

Klasifikasi Menggunakan Algoritma C4.5 dalam Memprediksi Status Keorganisasian Mahasiswa Universitas Amikom Yogyakarta

Classification using the C4.5 Algorithm in Predicting Students Organizational Status Amikom University Yogyakarta

Neni Firda Wardani Tan, Kusnawi*

Informatika, Ilmu Komputer, Universitas AMIKOM Yogyakarta

Jl. Padjajaran, Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta Indonesia 55283

*e-mail: khusnawi@amikom.ac.id

(received: 29 Mei 2022, revised: 3 Juni 2022, accepted: 21 Juli 2022)

Abstrak

Keikutsertaan dalam keorganisasian selama menjadi mahasiswa merupakan salah satu sarana yang penting bagi mahasiswa dalam pengembangan diri. Penelitian ini bertujuan untuk menganalisa kecenderungan seorang mahasiswa dalam berorganisasi, apakah Mahasiswa termasuk kedalam kategori mahasiswa aktif organisasi atau mahasiswa tidak aktif organisasi dengan menggunakan klasifikasi algoritma Decision Tree C4.5. Atribut yang terdapat di dalamnya yaitu, Keorganisasian Luar Kampus, Keorganisasian Pendidikan Sebelumnya, *Public Speaking*, *Problem Solving*, *Confidence Level*, dan Kepribadian. Dari kuesioner yang disebar pada bulan Oktober 2021 terkumpul sebanyak 203 data mentah dan sebanyak 167 data yang dapat digunakan untuk penelitian ini setelah dilakukannya pra-pemrosesan data. Dari hasil pengujian didapatkan *rule* atau pohon keputusan yang dapat dijadikan aturan dalam menentukan mahasiswa aktif organisasi. Nilai *accuracy* pada pengujian model ini dengan menggunakan *cross validation* menghasilkan nilai sebesar 59.27% atau dengan perbandingan *data training* : *data testing* sebesar 4 : 1 atau 80% : 20%.

Kata kunci: Data Mining, Klasifikasi, Decision Tree C4.5, Organisasi.

Abstract

Participation in an organization as a college student is an important way to self-improvement. This research analyzes students' learning in an organization and whether the student is active or not active in an organization using the Decision Tree C4.5 algorithm. Attributes analyzed were out-of-campus organization, previous organization experience, public speaking, problem-solving, confidence level, and personality. From a question from propagating on October 2021, the researcher collect 203 raw data, with 167 processes ones that were used in this research. The test conducted results in a tree or the decision tree that could be used to decide how active a student is in an organization. The accuracy value of this test using cross-validation resulted in a score of 59.27% or in comparison to the data training: data testing of 4:1 or 80%: 20%.

Keywords: Data Mining, Classification, Decision Tree C4.5, Organization.

1 Pendahuluan

Keikutsertaan dalam keorganisasian selama menjadi mahasiswa merupakan salah satu sarana yang penting bagi mahasiswa dalam pengembangan diri. Keaktifan mahasiswa dalam berorganisasi

mempunyai pengaruh yang besar terhadap prestasi belajar. Mahasiswa aktif berorganisasi dapat membagi waktunya antara pembelajaran dalam perkuliahan dan keorganisasian. Kebiasaan itulah yang akan menjadi pengalaman sangat berharga bagi mahasiswa. Namun, pada nyatanya terdapat juga banyak mahasiswa yang sama sekali tidak mengikuti keorganisasian di kampus. Mahasiswa inilah yang kegiatannya hanya sebatas kuliah dan pulang [1].

Banyak faktor yang mempengaruhi tingkat kecenderungan berorganisasi seorang mahasiswa, contohnya *confidence level* dan *public speaking*, beberapa mahasiswa merasa tidak percaya diri jika harus mengutarakan pendapat dan opininya kepada orang lain dan lebih memilih untuk menutup diri dan menghindari keikutsertaan dalam berorganisasi. Selain itu keorganisasian luar kampus, mahasiswa aktif organisasi akan menyadari betapa pentingnya berorganisasi di dalam maupun di luar kampus. Di luar kampus mahasiswa aktif akan mengikuti keorganisasian misal pelatihan, camp, atau bahkan keorganisasian diskusi [2]. Keorganisasian pendidikan sebelumnya, keorganisasian pendidikan sebelumnya ini merupakan keikutsertaan mahasiswa selama menjadi siswa sekolah menengah atas. Mahasiswa yang sebelumnya aktif organisasi pada masa SMA cenderung akan mengikuti kembali keorganisasian selama masa perkuliahan, mereka akan mencari keorganisasian yang sama di bidangnya. Selain faktor di atas yaitu tingkat pengetahuan tentang problem solving dan manajemen konflik atau pengetahuan tentang pemecahan suatu masalah, untuk faktor yang kedua ini mahasiswa yang gemar berorganisasi pasti akan lebih memahami bagaimana harus bersikap jika mendapati situasi dalam masalah. Faktor kepribadian seorang mahasiswa, setiap makhluk hidup pasti mempunyai kepribadian yang berbeda-beda, dan sangat mendukung dalam keaktifan seorang mahasiswa [3].

Atribut yang terdapat dalam penelitian ini yaitu, Keorganisasian Luar Kampus, Keorganisasian Pendidikan Sebelumnya, *Public Speaking*, *Problem Solving*, *Confidence Level*, dan Kepribadian yang mencakup Koleris, Sanguinis, Melankolis, dan Flegmatis. Tujuan dari penelitian ini adalah, mengetahui kinerja model dan mengetahui fakta kecenderungan mahasiswa dalam berorganisasi di Universitas AMIKOM Yogyakarta.

2 Tinjauan Literatur

Dalam penelitian dengan sumber dataset yang digunakan berasal dari data.go.id dari tahun 2009-2010. Beberapa atribut yang digunakan antara lain penduduk_miskin, pembiayaan_provinsi, pendapatan_provinsi, belanja_provinsi, tpak, rasio_gini, tingkat_putus_sekolah. Hasilnya adalah adanya hubungan antara atribut yang diambil dengan tingkat anak putus sekolah, yang mana itu merupakan hal yang penting. Berdasarkan hasil dari penelitian yang telah dilakukan, bahwa klasifikasi provinsi berdasarkan pada tingkat putus sekolah dasar dapat diselesaikan dengan menggunakan algoritma C4.5 dengan tingkat akurasi yang tinggi yaitu 71.2%. Tingkat akurasi tersebut dapat saja meningkat dengan adanya penambahan data, atribut, atau dengan menggunakan optimisasi pada algoritma [4].

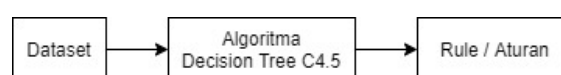
Untuk Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun, dilakukan pemeriksaan fisik dan penyebaran kuesioner terhadap 360 siswa sekolah dasar kelas 1 sampai dengan kelas 6 dengan rentan usia antara 7-12 tahun dan diperoleh hasil bahwa 79,7% siswa memiliki berat badan normal, 12,5% siswa mengalami kekurangan berat badan, dan 7,8% siswa mengalami kelebihan berat badan. Dari kondisi tersebut, pengujian lebih lanjut menggunakan *Decision Tree* menunjukkan bahwa faktor usia, berat badan, tinggi badan, BMR, dan BMI memiliki kontribusi pada penentuan kebutuhan energi pada anak, dan jenis kelamin mempengaruhi pada proses penentuan kebutuhan konsumsi karbohidrat, protein, lemak, dan serat [5].

Mengklasifikasikan tingkat pemahaman mahasiswa di STIKOM Tunas Bangsa. Data yang digunakan dalam penelitian ini diperoleh dari mahasiswa semester genap dengan sampel data yang terkumpul sebanyak 165 mahasiswa. Atribut yang digunakan dalam penelitian ini ada 5 yaitu *Communication* (C1), *Learning Atmosphere* (C2), *Instructional Media* (C3), *Appearance* (C4), dan *How to Teach* (C5). Metode yang digunakan dalam penelitian ini adalah Algoritma C4.5 dan diproses menggunakan software RapidMiner dalam membuat pohon keputusan. Dari hasil penelitian diperoleh 8 aturan untuk menentukan level pemahaman mahasiswa, empat aturan untuk mahasiswa yang dapat memahami materi dalam perkuliahan dan 4 aturan untuk mahasiswa yang tidak dapat memahami materi dalam perkuliahan. Algoritma C4.5 yang diterapkan untuk mengetahui tingkat pemahaman mahasiswa memiliki tingkat akurasi sebesar 87,50% [6].

Analisa Kinerja Algoritma C4.5 dalam memprediksi hasil belajar yang dilakukan bertujuan untuk memprediksi hasil belajar mahasiswa sistem informasi STMIK Royal Kisaran berdasarkan riwayat akademik mahasiswa pada saat sekolah menengah atas. Penelitian ini akan menginformasikan kepada mahasiswa untukantisipasi secara dini dalam mengikuti perkuliahan agar mendapatkan hasil belajar yang maksimal. Ada 4 atribut yaitu Sekolah, Jurusan, Peringkat, dan IPK yang menjadi atribut hasil. Atribut Sekolah terdiri dari SMA, Aliyah, dan SMK. Atribut Jurusan terdiri dari IPA, IPS, dan Teknik. Atribut Peringkat terdiri dari >10, 10, dan 5. Sedangkan atribut hasil atau IPK terdiri dari Sangat Baik, Baik, dan Kurang. Di dalam penelitian ini juga dijelaskan bahwa hasil pengujian dengan *tools rapidminer* menghasilkan hasil yang sama dengan hasil pengujian secara manual[7].

