

MULTI-CRITERIA DECISION MAKING DALAM SELEKSI FITUR ENSEMBLE UNTUK PREDIKSI CACAT PERANGKAT LUNAK

Muhammad Fikri¹, Rudy Herteno^{*2}, Radityo Adi Nugroho³, Setyo Wahyu Saputro⁴, Friska Abadi⁵

^{1,2,3,4,5}Universitas Lambung Mangkurat, Kalimantan Selatan

Email: ¹2111016310018@mhs.ulm.ac.id, ²rudy.herteno@ulm.ac.id, ³radityo.adi@ulm.ac.id,

⁴setyo.saputro@ulm.ac.id, ⁵friska.abadi@ulm.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 20 Maret 2025, diterima untuk diterbitkan: 10 Desember 2025)

Abstrak

Prediksi cacat perangkat lunak merupakan upaya strategis dalam meningkatkan kualitas produk melalui identifikasi dini modul yang berpotensi cacat. Kinerja prediksi dipengaruhi oleh pemilihan fitur, karena informasi yang berlebihan dan tidak relevan dapat mempengaruhi kualitas pembelajaran model. Seleksi fitur ensemble dinilai efektif dalam menyeleksi fitur yang relevan dengan menggabungkan beberapa metode seleksi fitur berbasis filter. Diperlukan mekanisme integrasi untuk menyatukan hasil dari empat teknik filter—Mutual Information, Fisher Score, Uncertainty dan Relief. Penelitian ini membandingkan empat metode Multi-Criteria Decision Making—TOPSIS, VIKOR, EDAS, dan WASPAS—yang bekerja dengan meranking nilai relevansi fitur hasil seleksi filter tersebut. Sepuluh fitur teratas dari tiap metode kemudian dievaluasi menggunakan model Random Forest dengan metrik AUC melalui K-Fold cross-validation. Dari 12 dataset NASA MDP yang diuji, TOPSIS menunjukkan kinerja paling konsisten dan terbaik dengan nilai rata-rata AUC sebesar 0,8038. Temuan ini menegaskan pentingnya pemilihan metode integrasi yang tepat dalam meningkatkan akurasi prediksi cacat perangkat lunak dan memberikan panduan bagi pengembangan model yang lebih efektif.

Kata kunci: *prediksi cacat perangkat lunak, seleksi fitur ensemble, MCDM, random forest*

MULTI-CRITERIA DECISION MAKING IN ENSEMBLE FEATURE SELECTION FOR SOFTWARE DEFECT PREDICTION

Abstract

Software defect prediction is a strategic effort to improve product quality through early identification of potentially defective modules. Prediction performance is influenced by feature selection, because redundant and irrelevant information can affect the quality of model learning. Ensemble feature selection is considered effective in selecting relevant features by combining several filter-based feature selection methods. An integration mechanism is needed to unify the results of four filter techniques—Mutual Information, Fisher Score, Uncertainty and Relief. This study compares four Multi-Criteria Decision Making methods—TOPSIS, VIKOR, EDAS, and WASPAS—which work by ranking the relevance values of the filter-selected features. The top ten features from each method are then evaluated using the Random Forest model with the AUC metric through K-Fold cross-validation. Of the 12 NASA MDP datasets tested, TOPSIS showed the most consistent and best performance with an average AUC value of 0.8038. These findings emphasize the importance of choosing the right integration method in improving the accuracy of software defect prediction and provide guidance for the development of more effective models.

Keywords: *software defect prediction, ensemble feature selection, MCDM, random forest*

1. PENDAHULUAN

Pengembangan perangkat lunak telah berkembang secara signifikan, ditandai oleh kompleksitas pengkodean dan implementasi yang semakin tinggi sehingga menuntut perhatian cermat untuk memastikan hasil yang bebas dari cacat (Nabella, 2024). Cacat perangkat lunak secara umum didefinisikan sebagai penyimpangan dari spesifikasi

atau persyaratan yang seharusnya dimiliki oleh perangkat lunak, cacat ini dapat menyebabkan perangkat lunak gagal menjalankan fungsi yang biasanya dilaporkan selama fase pengujian perangkat lunak (Ramadhani, 2024). Dalam konteks ini, Software Defect Prediction (SDP) merupakan tugas yang sangat krusial dalam tahap pengujian perangkat lunak karena memiliki banyak tantangan untuk

mengidentifikasi modul-modul yang rentan terhadap cacat (Ghinaya, 2024).

Sebagai contoh, *Metric Data Program* (MDP) oleh NASA merupakan dataset yang memuat beragam karakteristik perangkat lunak, termasuk variasi dalam ukuran kode, kompleksitas, dan fungsionalitas, sehingga memberikan gambaran menyeluruh mengenai tantangan dalam pengembangan perangkat lunak (Akbar, 2024). Namun, perlu diperhatikan bahwa dataset ini memiliki sejumlah keterbatasan, seperti keberadaan atribut yang tidak relevan, tingginya dimensi data, serta distribusi kelas yang tidak seimbang, yang dapat memengaruhi performa model prediksi (Rahman, 2024). Menurut Frieyadie et al. (2024), kualitas data dalam dataset NASA MDP berperan signifikan dalam menentukan efektivitas model SDP, terutama karena tingginya jumlah fitur dan ketidakseimbangan distribusi kelas yang dapat mempersulit proses pembelajaran model.

Permasalahan ini mempertegas pentingnya seleksi fitur sebagai langkah prasyarat dalam membangun model SDP yang efektif. Seleksi fitur tidak hanya mengurangi dimensi data dengan mempertahankan fitur informatif, tetapi juga meminimalkan risiko overfitting dan meningkatkan interpretabilitas model (Malekipirbazari, 2021). Dalam penelitian ini, pendekatan ensemble diterapkan dengan mengombinasikan empat metode seleksi fitur berbasis filter—ReliefF, Mutual Information, Fisher Score, dan Uncertainty—metode-metode yang menurut studi Hashemi (2020) menilai setiap fitur secara independen dengan kriteria statistik tertentu. Pendekatan ini bertujuan memanfaatkan kekuatan masing-masing teknik dan mengurangi sensitivitas terhadap kelemahan yang dimiliki metode tunggal.

Namun, metode seleksi fitur konvensional seperti filter-based approaches memiliki kelemahan mendasar. Seperti diungkapkan Hamid (2021), metode filter cenderung mengabaikan interaksi antar fitur karena evaluasinya dilakukan secara individual, sehingga tidak mampu menghasilkan jumlah fitur optimal yang relevan untuk klasifikasi.

