

PREDIKSI MAHASISWA NON-AKTIF MENGGUNAKAN ALGRITMA C4.5 (STUDI KASUS : UNIVERSITAS MUHAMMADIYAH SUKABUMI)

Leonie Siti Rahma, Asep Budiman Kusdinar, Prajoko

Teknik Informatika, Universitas Muhammadiyah Sukabumi

Jl. R. Syamsudin, S.H No.50,Cikole,Kec.Cikole, Kota Sukabumi,Jawa Barat 43113

leoniesr@ummi.ac.id

ABSTRAK

Dalam lingkungan perguruan tinggi di Indonesia, tingkat keberlanjutan studi mahasiswa menjadi perhatian penting, mengingat tingginya angka mahasiswa yang tidak melanjutkan studi. Untuk mengatasi permasalahan ini, machine learning khususnya menggunakan algoritma C4.5 digunakan untuk memprediksi potensi mahasiswa nonaktif. Dataset mahasiswa diambil dari Biro Kemahasiswaan dan Akademik Universitas Muhammadiyah Sukabumi, setelah melalui proses preprocessing, digunakan untuk melatih dan menguji model yang bertujuan untuk mengidentifikasi pola kritis yang dapat memprediksi keputusan mahasiswa untuk tidak melanjutkan studi. Algoritma C4.5 dipilih karena kemampuannya menghasilkan model yang mudah diinterpretasikan, sehingga memudahkan pengambil kebijakan dan staf akademik untuk mengambil tindakan proaktif. Evaluasi model dilakukan dengan menggunakan metrik akurasi yang menghasilkan 96.8%, presisi 98%, recall 94.2%, dan F1-score 96.5%, yang memberikan gambaran komprehensif tentang kinerja model dalam mengidentifikasi mahasiswa yang berpotensi tidak aktif. Penelitian ini memberikan kontribusi terhadap pengembangan sistem prediksi dan dapat menjadi dasar solusi keputusan berbasis data untuk meningkatkan tingkat keberlanjutan studi mahasiswa di perguruan tinggi di Indonesia.

Kata kunci : Algoritma C4.5, Prediksi, Mahasiswa Nonaktif

1. PENDAHULUAN

Pendidikan adalah landasan utama bagi masyarakat Indonesia yang mendambakan kemajuan dan kesejahteraan. Banyak orang berusaha mencapai pendidikan tinggi, seperti kuliah di perguruan tinggi. Perguruan tinggi menawarkan lingkungan belajar yang berbeda dengan jenjang sebelumnya, memberikan kebebasan dan semangat belajar. Keberhasilan mahasiswa di perguruan tinggi sangat dipengaruhi oleh motivasi dan keinginannya untuk meningkatkan pengetahuan dan keterampilan. Prestasi dan keterlibatan mahasiswa juga mempengaruhi reputasi dan keunggulan universitas. Sebaliknya, mahasiswa yang tidak aktif dan berkualitas rendah dapat menghambat kemampuan perguruan tinggi dalam menyediakan pendidikan bermutu.[1]

Menurut panduan akademik, mahasiswa berstatus tidak aktif merupakan mahasiswa yang tidak tercatat menjadi mahasiswa aktif dan belum memenuhi Kartu Rencana Studi (KRS) selama beberapa semester berturut-turut serta belum memperoleh izin resmi dari program studi. Mahasiswa yang berstatus tidak aktif tidak mempunyai hak-hak yang biasanya dimiliki oleh siswa aktif. Oleh karena itu, munculnya kondisi ambigu ini mampu diprediksi dengan melibatkan beberapa kemungkinan, seperti dikeluarkan, dipindahkan, mengundurkan diri, putus sekolah, meninggal, lulus, atau putus sekolah (DO).[2]

Langkah penting yang harus dilakukan untuk mencegah terjadinya siswa yang tidak aktif adalah dengan mengumpulkan data siswa yang tidak aktif dan memperkirakan kemungkinan terjadinya

ketidakaktifan selama pembelajaran. Melalui proses prediksi ini, kita dapat mengetahui pokok permasalahannya.[3]

Algoritma C4.5 merupakan salah satu algoritma pohon keputusan klasifikasi yang umum dan efektif. Algoritma ini beroperasi dengan membangun pohon keputusan berdasarkan nilai atribut dari data yang ada. C4.5 Memanfaatkan konsep entropi dan perolehan informasi untuk memilih atribut yang memberikan informasi terbaik untuk memisahkan data ke dalam kategori yang berbeda. Dalam kasus prediksi ketidakaktifan siswa, algoritma C4.5 akan menganalisis berbagai atribut siswa seperti prestasi akademik, kehadiran, dan faktor lainnya untuk memprediksi kemungkinan siswa menjadi tidak aktif. Algoritme ini terkenal dengan keakuratan dan kemampuannya dalam menangani data yang hilang serta atribut yang berkelanjutan, sehingga ideal untuk aplikasi ini.[4]

Oleh karena itu, menggunakan algoritma c4.5 memiliki potensi yang cukup bagus dan efektif dalam memprediksi mahasiswa nonaktif.