Decision Tree Algoritma C4.5

Algoritma C4.5 dipopulerkan pada tahun 1993 oleh Ross Quinlan, yang juga mengenalkan algoritma ID3, algoritma yang merupakan versi algoritma sebelumnya[8]. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan atau *Decision Tree* yang diperoleh berdasarkan hasil dataset seperti yang tertera pada Gambar 1 berikut ini.



Gambar 1. Konsep Decision Tree

Untuk membentuk pohon keputusan adalah dengan memilih atribut sebagai *root node* didasarkan pada nilai information gain tertinggi dari atribut-atribut yang ada. Sebelum perhitungan information gain, akan dilakukan perhitungan entropy. Entropy merupakan distribusi probabilitas untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data. Semakin tinggi tingkat entropy dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Perhitungan information gain menggunakan rumus 2, sedangkan entropy menggunakan rumus 1.

Rumus menghitung Entropy :

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

Keterangan :

S : Himpunan kasus

n : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

Rumus menghitung Gain :

$$Gaint(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_1|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S₁| : Jumlah kasus pada partisi ke-i

|S₂| : Jumlah kasus dalam S

Data Mining

Data Mining merupakan kegiatan menganalisis dataset dalam jumlah besar untuk menemukan relasi yang tidak diketahui sebelumnya dan merangkum dalam bentuk data baru yang dapat dipahami dan bermanfaat bagi pemilik data. Berdasarkan tipenya pekerjaannya, data mining dibagi menjadi 5, yaitu *Exploratory Data Analysis (EDA)*, *Descriptive Modelling*, *Predictive Modelling*, Penemuan Pola dan aturan, serta Pemanggilan Konten[9].

Dalam data mining terdapat dua pendekatan metode pelatihan, yaitu [10]:

1) *Unsupervised Learning*

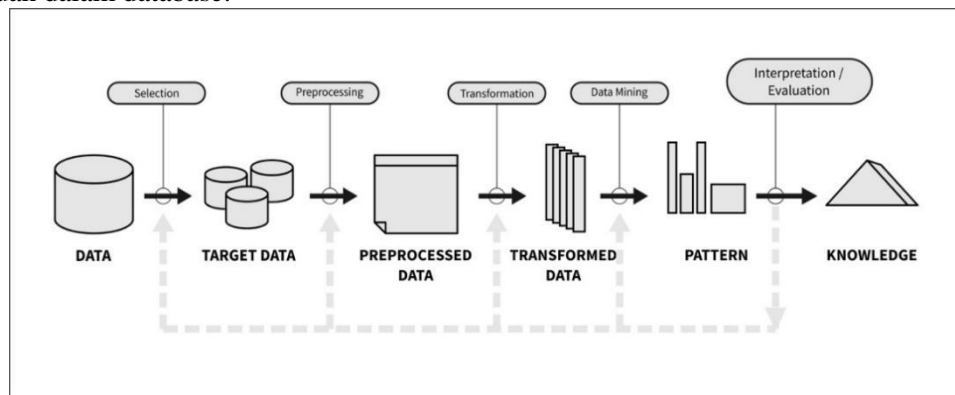
Metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.

2) *Supervised Learning*

Metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses training.

Tahapan Data Mining

Berikut seperti tertera pada Gambar 2 adalah tahapan dalam proses data mining atau sering juga disebut proses *Knowledge Discovery in Database (KDD)* merupakan proses penemuan pengetahuan dalam database.



Gambar 2. Tahapan Data Mining

Tahap-tahap Data Mining yaitu [11]:

- 1) Pembersihan Data (*Data Cleaning*)
Proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
- 2) Transformasi Data (*Data Transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
- 3) Proses Mining
Proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- 4) Evaluasi Pola (*Pattern Evaluation*)
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.
- 5) Presentasi Pengetahuan (*Knowledge Presentation*)
Memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

Rapid Miner

RapidMiner atau yang sebelumnya dikenal dengan nama Yale (*Yet Another Learning Environment*) awalnya ditemukan dan dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer pada tahun 2001 di Artificial Intelligence Unit of the Technical University of Dortmund [12].

RapidMiner merupakan platform perangkat lunak yang dikembangkan oleh perusahaan dengan nama yang sama yaitu RapidMiner yang menyediakan lingkungan terintegrasi untuk *data preparation, machine learning, deep learning, text mining, dan predictive analytics*. Platform perangkat lunak ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian pendidikan, pelatihan, pembuatan prototype cepat, dan pengembangan aplikasi dan mendukung semua langkah dari proses *machine learning* termasuk persiapan data, visualisasi hasil, validasi model, dan pengoptimalan [13].

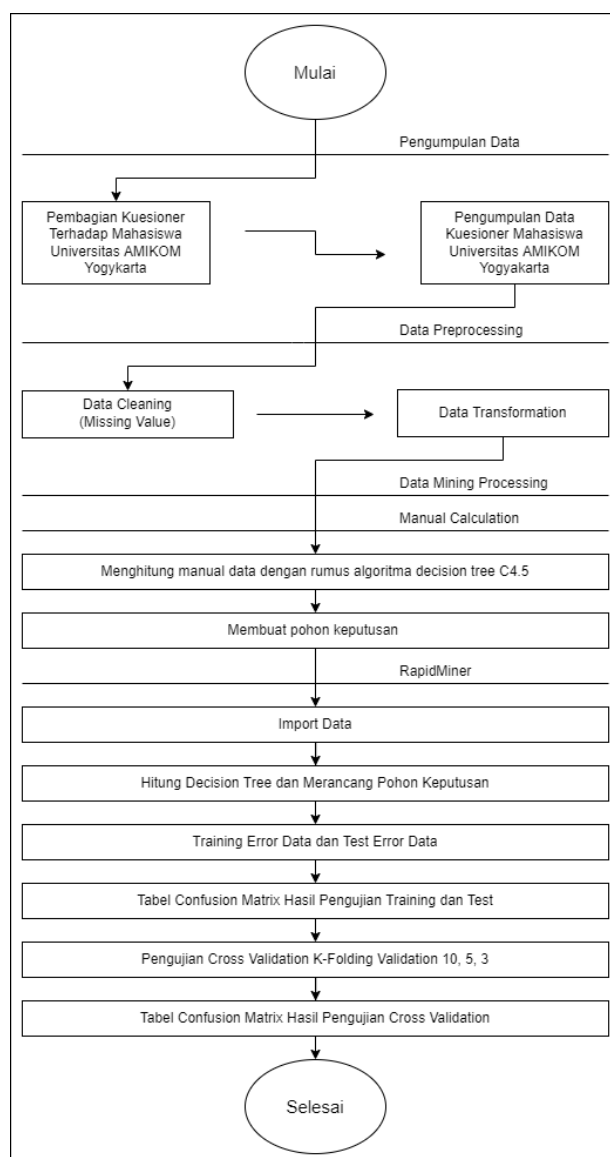
Keorganisasian

Keorganisasian atau istilah lainnya UKM (Unit Kegiatan Mahasiswa) dalam prosesnya adanya faktor-faktor yang mempengaruhi minat Mahasiswa dalam berorganisasi, setidaknya menghasilkan atau terdapat terdapat 5 faktor yang mempengaruhi minat mahasiswa berorganisasi, yaitu Kegiatan Menarik, Motivasi Dari Senior, Teman Baru, Kemampuan Berkomunikasi, dan Pengembangan di Dunia Kerja [14].

Hal lainnya adalah masalah Problem Solving pada Mahasiswa yang aktif berorganisasi, menunjukkan bentuk bahwa problem solving pada mahasiswa yang aktif berorganisasi antara lain memandang masalah sebagai hal yang positif, menganggap masalah sebagai cara untuk mengevaluasi diri, dan sebagai pengalaman untuk kehidupan. Hambatan utama yang dihadapi adalah anggota yang kurang dapat bekerjasama. Cara yang dilakukan untuk menyelesaikan masalah antara lain dengan mengadakan rapat, berbicara langsung dengan yang bermasalah, dan mendekati diri kepada anggota. Selanjutnya faktor yang mempengaruhi kemampuan problem solving adalah jenis kelamin, jabatan dalam organisasi, emosi, situasional, tingkah laku meniru serta bidang dalam organisasi [15].

3 Metode Penelitian

Pada metode penelitian ini dilakukan dengan berbagai langkah penelitian yang dapat menghasilkan model pohon keputusan serta pengujian untuk menghasilkan tingkat akurasi dalam memprediksi kecenderungan seorang mahasiswa dalam berorganisasi, apakah mahasiswa termasuk kedalam kategori mahasiswa aktif organisasi atau mahasiswa tidak aktif organisasi. Adapun langkah penelitian ini adalah sebagaimana pada Gambar 3 berikut.



