

OPTIMALISASI PREDIKSI LAMA STUDI MAHASISWA MENGGUNAKAN ROUGH SET DAN CASE-BASED REASONING

Muh. Nurtanzis Sutoyo^{*1}, Rabiah Adawiyah²

^{1,2}Universitas Sembilanbelas November Kolaka, Kolaka

Email: ¹mns.usn21@gmail.com, ²bia.fti@usn.ac.id

*Penulis Korespondensi

(Naskah masuk: 11 April 2025, diterima untuk diterbitkan: 18 September 2025)

Abstrak

Prediksi lama studi mahasiswa menjadi aspek penting dalam perencanaan akademik, evaluasi kinerja, serta identifikasi potensi keterlambatan kelulusan. Penelitian ini mengusulkan pendekatan hibrida dengan mengintegrasikan metode Rough Set dan Case-Based Reasoning untuk meningkatkan akurasi prediksi kelas lama studi mahasiswa. Metode Rough Set digunakan untuk mengekstraksi aturan klasifikasi berbasis kombinasi atribut IPK, status bekerja, dan status beasiswa, serta menghitung probabilitas kelas pada boundary region. Di sisi lain, metode CBR dimanfaatkan untuk menghitung similarity antar kasus berdasarkan kemiripan atribut, termasuk jumlah SKS yang dinormalisasi. Hasil prediksi dilakukan melalui integrasi probabilitas dari Rough Set dan similarity CBR menggunakan bobot kombinasi sebesar 0.6 dan 0.4. Pada pengujian kasus baru, diperoleh lima kasus historis paling mirip dengan similarity 0.97, empat di antaranya tergolong “Sangat Terlambat” dan satu “Terlambat”. Sementara itu, probabilitas dari Rough Set menunjukkan distribusi 0.667 untuk “Sangat Terlambat” dan 0.333 untuk “Terlambat”. Hasil integrasi memberikan skor akhir sebesar 0.720 untuk “Sangat Terlambat” dan 0.280 untuk “Terlambat”, yang menunjukkan sistem prediksi cenderung kuat terhadap kategori “Sangat Terlambat”. Pendekatan gabungan ini terbukti efektif dalam menggabungkan kekuatan generalisasi dari Rough Set dan fleksibilitas adaptif dari CBR, sehingga dapat digunakan sebagai sistem pendukung keputusan dalam evaluasi akademik berbasis data historis.

Kata kunci: *Prediksi Lama Studi, Rough Set, Case-Based Reasoning, Boundary Region, Similarity*

OPTIMIZATION OF STUDENT STUDY DURATION PREDICTION USING ROUGH SET AND CASE-BASED REASONING

Abstract

The forecast of student study length is essential for academic planning, performance assessment, and recognizing possible graduation delays. This study presents a hybrid methodology that combines Rough Set theory and Case-Based Reasoning techniques to enhance the precision of predicting student study length classifications. The Rough Set approach is employed to derive classification rules from combinations of attributes, including GPA, job status, and scholarship status, as well as to compute class probabilities inside the boundary region. Simultaneously, the CBR approach is utilized to assess similarity between cases based on attribute similarity, including normalized credit hours (SKS). The prediction results are produced by integrating Rough Set probability and CBR similarity, utilizing weighted values of 0.6 and 0.4, respectively. In the test case, five historical cases with similarity scores of 0.97 were identified, four classified as “Very Late” and one as “Late”. Rough Set probability were 0.667 for “Very Late” and 0.333 for “Late”. The conclusive integrated scores were 0.720 for “Very Late” and 0.280 for “Late”, signifying that the algorithm predominantly forecasts the “Very Late” category. This hybrid methodology adeptly integrates the generalization capabilities of Rough Set theory with the adaptive versatility of Case-Based Reasoning, rendering it appropriate as a decision support system for academic assessment grounded in historical data.

Keywords: *Prediction of Study Duration, Rough Set Theory, Case-Based Reasoning, Boundary Region, Similarity*

1. PENDAHULUAN

Lama studi mahasiswa adalah salah satu faktor penting dalam mengukur seberapa efektif pendidikan

tinggi. Semakin lama waktu yang dibutuhkan siswa untuk menyelesaikan kelas, semakin besar efeknya terhadap efisiensi akademik, perencanaan sumber daya kampus, dan biaya pendidikan. Dalam evaluasi

akreditasi program studi dan laporan kinerja perguruan tinggi, isu-isu lama juga sering menjadi perhatian.

Indeks prestasi kumulatif (IPK), jumlah SKS yang ditempuh setiap semester, status kerja mahasiswa, dan kondisi sosial ekonomi adalah beberapa faktor yang memengaruhi lama studi mahasiswa. Namun, tidak banyak sistem yang dapat memprediksi lama studi berdasarkan riwayat akademik dan kondisi individu siswa. Meskipun demikian, memiliki kemampuan untuk memperkirakan prospek masa depan pendidikan sejak awal dapat membantu kampus mengambil tindakan, memberikan bimbingan, dan mengoptimalkan layanan akademik.

Dalam penelitian ini, masalah yang ingin diselesaikan adalah bagaimana membuat model prediksi lama studi mahasiswa yang akurat dan dapat menghasilkan aturan klasifikasi yang dapat dipahami oleh pengguna (*interpretable*). Selain itu, diharapkan dapat memberikan prediksi yang didasarkan pada kemiripan dengan kasus-kasus sebelumnya, yang didasarkan pada pengalaman siswa sebelumnya.

Penelitian ini bertujuan untuk mengembangkan model hybrid yang menggabungkan metode *Rough Set Theory* (RST) dan *Case-Based Reasoning* (CBR) untuk memprediksi lama studi mahasiswa. Metode Rough Set digunakan untuk mengekstraksi aturan dan mereduksi atribut yang tidak relevan, sementara CBR digunakan untuk menemukan prediksi berdasarkan kasus yang memiliki kemiripan tinggi. Kontribusi utama dari penelitian ini adalah menghadirkan pendekatan prediksi yang tidak hanya akurat, tetapi juga efisien dan mudah diinterpretasikan oleh pengguna (akademisi atau dosen wali).

