

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA PENERIMA BANTUAN SOSIAL COVID-19

Eugenius Edsel Barito ¹⁾ Jap Tji Beng ²⁾ Desi Arisandi ³⁾

¹⁾²⁾³⁾ Teknik Informatika, FTI, Universitas Tarumanagara

Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia

¹⁾email : eugeniusedsel14@gmail.com, ²⁾email : t.jap@untar.ac.id, ³⁾email : desi@fti.untar.ac.id

ABSTRACT

Dengan adanya pandemi Covid-19 yang mempengaruhi kondisi sosial-ekonomi masyarakat, banyak mahasiswa yang berkuliah di Universitas Tarumanagara membutuhkan bantuan sosial agar dapat memenuhi kebutuhan hidup. Pembuatan aplikasi ini bertujuan untuk membantu melakukan proses klasifikasi mahasiswa Universitas Tarumanagara yang termasuk calon penerima bantuan sosial Covid-19. Karena terbatasnya jumlah bantuan yang dapat diberikan kepada mahasiswa maka dari itu akurasi data penerima bantuan sosial diperlukan agar penyaluran bantuan sosial dalam upaya mengatasi dampak pandemi Covid-19 bisa tepat sasaran sehingga para mahasiswa bisa mendapatkan bantuan secara adil. Parameter atau kriteria yang digunakan untuk klasifikasi mahasiswa penerima bantuan sosial Covid-19 adalah status kondisi finansial, Jumlah anggota keluarga, status kepemilikan tempat tinggal, Status pekerjaan kepala keluarga atau penanggung biaya hidup calon penerima bantuan, dan Status kesehatan calon penerima dan keluarganya terkait dengan Covid-19. Metode yang digunakan untuk mengolah data mahasiswa adalah pohon keputusan dengan Algoritma C4.5. Dengan mengumpulkan 500 data mahasiswa Universitas Tarumanagara digunakan 400 mahasiswa sebagai data pelatihan dan 100 mahasiswa sebagai data pengujian dan hasil dari lima kali percobaan pengujian data tersebut menunjukkan bahwa aplikasi ini sudah berfungsi dengan baik dalam mengklasifikasi mahasiswa calon penerima bantuan sosial dengan rata-rata accuracy sebesar 89%, precision sebesar 90.16%, dan recall sebesar 83.27%.

Key words

Algoritma C4.5, Bantuan Sosial Covid-19, Data Mining, Klasifikasi, Pohon Keputusan

1. Pendahuluan

1.1 Latar Belakang

Data mining merupakan suatu proses untuk menemukan pola dan pengetahuan yang menarik dalam jumlah data yang besar. Sumber data dapat mencakup basis data, gudang data, web, penyimpanan informasi lainnya, atau data yang secara dinamis dialirkan ke dalam sistem [1].

Algoritma C4.5 merupakan salah satu algoritma dalam pohon keputusan (decision tree) yang digunakan untuk klasifikasi. Algoritma C4.5 merupakan pengembangan lebih lanjut dari algoritma ID3 yang diusulkan oleh Quinlan. Algoritma C4.5 menggunakan information gain dan gain ratio sebagai ukuran saat memilih atribut [2].

Algoritma C4.5 dipilih untuk pembuatan pohon keputusan dalam perancangan aplikasi ini karena Algoritma C4.5 merupakan algoritma klasifikasi yang paling banyak digunakan dalam machine learning dan pemrosesan data serta kinerjanya termasuk dalam salah satu algoritma yang tercepat dibandingkan dengan algoritma lainnya. Kelebihan dari Algoritma C4.5 adalah dapat menangani atribut kontinu dan diskrit, dapat menangani data pelatihan dengan nilai atribut yang hilang, serta dapat memangkas pohon setelah dibuat. Pada Algoritma C4.5 data diolah dari model pohon keputusan menjadi aturan yang sederhana sehingga mudah dipahami [3]. Algoritma C4.5 dapat diterapkan untuk penelitian masalah yang terjadi di kehidupan nyata seperti salah satunya masalah yang ditimbulkan oleh pandemi Covid-19.

Pandemi Covid-19 berdampak tidak hanya pada kesehatan, tetapi juga pada situasi sosial dan ekonomi, baik individu maupun keluarga. Dalam jangka pendek, dampak kesehatan ditunjukkan dengan angka penyebaran yang makin meningkat dengan tingkat kematian korban di Indonesia juga semakin tinggi. Kondisi ini akan berdampak pada anjloknya aktivitas perekonomian domestik, yang tidak menutup kemungkinan akan menurunkan kesejahteraan masyarakat. Pandemi Covid-19 juga melanda banyak sektor usaha, menyebabkan terjadinya pemutusan hubungan kerja, dan pengurangan lapangan kerja. Dikhawatirkan ketidakstabilan sosial dapat terjadi jika situasi ini tidak diantisipasi dengan baik.

Selain itu dengan adanya pandemi Covid-19, seluruh pelajar atau mahasiswa yang berada di daerah padat penduduk seperti di wilayah DKI Jakarta juga terpengaruh terhadap kondisi sosial-ekonomi yang memburuk. Salah satu Universitas besar di Indonesia yang berada di wilayah DKI Jakarta adalah Universitas Tarumanagara. Masih banyak mahasiswa yang berkuliah di Universitas Tarumanagara membutuhkan bantuan sosial agar dapat memenuhi kebutuhan hidup pada saat kondisi pandemi seperti ini.