2. TINJAUAN PUSTAKA

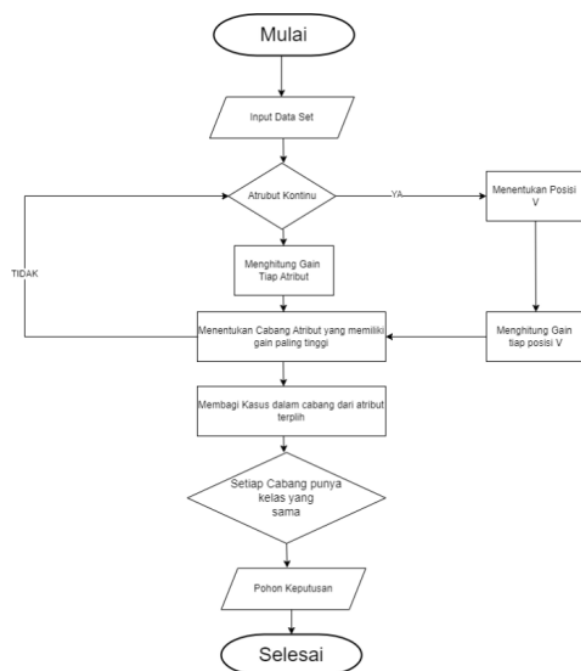
2.1. Machine Learning

Machine Learning (ML), cabang dari Artificial Intelligence (AI), untuk mengembangkan sistem komputer yang dapat belajar dan membuat keputusan otomatis berdasarkan data dan pengalaman, tanpa program eksplisit dari manusia Mencakup pembelajaran dan adaptasi, ekstraksi fitur otomatis, pengelolaan data besar, kemampuan generalisasi, interpretabilitas model, keamanan, etika, serta perkembangan algoritma baru. Selain itu,

memungkinkan sistem untuk mengidentifikasi pola atau tren baru, memperluas aplikasi ML dalam pengambilan keputusan. Dari banyaknya bagian atau subset *Machine Learning*, bagian berikut adalah yang paling sering dibahas antara nya yaitu *Supervised Learning*, *Unsupervised Learning* dan *Reinforcement Learning*. [5]

2.2. Algoritma C4.5

Metode algoritma C4.5 merupakan pendekatan yang sering digunakan dalam klasifikasi data yang melibatkan atribut numerik dan kategorikal. Hasil dari proses klasifikasi ini berupa aturan-aturan yang dapat digunakan untuk memprediksi nilai atribut pada data diskrit baru. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, dengan pembaruan dalam penyelesaian data yang hilang, kemampuan menangani data berkelanjutan, dan penggunaan pemangkasan untuk meningkatkan efisiensi model. Dengan kata lain, C4.5 memberikan solusi yang lebih kompleks untuk menangani varietas data yang lebih kompleks. [6]



Gambar 1. Flowcart Algoritma C4.5

Dalam induksi pohon keputusan, penting untuk menentukan kondisi pengujian pada node. Ada tiga kelompok penting dalam persyaratan pengujian node:

- Fitur Biner**
Sebuah fitur dengan hanya dua nilai berbeda, yang memerlukan pengujian dengan dua opsi cabang ketika sebuah node (root atau internal).
- Fitur Kategorial**
Fitur dengan nilai yang diklasifikasikan (nominal atau terurut), dapat memiliki beberapa nilai berbeda. Pengujian dapat berupa pemisahan biner atau pemisahan ganda.

c. Fitur Numerik

Untuk fitur numerik, kondisi pengujian node diwakili oleh pengujian perbandingan ($A < v$) atau ($A \geq v$) dengan hasil biner, atau pengujian ganda untuk hasil dengan beberapa nilai. Dalam kasus solusi biner, algoritma memeriksa semua kemungkinan posisi solusi v dan memilih posisi terbaik v . Untuk multi-mode, algoritma harus memeriksa semua kemungkinan rentang nilai kontinu. [6]

2.3. Mahasiswa Non-aktif

Mahasiswa yang berstatus Non-aktif adalah mereka yang tidak melakukan registrasi administrasi pada setiap awal semester ganjil. Perlu diketahui bahwa berdasarkan peraturan akademik, mahasiswa yang berstatus tidak aktif selama empat semester berturut-turut akan dianggap mahasiswa non aktif. [3]

