

Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma *Decision Tree Classifier*

Abdul Rahim¹, Pareza Alam Jusia²

a3m.nix@gmail.com, parezaalam@unama.ac.id

¹Universitas Dinamika Bangsa

Informasi Artikel

Diterima : 15 Jan 2024

Direview : 18 Jan 2024

Disetujui : 20 Feb 2024

Kata Kunci

Prediksi, Prediksi Mahasiswa Non Aktif, Algoritma Decision Tree, model prediksi

Abstrak

Dengan pertumbuhan jumlah mahasiswa yang semakin dinamis, kebutuhan untuk menerapkan strategi preventif guna meningkatkan tingkat retensi mahasiswa menjadi semakin penting. Penelitian ini bertujuan untuk mengembangkan model yang dapat digunakan untuk mendeteksi mahasiswa yang berpotensi status akademiknya menjadi non-aktif menggunakan algoritma *Decision Tree Classifier* di lingkungan Universitas Dinamika Bangsa. Data yang digunakan dalam penelitian ini mencakup beragam variabel seperti data pribadi mahasiswa, nilai akademik dan informasi demografis lainnya. Proses pemodelan menggunakan *Decision Tree Classifier* dilakukan dengan memanfaatkan data historis mahasiswa untuk melatih model dalam mengklasifikasikan mahasiswa yang berpotensi non-aktif. Selanjutnya, model ini diuji coba pada data mahasiswa baru untuk menguji tingkat akurasi dan efektivitasnya. Hasil penelitian ini menunjukkan bahwa algoritma *Decision Tree Classifier* mampu memberikan kontribusi yang signifikan dalam prediksi mahasiswa yang berpotensi non-aktif dengan tingkat akurasi 95.63% dengan variabel yang paling berpengaruh adalah indeks prestasi semester 3, indeks prestasi semester 2 dan umur saat diterima.

Keywords

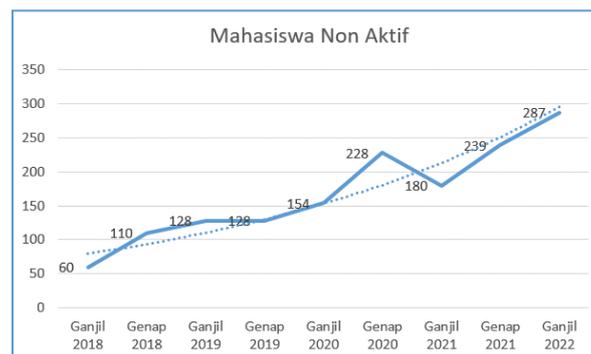
Prediction, Prediction of Inactive Students, Prediction Model

Abstrak

With the dynamic growth in the number of students, the need to implement preventive strategies to enhance student retention rates has become increasingly crucial. This research aims to develop a model for detecting students who are potentially at risk of becoming non-active academically, utilizing the Decision Tree Classifier algorithm within the environment of Universitas Dinamika Bangsa. The data employed in this study encompasses various variables, including students' personal information, academic grades, and other demographic details. The modeling process using the Decision Tree Classifier involves leveraging historical student data to train the model in classifying potentially non-active students. Subsequently, this model is tested on new student data to assess its accuracy and effectiveness. The findings of this research indicate that the Decision Tree Classifier algorithm significantly contributes to predicting potentially non-active students, achieving an accuracy rate of 95.63%. The most influential variables in this prediction include the Grade Point index for the third semester, grade point index for the second semester, and the age at admission.

A. Pendahuluan

Mahasiswa menjadi non aktif merupakan isu yang penting dan seringkali terjadi di berbagai perguruan tinggi, termasuk Universitas Dinamika Bangsa. Tingkat non aktif yang tinggi dapat mempengaruhi reputasi universitas dan juga menimbulkan dampak negatif bagi mahasiswa yang terkena dampaknya. Dalam era digital saat ini, perkembangan teknologi, khususnya dalam bidang prediksi data, mengalami kemajuan pesat. Salah satu inovasi yang semakin berkembang adalah teknologi prediksi menggunakan *machine learning* atau mesin pembelajaran. Penerapan metode prediksi dapat menjadi solusi alternatif dalam membantu mendeteksi mahasiswa yang berpotensi non aktif.



Gambar 1. Grafik Data Mahasiswa Non Aktif

Penerapan algoritma decision tree pada data akademik memungkinkan dapat dilakukannya prediksi terhadap mahasiswa yang berpotensi menjadi non aktif, sehingga universitas dapat memberikan intervensi yang sesuai dan tepat waktu.

Penelitian sebelumnya dilakukan oleh khafiizh hastuti yang berjudul analisis algoritma decision tree untuk prediksi mahasiswa non aktif. Penelitian menggunakan data akademik sebanyak 3.861 mahasiswa dan menggunakan decision tree dengan algoritma C4.5. Penelitian menghasilkan model dengan tingkat akurasi 95,29%[1].