Pada saat ini mekanisme pemberian bantuan di Universitas Tarumanagara salah satunya berupa bantuan uang kuliah tunggal (UKT) dari pemerintah untuk mahasiswa yang membutuhkan. Bantuan UKT mahasiswa diberikan dalam bentuk uang tunai sebesar Rp2.400.000,00 (dua juta empat ratus ribu rupiah) per mahasiswa yang didukung oleh dokumen yang valid. Ketentuan umum yang berlaku dalam penentuan penerima bantuan ini adalah aktif sebagai mahasiswa, tidak sedang menerima beasiswa dari instansi lain, tidak mendapatkan sanksi akademik, dan mengisi informasi mengenai mahasiswa dan melengkapi berkas yang diperlukan seperti surat keterangan mengalami kendala finansial. Pada pemberian bantuan UKT untuk semester ganjil 2020/2021 terdapat 500 mahasiswa Universitas Tarumanagara yang mendapatkan bantuan tersebut. Kendala yang ada sehingga dibutuhkan perancangan sistem ini adalah terbatasnya jumlah bantuan yang dapat diberikan kepada mahasiswa. Maka dari itu akurasi data penerima bantuan sosial diperlukan agar penyaluran bantuan sosial dalam upaya mengatasi dampak pandemi Covid-19 tepat sasaran sehingga para mahasiswa yang membutuhkan bisa mendapatkan bantuan secara adil.

Kriteria untuk menentukan calon penerima bantuan sosial Covid-19 dikatakan mendapatkan atau tidak mendapatkan bantuan sosial ditentukan berdasarkan kesimpulan yang berasal dari beberapa situs web oleh pemerintah mengenai syarat atau kebijakan untuk mendapatkan bantuan sosial Covid-19 yang diterapkan kepada masyarakat, dalam hal ini syarat atau kriteria tersebut diterapkan kepada mahasiswa Universitas Tarumanagara. Kriteria tersebut antara lain adalah calon penerima harus merupakan mahasiswa aktif yang berkuliah di Universitas Tarumanagara dalam masa pandemi Covid-19. Kepala keluarga atau penanggung biaya hidup calon penerima mengalami kendala finansial. Jumlah anggota keluarga calon penerima bantuan. Status tempat tinggal calon penerima berupa rumah milik sendiri/keluarga atau sewa. Status pekerjaan kepala keluarga atau penanggung biaya hidup calon penerima di masa pandemi Covid-19. Status kesehatan calon penerima dan keluarganya terkait dengan Covid-19.

Dasar dari pemilihan kriteria calon penerima bantuan sosial mengalami kendala finansial dan jumlah anggota keluarga calon penerima bantuan adalah berasal dari program pemerintah dalam pemberian bantuan UKT (Uang Kuliah Tunggal) untuk mahasiswa yang membutuhkan. Dasar dari pemilihan kriteria status tempat tinggal calon penerima bantuan berasal dari program pemerintah yang sudah ada sebagai bentuk bantuan sosial rumah tidak layak huni (RTLH) di Boyolali kepada masyarakat yang kurang mampu. Dasar dari pemilihan kriteria status pekerjaan kepala keluarga atau penanggung biaya hidup calon penerima di masa pandemi Covid-19 berasal dari program pemerintah yang sudah ada yang bernama Bantuan Langsung Tunai (BLT). Bantuan Langsung Tunai (BLT) adalah bantuan yang berasal dari alokasi dana desa pada Anggaran Pendapatan Belanja Desa (APB Desa) yang akan diberikan oleh

pemerintah kepada masyarakat yang kehilangan mata pencaharian karena pandemik virus corona atau Covid-19. Dasar dari pemilihan kriteria status kesehatan calon penerima dan keluarganya terkait dengan Covid-19 berasal dari program pemerintah yang sudah ada seperti sebagai bentuk kepedulian kepada pasien yang telah sembuh dari infeksi virus covid-19, Pemerintah Kota Solok melalui Dinas Sosial menyerahkan bantuan berupa uang sebesar Rp 1.000.000,- untuk masing masing penyintas covid-19 di Kota Solok.

1.3 Tujuan Rancangan

1. Untuk menerapkan dan melakukan analisis teknik klasifikasi menggunakan metode pohon keputusan (decision tree) dengan Algoritma C4.5 pada calon penerima bantuan sosial Covid-19 agar mendapatkan bantuan sosial untuk memenuhi kebutuhan hidup dalam masa pandemi Covid-19.
2. Untuk memberikan informasi bagi pemberi bantuan sosial Covid-19 mengenai calon penerima bantuan agar dapat memberi bantuan dengan tepat sasaran.

1.4 Batasan Rancangan

1. Data yang diambil hanya data responden yang merupakan mahasiswa yang aktif berkuliah di Universitas Tarumanagara dalam masa pandemi Covid-19, dan aplikasi ini dapat diakses oleh mahasiswa Universitas Tarumanagara yang terdaftar di dalam basis data yang dikelola oleh administrator. Alasan pembatasan rancangan tersebut adalah karena terbatasnya penulis dalam mengumpulkan data dalam skala besar dalam masa pandemi Covid-19.
2. Hanya membahas tentang klasifikasi mahasiswa penerima bantuan sosial Covid-19 dan Metode yang digunakan untuk mengolah data responden adalah pohon keputusan (decision tree) dengan Algoritma C4.5.
3. Parameter yang digunakan untuk klasifikasi mahasiswa penerima bantuan sosial Covid-19 adalah status atau kondisi finansial calon penerima bantuan akibat Covid-19, Jumlah anggota keluarga inti yang ditanggung oleh kepala keluarga atau penanggung biaya hidup keluarga calon penerima bantuan, status kepemilikan tempat tinggal calon penerima bantuan, Status pekerjaan kepala keluarga atau penanggung biaya hidup calon penerima di masa pandemi Covid-19, dan Status kesehatan calon penerima dan keluarganya terkait dengan Covid-19.