2.4. Prediksi

Prediksi adalah suatu tindakan untuk meramalkan atau memperkirakan apa yang akan terjadi di masa depan berdasarkan analisis data historis atau teknik data. Dalam konteks pengambilan keputusan, prediksi membantu dalam memprediksi atau mengantisipasi kejadian di masa depan, sehingga memungkinkan pengambilan keputusan yang lebih tepat dan tepat sasaran. [7]

2.5. CRISP- DM

Metode CRISP-DM (Cross-Industry Standard Process for Data Mining) memberikan panduan langkah demi langkah dalam proses pengumpulan data. [8]

a. Business Understanding

Tahap ini memerlukan pemahaman konteks bisnis yang diteliti secara mendalam. Hal ini mencakup pemahaman tentang jenis bisnis, situasi bisnis saat ini, tujuan bisnis, serta kendala dan skenario yang relevan.

b. Data Understanding

Setelah memahami konteks bisnis, langkah selanjutnya adalah mengumpulkan data yang relevan. Data ini kemudian dianalisis untuk memahami karakteristiknya dan mengidentifikasi masalah atau kekurangan, seperti nilai yang hilang atau data yang tidak konsisten.

c. Data Preparation

Tahap awal persiapan data mencakup Identifikasi dan ekstraksi data yang paling relevan dengan tujuan penelitian, serta penghapusan kesalahan atau ketidakakuratan yang ada dalam data secara cermat. Selain itu, proses ini mencakup penyesuaian yang diperlukan terhadap format data agar selaras dengan persyaratan alat analisis yang dipilih. Pada akhirnya, tujuannya adalah untuk memastikan bahwa data sudah siap dan siap untuk analisis selanjutnya.

d. *Modeling*

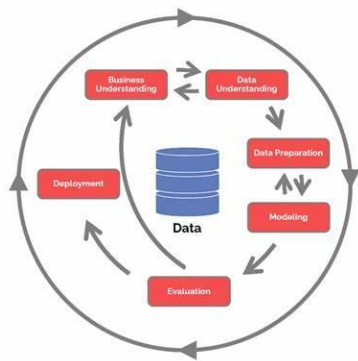
Pada fase ini berlangsung pemilihan dan penerapan teknik analisis data. Pilihan algoritma atau metode analisis data bergantung pada pertanyaan spesifik yang ingin dijawab dan atribut unik dari data yang diperiksa. Misalnya, kategorisasi data dapat dicapai melalui pemanfaatan algoritma klasifikasi.

e. *Evaluation*

Setelah penerapan model pada data, evaluasi menyeluruh dilakukan untuk memverifikasi kesesuaiannya dengan tujuan awal. Evaluasi ini mencakup analisis komprehensif tentang seberapa efektif model atau analisis tersebut menjawab pertanyaan penelitian dan mencapai tujuan bisnis yang telah ditentukan.

f. *Deployment*

Jika hasil evaluasi sesuai dengan harapan, langkah terakhir adalah mengimplementasikan hasil analisis. Ini bisa berupa penyajian temuan dalam laporan, pembuatan rekomendasi, atau penggunaan informasi untuk membuat keputusan strategis dalam konteks bisnis yang relevan.



Gambar 2. tahapan Metode CRISP-DM

2.6. Google Colaboratory

Google Colab, juga dikenal sebagai Google Colaboratory, adalah alat canggih untuk menyimpan, menulis, dan berbagi program melalui Google Drive. Fungsinya mirip dengan Jupyter Notebook berbasis cloud, yang dapat diakses melalui browser seperti Mozilla Firefox dan Google Chrome. Dengan Google Colab, pengguna dapat menjalankan kode Python tanpa perlu melakukan instalasi dan penyiapan tambahan, karena semua pengaturan dan penyesuaian ditangani di cloud.[9]

2.7. Flask

Flask merupakan pengembang yang dapat dengan mudah dan cepat membangun situs web, sekaligus memiliki kebebasan untuk menyesuaikan fungsionalitas untuk memenuhi kebutuhan spesifik proyek yang kita inginkan.[10]

3. METODE PENELITIAN

Pada tahap ini menggunakan metode *crisp-dm*. Dengan menggunakan metode *CRISP-DM* dalam penelitian ini, pendekatan sistematis dan efisien akan diambil untuk memastikan bahwa setiap langkah proses penelitian, mulai dari mengidentifikasi masalah hingga mengimplementasikan hasil, dilaksanakan secara terstruktur.

3.1. Business Understanding (Pemahaman Bisnis)

Tujuan utama penulis saat ini adalah memanfaatkan algoritma *c4.5* untuk mengidentifikasi siswa non-aktif. Pilihan ini didasarkan pada pengenalan pola perilaku bersama di antara berbagai kategori siswa non-aktif yang memerlukan deteksi.