Beberapa penelitian sebelumnya telah membahas tentang RST, CBR, maupun lama studi. Diantaranya RST untuk menganalisis karakter tertanggung dalam sistem klaim asuransi swasta di Indonesia (Ekaputrie and Fauzan, 2023). Dalam penelitian lain, RST digunakan untuk mengurangi data sensor melalui pendekatan granularitas informasi lokal yang adil (*justifiable granularity*), yang memungkinkan klasifikasi dan reduksi atribut secara lebih akurat pada data sensor yang tidak teratur (Fan et al., 2023). Selain itu, RST juga diterapkan untuk mengatasi masalah data tidak lengkap dalam analisis pelanggan (Gholami et al., 2018). RST digunakan untuk mengklasifikasikan dokumen teks berbasis fitur morfologis dan frekuensi kata (Yang et al., 2022). RST untuk pengambilan keputusan dalam lingkungan sistem informasi geografis (Yang et al., 2022). Selanjutnya RST digunakan sebagai alat analisis untuk identifikasi atribut yang signifikan dalam diagnosis medis (Alkhazaleh and Marei, 2021). RST diperluas dengan menggabungkan teori soft set dan neutrosophic set menjadi neutrosophic soft rough set, yang mampu menangani data yang tidak pasti, tidak lengkap, dan ambigu tanpa menggunakan full soft set, serta digunakan untuk

menyusun pendekatan klasifikasi dan generalisasi data (Alkhazaleh and Marei, 2021). Serta dikembangkan metode RST baru melalui pendekatan increasing dan decreasing approximation yang lebih akurat dibandingkan model klasik, dan diterapkan pada konteks pengambilan keputusan dalam situasi tidak pasti dengan konversi relasi parsial menjadi relasi total (El-Sheikh et al., 2023).

Penggunaan RST dalam studi-studi tersebut menunjukkan fleksibilitas dan kekuatannya dalam menangani ketidakpastian dan ketidaktepatan data, menjadikannya metode yang efektif untuk mendukung pengambilan keputusan di berbagai domain.

Sedangkan untuk CBR digunakan dalam berbagai penelitian untuk menyelesaikan masalah baru dengan mengandalkan pengalaman masa lalu yang tersimpan dalam basis kasus. Diantaranya pendekatan CBR digunakan untuk mendukung keputusan pada sistem diagnosis penyakit dengan mempertimbangkan kemiripan antar kasus (Ye, 2024). CBR dioptimalkan untuk pengambilan keputusan pada sistem rekomendasi yang bersifat personal dan relevan (Chun and Ko, 2020). CBR diterapkan dalam domain manajemen energi (Chun and Jang, 2022). CBR dimanfaatkan untuk menjelaskan hasil prediksi model pembelajaran mesin yang kompleks (Zeng et al., 2023). Selanjutnya integrasi CBR dengan pembelajaran serupa dalam konteks pemeliharaan sistem (Caro-Martinez et al., 2019). CBR yang berfokus pada peningkatan kinerja klasifikasi dengan merawat basis kasus secara selektif berdasarkan reputasi tiap kasus (Nakhjiri et al., 2020).

Selanjutnya CBR digunakan untuk merancang sistem rekomendasi diagnosis keperawatan berbasis data sebelumnya dengan mengintegrasikan teknik klasifikasi dan metode adaptasi untuk menangani variasi kasus yang luas (Gigih et al., 2020). Mengembangkan sistem penjadwalan tugas dinamis untuk pegawai dalam konteks industri (Jones and Hardiyanti, 2021). Kemudian CBR diterapkan dalam sistem prediksi harga rumah dengan mempertimbangkan karakteristik properti sebelumnya (Nurfalinda and Uperati, 2020). Pendekatan CBR dimanfaatkan dalam evaluasi kelayakan kredit mikro untuk pelaku UMKM (Safitri et al., 2018). Di bidang kesehatan CBR digunakan untuk mengembangkan sistem diagnosis penyakit mata berbasis citra retina (Minarni et al., 2018). CBR digunakan untuk membantu dokter dalam pengambilan keputusan medis dengan memanfaatkan data prosedur medis (Paruchuri and Granville, 2020). Terakhir, penerapan CBR dalam konteks hukum dengan menggabungkan logika klasifikasi biner untuk menjelaskan alasan di balik putusan berbasis preseden, sehingga memperkuat transparansi sistem pengambilan keputusan otomatis dalam ranah legal (Liu et al., 2022).

Berbagai pendekatan telah digunakan dalam penelitian untuk memprediksi dan menganalisis lama

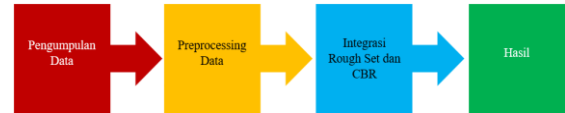
studi. Lama studi dianalisis menggunakan pendekatan *Cox Proportional Hazard Model* dan *Kaplan-Meier Curve* untuk mempelajari faktor-faktor yang mempengaruhi lama studi mahasiswa di tingkat universitas, dengan mempertimbangkan variabel-variabel seperti gender, nilai akademik, dan status ekonomi (Feely et al., 2020). Menerapkan metode *Random Forest* untuk klasifikasi dan prediksi ketepatan waktu kelulusan mahasiswa (Morris et al., 2022). Lama studi menggunakan *multinomial logistic regression* untuk menilai pengaruh berbagai faktor seperti IPK, status kerja, dan beasiswa (Edde et al., 2023). Berbeda dari sebelumnya, prediksi lama studi pada konteks internasional dengan pendekatan *survival analysis*, yang mempertimbangkan variabel-variabel sosial-ekonomi dan akademik sebagai penentu utama (Zulfan et al., 2023). Terakhir, prediksi lama studi pada uji klinis berbasis *time-to-event analysis* dengan pendekatan probabilistik (Machida et al., 2021).

Namun, hingga saat ini, belum banyak penelitian yang mengintegrasikan RST dan CBR secara bersamaan dalam konteks prediksi lama studi mahasiswa. Integrasi RST dan CBR dipilih karena RST mampu menyederhanakan data serta menghasilkan aturan klasifikasi yang interpretable, sementara CBR memberikan prediksi berbasis pengalaman kasus serupa, sehingga kombinasi keduanya menghasilkan model prediksi yang lebih akurat, adaptif, dan transparan. Oleh karena itu, penelitian ini mencoba mengisi celah tersebut dengan membangun sistem prediktif berbasis rule dan pengalaman kasus.