Untuk mengatasi keterbatasan ini, ensemble feature selection diusulkan sebagai solusi. Bayne (2025) menuliskan bahwa pendekatan ensemble mampu mengurangi bias yang didapatkan dari satu metode tertentu dengan meintegrasikan hasil dari beberapa metode seleksi fitur. Ide ini sejalan dengan pernyataan Hashemi (2021) yang mengutip prinsip "Two heads are better than one", di mana variasi metode dalam ensemble menciptakan sinergi untuk menghasilkan subset fitur lebih baik.

Namun, kombinasi hasil dari berbagai filter dalam ensemble memerlukan mekanisme integrasi yang sistematis. Di sinilah Multi-Criteria Decision Making (MCDM) berperan penting. Seperti dijelaskan Hashemi (2020), MCDM merupakan alat yang ampuh untuk menentukan preferensi

berdasarkan beberapa kriteria atau pendapat. memperkuat argumen ini dengan menyatakan bahwa MCDM tidak hanya membantu memilih opsi terbaik, tetapi juga menyediakan pendekatan metodologis untuk mengelola konflik dalam pengambilan keputusan kompleks. Sehingga, metode ini cocok digunakan sebagai kerangka integrasi dalam ensemble.

MCDM menawarkan beragam metode dengan pendekatan yang berbeda-beda, masing-masing memiliki keunggulan tersendiri dalam menyelesaikan masalah multi-kriteria. Penelitian terdahulu telah mengimplementasikan beberapa metode MCDM dan memberikan pernyataan mengenai metode yang digunakannya. Seperti, Kaur (2023) pada penelitiannya menunjukkan bahwa TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) mampu memilih fitur yang sesuai sehingga meningkatkan performa prediksi. Sementara itu, Hashemi (2021) menerapkan VIKOR dalam penelitiannya dan menemukan bahwa metode ini mampu menghasilkan accuracy, F-score, dan algorithm run-time yang lebih efektif dibanding metode lainnya. Di sisi lain, Abellana (2022) mengatakan bahwa EDAS (Evaluation based on Distance from Average Solution) adalah salah satu metode yang akan yang dikenal dalam konteks ensemble. Selain itu, Alves (2024) mengatakan WASPAS (Weighted Aggregated Sum Product Assessment) merupakan kombinasi dari WSM (Weighted Sum Method) dan WPM (Weighted Product Method) yang efektif dalam situasi di mana bobot kriteria diketahui atau dapat diperkirakan secara akurat.

Namun, efektivitas masing-masing metode MCDM tersebut belum dievaluasi secara komparatif dalam konteks spesifik Software Defect Prediction (SDP), terutama pada dataset seperti NASA MDP. Sehingga perlu dilakukan pengujian untuk mengetahui metode mana yang mampu menyeleksi fitur yang paling relevan. Pengujian dilakukan dengan menggunakan metode Random Forest sebagai metode pengklasifikasi. Random Forest dipilih untuk karena keunggulannya dalam menangani data dengan banyak fitur, mencegah overfitting, dan menghasilkan prediksi yang stabil bahkan pada dataset tidak seimbang (Salman et al., 2024).

Selanjutnya, hasil pengklasifikasian dinilai berdasarkan nilai Area Under Curve (AUC). Metode yang menghasilkan nilai AUC yang lebih tinggi dapat diartikan sebagai metode yang lebih efektif untuk menentukan fitur sesuai dengan relevansinya. Hal ini sejalan dengan pernyataan Çorbacioğlu & Aksel (2023) yang menyatakan bahwa nilai AUC yang mendekati 1 menunjukkan kemampuan model yang lebih baik dalam membedakan antara kelas positif dan negatif. Selain itu, penelitian sebelumnya juga merekomendasikan AUC sebagai metrik evaluasi

yang lebih sesuai untuk dataset dengan distribusi kelas yang tidak seimbang. (Pasha, 2020)

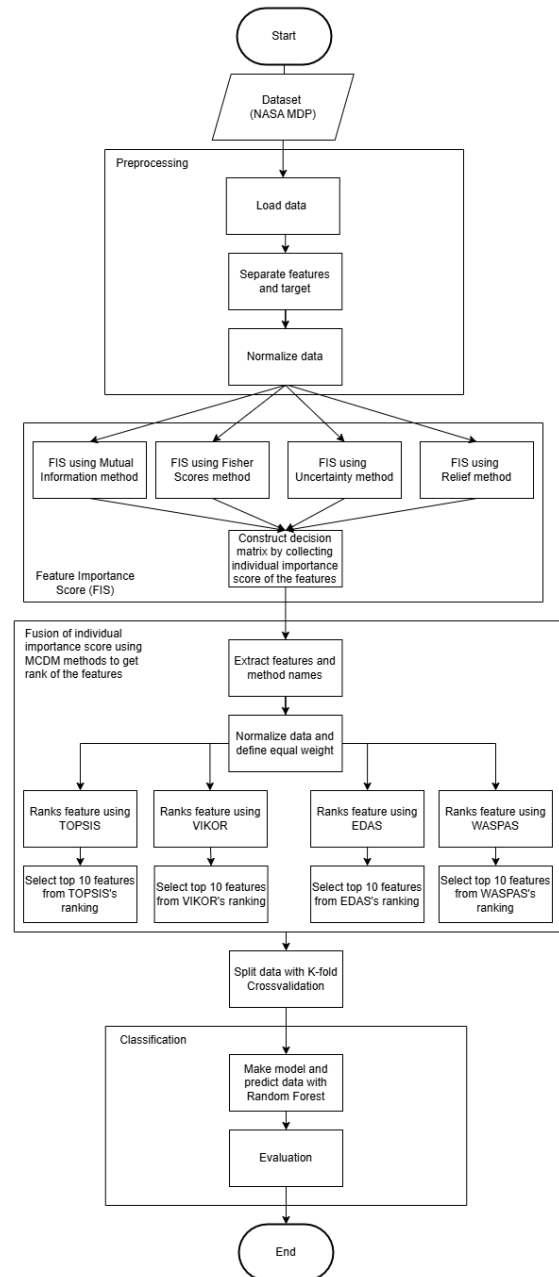
Beberapa penelitian sebelumnya telah menerapkan pendekatan ensemble feature selection (EFS) yang dikombinasikan dengan metode Multi-Criteria Decision Making (MCDM) untuk meningkatkan kinerja prediksi dalam berbagai domain, termasuk prediksi cacat perangkat lunak. Kaur (2023) menggunakan pendekatan MCDM-EFS dengan empat metode seleksi fitur (Chi-squared, Korelasi, Information Gain, dan Random Forest) yang digabungkan melalui TOPSIS dan diuji pada dataset NASA MDP dan Mylyn, sehingga menunjukkan peningkatan performa prediksi cacat hingga 16,29%. Selanjutnya, Hashemi (2020) mengusulkan metode MFS-MCDM berbasis TOPSIS untuk seleksi fitur multi-label dan mengujinya pada dataset Mulan dan Meka, yang menunjukkan keunggulan metode ini dibandingkan metode lain dengan skor tertinggi dalam uji Friedman. Kemudian, Hashemi (2021) mengembangkan EFS-MCDM berbasis VIKOR untuk seleksi fitur dengan lima filter berbeda, yang dilaporkan meningkatkan akurasi klasifikasi pada 8 dari 10 dataset berdimensi tinggi. Selain itu, Hamid (2021) memperkenalkan teknik Ensemble-PSO-SVM yang memadukan empat filter (Information Gain, Gain Ratio, Chi-squared, dan Relief-F) dengan optimasi PSO, sehingga mencapai akurasi hingga 96,15% pada dataset kanker. Abellana (2022) menggunakan metode EDAS dalam ensemble seleksi fitur untuk klasifikasi biner dan menunjukkan bahwa algoritme tersebut mampu mengurangi ukuran dataset tanpa mengorbankan kinerja klasifikasi. Temuan-temuan tersebut menunjukkan potensi besar metode MCDM dalam proses seleksi fitur ensemble. Namun, perbandingan langsung antara berbagai metode MCDM dalam konteks spesifik prediksi cacat perangkat lunak, khususnya pada dataset seperti NASA MDP, masih jarang dilakukan dan perlu diteliti lebih lanjut.

