

PREDIKSI KELULUSAN TEPAT WAKTU PESERTA MAGANG MENGUNAKAN ALGORITMA C4.5 *DECISION TREE* PADA PERUSAHAAN XYZ

Muhammad Azmi Fadhlurrahman^{1*}, Titin Fatimah²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, DKI Jakarta,

IndonesiaEmail: ^{1*}mazmi9999@gmail.com, ²titin.fatimah@budiluhur.ac.id
(* : corresponding author)

Abstrak- Pelatihan kerja atau magang adalah keseluruhan kegiatan yang dilakukan untuk memberikan, memperoleh, meningkatkan, dan memaksimalkan serta menumbuhkan kompetensi kerja, produktivitas, kedisiplinan, perubahan pola pikir, dan etos bekerja pada tingkat keterampilan dan keahlian tertentu yang sesuai dengan tingkatan dan kualifikasi posisi atau pekerjaan. Belum adanya suatu sistem yang mampu melakukan prediksi kelulusan tepat waktu peserta magang di Perusahaan XYZ. Penelitian ini bertujuan untuk merancang suatu sistem yang mampu untuk membantu pengambil keputusan untuk mempercepat proses pengambilan keputusan dalam menentukan kelulusan tepat waktu peserta magang di Perusahaan XYZ. Data mining dapat menjadi solusi untuk memprediksi apakah peserta magang berpotensi akan lulus tepat waktu atau tidak. Oleh karena itu, pada penelitian ini akan mengolah data peserta magang untuk memprediksi peserta magang yang berpotensi lulus tepat waktu dengan menggunakan metode Decision Tree pada algoritme C4.5. Untuk pengujian menggunakan algoritme C4.5 dengan membagi *dataset* sebesar 80% untuk data latih dan 20% untuk data uji. Setelah dilakukan pembagian *dataset* akan dilakukan proses perhitungan dan pembentukan pohon keputusan, dan akan menghasilkan sebuah keputusan akhir berupa prediksi kelulusan tepat waktu atau terlambat. Bahasa pemrograman menggunakan bahasa pemrograman PHP menggunakan framework Codeigniter. Hasil penelitian menunjukkan bahwa pohon keputusan yang dihasilkan dari algoritme C4.5 memiliki akurasi sebesar 88.88%, dengan menggunakan algoritme C4.5 berhasil meningkatkan akurasi algoritma C4.5 sebesar 3.88% dari penelitian sebelumnya yang menggunakan metode Naive Bayes pada kasus yang sama, sehingga menghasilkan hasil yang lebih baik dalam memprediksi peserta magang yang berpotensi tepat waktu.

Kata Kunci: Magang, Data Mining, C4.5.

PREDICTION OF ON-TIME GRADUATION OF INTERNATIONAL PARTICIPANTS USING THE C4.5 DECISION TREE ALGORITHM IN XYZ COMPANIES

Abstract- Job training or apprenticeship is the whole activity carried out to provide, obtain, improve, and maximize and grow work competence, productivity, discipline, change of mindset, and work ethic at a certain level of skills and expertise in accordance with the level and qualifications of the position or job. There is no system capable of predicting the timely graduation of apprentices at Company XYZ. This study aims to design a system capable of helping decision makers to speed up the decision-making process in determining the timely graduation of apprentices at XYZ Company. Data mining can be a solution for predicting whether potential apprentices will graduate on time or not. Therefore, this study will process apprentice data to predict apprentices who have the potential to graduate on time using the Decision Tree method in the C4.5 algorithm. For testing using the C4.5 algorithm by dividing the dataset by 80% for training data and 20% for test data. After dividing the dataset, the process of calculating and forming a decision tree will be carried out, and will produce a final decision in the form of a prediction of graduation on time or late. The programming language uses the PHP programming language using the CodeIgniter framework. The results showed that the decision tree generated from the C4.5 algorithm had an accuracy of 88.88%, using the C4.5 algorithm succeeded in increasing the accuracy of the C4.5 algorithm by 3.88% from previous studies using the Naive Bayes method in the same case, resulting in better results in predicting potential timely apprentices.

Keywords: Internship, Data Mining, C4.5.

1. PENDAHULUAN

Perluasan teknologi khususnya di bidang informasi, memiliki efek mendalam pada setiap aspek kehidupan manusia. Kemajuan teknologi yang berhubungan dengan komputer memiliki dampak yang signifikan terhadap

keberadaan manusia. Bahkan saat ini komputer dapat mengambil peran manusia sebagai alat untuk melakukan suatu tugas. Pelatihan kerja atau magang adalah keseluruhan kegiatan yang dilakukan untuk memberikan, memperoleh, meningkatkan, dan memaksimalkan serta menumbuhkan kompetensi kerja, produktivitas, kedisiplinan, perubahan pola pikir, dan etos bekerja pada tingkat keterampilan dan keahlian tertentu yang sesuai dengan tingkatan dan kualifikasi posisi atau pekerjaan.

