

SELEKSI BEASISWA MENGGUNAKAN ANALYTICAL HIERARCHY PROCESS DAN PENILAIAN RUBRIK DENGAN EXPLAINABILITY SERTA AUDIT FAIRNESS

Scholarship Selection Using Analytical Hierarchy Process and Rubric Scoring with Explainability and Fairness Auditing

Amanda Listiana
Puspanagara^{1*}

Fathoni Mahardika²

Dani Indra Junaedi³

Asep Saeppani⁴

^{*1-4} Program Studi Informatika,
Fakultas Teknologi Informasi,
Universitas Sebelas April,
Sumedang, Indonesia

*email:

amandalpn394@gmail.com

Abstrak

Seleksi beasiswa di perguruan tinggi sering terdistorsi oleh subjektivitas, penerapan kriteria yang tidak konsisten, dan rendahnya transparansi ketika penilaian dilakukan manual atau berbasis *spreadsheet*. Penelitian ini mengembangkan dan memvalidasi sistem pendukung keputusan yang memadukan *Analytical Hierarchy Process* untuk pembobotan kriteria dengan penilaian berbasis rubrik untuk menilai pelamar, disertai fitur *explainability* dan *fairness auditing*. Sistem menampilkan kontribusi skor per kriteria, menghasilkan justifikasi naratif deterministik yang diturunkan dari level rubrik dan bobot, serta merangkum dampak kelompok melalui *dashboard* audit. Artefak dirancang, diimplementasikan, dan dievaluasi menggunakan *Design Science Research Methodology* pada uji coba *sandbox* menggunakan 100 pelamar sintesis dalam tiga konfigurasi kebijakan. Pada kebijakan dasar, skor ternormalisasi berada pada rentang 0.48-0.93 dengan rata-rata 0.774, dan pelamar terkelompok menjadi 64 direkomendasikan, 34 *borderline*, dan 2 tidak direkomendasikan. Dibandingkan dengan keputusan referensi pakar, sistem menunjukkan kesepakatan tinggi (akurasi 0.96; presisi 0.953; *recall* 0.983; F1-score 0.96) serta keselarasan peringkat kuat (Spearman 0.97; Kendall 0.86). Audit keadilan berdasarkan kelompok semester mengindikasikan disparitas pada kebijakan dasar, dengan tingkat direkomendasikan 0% untuk semester 1-2, 52% untuk semester 3-4, dan 96% untuk semester 5-8. Hasil ini menunjukkan bahwa pendekatan hibrida meningkatkan konsistensi dan transparansi penilaian sekaligus menyediakan bukti untuk meninjau pertukaran antara prestasi, kebutuhan, dan dampak kelompok, dengan keputusan akhir tetap berada pada komite.

Kata Kunci:

Analytical Hierarchy Process
Sistem Pendukung Keputusan
Rubric Scoring
Explainability
Fairness Auditing

Keywords:

Analytical Hierarchy Process
Decision Support System
Rubric Scoring
Explainability
Fairness Auditing

Abstract

Scholarship selection in higher education can be distorted by subjectivity, inconsistent criterion use, and limited transparency when committees rely on manual review or spreadsheet scoring. This study developed and validated a decision support system that combined the Analytic Hierarchy Process for criterion weighting with rubric-based scoring for applicant assessment, while supporting explainability and fairness auditing. Following the Design Science Research Methodology, the artifact was designed, implemented, and evaluated in a controlled sandbox trial using 100 synthetic applicants under three policy configurations. Under the baseline policy, normalized scores range from 0.48-0.93 with a mean of 0.774, and candidates are grouped into 64 recommended, 34 *borderline*, and 2 not recommended cases. When compared with an expert reference decision set, the system achieves an accuracy of 0.96 and an F1 score of 0.96; precision and recall are 0.953 and 0.983, respectively. The ranking is strongly aligned with the expert ordering (Spearman $\rho = 0.97$; Kendall $\tau = 0.86$). A fairness audit by semester reveals disparities in the baseline policy, with recommended rates of 0% for semesters 1-2, 52% for semesters 3-4, and 96% for semesters 5-8. Overall, the proposed hybrid approach improves scoring consistency and transparency, and it provides actionable evidence for committees to review trade-offs between merit, need, and group impacts while keeping final decisions under human oversight.