2. METODE PENELITIAN

Langkah-langkah pada penelitian ini terdiri dari beberapa tahap dan dapat ditunjukkan oleh Gambar 1.

2.1 Dataset

Penelitian ini menggunakan data *Metric Data Program* (MDP) yang disediakan oleh *National Aeronautics and Space Administration* (NASA) atau yang biasa dikenal dengan akronim NASA MDP. Dataset NASA MDP yang digunakan adalah dataset versi D'' yang dapat diakses melalui [Github|NASADefectDataset](#). Dataset tersebut terbagi menjadi 12 dataset yang masing-masing memiliki jumlah fitur dan instance yang berbeda, seperti yang ditunjukkan dalam Table 1.



Gambar 1. Langkah-langkah Penelitian

Tabel 1 Dataset NASA MDP D''

Dataset	Number of features	Amount of data
CM1	38	327
JM1	22	7720
KC1	22	1162
KC3	40	194
MC1	39	1952
MC2	40	124
MW1	38	250
PC1	38	679
PC2	37	722
PC3	38	1053
PC4	38	1270
PC5	39	1694

2.2 Pre-processing

Pada tahap ini, dilakukan serangkaian proses untuk mempersiapkan data sebelum digunakan pada tahap selanjutnya. Pertama-tama data yang telah dimuat dari dataset akan dipisahkan menjadi data fitur dan label. Data label “Defective” pada dataset yang awalnya berupa “N” dan “Y” diubah ke dalam bentuk numerik berupa 0 dan 1 agar bisa digunakan tahap-tahap berikutnya. Selanjutnya, data fitur dilakukan normalisasi dengan menggunakan metode normalisasi z-score agar setiap fitur memiliki skala yang seragam.

2.3 Seleksi Fitur

Penyeleksian fitur adalah salah satu langkah penting yang menjawab tantangan kumpulan data berdimensi tinggi. Tantangan ini melibatkan dua aspek: variabel yang tidak relevan, di mana fitur tertentu (variabel independen) tidak memengaruhi fitur target (variabel dependen), dan variabel redundan, di mana variabel independen menunjukkan korelasi tinggi dan dapat dihilangkan. Metode pemilihan fitur telah menjadi lebih menonjol karena kemampuannya untuk meningkatkan akurasi prediksi dan mengurangi waktu pembuatan model (Ali, 2023).

Pada tahap ini, penyeleksian fitur tidak dilakukan secara langsung. Melainkan dengan memberikan nilai relevansi, fitur diurutkan berdasarkan seberapa berpengaruh fitur tersebut. Penilaian relevansi fitur dilakukan dengan menggunakan 4 metode seleksi fitur berbasis filter untuk menilai fitur mana yang paling relevan. Metode yang digunakan pada tahap ini adalah Mutual Information, Fisher Score, Uncertainty dan Relief yang sebelumnya juga digunakan oleh Hashemi (2020) pada penelitiannya mengenai seleksi fitur ensemble. Fitur yang redundan diberikan nilai yang lebih rendah dan fitur yang paling relevan diberikan nilai yang lebih tinggi.

2.4 MCDM

Multi-Criteria Decision Making (MCDM) adalah salah satu metode pengambilan keputusan yang bertujuan untuk menentukan alternatif terbaik dengan mempertimbangkan lebih dari satu kriteria dalam proses seleksi. MCDM dikategorikan ke dalam dua subkategori umum, yaitu *multi-attribute decision making* (MADM) dan *multi-objective decision making* (MODM). MODM berfokus pada ruang keputusan berkelanjutan dengan jumlah alternatif yang tak terbatas dan juga dikenal sebagai masalah pengambilan keputusan berkelanjutan. Sedangkan, MADM dikenal sebagai masalah diskrit dan berfokus pada masalah dengan alternatif keputusan yang diketahui secara eksplisit dengan jumlah terbatas. (Taherdoost & Madanchian, 2023)

Pada tahap ini, metode MCDM yang digunakan masuk ke kategori masalah diskrit. MCDM digunakan untuk menentukan fitur terbaik yang

dipertimbangkan menggunakan nilai relevansi sebagai kriteria. Fitur yang menjadi opsi terbaik diurutkan menjadi sebuah ranking. Hal ini sesuai dengan pernyataan Abellana (2022), kelebihan utama MCDM terletak pada kemampuannya menyajikan pendekatan terorganisir, baik untuk memilih opsi terbaik maupun meranking seluruh alternatif secara hierarkis. Adapun metode MCDM yang digunakan adalah TOPSIS, VIKOR, EDAS dan WASPAS.