Selama ini dalam menentukan kelulusan peserta pelatihan atau magang di perusahaan XYZ hanya berdasarkan penilaian yang dilakukan dengan melihat dari beberapa faktor, seperti keaktifan, sikap, rating trainer, kontribusi artikel, sikap, dan nilai akhir proyek. Namun hal ini tidak jarang keputusan yang diambil kurang akurat. Belum adanya suatu sistem yang mampu menangani permasalahan ini.

Oleh karena permasalahan diatas maka perlu dirancang suatu sistem yang mampu melakukan prediksi kelulusan tepat waktu terhadap peserta magang, yang diharapkan dapat membantu melakukan penilaian dan pengambilan keputusan dalam mendapatkan informasi sebagai bahan pertimbangan untuk pengambilan keputusan dalam menentukan kelulusan tepat waktu peserta magang.

Data adalah bahan mentah yang perlu dilakukan pengolahan sehingga menghasilkan informasi atau keterangan, baik kualitatif maupun kuantitatif yang menunjukkan fakta sehingga dapat memberi manfaat bagi peneliti atau memberi gambaran kepada peneliti tentang kondisi atau suatu keadaan[1].

Data mining merupakan suatu ilmu yang menguraikan hasil penemuan berupa pengetahuan pada sekumpulan data, dan dalam proses data mining menggunakan teknik statistik matematika, kecerdasan buatan, dan mesin pembelajaran untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat[2]. *Data mining* merupakan kegiatan meliputi pengumpulan, pemakaian data historis dalam menemukan keteraturan, pola, atau hubungan dalam dataset berukuran besar. Output dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan[3].

Data mining memiliki beberapa fungsi yaitu, *classification*, *clustering*, *association*, *regression*, *forecasting*, *sequence analysis*, dan *deviation analysis*[4].

KDD (*Knowledge Discovery in Database*) merupakan tahapan dari proses data mining, KDD (*Knowledge Discovery in Database*) ini juga membatasi penggunaan metode ilmiah dalam knowledge mining[5]. Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya agar dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu[6]. *Decision Tree* adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan hasil dari pengujian dan leaf node (simpul terminal) menyatakan kelas-kelas atau distribusi kelas[7]. Algoritma C4.5 merupakan algoritme pohon keputusan, banyak sekali perkembangannya tetapi yang sering dipakai adalah ID3 dan C4.5 yang sama-sama membangun pohon keputusan, oleh karena itu keduanya mempunyai prinsip yang sama[8].

Entropy merupakan ukuran ketidakpastian, yakni perbedaan keputusan terhadap nilai atribut tertentu. Semakin tinggi nilai entropy, semakin tinggi perbedaan keputusan[9]. Gain merupakan salah satu langkah pemilihan atribut yang digunakan untuk memilih tes atribut setiap simpul pada pohon keputusan atau dengan kata lain gain merupakan tingkat pengaruh suatu atribut terhadap keputusan atau ukuran efektifitas suatu variabel dalam mengklasifikasi data[9].

Dari penelitian yang sudah dilakukan oleh (Lamramot et al., 2022) yaitu Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritme C4.5. dataset yang digunakan berjumlah 1010 data karyawan. Pada tahap *preprocessing* pada penelitian sebelumnya terdapat tahapan data seleksi dan data transformasi. Hasil pengujian dengan menggunakan metode *confusion matrix* dan mendapatkan hasil akurasi sebesar 94.66%.

Selain itu penelitian yang dilakukan oleh (Yunianita et al., 2018) yaitu Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritme Pohon Keputusan C45 dengan dataset yang berasal dari Teknik Informatika UII angkatan 2010 hingga 2013. Dengan menggunakan teknik *filter* dan *wrapper* terhadap subset atribut, dan penerapan rules ke dalam aplikasi. Memberikan hasil akurasi yang cukup baik sebesar 73.9%.