1.4 Rancangan yang Sudah Dibuat

1. Salah satu rancangan yang sudah dibuat sebelumnya adalah perancangan sistem oleh Stefanny Claudia bersama Tri Sutrisno yang terbit pada Jurnal Ilmu Komputer Dan Sistem Informasi (JIKSI) Vol.6 No.2 Agustus 2018, yang berjudul "Analisis Rekomendasi Peminatan Menggunakan Metode Decision Tree Dengan Algoritma C4.5". Hasil perancangan sistem ini menunjukkan perbandingan akurasi dari kelima model pohon yang dipilih dari data pelatihan dan

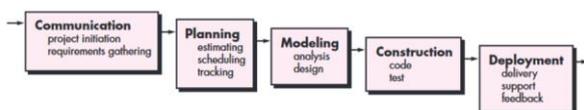
preferensi tesis siswa dengan presentase rata-rata akurasi dari kelima percobaan sebesar 72,6227% [4]. Persamaan perancangan sistem ini dengan skripsi penulis yaitu, menerapkan metode decision tree dengan Algoritma C4.5. Perbedaan perancangan sistem ini dengan skripsi penulis adalah perancangan sistem ini menerapkan Algoritma C4.5 untuk analisis rekomendasi peminatan, sedangkan penulis menerapkan Algoritma C4.5 untuk klasifikasi mahasiswa penerima bantuan sosial Covid-19.

2. Landasan Teoritik

2.1 Sistem yang Dirancang

Sistem yang dirancang merupakan aplikasi untuk menentukan mahasiswa penerima bantuan sosial Covid-19 menggunakan metode pohon keputusan (decision tree) dengan Algoritma C4.5 yang berbasis web. Sistem ini akan dirancang menggunakan bahasa pemrograman C# dan framework web yang digunakan adalah ASP.Net. Perangkat lunak yang digunakan untuk mengembangkan aplikasi ini adalah Microsoft Visual Studio. Sistem basis data yang digunakan untuk mengintegrasikan data dalam aplikasi ini adalah Microsoft SQL Server.

Proses perancangan sistem pada aplikasi ini menggunakan SDLC (Systems development life cycle) model air terjun (waterfall). Model air terjun (*waterfall*) atau yang sering disebut sebagai siklus hidup klasik (classic life cycle), dimana pendekatan dideskripsikan secara sistematis dan juga berurutan dalam pengembangan perangkat lunak, yang dimulai dengan spesifikasi kebutuhan pengguna kemudian dilanjutkan dengan tahapan perencanaan (planning), pemodelan (modeling), konstruksi (construction), dan penyerahan sistem kepada pengguna (deployment), lalu diakhiri dengan dukungan pada perangkat lunak yang telah selesai dibuat [6].



Gambar 1 Model Air Terjun

Tahapan-tahapan dari model air terjun dalam pengembangan sistem pada aplikasi ini adalah sebagai berikut :

1. Komunikasi (Communication)

Pada tahapan ini diperlukan komunikasi yang bertujuan untuk memahami perangkat lunak yang diharapkan oleh pengguna dan batasan perangkat lunak tersebut. Selain itu juga dilakukan pengumpulan Informasi mengenai hal-hal yang dibutuhkan seperti perangkat keras dan perangkat lunak yang akan digunakan dalam pembuatan aplikasi.

2. Perencanaan (Planning)

Pada tahapan ini diperlukan perencanaan mengenai estimasi penjadwalan pembuatan aplikasi dan melacak hal apa saja yang perlu disiapkan dalam pembuatan aplikasi.

3. Pemodelan (Modeling)

Pada tahapan ini diperlukan analisis pemodelan dan membuat desain antarmuka perangkat lunak untuk mendefinisikan arsitektur sistem secara keseluruhan.

4. Konstruksi (Construction)

Pada tahapan ini diperlukan proses pengembangan program dengan pengkodean (coding) dan proses pengujian program untuk mengecek kesalahan pada program agar dapat diperbaiki atau dikoreksi.

5. Penyerahan (Deployment)

Pada tahapan akhir ini dilakukan proses penyerahan aplikasi kepada pengguna, perangkat lunak yang sudah dibuat dapat digunakan dan dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada langkah sebelumnya tetapi ditemukan dari umpan balik pengguna.

2.2 Landasan Teoretik

2.2.1 Bantuan Sosial COVID-19

Menurut Peraturan Presiden Republik Indonesia Nomor 63 Tahun 2017 Tentang penyaluran Bantuan Sosial Secara Non Tunai, bantuan sosial adalah bantuan berupa uang, barang, atau jasa kepada seseorang, keluarga, kelompok atau masyarakat miskin, tidak mampu, dan/atau rentan terhadap risiko sosial.

Covid-19 adalah penyakit menular yang disebabkan oleh jenis coronavirus yang baru ditemukan. Virus baru dan penyakit yang disebabkan ini tidak dikenal sebelum mulainya wabah di Wuhan, Tiongkok, bulan Desember 2019. Covid-19 ini sekarang menjadi sebuah pandemi yang terjadi di banyak negara di seluruh dunia.

2.2.2 Data Mining

Data mining adalah proses menemukan informasi yang berguna secara otomatis dalam penyimpanan data yang besar. Teknik data mining digunakan untuk mengeksplorasi basis data besar untuk menemukan pola baru yang berguna yang mungkin tidak diketahui.

Data mining didefinisikan sebagai proses menemukan informasi tersembunyi dalam database. Selain itu data mining juga biasa dikenal sebagai analisis data eksplorasi, penemuan yang didorong data, dan pembelajaran deduktif. Berdasarkan tugas dasar yang dapat dilakukan, data mining dapat dibagi menjadi beberapa kelompok, seperti teknik klasifikasi, regresi, prediksi, clustering, dan asosiasi [7].

2.2.3 Klasifikasi

Klasifikasi adalah salah satu teknik dalam data mining yang memetakan data ke dalam kelompok atau kelas yang telah ditentukan. Klasifikasi termasuk metode pembelajaran terawasi (supervised learning) karena memerlukan data pelatihan (data training) untuk membuat aturan yang mengklasifikasikan data pengujian (data testing) ke dalam kelas atau kelompok tertentu.