2.4.1 TOPSIS

Proses dalam metode TOPSIS terdiri dari beberapa langkah utama. Pertama, dilakukan normalisasi data untuk memastikan bahwa semua fitur memiliki skala yang seragam. Kedua, bobot diberikan pada setiap fitur sesuai dengan tingkat kepentingannya. Selanjutnya, solusi ideal positif dan negatif dihitung berdasarkan nilai terbaik dan terburuk dari setiap fitur. Setelah itu, jarak antara setiap alternatif dengan solusi ideal positif (D_i^+) dan solusi ideal negatif (D_i^-) dihitung menggunakan rumus Euclidean Distance pada persamaan (2) dan (3).

$$D_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2} \quad (2)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2} \quad (3)$$

Di mana D_i^+ adalah jarak ke solusi ideal positif dan D_i^- adalah jarak ke solusi ideal negatif. Langkah terakhir adalah menghitung skor preferensi C_i menggunakan rumus pada persamaan (4).

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (4)$$

Nilai C_i berada dalam rentang 0 hingga 1, di mana fitur dengan nilai tertinggi dianggap paling relevan diantara fitur lainnya (Janane, Ouaderhman, & Chamlal, 2023).

2.4.2 VIKOR

Proses perhitungan dalam VIKOR melibatkan dua ukuran utama, yaitu jarak agregat ke solusi ideal terbaik (S_i) dan jarak maksimum ke solusi terburuk (R_i), yang didefinisikan pada persamaan (5) dan (6).

$$S_i = \sum_{j=1}^n w_j \frac{(f_j^+ - f_{ij})}{(f_j^+ - f_j^-)} \quad (5)$$

$$R_i = \max_j \left[w_j \frac{(f_j^+ - f_{ij})}{(f_j^+ - f_j^-)} \right] \quad (6)$$

Di mana w_j adalah bobot kriteria, f_{ij} adalah nilai alternatif i pada kriteria j , dan f_j^+ serta f_j^- masing-masing adalah nilai terbaik dan terburuk untuk

kriteria j . Setelah itu, nilai indeks VIKOR (Q_i) dihitung dengan rumus pada persamaan (7).

$$Q_i = v \frac{S_i - S^*}{S^- - S^*} + (1 - v) \frac{R_i - R^*}{R^- - R^*} \quad (7)$$

Di mana v adalah parameter keseimbangan kompromi (biasanya $v=0.5$). Alternatif dengan nilai Q_i terendah dipilih sebagai solusi terbaik (Silva et al., 2024).

2.4.3 EDAS

Proses perhitungan EDAS terdiri dari beberapa langkah utama. Pertama, nilai rata-rata (A_j) dari setiap kriteria ditentukan berdasarkan semua alternatif. Selanjutnya, dihitung nilai PDA dan NDA menggunakan rumus pada persamaan (8) dan (9).

$$PDA_{ij} = \max\left(\frac{X_{ij} - A_j}{A_j}, 0\right) \quad (8)$$

$$NDA_{ij} = \max\left(\frac{A_j - X_{ij}}{A_j}, 0\right) \quad (9)$$

Di mana X_{ij} adalah nilai alternatif i pada kriteria j , dan A_j adalah solusi rata-rata untuk kriteria j . Nilai akhir atau disebut AS_i (Assessment Score) dari setiap alternatif dihitung dengan agregasi bobot kriteria menggunakan rumus pada persamaan (10)

$$AS_i = 0.5 \left(\frac{SP_i}{\max(SP)} + 1 - \frac{SN_i}{\max(SN)} \right) \quad (10)$$

Di mana SP_i dan SN_i masing-masing merupakan agregasi dari PDA dan NDA berdasarkan bobot kriteria yang telah ditentukan. Nilai AS_i tertinggi menunjukkan alternatif terbaik dalam pemilihan (Silva et al., 2024).

2.4.4 WASPAS

Proses perhitungan WASPAS diawali dengan normalisasi data agar semua nilai berada dalam skala yang seragam. Selanjutnya, dua skor utama dihitung nilai WSM (Q_1) dan nilai WPM (Q_2), dengan menggunakan rumus pada persamaan (11) dan (12).

$$Q_{1i} = \sum_{j=1}^n w_j \cdot X_{ij} \quad (11)$$

$$Q_{2i} = \prod_{j=1}^n X_{ij}^{w_j} \quad (12)$$

Di mana X_{ij} adalah nilai normalisasi alternatif i pada kriteria j , dan w_j adalah bobot kriteria. Setelah itu, kedua skor ini digabungkan menggunakan parameter λ dengan menggunakan rumus pada persamaan (13).

$$Q = \lambda Q_{1i} + (1 - \lambda) Q_{2i} \quad (13)$$

Nilai λ biasanya diatur sebesar 0.5 untuk menyeimbangkan keunggulan dari WSM dan WPM.

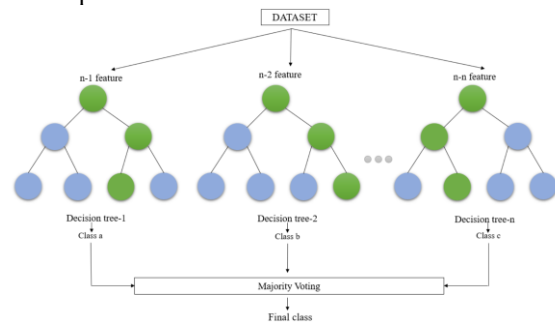
Alternatif dengan nilai Q_i tertinggi dianggap sebagai alternatif terbaik. (Silva et al., 2024).

2.5 K-Fold Cross-Validation

Cross-validation adalah salah satu teknik yang berperan penting dalam pembelajaran mesin untuk mengevaluasi performa model untuk menghindari overfitting, dan memastikan generalisasi yang baik. Salah satu metode yang umum digunakan adalah K-fold cross-validation, di mana dataset dibagi menjadi K subset atau "fold". Model dilatih pada K-1 fold dan diuji pada fold tersisa, dengan proses ini diulang K kali. Rata-rata hasil dari semua iterasi digunakan sebagai evaluasi akhir (Barry, 2023).