3.2. Data Understanding (Pemahaman Data)

Pada tahap ini, penulis menggunakan data sekunder yang bersumber dari Biro Akademik dan Kemahasiswaan Universitas Muhammadiyah Sukabumi. Dataset ini mencakup informasi komprehensif mengenai pendaftaran siswa dari tahun 2017 hingga 2023, mencakup siswa aktif dan tidak aktif dalam tujuh tahun terakhir.

3.3. Data Preparation (Pengolahan Data Awal)

Pada tahap awal pengolahan data, dataset yang berisi siswa aktif dan tidak aktif menjalani persiapan untuk memudahkan proses klasifikasi. Variabel keputusan diidentifikasi, sedangkan informasi yang tidak relevan, seperti Nama dan Program Studi, dihilangkan.

a. Data Integrasi

Proses pengintegrasian data melibatkan penggabungan data akademik dan kemahasiswaan yang sebelumnya dipisahkan menjadi satu kumpulan data terpadu. Integrasi ini memungkinkan pemahaman siswa secara holistik, karena menggabungkan informasi akademik dan kemahasiswaan ke dalam satu sumber.

b. Seleksi Fitur (Atribut/Data Selection)

Proses pemilihan fitur, disebut juga pemilihan atribut atau data, melibatkan identifikasi variabel atau atribut tertentu yang akan digunakan dalam proses klasifikasi. Fitur-fitur pilihan ini berperan penting dalam pengambilan keputusan dan mencakup informasi seperti KTP, Nama, Jenis Kelamin, program studi, status pekerjaan, pekerjaan orang tua, pendapatan orang tua, IPK Semester 1 sd 4, SKS 1 sd 4, status skripsi, dan status akademis.

c. Data Cleansing

Selama tahap pembersihan data, data yang tidak lengkap dihilangkan. Proses pemurnian ini mencakup berbagai tugas seperti menghilangkan duplikat data, memastikan konsistensi data, dan memperbaiki kesalahan yang ada dalam kumpulan data.

3.4. Modelin

Pada proses pemodelan dengan algoritma C4.5 memerlukan pembuatan pohon keputusan menggunakan data pelatihan untuk membuat prediksi terhadap variabel keputusan, dan kemudian menilai kinerja model pada data pengujian untuk memastikan prediksi akurat pada data baru.

3.5. Evaluation

Dengan Python, pemanfaatan algoritma C4.5 untuk pemrosesan kumpulan data mencakup berbagai langkah, termasuk mengimpor data CSV, memilih fitur dan variabel keputusan yang relevan, mempartisi data ke dalam kumpulan pelatihan dan pengujian, melatih model, menilai akurasi, dan memvisualisasikan pohon keputusan. untuk mendapatkan wawasan tentang aturan yang mendasarinya.

3.6. Deployment

Untuk meningkatkan kemudahan penggunaan, peneliti telah menciptakan antarmuka berbasis web yang dapat memperkirakan siswa yang tidak aktif secara akurat. Selama proses pengembangan, para peneliti menggunakan flask sebagai editor teks pilihan mereka untuk melakukan pemodelan dan mengatur struktur serta logika program.

4. HASIL DAN PEMBAHASAN

Dalam tahap ini, kita akan mengeksplorasi pelaksanaan dan evaluasi upaya penelitian yang berfokus pada peningkatan ketepatan algoritma C4.5 dalam memprediksi hasil bagi siswa yang tidak terlibat secara aktif. Biro Akademik dan Kemahasiswaan Universitas Muhammadiyah Sukabumi menyediakan data mahasiswa yang diperlukan untuk penelitian ini, yang mencakup berbagai atribut. Dan juga akan disajikan penjelasan lengkap mengenai pemanfaatan algoritma C4.5 dalam memprediksi kuantitas mahasiswa tidak aktif di Universitas Muhammadiyah Sukabumi, meliputi tata cara dan luarannya.

4.1. Business Understanding

Dengan pemahaman yang mendalam tentang data mahasiswa non-aktif dan penerapan strategi preventif yang tepat, Universitas Muhammadiyah Sukabumi dapat meningkatkan kualitas pendidikan, mengurangi jumlah mahasiswa non-aktif, dan mendukung kesuksesan akademis serta profesional para mahasiswanya. Hal ini tidak hanya akan membantu institusi dalam menjaga tingkat kelulusan tepat waktu tetapi juga meningkatkan akreditasi dan mutu pendidikan secara keseluruhan.