Pemilihan metode Rough Set dan Case-Based Reasoning dalam penelitian ini didasarkan pada kekuatan masing-masing pendekatan. Rough Set unggul dalam menyederhanakan data serta mengekstrak aturan keputusan tanpa memerlukan parameter eksternal seperti derajat keanggotaan, sementara CBR mampu memberikan hasil prediksi berbasis pengalaman nyata dari kasus-kasus terdahulu yang serupa. Dengan menggabungkan kedua metode ini, diharapkan sistem yang dihasilkan dapat lebih adaptif, efisien, dan dapat memberikan prediksi yang lebih tepat. Sedangkan tujuan penelitian ini adalah mengembangkan model prediksi lama studi mahasiswa berbasis pendekatan hibrida dengan mengintegrasikan metode Rough Set dan Case-Based Reasoning untuk menghasilkan sistem prediksi yang akurat, efisien, adaptif, serta mudah diinterpretasikan bagi kebutuhan evaluasi dan perencanaan akademik.

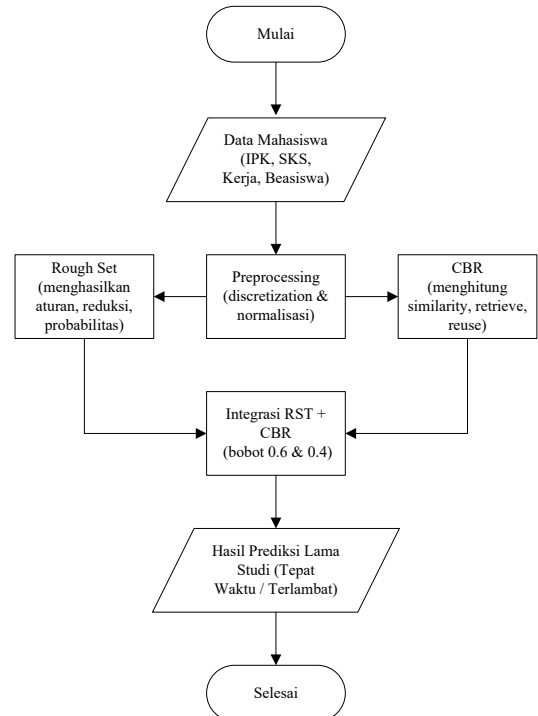
2. METODE PENELITIAN

Tahapan penelitian ini dirancang secara sistematis untuk menggambarkan alur proses mulai dari pengumpulan data hingga menghasilkan prediksi akhir. Secara lebih jelas, tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Untuk memperjelas alur integrasi metode Rough Set dan Case-Based Reasoning dalam penelitian ini, disajikan flowchart pada Gambar 2 yang menggambarkan hubungan antar tahapan utama hingga menghasilkan prediksi lama studi mahasiswa.



Gambar 2. Flowchart Integrasi Rough Set dan CBR

Pendekatan hybrid yang menggabungkan metode RST dan CBR digunakan dalam penelitian ini untuk melakukan prediksi lama studi siswa berdasarkan data akademik historis. Prediksi dilakukan dalam dua tahap utama: pembentukan aturan dengan RST dan prediksi berbasis kasus serupa dengan CBR. Metode ini digunakan untuk mengoptimalkan akurasi prediksi dan mempercepat proses klasifikasi dengan mengurangi atribut yang tidak penting.

2.1. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari sistem informasi akademik salah satu perguruan tinggi di Indonesia yaitu USN Kolaka pada tahun akademik 2024/2025 gasal. Dataset terdiri dari data riwayat akademik mahasiswa yang telah menyelesaikan studi, dengan jumlah 126 data mahasiswa Program Studi Sistem Informasi. Setiap data mencakup atribut seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS per semester, status bekerja, status beasiswa, durasi studi (dalam semester), dan label kelas berupa kategori "Tepat Waktu" atau "Terlambat".

Alasan penggunaan atribut Indeks Prestasi Kumulatif (IPK), jumlah SKS per semester, status bekerja, status beasiswa, dan durasi studi adalah karena kelima faktor tersebut secara langsung mencerminkan capaian akademik, beban studi, serta kondisi sosial-ekonomi mahasiswa yang berpengaruh signifikan terhadap ketepatan waktu penyelesaian studi, sehingga relevan dijadikan variabel prediktor dalam membangun model klasifikasi lama studi.

Adapun contoh data yang digunakan dalam penelitian ini meliputi atribut Indeks Prestasi Kumulatif (IPK), jumlah SKS per semester, status bekerja, status beasiswa, serta durasi studi mahasiswa sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Contoh Data Penelitian

Mahasiswa	SKS	IPK	Status Bekerja	Status Beasiswa
MHS001	143	2.87	Ya	Tidak
MHS002	150	3.39	Ya	Tidak
MHS003	152	3.22	Ya	Tidak
MHS004	145	2.79	Ya	Tidak
MHS005	143	3.28	Ya	Tidak
MHS006	146	3.38	Ya	Ya
MHS007	152	3.14	Ya	Tidak
MHS008	152	3.16	Ya	Tidak
MHS009	148	2.85	Ya	Ya
MHS010	152	3.61	Tidak	Tidak
...
MHS125	148	3.88	Tidak	Tidak
MHS126	151	3.82	Tidak	Tidak

Data yang digunakan adalah kombinasi data numerik dan kategorik. Oleh karena itu, proses disretisasi dilakukan pada data numerik sehingga metode Rough Set dapat digunakan untuk memprosesnya. Selain itu, data numerik dinormalisasi untuk memastikan kesetaraan skala selama perhitungan similarity pada tahap CBR.

2.2. Tahapan Rough Set

Metode Rough Set digunakan sebagai pendekatan awal dalam proses prediksi untuk menyederhanakan atribut (*feature selection*) dan menghasilkan aturan keputusan berdasarkan data historis mahasiswa.

Tahapan pertama dalam metode ini adalah membentuk *information system* yang direpresentasikan sebagai pasangan $S = (U, A)$, di mana U adalah himpunan objek (dalam hal ini, mahasiswa), dan A adalah himpunan atribut yang mencakup data seperti IPK, jumlah SKS, status beasiswa, dan lain sebagainya. Setiap atribut $a \in A$ berfungsi sebagai pemetaan dari objek ke domain nilainya, yaitu $a : U \rightarrow Va$.

Karena Rough Set bekerja optimal pada data diskrit atau kategorikal, maka dilakukan disretisasi terhadap atribut numerik menggunakan metode *equal-width binning*, agar nilai-nilai numerik seperti

IPK dapat dikelompokkan ke dalam interval tertentu (contoh: "rendah", "sedang", "tinggi").