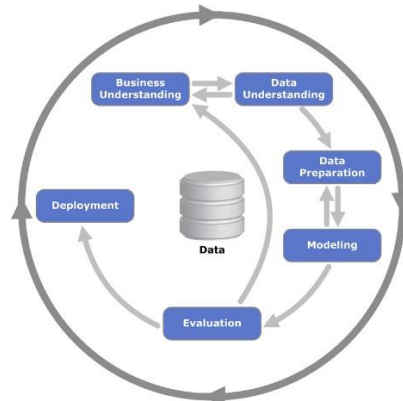
Sedangkan penelitian yang dilakukan oleh (Sucipto 2015) algoritme klasifikasi *data mining* dengan model algoritme C4.5 dilakukan dengan pengujian-pengujian yang terukur melalui uji AUC, ROC dan T-Test dengan bantuan *software RapidMiner* yang telah dilakukan pengujian dengan menggunakan parameter biodata nasabah sebanyak 1312 data, yang menghasilkan akurasi sebesar secara keseluruhan nilai hasil validasi yaitu *accuracy* = 91.06%, *precision* = 100.00%, dan *recall* = 78.00%.

Penelitian ini menggunakan metode algoritme C4.5 *Decision Tree*, dengan menggunakan algoritme C4.5 berhasil meningkatkan akurasi algoritma C4.5 sebesar 3.88% dari penelitian sebelumnya yang menggunakan metode Naive Bayes pada kasus yang sama, sehingga menghasilkan hasil yang lebih baik dalam memprediksi peserta magang yang berpotensi tepat waktu.

2. METODE PENELITIAN

2.1 CRISP-DM

Pendekatan Data Mining CRISP-DM digunakan dalam pekerjaan ini sebagai solusi masalah umum untuk bisnis dan penelitian. Enam langkah dari teknik ini adalah Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Penyebaran[10].



Gambar 1. CRISP-DM

Berdasarkan Gambar 1, berikut adalah penjelasan dari setiap tahapan yang ada:

a. Business Understanding

Pada tahap ini yaitu memahami tujuan dan kebutuhan dari masalah yang dihadapi. Tujuan dari penelitian ini adalah untuk memprediksi kelulusan tepat waktu peserta magang di masa yang akan datang agar memiliki perkiraan yang akurat dan dapat digunakan sebagai bahan evaluasi dari sistem yang sudah ada.

b. Data Understanding

Tahap ini diawali dengan pengumpulan data peserta magang di Perusahaan XYZ pada periode tahun 2022, dilanjutkan dengan proses pemahaman terhadap data tersebut untuk menemukan bagian-bagian yang menarik untuk digunakan pada tahap awal.

c. Data Preparation

Pada tahap ini dilakukan pengolahan data agar dapat mencapai hasil yang diinginkan. Ada beberapa langkah dalam pengolahan data yaitu melakukan *data selection* atau pemilihan data yang relevan dengan penelitian dan *data transformation* yaitu melakukan perubahan nilai pada beberapa atribut dan menambahkan 1 atribut baru sebagai atribut kelas.

d. Modelling

Pada tahap pemodelan, teknik pemodelan dipilih dan diterapkan, dan beberapa parameter disesuaikan untuk mendapatkan nilai yang optimal. Pada penelitian ini digunakan algoritma *C4.5 Decision Tree*. Dataset dibagi menjadi data latih dan data uji, dengan presentase data latih sebesar 80% dan data uji sebesar 20%.

e. Evaluation

Langkah evaluasi dilakukan untuk mengetahui performa dari model yang digunakan, model yang dihasilkan sebelumnya diuji dengan menggunakan data testing sebagai data masukan. Pada penelitian ini akan menggunakan pengujian dengan melihat dari nilai akurasi yang dihasilkan dari model yang digunakan.

f. Deployment

Pada tahap ini, pengetahuan atau informasi yang diperoleh disajikan dalam bentuk khusus untuk digunakan pengguna akhir. Fase penyebaran dapat berupa pembuatan aplikasi atau dalam bentuk laporan.

3. HASIL DAN PEMBAHASAN

Analisis, implementasi atau hasil pengujian, dan diskusi tentang subjek studi potensial yang mungkin dilakukan sebelum memilih teknik semuanya termasuk dalam bagian ini. Selain itu, area ini mencakup penjelasan dalam bentuk teks, ilustrasi, tabel, dan lainnya.

3.1 Business Understanding

Terdapat sebanyak 225 data peserta magang pada periode 2022. Penelitian ini memprediksi tingkat kelulusan tepat waktu peserta magang pada Perusahaan XYZ dengan menggunakan metode algoritme *Decision Tree C4.5*.