4.2. Data Understanding

Data yang dikumpulkan mencakup beberapa variabel penting:

- NIK (Nomor Induk Mahasiswa) : Identifikasi unik untuk setiap mahasiswa.

- Nama Lengkap : Nama lengkap dari mahasiswa non-aktif.
- Jenis Kelamin : Informasi mengenai apakah mahasiswa tersebut pria atau wanita.
- Pekerjaan Orang Tua : Jenis pekerjaan dari orang tua mahasiswa, memberikan wawasan tentang latar belakang ekonomi keluarga.
- Prodi (Program Studi) : Jurusan yang diambil oleh mahasiswa selama kuliah.
- SKS (Satuan Kredit Semester) : Jumlah SKS yang telah ditempuh oleh mahasiswa.
- IPS (Indeks Prestasi Semester) : Nilai rata-rata prestasi akademis mahasiswa.
- Umur: Usia mahasiswa non-aktif.
- Kota Asal : Kota tempat mahasiswa berasal.
- Agama : Informasi mengenai agama mahasiswa.
- Asal Sekolah : Nama sekolah asal mahasiswa sebelum masuk perguruan tinggi.
- Jalur Masuk : Jalur masuk mahasiswa ke perguruan tinggi, seperti jalur ujian masuk atau jalur prestasi.
- Pekerjaan Ayah : Jenis pekerjaan atau profesi yang dijalani oleh ayah mahasiswa.
- Gaji Ayah : Pendapatan bulanan atau tahunan ayah, dikategorikan dalam beberapa tingkat pendapatan.
- Status : Status mahasiswa, seperti aktif atau non-aktif.

Data ini akan diolah menggunakan algoritma C4.5 untuk membangun pohon keputusan, yang membantu dalam klasifikasi atau prediksi berdasarkan aturan yang dihasilkan. Algoritma ini akan mengidentifikasi pola atau faktor yang mempengaruhi seorang mahasiswa menjadi non-aktif. Penelitian menggunakan data mahasiswa dari tahun akademik 2017 hingga 2023.

4.3. Data Preparation

4.3.1. Seleksi Fitur (Atribut/Data Selection)

Pada tahap ini adalah pemilihan atribut data yang akan digunakan untuk pemodelan

Tabel 1. seleksi atribut

Atribut		Detail Penggunaan
NIM	√	ID
Nama	√	NO
Prodi	√	NO
Umur	X	NO
JK	X	NO
Kota Asal	X	NO
Agama	X	NO
Pekerjaan Ayah	√	Nilai Model
Penghasilan Ayah	√	Nilai Model
Asal Sekolah	X	NO
Jalur Masuk	X	NO
IPS1	√	Nilai Model
IPS2	√	Nilai Model
IPS3	√	Nilai Model

Atribut		Detail Penggunaan
IPS4	✓	Nilai Model
IPS5	X	NO
IPS6	X	NO
IPS7	X	NO
IPS8	X	NO
IPS9	X	NO
SKS1	✓	Nilai Model
SKS2	✓	Nilai Model
SKS3	✓	Nilai Model
SKS4	✓	Nilai Model
Status	✓	Label Target

Tabel di atas memberikan informasi tentang atribut-atribut yang akan digunakan dalam penelitian. Tanda centang (✓) menunjukkan atribut yang akan digunakan, sedangkan tanda silang (X) menandakan atribut yang akan dieliminasi pada tahap persiapan data. Setelah pemilihan atribut, data tersebut akan dikonversikan agar memudahkan proses pemodelan.

4.3.2. Data Cleansing

Selama tahap ini, data yang tidak lengkap dihilangkan melalui proses pembersihan menyeluruh. Proses ini mencakup berbagai tugas seperti penghapusan data duplikat, verifikasi konsistensi data, dan perbaikan kesalahan yang ada dalam data.

```
data.isna().sum()
```

```

No      0
NIM     0
Nama    0
Program_Studi  0
Status  0
Penghasilan_ayah  0
pekerjaan_ayah  0
IPS_1   0
IPS_2   0
IPS_3   0
IPS_4   0
SKS_1   0
SKS_2   0
SKS_3   0
SKS_4   0
dtype: int64

```

Gambar 3. Mengecek Missing Value

```

duplicated = data[data.duplicated()]
len(duplicated)

```

```
0
```

Gambar 4. Mengecek Duplikasi Data

```
data = data.drop(columns=['Nama', 'No', 'NIM'])
```

```

Program_Studi  Status  Penghasilan_ayah  pekerjaan_ayah  IPS_1  IPS_2  IPS_3  IPS_4  SKS_1  SKS_2  SKS_3  SKS_4
0      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
1      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
2      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
3      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
4      1      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22