Langkah selanjutnya adalah pembentukan relasi tak terbedakan (*indiscernibility relation*). Relasi ini dibangun berdasarkan subset atribut $B \subseteq A$, dan dua objek dikatakan tak terbedakan jika memiliki nilai atribut yang sama pada seluruh atribut dalam B . Relasi ini dinyatakan dengan rumus (El-Sheikh et al., 2023):

$$IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, a(x) = a(y)\} \quad (1)$$

Setelah relasi terbentuk, dilakukan proses aproksimasi kasar terhadap kelas keputusan dengan membentuk aproksimasi bawah (*lower approximation*) dan aproksimasi atas (*upper approximation*). Untuk setiap himpunan objek $X \subseteq U$, aproksimasi bawah adalah himpunan objek yang pasti termasuk dalam X , sedangkan aproksimasi atas adalah himpunan objek yang mungkin termasuk dalam X . Keduanya dirumuskan sebagai berikut:

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\} \quad (2)$$

$$\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\} \quad (3)$$

Kelas ekivalensi $[x]_B$ menunjukkan sekumpulan objek yang tak terbedakan terhadap objek x berdasarkan atribut B .

Dari hasil aproksimasi tersebut, dibentuklah *positive region* yaitu objek-objek yang dapat diklasifikasikan secara pasti ke dalam kelas keputusan D . Positive region dihitung dengan:

$$POS_B(D) = \bigcup_{x \in U/D} \underline{B}(X) \quad (4)$$

Nilai derajat dependency atau ketergantungan antara atribut kondisi B terhadap keputusan D dihitung menggunakan rasio jumlah objek pada positive region terhadap total objek dalam U , dengan rumus:

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|} \quad (5)$$

Apabila nilai $\gamma_B(D) = 1$, berarti seluruh objek dalam data dapat diklasifikasikan secara pasti menggunakan atribut B , dan tidak ada ambiguitas.

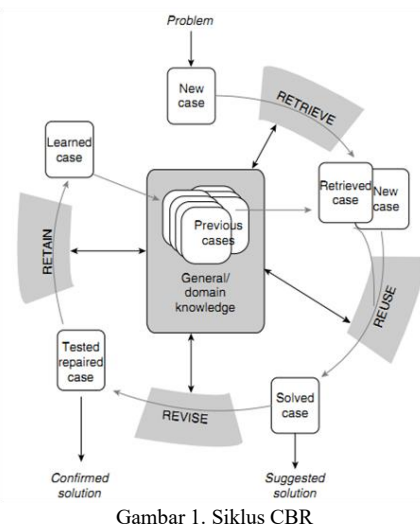
Langkah berikutnya adalah melakukan reduksi atribut (*Reduct*), yaitu mencari subset minimal dari atribut A yang tetap mempertahankan nilai dependency yang sama seperti atribut lengkap. Reduct diperoleh melalui proses eliminasi iteratif terhadap atribut yang tidak berpengaruh signifikan terhadap positive region.

Terakhir, dilakukan ekstraksi aturan keputusan (*decision rule generation*) berdasarkan positive region yang terbentuk. Aturan-aturan ini dihasilkan dalam bentuk logika *if-then* yang dapat diinterpretasikan secara langsung. Contoh aturan

yang mungkin terbentuk adalah $IF\ IPK = Tinggi\ AND\ SKS = Banyak\ THEN\ Lama\ Studi = Tepat\ Waktu$.

2.3. Tahapan Case-Based Reasoning (CBR)

Metode CBR digunakan sebagai pendekatan prediksi berbasis pengalaman, dengan prinsip bahwa masalah baru dapat diselesaikan dengan merujuk pada solusi dari kasus-kasus terdahulu yang serupa. Dalam penelitian ini, CBR digunakan untuk memprediksi lama studi mahasiswa berdasarkan kemiripan atribut dengan mahasiswa sebelumnya yang telah menyelesaikan studi. Proses CBR terdiri dari empat tahap utama yang dikenal dengan istilah 4R, yaitu *Retrieve*, *Reuse*, *Revise*, dan *Retain*.



Gambar 1. Siklus CBR

Tahap pertama *Retrieve*, adalah proses pengambilan kasus lama yang paling mirip dengan kasus baru. Untuk menjaga relevansi perbandingan, pengambilan kasus dibatasi hanya pada mahasiswa yang berasal dari program studi yang sama. Hal ini dilakukan dengan menambahkan kondisi filter pada basis kasus saat pencarian, sehingga hanya kasus dengan prodi yang identik yang dipertimbangkan dalam proses similarity. Rumus perhitungan similarity yang digunakan adalah sebagai berikut (Nurfalinda and Uperati, 2020):

$$Similarity(T, S) = \frac{\sum_{i=1}^n f(T, S) \cdot w_i}{\sum_{i=1}^n w_i} \quad (6)$$

Dalam rumus tersebut, T adalah kasus baru, S adalah kasus lama dalam basis kasus, T_i dan S_i adalah nilai atribut ke- i dari masing-masing kasus, dan w_i adalah bobot atribut ke- i yang diperoleh dari hasil Rough Set. Fungsi kecocokan $f(T_i, S_i)$ didefinisikan sebagai:

$$f(T_i, S_i) = \begin{cases} 1, & \text{Jika } T_i = S_i \\ 0, & \text{Jika } T_i \neq S_i \end{cases} \quad (7)$$

Jika data bersifat numerik dan tidak dilakukan diskritisasi, maka rumus similarity dapat diganti menggunakan pendekatan jarak Euclidean yang dinormalisasi sebagai berikut:

$$Similarity(T, S) = \frac{\sqrt{\sum_{i=1}^n (T_i - S_i)^2}}{d_{max}} \quad (8)$$

di mana d_{max} adalah jarak maksimum antar dua kasus dalam dataset, yang digunakan untuk menjaga nilai similarity tetap dalam rentang 0 hingga 1.

Tahap kedua adalah *Reuse*, yaitu pemanfaatan solusi dari kasus yang paling mirip. Jika hanya satu kasus yang diambil, maka kelas dari kasus tersebut akan langsung digunakan. Namun, jika digunakan lebih dari satu kasus, maka dilakukan proses voting terhadap k kasus dengan similarity tertinggi untuk menentukan kelas prediksi. Proses ini dapat dinyatakan dengan rumus:

$$Class(T) = mode(\{Class(S_1), \dots, Class(S_k)\}) \quad (9)$$

Tahap ketiga adalah *Revise*, yaitu pengecekan ulang terhadap solusi yang dihasilkan. Jika prediksi tidak sesuai dengan kondisi nyata atau validasi pakar, maka solusi dapat disesuaikan. Dalam penelitian ini, tahap revisi tidak dilakukan secara eksplisit karena fokus berada pada evaluasi akurasi sistem.