Gambar 3. Langkah Penelitian

Berikut adalah penjelasan langkah-langkah penelitian yang tertera pada gambar 3 :

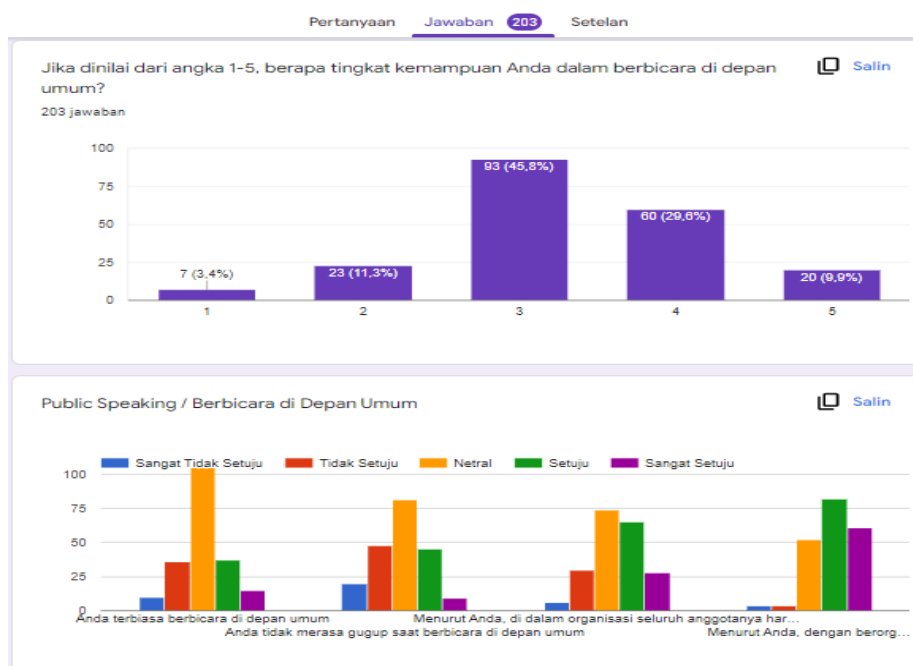
1. Penulis membuat kuesioner dan selanjutnya kuesioner tersebut akan digunakan sebagai bahan atau data awal untuk menentukan jumlah aktif dan tidak aktif pada persentase keaktifan Mahasiswa Universitas AMIKOM Yogyakarta
2. Mengumpulkan hasil dari kuesioner yang telah disebarakan
3. Preprocessing Data, terdiri dari Cleaning dan Transformation, yaitu membersihkan dan mengubah data kuesioner guna menyesuaikan dengan dataset yang dibutuhkan
4. Data Mining Processing, terdiri dari Manual Calculation dan proses menggunakan RapidMiner
5. Manual Calculation, menghitung nilai entropy dan nilai gain, selanjutnya membuat pohon keputusan
6. Proses Data Mining menggunakan RapidMiner
7. Import data, terapkan operator decision tree dan dihasilkan pohon keputusan
8. Training Error dan Test Error, untuk mengetahui accuracy model dari tabel confusion matrix
9. Cross Validation dengan k-fold untuk pengujian model secara berulang guna mengetahui accuracy pada tabel confusion matrix
10. Selesai

4 Hasil dan Pembahasan

Dalam memprediksi status keorganisasian mahasiswa dengan menggunakan data kuesioner sampai mendapatkan dataset final terdiri dari enam atribut meliputi Keorganisasian Luar Kampus Keorganisasian Pendidikan Sebelumnya, kemampuan *public speaking*, *problem solving*, *confidence level* dan Kepribadian.

4.1 Kuesioner

Data yang akan digunakan dalam penelitian ini adalah bersumber dari data kuesioner yang disebarakan kepada Mahasiswa Universitas AMIKOM Yogyakarta pada bulan Oktober 2021. Kuesioner ini dibuat menggunakan Google Formulir atau perangkat lunak administrasi survei. Data yang berhasil dikumpulkan dalam pembagian kuesioner ini adalah sebanyak 203 data. Setelah dilakukan preprocessing data akhirnya diputuskan untuk digunakannya data hasil pembagian kuesioner sebanyak 167 dataset seperti tertera pada Gambar 4.



Gambar 4. Hasil Kuesioner

4.2 Data Preprocessing

Data preprocessing dilakukan untuk mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Berikut proses yang dilakukan yaitu data cleaning, mengubah data sesuai kebutuhan klasifikasi.

Data Cleaning

Pembersihan data atau data cleaning yang bertujuan untuk membuang dataset jika terdapat data yang tidak relevan atau data yang tidak konsisten. Hasilnya bisa diperoleh sesuai pada Gambar 5 dan Gambar 6 berikut.

1) Sebelum Cleaning

Nama	NPM	Jenis Kelamin	Apakah Berapa nilai	Apakah	Apakah	Jil	Public	Public	Public	Public	Jil	Proble	Proble	Proble	Proble	Proble	Jil	Confide	
Neni Firda Wardani Tan	17.11.1574	Perempuan	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Tidak	Netral	Sangat	Sangat	4	Netral	Netral	Netral	Netral	Sangat	3	Setuju
FIQIH BRIAN KARLINGC	17.11.1569	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	3	Netral	Netral	Netral	Netral	3	Netral	Netral	Netral	Netral	Netral	3	Netral
Fahmi Abdurrabi	17.11.1549	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	4	Setuju	Setuju	Setuju	Sangat	4	Setuju	Setuju	Setuju	Sangat	Sangat	4	Setuju
Alfian Difa'ul Amien	17.11.1584	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Tidak	3	Netral	Netral	Setuju	Sangat	3	Netral	Setuju	Setuju	Setuju	Setuju	4	Setuju
Anonim	404	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Tidak	2	Netral	Tidak	Setuju	Setuju	3	Netral	Netral	Netral	Netral	Netral	3	Netral
Linggar Brilly Ardha	21.11.4017	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	2	Netral	Tidak	Tidak	Netral	3	Netral	Setuju	Setuju	Netral	Setuju	3	Setuju
Putra Ramadhan	17.95.0033	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	4	Netral	Netral	Tidak	Setuju	3	Netral	Setuju	Setuju	Netral	Setuju	4	Tidak
Ilham temen beruang	17.11.1590	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	5	Sangat	Setuju	Tidak	Setuju	5	Netral	Tidak	Sangat	Setuju	Setuju	4	Netral
Fathan azka pradana	19.11.3089	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Netral	Setuju	3	Netral	Netral	Setuju	Setuju	Setuju	3	Setuju
Fauzan Sukri	17.11.1551	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	3	Netral	Tidak	Netral	Setuju	4	Netral	Setuju	Setuju	Setuju	Setuju	4	Netral
Hasbi Hawarizmi	21.11.4411	Laki-Laki	Tidak	IPK $< 2,75$	Tidak	Ya	4	Setuju	Setuju	Netral	Sangat	3	Tidak	Setuju	Netral	Tidak	Netral	3	Tidak
rakha	17.11.1571	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	2	Netral	Tidak	Setuju	Netral	3	Netral	Netral	Netral	Netral	Netral	2	Netral
Inge Sekar Widatik	17.11.1559	Perempuan	Tidak	IPK $< 2,75$	Ya	Ya	4	Setuju	Netral	Setuju	Setuju	4	Setuju	Netral	Setuju	Netral	Setuju	4	Setuju
Setiawan Eko Nugroho	17.11.1555	Laki-Laki	Ya	IPK $\geq 2,75$	Ya	Ya	5	Netral	Netral	Netral	Netral	4	Setuju	Setuju	Setuju	Setuju	Setuju	5	Netral
Rita Ayuning Tyas	20.11.3306	Perempuan	Ya	IPK $< 2,75$	Ya	Ya	4	Netral	Netral	Setuju	Sangat	4	Netral	Netral	Setuju	Netral	Sangat	4	Setuju
Maulana Rizky Rahmad	21.11.4465	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Sangat	3	Setuju	Setuju	Setuju	Setuju	Setuju	Setuju	3	Setuju
sukma	17.11.1536	Perempuan	Tidak	IPK $\geq 2,75$	Tidak	Ya	4	Setuju	Netral	Netral	Setuju	5	Sangat	Setuju	Setuju	Setuju	Netral	4	Sangat
Falahul Fadli	17.11.1563	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Setuju	Setuju	4	Netral	Netral	Setuju	Netral	Setuju	4	Netral
Putra Arya Darmawan	21.11.4404	Laki-Laki	Ya	IPK $\geq 2,75$	Ya	Ya	3	Netral	Netral	Setuju	Setuju	4	Netral	Netral	Setuju	Netral	Netral	5	Netral
Nurcholish Dwilestanto	17.11.1567	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Tidak	Sangat	Setuju	4	Setuju	Sangat	Setuju	Setuju	Sangat	3	Setuju