Submit Tgl.: 06-Januari-2026

Diterima Tgl.: 10-Januari-2026

Diterbitkan Tgl.: 12-Januari-2026

Cara mengutip Puspanagara, A. L., Mahardika, F., Junaedi, D. I., & Saeppani, A. (2026). Seleksi Beasiswa Menggunakan Analytical Hierarchy Process dan Penilaian Rubrik dengan Explainability Serta Audit Fairness. *Jurnal Informatika, Multimedia Dan Teknik*, 2(2), 92–99. <https://doi.org/10.71456/jimt.v2i2.1588>

PENDAHULUAN

Keterbatasan finansial masih menjadi penghalang utama bagi banyak mahasiswa untuk memperoleh dan mempertahankan akses terhadap pendidikan tinggi. Program beasiswa bertujuan menutup kesenjangan tersebut, namun proses seleksi kerap terdistorsi oleh subjektivitas, penggunaan kriteria yang tidak konsisten, serta transparansi yang terbatas ketika panitia mengandalkan penilaian manual atau skoring berbasis *spreadsheet* (Mulyaningsih et al., 2022). Pada saat yang sama, ekosistem donasi daring dan *crowdfunding* berkembang pesat, sehingga menuntut mekanisme seleksi yang akuntabel dan dapat dijelaskan kepada para pemangku kepentingan (Dethier et al., 2024; Ghoorah et al., 2025). Dalam konteks ini, sistem pendukung keputusan yang terstruktur dan transparan diperlukan agar penilaian lebih konsisten, alasan keputusan dapat ditelusuri, dan kebijakan seleksi dapat dievaluasi dampaknya.

Pendekatan *multi-criteria decision-making* (MCDM) telah digunakan untuk rekomendasi penerima beasiswa. Salah satu yang paling banyak diadopsi adalah *Analytical Hierarchy Process* (AHP) untuk menentukan bobot kriteria melalui perbandingan berpasangan (Anggrawan et al., 2022). AHP telah banyak ditinjau dan diaplikasikan dalam beragam domain, serta memiliki mekanisme uji konsistensi untuk meminimalkan inkonsistensi penilai (Madzik & Falat, 2022; Salomon & Gomes, 2024). Namun, bobot kriteria saja tidak cukup, kualitas penilaian pada level pelamar juga dipengaruhi oleh bagaimana rubrik dipakai secara konsisten. Penilaian berbasis rubrik terbukti mendukung konsistensi, pembelajaran terarah, dan efikasi diri, baik pada pendidikan tinggi maupun K-12 (English et al., 2022; Garcia et al., 2023; Jonsson et al., 2025; Teh, 2025). Karena itu, integrasi AHP untuk bobot dan rubrik untuk skor pelamar berpotensi menghasilkan seleksi yang lebih terstandar dan dapat dipertanggungjawabkan. Di sisi lain, perkembangan *explainable multi-criteria decision-making* menunjukkan kebutuhan untuk menyajikan

alasan keputusan yang mudah dipahami pengguna (Shi & Yao, 2025).

Meskipun sistem seleksi yang terotomatisasi dapat meningkatkan efisiensi, ia juga memunculkan tuntutan *explainability* dan *fairness*. Bidang *Explainable Artificial Intelligence* (XAI) menekankan pentingnya transparansi atas alasan keputusan dan keterbatasan model (Longo et al., 2024). Dalam pengambilan keputusan otomatis, *fairness* dan *explainability* sering dipandang saling terkait, terutama ketika keputusan memengaruhi akses terhadap sumber daya publik (Deck et al., 2024; Kirat et al., 2023). Literatur tentang *fairness* menyoroti adanya beragam definisi dan metrik yang dapat menimbulkan ketegangan, sehingga audit dan pelaporan menjadi krusial (Alves et al., 2023; DeLuca, 2025; Naghiaei et al., 2021). Dalam konteks rekomendasi dan pemeringkatan, penelitian *fairness* berkembang pesat, termasuk metrik *fairness*, mitigasi bias, dan arah riset ke depan (Chen, 2025; Jannach et al., 2024; Jin et al., 2023; Pitoura et al., 2022). Selain itu, bias juga dapat muncul dalam asesmen berbasis manusia seperti ujian lisan atau *peer assessment*, sehingga kontrol proses penilaian tetap relevan bahkan ketika sistem bersifat semi-otomatis (Syverud, 2025; Tanaka, 2025). Oleh karena itu, studi ini mengusulkan pendekatan hibrida AHP–rubrik yang dilengkapi modul *explainability* dan *dashboard fairness auditing* untuk mendukung transparansi dan akuntabilitas seleksi beasiswa, sambil tetap mempertahankan keputusan akhir pada panitia.