```

Gambar 5. Hasil Menghapus kolom yang tidak diperlukan

4.4. Modeling

Pada tahap ini akan dilakukan pengimplementasian algoritma C4.5 untuk pemodelan data set.

```
data = data.drop(columns=['No', 'NIM'])
```

Mengapus Model yang tidak diperlukan

```
ips_columns = ['IPS_1', 'IPS_2', 'IPS_3', 'IPS_4']
```

```

for col in ips_columns:
    data[col] = data[col].str.replace(',', '.').astype(float)

```

Mengubah IPS1-4 menjadi Plot

```

Program_Studi  Status  Penghasilan_ayah  pekerjaan_ayah  IPS_1  IPS_2  IPS_3  IPS_4  SKS_1  SKS_2  SKS_3  SKS_4
0      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
1      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
2      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
3      0      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22
4      1      0      0      4      7      1.9  4.0  2.0  3.4  19      21      19      22

```

Gambar 6. Hasil Plot

```

print("Perubahan pada kolom 'Status':")
print(status_mapping)

```

```

Perubahan pada kolom 'Status':
{'Aktif': 0, 'Non-Aktif': 1}

```

Gambar 7. Perubahan Kolom Status

```

clf = DecisionTreeClassifier(criterion='entropy',
                             random_state=42)
# Menggunakan entropy untuk mendekati C4.5
clf.fit(X_train, y_train)

```

```
y_pred = clf.predict(X_test)
```

```

# Definisikan model Decision Tree dengan kriteria 'entropy'
model = DecisionTreeClassifier(criterion='entropy', random_state=42)

```

```
#metrik
```

```

print("Accuracy:", accuracy_score(y_test, y_pred))
print("Presisi:", precision_score(y_test, y_pred))
print("Recall:",

```

```

recall score(y_test, y_pred))
print("Classification
Report:\n",
classification_report(y_test,
y_pred))

# Visualisasi pohon keputusan
plt.figure(figsize=(8, 10)) #
Sesuaikan ukuran gambar sesuai
kebutuhan
plot_tree(clf, filled=True,
feature_names=X_train.columns,
class_names=['Non Aktif',
'Aktif'], rounded=True)
plt.title("Pohon Keputusan")
plt.show()

```

Cuplikan kode ini menunjukkan pembuatan, pelatihan, dan evaluasi model pohon keputusan dengan Python menggunakan perpustakaan scikit-learn. Model dibangun dengan kriteria pemisahan berdasarkan entropi, mirip dengan algoritma C4.5. Untuk memulai, model pohon keputusan dibuat menggunakan `DecisionTreeClassifier`, yang menetapkan `criterion='entropy'` untuk menunjukkan penggunaan entropi sebagai kriteria pemisahan. Selain itu, `random\_state=42` diatur untuk memastikan hasil yang dapat direproduksi saat menjalankan model.

```

Accuracy: 0.9680851063829787
Presisi: 0.9878048780487805
Recall: 0.9418604651162791
Classification Report:

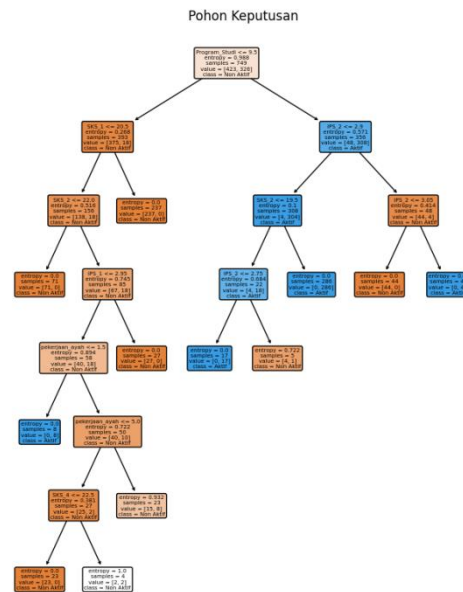
```

	precision	recall	f1-score	support
0	0.95	0.99	0.97	102
1	0.99	0.94	0.96	86
accuracy			0.97	188
macro avg	0.97	0.97	0.97	188
weighted avg	0.97	0.97	0.97	188

Gambar 8. Tabel Akurasi, Presisi, Recall, dan F1-Score

4.5. Evaluation

Mengukur kinerja model dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Memberikan gambaran seberapa baik model dapat memprediksi mahasiswa berpotensi non-aktif berdasarkan data yang ada. Penting untuk dicatat bahwa pengujian model dilakukan setelah data melalui tahap preprocessing agar model dapat memberikan hasil yang lebih akurat dan bermakna. Selain itu, proses ini membantu memahami kehandalan model dalam membuat prediksi yang relevan terkait mahasiswa yang berpotensi menjadi non-aktif.