Teknik klasifikasi (atau pengklasifikasi) adalah pendekatan sistematis untuk membangun model klasifikasi dari sekumpulan data input. Contohnya termasuk pengklasifikasi pohon keputusan (decision tree

classifiers), pengklasifikasi berbasis aturan (rule-based classifiers), jaringan saraf (neural networks), mesin vektor pendukung (support vector machines), dan pengklasifikasi naif Bayes (naive Bayes classifiers)[8].

2.2.4 Pohon Keputusan

Pohon keputusan (decision tree) adalah teknik pemodelan prediktif yang digunakan dalam tugas klasifikasi, pengelompokan, dan prediksi. Pohon keputusan menggunakan teknik "divide and conquer" untuk membagi ruang eksplorasi masalah menjadi beberapa subset. Sebuah Pohon keputusan (decision tree) adalah pohon di mana akar dan setiap simpul internal diberi label dengan sebuah pertanyaan. Busur yang berasal dari setiap node mewakili semua kemungkinan jawaban untuk pertanyaan terkait. Setiap simpul daun mewakili solusi solusi yang dapat diprediksi untuk masalah yang sedang dipertimbangkan.

2.2.5 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang dapat digunakan untuk membangun pohon keputusan. Di antara algoritma pohon keputusan, metode ID3 dan C4.5 milik J. Ross Quinlan mungkin merupakan yang paling populer dan banyak digunakan dalam machine learning [9]. Algoritma C4.5 merupakan evolusi dari ID3 yang dikembangkan oleh Quinlan pada tahun 1993. Algoritma C4.5 secara rekursif memisahkan data untuk menghasilkan pohon keputusan untuk kumpulan data tertentu. Algoritma C4.5 memperhitungkan file atribut kategorikal dan numerik. Untuk setiap jenis atribut, C4.5 menghitung perolehan information gain dan gain ratio untuk menentukan nilai tertinggi.

Algoritma C4.5 telah membuat beberapa perubahan untuk meningkatkan algoritma ID3. Beberapa di antaranya adalah dapat menangani data pelatihan dengan nilai atribut yang hilang, dapat memproses nilai atribut yang berbeda, dapat memotong pohon keputusan yang dihasilkan dan dapat mengelola atribut dengan nilai diskrit dan kontinu [10]. Diagram alir (Flowchart) untuk algoritma C4.5 dapat dilihat pada **Gambar 2**.



Gambar 2 Diagram Alir Algoritma C4.5

Tahapan-tahapan dalam perhitungan algoritma C4.5 adalah sebagai berikut :

Langkah pertama adalah menghitung nilai info atau sering disebut juga sebagai nilai entropy pada setiap kategori atribut. Rumus perhitungan dari info atau entropy dapat dihitung dengan persamaan :

$$Info(D) = - \sum_{i=1}^m p_i \times \log_2(p_i) \dots\dots\dots (1)$$

Keterangan:

D = himpunan kasus

m = partisi data D

p_i = probabilitas dari sum(ya) dibagi dengan total kasus

Setelah menghitung nilai info (entropy) dari semua kategori atribut pada kumpulan data (dataset) maka dibutuhkan nilai info dari masing-masing atribut itu sendiri untuk menghitung nilai gain.

Rumus perhitungan nilai info pada setiap atribut dapat dihitung dengan persamaan :

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \dots\dots\dots (2)$$

Keterangan:

D = himpunan kasus

A = Atribut

v = partisi data D

$|D_j|$ = jumlah kasus pada partisi ke j

$|D|$ = jumlah kasus dalam D

Selanjutnya akan dilakukan perhitungan information gain pada setiap atribut dengan persamaan :

$$Gain(A) = Info(D) - Info_A(D) \dots\dots\dots (3)$$

Keterangan:

A = Atribut
 D = himpunan kasus

Algoritma C4.5 merupakan pengembangan dari algoritma ID3, yang menggunakan perluasan perolehan informasi yang dikenal sebagai gain ratio yang diterapkan semacam normalisasi pada information gain menggunakan nilai "split information". Rumus perhitungan nilai split info dapat dihitung dengan persamaan :

$$SplitInfo_A(D) = -\sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \dots\dots\dots (4)$$

Keterangan:
 D = ruang sampel
 A = Atribut
 v = banyak partisi data D
 Dj = jumlah sampel untuk atribut ke j

Setelah mendapatkan nilai split info, langkah berikutnya adalah menentukan nilai gain ratio yang dapat dihitung dengan persamaan :

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)} \dots\dots\dots (5)$$

Keterangan:
 A = Atribut
 D = himpunan kasus

Langkah selanjutnya adalah menentukan atribut dengan nilai gain ratio terbesar untuk dipilih sebagai atribut pemisahan yang akan dijadikan akar (*root*) pohon keputusan. Kemudian langkah berikutnya adalah menghitung kembali nilai gain ratio terbesar dari seluruh atribut dengan kondisi tidak mengikutsertakan atribut yang sudah terpilih sebagai akar (*root*) atau cabang dalam pohon keputusan untuk menemukan atribut baru yang akan dijadikan cabang baru dalam pohon keputusan, proses ini dilakukan hingga menghasilkan nilai gain sama dengan nol untuk semua atribut yang tersisa.

2.2.6 Confusion Matrix

Confusion Matrix adalah alat yang berguna untuk menganalisis seberapa baik model klasifikasi yang telah dibuat. Confusion Matrix merupakan sebuah tabel yang terdiri dari sekumpulan data pengujian yang telah diklasifikasikan benar (*positive*) dan tidak benar (*negative*) oleh model klasifikasi. Tabel confusion matrix dapat dilihat pada **Tabel 1**.

