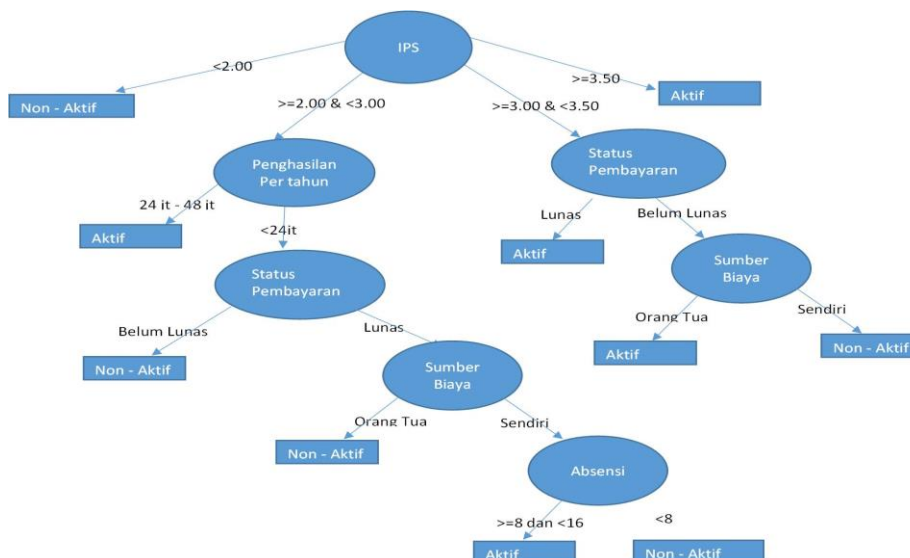
$$\text{Gain (Total, Absensi)} = 0,9135 - \left(\left(\frac{4}{70} \right) \times 0 \right) + \left(\left(\frac{66}{70} \right) \times 0,8660 \right) + \left(\left(\frac{0}{70} \right) \times 0 \right) = 0,0970$$

$$\text{Gain (Total, Sumber Biaya)} = 0,9135 - \left(\left(\frac{17}{70} \right) \times 0,7871 \right) + \left(\left(\frac{0}{70} \right) \times 0 \right) + \left(\left(\frac{53}{70} \right) \times 0,9414 \right) + \left(\left(\frac{0}{70} \right) \times 0 \right) = 0,0095$$

$$\text{Gain (Total, Penghasilan per tahun)} = 0,9135 - \left(\left(\frac{63}{70} \right) \times 0,9183 \right) + \left(\left(\frac{6}{70} \right) \times 0,6500 \right) + \left(\left(\frac{0}{70} \right) \times 0 \right) + \left(\left(\frac{1}{70} \right) \times 0 \right) = 0,0313$$

$$\text{Gain (Total, Status Pembayaran)} = 0,9135 - \left(\left(\frac{50}{70} \right) \times 0,4690 \right) + \left(\left(\frac{20}{70} \right) \times 0,4690 \right) = 0,4445$$