Penelitian lain dilakukan oleh Dewa Made Aryadi Mertha Sanjaya, penelitian ini melakukan prediksi mahasiswa non aktif menggunakan algoritma C4.5. Penelitian ini menggunakan 558 data training dan 62 data testing dengan menggunakan aplikasi orange. Algoritma C4.5 dengan dataaset yang digunakan mendapat skor akurasi (CA) sebesar 92.5%, F1 Score sebesar 91.2%, precision sebesar 92.1%. recall sebesar 92.5%, specificity sebesar 47.3%, dan AUC (Area Under Curve) sebesar 88.4% atau dapat dikategorikan good classification. Penelitian ini juga memberikan hasil bahwa atribut yang paling mempengaruhi mahasiswa berpotensi non-aktif adalah mahasiswa dengan nilai IPK 0 – 1.99 atau 2.00 – 2.75, jenis tinggal mahasiswa adalah kontrakan/kos dan Jarak dengan variabel cukup dekat (6 – 10 KM), cukup jauh (11 – 15 KM), atau sangat jauh (> 20 KM) sedangkan atribut yang tidak terlalu mempengaruhi mahasiswa berpotensi non-aktif adalah atribut status pembayaran SPP[2].

Penelitian lain yang dilakukan oleh khaafizh hastuti melakukan komparasi algoritma untuk memprediksi status mahasiswa menjadi non aktif. penelitian ini dilakukan analisis komparasi empat algoritma klasifikasi data mining yaitu logistic regression, decision tree, naïve bayes dan neural network dengan menggunakan

3681 data set mahasiswa yang terdiri atas data demografi dan akademik mahasiswa sehingga dapat diketahui algoritma yang paling akurat untuk memprediksi mahasiswa non-aktif. Hasil penelitian ini yaitu *decision tree* merupakan algoritma yang paling akurat untuk melakukan prediksi mahasiswa non aktif dengan tingkat akurasi mencapai 95,29%[3].

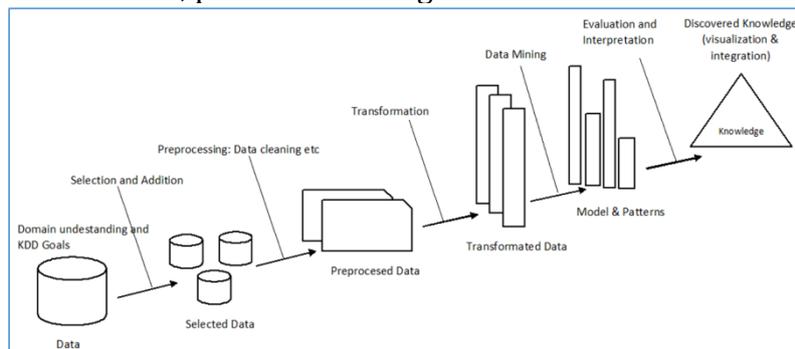
Penelitian yang penulis lakukan ini mencoba melakukan prediksi algoritma *decision tree* pada data akademik mahasiswa di universitas dinamika bangsa yang diusulkan sebagai alternatif metode dalam mengatasi masalah mahasiswa non aktif. Diharapkan hasil dari penelitian ini dapat menghasilkan model yang dapat digunakan untuk melakukan prediksi status akademik mahasiswa menjadi non aktif. Selain itu, hasil penelitian ini juga dapat memberikan kontribusi bagi pengembangan teknologi dan metode dalam bidang deteksi dini mahasiswa yang berpotensi non aktif.

Mahasiswa non aktif adalah mahasiswa yang tidak aktif mengikuti perkuliahan selama periode tertentu. Menurut Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 34 Tahun 2017 tentang Perubahan Atas Peraturan Menteri Pendidikan dan Kebudayaan Nomor 49 Tahun 2014 tentang Standar Nasional Pendidikan Tinggi, mahasiswa non aktif adalah mahasiswa yang telah terdaftar dan telah mengikuti perkuliahan pada tahun akademik berjalan, tetapi tidak mengikuti perkuliahan pada semester berikutnya.

Dalam buku peraturan akademik universitas dinamika bangsa, disebutkan bahwa “Seorang Mahasiswa dapat dinyatakan tidak aktif kuliah apabila tidak melakukan prosedur aktif (membayar uang kuliah, mengisi dan mengambil KRS sampai dengan batas waktu yang ditentukan) dan tidak mengajukan permohonan cuti kuliah”[4].

Data mining merupakan proses pengambilan informasi dari sekumpulan data yang besar. Data mining dapat digunakan untuk menggali informasi-informasi dari sekumpulan data berukuran besar yang kemudian dapat digunakan untuk pengambilan keputusan. [2]

Data mining, sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. [5]



Gambar 2. *Knowledge Discovering in Database* [6]

Berikut adalah penjelasan dari setiap tahapan pada KDD: [7]

- 1) Domain Understanding and KDD Goals.

Tujuan ditentukan dari sudut pandang user dan digunakan untuk mengembangkan dan pemahaman tentang domain aplikasi dan pengetahuan sebelumnya.

