

Deteksi Anomali Integritas Akademik Siswa Menggunakan Isolation Forest dan Data Mining

Yesi Wulan Puspitasari¹, Achmad Agus Athok Miftachuddin²

^{1,2} Program Studi Informatika, Fakultas Teknologi Informasi, Universitas KH. A. Wahab Hasbullah, Indonesia

Email: ¹syubban230@gmail.com, ²agusathok@unwaha.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mendeteksi anomali integritas akademik siswa kelas XII pada sebuah Madrasah Aliyah Negeri (MAN) menggunakan pendekatan analisis data mining berbasis Python. Sistem pemantauan akademik yang selama ini diterapkan masih dilakukan secara manual sehingga berpotensi menyebabkan keterlambatan dalam mengidentifikasi siswa yang memerlukan pendampingan. Penelitian ini menerapkan algoritma Isolation Forest dengan pendekatan unsupervised learning untuk menganalisis data nilai UTS, UAS, tugas akhir, rata-rata rapor, persentase kehadiran, jumlah pelanggaran disiplin, dan konseling BK dari 650 siswa. Proses analisis menggunakan metodologi CRISP-DM yang mencakup enam tahapan, mulai dari pemahaman permasalahan hingga penerapan hasil analisis. Hasil deteksi anomali selanjutnya divalidasi menggunakan pendekatan rule-based sehingga menghasilkan proses hybrid validation yang lebih objektif dan kontekstual. Model berhasil mengidentifikasi 52 siswa (8%) sebagai data anomali. Melalui hybrid validation, diperoleh 544 siswa (83,69%) berada pada kategori Aman, 97 siswa (14,92%) pada kategori Perlu Monitoring, dan 9 siswa (1,39%) pada kategori Perlu Tindak Lanjut. Hasil penelitian divisualisasikan melalui dashboard berbasis Streamlit yang dapat dimanfaatkan sebagai pendukung pengambilan keputusan oleh pihak sekolah dalam melakukan pemantauan akademik secara objektif dan berbasis data.

Kata Kunci: Data Mining, Deteksi Anomali, Isolation Forest, Integritas Akademik, Hybrid Validation, Streamlit, Madrasah Aliyah Negeri

ABSTRACT

This study aimed to detect academic integrity anomalies among twelfth-grade students at an Islamic senior high school (Madrasah Aliyah Negeri) using a Python-based data mining approach. Academic monitoring had previously been conducted manually, which risked delaying the identification of students who required guidance. This research applied the Isolation Forest algorithm with an unsupervised learning approach to analyze midterm and final exam scores, project scores, report averages, attendance percentages, disciplinary violation records, and school counseling records from 650 students. The analysis followed the CRISP-DM methodology, covering six stages from business understanding to deployment. The anomaly detection results were further validated using a rule-based approach, producing a more objective and contextual hybrid validation process. The model identified 52 students (8%) as anomalous data. Hybrid validation classified 544 students (83.69%) as Safe, 97 students (14.92%) as Needing Monitoring, and 9 students (1.39%) as Needing Follow-up. The results were visualized through a Streamlit-based dashboard that can support school decision-making in objective, data-driven academic monitoring.

Keywords: Data Mining, Anomaly Detection, Isolation Forest, Academic Integrity, Hybrid Validation, Streamlit, Madrasah Aliyah Negeri.

Penulis Korespondensi:

Yesi Wulan Puspitasari

Email: syubban230@gmail.com

Article Info

Diterima: 25 Juli 2026

Direvisi: 30 Juli 2026

Disetujui: 6 Agustus 2026

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Integritas akademik merupakan fondasi utama dalam proses pendidikan yang mencakup nilai kejujuran, tanggung jawab, dan keadilan dalam seluruh kegiatan belajar-mengajar. Di era digital, menjaga integritas akademik menjadi tantangan yang semakin kompleks karena kemudahan akses teknologi informasi membuka peluang terjadinya pelanggaran akademik, seperti plagiarisme digital, berbagi jawaban melalui media sosial, hingga inkonsistensi nilai yang mengindikasikan ketidakwajaran proses belajar siswa. Pelanggaran tersebut sering kali baru disadari setelah berdampak negatif terhadap kualitas pembelajaran secara keseluruhan [1].