Gambar 5. Data sebelum Cleaning

2) Setelah Cleaning

Nama	NPM	Jenis Kelamin	Apakah Berapa nilai	Apakah	Apakah	Jil	Public	Public	Public	Public	Jil	Proble	Proble	Proble	Proble	Proble	Jil	Confide	
Neni Firda Wardani Tan	17.11.1574	Perempuan	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Tidak	Netral	Sangat	Sangat	4	Netral	Netral	Netral	Netral	Sangat	3	Setuju
FIQIH BRIAN KARLINGC	17.11.1569	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	3	Netral	Netral	Netral	Netral	3	Netral	Netral	Netral	Netral	Netral	3	Netral
Fahmi Abdurrabi	17.11.1549	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	4	Setuju	Setuju	Setuju	Sangat	4	Setuju	Setuju	Setuju	Sangat	Sangat	4	Setuju
Alfian Difa'ul Amien	17.11.1584	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Tidak	3	Netral	Netral	Setuju	Sangat	3	Netral	Setuju	Setuju	Setuju	Setuju	4	Setuju
Linggar Brilly Ardha	21.11.4017	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	2	Netral	Tidak	Tidak	Netral	3	Netral	Setuju	Setuju	Netral	Setuju	3	Setuju
Putra Ramadhan	17.95.0033	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	4	Netral	Netral	Tidak	Setuju	3	Netral	Setuju	Setuju	Netral	Setuju	4	Tidak
Ilham temen beruang	17.11.1590	Laki-Laki	Ya	IPK $\geq 2,75$	Tidak	Ya	5	Sangat	Setuju	Tidak	Setuju	5	Netral	Tidak	Sangat	Setuju	Setuju	4	Netral
Fathan azka pradana	19.11.3089	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Netral	Setuju	3	Netral	Netral	Setuju	Setuju	Setuju	3	Setuju
Fauzan Sukri	17.11.1551	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	3	Netral	Tidak	Netral	Setuju	4	Netral	Setuju	Setuju	Setuju	Setuju	4	Netral
Hasbi Hawarizmi	21.11.4411	Laki-Laki	Tidak	IPK $< 2,75$	Tidak	Ya	4	Setuju	Setuju	Netral	Sangat	3	Tidak	Setuju	Netral	Tidak	Netral	3	Tidak
rakha	17.11.1571	Laki-Laki	Tidak	IPK $\geq 2,75$	Ya	Ya	2	Netral	Tidak	Setuju	Netral	3	Netral	Netral	Netral	Netral	Netral	2	Netral
Inge Sekar Widatik	17.11.1559	Perempuan	Tidak	IPK $< 2,75$	Ya	Ya	4	Setuju	Netral	Setuju	Setuju	4	Setuju	Netral	Setuju	Netral	Setuju	4	Setuju
Setiawan Eko Nugroho	17.11.1555	Laki-Laki	Ya	IPK $\geq 2,75$	Ya	Ya	5	Netral	Netral	Netral	Netral	4	Setuju	Setuju	Setuju	Setuju	Setuju	5	Netral
Rita Ayuning Tyas	20.11.3306	Perempuan	Ya	IPK $< 2,75$	Ya	Ya	4	Netral	Netral	Setuju	Sangat	4	Netral	Netral	Setuju	Netral	Sangat	4	Setuju
Maulana Rizky Rahmad	21.11.4465	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Sangat	3	Setuju	Setuju	Setuju	Setuju	Setuju	Setuju	3	Setuju
sukma	17.11.1536	Perempuan	Tidak	IPK $\geq 2,75$	Tidak	Ya	4	Setuju	Netral	Netral	Setuju	5	Sangat	Setuju	Setuju	Setuju	Netral	4	Sangat
Falahul Fadli	17.11.1563	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Netral	Setuju	Setuju	4	Netral	Netral	Setuju	Netral	Setuju	4	Netral
Putra Arya Darmawan	21.11.4404	Laki-Laki	Ya	IPK $\geq 2,75$	Ya	Ya	3	Netral	Netral	Setuju	Setuju	4	Netral	Netral	Setuju	Netral	Netral	5	Netral
Nurcholish Dwilestanto	17.11.1567	Laki-Laki	Tidak	IPK $\geq 2,75$	Tidak	Ya	3	Netral	Tidak	Sangat	Setuju	4	Setuju	Sangat	Setuju	Setuju	Sangat	3	Setuju
Na amma Fidausiya	17.95.0024	Perempuan	Ya	IPK $\geq 2,75$	Tidak	Ya	4	Sangat	Tidak	Setuju	Sangat	4	Setuju	Setuju	Setuju	Tidak	Sangat	3	Tidak

Gambar 6. Data setelah Cleaning

Data Transformation

Transformasi data disini bertujuan untuk mengubah data ke dalam format yang sesuai yang dibutuhkan dalam pemrosesan data mining. Mengubah skala pengukuran data asli menjadi bentuk yang lain sesuai hasil pada Gambar 7 data sebagai data asli dan Gambar 8 adalah hasil tranformasi datanya. Selanjutnya pada tahap ini dilakukan juga penghitungan nilai rata-rata pada tiap data atribut yang mengandung nilai lebih dari satu, diantaranya: *Public Speaking* (X3), *Problem Solving* (X4), *Confidence Level* (X5), Kepribadian (X6){Koleris, Sanguinis, Melankolis, Phlegmatis} seperti tertera pada Gambar 9.

1) Sebelum Transformation

	Jika dinilai dari angka 1-5, Public Speaking / Berbicara	Public Speaking / Berbicara	Public Speaking / Berbicara	Public Speaking / Berbicara	Jika dinilai dari angka 1-5, Problem Solving / Penyelesaian	Problem Solving / Penyelesaian
3	Netral	Netral	Netral	Netral	3	Netral
4	Setuju	Setuju	Setuju	Sangat Setuju	4	Setuju
3	Netral	Netral	Setuju	Sangat Setuju	3	Netral
2	Netral	Tidak Setuju	Tidak Setuju	Netral	3	Netral
4	Netral	Netral	Tidak Setuju	Setuju	3	Netral
3	Netral	Netral	Netral	Setuju	3	Netral
3	Netral	Tidak Setuju	Netral	Setuju	4	Netral
4	Setuju	Setuju	Netral	Sangat Setuju	3	Tidak Setuju
4	Setuju	Netral	Setuju	Setuju	4	Setuju
5	Netral	Netral	Netral	Netral	4	Setuju
4	Netral	Netral	Setuju	Sangat Setuju	4	Netral
3	Netral	Netral	Netral	Sangat Setuju	3	Setuju
4	Setuju	Netral	Netral	Setuju	5	Sangat Setuju
3	Netral	Netral	Setuju	Setuju	4	Netral
3	Netral	Netral	Setuju	Setuju	4	Netral
3	Netral	Tidak Setuju	Sangat Setuju	Setuju	4	Setuju
4	Sangat Setuju	Tidak Setuju	Setuju	Sangat Setuju	4	Setuju
4	Netral	Setuju	Tidak Setuju	Sangat Setuju	4	Netral
3	Netral	Netral	Netral	Setuju	3	Netral
3	Netral	Tidak Setuju	Setuju	Netral	3	Netral
3	Netral	Netral	Setuju	Sangat Setuju	4	Setuju
5	Setuju	Setuju	Sangat Tidak Setuju	Sangat Tidak Setuju	5	Sangat Tidak Setuju
4	Setuju	Setuju	Tidak Setuju	Sangat Setuju	5	Setuju
3	Netral	Setuju	Setuju	Sangat Setuju	3	Netral

Gambar 7. Data sebelum Transformation

2) Setelah Transformation

	Jika dinilai dari angka 1-5, Public Speaking / Berbicara	Public Speaking / Berbicara	Public Speaking / Berbicara	Public Speaking / Berbicara	Jika dinilai dari angka 1-5, Problem Solving / Penyelesaian	Problem Solving / Penyelesaian
3	3	3	3	3	3	3
4	4	4	4	4	4	4
3	3	3	3	3	3	3
2	3	2	2	3	3	3
4	3	3	2	4	3	3
3	3	3	3	3	3	3
3	3	2	3	4	4	3
4	4	4	3	5	3	2
4	4	3	4	4	4	4
5	3	3	3	3	4	4
4	3	3	4	5	4	3
3	3	3	3	3	5	4
4	4	3	3	4	5	5
3	3	3	4	4	4	3
3	3	3	4	4	4	3
3	3	3	5	4	4	5
4	5	2	4	5	4	4
4	3	4	2	5	4	3
3	3	3	3	4	3	3
3	3	2	4	3	3	3
3	3	3	4	5	4	3
5	4	4	1	1	5	1
4	4	4	2	5	5	4
3	3	4	4	5	3	3