2.6 Random Forest

Random Forest (RF), yang diperkenalkan oleh Breiman (2001), adalah algoritma pembelajaran ensemble yang menggabungkan beberapa pohon prediktor. Setiap pohon bergantung pada nilai vektor acak yang diambil secara independen dan dengan distribusi yang sama yang diterapkan secara seragam pada semua pohon di hutan. Algoritma ini bekerja dengan membuat kumpulan pohon keputusan secara acak dan menggunakan hasil mayoritas dari setiap pohon untuk menghasilkan prediksi akhir (Kibria & Matin, 2022). Ilustrasi dari cara Random Forest dapat dilihat pada Gambar 2.



Gambar 2. Random Forest

2.7 Area Under Curve

AUC (Area Under Curve) adalah metrik evaluasi model klasifikasi yang diukur dari luas di bawah kurva ROC (Receiver Operating Characteristic), di mana sumbu X menunjukkan False Positive Rate (FPR) dan sumbu Y True Positive Rate (TPR). Nilai AUC berkisar 0–1: mendekati 1 menandakan kemampuan model membedakan kelas positif dan negatif dengan baik, sedangkan 0,5 setara dengan tebakan acak. Umumnya, $AUC > 0,8$ dianggap baik untuk analisis lebih lanjut (Krzywicka & Wosiak, 2023; Çorbacıoğlu & Aksel, 2023).

Tabel 2 Sample Dataset Setelah Preprocessing

No	LOC_BLANK	BRANCH_COUNT	CALL_PAIRS	LOC_COMMENTS	...	LOC_TOTAL	Defective
1	-0,0933	-0,0243	-0,2490	1,3414	...	-0,3028	0
2	-0,3682	-0,2812	-0,6933	-0,3910	...	-0,0324	0
3	-0,0933	0,1042	0,6396	0,2689	...	-0,1526	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1952	-0,5057	-0,2812	-0,4712	-0,3910	...	-0,4831	0

Tabel 3. Ranking Fitur Berdasarkan MCDM

No	TOPSIS	VIKOR	EDAS	WASPAS
1	PERCENT_COMMENTS	PERCENT_COMMENTS	PERCENT_COMMENTS	PERCENT_COMMENTS
2	HALSTEAD_CONTENT	LOC_COMMENTS	LOC_CODE_AND_COMMENT	LOC_CODE_AND_COMMENT
3	LOC_COMMENTS	HALSTEAD_CONTENT	LOC_COMMENTS	NUMBER_OF_LINES
4	LOC_CODE_AND_COMMENT	LOC_CODE_AND_COMMENT	HALSTEAD_CONTENT	LOC_COMMENTS
5	CYCLOMATIC_DENSITY	NUMBER_OF_LINES	NUMBER_OF_LINES	HALSTEAD_CONTENT
6	NUMBER_OF_LINES	LOC_BLANK	CALL_PAIRS	MODIFIED_CONDITION_COUNT
7	MODIFIED_CONDITION_COUNT	CALL_PAIRS	LOC_BLANK	NUM_OPERATORS
8	MAINTENANCE_SEVERITY	NUM_UNIQUE_OPERANDS	CYCLOMATIC_DENSITY	HALSTEAD_VOLUME
9	NUM_OPERATORS	MODIFIED_CONDITION_COUNT	MODIFIED_CONDITION_COUNT	LOC_BLANK
10	LOC_EXECUTABLE	CYCLOMATIC_DENSITY	NUM_OPERATORS	CALL_PAIRS

AUC menjadi metrik yang populer dalam berbagai tugas klasifikasi, terutama pada dataset yang memiliki ketidakseimbangan kelas. Pada dataset seperti ini, di mana jumlah sampel kelas mayoritas jauh lebih banyak dibandingkan dengan kelas minoritas, metrik seperti accuracy cenderung memberikan hasil yang bias terhadap kelas yang dominan. Oleh karena itu, AUC lebih disarankan karena mampu mengevaluasi performa model di berbagai tingkat ambang batas klasifikasi (Pasha, 2020).

3. HASIL PEMBAHASAN

Penelitian ini menggunakan versi D" dari kumpulan data NASA MDP, yang terdiri atas 12 file berisi berbagai fitur dan label (cacat dan tidak cacat). Setiap berkas diubah menjadi format tabular dan dinormalisasi untuk memastikan keseragaman skala sebelum proses selanjutnya. Data yang telah siap digunakan dapat dilihat pada Tabel 2.

Pada tahapan berikutnya, setiap fitur dinilai menggunakan empat teknik filter—Mutual Information, Fisher Score, Uncertainty, dan Relief—sehingga terbentuk empat skor kepentingan (Feature Importance Score/FIS) per fitur. Keempat kumpulan skor ini disusun menjadi matriks keputusan berukuran $m \times 4$, di mana m adalah jumlah fitur asli. Matriks tersebut dapat dilihat pada persamaan (14) dan (15).

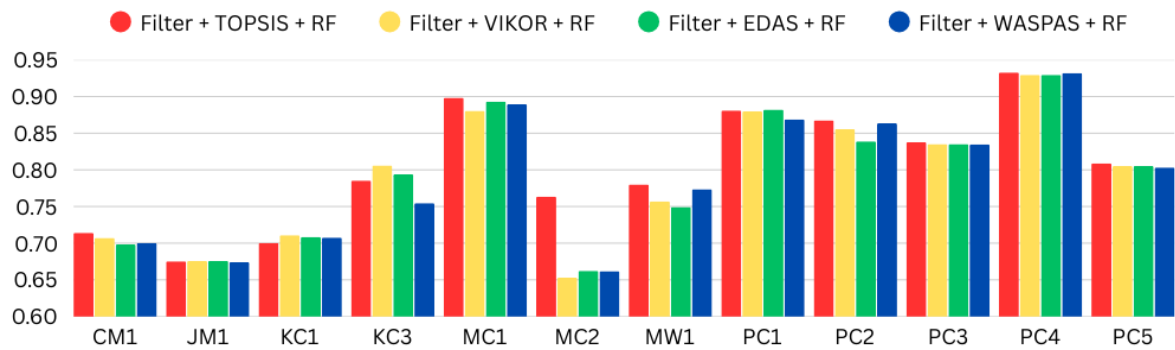
$$\begin{matrix}
 & FIS1 & FIS2 & FIS3 & FIS4 \\
 \begin{bmatrix} 0,0135 \\ 0,1336 \\ \vdots \\ 1,0000 \end{bmatrix} & \begin{bmatrix} 0,7550 \\ 0,4741 \\ \vdots \\ 0,1950 \end{bmatrix} & \begin{bmatrix} 0,0049 \\ 0,0046 \\ \vdots \\ 0,0165 \end{bmatrix} & \begin{bmatrix} 0,0099 \\ 0,0304 \\ \vdots \\ 0,0170 \end{bmatrix}
 \end{matrix} \quad (14)$$

$$\begin{bmatrix} 0,0135 & 0,7550 & 0,0049 & 0,0099 \\ 0,1336 & 0,4741 & 0,0046 & 0,0304 \\ \vdots & \vdots & \vdots & \vdots \\ 1,0000 & 0,1950 & 0,0165 & 0,0170 \end{bmatrix} \quad (15)$$