Tabel 1 Confusion Matrix

Kelas Yang Sebenarnya	Kelas Hasil Klasifikasi			
		Yes	No	Total
Yes	TP	FN	P	
No	FP	TN	N	
Total	P'	N'	P + N	

Pada Tabel 1, P merupakan jumlah data positif yang sebenarnya dan N merupakan jumlah data negatif yang sebenarnya. True positives (TP) merupakan jumlah data positif pada kelas yang sebenarnya dan positif juga pada

kelas hasil klasifikasi. True Negatives (TN) merupakan jumlah data negatif pada kelas yang sebenarnya dan negatif juga pada kelas hasil klasifikasi. False Positives (FP) merupakan jumlah data negatif pada kelas yang sebenarnya dan positif pada kelas hasil klasifikasi. False Negatives (FN) merupakan jumlah data positif pada kelas yang sebenarnya dan negatif pada kelas hasil klasifikasi. Selain itu, P' adalah jumlah data yang diberi label positif (TP + FP) dan N' adalah jumlah data yang diberi label negatif (TN + FN).

Untuk permasalahan dalam klasifikasi, pengukuran yang biasa digunakan adalah accuracy, precision dan recall. Accuracy (Keakuratan) hasil klasifikasi pada data pengujian yang diberikan adalah persentase data pengujian yang diklasifikasikan dengan benar oleh model klasifikasi yang telah dibuat. Rumus dari accuracy ini adalah sebagai berikut,

$$Accuracy = \frac{TP+TN}{P+N} \dots\dots\dots (6)$$

Keterangan :
 TP = jumlah data True Positives
 TN = jumlah data True Negatives
 P = jumlah data positif yang sebenarnya
 N = jumlah data negatif yang sebenarnya

Ukuran precision dan recall juga banyak digunakan dalam klasifikasi. Precision (*presisi*) dapat dianggap sebagai ukuran ketepatan antara informasi yang didapatkan oleh pengguna dengan informasi yang diberikan oleh sistem. Rumus untuk menghitung precision adalah sebagai berikut,

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (7)$$

Keterangan :
 TP = jumlah data True Positives
 FP = jumlah data False Positives

Sedangkan recall adalah ukuran kelengkapan data atau tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Rumus untuk recall adalah sebagai berikut,

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (8)$$

Keterangan :
 TP = jumlah data True Positives
 FN = jumlah data False Negatives

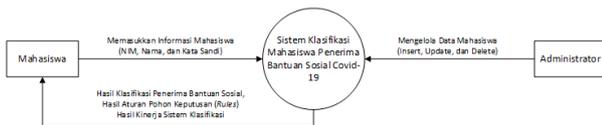
3. Hasil dan Pembahasan

3.1. Perancangan Sistem

3.1.1. Data Flow Diagram (DFD)

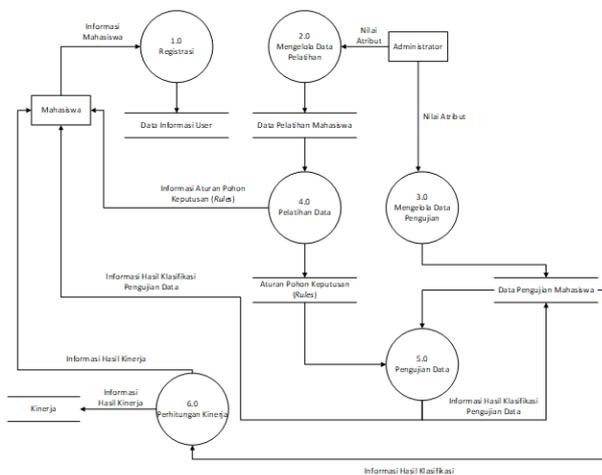
Perancangan DFD dari sistem perangkat lunak pada aplikasi ini berupa diagram konteks (Context Diagram) yang terdiri dari dua entitas yaitu mahasiswa dan direktorat kemahasiswaan serta sebuah proses berupa sistem klasifikasi mahasiswa penerima bantuan

Covid-19. Entitas mahasiswa dapat memasukkan informasi seperti NIM, nama, dan password. Entitas direktorat kemahasiswaan dapat mengelola data mahasiswa seperti menambahkan, mengedit dan menghapus data mahasiswa. Selain itu entitas mahasiswa juga dapat melihat hasil klasifikasi penerima bantuan sosial dan hasil kinerja dari proses sistem klasifikasi mahasiswa penerima bantuan sosial Covid-19. Diagram konteks pada aplikasi ini dapat dilihat pada **Gambar 3**.



Gambar 3 Diagram Konteks

Perancangan selanjutnya setelah diagram konteks adalah perancangan DFD level 0. Pada DFD level 0 dalam aplikasi ini terdapat dua entitas, enam proses, dan lima data store. Entitas mahasiswa dapat melakukan proses registrasi yang akan disimpan di dalam data store data informasi user. Kemudian Entitas direktorat kemahasiswaan dapat mengelola data pelatihan dan data pengujian yang masing-masing juga akan disimpan di dalam data store. Lalu data pelatihan mahasiswa digunakan di dalam proses pelatihan data yang akan menghasilkan aturan pohon keputusan (rules) yang akan disimpan di dalam data store. Setelah itu aturan pohon keputusan ini digunakan untuk proses pengujian data mahasiswa dan menghasilkan hasil klasifikasi yang akan disimpan dalam data store data pengujian mahasiswa dan dapat dilihat oleh entitas mahasiswa. Hasil klasifikasi juga digunakan untuk perhitungan kinerja yang juga disimpan dalam data store dan dapat dilihat oleh entitas mahasiswa. Perancangan DFD Level 0 pada aplikasi ini dapat dilihat pada **Gambar 4**.