Gambar 8. Data Setelah Transformation

3) Perhitungan Rata-Rata

PUBLIC SPEAKING (X3)	PROBLEM SOLVING (X4)	CONFIDENCE LEVEL (X5)	KOLERIS (X6)	SANGUINIS (X7)	MELANKOLIS (X8)	PHLEGMATIS (X9)
3,00	3,00	3,00	3,00	3,00	3,00	3,00
4,20	4,33	3,80	4,67	4,00	2,86	3,67
3,60	3,67	4,00	3,17	3,25	3,43	2,83
2,40	3,50	3,20	3,17	3,75	3,71	3,50
3,20	3,50	3,80	3,00	3,25	2,71	4,17
3,20	3,50	3,80	3,67	3,25	3,43	3,17
3,00	3,67	3,00	2,00	5,00	4,00	3,83
4,00	2,83	3,40	3,00	3,25	2,71	4,17
3,80	3,67	3,80	3,83	4,50	3,43	3,67
3,40	4,00	3,60	4,50	4,75	3,29	3,00
3,80	3,67	4,00	3,50	3,75	3,57	4,00
3,40	3,83	3,60	3,33	2,75	3,14	3,67
3,60	4,17	4,40	3,67	4,00	3,00	3,67
3,40	3,50	3,60	3,50	3,75	3,00	3,33
3,40	3,33	3,40	3,00	4,50	2,71	3,67
3,40	4,33	4,40	3,83	3,75	4,00	4,33
4,00	3,83	2,80	3,33	3,00	3,57	3,00
3,60	3,50	3,60	3,83	2,50	3,29	3,17
3,20	3,17	3,80	3,50	3,00	3,57	3,33
3,00	2,83	3,00	2,83	3,75	3,57	3,33
3,60	3,83	4,00	3,33	4,25	3,57	4,83
3,00	2,67	3,60	4,00	3,25	2,29	3,00
3,80	4,17	4,60	5,00	3,75	3,29	3,33
3,80	3,33	3,20	3,00	3,00	3,43	3,67

Gambar 9. Hasil hitung rata-rata

Hitung Nilai Y / Label

Rumus menghitung nilai Y / Label :

$$= IF(AND(Keorganisasian Dalam Kampus="Ya";$$

$$Kehadiran Keorganisasian Dalam Kampus="Ya");"Ya";"Tidak")$$

(3)

Jika nilai Keorganisasian Dalam Kampus = “Ya” dan Kehadiran Keorganisasian Dalam Kampus = “Ya”, maka nilai Mahasiswa Aktif Organisasi = “Ya”, jika Keorganisasian Dalam Kampus dan atau Kehadiran Keorganisasian Dalam Kampus = “Tidak” maka nilai Mahasiswa Aktif Organisasi = “Tidak”. Berikut hasil perhitungan nilai label seperti pada Gambar 10 berikut.

Keorganisasian Dalam Kampus	Kehadiran Keorganisasian Dalam Kampus	Mahasiswa Aktif Organisasi
Tidak	Tidak	Tidak
Ya	Ya	Ya
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Ya	Ya	Ya
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Ya	Tidak	Tidak
Ya	Ya	Ya
Ya	Ya	Ya
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Ya	Tidak	Tidak
Ya	Ya	Ya
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Ya	Ya	Ya
Ya	Ya	Ya
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Tidak	Tidak	Tidak
Ya	Ya	Ya

Gambar 10. Setelah Hitung Nilai Label

Dataset Final dan Atribut

Data akhir yang diperoleh adalah dataset yang akan digunakan dalam proses data mining beserta atribut yang diperoleh seperti yang tertuang pada Tabel 1 dan Tabel 2 berikut ini.

Tabel 1. Dataset Final

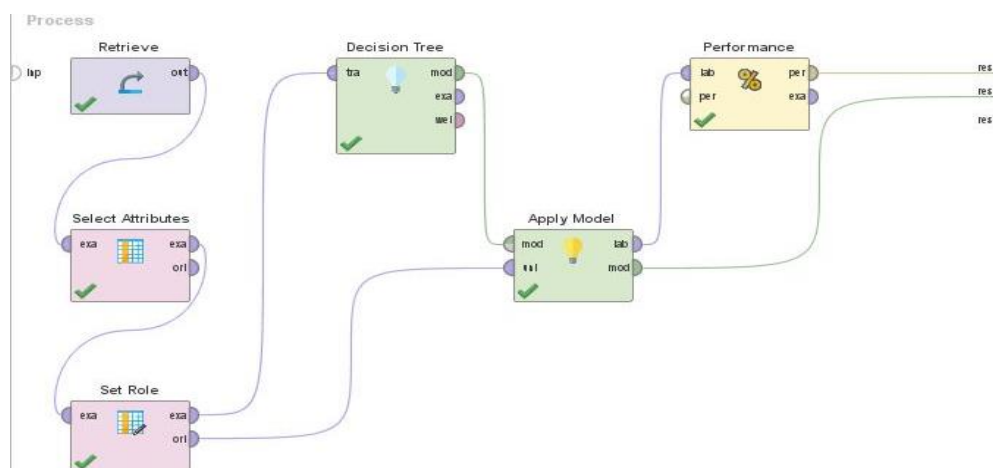
No	Keorganisasi an Luar Kampus (X1)	Keorganisasian Pendidikan Sebelumnya (X2)	Public Speaking (X3)	Problem Solving (X4)	Confidence Level (X5)	Kepribadian (X6)	Mahasiswa Aktif Organisasi (Y)
1	Tidak	Tidak	4	4	4	Koleris	Ya
2	Tidak	Tidak	3	3	4	Melankolis	Tidak
3	Tidak	Ya	4	4	3	Koleris	Tidak
4	Tidak	Ya	3	3	3	Phlegmatis	Ya
5	Tidak	Ya	3	4	4	Koleris	Ya
6	Ya	Ya	4	3	4	Sanguinis	Tidak
7	Ya	Ya	4	3	4	Sanguinis	Tidak
8	Tidak	Tidak	3	3	3	Sanguinis	Tidak
9	Tidak	Ya	3	3	3	Koleris	Tidak
10	Tidak	Ya	4	4	4	Sanguinis	Tidak
....
167	Ya	Ya	2	2	2	Phlegmatis	Tidak

Tabel 2. Tabel Keterangan Atribut

Set Role	Atribut	Keterangan
ID	NPM	Menginformasikan Nomor Pokok Mahasiswa
X₁	Keorganisasian Luar Kampus	Menginformasikan keikutsertaan mahasiswa dalam keorganisasian di luar kampus
X₂	Keorganisasian Pendidikan Sebelumnya	Menginformasikan keikutsertaan mahasiswa dalam keorganisasian pada saat masih dalam jenjang pendidikan sebelumnya
X₃	Public Speaking	Menginformasikan kemampuan mahasiswa dalam berbicara di depan umum
X₄	Problem Solving	Menginformasikan kemampuan mahasiswa dalam penyelesaian suatu masalah
X₅	Confidence Level	Menginformasikan tingkat kepercayaan diri mahasiswa
X₆	Kepribadian	Menginformasikan kategori kepribadian mahasiswa : Koleris, Sanguinis, Melankolis, atau Phlegmatis
Label	Mahasiswa Aktif Organisasi	Menginformasikan hasil data kuesioner

4.3 Proses Data Mining Decision Tree C4.5

Berdasarkan hasil perhitungan baik yang dilakukan secara manual dan menggunakan RapidMiner sebagai prosesnya maka dapat menggambarkan hasil tahapan pada Gambar 11, Rule yang diperoleh beserta pohon keputusan pada Gambar 12.



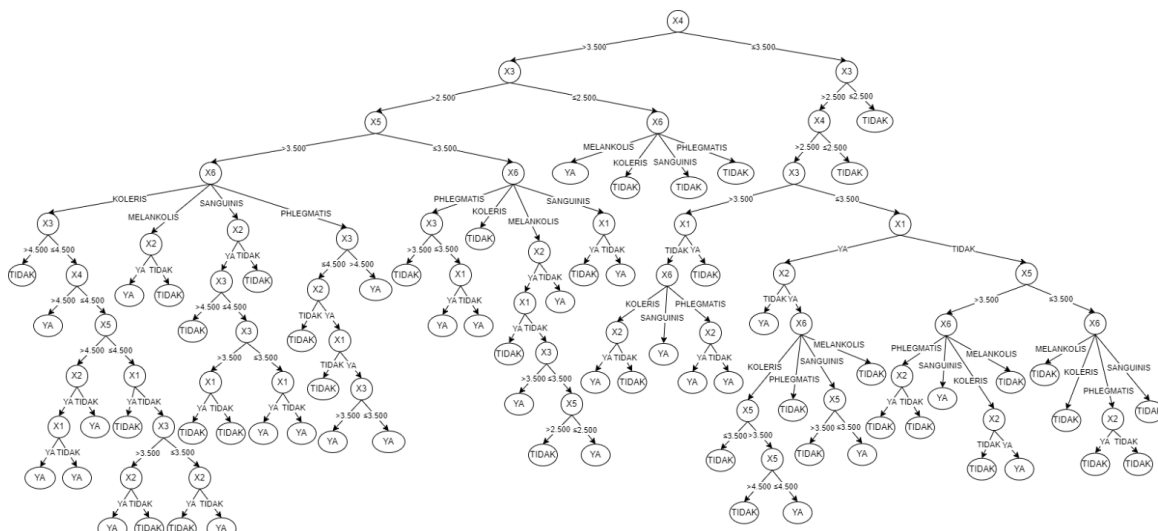
Gambar 11. Proses Data Mining dengan RapidMiner

Keterangan :

- 1) *Retrieve* : Operator untuk memuat dataset ke dalam process dihubungkan dengan *Select Attributes*
- 2) *Select Attributes* : Operator untuk memilih atribut yang akan masuk ke dalam process dihubungkan dengan *Set Role*
- 3) *Set Role* : Operator untuk memilih atribut target / prediksi dihubungkan dengan *Decision Tree*
- 4) *Decision Tree* : Operator Modeling dihubungkan dengan *Apply Model*
- 5) *Apply Model* : Operator untuk menguji dataset dihubungkan dengan *Performance*
- 6) *Performance* : Operator untuk evaluasi kinerja, operator ini memberikan daftar nilai kriteria kinerja,