Tahap terakhir adalah *Retain*, yaitu penyimpanan kasus baru beserta prediksi hasil klasifikasi ke dalam basis kasus. Hal ini bertujuan agar sistem dapat belajar dari pengalaman baru, dan basis kasus terus berkembang seiring waktu sehingga mampu memberikan prediksi yang lebih akurat dan relevan di masa depan.

2.4. Integrasi RST dan CBR

Proses prediksi pada penelitian ini memanfaatkan pendekatan gabungan antara RST dan CBR. RST digunakan untuk menghasilkan aturan-aturan klasifikasi berdasarkan kombinasi atribut yang terdapat dalam data pelatihan. Pada kombinasi atribut yang tidak memiliki kejelasan kelas (boundary region), dilakukan perhitungan nilai probabilitas kemunculan setiap kelas berdasarkan frekuensi kemunculannya dalam data.

Aturan-aturan inilah yang kemudian dihitung nilai probabilitasnya melalui pendekatan Rough Set, dan selanjutnya dikombinasikan dengan hasil perhitungan similarity pada proses CBR. CBR digunakan untuk mengukur tingkat kemiripan kasus baru terhadap kasus-kasus lama yang relevan dalam basis kasus. Proses ini memberikan fleksibilitas tambahan dalam menghadapi variasi data mahasiswa yang tidak selalu identik secara atribut.

Untuk memperoleh skor akhir prediksi, dilakukan integrasi antara nilai probabilitas RST dan similarity CBR menggunakan formula:

$$Score = (\alpha \cdot P_{RST}) + ((1 - \alpha) \cdot Similarity_{CBR}) \quad (10)$$

dengan nilai bobot $\alpha = 0.6$, yang berarti probabilitas dari Rough Set diberikan bobot lebih besar dibandingkan nilai similarity dari CBR. Hal ini didasarkan pada pertimbangan bahwa probabilitas RST merepresentasikan distribusi kelas berdasarkan

keseluruhan data historis, sehingga lebih stabil dan general. Di sisi lain, similarity CBR bersifat lokal karena hanya bergantung pada kedekatan dengan kasus-kasus individu, sehingga cenderung sensitif terhadap outlier.

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil dari implementasi dan pengujian sistem prediksi lama studi mahasiswa menggunakan metode gabungan Rough Set dan Case-Based Reasoning. Tahapan pembahasan disesuaikan dengan tahapan dalam metode penelitian, yaitu mulai dari preprocessing data, pembentukan aturan dengan Rough Set, perhitungan similarity dengan CBR, hingga evaluasi akurasi sistem.

3.1. Hasil Preprocessing dan Disretisasi Data

Tahap awal yang dilakukan adalah preprocessing data, termasuk pembersihan nilai kosong dan disretisasi pada atribut numerik seperti IPK dan jumlah SKS. Disretisasi dilakukan menggunakan metode *equal-width binning*, dan hasilnya dikategorikan menjadi tiga level: Rendah, Sedang, dan Tinggi. Data yang digunakan dalam penelitian ini berasal dari Program Studi Sistem Informasi dengan jumlah total 126 data mahasiswa yang telah lulus. Hasil dari proses ini dapat dilihat pada Tabel 2, yang menunjukkan konversi data numerik menjadi kategori diskrit.

Tabel 2. Data Hasil Perprocessing

ID	IPK	Bekerja	Beasiswa	Lama Studi
MHS001	Sedang	Ya	Tidak	Sangat Terlambat
MHS002	Sedang	Ya	Tidak	Sangat Terlambat
MHS003	Sedang	Ya	Tidak	Sangat Terlambat
MHS004	Sedang	Ya	Tidak	Sangat Terlambat
MHS005	Sedang	Ya	Tidak	Sangat Terlambat
MHS006	Sedang	Ya	Tidak	Sangat Terlambat
...
MHS118	Tinggi	Tidak	Tidak	Terlambat
MHS119	Tinggi	Tidak	Tidak	Terlambat
MHS120	Tinggi	Tidak	Tidak	Terlambat
MHS121	Tinggi	Tidak	Tidak	Terlambat
MHS122	Tinggi	Tidak	Tidak	Terlambat
MHS123	Tinggi	Tidak	Ya	Terlambat
MHS124	Tinggi	Tidak	Ya	Terlambat
MHS125	Tinggi	Tidak	Ya	Terlambat
MHS126	Tinggi	Tidak	Tidak	Tepat Waktu

Konversi data numerik IPK menjadi kategori didasarkan pada Peraturan Rektor Universitas Sembilanbelas November Kolaka Nomor 02 Tahun

2024 tentang Pedoman Akademik Universitas Sembilanbelas November Kolaka.

3.2. Hasil Ekstraksi Rough Set

Setelah data berhasil didisretisasi, dilakukan pembentukan relasi tak terbedakan, aproksimasi kasar, serta penghitungan positive region dan reduct menggunakan metode Rough Set. Aturan-aturan klasifikasi yang terbentuk dari *positive region* dirangkum dalam Tabel 3. Aturan ini diperoleh berdasarkan kombinasi atribut yang secara konsisten mengarah pada satu kelas keputusan tertentu (definitif), sehingga dapat dimasukkan ke dalam *lower approximation*. Dari total 126 data yang dianalisis, diperoleh sebanyak 18 aturan klasifikasi (*decision rules*) yang merepresentasikan pola-pola tetap dalam data mahasiswa terhadap kelas lama studi.

Tabel 3. Rule Klasifikasi Berdasarkan Rough Set

Rule	IPK	Bekerja	Beasiswa	Lama Studi
R1	Sedang	Ya	Tidak	Sangat Terlambat
R2	Sedang	Ya	Tidak	Terlambat
R3	Sedang	Ya	Ya	Sangat Terlambat
R4	Sedang	Ya	Ya	Terlambat
R5	Sedang	Tidak	Tidak	Sangat Terlambat
R6	Sedang	Tidak	Tidak	Terlambat
R7	Sedang	Tidak	Ya	Sangat Terlambat
R8	Sedang	Tidak	Ya	Terlambat
R9	Rendah	Ya	Ya	Sangat Terlambat
R10	Rendah	Ya	Tidak	Sangat Terlambat
R11	Rendah	Tidak	Tidak	Sangat Terlambat
R12	Rendah	Tidak	Tidak	Terlambat
R13	Tinggi	Ya	Tidak	Terlambat
R14	Tinggi	Ya	Tidak	Tepat Waktu
R15	Tinggi	Tidak	Tidak	Sangat Terlambat
R16	Tinggi	Tidak	Tidak	Terlambat
R17	Tinggi	Tidak	Ya	Terlambat
R18	Tinggi	Tidak	Tidak	Tepat Waktu

Kemudian kasus baru yang disajikan memiliki kombinasi atribut berupa IPK *Sedang*, berstatus *Bekerja*, dan *Tidak* menerima beasiswa, dengan label *Kelas Lama Studi* yang belum diketahui. Kombinasi ini kemudian diproses menggunakan pendekatan Rough Set untuk menghitung probabilitas kelas keputusan sebelum dilanjutkan ke tahap similarity pada Case-Based Reasoning (CBR).