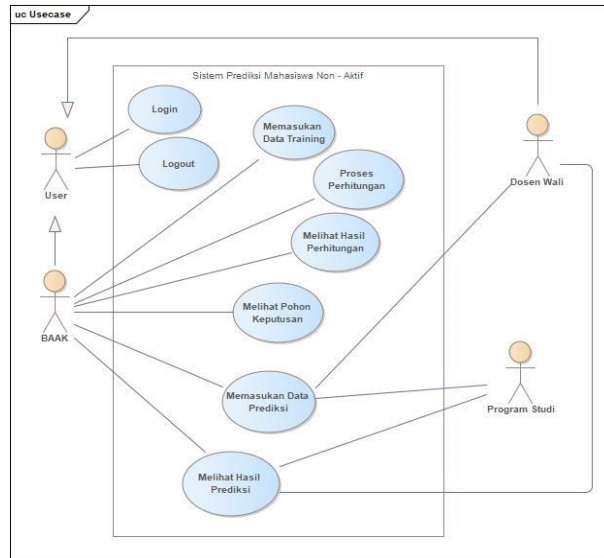
Dari hasil perhitungan nilai Entropy dan Gain seperti terlihat bahwa atribut yang memiliki nilai Gain tertinggi adalah atribut IPS yaitu 0,5604, dengan demikian atribut IPS menjadi root. simpul. Ada empat nilai yang terdapat pada atribut IPS yaitu $<2.00, \geq 2.00 \ \& \ <3.00, \geq 3.00 \ \& \ <3.50, \geq 3.50$. Dari keempat nilai tersebut dihitung untuk menemukan simpul cabang dari nilai atribut IPS. Cabang perhitungan ini diselesaikan dengan mencari jumlah kasus dalam kategori aktif dan tidak aktif dan nilai entropi dari semua kasus $IPS \geq 2.00 \ \& \ <3.00$, dan kemudian melakukan perhitungan untuk menemukan nilai entropi dan gain tertinggi dari nilai aktif. Setelah selesai menghitung nilai entropy , gain dan menemukan simpul untuk atribut nilai indeks prestasi semester (IPS) dilanjutkan menghitung nilai entropy, gain dan simpul variable dan atribut Absensi, Sumber Biaya, Penghasilan Per Tahun, dan Status Pembayaran. Berdasarkan hasil perhitungan yang telah dilakukan, diperoleh hasil berupa pohon keputusan (*decision tree*) yang dipergunakan sebagai dasar untuk memprediksi mahasiswa yang non – aktif.



Gambar 6. Pohon Keputusan Prediksi mahasiswa yang non – aktif

b. Desain

Tahapan yang terdapat dalam model komputasi Algoritma C. 45 dijadikan sebagai dasar dalam melakukan pengembangan sistem informasi prediksi mahasiswa non aktif, yang digambarkan ke dalam usecase diagram yang dapat dilihat pada gambar 3.



Gambar 7. Usecase Diagram

Struktur sistem pada prediksi mahasiswa non aktif digambarkan dengan *class diagram* yang di tunjukan pada gambar 4. Diagram ini terdiri dari 4 tabel, 6 proses dan 5 form. Tabel – tabel tersebut terdiri dari user, tbl_data_mahasiswa, atribut, iterasi_c45, mining_c45, pohon_keputusan_c45, dataujisingel dan datauji dan tbl_hasil. Kemudian proses – proses meliputi validasi login, proses logout, proses get data tbl_prediksi, proses input data tbl_prediksi, proses get data tbl_data_mahasiswa, proses input data tbl_data_mahasiswa, proses input data dataujisingel dan proses get data tbl_data_mahasiswa untuk form melihat hasil perhitungan C4.5.

2. Pembahasan

Pengujian terhadap *prototype* sistem prediksi mahasiswa non aktif dilakukan dengan menyebarkan kuisisioner kepada responden pengguna sistem. uji coba dilakukan agar mengetahui kesalahan, ketidaktepatan, ketidakjelasan dan kelayakan pada proses pembuatan sebuah produk atau pun sistem. Pengujian dilakukan menggunakan *instrument Post-study System Usability Quistionnaire (PSSUQ)* yang terdiri dari 19 item pertanyaan dan membandingkan total jawaban yang di peroleh dari responden dengan jawaban ideal yang diharapkan, Hasil uji produk pengguna dapat dilihat pada tabel 7.