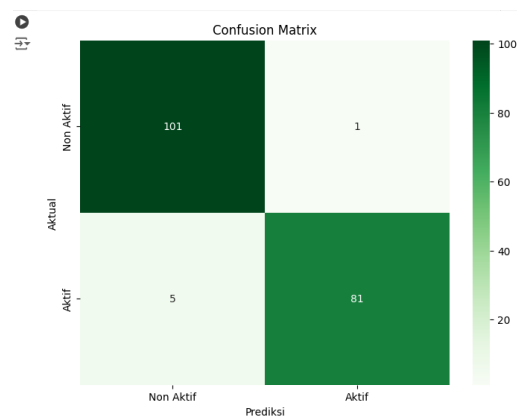


Gambar 9. Pohon keputusan

```

matrix=confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(matrix, annot=True, fmt='d',
cmap='Greens', xticklabels=['Non Aktif',
'Aktif'], yticklabels=['Non Aktif', 'Aktif'])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('Confusion Matrix')
plt.show()

```



Gambar 10. Confusion Matrix

- True Negatives (TN): 101 (Model memprediksi 101 sampel sebagai "Non Aktif" yang sebenarnya memang "Non Aktif". Ini menunjukkan prediksi negatif yang benar.)
- False Positives (FP): 1 (Model memprediksi 1 sampel sebagai "Aktif" padahal sebenarnya "Non Aktif". Ini menunjukkan prediksi positif yang salah.)
- False Negatives (FN): 5 (Model memprediksi 5 sampel sebagai "Non Aktif" padahal sebenarnya "Aktif". Ini menunjukkan prediksi negatif yang salah.)

- d. True Positives (TP): 81 (Model memprediksi 81 sampel sebagai "Aktif" yang sebenarnya memang "Aktif". Ini menunjukkan prediksi positif yang benar.)

Untuk menghitung nilai evaluasi pada bab 4.5 ini, kita bisa menggunakan beberapa metrik evaluasi yang umum digunakan dalam pemodelan prediksi. Metrik yang sering digunakan adalah akurasi, precision, recall, dan F1 score. Berikut adalah persamaan yang digunakan untuk menghitung masing-masing metrik tersebut:

a. Accuracy

Accuracy adalah proporsi dari total prediksi yang benar :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Precision

Precision adalah proporsi prediksi positif yang benar

$$Precision = \frac{TP}{TP + FP}$$

c. Recall

Recall adalah proporsi dari total positif yang terdeteksi dengan benar

$$Recall = \frac{TP}{TP + FN}$$

d. F1 score

F1 Score adalah rata-rata dari precision dan recall.

$$F1\ Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

Berikut adalah perhitungan hasil menggunakan Confusion Matrix :

a. Accuracy

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Akurasi = \frac{81 + 101 + 1 + 5}{81 + 101 + 1 + 5}$$

$$Akurasi = \frac{182}{188} = 0.968(96.8\%)$$

b. Precision

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{81}{81 + 1} = 0.988(98\%)$$

c. Recall

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{81}{81 + 5}$$

$$Recall = \frac{81}{86} = 0.942(94.2\%)$$

d. F1-Score

$$F1\ Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\ Score = 2x \frac{0.988 \times 0.942}{0.988 + 0.942} = 0.965(96.5\%)$$

4.6. Hasil Evaluasi

$$Akurasi = \frac{182}{188} = 0.968(96.8\%)$$

$$Precision = \frac{81}{81 + 1} = 0.988(98\%)$$

$$Recall = \frac{81}{86} = 0.942(94.2\%)$$

$$F1\ Score = 2x \frac{0.988 \times 0.942}{0.988 + 0.942} = 0.965(96.5\%)$$

Hasil evaluasi menunjukkan bahwa model pohon keputusan yang digunakan memiliki performa yang sangat baik dengan akurasi tinggi (96.8%), presisi yang sangat baik (98.8%), dan recall yang juga sangat baik (94.2%). Laporan klasifikasi menunjukkan bahwa model memiliki performa yang seimbang dan konsisten baik untuk kelas 0 maupun kelas 1, dengan nilai precision, recall, dan F1-score yang tinggi.

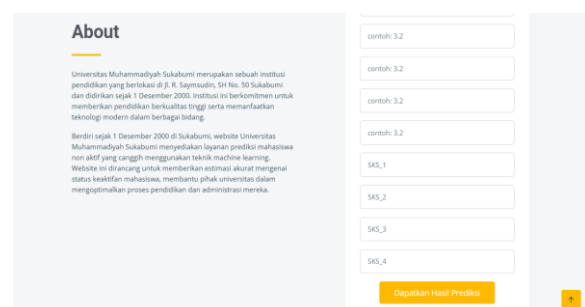
4.7. Deployment

Tahap deployment ini memastikan bahwa model machine learning tidak hanya berfungsi dengan baik secara teknis, tetapi juga memberikan nilai praktis bagi pengguna melalui prediksi yang akurat dan dapat diandalkan terkait status mahasiswa non-aktif.