Dalam metode Rough Set, kombinasi tersebut tidak ditemukan secara eksak dalam *positive region*, sehingga dianggap berada pada *boundary region*. Oleh karena itu, dilakukan perhitungan probabilitas kemunculan masing-masing kelas berdasarkan frekuensi kombinasi serupa dalam data latih. Berdasarkan hasil analisis, diketahui bahwa dari seluruh data yang memiliki kombinasi atribut yang sama, sebesar 66.7% (0.667) mengarah pada kelas "Sangat Terlambat", sementara sisanya mengarah 33.3% (0.333) pada kelas "Terlambat".

3.3. Hasil Perhitungan dan Prediksi CBR

Langkah selanjutnya adalah proses prediksi menggunakan CBR, yang didasarkan pada pencocokan kemiripan antar kasus. Tabel 4 menampilkan nilai similarity antara kasus uji dengan beberapa kasus dalam basis kasus, beserta hasil prediksi sistem.

Dalam proses ini, seluruh atribut termasuk jumlah SKS diikutsertakan sebagai bagian dari perhitungan similarity. Karena SKS merupakan atribut numerik, maka dilakukan proses normalisasi menggunakan metode Min-Max untuk mengubah nilai asli ke dalam rentang [0, 1], sehingga dapat dihitung kemiripannya secara proporsional dengan atribut lain yang bersifat kategorik. Setelah dinormalisasi, setiap atribut diberi bobot berdasarkan tingkat kontribusinya terhadap hasil prediksi, di antaranya IPK (0.35), status bekerja (0.20), status beasiswa (0.15), dan SKS (0.35).

Tabel 4. Nilai Similarity Antara Kasus

ID	SKS	IPK	Bekerja	Beasiswa	Similarity
MHS001	0.21	0.35	0.20	0.15	0.91
MHS002	0.27	0.35	0.20	0.15	0.97
MHS003	0.00	0.35	0.20	0.15	0.70
MHS004	0.27	0.35	0.20	0.15	0.97
MHS005	0.27	0.35	0.20	0.15	0.97
MHS006	0.18	0.35	0.20	0.15	0.88
MHS007	0.18	0.35	0.20	0.15	0.88
MHS008	0.27	0.35	0.20	0.15	0.97
MHS009	0.12	0.35	0.20	0.15	0.82
MHS010	0.12	0.35	0.20	0.15	0.82
...
MHS118	0.12	0.00	0.00	0.15	0.27
MHS119	0.15	0.00	0.00	0.15	0.30
MHS120	0.24	0.00	0.00	0.15	0.39
MHS121	0.24	0.00	0.00	0.15	0.39
MHS122	0.24	0.00	0.00	0.15	0.39
MHS123	0.12	0.00	0.00	0.00	0.12
MHS124	0.24	0.00	0.00	0.00	0.24
MHS125	0.24	0.00	0.00	0.00	0.24
MHS126	0.15	0.00	0.00	0.15	0.30

Nilai similarity akhir dihitung berdasarkan kombinasi nilai kemiripan dari masing-masing atribut, dan lima kasus dengan similarity tertinggi kemudian dipilih untuk mendukung proses prediksi dirangkum pada Tabel 5.

Tabel 5. Nilai Similarity Tertinggi

ID	Similarity	Kelas Lama Studi
MHS002	0.97	Sangat Terlambat
MHS004	0.97	Sangat Terlambat
MHS005	0.97	Sangat Terlambat
MHS008	0.97	Sangat Terlambat
MHS015	0.97	Terlambat

Berdasarkan hasil perhitungan similarity antara kasus uji dan basis kasus menggunakan metode CBR, diperoleh lima kasus dengan tingkat kemiripan tertinggi sebesar 0.97. Dari lima kasus tersebut,

empat kasus termasuk dalam kelas “Sangat Terlambat” dan satu kasus termasuk dalam kelas “Terlambat”. Oleh karena itu, dapat dihitung probabilitas untuk masing-masing kelas berdasarkan proporsi kemunculannya dalam lima kasus tersebut, yaitu sebesar 0.80 (80%) untuk kelas “Sangat Terlambat” dan 0.20 (20%) untuk kelas “Terlambat”. Nilai probabilitas ini mencerminkan kecenderungan keputusan berdasarkan kedekatan lokal terhadap kasus-kasus historis yang paling mirip dengan kasus baru.

3.4. Integrasi RST dan CBR

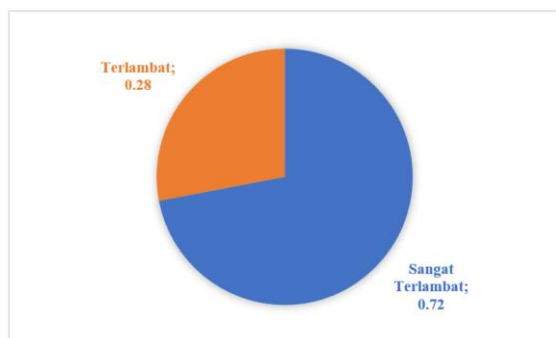
Setelah diperoleh nilai probabilitas dari metode Rough Set dan metode Case-Based Reasoning (CBR), langkah selanjutnya adalah melakukan integrasi kedua pendekatan tersebut untuk menghasilkan skor akhir prediksi. Nilai probabilitas yang dihasilkan dari Rough Set adalah 0.667 untuk kelas “Sangat Terlambat” dan 0.333 untuk kelas “Terlambat”, berdasarkan distribusi frekuensi kombinasi atribut dalam boundary region. Sementara itu, hasil probabilitas dari CBR, yang dihitung dari lima kasus dengan tingkat similarity tertinggi sebesar 0.97, menunjukkan bahwa empat kasus tergolong “Sangat Terlambat” dan satu kasus “Terlambat”, sehingga diperoleh probabilitas 0.80 untuk “Sangat Terlambat” dan 0.20 untuk “Terlambat”. Kedua probabilitas tersebut kemudian diintegrasikan menggunakan bobot gabungan, yaitu 0.6 untuk RST dan 0.4 untuk CBR.