Salah satu Madrasah Aliyah Negeri (MAN) di Indonesia yang menjadi objek penelitian ini memiliki 650 siswa kelas XII yang terbagi dalam beberapa kelas paralel. Meskipun belum pernah terjadi kasus pelanggaran akademik yang bersifat serius, pihak sekolah menyadari pentingnya sistem pemantauan akademik yang proaktif [2]. Sistem monitoring yang selama ini diterapkan masih mengandalkan pendekatan manual berupa observasi langsung di kelas dan rekap nilai berkala, sehingga sulit menganalisis ribuan catatan akademik, meliputi nilai UTS, UAS, tugas akhir, rata-rata rapor, data kehadiran, pelanggaran disiplin, dan riwayat konseling BK, secara objektif dan tepat waktu. Akibatnya, siswa yang sesungguhnya membutuhkan perhatian khusus sering kali tidak teridentifikasi secara tepat [3].

Data mining menawarkan pendekatan yang lebih sistematis dan berbasis komputasi untuk mengatasi keterbatasan tersebut, karena mampu mengekstraksi pola dan pengetahuan tersembunyi dari kumpulan data akademik yang besar [4], [5]. Salah satu teknik yang sesuai untuk permasalahan ini adalah deteksi anomali, yaitu metode yang dirancang untuk mengidentifikasi pola data yang menyimpang dari kebiasaan normal [6], misalnya siswa dengan nilai akademik tinggi namun kehadiran rendah, atau siswa dengan selisih nilai UTS dan UAS yang sangat besar.

Penelitian ini menerapkan algoritma *Isolation Forest* sebagai metode utama deteksi anomali. *Isolation Forest* bekerja berdasarkan prinsip bahwa data anomali lebih mudah diisolasi dibandingkan data normal karena memiliki karakteristik yang berbeda dari mayoritas data [7]. Algoritma ini dipilih karena tidak memerlukan data berlabel (unsupervised learning) sehingga sesuai diterapkan pada data akademik sekolah yang umumnya tidak memiliki label kelas anomali secara eksplisit. Proses analisis data mengikuti metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*), yang telah terbukti efektif pada berbagai penelitian data mining di bidang pendidikan maupun industri [8], [9].

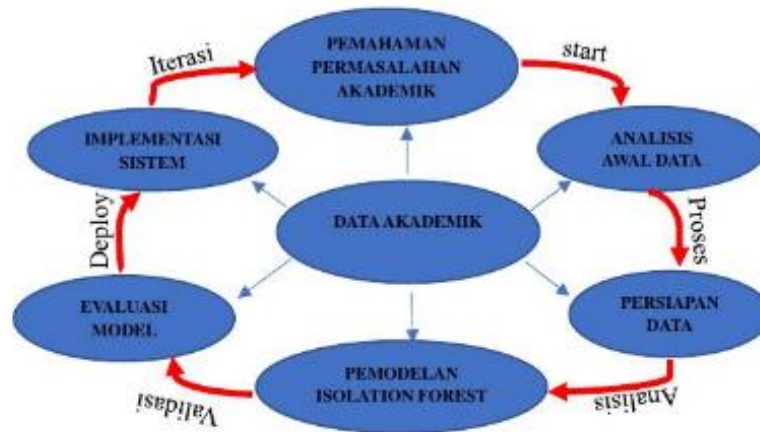
Penelitian terkait deteksi anomali pada data pendidikan telah banyak dilakukan, misalnya penerapan *Isolation Forest* yang dioptimalkan untuk mengenali perilaku menyimpang siswa pada data mining manajemen pendidikan [10], deteksi menggunakan *Extended Isolation Forest* [11], serta pendekatan LOF-GMM untuk perilaku abnormal siswa [12]. Penelitian lain juga mengembangkan sistem peringatan dini pelanggaran integritas akademik berbasis *machine learning* [13] dan kerangka deteksi *hybrid* yang menggabungkan pendekatan anomali dan berbasis aturan [14]. Berbeda dengan penelitian sebelumnya, kebaruan penelitian ini terletak pada integrasi algoritma *Isolation Forest* dengan mekanisme *hybrid validation* berbasis *rule-based* yang disusun berdasarkan kebijakan akademik sekolah. Selain itu, hasil deteksi divisualisasikan melalui dashboard monitoring berbasis Streamlit sehingga proses identifikasi siswa yang memerlukan perhatian khusus dapat dilakukan secara lebih objektif, kontekstual, dan mendukung pengambilan keputusan di lingkungan sekolah.