Gambar 11. tampilan home

Halaman home ini menampilkan penerapan model prediksi yang terintegrasi di situs Web "Mahasiswa Aktif dan Non Aktif". Model ini digunakan untuk memprediksi hasil prediksi mahasiswa non aktif.



Gambar 12. tampilan about

4.8. Pengujian Fungsionalitas

Pengujian Fungsionalitas didefinisikan sebagai sebuah kegiatan untuk menguji apakah website yang dibangun sudah layak digunakan atau belum layak

digunakan. Adapun hasil pengujian Website Prediksi Mahasiswa Aktif dan Non Aktif yang telah dilakukan yaitu sebagai berikut:

Tabel 2. Pengujian Fungsional

Pengujian Halaman Home				
No	Pengujian Fungsi	Deskripsi	Output	Hasil Pengujian
1	Halaman Home	Peneliti melakukan <i>Klik</i> pada halaman home	Jika meng klik <i>home</i> dengan benar, maka peneliti akan beralih ke halaman Website Prediksi mahasiswa Aktif dan Non aktif	SUKSES
2	Halaman About	Peneliti melakukan <i>Klik</i> pada <i>Get Started</i>	Jika menekan form <i>Get strated</i> maka Website menampilkan halaman <i>About</i>	SUKSES
Pengujian Halaman Admin (Profile)				
1	Halaman About	Peneliti melakukan <i>Klik</i> pada <i>Get Started</i>	Jika mengisi form dari Program studi sampai sks 4 maka akan menghasilkan hasil dari prediksi	SUKSES

5. KESIMPULAN DAN SARAN

Integrasi algoritma C4.5 dalam website untuk memprediksi mahasiswa non-aktif telah dilakukan dengan optimal dan efisien. Algoritma ini mampu memberikan prediksi yang akurat mengenai status mahasiswa, sehingga dapat digunakan sebagai alat bantu dalam pengambilan keputusan terkait status keaktifan mahasiswa dan mahasiswa yang non-aktif. Dengan menerapkannya metode algoritma C4.5 pada sistem prediksi mahasiswa non – aktif ini menghasilkan nilai akurasi dari sistem prediksi dengan menggunakan confusion matrix sebesar 96.8%

DAFTAR PUSTAKA

- [1] D. Made Aryadi Mertha Sanjaya, A. A. Istri Ita Paramitha, and N. Widya Utami, "Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5: Studi Kasus STMIK Primakara," *J. Ilm. Ilmu Terap. Univ. Jambi*, vol. 6, no. 1, pp. 84–97, 2022.
- [2] P. I. Lestari and M. Andriansyah, "Analisis K-Nearest Neighbor Berdasarkan Forward Selection Untuk Prediksi Status Mahasiswa Non Aktif Pada STMIK Bani Saleh," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 181–186, 2021, [Online]. Available: <http://pddiktiadmin.kemdikbud.go.id/admin/kemahasiswaan/d>.
- [3] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5," *J. Inf. Teknol.*, vol. 2, pp. 23–29, 2020, doi: 10.37034/jidt.v2i1.22.
- [4] A. Salam, F. B. Nugroho, and J. Zeniarja, "Implementasi Algoritma K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Mahasiswa Non Aktif Universitas Dian Nuswantoro Semarang," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 69–76, 2020, doi: 10.33633/joins.v5i1.3351.
- [5] E. Retnoningsih and R. Pramudita, "Menenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [6] I. Irmayansyah and E. Lastrini, "Penerapan Metode Algoritma C4.5 Untuk Prediksi Mahaiswa Non Aktif," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 11, no. 2, pp. 95–106, 2021, doi: 10.36350/jbs.v11i2.119.
- [7] R. Habibi and A. Suryansah, *Aplikasi prediksi jumlah kebutuhan perusahaan*. Kreatif, 2020.
- [8] Y. Suhandi, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. dan Komput.*, vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.
- [9] A. H. Pratama, *BELAJAR MUDAH DAN SINGKAT MACHINE LEARNING: Panduan Praktis dengan Studi Kasus, Kode Program, dan Dataset*. Penerbit Andi, 2024.
- [10] E. Ghanbari and S. Najafzadeh, "Machine Learning," *Mach. Learn. Big Data Concepts, Algorithms, Tools Appl.*, pp. 155–207, 2020, doi: 10.1002/9781119654834.ch7.