- 2) Selection and Additions
Tahap kedua berfokus pada penentuan data target dan subset dari data sampel atau variabel.
- 3) Preprocessing: Data Cleaning etc
Pembersihan dan preprocessing data merupakan operasi dasar untuk menyelesaikan data yang konsisten tanpa noisy.
- 4) Transformation
Transformasi data dari satu bentuk ke bentuk lainnya sehingga data diimplementasikan dengan mudah.
- 5) Data Mining (Chosing the Suitable Data Mining Task)
Memilih metode data mining yang sesuai berdasarkan tujuan tertentu yang telah didefinisikan pada tahap pertama, contoh dari metode data mining adalah classification, regression, clustering dan summarization.
- 6) Data Mining (Chosing the Suitable Data Mining Algorithm)
Memilih algoritma yang tepat untuk pencarian pola-pola data, algoritma yang dipilih berdasarkan kecocokan kriteria dengan metode data mining.
- 7) Data Mining (Imploying Data Mining Algorithm)
Pada tahap ini algoritma yang telah dipilih diimplementasikan.
- 8) Evaluation and Interpretation
Tahap ini berfokus pada interpretasi dan evaluasi yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan hipotesa yang ada sebelumnya.
- 9) Discovered Knowledge
Penggunaan pengetahuan yang ditemukan dari proses KDD, dimana memutuskan apa yang akan dilakukan dengan pengetahuan dihasilkan.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat di perkecil. [8]

Decision Tree atau Pohon keputusan adalah konsep mengubah data menjadi aturan - aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi simple, sehingga pengambilan keputusan akan lebih memudahkan solusi dari sebuah permasalahan[9].

Menurut scikit-learn.org, *Decision Tree Classifier* (DTC) merupakan algoritma yang membentuk sebuah pohon keputusan (decision tree). Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan[10].

Ada beberapa algoritma dalam membuat pohon keputusan yaitu [10]:

1. ID3 (Iterative Dichotomiser 3)
Dikembangkan pada tahun 1986 oleh Ross Quinlan. Membangun pohon multiarah (multiway tree), di mana setiap node membelah data menjadi

beberapa cabang berdasarkan fitur kategorikal yang menghasilkan perolehan informasi (information gain) terbesar. Pohon dibangun hingga ukuran maksimumnya, lalu dilakukan pemangkasan (pruning) untuk meningkatkan kemampuan generalisasi pada data baru.

2. C4.5

Merupakan penerus algoritma ID3. Menghilangkan batasan pada fitur kategorikal dengan mendefinisikan atribut diskrit secara dinamis dari variabel numerik. Mengubah pohon hasil pelatihan menjadi set aturan if-then. Melakukan evaluasi akurasi setiap aturan untuk menentukan urutan penerapannya. Melakukan pemangkasan dengan menghapus prasyarat aturan jika akurasinya meningkat tanpa prasyarat tersebut.

3. C5.0

Versi terbaru yang dikembangkan oleh Quinlan, dirilis dengan lisensi berbayar. Menggunakan lebih sedikit memori dan menghasilkan set aturan yang lebih kecil dibandingkan C4.5. Memiliki tingkat akurasi yang lebih tinggi dibandingkan C4.5.

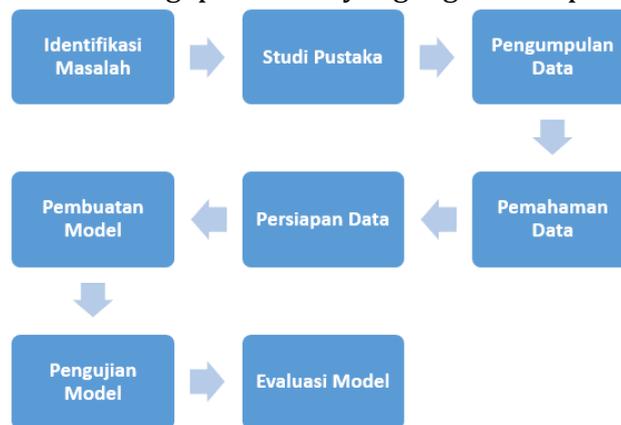
4. CART (Classification and Regression Trees)

Mirip dengan C4.5, tetapi mendukung variabel target numerik (untuk regresi) dan tidak menghasilkan set aturan. Membangun pohon biner (binary tree) menggunakan fitur dan ambang batas (threshold) yang menghasilkan perolehan informasi terbesar pada setiap node.

Penelitian ini menggunakan algoritma CART (Classification and Regression Trees) untuk membangun model pohon keputusan.

B. Metode Penelitian

Berikut adalah metodologi penelitian yang digunakan pada penelitian ini:



Gambar 3. Metodologi Penelitian

Adapun tahapan dalam penelitian ini dapat dijelaskan sebagai berikut:

1. Identifikasi Masalah: Penelitian dimulai dengan mengidentifikasi permasalahan yang akan menjadi fokus utama.
2. Studi Pustaka: Dilakukan kajian literatur terkait algoritma machine learning untuk klasifikasi dan prediksi data guna memahami konsep dan metode yang telah ada.