Matriks keputusan tersebut kemudian diproses melalui empat method Multi-Criteria Decision Making (MCDM): TOPSIS, VIKOR, EDAS, dan WASPAS. Masing-masing metode menghasilkan peringkat fitur independen. Tabel 3 menyajikan data peringkat sepuluh fitur teratas dari setiap metode. Dari setiap peringkat, dipilih sepuluh fitur teratas, sehingga terbentuk empat sub-dataset yang siap untuk digunakan pada tahap pembuatan model prediksi.

Setiap sub-dataset selanjutnya diolah secara mandiri: dilakukan split data menggunakan K-Fold cross-validation, kemudian model prediksi dibangun dengan menggunakan metode klasifikasi Random Forest, dan hasil prediksi dievaluasi dengan metrik AUC. Proses split–training–evaluasi ini diulang empat kali sehingga diperoleh distribusi skor AUC untuk masing-masing metode MCDM. Hasil performa masing-masing metode MCDM pada 12 dataset disajikan dalam Tabel 4.

Nilai AUC yang dihasilkan oleh empat metode yang dibandingkan dalam Tabel 4, di mana setiap hasil ditampilkan dalam kolom yang berbeda. yang digunakan ditampilkan di kolom pertama, dan nilai AUC yang diperoleh dari klasifikasi Random Forest dengan penyeleksian fitur berbasis filter yang diintegrasikan dengan TOPSIS, VIKOR, EDAS, dan WASPAS ditampilkan di kolom kedua hingga kelima secara berurutan. Data nilai AUC tertinggi ditandai dengan cetak tebal.



Gambar 3. Diagram Nilai AUC

Tabel 4. Nilai AUC

Dataset	Filter + TOPSIS + RF	Filter + VIKOR + RF	Filter + EDAS + RF	Filter + WASPAS + RF
CM1	0,7142	0,7072	0,6982	0,7003
JM1	0,6754	0,6756	0,6758	0,6741
KC1	0,7003	0,7110	0,7084	0,7075
KC3	0,7853	0,8061	0,7940	0,7547
MC1	0,8986	0,8804	0,8932	0,8896
MC2	0,7634	0,6530	0,6621	0,6619
MW1	0,7801	0,7572	0,7493	0,7736
PC1	0,8809	0,8797	0,8823	0,8690
PC2	0,8678	0,8556	0,8389	0,8635
PC3	0,8378	0,8354	0,8351	0,8347
PC4	0,9328	0,9296	0,9296	0,9322
PC5	0,8090	0,8050	0,8053	0,8032

Berdasarkan data yang terdapat pada Tabel 4, model yang menggunakan metode TOPSIS secara konsisten menghasilkan nilai AUC paling tinggi pada sebagian besar dataset, yaitu di 8 dari total 12 dataset yang diujikan. Khususnya pada dataset CM1, MC1, MC2, MW1, PC2, PC3, PC4, dan PC5, TOPSIS menampilkan performa terbaik. Namun, terdapat empat dataset—JM1, KC1, KC3, dan PC1—di mana metode lain memiliki AUC yang lebih tinggi. Sehingga, VIKOR memiliki performa tertinggi di 2 dataset. EDAS di 2 dataset dan WASPAS di 0 dataset. Adapun penyajian data berupa diagram dapat dilihat pada Gambar 3.

Tabel 5. Selisih Nilai AUC

Dataset	Filter + TOPSIS + RF	Filter + VIKOR + RF	Filter + EDAS + RF	Filter + WASPAS + RF
CM1	0,0000	-0,0070	-0,0160	-0,0139
JM1	-0,0004	-0,0002	0,0000	-0,0017
KC1	-0,0107	0,0000	-0,0026	-0,0035
KC3	-0,0208	0,0000	-0,0122	-0,0515
MC1	0,0000	-0,0182	-0,0053	-0,0090
MC2	0,0000	-0,1104	-0,1012	-0,1015
MW1	0,0000	-0,0229	-0,0308	-0,0065
PC1	-0,0015	-0,0026	0,0000	-0,0133
PC2	0,0000	-0,0122	-0,0288	-0,0043
PC3	0,0000	-0,0024	-0,0027	-0,0031
PC4	0,0000	-0,0033	-0,0032	-0,0007
PC5	0,0000	-0,0040	-0,0037	-0,0058

Tabel 5 menunjukkan selisih nilai AUC, dihitung dengan mengurangi nilai AUC masing-masing metode dengan nilai AUC tertinggi yang dicapai pada tiap dataset. Selisih nilai AUC yang lebih besar menunjukkan bahwa kinerja metode tersebut semakin jauh dari nilai optimal yang dicapai oleh metode terbaik di tiap dataset. Susunan kolom pada Tabel 5 sama dengan Tabel 4. Data selisih nilai AUC terbesar ditandai dengan cetak tebal.

Berdasarkan data yang terdapat pada Tabel 5, metode WASPAS memiliki selisih terbesar di 5 dari 12 dataset, mengindikasikan bahwa performanya cenderung lebih rendah dibandingkan metode lain. Sementara itu, baik metode EDAS maupun VIKOR masing-masing menunjukkan selisih terbesar pada 3 dataset, menempatkan keduanya di posisi menengah. Di sisi lain, metode TOPSIS hanya memiliki satu dataset dengan selisih terbesar, yang mengonfirmasi bahwa metode ini secara konsisten mendekati nilai AUC tertinggi di sebagian besar dataset. Temuan ini menegaskan keunggulan TOPSIS dalam mempertahankan performa prediktif yang mendekati optimal, sedangkan WASPAS cenderung menunjukkan kinerja yang kurang memuaskan dalam beberapa kasus.