Berdasarkan uraian tersebut, penelitian ini bertujuan membangun sistem deteksi anomali integritas akademik siswa kelas XII menggunakan algoritma *Isolation Forest* dengan pendekatan data mining berbasis *Python*. Sistem ini mengintegrasikan hasil deteksi dengan mekanisme *hybrid validation* dan menyajikannya melalui dashboard monitoring berbasis Streamlit. Hasil penelitian diharapkan mampu membantu pihak sekolah mengidentifikasi secara dini siswa yang memerlukan pendampingan akademik secara lebih objektif dan berbasis data, sekaligus memberikan kontribusi pada pengembangan penelitian di bidang *educational data mining* di Indonesia [9].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode pengembangan sistem berbasis data mining untuk deteksi anomali integritas akademik siswa. Sumber data terdiri atas data primer, yaitu hasil wawancara dan diskusi dengan pihak sekolah terkait indikator perilaku akademik dan kebijakan integritas, serta data sekunder berupa dokumen akademik siswa yang mencakup nilai UTS, UAS, tugas akhir, rata-rata rapor, riwayat kehadiran, catatan bimbingan konseling, dan catatan pelanggaran disiplin dari 650 siswa kelas XII.

Proses analisis data mengikuti kerangka kerja CRISP-DM yang terdiri atas enam tahapan, yaitu pemahaman permasalahan akademik, analisis awal data akademik, persiapan data, pemodelan, evaluasi, dan penyebaran hasil, sebagaimana ditunjukkan pada Gambar 1 [8].



Gambar 1. Tahapan Metode CRISP-DM

Pada tahap persiapan data dilakukan *data cleaning* untuk memeriksa *missing value* dan duplikasi data; hasil pemeriksaan menunjukkan *dataset* tidak memiliki *missing value* maupun data duplikat. Selanjutnya dilakukan normalisasi menggunakan metode Min-Max Scaling pada rentang 0-1, *encoding* data kategorikal menjadi numerik, serta *feature engineering* untuk membentuk 10 fitur turunan sebagaimana ditunjukkan pada Tabel 1.]]

Tabel 1. Fitur Hasil Feature Engineering

No	Nama Fitur	Keterangan
1	AVG_NILAI_AKADEMIK	Rata-rata nilai UTS, UAS, dan tugas akhir
2	NILAI_STD	Standar deviasi nilai akademik siswa
3	SELISIH_UTS_UAS	Selisih antara nilai UTS dan UAS
4	SELISIH_UTS_AKHIR	Selisih antara nilai UTS dan tugas akhir
5	SELISIH_UAS_AKHIR	Selisih antara nilai UAS dan tugas akhir
6	PERSENTASE_KEHADIRAN	Persentase kehadiran siswa
7	JUMLAH_PELANGGARAN_DISIPLIN	Jumlah pelanggaran disiplin siswa
8	JUMLAH_KONSELING_BK	Jumlah sesi konseling BK yang diikuti
9	RISK_SCORE	Skor risiko awal berbasis kombinasi indikator
10	RATIO_NILAI_KEHADIRAN	Rasio antara nilai akademik dan kehadiran