$$Score_{ST} = (0.6 \times 0.667) + (0.4 \times 0.80) = 0.720$$

$$Score_T = (0.6 \times 0.333) + (0.4 \times 0.20) = 0.280$$

Dari hasil integrasi kedua pendekatan, dapat disimpulkan bahwa sistem prediksi menunjukkan kecenderungan kuat terhadap kelas “Sangat Terlambat”, dengan skor akhir sebesar 0.720 (72%), yang berasal dari kombinasi probabilitas Rough Set dan similarity CBR. Skor ini secara signifikan lebih tinggi dibandingkan dengan kelas “Terlambat”, yang hanya memperoleh skor sebesar 0.280 (28%). Hasil ini mengindikasikan bahwa berdasarkan distribusi data historis dan kemiripan kasus, mahasiswa dengan atribut berupa IPK *Sedang*, berstatus *Bekerja*, dan *Tidak* menerima beasiswa termasuk dalam kategori “Sangat Terlambat”.

Hasil akhir dari integrasi antara pendekatan Rough Set dan Case-Based Reasoning (CBR) dapat dilihat pada Gambar 1. Gambar tersebut menunjukkan proses penggabungan nilai probabilitas yang diperoleh dari aturan klasifikasi Rough Set dengan nilai similarity dari hasil perhitungan CBR. Kombinasi ini tidak hanya memperkuat hasil prediksi, tetapi juga memberikan justifikasi berbasis aturan dan kedekatan kasus yang transparan dan dapat dipertanggungjawabkan.



Gambar 1. Hasil Integrasi RST dan CBR

3.5. Analisis dan Pembahasan

Hasil implementasi menunjukkan bahwa integrasi metode Rough Set dan Case-Based Reasoning (CBR) mampu memberikan prediksi lama studi mahasiswa yang tidak hanya akurat tetapi juga dapat diinterpretasikan. Pendekatan Rough Set berhasil menghasilkan aturan-aturan klasifikasi berdasarkan kombinasi atribut IPK, status bekerja, dan status beasiswa, yang menyederhanakan data dan mengurangi dimensi tanpa mengorbankan informasi penting. Pada kasus baru yang diuji, kombinasi atribut tersebut tidak ditemukan dalam positive region, sehingga dimasukkan ke boundary region dan dihitung probabilitasnya. Hasilnya menunjukkan bahwa 66,7% dari kombinasi serupa mengarah pada kelas “Sangat Terlambat”.

Di sisi lain, pendekatan CBR berhasil menyesuaikan prediksi berdasarkan kemiripan lokal terhadap kasus-kasus historis. Lima kasus dengan similarity tertinggi (0.97) ditemukan, empat di antaranya berasal dari kelas “Sangat Terlambat” dan satu dari “Terlambat”. Dengan demikian, probabilitas prediksi berbasis CBR adalah 80% untuk “Sangat Terlambat” dan 20% untuk “Terlambat”. Hal ini menunjukkan bahwa prediksi berdasarkan kedekatan kasus memiliki kecenderungan yang kuat terhadap kelas dominan, meskipun jumlah data terbatas dan kasus serupa tidak selalu identik secara keseluruhan.

Integrasi kedua pendekatan dilakukan dengan formula gabungan, yaitu 60% bobot diberikan untuk probabilitas hasil Rough Set dan 40% untuk similarity CBR. Hasil penggabungan menunjukkan skor tertinggi sebesar 0.720 pada kelas “Sangat Terlambat”, sedangkan skor pada kelas “Terlambat” hanya sebesar 0.280. Ini mengindikasikan bahwa sistem prediksi memiliki keyakinan kuat terhadap klasifikasi “Sangat Terlambat”. Keputusan ini mencerminkan keunggulan pendekatan hybrid dalam menggabungkan kekuatan generalisasi dari RST dan kekuatan adaptif dari CBR, sehingga menghasilkan sistem prediksi yang lebih tangguh terhadap variasi data dan tetap mempertahankan interpretabilitasnya.

Untuk menilai kinerja model prediksi yang dibangun, dilakukan evaluasi menggunakan confusion matrix dengan menghitung metrik akurasi, precision, recall, dan F1-score pada setiap kelas hasil prediksi.

Berdasarkan perhitungan, kelas Sangat Terlambat memperoleh precision sebesar 0.495, recall 0.911, dan F1-score 0.642, yang menunjukkan model mampu mengenali sebagian besar mahasiswa dengan kategori tersebut. Sementara itu, kelas Terlambat memiliki precision 0.700, recall 0.215, dan F1-score 0.329, sehingga meskipun prediksi yang dihasilkan cukup tepat ketika model memilih kelas ini, namun masih banyak kasus Terlambat yang salah diklasifikasikan ke kelas lain. Adapun kelas Tepat Waktu tidak terprediksi sama sekali sehingga precision, recall, dan F1-score bernilai 0. Secara keseluruhan, nilai macro average diperoleh sebesar precision 0.398, recall 0.375, dan F1-score 0.324, yang menunjukkan bahwa model cenderung lebih baik dalam mengenali kelas dominan (Sangat Terlambat) dibandingkan kelas minoritas (Tepat Waktu).

4. KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi metode Rough Set dan Case-Based Reasoning (CBR) dapat digunakan secara efektif untuk memprediksi lama studi mahasiswa berdasarkan atribut IPK, status bekerja, status beasiswa, dan jumlah SKS. Pendekatan Rough Set digunakan untuk membentuk aturan klasifikasi dan menghitung probabilitas pada boundary region, sedangkan CBR digunakan untuk mengukur tingkat kemiripan kasus baru dengan kasus-kasus historis. Hasil integrasi menunjukkan bahwa skor tertinggi diperoleh pada kelas “Sangat Terlambat” dengan nilai 0.720, mengindikasikan bahwa mahasiswa dengan atribut yang sama memiliki kecenderungan kuat terhadap kategori tersebut. Penggabungan dua pendekatan ini memungkinkan sistem prediksi tidak hanya mempertimbangkan distribusi global dari data historis, tetapi juga memperhatikan kedekatan lokal antar kasus. Dengan demikian, sistem menjadi lebih adaptif, akurat, dan tetap dapat dijelaskan secara logis, menjadikannya solusi yang menjanjikan dalam mendukung evaluasi dan perencanaan akademik.