3.2 Data Understanding

Penelitian ini dilakukan dengan menggunakan data sekunder yang didapati dari internal Perusahaan XYZ. Data yang di dapatkan berupa file *excel* yang berisikan *record* data dalam rentang waktu yang diambil pada penelitian ini yaitu selama periode Januari – Desember tahun 2022. Berisikan 14 atribut yaitu Nama Lengkap, Jenis Kelamin, Status, Divisi, Asal Sekolah, Alur Masuk, Durasi, Sikap, Pengetahuan, Keterampilan, Keaktifan, *Rating Trainer*, *Projek Akhir* dan *Artikel*. Data awal seperti pada Gambar 2. Dari gambar 2 terlihat data peserta magang periode tahun 2022 dengan jumlah data sebanyak 225 baris data dan atribut berjumlah 14 atribut.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Nama Lengkap	Jenis Kelamin	Status	Divisi	Asal Sekolah	Alur Masuk	Durasi	Sikap	Pengetahuan	Keterampilan	Keaktifan	Rating Trainer	Projek Akhir	Artikel	
2	Betlin Puji	perempuan	mahasiswa	Art & Illust	Univ Sahid	Mandiri	2 Bulan	90	80	95	85	4.65	83	Published	
3	Andika Eki	laki-laki	mahasiswa	3D Creatio	Univ Bina	Premium	3 Bulan	92	90	90	90	4.72	76	Inpublished	
4	Krisna Agul	laki-laki	SMK	3D Creatio	SMK Nege	Mandiri	4 Bulan	95	91	93	92	4.55	98	Published	
5	An Nasti A	perempuan	mahasiswa	Game Dev	Univ Dian	Mandiri	2 Bulan	95	92	88	95	4.80	97	Published	
6	Oktavian I	perempuan	smk	Game Dev	SMK Nege	Mandiri	3 Bulan	90	94	91	88	4.92	97	Published	
7	martinus r	laki-laki	smk	Game Dev	smk telek	Premium	3 Bulan	75	95	90	85	4.26	89	Published	
8	vensensius	laki-laki	smk	Programms	smk telek	Mandiri	4 Bulan	96	91	96	84	4.33	84	Published	
9	Ryan Sbas	laki-laki	mahasiswa	Art & Illust	Univ Merc	Mandiri	3 Bulan	94	90	93	95	4.75	97	Published	
10	Khadana r	laki-laki	mahasiswa	game des	Univ Dian	Premium	2 Bulan	95	89	94	91	4.56	90	Published	
11	Alira Cahy	perempuan	mahasiswa	Digital Mai	Univ Nege	Mandiri	1 Bulan	95	95	95	93	4.55	77	Published	
12	Rahadika I	laki-laki	mahasiswa	Video & Ai	UDINUS	Premium	2 Bulan	90	93	85	92	4.20	89	Published	
13	Dimas Adil	laki-laki	mahasiswa	QA & Test	Univ Telko	Mandiri	3 Bulan	94	75	87	90	4.77	87	Published	
14	Mochama	laki-laki	mahasiswa	Programm	Univ Tekn	Mandiri	2 Bulan	92	94	92	90	4.63	87	Inpublished	
15	Annis Nu	perempuan	smk	3D Creatio	SMK Telek	Premium	4 Bulan	96	88	93	90	4.54	82	Published	
16	Elvina Sarz	perempuan	mahasiswa	QA & Test	Univ Telko	Mandiri	2 Bulan	97	84	94	90	4.93	85	Published	
17	Christian C	laki-laki	mahasiswa	Game Dev	Univ Dian	Premium	3 Bulan	94	94	92	88	4.25	95	Published	
18	Fahd Sjarif	laki-laki	mahasiswa	Programm	STMM Yq	Premium	1 Bulan	95	90	95	80	4.51	98	Published	
19	Ektin Muh	laki-laki	mahasiswa	3D Creatio	Univ Dian	Premium	2 Bulan	97	91	95	85	4.60	93	Published	
20	Toto M	perempuan	mahasiswa	Digital Mai	Univ Nege	Mandiri	2 Bulan	97	91	95	85	4.60	93	Published	

Gambar 2. Data Awal Peserta Magang

3.3 Data Preparation

Pada tahapan ini diterapkan pengolahan data agar dapat mencapai hasil yang diinginkan. Ada beberapa langkah dalam pengolahan data untuk mempersiapkan semua kegiatan yang dibutuhkan yaitu:

a. Data Selection

Pada tahap ini dilakukan pemilihan data yang relevan dengan penelitian yang diperoleh dari *dataset* awal yang diterima. Sebelumnya terdapat 14 atribut data kemudian dipilih 7 atribut data yang digunakan pada penelitian ini, yaitu Sikap, Pengetahuan, Keterampilan, Keaktifan, *Rating Trainer*, *Projek Akhir* dan *Artikel*.

b. Data Transformation

Pada tahap ini, ditambahkan 1 atribut baru, yaitu lulus, yang merupakan label kelas yang berisi kategori "Tepat Waktu" dan "Terlambat", dibuat atau ditambahkan untuk memenuhi *Attribute/Feature Construction* yang diperlukan.