3. Pemahaman Data: Setelah studi pustaka, dilakukan analisis terhadap data akademik universitas dinamika bangsa. Tahap ini menghasilkan atribut-atribut yang akan digunakan pada langkah-langkah berikutnya.
4. Persiapan Data: Melibatkan proses cleaning data, termasuk penanganan nilai null, missing value, dan kategorisasi data. Data dibagi menjadi set pelatihan dan pengujian.
5. Pelatihan Model: Membuat model menggunakan algoritma decision tree classifier berdasarkan temuan dari studi pustaka. Model ini akan dilatih menggunakan data pelatihan.
6. Pengujian Model: Model yang telah dilatih diuji dengan menggunakan data pengujian, dan hasilnya dievaluasi melalui nilai akurasi serta metrik lainnya.
7. Evaluasi Model: Tahap akhir melibatkan evaluasi keseluruhan performa model, termasuk identifikasi kelebihan dan kekurangan. Potensi perbaikan dan pengembangan model juga dapat dijelaskan.

Tabel 1. Data Mahasiswa Non Aktif Berdasarkan Semester

Semester Ke	Jumlah Non Aktif
1	39
2	341
3	308
4	349
5	250
6	188
7	121

Data pada tabel 1 merupakan data jumlah mahasiswa non aktif berdasarkan semester yang sedang ditempuh mahasiswa. Berdasarkan data di tabel 1 dapat dilihat bahwa kasus mahasiswa menjadi non aktif banyak terjadi pada semester ke 2, 3 dan semester 4. Dari data ini dapat dikatakan bahwa model yang akan dibuat harus dapat melakukan prediksi status akademik mahasiswa di semester 4 dengan menggunakan data semester 1 sampai semester 3.

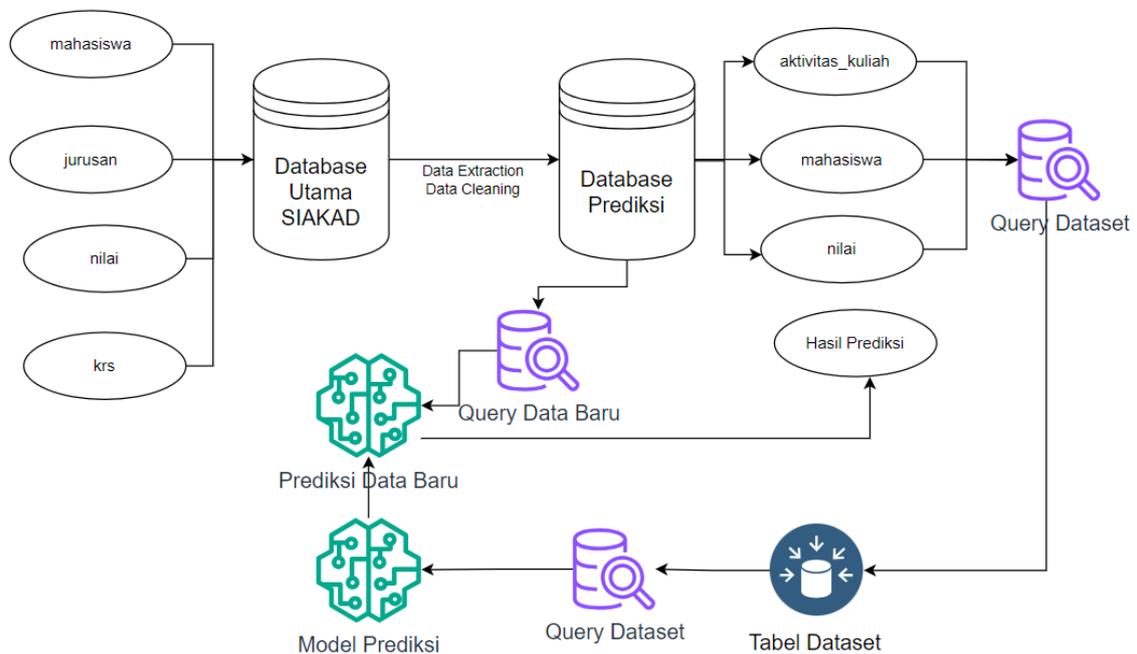
Setelah melakukan analisis jumlah mahasiswa non aktif berdasarkan semester berjalan mahasiswa. Tahapan selanjutnya adalah melakukan analisis dan eksperimen atribut yang akan digunakan untuk pelatihan model. Penelitian ini menggunakan atribut sebagai berikut:

Tabel 2. Atribut Model

Atribut	Value
Program Studi	TI,SI,SK
Jenis Kelamin	L, P
Umur Saat Mendaftar	<=20, <=25, <=30, >30
Kota Asal	Jambi, Luar Jambi
Sumber Biaya	Mandiri, Beasiswa, orang tua, lainnya
Pekerjaan Orang Tua	Wiraswasta, TNI/POLRI, PNS, Pegawai Swasta, Pensiunan, lainnya
Status marital	Menikah, Lajang
Status Sekolah Asal	Negeri, Swasta
Status Kerja	Sudah Kerja, Belum Kerja
Indeks Prestasi Semester 1	Kurang, Cukup, Memuaskan, Sangat Memuaskan, Dengan