Rule yang dihasilkan :

X4 > 3.500	X4 ≤ 3.500
X3 > 2.500	X3 > 2.500
X5 > 3.500	X4 > 2.500
X6 = Koleris	X3 > 3.500
X3 > 4.500: Tidak {Ya=0, Tidak=3}	X1 = Tidak
X3 ≤ 4.500	X6 = Koleris
X4 > 4.500: Ya {Ya=2, Tidak=0}	X2 = Tidak: Tidak {Ya=0, Tidak=1}
X4 ≤ 4.500	X2 = Ya: Ya {Ya=1, Tidak=0}
X5 > 4.500	X6 = Phlegmatis
X2 = Tidak: Ya {Ya=1, Tidak=0}	X2 = Tidak: Ya {Ya=1, Tidak=0}
X2 = Ya	X2 = Ya: Ya {Ya=1, Tidak=1}
X1 = Tidak: Ya {Ya=1, Tidak=1}	X6 = Sanguinis: Ya {Ya=1, Tidak=1}
X1 = Ya: Ya {Ya=1, Tidak=1}	X1 = Ya: Tidak {Ya=0, Tidak=6}
X5 ≤ 4.500	X3 ≤ 3.500
X1 = Tidak	X1 = Tidak
X3 > 3.500	X5 > 3.500
X2 = Tidak: Tidak {Ya=1, Tidak=2}	X6 = Koleris
X2 = Ya: Ya {Ya=2, Tidak=2}	X2 = Tidak: Tidak {Ya=0, Tidak=2}
X3 ≤ 3.500	X2 = Ya: Ya {Ya=1, Tidak=0}
X2 = Tidak: Ya {Ya=1, Tidak=1}	X6 = Melankolis: Tidak {Ya=0, Tidak=4}
X2 = Ya: Tidak {Ya=1, Tidak=3}	X6 = Phlegmatis
X1 = Ya: Tidak {Ya=4, Tidak=8}	X3 ≤ 3.500
X6 = Melankolis	X1 = Tidak: Ya {Ya=1, Tidak=1}
X2 = Tidak: Tidak {Ya=0, Tidak=1}	X1 = Ya: Ya {Ya=1, Tidak=1}
X2 = Ya: Ya {Ya=3, Tidak=0}	X6 = Sanguinis
X6 = Phlegmatis	X1 = Tidak: Ya {Ya=2, Tidak=0}
X3 > 4.500: Ya {Ya=1, Tidak=0}	X1 = Ya: Tidak {Ya=0, Tidak=2}
X3 ≤ 4.500	X3 ≤ 2.500
X2 = Tidak: Tidak {Ya=0, Tidak=1}	X6 = Koleris: Tidak {Ya=0, Tidak=1}
X2 = Ya	X6 = Melankolis: Ya {Ya=1, Tidak=0}
X1 = Tidak: Tidak {Ya=2, Tidak=4}	X6 = Phlegmatis: Tidak {Ya=0, Tidak=2}
X1 = Ya	X6 = Sanguinis: Tidak {Ya=0, Tidak=2}
X3 > 3.500: Ya {Ya=1, Tidak=1}	X2 = Tidak: Tidak {Ya=0, Tidak=2}
X3 ≤ 3.500: Ya {Ya=1, Tidak=0}	X2 = Ya: Tidak {Ya=1, Tidak=2}
X6 = Sanguinis	X6 = Sanguinis: Ya {Ya=1, Tidak=0}
X2 = Tidak: Tidak {Ya=0, Tidak=2}	X5 ≤ 3.500
X2 = Ya	X6 = Koleris: Tidak {Ya=0, Tidak=2}
X3 > 4.500: Tidak {Ya=0, Tidak=1}	X6 = Melankolis: Tidak {Ya=1, Tidak=4}
X3 ≤ 4.500	X6 = Phlegmatis
X3 > 3.500	X2 = Tidak: Tidak {Ya=0, Tidak=2}
X1 = Tidak: Tidak {Ya=1, Tidak=3}	X2 = Ya: Tidak {Ya=1, Tidak=3}
X1 = Ya: Tidak {Ya=3, Tidak=4}	X6 = Sanguinis: Tidak {Ya=0, Tidak=8}
X3 ≤ 3.500	X1 = Ya
X1 = Tidak: Ya {Ya=2, Tidak=2}	X2 = Tidak: Ya {Ya=1, Tidak=0}
X1 = Ya: Ya {Ya=3, Tidak=2}	X2 = Ya
X5 ≤ 3.500	X6 = Koleris
X6 = Koleris: Tidak {Ya=0, Tidak=3}	X5 > 3.500
X6 = Melankolis	X5 > 4.500: Tidak {Ya=0, Tidak=1}
X2 = Tidak: Ya {Ya=1, Tidak=0}	X5 ≤ 4.500: Ya {Ya=1, Tidak=1}
X2 = Ya	X5 ≤ 3.500: Tidak {Ya=0, Tidak=1}
X1 = Tidak	X6 = Melankolis: Tidak {Ya=0, Tidak=2}
X3 > 3.500: Ya {Ya=1, Tidak=0}	X6 = Phlegmatis: Tidak {Ya=0, Tidak=1}
X3 ≤ 3.500	X6 = Sanguinis
X5 > 2.500: Tidak {Ya=0, Tidak=5}	X5 > 3.500: Tidak {Ya=0, Tidak=1}
X5 ≤ 2.500: Ya {Ya=1, Tidak=0}	X5 ≤ 3.500: Ya {Ya=1, Tidak=0}
X1 = Ya: Tidak {Ya=0, Tidak=3}	X4 ≤ 2.500: Tidak {Ya=0, Tidak=2}
X6 = Phlegmatis	X3 ≤ 2.500: Tidak {Ya=0, Tidak=6}
X3 > 3.500: Tidak {Ya=0, Tidak=1}	



Gambar 12. Hasil Rule dan Pohon Keputusan

4.4 Training dan Testing

Proses training meliputi training error dan testing error dengan menggunakan confusion matrix dari hasil prediksi dan data actual yang diperoleh.

Training Error

Training error berfungsi untuk menghitung dan mengetahui error pada model klasifikasi. Nilai confusion matrix training error dihasilkan dari data yang sama untuk training model dan juga testing model seperti tertera pada Tabel 3.

Tabel 3. Tabel Confusion Matrix Training Error

Prediksi	Data Aktual	
	Ya	Tidak
Ya	TN : 33	FN : 12
Tidak	FP : 18	TP : 104

Keterangan :

- TN : True Negative
- FP : False Positive
- FN : False Negative
- TP : True Positive

Perhitungan Confusion Matrix :

1. Accuracy

$$= \left(\frac{TP + TN}{Total} \right) = \left(\frac{104 + 33}{167} \right) = 0.82035 = 82.04\%$$

Tingkat keakuratan klasifikasi sebesar 82.04%

2. Misclassification Rate

$$= \left(\frac{FP + FN}{Total} \right) = \left(\frac{18 + 12}{167} \right) = 0.1796 = 17.96\%$$

Tingkat kesalahan klasifikasi sebesar 17.96%

3. True Positive Rate/Recall/Sensitivity

$$= \left(\frac{TP}{Data\ Tidak} \right) = \left(\frac{TP}{TP + FN} \right) = \left(\frac{104}{104 + 12} \right) = 0.89655 = 89.66\%$$

Maka yang sebenarnya “Tidak” tingkat prediksinya “Tidak” adalah sebesar 89.66%

4. False Positive Rate

$$= \left(\frac{FP}{Data\ Ya} \right) = \left(\frac{FP}{FP + TN} \right) = \left(\frac{18}{18 + 33} \right) = 0.35294 = 35.29\%$$

Maka yang sebenarnya “Ya” tingkat prediksinya “Tidak” adalah sebesar 35.29%

5. True Negative Rate/Specificity

$$= \left(\frac{TN}{Data\ Ya} \right) = \left(\frac{TN}{FP + TN} \right) = \left(\frac{33}{18 + 33} \right) = 0.64705 = 64.71\%$$

Maka yang sebenarnya “Ya” tingkat prediksinya “Ya” adalah sebesar 64.71%

6. False Negative Rate

$$= \left(\frac{FN}{Data\ Tidak} \right) = \left(\frac{FN}{TP + FN} \right) = \left(\frac{12}{104 + 12} \right) = 0.10344 = 10.34\%$$

Maka yang sebenarnya “Tidak” tingkat prediksinya “Ya” adalah sebesar 10.34%

7. Precision

$$= \left(\frac{TP}{Prediksi\ Tidak} \right) = \left(\frac{TP}{TP + FP} \right) = \left(\frac{104}{104 + 18} \right) = 0.85245 = 85.25\%$$

Tingkat kebenaran dalam memprediksi “Tidak” adalah sebesar 85.44ra%

8. Prevelance

$$= \left(\frac{Data\ Tidak}{Total} \right) = \left(\frac{TP + FN}{Total} \right) = \left(\frac{104 + 12}{167} \right) = 0.69461 = 69.46\%$$

Testing Error

Berbeda dengan training error, testing error akan menggunakan 2 data yang terpisah. Dataset akan dibagi menjadi 2 bagian yaitu data latih dan data uji dengan menggunakan operator split data pada rapidminer. Dalam praktiknya dataset akan di bagi menjadi 2 bagian dengan ratio data training sebesar 0.75 dan data testing sebesar 0.25 yang tertera pada Tabel 4.