Kemudian di tahap ini dilakukan transformasi/perubahan nilai untuk 6 atribut yaitu Sikap, Pengetahuan, Keterampilan, Keaktifan, *Rating Trainer*, *Projek Akhir*, sedangkan untuk atribut *Artikel* tidak dilakukan transformasi data. Perubahan dibagi menjadi 2 yaitu tinggi dan rendah. Transformasi data dapat dilihat pada Tabel 1.

Tabel 1 Tranformasi Data

No	Nama Atribut	Nilai
1	Sikap	Tinggi (>77) Rendah (<77)
2	Pengetahuan	Tinggi (>77) Rendah (<77)
3	Keterampilan	Tinggi (>77) Rendah (<77)
4	Keaktifan	Tinggi (>77) Rendah (<77)
5	Rating Trainer	Tinggi (>77) Rendah (<77)
6	Projek Akhir	Tinggi (>77) Rendah (<77)

3.4 Modelling

Setelah melakukan tahapan *preprocessing*, kemudian untuk membuat model untuk memprediksi kelulusan peserta magang, pendekatan C4.5 dipilih sebagai metode. Beberapa contoh rule yang diperoleh sebagai berikut:

Keterampilan = rendah

| Sikap = tinggi

| | Pengetahuan = tinggi

| | | Keaktifan = rendah

| | | | Artikel = *Unpublished* : terlambat (2.0)

| | | | Artikel = *Published*

| | | | Rating Trainer = tinggi

| | | | | Projek Akhir = tinggi : tepat waktu (1.0)

| | | | | Projek Akhir = rendah : terlambat (1.0)

| | | | Rating Trainer = rendah : terlambat (1.0)

| | | Keaktifan = tinggi

Number of Leaves : 17

Size of tree : 22

Setiap atribut memiliki kelas setelah pohon terbentuk berdasarkan perkiraan entropi dan *gain* untuk setiap masing-masing atribut. Berikut ini merupakan contoh perhitungan dari pembentukan *tree*:

a. Perhitungan Entropy

$$\begin{aligned}
 \text{Entropy Total} &= (-185/225 * (\text{LOG}_2(185/225))) + (-40/225 * (\text{LOG}_2(40/225))) \\
 &= (-0.822222 * \text{LOG}_2(0.822222)) + (-0.177777 * \text{LOG}_2(0.177777)) \\
 &= (-0.822222 * -0.085010) + (-0.177777 * -0.750124) \\
 &= 0.069897 + 0.133354 \\
 &= 0.675191
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy Keterampilan} &= (-164/216 * \text{LOG}_2(164/216)) + (-52/216 * \text{LOG}_2(52/216)) \\
 &= (-0.759259 * \text{LOG}_2(0.759259)) + (-0.240740 * \text{LOG}_2(0.240740)) \\
 &= (-0.759259 * -0.119610) + (-0.240740 * -0.618451) \\
 &= 0.090814 + 0.148885 \\
 &= 0.796269
 \end{aligned}$$

b. Perhitungan *Gain*

$$\begin{aligned}\text{Gain Keterampilan} &= 0.675191 - ((29/225 * 0.929364) + (196/225 * 0.491237)) \\ &= 0.675191 - (0.119783 + 0.427921) \\ &= 0.675191 - 0.547704 \\ &= 0.127484\end{aligned}$$

c. Penentuan *Node*

Karena *gain* tertinggi adalah atribut Keterampilan, maka Keterampilan menjadi node dan referensi untuk entropi dan mendapatkan perhitungan yang mengikuti. Ulangi proses ini sampai semua atribut memiliki kelas masing-masing.