METODE PENELITIAN

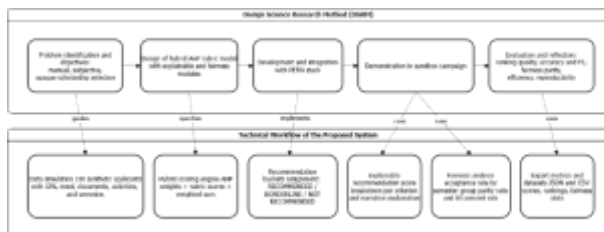
Alat dan Bahan

Alat yang digunakan meliputi aplikasi web LuminaFund sebagai antarmuka seleksi dan pemantauan; modul *Analytical Hierarchy Process* untuk perbandingan berpasangan, perhitungan bobot kriteria, dan rasio konsistensi; modul penilaian rubrik untuk menghitung skor pelamar berdasarkan level rubrik pada tiap kriteria; modul *explainability* yang menghasilkan rincian kontribusi per kriteria dan narasi alasan rekomendasi

secara deterministik; *dashboard fairness auditing* untuk menghitung indikator keadilan antar kelompok berdasarkan semester; skrip *seeding* yang menghasilkan 100 pelamar sintetis dengan atribut yang selaras dengan kriteria seperti IPK, indikator kebutuhan ekonomi, kelengkapan dokumen, riwayat beasiswa, keterlibatan organisasi, dan semester. Seluruh komponen dijalankan pada lingkungan komputasi standar (laptop/PC) dengan basis data aplikasi untuk menyimpan profil pelamar, bobot AHP, skor rubrik, serta hasil rekomendasi.

Metode Pelaksanaan

Penelitian ini menggunakan *Design Science Research Methodology* (DSRM) yang mencakup identifikasi masalah, penetapan tujuan solusi, desain dan pengembangan artefak, demonstrasi, evaluasi, serta komunikasi hasil. Artefak yang dikembangkan adalah aplikasi web LuminaFund yang mengintegrasikan pembobotan AHP, penilaian rubrik, modul *explainability*, dan *dashboard fairness auditing*. Alur DSRM dan *pipeline* hibrida AHP–rubrik dirangkum pada Gambar 1 (Al-Habaibeh et al., 2021).



Gambar 1. Alur DSRM dan *pipeline* hibrida AHP–rubrik untuk seleksi beasiswa

Tahap desain dan pengembangan dimulai dengan perumusan kriteria seleksi dan penyusunan matriks perbandingan berpasangan oleh evaluator. Bobot kriteria dihitung menggunakan metode *eigenvector* dan hanya konfigurasi dengan *consistency ratio* < 0.10 yang digunakan. Konsistensi penilai dievaluasi menggunakan *consistency ratio* (CR), dan matriks yang tidak konsisten perlu diperbaiki agar keputusan lebih reliabel (Madzik & Falat, 2022; Salomon & Gomes, 2024). Selanjutnya, rubrik penilaian disusun untuk memetakan kinerja

pelamar ke level rubrik pada setiap kriteria. Rubrik dipilih karena mampu meningkatkan konsistensi penilaian dan memberikan umpan balik yang lebih terstruktur (English et al., 2022; Garcia et al., 2023; Jonsson et al., 2025; Teh, 2025). Kemudian skor total dihitung sebagai penjumlahan berbobot:

$$S_i = \sum_{j=1}^m \omega_j r_{ij} \quad (1)$$

Dengan w_j menyatakan bobot kriteria ke- j , r_{ij} skor rubrik pelamar i pada kriteria ke- j , dan m jumlah kriteria. Skor kemudian dinormalisasi ke interval $[0,1]$, pelamar diranking menurun berdasarkan skor, dan hasil dirangkum ke tiga *bucket* rekomendasi: *RECOMMENDED*, *BORDERLINE*, dan *NOT RECOMMENDED*. Pada konfigurasi dasar, ambang $T_{high} = 0.75$ dan $T_{low} = 0.55$ digunakan; untuk menjaga keterbandingan antar-skenario kebijakan, nilai ambang yang sama diterapkan pada semua skenario.