Pada tahap pemodelan, fitur hasil normalisasi terlebih dahulu direduksi dimensinya menggunakan Principal Component Analysis (PCA) untuk mengurangi redundansi antar fitur dan noise sehingga proses deteksi anomali berjalan lebih stabil. Algoritma *Isolation Forest* kemudian diterapkan secara unsupervised dengan konfigurasi parameter $n_estimators = 200$, $contamination = 0,08$, $max_samples = auto$, dan $random_state = 42$ [7], [11]. Nilai $contamination$ sebesar 8% digunakan sebagai estimasi proporsi data yang berpotensi memiliki pola akademik berbeda dari mayoritas siswa. Keluaran model berupa label deteksi (normal/anomali), anomaly score, dan probabilitas anomali untuk setiap siswa.

Hasil deteksi *Isolation Forest* selanjutnya divalidasi menggunakan pendekatan rule-based yang disusun berdasarkan kebijakan akademik sekolah, misalnya selisih nilai UTS dan UAS lebih dari 20, kombinasi nilai akademik tinggi dengan kehadiran rendah, serta jumlah pelanggaran disiplin dan konseling BK. Setiap aturan menghasilkan skor (rule flag) yang kemudian digabungkan dengan anomaly score melalui persamaan (1) untuk menghasilkan Hybrid Risk Score.

$$\text{Hybrid Risk Score} = (0,6 \times \text{Anomaly Score}_{\text{norm}} + (0,4 \times \text{Rule Flag}_{\text{norm}}) \quad (1)$$

Bobot 60% diberikan pada hasil *Isolation Forest* sebagai metode analisis utama berbasis machine learning, sedangkan bobot 40% diberikan pada rule-based validation sebagai mekanisme validasi tambahan. Berdasarkan nilai Hybrid Risk Score, siswa dikelompokkan ke dalam tiga kategori risiko akademik, yaitu Aman, Perlu Monitoring, dan Perlu Tindak Lanjut. Seluruh tahapan analisis diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Scikit-learn, Matplotlib, dan Seaborn, dan hasilnya divisualisasikan melalui dashboard interaktif berbasis Streamlit.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan hasil-hasil penelitian dan sekaligus diberikan pembahasan yang komprehensif. Hasil dapat disajikan dalam bentuk gambar, grafik, tabel dan lain-lain yang memudahkan pembaca. Pembahasan dapat dilakukan dalam beberapa sub-bab.

3.1. Hasil Persiapan Data

Dataset penelitian terdiri atas 650 data siswa kelas XII dengan 10 atribut utama, meliputi nilai rata-rata UTS, UAS, nilai tugas akhir, rata-rata rapor, jumlah pelanggaran disiplin, persentase kehadiran, jumlah konseling BK, dan tanggal ujian. Hasil pemeriksaan data menunjukkan tidak ditemukan *missing value* maupun duplikasi data sehingga seluruh data dapat digunakan pada proses analisis, sebagaimana dirangkum pada Tabel 2.

Tabel 2. Informasi Dataset Penelitian

Keterangan	Hasil
Jumlah Data	650 siswa
Jumlah Atribut	10 atribut
Metode Analisis	CRISP-DM
Algoritma Utama	Isolation Forest
Bahasa Pemrograman	Python
Dashboard Visualisasi	Streamlit

Berdasarkan Tabel 2, seluruh data memenuhi kebutuhan proses pemodelan sehingga tahap *preprocessing* dapat dilanjutkan tanpa proses imputasi maupun penghapusan data.