3.5 Tampilan Layar

a. Halaman *Import Dataset*

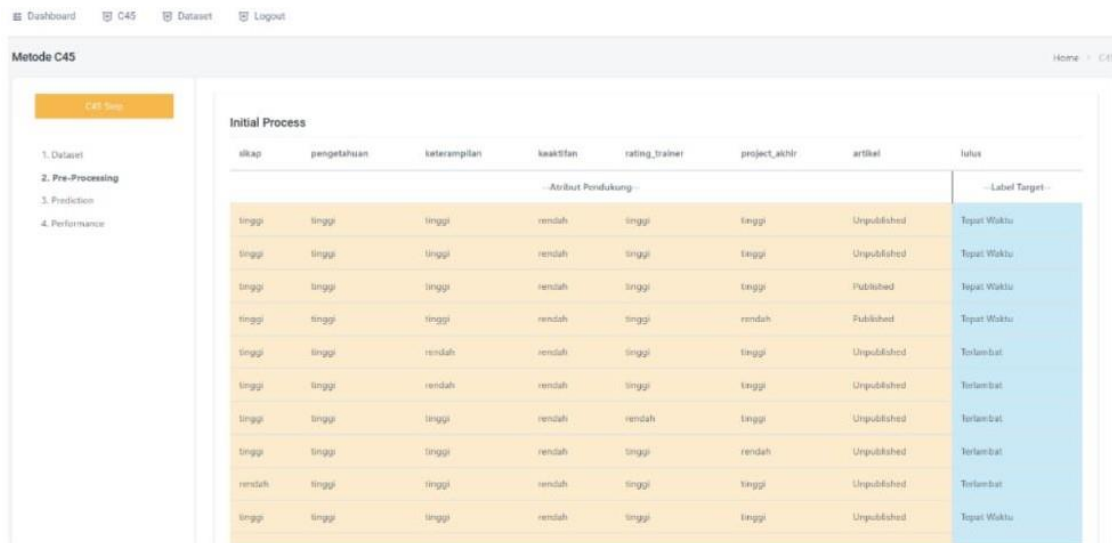
Untuk halaman *import dataset* ditunjukkan pada gambar 3. Halaman *import dataset* adalah halaman yang digunakan *user* untuk melakukan *import dataset*, dan data akan ditampilkan dalam bentuk tabel. Data yang dapat dilakukan *import* adalah data dengan tipe file ekstensi *.xls*.

sikap	pengetahuan	keterampilan	keaktifan	rating_trainer	project_akhir	artikel	kelas
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Published	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	rendah	Published	Tepat Waktu
tinggi	tinggi	rendah	rendah	tinggi	tinggi	Unpublished	Terlambat
tinggi	tinggi	rendah	rendah	tinggi	tinggi	Unpublished	Terlambat
tinggi	tinggi	tinggi	rendah	rendah	tinggi	Unpublished	Terlambat
tinggi	tinggi	tinggi	rendah	tinggi	rendah	Unpublished	Terlambat

Gambar 3. Halaman *Import Dataset*

b. Halaman *Preprocessing*

Untuk halaman *preprocessing* ditunjukkan pada gambar 4. Halaman *preprocessing* adalah halaman yang menunjukkan proses *preprocessing*, dimana data ditransformasi dan dibagi menjadi dua yaitu atribut pendukung dan data label sebagai target untuk melakukan proses prediksi.