	Pujian
Indeks Prestasi Semester 2	Kurang, Cukup, Memuaskan, Sangat Memuaskan, Dengan Pujian
Indeks Prestasi Semester 3	Kurang, Cukup, Memuaskan, Sangat Memuaskan, Dengan Pujian
Status Akademik Semester 4	Aktif, Non-Aktif

Pada penelitian ini jenis dataset yang akan digunakan langsung dari database MySQL dengan sumber data yang akan digunakan adalah database sistem informasi akademik universitas dinamika bangsa. Data akademik akan di duplikat ke database baru menggunakan bahasa pemrograman python lalu berdasarkan data yang telah di duplikat maka akan dibuat dataset sehingga proses pembuatan model tidak menggantu database utama.



Gambar 4. Arsitektur dataset

Pada gambar 4 digambarkan bahwa dataset yang akan digunakan untuk melatih dan menguji model terkoneksi langsung ke database MySQL. Arsitektur ini dibuat dengan tujuan melakukan ekstraksi data dari database sistem informasi akademik, hasil dari proses ini adalah didapatnya informasi aktivitas kuliah setiap mahasiswa untuk setiap semester. Data pada database prediksi kemudian dapat digunakan untuk melatih dan menguji model. Setelah model memiliki kemampuan untuk melakukan prediksi maka sistem dapat mengambil data baru kemudian dilakukan prediksi menggunakan model yang sudah ada dan hasilnya di simpan di tabel hasil prediksi.

C. Hasil dan Pembahasan

Data yang digunakan di penelitian ini adalah data mahasiswa dari tahun 2017 sampai tahun 2022. Jumlah data yang digunakan sebanyak 3773, data ini merupakan hasil ekstraksi dari data awal.

Tabel 3. Sampel Data Mahasiswa

No	Prodi	Kelamin	Umur	Kota Asal	Sumber Biaya	Pekerjaan Ortu	Status Marital	Sekolah Asal	IPS1	IPS2	IPS3	Status Semester 4
1	SI	L	20	Jambi	Mandiri	Swasta	Menikah	Swasta	3	2	2	Aktif
2	TI	P	21	Luar Jamb	Orang Tua	Negeri	Belum	Negeri	2	3	3	Non Aktif
3	TI	P	21	Luar Jamb	Orang Tua	Negeri	Belum	Negeri	2	3	3	Non Aktif

Proses selanjutnya data akan di *cleaning* dan di transformasi. Data umur akan dikelompokkan ≤ 20 , ≤ 25 , ≤ 30 , > 30 , data IPS atau indeks prestasi semester dikelompokkan sesuai aturan akademik seperti berikut: Kurang, Cukup, Memuaskan, Sangat Memuaskan, Dengan Pujian.

Tabel 4. Kategori Nilai IPS

No	Nilai	Kategori
1	≥ 3.5	Dengan Pujian
2	$\geq 3.0 \ \&\& \ \leq 3.4$	Sangat Memuaskan
3	$\geq 2.6 \ \&\& \ \leq 2.9$	Memuaskan
4	$\geq 2.0 \ \&\& \ \leq 2.5$	Cukup
5	≤ 1.9	Kurang

Penelitian ini menggunakan tools bantu bahasa pemrograman python, dengan library scikit-learn untuk melatih dan menguji model menggunakan algoritma *decision tree classifier*. Library yang digunakan pada penelitian ini sebagai berikut:

1. Scikit-Learn digunakan untuk melatih dan menguji model algoritma decision tree
2. Mysql.connector digunakan untuk menghubungkan python dengan database mysql
3. Pandas digunakan untuk mengelola dan menganalisis data

Tahapan berikutnya adalah merancang kode program untuk melatih dan mengujimodel. Kode program dapat dibagi menjadi seperti berikut:

1. Load / Import library
2. Buat Koneksi ke database MySQL
3. Query Data Untuk Training dan Testing
4. Kategorikal data dan data *Cleaning*
5. Pisahkan fitur dan target
6. Pisahkan data traning dan data testing
7. Buat model dan lakukan training
8. Lakukan prediksi dengan data testing
9. Evaluasi model dengan akurasi, laporan klasifikasi dan confusion matrix

Kode yang digunakan pada penelitian ini seperti berikut:

```

from sklearn.tree import DecisionTreeClassifier

from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score, classification_report

import mysql.connector

import pandas as pd

from sklearn.metrics import confusion_matrix