3.2. Hasil Deteksi Anomali Menggunakan *Isolation Forest*

Algoritma *Isolation Forest* diterapkan pada *dataset* yang terdiri atas 650 data siswa menggunakan sepuluh fitur hasil *feature engineering*. Berdasarkan hasil pemodelan, sebanyak 598 siswa (92%) dikategorikan sebagai data normal, sedangkan 52 siswa (8%) terdeteksi sebagai data anomali. Persentase tersebut sesuai dengan nilai *contamination* sebesar 0,08 yang digunakan pada model. Ringkasan hasil deteksi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Deteksi Anomali Menggunakan *Isolation Forest*

Kategori	Jumlah Siswa	Persentase
Normal	598	92%
Anomaly	52	8%
Total	650	100%

Hasil tersebut menunjukkan bahwa sebagian besar siswa memiliki pola akademik yang masih sesuai dengan karakteristik mayoritas data, sedangkan data yang terdeteksi sebagai anomali memiliki pola yang lebih mudah diisolasi oleh algoritma *Isolation Forest* [5]. Berdasarkan hasil proses deteksi menggunakan algoritma *Isolation Forest*, sebagian besar data siswa berhasil diklasifikasikan sebagai data normal. Kategori normal menunjukkan bahwa karakteristik akademik siswa masih berada dalam pola yang umum berdasarkan kombinasi nilai akademik, tingkat kehadiran, jumlah pelanggaran disiplin, dan riwayat konseling BK. Visualisasi data normal pada Gambar 2 memberikan gambaran bahwa mayoritas siswa memiliki karakteristik yang seragam sehingga tidak teridentifikasi sebagai anomali oleh model.

5 CONTOH SISWA NORMAL

5 Data Normal

AVG_NILAI_AKADEMIK	PERSENTASE KEHADIRAN	JUMLAH PELANGGARAN DISIPLIN	JUMLAH KONSELING BK	RISK_SCORE	KETERANGAN	KARAKTERISTIK
91.250000	86	0.000000	0	14.000000	Normal	Nilai akademik baik Tidak ada pelanggaran Konseling BK rendah Risk Score rendah
83.000000	95	2.000000	1	13.000000	Normal	Nilai akademik baik Kehadiran tinggi Konseling BK rendah Risk Score rendah
85.000000	87	0.000000	0	13.000000	Normal	Nilai akademik baik Tidak ada pelanggaran Konseling BK rendah Risk Score rendah
79.000000	87	1.000000	1	18.000000	Normal	Pelanggaran rendah Konseling BK rendah
82.500000	81	1.000000	0	22.000000	Normal	Nilai akademik baik Pelanggaran rendah Konseling BK rendah

Gambar 2. Data Hasil Deteksi Normal

Berdasarkan Gambar 2, siswa pada kategori normal umumnya memiliki pola akademik yang stabil dengan rata-rata nilai yang konsisten, tingkat kehadiran yang tinggi, serta jumlah pelanggaran disiplin dan konseling BK yang relatif rendah.

Karakteristik tersebut menunjukkan bahwa data berada dalam distribusi mayoritas sehingga tidak teridentifikasi sebagai anomali oleh algoritma *Isolation Forest*.

Selain mengidentifikasi data normal, algoritma *Isolation Forest* juga mendeteksi sejumlah data yang memiliki karakteristik berbeda dari mayoritas siswa. Data tersebut diklasifikasikan sebagai anomali karena menunjukkan penyimpangan pada satu atau beberapa indikator akademik maupun nonakademik. Visualisasi pada Gambar 3 memperlihatkan contoh data siswa yang termasuk dalam kategori anomali berdasarkan hasil proses deteksi model.

AVG_NILAI_AKADEMIK	PERSENTASE KEHADIRAN	JUMLAH PELANGGARAN DISIPLIN	JUMLAH KONSELING BK	RISK_SCORE	ANOMALY_PROB	KETERANGAN	KARAKTERISTIK	
649	87.00	95	5.0	0	20.0	0.53	Anomali	Selisih UTS-UAS besar\nPelanggaran tinggi
555	83.75	80	3.0	0	29.0	0.52	Anomali	Pelanggaran tinggi\nRisk Score tinggi
648	88.75	81	3.0	3	34.0	0.52	Anomali	Nilai tinggi tetapi kehadiran rendah\nPelanggaran tinggi\nBK sering\nRisk Score tinggi
318	80.00	100	0.0	0	0.0	0.52	Anomali	Selisih UTS-UAS besar
134	92.00	84	3.0	1	27.0	0.51	Anomali	Nilai tinggi tetapi kehadiran rendah\nPelanggaran tinggi\nRisk Score tinggi