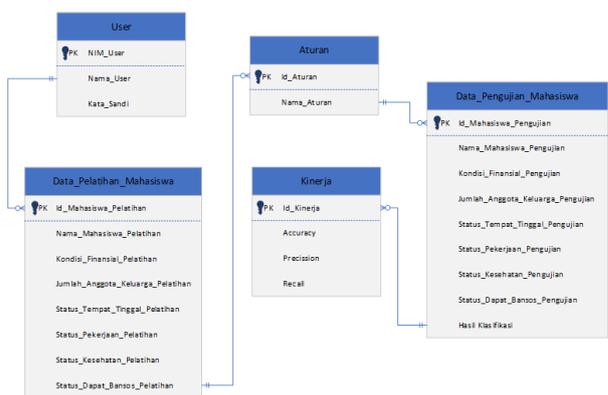


Gambar 4 DFD Level 0

3.1.2. Entity Relationship Diagram (ERD)

Perancangan ERD pada aplikasi ini terdiri dari lima entitas yang berupa user, data pelatihan mahasiswa, aturan, data pengujian mahasiswa, dan kinerja. Pada entitas user terdapat atribut NIM user sebagai primary key, nama user, status user, dan kata sandi. Pada entitas data pelatihan mahasiswa terdapat atribut id mahasiswa pelatihan sebagai primary key, nama mahasiswa pelatihan, kondisi finansial pelatihan, jumlah anggota keluarga pelatihan, status tempat tinggal pelatihan, status pekerjaan pelatihan, status kesehatan pelatihan dan status dapat bansos pelatihan. Pada entitas aturan terdapat atribut id aturan sebagai primary key dan nama aturan. Pada entitas data pengujian mahasiswa terdapat atribut id mahasiswa pengujian sebagai primary key, nama mahasiswa pengujian, kondisi finansial pengujian, jumlah anggota keluarga pengujian, status tempat tinggal pengujian, status pekerjaan pengujian, status kesehatan pengujian dan status dapat bansos pengujian dan hasil klasifikasi. Pada entitas kinerja terdapat atribut id kinerja sebagai primary key, accuracy, precision dan recall.

Relationship atau hubungan antar entitas tersebut adalah entitas user yang terbagi menjadi dua yaitu, user sebagai administrator yang dapat mengelola data pelatihan dan data pengujian, sedangkan untuk user sebagai mahasiswa dapat melihat hasil klasifikasi dan hasil kinerja, entitas data pelatihan mahasiswa digunakan untuk membuat aturan pohon keputusan. Entitas aturan digunakan untuk mengolah data pengujian untuk mendapatkan hasil klasifikasi. Entitas data pengujian mahasiswa digunakan untuk menguji kinerja program berdasarkan dengan hasil klasifikasi yang didapat. Entitas kinerja digunakan untuk menunjukkan keakuratan program hasil klasifikasi. ERD dari sistem perangkat lunak pada aplikasi ini dapat dilihat pada **Gambar 5**.



Gambar 5 Entity Relationship Diagram

3.2. Teknik Pengumpulan Data

Karena keterbatasan penulis dalam mengumpulkan data dalam jumlah besar dalam masa pandemi Covid-19 ini maka pengumpulan data dilakukan dengan menggunakan kuesioner/angket. Jumlah data responden angket yang akan digunakan sebagai data pelatihan dan pengujian dalam pembuatan aplikasi ini adalah sekitar 500 mahasiswa yang akan disebarakan melalui media

sosial dan email yang ditujukan kepada mahasiswa universitas tarumanagara yang aktif kuliah pada saat masa pandemi Covid-19 berlangsung dan kuesioner ini dikumpulkan pada semester genap 2020/2021. Kemudian dari 500 data yang telah terkumpul akan dibagi 80% untuk data pelatihan dan 20% untuk data pengujian sehingga jumlah data pelatihan sebesar 400 data mahasiswa dan jumlah data pengujian sebesar 100 data mahasiswa. Kriteria calon penerima bantuan sosial covid-19 dapat dilihat pada Tabel 2.

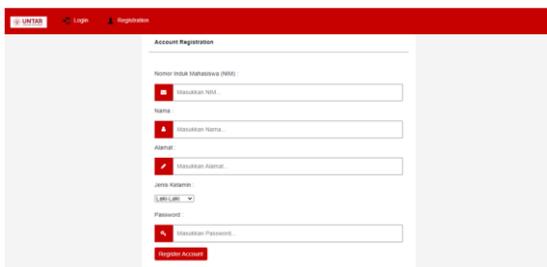
Tabel 2 Kriteria Calon Penerima Bantuan Sosial Covid-19

No.	Kriteria Calon Penerima Bantuan Sosial Covid-19	Pilihan
1	Mahasiswa mengalami kendala finansial di masa pandemi Covid-19.	Ya / Tidak
2	Jumlah anggota keluarga mahasiswa calon penerima bansos	Jumlah anggota keluarga
3	Status kehilangan pekerjaan kepala keluarga atau penanggung biaya hidup mahasiswa calon penerima bansos	Ya / Tidak
4	Status keadaan tempat tinggal mahasiswa calon penerima bansos	Milik Sendiri / Sewa
5	Status pernah terpapar Covid-19 mahasiswa calon penerima bansos dan anggota keluarga	Ya / Tidak