```

```
def db_connect():
    config = {
        'host': 'localhost',
        'user': 'root',
        'password': 'password_db',
        'database': 'db_predict'
    }

    try:
        conn = mysql.connector.connect(**config)
        if conn.is_connected():
            print("Terhubung ke MySQL!")
            return conn
    except mysql.connector.Error as err:
        print(f"Kesalahan MySQL: {err}")
        return None

# buat koneksi ke database
conn = db_connect()

query = "select nama_jurusan, jenis_kelamin, umur_saat_mendaftar, kota_asal,pekerjaan_orang_tua,
status_marital, status_sekolah_asal, status_kerja, ips1, ips2, ips3, a.status_akademik as
status_akademik from mahasiswa_trains2 a, aktivitas_kuliahs b where a.nim=b.nim && b.semester_ke
<= 3 && a.status_akademik in ('Aktif', 'Non-Aktif') AND a.nama_jurusan IN ('si', 'ti','sk') group by
a.nim"

# baca data menggunakan pandas
df = pd.read_sql(query, conn)

# tampilkan total data yang diambil
print(f"Total data: {df.shape[0]}")

# tampilkan jumlah data berdasarkan status akademik
print(df['status_akademik'].value_counts())

# kategorikal umur
bins = [0, 20, 25, 30, float('inf')]
labels = ['<=20', '<=25', '<=30', '>30']
df['umur_group'] = pd.cut(df['umur_saat_mendaftar'],
                           bins=bins, labels=labels, right=False)
X = df.drop(['umur_saat_mendaftar'], axis=1)
```

```
# kategorikal ips
df['ips1'] = pd.to_numeric(df['ips1'], errors='coerce')
df['ips2'] = pd.to_numeric(df['ips2'], errors='coerce')
df['ips3'] = pd.to_numeric(df['ips3'], errors='coerce')
ips_bins = [0, 2.0, 2.76, 3.01, 3.5, float('inf')]
ips_labels = ['Kurang', 'Cukup', 'Memuaskan',
              'Sangat Memuaskan', 'Dengan Pujian']

# Apply categorization to ips columns
df['ips1_category'] = pd.cut(
    df['ips1'], bins=ips_bins, labels=ips_labels, right=False)
df['ips2_category'] = pd.cut(
    df['ips2'], bins=ips_bins, labels=ips_labels, right=False)
df['ips3_category'] = pd.cut(
    df['ips3'], bins=ips_bins, labels=ips_labels, right=False)

# Drop ips1, ips2, ips3 karena sudah diganti dengan label baru
df.drop(['ips1', 'ips2', 'ips3'], axis=1, inplace=True)

# Memisahkan fitur dan target
# Drop umur_saat_mendaftar karena sudah dielompokkan
X = df.drop(['status_akademik'], axis=1)
y = df['status_akademik']

# Melakukan encoding pada fitur kategorikal
X_encoded = pd.get_dummies(X, drop_first=True)

# Memisahkan data menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(
    X_encoded, y, test_size=0.2, random_state=42)

# Buat model Decision Tree
dt_model = DecisionTreeClassifier(random_state=42)

# Latih model pada data pelatihan
dt_model.fit(X_train, y_train)

# Lakukan prediksi pada data pengujian
y_pred = dt_model.predict(X_test)
```

```
# Dapatkan struktur pohon keputusan
tree_structure = dt_model.tree_

# Hitung jumlah cabang dalam pohon
num_nodes = tree_structure.node_count

# Tampilkan jumlah cabang
print(f"Jumlah Cabang dalam Pohon Keputusan: {num_nodes}")

# Evaluasi performa model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)

print(f'Accuracy: {accuracy}')
print('Classification Report:\n', classification_rep)
print('Confusion Matrix:\n', conf_matrix)
```

Pada penelitian ini dilakukan 2 eksperimen dataset, pada percobaan pertama, pelatihan dan pengujian model mendapat tingkat akurasi 0.944 dengan laporan klasifikasi sebagai berikut:

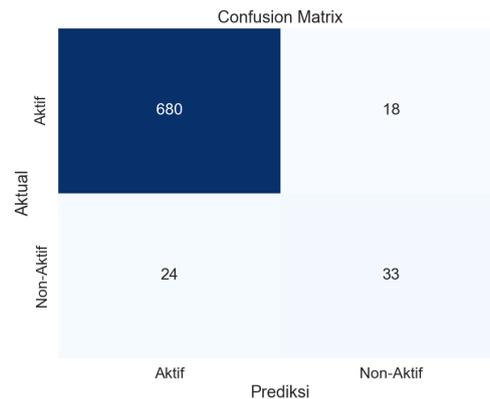
```
Accuracy: 0.9443708609271523
Classification Report:
              precision    recall  f1-score   support

   Aktif         0.97         0.97         0.97         698
  Non-Aktif      0.65         0.58         0.61          57

 accuracy                0.94         755
 macro avg              0.81         0.78         0.79         755
 weighted avg           0.94         0.94         0.94         755
```

Gambar 5. Hasil Percobaan 1 Klasifikasi Data

Dalam penelitian ini, dilakukan pengujian dengan menggunakan metode klasifikasi *decision tree* untuk mengidentifikasi status akademik mahasiswa. Hasil pengujian menunjukkan tingkat akurasi sebesar 94.44%, yang menandakan tingkat ketepatan model dalam memprediksi kategori status akademik mahasiswa. Pada kategori status akademik Aktif, model menunjukkan tingkat presisi dan recall sebesar 97%, serta nilai f1-score sebesar 97%. Sementara itu, kategori status akademik Non-Aktif memiliki tingkat presisi sebesar 65%, recall sebesar 58%, dan f1-score sebesar 61%.