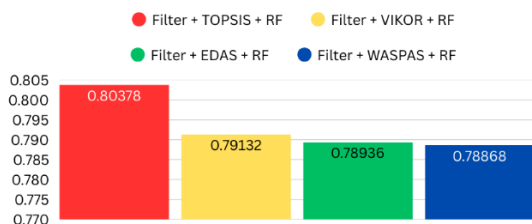
Tabel 4 menunjukkan rata-rata nilai AUC, dihitung dengan menjumlahkan nilai AUC dari tiap dataset dan dibagi dengan jumlah dataset. Susunan kolom pada Tabel 4 sama dengan Tabel 2 dan Tabel 3. Data rata-rata nilai AUC tertinggi ditandai dengan cetak tebal.

Tabel 6. Rata-rata Nilai AUC

Dataset	Filter + TOPSIS + RF	Filter + VIKOR + RF	Filter + EDAS + RF	Filter + WASPAS + RF
Average	0,8038	0,7913	0,7894	0,7887

Berdasarkan data yang terdapat pada Tabel 4, metode TOPSIS mencatatkan rata-rata nilai AUC tertinggi sebesar 0,8038, diikuti oleh metode VIKOR dengan 0,7913, EDAS dengan 0,7894, dan WASPAS dengan 0,7887. Perbedaan ini mengindikasikan bahwa TOPSIS secara umum mampu menangkap aspek penting dari data yang berpengaruh pada kinerja model, meskipun metode-metode lainnya

tetap memberikan hasil yang kompetitif. Adapun penyajian gambar berupa diagram dapat dilihat pada Gambar 4.



Gambar 4 Diagram Rata-rata Nilai AUC

Temuan penelitian ini membuka peluang baru untuk mengeksplorasi integrasi hasil seleksi fitur dalam konteks prediksi cacat perangkat lunak, sekaligus menyoroti keunggulan TOPSIS sebagai kandidat yang menjanjikan. Secara keseluruhan, TOPSIS menunjukkan performa unggul di sebagian besar kasus—baik dari segi dominasi per dataset maupun rata-rata keseluruhan—yang menegaskan pentingnya pemilihan metode pemeringkatan yang sesuai dengan karakteristik data. Meskipun satu metode mungkin unggul secara agregat, pendekatan lain dapat menunjukkan performa yang lebih baik pada dataset tertentu. Dengan demikian, penelitian ini tidak hanya menyediakan landasan praktis bagi pengembangan model prediksi yang lebih adaptif, tetapi juga mendorong penelitian lanjutan untuk mengoptimalkan strategi integrasi fitur dalam menghadapi kompleksitas data nyata.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi efektivitas metode Multi-Criteria Decision Making (MCDM) dalam mengintegrasikan hasil seleksi fitur berbasis filter pada prediksi cacat perangkat lunak. Dengan menggunakan empat metode MCDM, yaitu TOPSIS, VIKOR, EDAS, dan WASPAS, model prediksi dibangun menggunakan algoritma Random Forest dan dievaluasi berdasarkan metrik AUC.

Hasil penelitian menunjukkan bahwa metode TOPSIS secara konsisten memberikan performa terbaik di sebagian besar dataset dengan nilai AUC tertinggi dalam 8 dari 12 dataset yang diuji. Selain itu, metode ini juga memiliki rata-rata nilai AUC tertinggi dibandingkan dengan metode lainnya, yakni sebesar 0,8038. Di sisi lain, WASPAS menunjukkan kinerja yang kurang optimal dengan selisih nilai AUC terbesar dalam 5 dataset, sementara VIKOR dan EDAS menempati posisi menengah. Hasil ini menegaskan bahwa pemilihan metode pemeringkatan yang tepat berpengaruh signifikan terhadap kinerja model, di mana TOPSIS terbukti lebih unggul secara keseluruhan.

Sebagai tindak lanjut, penelitian ini membuka peluang untuk eksplorasi lebih lanjut dalam mengoptimalkan integrasi seleksi fitur menggunakan metode MCDM yang lebih kompleks atau adaptif.

Selain itu, pengujian terhadap dataset lain dengan karakteristik berbeda dapat memberikan wawasan lebih dalam mengenai keandalan metode yang digunakan. Penelitian di masa depan juga dapat mempertimbangkan kombinasi MCDM dengan teknik machine learning yang lebih canggih untuk meningkatkan akurasi prediksi cacat perangkat lunak.

DAFTAR PUSTAKA

- ABELLANA, D.P.M., ROXAS, R.R., LAO, D.M., & MAYOL, P.E., LEE, S. 2022. Ensemble feature selection in binary machine learning classification: A novel application of the evaluation based on distance from average solution (EDAS) method. *Mathematical Problems in Engineering*, <https://doi.org/10.1155/2022>
- AKBAR, A. M., HERTENO, R., SAPUTRO, S. W., FAISAL, M. R., & NUGROHO, R. A. 2024. Optimizing Software Defect Prediction Models: Integrating Hybrid Grey Wolf and Particle Swarm Optimization for Enhanced Feature Selection with Popular Gradient Boosting Algorithm. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(2), 169-181. <https://doi.org/10.35882/jeeemi.v6i2.388>
- ALI, M., MAZHAR, T., SHAHZAD, T., GHADI, Y., MOHSIN, S., AKBAR, S., & ALI, M. 2023. Analysis of Feature Selection Methods in Software Defect Prediction Models. *IEEE Access*, 11, 145954-145974. <https://doi.org/10.1109/ACCESS.2023.3343249>
- ALVES, MA, OLIVEIRA, BAS, & GUIMARÃES, FG 2024. Ensemble ranking: An aggregation of multiple multicriteria methods and scenarios and its application to power generation planning. *Decision Analytics Journal*, Elsevier, <https://doi.org/10.1016/j.dajour.2024.100435>
- BARRY, M. , NDERU, L. AND GICHUHI, A. 2023. A Hybrid Spatial Dependence Model Based on Radial Basis Function Neural Networks (RBFNN) and Random Forest (RF). *Journal of Data Analysis and Information Processing*, 11, 293-309. <https://doi.org/10.4236/jdaip.2023.113015>
- BAYNE, C., MCGROSSO, D., SANCHEZ, C., ROSSITTO, L.A., PATTERSON, M., GONZALEZ, C., BAUS, C., VOLK, C., ZHAO, H. N., DORRESTEIN, P., NIZET, V., SAKOULAS, G., GONZALEZ, D. J., & ROSE, W. 2025. Multi-omic signatures of host response associated with presence, type, and outcome of enterococcal bacteremia. *mSystems*, 10(2), e0147124. <https://doi.org/10.1128/msystems.01471-24>