3.3. Tampilan Antarmuka



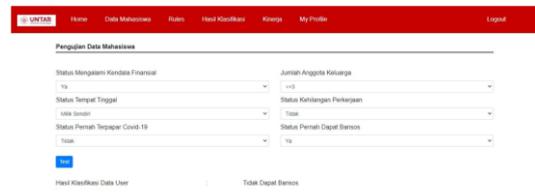
Gambar 6 Tampilan Halaman Login



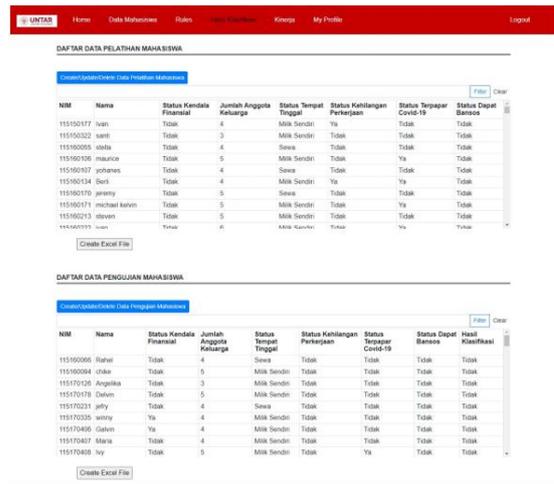
Gambar 7 Tampilan Halaman Registrasi



Gambar 8 Tampilan Halaman Home



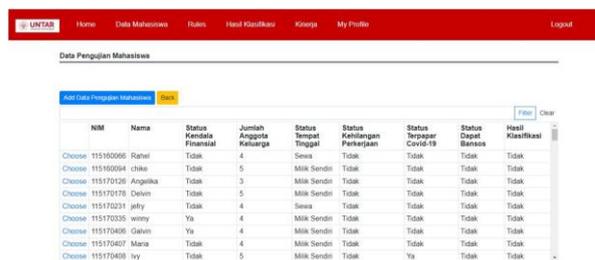
Gambar 9 Tampilan Halaman Data Mahasiswa



Gambar 10 Tampilan Halaman Data Mahasiswa (Admin)



Gambar 11 Tampilan Halaman Data Pelatihan Mahasiswa (Admin)



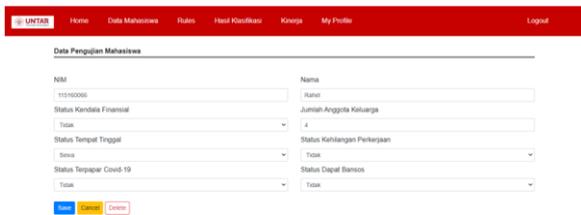
Gambar 12 Tampilan Halaman Data Pengujian Mahasiswa (Admin)



Gambar 13 Tampilan Halaman Add Data Pengujian Mahasiswa (Admin)



Gambar 18 Tampilan Halaman Kinerja



Gambar 14 Tampilan Halaman Update dan Delete Data Pengujian Mahasiswa (Admin)



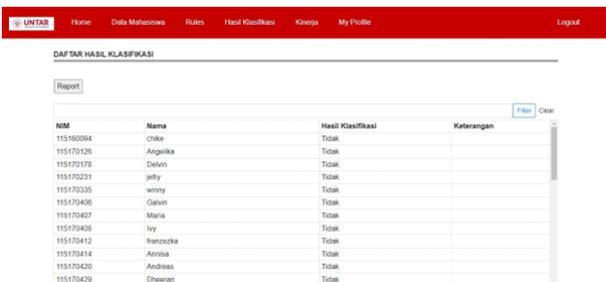
Gambar 19 Tampilan Halaman My Profile



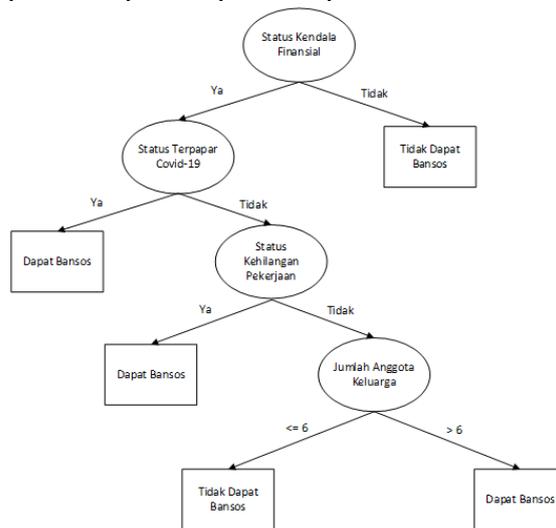
Gambar 15 Tampilan Halaman Rules

3.3 Hasil Pohon Keputusan

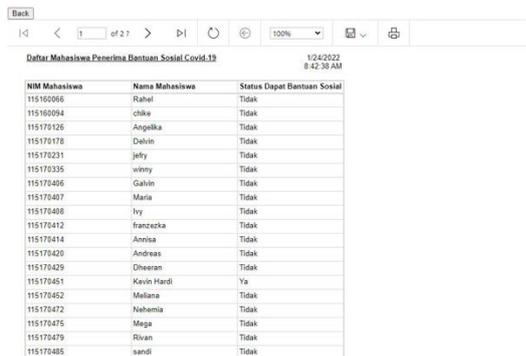
Setelah proses perhitungan algoritma C4.5 terhadap data pelatihan mahasiswa selesai dilakukan oleh program maka akan dihasilkan pohon keputusan yang kemudian akan disederhanakan menjadi aturan-aturan (*rules*) untuk mengklasifikasi data pengujian mahasiswa. Berikut merupakan hasil pohon keputusan dan aturan dari percobaan pertama pada data pelatihan mahasiswa.