Tabel 4. Tabel Confusion Matrix Testing Error

Prediksi	Data Aktual	
	Ya	Tidak
Ya	TN : 2	FN : 4
Tidak	FP : 11	TP : 25

Keterangan :

- TN : True Negative • FP : False Positive • FN : False Negative • TP : True Positive
- Perhitungan Confusion Matrix :

1. Accuracy

$$= \left(\frac{TP + TN}{Total} \right) = \left(\frac{25 + 2}{42} \right) = 0.64285 = 64.29\%$$

Tingkat keakuratan klasifikasi sebesar 64.29%

2. Misclassification Rate

$$= \left(\frac{FP + FN}{Total} \right) = \left(\frac{11 + 4}{42} \right) = 0.35714 = 35.71\%$$

Tingkat kesalahan klasifikasi sebesar 35.71%

3. True Positive Rate/Recall/Sensitivity

$$= \left(\frac{TP}{Data\ Tidak} \right) = \left(\frac{TP}{TP + FN} \right) = \left(\frac{25}{25 + 4} \right) = 0.86206 = 86.21\%$$

Maka yang sebenarnya “Tidak” tingkat prediksinya “Tidak” adalah sebesar 86.21%

4. False Positive Rate

$$= \left(\frac{FP}{Data\ Ya} \right) = \left(\frac{FP}{FP + TN} \right) = \left(\frac{11}{11 + 2} \right) = 0.84615 = 84.62\%$$

Maka yang sebenarnya “Ya” tingkat prediksinya “Tidak” adalah sebesar 84.62%

5. True Negative Rate/Specificity

$$= \left(\frac{TN}{Data\ Ya} \right) = \left(\frac{TN}{FP + TN} \right) = \left(\frac{2}{11 + 2} \right) = 0.15384 = 15.38\%$$

Maka yang sebenarnya “Ya” tingkat prediksinya “Ya” adalah sebesar 15.38%

6. False Negative Rate

$$= \left(\frac{FN}{Data\ Tidak} \right) = \left(\frac{FN}{TP + FN} \right) = \left(\frac{4}{25 + 4} \right) = 0.13793 = 13.79\%$$

Maka yang sebenarnya “Tidak” tingkat prediksinya “Ya” adalah sebesar 13.79%

7. Precision

$$= \left(\frac{TP}{Prediksi\ Tidak} \right) = \left(\frac{TP}{TP + FP} \right) = \left(\frac{25}{25 + 11} \right) = 0.69444 = 69.44\%$$

Tingkat kebenaran dalam memprediksi “Tidak” adalah sebesar 69.44%

8. Prevelance

$$= \left(\frac{Data\ Tidak}{Total} \right) = \left(\frac{TP + FN}{Total} \right) = \left(\frac{25 + 4}{42} \right) = 0.69047 = 69.05\%$$

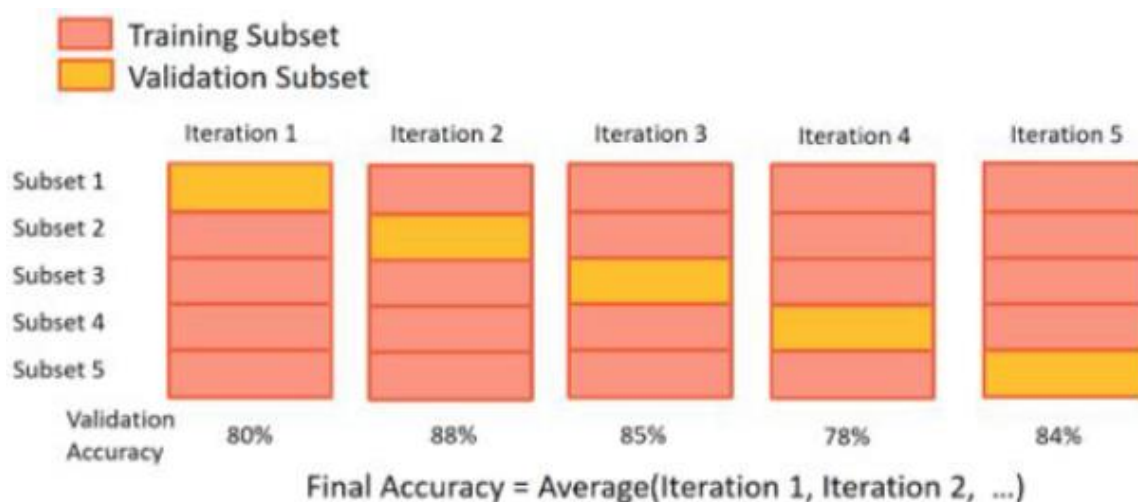
Berikut adalah hasil perbandingan dari hasil training error dan testing error seperti pada Tabel 5.

Tabel 5. Perbandingan Confusion Matrix Training Error dan Testing Error

	Training Error	Test Error
Accuracy	82.04%	64.29%
Misclassification Rate	17.96%	35.71%
True Positive Rate	89.66%	86.21%
False Positive Rate	35.29%	84.62%
True Negative Rate	64.71%	15.38%
False Negative Rate	10.34%	13.79%
Pecision	85.25%	69.44%
Prevelance	69.46%	69.05%

4.5 Pengujian Model

Pengujian akan dilakukan untuk menguji kinerja daripada metode Algoritma Decision Tree C4.5 yang diterapkan pada penelitian ini. Pengujian model menggunakan operator Cross Validation dengan menerapkan K-Fold Validation[16]. Model ini menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K k-fold seperti tertera pada Gambar 13 berikut.

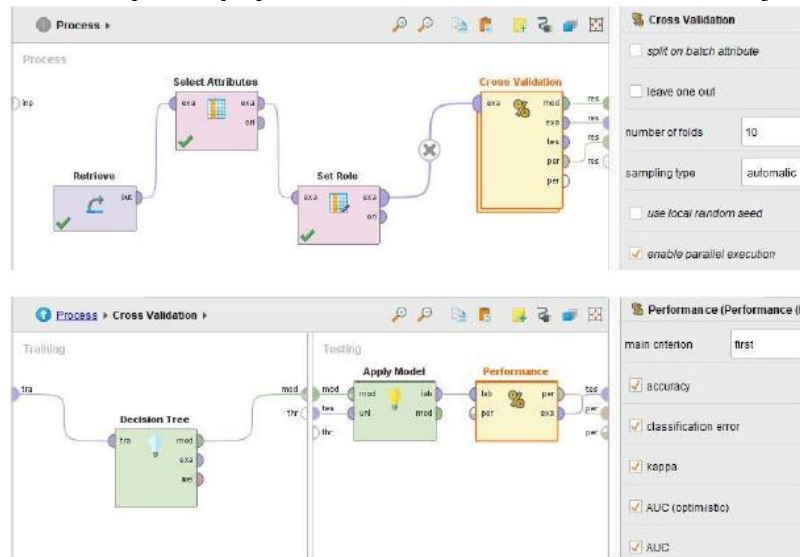


Gambar 13. Konsep K-Fold Validation

K-Fold Validation, membagi data menjadi 2 bagian yaitu data training dan data testing dengan perbandingan tertentu. K-Fold Validation akan membagi dataset menjadi beberapa bagian terpisah dengan masing-masing ukuran yang sama, selanjutnya data akan di uji dan di hitung nilai rata-rata test error. Pengujian model dengan menerapkan *K-Fold Validation* dilakukan secara 3 kali dengan nilai subset yang berbeda-beda, yaitu:

1) K-Fold Validation 10

Dataset akan dibagi menjadi 10 bagian yang terdiri dari 9 sampel *data training* dan 1 sampel *data testing*. Mekanisme dan prosesnya pada Gambar 14 dan Tabel 6 adalah hasil uji 1.



Gambar 14. Uji 1

Tabel 6. Tabel Confusion Matrix Uji 1

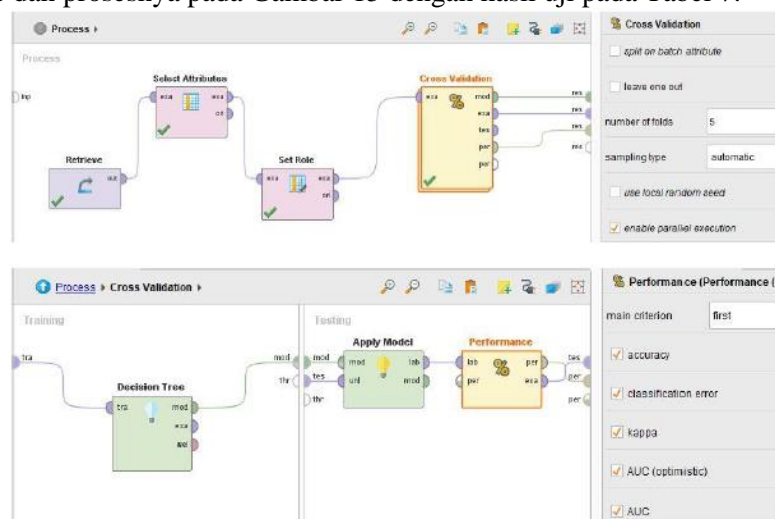
	Ya	Tidak
Prediksi Ya	TN : 11	FN : 33
Prediksi Tidak	FP : 40	TP : 83

$$= \left(\frac{TP + TN}{Total} \right) = \left(\frac{83 + 11}{167} \right) = 0.5636 = 56.36\%$$

Hasil *confusion matrix* dari uji 1 (k-fold validation 10) menghasilkan nilai accuracy sebesar 56.36%.