Untuk *explainability*, sistem menampilkan rincian kontribusi skor per kriteria ($w_j \times r_{ij}$) dan menghasilkan narasi alasan rekomendasi. Narasi dibangun menggunakan *template* deterministik, sehingga alasan yang disajikan konsisten dan mudah diaudit. Pendekatan ini sejalan dengan agenda XAI yang menekankan transparansi dan komunikasi alasan keputusan kepada pengguna (Longo et al., 2024), serta kebutuhan menyeimbangkan akurasi dan keterjelasan dalam keputusan berbantuan AI (van der Veer et al., 2021). Pemilihan strategi *explainability* juga mempertimbangkan kerangka pemilihan metode XAI berbasis MCDM (Matejova & Paralic, 2025).

Untuk *fairness auditing*, studi ini mengadopsi pendekatan audit algoritma dengan membandingkan *outcome* antar kelompok yang relevan secara kebijakan (Brown et al., 2021). Kelompok yang diaudit ditetapkan berdasarkan semester (1–2, 3–4, dan 5–8). Indikator utama yang digunakan adalah *acceptance rate* (AR) per

kelompok untuk *bucket RECOMMENDED*, dan kemudian dihitung *disparity index* (DI) relatif terhadap kelompok dengan AR tertinggi. Selain itu, aturan praktis 80% rule digunakan sebagai penanda awal potensi disparitas, sebagaimana umum dipakai dalam audit fairness (Jannach et al., 2024; Jin et al., 2023; Pitoura et al., 2022). Disparitas diringkas melalui metrik parity berikut dan kelompok dengan parity <0.80 ditandai untuk ditinjau.

$$\text{Parity}_g = \frac{a_g}{\max_h a_h} \quad (2)$$

Demonstrasi dan evaluasi dilakukan pada kampanye *sandbox* dengan 100 pelamar sintetis. Dataset sintetis dipilih untuk menguji perilaku sistem secara terkendali sebelum diterapkan pada data riil. Tiga skenario kebijakan diuji: (A) kebijakan dasar yang menyeimbangkan merit (prestasi) dan need (kebutuhan), (B) kebijakan berfokus kebutuhan, dan (C) kebijakan berfokus merit. Untuk mengevaluasi kualitas rekomendasi, hasil sistem dibandingkan terhadap himpunan keputusan referensi pakar menggunakan metrik akurasi, presisi, *recall*, dan F1-score; serta metrik keselarasan peringkat menggunakan *Spearman's ρ* dan *Kendall's τ*.

HASIL DAN PEMBAHASAN

Pada skenario kebijakan dasar (A), distribusi skor ternormalisasi berada pada rentang 0.48–0.93 dengan rata-rata 0.774 (SD 0.100) dan koefisien *Gini* 0.072. Dengan ambang $T_{high} = 0.75$ dan $T_{low} = 0.55$, sistem menghasilkan 64 pelamar pada *bucket RECOMMENDED*, 34 pada *BORDERLINE*, dan 2 pada *NOT RECOMMENDED*. Ringkasan statistik skor serta ukuran *bucket* pada tiap skenario disajikan pada Tabel I.

Tabel I. Statistik skor ternormalisasi dan ukuran *bucket* pada tiap skenario.

Scenario	Avg	SD	G	Rec	Bor	Not
A	0.774	0.100	0.072	64	34	2

B	0.784	0.095	0.068	68	29	3
C	0.764	0.113	0.082	61	34	5

Catatan: Avg = rata-rata skor; SD = simpangan baku; G = koefisien *Gini*; Rec/Bor/Not = jumlah pelamar per *bucket*.

Perubahan bobot kebijakan memengaruhi distribusi skor dan komposisi *bucket*. Skenario berfokus kebutuhan (B) menghasilkan rata-rata sedikit lebih tinggi dan jumlah *RECOMMENDED* terbesar, sedangkan skenario berfokus merit (C) menunjukkan dispersi skor lebih tinggi dan jumlah *NOT RECOMMENDED* terbanyak. Dari sisi antarmuka, LuminaFund menyediakan *selection interface* yang menampilkan skor, status *bucket*, dan tautan rincian kontribusi (Gambar 2).