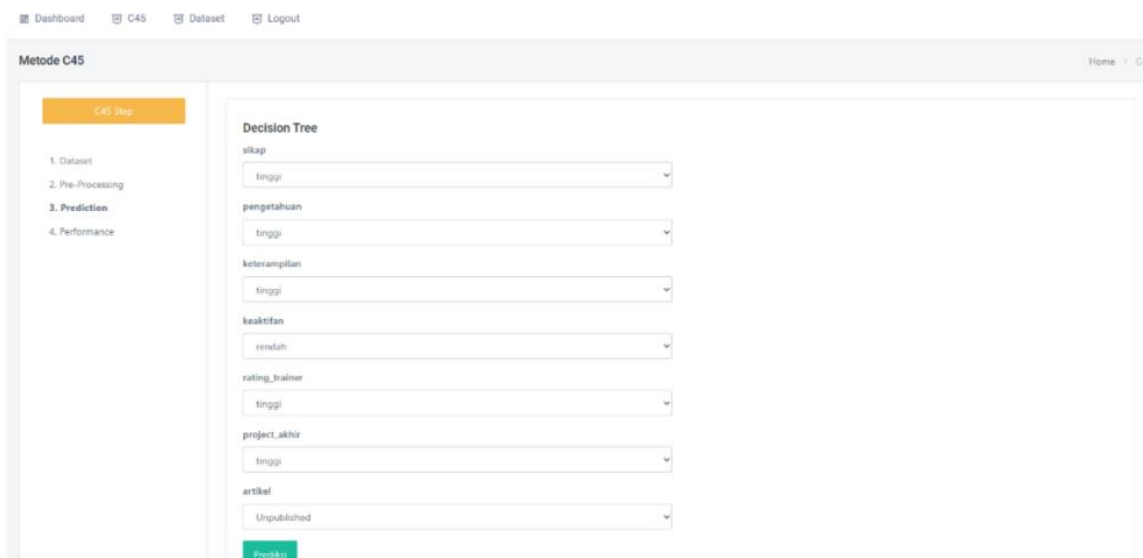


sikap	pengetahuan	keterampilan	keaktifan	rating_trainer	project_akhir	artikel	lulus
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Published	Tepat Waktu
tinggi	tinggi	tinggi	rendah	tinggi	rendah	Published	Tepat Waktu
tinggi	tinggi	rendah	rendah	tinggi	tinggi	Unpublished	Terlambat
tinggi	tinggi	rendah	rendah	tinggi	tinggi	Unpublished	Terlambat
tinggi	tinggi	tinggi	rendah	rendah	tinggi	Unpublished	Terlambat
tinggi	tinggi	tinggi	rendah	tinggi	rendah	Unpublished	Terlambat
rendah	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Terlambat
tinggi	tinggi	tinggi	rendah	tinggi	tinggi	Unpublished	Tepat Waktu

Gambar 4. Halaman *Preprocessing*

c. Halaman Prediksi

Untuk halaman prediksi ditunjukkan pada gambar 5. Halaman prediksi berupa *form input* yang harus diisi terlebih dahulu oleh *user* sebelum nantinya *user* menekan tombol hijau untuk melakukan prediksi.



Decision Tree
sikap: tinggi
pengetahuan: tinggi
keterampilan: tinggi
keaktifan: rendah
rating_trainer: tinggi
project_akhir: tinggi
artikel: Unpublished
Prediksi

Gambar 5. Halaman Prediksi

d. Hasil Prediksi

Untuk hasil prediksi ditunjukkan gambar 6. Pada halaman ini ditunjukkan hasil dari *form* prediksi setelah *user* menekantombol hijau. Diperlihatkan bentuk dari *tree* yang terbentuk dan hasil dari prediksi yang telah dilakukan. Pohon yang terbentuk adalah dari hasil file yang telah di *import* sebelumnya, dalam bagian akhir pohon terdapat hasil akhir keputusan dari prediksi bernilai tepat waktu atau terlambat.

sikap = tinggi
pengetahuan = tinggi
keaktifan = rendah
artikel = Unpublished : Terlambat (2.0)
artikel = Published
rating_trainer = tinggi
project_akhir = tinggi : Tepat Waktu (1.0)
project_akhir = rendah : Terlambat (1.0)
rating_trainer = rendah : Terlambat (1.0)
keaktifan = tinggi
rating_trainer = tinggi
project_akhir = tinggi : Tepat Waktu (5.0)
project_akhir = rendah
artikel = Unpublished : Terlambat (1.0)
artikel = Published : Tepat Waktu (1.0)
rating_trainer = rendah
project_akhir = tinggi
artikel = Unpublished : Terlambat (1.0)
artikel = Published : Tepat Waktu (1.0)
project_akhir = rendah : Terlambat (1.0)
pengetahuan = rendah
rating_trainer = tinggi
keaktifan = rendah : Terlambat (1.0)
keaktifan = tinggi
project_akhir = tinggi
artikel = Unpublished : Terlambat (1.0)
artikel = Published : Tepat Waktu (1.0)
project_akhir = rendah : Terlambat (1.0)
rating_trainer = rendah : Terlambat (2.0)
sikap = rendah
pengetahuan = tinggi
project_akhir = tinggi
keaktifan = rendah : Terlambat (1.0)
keaktifan = tinggi
rating_trainer = tinggi
artikel = Unpublished : Terlambat (1.0)
artikel = Published : Tepat Waktu (1.0)
rating_trainer = rendah : Terlambat (1.0)
project_akhir = rendah : Terlambat (2.0)
pengetahuan = rendah : Terlambat (2.0)