2) K-Fold Validation 5

Dataset akan dibagi menjadi 5 bagian yang terdiri dari 4 sampel *data training* dan 1 sampel *data testing*. Mekanisme dan prosesnya pada Gambar 15 dengan hasil uji pada Tabel 7.



Gambar 15. Uji 2

Tabel 7. Tabel Confusion Matrix Uji 2

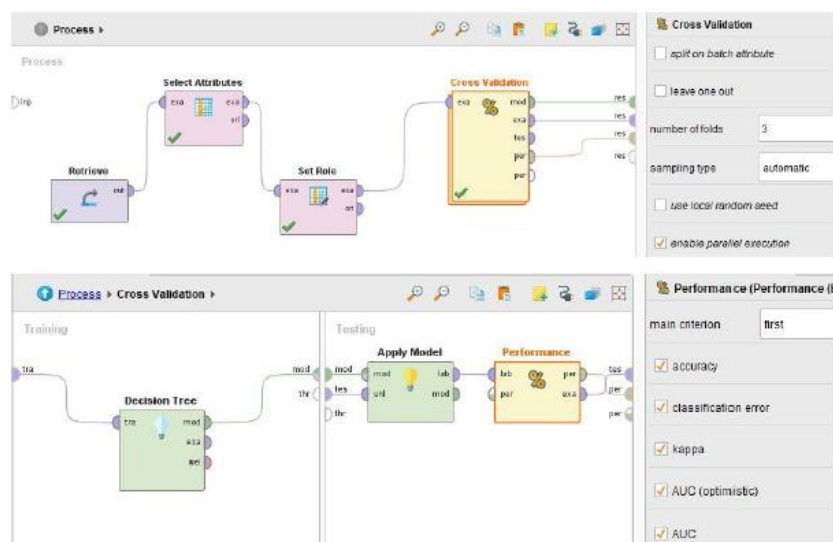
	Ya	Tidak
Prediksi Ya	TN : 11	FN : 28
Prediksi Tidak	FP : 40	TP : 88

$$= \left(\frac{TP + TN}{Total} \right) = \left(\frac{88 + 11}{167} \right) = 0.5927 = 59.27\%$$

Hasil *confusion matrix* dari uji 2 (k-fold validation 5) menghasilkan nilai accuracy sebesar 59.27%.

3) K-Fold Validation 3

Dataset akan dibagi menjadi 3 bagian yang terdiri dari 2 sampel *data training* dan 1 sampel *data testing*. Perhatikan prosesnya pada Gambar 16 dan hasilnya tertera pada Tabel 8.



Gambar 16. Uji 3

Tabel 8. Tabel Confusion Matrix Uji 3

	Ya	Tidak
Prediksi Ya	TN : 8	FN : 33
Prediksi Tidak	FP : 43	TP : 83

$$= \left(\frac{TP + TN}{Total} \right) = \left(\frac{83 + 8}{167} \right) = 0.5450 = 54.50\%$$

Hasil *confusion matrix* dari uji 3 (k-fold validation 3) menghasilkan nilai accuracy sebesar 54.50%. Berdasarkan hasil uji secara keseluruhan hasil perbandingannya ada pada tabel 9 berikut.

Tabel 9. Perbandingan Accuracy Uji 1, Uji 2, Uji 3

	K-Fold Validation	Training : Testing	Accuracy
Uji 1	k = 10	9 : 1	56.36%
Uji 2	k = 5	4 : 1	59.27%
Uji 3	k = 3	2 : 1	54.50%

Nilai accuracy tertinggi yang dihasilkan pada pengujian model terdapat pada uji 2 yaitu k-fold validation 5 dengan perbandingan data training : data testing yaitu 4 : 1, yang menghasilkan nilai accuracy sebesar 59.27%.

5 Kesimpulan

Implementasi data mining dengan menggunakan metode algoritma decision tree c4.5 dapat menghasilkan rule untuk memprediksi Mahasiswa status untuk aktif dalam organisasi berdasarkan atribut Keorganisasian Luar Kampus, Keorganisasian Pendidikan Sebelumnya, *Public Speaking*, *Problem Solving*, *Confidence*, dan Kepribadian. Rule yang telah dihasilkan dapat digunakan untuk memprediksi kecenderungan Mahasiswa Aktif Organisasi berdasarkan pengujian model. Pengujian model dengan menggunakan *cross validation* menghasilkan nilai akurasi tertinggi yaitu sebesar 59.27% dengan menggunakan k-fold validation 5, yang berarti dataset dibagi menjadi 5 bagian yang terdiri dari 4 sampel data training dan 1 sampel data

<http://sistemasi.ftik.unisi.ac.id>

training. Dapat disimpulkan bahwa model akan optimal jika perbandingan data training : data testing yaitu sebesar 4 : 1 atau 80% : 20%.

Referensi

- [1] A. Sholikhah, “Pengaruh Keaktifan Mahasiswa Dalam Organisasi Terhadap Prestasi Belajar Mahasiswa Jurusan Pendidikan Ekonomi FE Unesa Angkatan 2015 | Jurnal Pendidikan Ekonomi (JUPE).” [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jupe/article/view/24509>. [Accessed: 05-Jun-2022].
- [2] P. Bayina, R. Nurdin¹, S. Likuallo², and A. Meiliska, “Faktor-Faktor yang Mempengaruhi Minat Mahasiswa Berorganisasi,” *J. Manaj. dan Organ. Rev.*, vol. 2, no. 2, pp. 122–131, Nov. 2020.
- [3] J. F. Kimberly, D. B. Prakoso, and T. C. Efrata, “Peran Individual Innovation Capability, Motivasi Intrinsik, dan Self-Efficacy Terhadap Kinerja Individu dalam Organisasi Mahasiswa,” *Media Mahard.*, vol. 17, no. 2, pp. 231–243, Jan. 2019.
- [4] A. G. Pertiwi, T. Widyaningtyas, and U. Pujiyanto, “Classification of province based on dropout rate using C4.5 algorithm,” *Proc. - 2017 Int. Conf. Sustain. Inf. Eng. Technol. SIET 2017*, vol. 2018-Janua, pp. 410–413, Feb. 2018.
- [5] S. Supangat, A. R. Amna, and T. Rahmawati, “Implementasi Decision Tree C4.5 Untuk Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun,” *Teknika*, vol. 7, no. 2, pp. 73–78, Nov. 2018.
- [6] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” *J. Phys. Conf. Ser.*, vol. 1255, no. 1, p. 012005, Aug. 2019.
- [7] M. A. Sembiring, M. F. L. Sibuea, and A. Sapta, “Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar,” *J. Sci. Soc. Res.*, vol. 1, no. 1, pp. 73–79, Feb. 2018.
- [8] J. P. Jiawei Han, Micheline Kamber, *Data mining: Data mining concepts and techniques*. 2014.
- [9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, “Data Mining: Practical Machine Learning Tools and Techniques,” *Data Min. Pract. Mach. Learn. Tools Tech.*, pp. 1–621, Nov. 2016.
- [10] J. N.Undavia, P. M. Dolia, and N. P. Shah, “Prediction of Graduate Students for Master Degree based on Their Past Performance using Decision Tree in Weka Environment,” *Int. J. Comput. Appl.*, vol. 74, no. 11, pp. 23–29, Jul. 2013.
- [11] P.-N. Tan, M. Steinbach, and V. Kumar, “Association Analysis: Basic Concepts and Algorithms,” *Introd. to Data Min.*, pp. 327–414, 2005.
- [12] V. Vyas and V. Uma, “An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner,” in *Procedia Computer Science*, 2018.
- [13] R. K. Markus Hofmann, “RapidMiner: Data Mining Use Cases and Business Analytics Applications Chapman & Hall/CRC Data Mining and Knowledge Discovery: Amazon.de: Hofmann, Markus, Klinkenberg, Ralf: Fremdsprachige Bücher.”
- [14] R. Suwandar, J. Manajemen, and S. Nitro Makassar, “Prosiding SENTIA 2017- Politeknik Negeri Malang Analisis Faktor-Faktor Mahasiswa dalam Memilih Unit Kegiatan Mahasiswa Perguruan Tinggi di Kota Makassar.”
- [15] W. D. Pratisti, A. Agung, and A. Ardeliaputri, “Potret Strategi Pemecahan Masalah pada Mahasiswa yang Aktif Berorganisasi,” *Ikat. Psikol. Perkemb. Indones.*, 2013.
- [16] P. Jiang and J. Chen, “Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation,” *Neurocomputing*, vol. 198, pp. 40–47, Jul. 2016.