Gambar 6. Confusion Matrix 1

Berdasarkan gambar confusion matrix dapat dijelaskan sebagai berikut:

1. True Positive (TP) 680 merupakan jumlah instance mahasiswa aktif yang diklasifikasikan dengan benar oleh model sebagai data mahasiswa aktif.
2. False Negative (FN) 18 adalah jumlah instance mahasiswa aktif yang sebenarnya ada, tetapi model salah mengklasifikasikannya sebagai data mahasiswa non aktif.
3. False Positive (FP) 24 adalah jumlah instance yang sebenarnya non aktif, tetapi model salah mengklasifikasikannya sebagai data mahasiswa aktif.
4. True Negative (TN) 33 adalah jumlah instance mahasiswa non aktif yang diklasifikasikan dengan benar oleh model sebagai mahasiswa non aktif.

Berdasarkan data pada confusion matrix dapat dijelaskan bahwa model hanya mampu melakukan prediksi mahasiswa non-aktif dengan benar sebanyak 33 data.

	Feature	Importance
24	ips3_category_Memuaskan	0.212328
25	ips3_category_Sangat Memuaskan	0.125898
23	ips3_category_Cukup	0.121262
26	ips3_category_Dengan Pujian	0.076045
0	umur_saat_mendaftar	0.068622
19	ips2_category_Cukup	0.049203
3	jenis_kelamin_P	0.040029
9	pekerjaan_orang_tua_wiraswasta	0.032299
5	pekerjaan_orang_tua_pegawai swasta	0.027702
2	nama_jurusan_ti	0.025839
7	pekerjaan_orang_tua_pns	0.025775
1	nama_jurusan_sk	0.020327
17	ips1_category_Sangat Memuaskan	0.019411
12	umur_group_<=25	0.019251
11	status_sekolah_asal_swasta	0.018066
4	kota_asal_luar_jambi	0.017279
22	ips2_category_Dengan Pujian	0.014030
18	ips1_category_Dengan Pujian	0.013692
6	pekerjaan_orang_tua_pensiun/purnawirawan	0.013453
15	ips1_category_Cukup	0.012364
20	ips2_category_Memuaskan	0.012249
8	pekerjaan_orang_tua_tni/polri	0.011881
16	ips1_category_Memuaskan	0.011663
21	ips2_category_Sangat Memuaskan	0.011332
14	umur_group_>30	0.000000
10	status_marital_menikah	0.000000
13	umur_group_<=30	0.000000

Gambar 7. Feature Importance

Berdasarkan gambar 7, fitur atau atribut yang paling berpengaruh terhadap prediksi model adalah indeks prestasi semester 3 dengan predikat memuaskan

dengan nilai 0.212 sedangkan umur ≤ 30 tidak memiliki pengaruh terhadap keputusan model.

Pada eksperimen kedua, penulis menambahkan nilai matakuliah semester 3 sebanyak 9 matakuliah. Data nilai berisi grade nilai mulai dari A, B+, B, C+, C, D+, D, E. Berikut hasil pengujian dengan tambahan atribut nilai semester 3:

```

Accuracy: 0.9562913907284768
Classification Report:

```

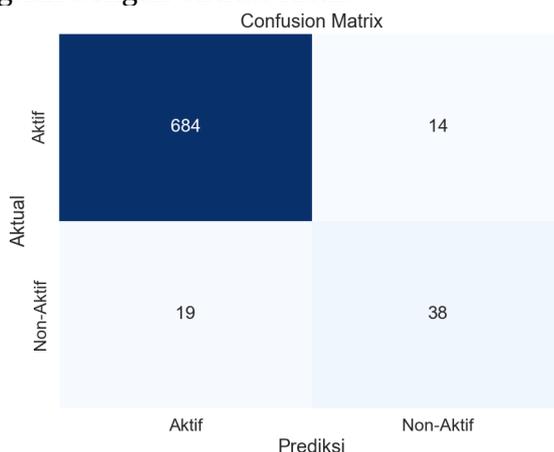
	precision	recall	f1-score	support
Aktif	0.97	0.98	0.98	698
Non-Aktif	0.73	0.67	0.70	57
accuracy			0.96	755
macro avg	0.85	0.82	0.84	755
weighted avg	0.95	0.96	0.96	755

Gambar 8. Hasil Percobaan ke 2 Klasifikasi Data

Pada eksperimen kedua menunjukkan peningkatan kinerja model dibandingkan dengan eksperimen pertama. Dalam eksperimen ini, tingkat akurasi model meningkat menjadi 95.63%, menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengklasifikasikan entitas berdasarkan status akademiknya.

Peningkatan tersebut terutama terlihat pada status akademik Aktif, di mana presisi, recall, dan f1-score semuanya mengalami peningkatan menjadi 97%, 98%, dan 98% masing-masing. Hal ini menunjukkan bahwa model lebih efektif dalam mengidentifikasi dan memprediksi entitas dengan status akademik Aktif.