Penelitian selanjutnya dapat memperluas jumlah dataset dengan melibatkan lebih banyak program studi atau universitas berbeda agar model lebih general. Penelitian lanjutan juga dapat mengintegrasikan metode optimasi, seperti genetic algorithm atau particle swarm optimization, untuk menentukan bobot kombinasi terbaik antara Rough Set dan Case-Based Reasoning. Disamping itu, eksplorasi metode discretization dan similarity measure lain dapat dilakukan guna meningkatkan performa model. Terakhir, sistem prediksi ini dapat dikembangkan menjadi aplikasi berbasis web agar lebih mudah digunakan oleh dosen wali maupun pihak akademik dalam melakukan monitoring dan bimbingan studi mahasiswa.

DAFTAR PUSTAKA

- ALKHAZALEH, S. AND MAREI, E., 2021. New Soft Rough Set Approximations. *International Journal of Fuzzy Logic and Intelligent Systems*, 21(2), pp.123–134.
- CARO-MARTINEZ, M., RECIO-GARCIA, J.A. AND JIMENEZ-DIAZ, G., 2019. An Algorithm Independent Case-Based Explanation Approach for Recommender Systems Using Interaction Graphs. In: *27th International Conference, ICCBR 2019 Otzenhausen, Germany, September 8–12, 2019 Proceedings*. pp.17–32.
- CHUN, S.H. AND JANG, J.W., 2022. A New Trend Pattern-Matching Method of Interactive Case-Based Reasoning for Stock Price Predictions. *Sustainability (Switzerland)*, 14(3).
- CHUN, S.H. AND KO, Y.W., 2020. Geometric case based reasoning for stock market prediction. *Sustainability (Switzerland)*, 12(17), pp.1–11.
- EDDE, M., HOUDE, F., THEAUD, G., DUMONT, M., GILBERT, G., HOUDE, J.C., MALTAIS, L., THÉBERGE, A., DOUMBIA, M., BEAUDOIN, A.M., LAPOINTE, E., BARAKOVIC, M., MAGON, S. AND DESCOTEAUX, M., 2023. Impact of follow ups, time interval and study duration in diffusion & myelin MRI clinical study in MS. *NeuroImage: Clinical*, 40(1), pp.1–13.
- EKAPUTRIE, A. AND FAUZAN, A., 2023. Analysis of the Character of the Insured by Policy Holders of Private Insurance Companies in Indonesia with the Rough Set Method. *EKSAKTA: Journal of Sciences and Data Analysis*, 4(2), pp.8–14.
- EL-SHEIKH, S., KANDIL, S. AND SHALIL, S., 2023. Increasing and Decreasing Soft Rough Set Approximations. *International Journal of Fuzzy Logic and Intelligent Systems*, 23(4), pp.425–435.
- FAN, X., MAO, X., CAI, T., SUN, Y., GU, P. AND JU, H., 2023. Sensor data reduction with novel local neighborhood information granularity and rough set approach. *Frontiers in Physics*, 11(1), pp.1–7.
- FEELY, A., WALL-WIELER, E., ROOS, L.L. AND LIX, L.M., 2020. Effect of study duration and outcome measurement frequency on estimates of change for longitudinal cohort studies in routinely-collected administrative data. *International Journal of Population Data Science*, 5(1), pp.1–8.
- GHOLAMI, A., SHEIKH, R., MIZANI, N. AND SANA, S.S., 2018. ABC analysis of the customers using axiomatic design and incomplete rough set. *RAIRO - Operations Research*, 52(4), pp.1219–1232.
- GIGIH, P., ANDRIANI, C., ISTIADI, I. AND ABD WAHAB, M.H., 2020. Case Service System at the Child Welfare Institution using the Case-Based Reasoning Method. *International Journal of Electrical, Energy and Power System Engineering*, 3(2), pp.40–45.
- JONES, A. AND HARDIYANTI, C., 2021. Case Based Reasoning using K-Nearest Neighbor with Euclidean Distance for Early Diagnosis of Personality Disorder. *IJISTECH (International Journal of Information System & Technology)*, 5(1), p.23.
- LIU, X., LORINI, E., ROTOLO, A. AND SARTOR, G., 2022. Modelling and Explaining Legal Case-Based Reasoners Through Classifiers. *Frontiers in Artificial Intelligence and Applications*, 362, pp.83–92.
- MACHIDA, R., FUJII, Y. AND SOZU, T., 2021. Predicting study duration in clinical trials with a time-to-event endpoint. *Statistics in Medicine*, 40(10), pp.2413–2421.
- MINARNI, M., WARMAN, I., YUHENDRA, Y. AND HANDAYANI, W., 2018. Comparison of case-based reasoning and dempster shafer on expert system of cassava disease identification. In: *MATEC Web of Conferences*. pp.1–5.
- MORRIS, K.A., HORNUM, S., CRYSTAL-ORNELAS, R., PENNINGTON, S.C. AND BOND-LAMBERTY, B., 2022. Soil Respiration Response to Simulated Precipitation Change Depends on Ecosystem Type and Study Duration. *Journal of Geophysical Research: Biogeosciences*, 127(11), pp.1–14.
- NAKHJIRI, N., SALAMÓ, M. AND SÀNCHEZ-MARRÈ, M., 2020. Reputation-Based Maintenance in Case-Based Reasoning. *Knowledge-Based Systems*, 193.
- NURFALINDA, N. AND UPERATI, A., 2020. Case Based Reasoning untuk Diagnosis Penyakit Ikan Kakap Putih. *Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan*, 9(1), pp.45–50.
- PARUCHURI, V.A. AND GRANVILLE, B.C., 2020. A Case-Based Reasoning System for Aiding Physicians in Decision Making. *Intelligent Information Management*, 12(02), pp.63–74.
- SAFITRI, N.A., RAHMATULLAH, S., DIRGANTARA, Y. AND FARIDA, I., 2018. Using Case-Based Reasoning To Solve Water Purification Problems. pp.107–110.
- YANG, X., WANG, X. AND KANG, J., 2022. Multi-granularity decision rough set attribute reduction algorithm under quantum particle swarm optimization. *IET Cyber-systems and Robotics*, 4(1), pp.25–37.
- YE, S., 2024. Research on case-based reasoning for determining maintenance intervals of civil aircraft systems. *Journal of Physics: Conference Series*, 2764(1), pp.6–11.

- ZENG, J., JING, G., ZHU, Q. AND SUN, H., 2023. Emergency Response Plan for Spontaneous Combustion Based on Case-Based Reasoning. *Processes*, 11(7), pp.1–14.
- ZULFAN, Z., PONCORINI, E. AND MARYANI, M., 2023. Relationship Between Gender, Study Duration and Burnout Among Medical Students. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, 12(4), p.465.