- ÇORBACIOĞLU, Ş., & AKSEL, G. 2023. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23, 195 - 198. https://doi.org/10.4103/tjem.tjem_182_23.
- FRIEYADIE, S., FONI, R., & SETIAWAN, A. 2024. Improving Software Defect Prediction Performance through Effective Forward Selection Method with Linear Regression for Handling Class Imbalance. 2024 International Conference on Information Technology Research and Innovation (ICITRI), 218-223. <https://doi.org/10.1109/ICITRI62858.2024.10699214>
- GHINAYA, H., HERTENO, R., FAISAL, M. R., FARMADI, A., & INDRIANI, F. 2024. Analysis of Important Features in Software Defect Prediction Using Synthetic Minority Oversampling Techniques (SMOTE), Recursive Feature Elimination (RFE) and Random Forest. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(3), 276-288. <https://doi.org/10.35882/jeeemi.v6i3.453>
- HAMID, T.H., SALLEHUDDIN, R., YUNOS, Z.M., & ALI, A. 2021. Ensemble Based Filter Feature Selection with Harmonize Particle Swarm Optimization and Support Vector Machine for Optimal Cancer Classification. <https://doi.org/10.1016/j.mlwa.2021.100054>
- HASHEMI, A., DOWLATSHAHI, M. B., & NEZAMABADI-POUR, H. 2020. MFS-MCDM: Multi-label feature selection using multi-criteria decision making. *Knowledge-Based Systems*, 106365. <https://doi.org/10.1016/j.knosys.2020.106365>
- HASHEMI, A., DOWLATSHAHI, M., & NEZAMABADI-POUR, H. 2021. Ensemble of feature selection algorithms: a multi-criteria decision-making approach. *International Journal of Machine Learning and Cybernetics*, 13, 49 - 69. <https://doi.org/10.1007/s13042-021-01347-z>.
- JANANE, F., OUADERHMAN, T., & CHAMLAL, H. 2023. A filter feature selection for high-dimensional data. *Journal of Algorithms & Computational Technology*, 17. <https://doi.org/10.1177/17483026231184171>.
- KAUR, K., & KUMAR, A.M. 2023. MCDM-EFS: A novel ensemble feature selection method for software defect prediction using multi-criteria decision making. *Intell. Decis. Technol.*, 17, 1283-1296. <https://doi.org/10.3233/IDT-230251>
- KIBRIA, H.B., & MATIN, A. 2022. The Severity Prediction of The Binary And Multi-Class Cardiovascular Disease - A Machine Learning-Based Fusion Approach. *Computational biology and chemistry*, 98, 107672. <https://doi.org/10.1016/j.compbiolchem.2022.107672>
- KRZYWICKA, M., & WOSIAK, A. 2023. Sensitivity of Standard Evaluation Metrics for Disease Classification and Progression Assessment Based on Whole-Body Imaging. *Procedia Computer Science*, 225, 4314-4323. <https://doi.org/10.1016/j.procs.2023.10.428>
- MALEKIPIRBAZARI, M., AKSAKALLI, V., SHAFQAT, W., & EBERHARD, A. 2021. Performance comparison of feature selection and extraction methods with random instance selection. *Expert Systems with Applications*, 179, 115072. <https://doi.org/10.1016/j.eswa.2021.115072>
- PASHA, S., MOHAMED, E., & PASHA, S. 2020. Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction. *IEEE Access*, 8, 184087-184108. <https://doi.org/10.1109/ACCESS.2020.3028714>.
- NABELLA, P., HERTENO, R., SAPUTRO, S.W., FAISAL, M.R., & ABADI, F. 2024. Impact of a Synthetic Data Vault for Imbalanced Class in Cross-Project Defect Prediction. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(2), 219-230. <https://doi.org/10.35882/jeeemi.v6i2.409>
- RAHMAN, M.N., NUGROHO, R.A., FAISAL, M.R., ABADI, F., & HERTENO, R. 2024. Optimized multi correlation-based feature selection in software defect prediction. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*. <https://doi.org/10.12928/telkomnika.v22i3.25793>
- RAMADHANI, A.A.P., NUGROHO, R.A., FAISAL, M.R., ABADI, F., & HERTENO, R. 2024. The Impact Of Software Metrics In NASA Metric Data Program Dataset Modules For Software Defect Prediction. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*. <https://doi.org/10.12928/TELKOMNIKA.V22I4.25787>

- SALMAN, H., KALAKECH, A., & STEITI, A. 2024. Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, Vol.2024, 69-79. <https://doi.org/10.58496/bjml/2024/007>.
- SHEPPERD, M., SONG, Q., SUN, Z., & MAIR, C. 2018. MDP data sets (D' and D'') - zipped up (Version 1). *figshare*. <https://doi.org/10.6084/m9.figshare.6071675.v1>
- SILVA, F.S. D., LIMA, M.P., CORUJO, D., VENÂNCIO NETO, A.J., & ESPOSITO, F. 2024. A Comprehensive Step-Wise Survey of Multiple Attribute Decision-Making Mobility Approaches. *IEEE Access*, 12, 108616-108656. <https://doi.org/10.1109/ACCESS.2024.3436074>
- TAHERDOOST, H., & MADANCHIAN, M. 2023. Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia*, 3(1), 77-87. <https://doi.org/10.3390/encyclopedia3010006>
- ZEINI, H. A., AL-JEZNAWI, D., IMRAN, H., BERNARDO, L. F. A., AL-KHAFAJI, Z., & OSTROWSKI, K. A. 2023. Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil. *Sustainability*, 15(2), 1408. <https://doi.org/10.3390/su15021408>
- ZHAO, Y., HUANG, Z., GONG, L., ZHU, Y., YU, Q., & GAO, Y. 2023. Evaluating the Impact of Data Transformation Techniques on the Performance and Interpretability of Software Defect Prediction Models. *IET Software*. <https://doi.org/10.1049/2023/6293074>.