Terdapat penurunan performa pada kelas Non-Aktif dibandingkan dengan eksperimen pertama, hasil tersebut masih menunjukkan tingkat presisi, recall, dan f1-score yang relatif tinggi sebesar 73%, 67%, dan 70% secara berturut-turut. Ini menandakan bahwa model tetap mampu mengklasifikasikan dengan baik entitas dengan status Non-Aktif, meskipun dengan tingkat ketidakpastian yang sedikit lebih tinggi dibandingkan dengan entitas Aktif.



Gambar 9. Confusion Matrix 2

Berikut adalah hasil dari percobaan kedua:

1. True Positive (TP): Pada eksperimen kedua, jumlah entitas yang benar-benar terklasifikasi sebagai mahasiswa aktif mengalami peningkatan dari 680 (eksperimen pertama) menjadi 684.

2. False Positive (FP): Terdapat penurunan pada jumlah entitas yang seharusnya diklasifikasikan sebagai mahasiswa non-aktif, tetapi salah diklasifikasikan sebagai positif, dari 18 (eksperimen pertama) menjadi 14.
3. False Negative (FN): Jumlah entitas yang seharusnya diklasifikasikan sebagai mahasiswa aktif, tetapi salah diklasifikasikan sebagai mahasiswa non-aktif, mengalami penurunan dari 24 (eksperimen pertama) menjadi 19.
4. True Negative (TN): Jumlah entitas yang benar-benar terklasifikasi sebagai mahasiswa non-aktif mengalami kenaikan dari 33 (eksperimen pertama) menjadi 38.

D. Simpulan

Pemanfaatan database sebagai sumber data untuk menyusun dataset memiliki keunggulan dalam memberikan representasi yang lebih akurat dan kontekstual terkait perilaku mahasiswa. Dalam konteks ini, penggunaan dataset dari MySQL memberikan kelebihan dalam mengakses dan mengelola atribut data mahasiswa dengan mudah dan efisien. Hasil pengujian dari dua eksperimen menunjukkan perkembangan dalam kinerja model. Eksperimen kedua berhasil meningkatkan tingkat akurasi secara konsisten, mencapai 95.63% dibandingkan dengan 94.44% pada eksperimen pertama. Analisis laporan klasifikasi menunjukkan peningkatan presisi, recall, dan f1-score, khususnya pada kelas Aktif, mengindikasikan kemajuan dalam memprediksi mahasiswa yang berpotensi non-aktif. Melalui hasil confusion matrix, eksperimen kedua menunjukkan peningkatan dalam mengenali mahasiswa yang sebenarnya aktif, dengan penurunan False Negative yang signifikan. Hal ini menggambarkan kemampuan model untuk memberikan prediksi yang lebih andal terkait mahasiswa yang berpotensi non-aktif, memungkinkan institusi pendidikan untuk mengambil tindakan preventif dengan lebih tepat waktu. Hasil penelitian ini dapat menjadi referensi untuk implementasi sistem prediksi yang dapat memberikan dukungan dalam pengambilan keputusan strategis di lingkungan pendidikan.

E. Ucapan Terima Kasih

Terima kasih kepada LPPM Universitas Dinamika Bangsa yang telah mendukung penelitian ini, Penelitian ini merupakan bagian dari program penelitian internal Dosen Universitas Dinamika Bangsa

F. Referensi

- [1] K. Hastuti and E. Y. Hidayat, "Analisis Algoritma Decision Tree untuk Prediksi Mahasiswa Non Aktif," *Semantik*, vol. 2013, no. November, pp. 211–216, 2013.
- [2] 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.
- [3] D. Goldblatt, "Foreword," *Semin. Neurol.*, vol. 14, no. 1, pp. 241–249, 1994, doi: 10.2307/j.ctv11hppt6.3.
- [4] U. D. Bangsa, *Peraturan Akademik*. 2020.
- [5] J. Susilo and Y. Syahra, "Implementasi Data Mining Untuk Menganalisa Pola Penjualan Menu Makanan Berdasarkan Permintaan Konsumen DI Restoran

- Wakaka Center Point Menggunakan Algoritma Apriori,” *Cyber Tech*, pp. 1–14, 2020.
- [6] R. H. Sukarna and Y. Ansori, “Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu,” *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022, doi: 10.47080/saintek.v6i1.1467.
- [7] I. A. Nikmatun and I. Waspada, “Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [8] D. Laia, E. Buulolo, and M. J. F. Sirait, “Implementasi Data Mining Untuk Memprediksi Pemesanan Driver Go-Jek Online Dengan Menggunakan Metode Naive Bayes (Studi Kasus: Pt. Go-Jek Indonesia),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 434–439, 2018, doi: 10.30865/komik.v2i1.972.
- [9] S. Samasil, Y. Yuyun, and H. Hazriani, “Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Naive Bayes Dan Decision Tree,” *J. Ilm. Ilmu Komput.*, vol. 8, no. 2, pp. 108–114, 2022, doi: 10.35329/jiik.v8i2.242.
- [10] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in {P}ython,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.