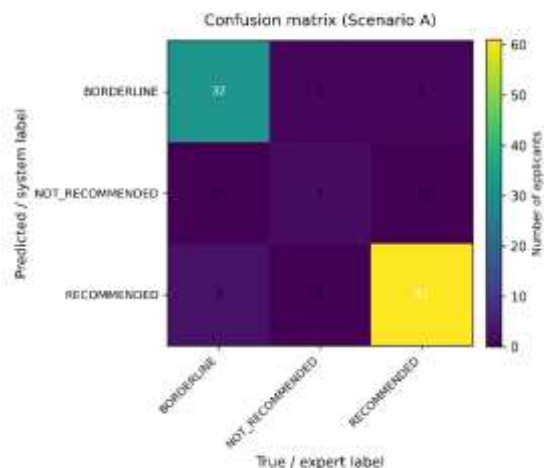


Gambar 2. Mentor-facing selection board di LuminaFund untuk Skenario A

Untuk menilai kesesuaian terhadap penilaian pakar, keputusan referensi digunakan sebagai pembanding pada klasifikasi biner *RECOMMENDED* versus non-*RECOMMENDED*. Pada Skenario A, sistem mencapai akurasi 0.96, presisi 0.953, *recall* 0.983, dan F1-score 0.96; keselarasan pemeringkatan juga tinggi (*Spearman's ρ* = 0.97; *Kendall's τ* = 0.86), sebagaimana diringkas pada Tabel II. *Confusion matrix* pada Gambar 3 menunjukkan sebagian besar kasus berada pada diagonal, sementara ketidaksesuaian terutama muncul di sekitar batas keputusan sehingga menegaskan perlunya peninjauan manusia untuk profil *borderline*.

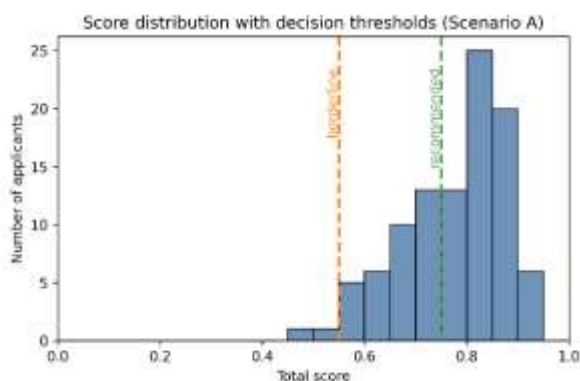
Tabel II. Metrik evaluasi pada Skenario A terhadap keputusan referensi pakar.

Metrik	Nilai
Akurasi	0.96
Presisi	0.953
Recall	0.983
F1-score	0.96
Spearman's ρ	0.97
Kendall's τ	0.86



Gambar 3. Confusion matrix rekomendasi skenario A dibanding keputusan referensi

Distribusi skor dan posisi ambang (T_{low} dan T_{high}) dapat dipantau untuk memahami sensitivitas kebijakan terhadap perubahan parameter. Visualisasi histogram skor dengan garis ambang pada skenario A ditunjukkan pada Gambar 4, yang membantu panitia melihat apakah sebagian besar pelamar terkonsentrasi di sekitar batas *borderline* atau tidak.



Gambar 4. Distribusi skor ternormalisasi dan ambang T_{low}/T_{high} pada skenario A

Untuk *explainability*, sistem menampilkan kontribusi skor per kriteria dan narasi alasan. Tabel III memberikan contoh *breakdown* kontribusi lima kriteria utama, yang memperlihatkan bagaimana bobot AHP memengaruhi penjumlahan skor akhir. Dengan mekanisme ini, panitia dapat menelusuri mengapa seorang pelamar berada pada *bucket* tertentu, sekaligus

mengidentifikasi kriteria mana yang paling menentukan. Pendekatan penjelasan deterministik ini selaras dengan prinsip XAI untuk meningkatkan kepercayaan dan akuntabilitas keputusan (Longo et al., 2024), serta kebutuhan *fairness* dan *explainability* dalam sistem keputusan otomatis (Kirat et al., 2023).

Tabel III. Breakdown skor per kriteria untuk satu pelamar *recommended* pada Skenario A.