Tabel 7. Hasil Uji Produk

| No | Aspek Penilaian | Skor | Skor Diharapkan | Kelayakan |
|----|-----------------|------|-----------------|-----------|
| 1 | Overall | 467 | 532 | 87,78% |
| 2 | Sysuse | 200 | 224 | 89,28% |
| 3 | Infoqual | 171 | 196 | 87,24 % |
| 4 | Interqual | 72 | 84 | 85,71 |

Dari hasil perhitungan ujicoba produk diperoleh hasil kelayakan keseluruhan *prototype* sistem (overall) sebesar 87,78%, berdasarkan tabel kategori kelayakan menurut Arikanto nilai tersebut masuk dalam kategori Sangat layak , yang artinya *prototype* yang di buat Sangat layak untuk diimplementasikan . Selain dilakukan pengukuran kelayakan *prototype* produk yang di kembangkan, dilakukan juga pengukuran keakuratan keluaran prediksi mahasiswa non aktif yang dihasilkan dari metode C.45. Menggunakan confussion matrix. Perhitungan nilai keakuratan dilakukan dengan membandingkan hasil prediksi data uji dengan data aktual atau data nyata yang sesuai dengan variabel yang sudah ditetapkan. Hasil perhitungan tersebut dapat dilihat dari tabel 7

Berdasarkan tabel 4.18 maka berikut adalah perhitungan akurasi dengan cara seperti dibawah ini:

$$\text{Akurasi} = \frac{59 + 13}{59 + 6 + 11 + 13} \times 100\% = 81\%$$

Berdasarkan perhitungan nilai akurasi, maka hasil akurasi yang didapatkan sebesar 81%

D. KESIMPULAN

Berdasarkan hasil penelitian yang telah diselesaikan, dapat ditarik kesimpulan sebagai berikut :

1. Dapat memberikan rekomendasi mahasiswa yang akan non – aktif berdasarkan hasil prediksi dengan variabel ips, absensi, penghasilan per tahun, sumber biaya, dan status pembayaran.
2. Dalam pembuatan aplikasi untuk memprediksikan mahasiswa non – aktif dengan menerapkan pemodelan Algoritma C4.5 menjadi lebih efektif untuk proses pendataan ketidakaktifan mahasiswa. Telah dilakukan uji pengguna dengan menggunakan PSSUQ, dengan sesuai kategori diantaranya kategori Skor kepuasan secara keseluruhan (OVERALL) sebesar 87,78%, kegunaan sistem (SYSUSE) sebesar 89,28%, kualitas informasi (INFOQUAL) sebesar 87,24% dan kualitas antarmuka (INTERQUAL) sebesar 85,71%.
3. Dengan menerapkannya metode algoritma C4.5 pada sistem prediksi mahasiswa non – aktif ini menghasilkan nilai akurasi dari sistem prediksi dengan menggunakan confusion matrix sebesar 81%

E. DAFTAR PUSTAKA

- [1] Arikunto, S. (2006). *Prosedur Penelitian Suatu Tindakan Praktik*. Jakarta: Rineka Cipta.
- [2] Andriani, A. (2012). Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout. *Seminar Nasional Matematika 2012*
- [3] Eko Prasetyo. (2013). Data Mining: Konsep Dan Aplikasi Menggunakan Matlab. In *Journal of Chemical Information and Modeling*.
- [4] F. Gorunescu, *Data Mining Concept, Models and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.
- [5] Hastuti, K. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*.
- [6] Himawan, D. (2011). Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang. *Fakultas Ilmu Komputer*.
- [7] Kusri and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009
- [8] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. In *Discovering Knowledge in Data: An Introduction to Data Mining*.
- [9] Prasetyo, Eko. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : Andi.
- [10] Rosa A.S dan M. Shalahudin. (2011). *Rekayasa Perangkat Lunak (Terstruktur & Berorientasi Objek)*. Politeknik Negri Sriwijaya.
- [11] Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif dan R & D*. Bandung: Alfabeta.
- [12] Turban, E., dkk, 2001, *Decision Support Systems and Intelligent Systems*, Andi Offset, Yogyakarta
- [13] Untari, D. (2014). Data Mining untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5. *Fakultas Ilmu Komputer Universitas Dian Nuswantoro*.
- [14] Van Blerkom, M. L. (2009). Measurement and statistics for teachers. In *Measurement and Statistics for Teachers*. <https://doi.org/10.4324/9781315464770>