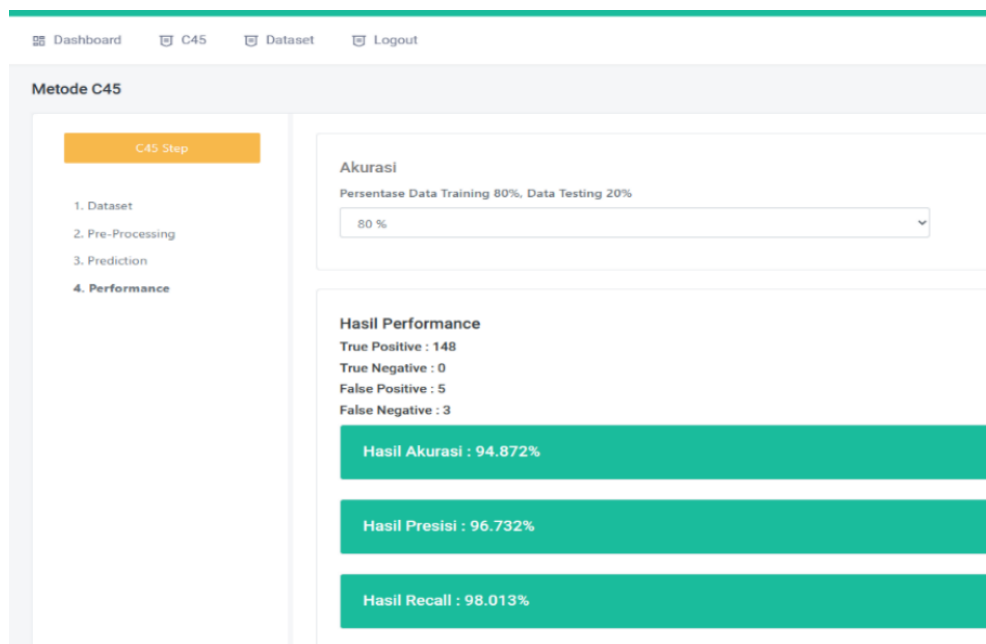
Hasil Prediksi

Tepat Waktu

Gambar 6. Hasil Prediksi

e. Halaman Hasil *Performance*

Untuk halaman hasil *performance* ditunjukan gambar 7. Pada halaman ini diperlihatkan hasil akhir dari *performance*, hasil yang tampil berupa nilai akurasi, presisi, dan *recall*. Hasil dari *performance* tergantung dari persentase dari *data training* dan *data testing* yang diinputkan oleh user.



Gambar 7. Halaman Hasil *Performance*

4. KESIMPULAN

Berdasarkan temuan penelitian, dapat disimpulkan bahwa model prediksi kelulusan tepat waktu dapat dibuat dengan menggunakan algoritma C4.5 *Decision Tree* dan 7 atribut prediktif, antara lain sikap, keterampilan, aktivitas, rating trainer, artikel, tugas akhir, dan lulus. Hasil akurasi sebesar 88,88%. Hasil berikut perlu dilakukan untuk pengembangan dan penerapan temuan penelitian ini: menjalankan pengujian dengan berbagai metode, menambahkan prosedur pembangkitan fitur sehingga nilai akurasi yang dihasilkan dapat ditingkatkan.

Untuk pengembangan dan implementasi hasil penelitian ini beberapa hal yang perlu dilakukan sebagai berikut: menguji dengan algoritma yang sama atau berbeda, dengan atribut yang berbeda dan jumlah atribut yang lebih sedikit dan jumlah data yang lebih banyak untuk prediksi kelulusan tepat waktu peserta magang dan dapat menghasilkan akurasi yang lebih tinggi dari penelitian ini, perlu dilakukan *training* terhadap *user* agar data yang dimasukan dapat terjaga validitasnya, perlu dilakukan *maintance* secara berkala untuk menghindari kerusakan pada sistem.

DAFTAR PUSTAKA

- [1] F. Gunadi and S. R. Widiyanto, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Perbandingan Data Warehouse Cloud Computing Menggunakan Konvensional Berbasis Kriptografi*, 2020.
- [2] A. Prasetyo *et al.*, "IMPLEMENTASI DATA MINING UNTUK ANALISIS DATA PENJUALAN DENGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS DAPOERIN'S) Program Studi Sistem Informasi [2]," vol. VIII, no. 2, 2020.
- [3] E. Tachi and N. & Andri, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," 2021.
- [4] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Faktor Exacta*, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [5] A. Sapitri1, N. Elisya2, N. M. Mustafa3, and M. Badrul4, "Makasar Jakarta Timur; (021) 28534471; 2Sistem Informasi; Universitas Nusa Mandiri," *Universitas Nusa Mandiri Jl. Jatiwaringin*, vol. 7, no. 1, p. 28534471, 2022.
- [6] A. Kristianus Lalo, P. Batarius, Y. Carmeneja, and H. Siki, "Implementasi Algoritma C4.5 Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia," 2021.
- [7] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," vol. 1, no. 2, pp. 63–69, 2019.
- [8] B. Ferdiansyah and L. goeirmanto, "Prediksi Loyalitas dalam Keterikatan Karyawan terhadap Perusahaan Menggunakan Algoritma C4.5* (Studi Kasus PT.XYZ)," vol. 8, no. 1, 2020.
- [9] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "PRISMA, Prosiding Seminar Nasional Matematika Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," vol. 3, pp. 64–71, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [10] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>