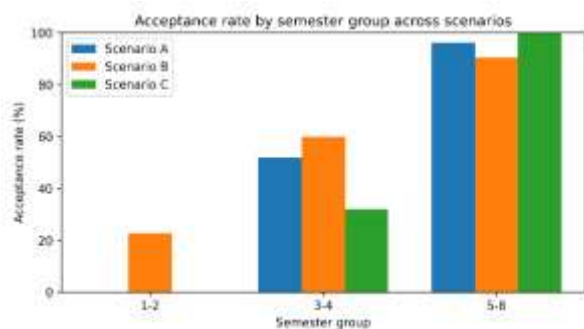
Kriteria	Bobot AHP	Skor Rubrik	Kontribusi ($w \times r$)
Rekam akademik/GPA	0.300	0.932	0.280
Kebutuhan finansial	0.300	0.950	0.285
Kelengkapan dokumen	0.100	0.650	0.065
Riwayat beasiswa	0.100	1.000	0.100
Semester aktif	0.200	1.000	0.200
Total	-	-	0.930

Hasil *fairness auditing* menunjukkan adanya disparitas yang besar ketika *outcome* dibandingkan antar kelompok semester pada kebijakan dasar. Pada skenario A, *acceptance rate RECOMMENDED* untuk semester 1–2 adalah 0% (0 dari 22), semester 3–4 sebesar 52% (13 dari 25), dan semester 5–8 sebesar 96% (51 dari 53). Rincian metrik *fairness* per kelompok disajikan pada Tabel IV, perbandingan *acceptance rate* antar kelompok dan skenario divisualisasikan pada Gambar 5. Temuan ini menunjukkan bahwa pilihan bobot dan ambang kebijakan dapat secara sistematis menguntungkan kelompok tertentu, sehingga audit *fairness* perlu dilakukan sebagai bagian dari proses pengambilan keputusan (Brown et al., 2021; Jannach et al., 2024; Jin et al., 2023).

Tabel IV. Audit *fairness* berdasarkan kelompok semester pada Skenario A.

Grp	N	Rec	AR	Par	Flag
1–2	22	0	0	0.00	Concern
3–4	25	13	0.52	0.54	Concern
5–8	53	51	0.96	1.00	OK

Catatan: AR = *acceptance rate* untuk *bucket RECOMMENDED*; *parity* dihitung menurut Persamaan (2); *parity* < 0.80 ditandai (*Concern*).



Gambar 5. Acceptance rate bucket RECOMMENDED per kelompok semester di tiap skenario

Secara konseptual, disparitas ini dapat berasal dari korelasi antara semester dan indikator merit tertentu pada data sintesis, atau dari kebijakan yang secara eksplisit lebih mengutamakan kriteria yang cenderung dimiliki mahasiswa senior. Karena rubrik dan AHP mengikat proses penilaian pada kriteria yang terstruktur, pergeseran bobot atau perubahan definisi level rubrik dapat digunakan sebagai instrumen kebijakan untuk menyeimbangkan antara merit dan need. Namun, literatur *fairness* juga menekankan bahwa tidak ada satu metrik yang universal, audit perlu mempertimbangkan konteks dan *trade-off* antar metrik (Alves et al., 2023; Naghiaei et al., 2021; Pitoura et al., 2022).

Implikasi praktis dari artefak ini adalah tersedianya jejak alasan atau audit *trail* yang memudahkan komunikasi keputusan kepada pelamar dan pemangku kepentingan. Di sisi lain, studi tentang keputusan berbantuan AI menunjukkan bahwa pengguna sering menuntut penjelasan yang ringkas dan mudah dipahami, bahkan ketika akurasi tinggi (van der Veer et al., 2021). Karena itu, narasi deterministik dan tampilan kontribusi kriteria di LuminaFund dirancang agar tetap operasional bagi panitia. Selain itu, praktik audit algoritma yang direkomendasikan dalam literatur menekankan perlunya pelaporan dan peninjauan berkala terhadap dampak kelompok (Brown et al., 2021).