Gambar 16 Tampilan Halaman Hasil Klasifikasi



Gambar 20 Hasil Pohon Keputusan



Gambar 18 Tampilan Report

Berdasarkan hasil pohon keputusan di atas dapat dibuat aturan sebagai berikut :

- IF (Status Kendala Finansial = Tidak) THEN (Hasil Klasifikasi Dapat Bansos = Tidak)
- IF (Status Kendala Finansial = Ya) AND (Status Terpapar Covid-19 = Ya) THEN (Hasil Klasifikasi Dapat Bansos = Ya)
- IF (Status Kendala Finansial = Ya) AND (Status Terpapar Covid-19 = Tidak) AND (Status

- Kehilangan Pekerjaan = Ya) THEN (Hasil Klasifikasi Dapat Bansos = Ya)
4. IF (Status Kendala Finansial = Ya) AND (Status Terpapar Covid-19 = Tidak) AND (Status Kehilangan Pekerjaan = Tidak) AND (Jumlah Anggota Keluarga > 6) THEN (Hasil Klasifikasi Dapat Bansos = Ya) ELSE (Hasil Klasifikasi Dapat Bansos = Tidak)

3.4 Hasil Pengujian Metode

Pengujian metode dilakukan dengan menggunakan 500 data mahasiswa yang telah terkumpul dengan perbandingan 80% data pelatihan dan 20% data pengujian sehingga didapatkan jumlah data sebesar 400 data pelatihan dan 100 data pengujian lalu data tersebut diuji lima kali dengan data pelatihan dan pengujian yang berbeda sehingga dapat ditentukan akurasi Algoritma C4.5 dengan melihat nilai rata-rata dari akurasi kelima percobaan tersebut.

Berdasarkan rumus perhitungan *Confusion Matrix* didapatkan hasil perhitungan sebagai berikut :

1. Hasil percobaan ke-1 menghasilkan Accuracy sebesar 98%, Precision Sebesar 96.77%, dan Recall sebesar 96.77%.
2. Hasil percobaan ke-2 menghasilkan Accuracy sebesar 94%, Precision Sebesar 84%, dan Recall sebesar 91.3%.
3. Hasil percobaan ke-3 menghasilkan Accuracy sebesar 83%, Precision Sebesar 83.33%, dan Recall sebesar 73.17%.
4. Hasil percobaan ke-4 menghasilkan Accuracy sebesar 85%, Precision Sebesar 91.49%, dan Recall sebesar 79.63%.
5. Hasil percobaan ke-5 menghasilkan Accuracy sebesar 85%, Precision Sebesar 95.24%, dan Recall sebesar 75.47%.

Dari lima hasil percobaan yang telah dilakukan didapatkan rata-rata nilai Accuracy sebesar 89%, Precision sebesar 90.16%, dan Recall sebesar 83.27%.

Tabel 3 Hasil Pengujian

Percobaan Ke-	Nilai Accuracy	Nilai Precision	Nilai Recall
1	98%	96.77%	96.77%.
2	94%	84%	91.3%.
3	83%	84%	73.17%.
4	85%	83.33%	79.63%
5	85%	95.24%	75.47%

4. Kesimpulan

1. Sistem yang dibuat dapat membantu mempermudah pengklasifikasian mahasiswa dalam menentukan calon penerima bantuan sosial Covid-19 melalui aturan yang terbentuk dari pohon keputusan dari hasil perhitungan algoritma C4.5.

2. Berdasarkan hasil pengujian metode dengan menggunakan 400 data pelatihan mahasiswa untuk menguji 100 data pengujian mahasiswa yang dilakukan sebanyak lima kali percobaan didapatkan rata-rata nilai Accuracy sebesar 89%, Precision sebesar 90.16%, dan Recall sebesar 83.27%. Aplikasi ini diharapkan dapat dikembangkan lagi seperti penambahan fitur dan kriteria baru dalam penentuan calon mahasiswa penerima bantuan sosial covid-19 sehingga dapat meningkatkan kinerja program. Fitur dan kriteria yang ditambahkan tersebut juga perlu dipertimbangkan lebih lanjut agar sesuai dengan kebutuhan pengguna.

REFERENSI

- [1] Han, Jiawei. Kamber, Micheline., Pei, Jian., 2012, "Data Mining: Concepts and Techniques, 3rd Edition", Morgan Kaufmann, Waltham.
- [2] Wang, Yanbin., Januari 2021, "Prediction of Rockburst Risk in Coal Mines Based on a Locally Weighted C4.5 Algorithm", IEEE Access, Vol.9, China.
- [3] Angkasa, Apriyudha., Fitriana, Devi. Februari 2020, "The Implementation of Classification Algorithm C4.5 in Determining the Illness Risk Level for Health Insurance Company in Indonesia", International Journal of Computer Applications, Vol. 177, No. 37, New York.
- [4] Claudia, Stefanny., Sutrisno, Tri. Agustus 2018 "Analisis Rekomendasi Peminatan Menggunakan Metode Decision Tree Dengan Algoritma C4.5", Jurnal Ilmu Komputer Dan Sistem Informasi (JIKSI). Vol. 6, No. 2, Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.
- [5] Hariati., Wati, Masna., Cahyono, Bambang., Desember 2018, "Penerapan Algoritma C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara", Jurnal Rekayasa Teknologi Informasi (JURTI), Vol. 2, No. 2, Kalimantan Timur.
- [6] Pressman, Roger S., 2010, "Software Engineering: A Practitioner's Approach, 7th Edition", McGraw-Hill, New York.
- [7] Dunham, Margaret H., 2003, "Data Mining Introductory and Advanced Topics", Prentice Hall, Upper Saddle River.
- [8] Tan, Pang-Ning., Stanbach, Michael., Kumar, Vipin., 2006 "Introduction to Data Mining", Pearson Addison Wesley, London.
- [9] Salzberg, Steven L., September 1994, "Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993". Machine Learning. Vol. 16, No. 3., Kluwer Academic Publishers, Boston.
- [10] Fakir, Y., Azalmd, M., Elaychi, R., April 2020 "Study of The ID3 and C4.5 Learning Algorithms", Journal of Medical Informatics and Decision Making. Vol. 1, No. 2.

Eugenius Edsel Barito, Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara Tahun 2022.

Jap Tji Beng, Dosen Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.

Desi Arisandi, Dosen Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.