Meskipun dievaluasi pada dataset sintesis, pendekatan ini memberikan kerangka implementasi yang dapat diperluas. Studi terkait *fairness* dan *explainability* menyoroti tantangan lintas disiplin,

termasuk aspek hukum dan tata kelola, sehingga penerapan pada data nyata perlu mempertimbangkan kebijakan institusi serta mekanisme moderasi eksternal bila dibutuhkan (Kirat et al., 2023; Murchan et al., 2025). Selain itu, agenda riset XAI 2.0 menekankan adanya tantangan terbuka dalam evaluasi dan integrasi penjelasan, terutama ketika sistem digunakan pada konteks berisiko tinggi (Longo et al., 2024).

KESIMPULAN

Penelitian ini menghasilkan artefak pada LuminaFund yang mengintegrasikan pembobotan kriteria menggunakan *Analytical Hierarchy Process*, penilaian berbasis rubrik, *explainability* deterministik, dan *fairness auditing*. Evaluasi pada kampanye *sandbox* dengan 100 pelamar sintesis di tiga skenario kebijakan menunjukkan bahwa pada kebijakan dasar skor ternormalisasi berada pada rentang 0.48–0.93 (rata-rata 0.774) dan terkelompok menjadi 64 RECOMMENDED, 34 BORDERLINE, dan 2 NOT RECOMMENDED, dengan kesepakatan tinggi terhadap referensi pakar (akurasi 0.96; F1-score 0.96) serta keselarasan peringkat kuat (kuat *Spearman's ρ* = 0.97, *Kendall's τ* = 0.86.). Di tingkat kelompok, audit *fairness* mengungkap disparitas rekomendasi berdasarkan semester (0% pada semester 1–2, 52% pada semester 3–4, dan 96% pada semester 5–8), sehingga tujuan *fairness* berpotensi memerlukan mekanisme kebijakan tambahan seperti penyesuaian rubrik atau kuota berbasis semester, bukan sekadar tuning bobot.

REFERENSI

- Al-Habaibeh, A., Christersson, E., Hossain, M. A., & Desouza, N. (2021). Human-centered design science research evaluation for AR games. *Frontiers in Virtual Reality*, 2, 713718. <https://doi.org/10.3389/frvir.2021.713718>
- Alves, G., Bernier, F., Couceiro, M., Makhlof, K., Palamidessi, C., & Zhioua, S. (2023). Survey on fairness notions and related tensions. *EURO Journal on Decision Processes*, 11, 100033. <https://doi.org/10.1016/j.ejdp.2023.100033>



- Anggrawan, A., Mayadi, M., Satria, C., & Putra, L. G. R. (2022). Scholarship recipients recommendation system using AHP and MOORA methods. *International Journal of Intelligent Engineering and Systems*, 15(2), 260–275. <https://doi.org/10.22266/ijies2022.0430.24>
- Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, 8(1), 1–15. <https://doi.org/10.1177/2053951720983865>
- Chen, W. (2025). FairDgcl: fairness-aware recommendation with dynamic graph contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(9), 5230–5242. <https://doi.org/10.1109/TKDE.2025.3580087>
- Deck, L., Schoeffer, J., De-Arteaga, M., & Kühl, N. (2024). A Critical Survey on Fairness Benefits of Explainable AI. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 1579–1595. <https://doi.org/10.1145/3630106.3658990>
- DeLuca, C. (2025). Diversity: a necessary imperative for assessment research. *Assessment in Education: Principles, Policy and Practice*, 32(3), 253–256. <https://doi.org/10.1080/0969594X.2025.2549158>
- Dethier, A., Delcourt, C., & Dessart, L. (2024). Donor perceptions of nonprofit organizations' transparency: Conceptualization and operationalization. *Nonprofit and Voluntary Sector Quarterly*, 53(5), 1230–1260. <https://doi.org/10.1177/08997640231211212>
- English, N., Robertson, P., Gillis, S., & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, 113, 101964. <https://doi.org/10.1016/j.ijer.2022.101964>
- Garcia, T., Tenorio, R., Garcia, P., Arquero, J. F., & Diaz, E. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analysis. *Educational Psychology Review*, 35, 113. <https://doi.org/10.1007/s10648-023-09823-4>
- Ghoorah, U., Mariyani-Squire, E., & Amin, S. Z. (2025). Relationships between financial transparency, trust, and performance: an examination of donors' perceptions. *Humanities and Social Sciences Communications*, 12, 315. <https://doi.org/10.1057/s41599-025-04640-2>
- Jannach, D., Lerche, L., Kovalerchuk, I., Zanker, M., Kuschig, P., & Jerhot, P. (2024). Fairness in recommender systems: Research landscape and future directions. *User Modeling and User-Adapted Interaction*, 34, 59–108. <https://doi.org/10.1007/s11257-023-09364-z>
- Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., & Pan, S. (2023). A Survey on Fairness-aware Recommender Systems. *Information Fusion*, 100, 101906. <https://doi.org/10.1016/j.inffus.2023.101906>
- Jonsson, A., Panadero, E., Pinedo, L., & Fernandez-Castilla, B. (2025). Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations. *Assessment in Education: Principles, Policy and Practice*, 32(2), 192–211. <https://doi.org/10.1080/0969594X.2025.2486947>
- Kirat, Th., Tambou, O., Do, V., & Tsoukiàs, A. (2023). Fairness and explainability in automatic decision-making systems. A challenge for computer science and law. *EURO Journal on Decision Processes*, 11, 100036. <https://doi.org/10.1016/j.ejdp.2023.100036>
- Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. Del, Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106, 102301. <https://doi.org/10.1016/j.inffus.2024.102301>
- Madzik, P., & Falat, L. (2022). State-of-the-art on analytic hierarchy process in the last 40 years: Literature review based on Latent Dirichlet Allocation topic modelling. *PLOS ONE*, 17(5), e0268777. <https://doi.org/10.1371/journal.pone.0268777>
- Matejova, M., & Paralic, J. (2025). A multi-criteria decision-making approach for the selection of explainable AI methods. *Machine Learning and Knowledge Extraction*, 7(4), 158. <https://doi.org/10.3390/make7040158>
- Mulyaningsih, T., Dong, S. X., Miranti, R., Daly, A., & Purwaningsih, Y. (2022). Targeted scholarship for higher education and academic performance: Evidence from Indonesia. *International Journal of Educational Development*, 88, 102510. <https://doi.org/10.1016/j.ijedudev.2021.102510>
- Murchan, D., Shaw, S., & Likhovtseva, E. (2025). Policy and practice in relation to external moderation of school-based assessment in 13 examination systems internationally. *Assessment in Education: Principles, Policy and Practice*. <https://doi.org/10.1080/0969594X.2025.2562814>
- Naghiaei, M., Kamishima, T., Kameyada, H., & Sakuma, J. (2021). Fairness metrics and bias mitigation strategies for rating predictions. *Information Processing & Management*, 58(5), 102645. <https://doi.org/10.1016/j.ipm.2021.102646>

- Pitoura, G., Siachamis, K., Stefanidis, K., & Tsaparas, P. (2022). Fairness in rankings and recommendations: An overview. *The VLDB Journal*, 31, 431–454. <https://doi.org/10.1007/s00778-021-00697-y>
- Salomon, V. A. P., & Gomes, L. F. A. M. (2024). Consistency Improvement in the Analytic Hierarchy Process. *Mathematics*, 12(6), 828. <https://doi.org/10.3390/math12060828>
- Shi, C., & Yao, Y. (2025). Explainable multi-criteria decision-making: A three-way decision perspective. *International Journal of Approximate Reasoning*, 187, 109528. <https://doi.org/10.1016/j.ijar.2025.109528>
- Syverud, M. S. (2025). Oral exams in four Norwegian secondary schools - characteristics and variations in practice and possible threats to validity and fairness. *Assessment in Education: Principles, Policy and Practice*. <https://doi.org/10.1080/0969594X.2025.2563722>
- Tanaka, M. (2025). Friendship bias in peer assessment of EFL oral presentations. *Assessment in Education: Principles, Policy and Practice*, 1–23. <https://doi.org/10.1080/0969594X.2025.2570248>
- Teh, L. J. (2025). Exploring postgraduate students' experience with rubric-referenced assessment: limitations and solutions. *Pertanika Journal of Social Sciences and Humanities*, 33(2), 541–562. <https://doi.org/10.47836/pjssh.33.2.03>
- van der Veer, S. N., Dowrick, C. J., Genao, M. G. M., O'Connor, M. J., & Teodorczuk, A. (2021). Trading off accuracy and explainability in AI decision-making: Findings from two citizens' juries. *Journal of the American Medical Informatics Association*, 28(10), 2128–2138. <https://doi.org/10.1093/jamia/ocab118>