Gambar 3. Data Hasil Deteksi Anomali

Berdasarkan Gambar 3, data yang terdeteksi sebagai anomali menunjukkan karakteristik yang berbeda dibandingkan mayoritas siswa, seperti selisih nilai akademik yang tinggi, tingkat kehadiran yang rendah, jumlah pelanggaran disiplin yang lebih banyak, atau nilai *Risk Score* yang tinggi. Temuan tersebut mengindikasikan adanya pola akademik yang tidak umum sehingga memerlukan proses validasi lebih lanjut menggunakan pendekatan *rule-based* yang dijelaskan pada Subbagian 3.3.

3.3 Hasil Hybrid Validation

Hasil deteksi *Isolation Forest* kemudian digabungkan dengan *Rule-Based Validation* melalui proses *Hybrid Scoring*. Distribusi kategori risiko akademik hasil *hybrid validation* ditunjukkan pada Tabel 4 dan Gambar 4.

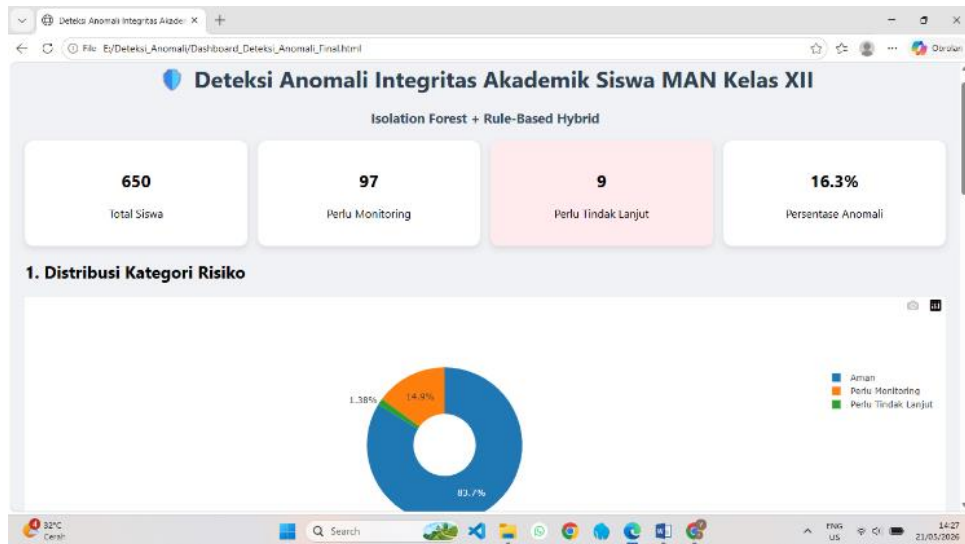
Tabel 4. Distribusi Kategori Risiko Hasil Hybrid Validation

Kategori Risiko	Jumlah Siswa	Persentase
Aman	544	83,69%
Perlu Monitoring	97	14,92%
Perlu Tindak Lanjut	9	1,39%
Total	650	100%

Berdasarkan Tabel 4, sebagian besar siswa (83,69%) berada pada kategori Aman, menunjukkan bahwa pola akademik mayoritas siswa masih berada dalam batas kewajaran. Sebanyak 97 siswa (14,92%) masuk kategori Perlu Monitoring, sedangkan 9 siswa (1,39%) termasuk kategori Perlu Tindak Lanjut. Kelompok Perlu Tindak Lanjut menunjukkan karakteristik rata-rata nilai akademik tertinggi (87,15) namun disertai tingkat kehadiran yang relatif rendah (sekitar 83%) serta jumlah pelanggaran disiplin yang lebih tinggi dibandingkan kelompok lain, sehingga menjadi prioritas verifikasi dan pendampingan lebih lanjut oleh pihak sekolah.

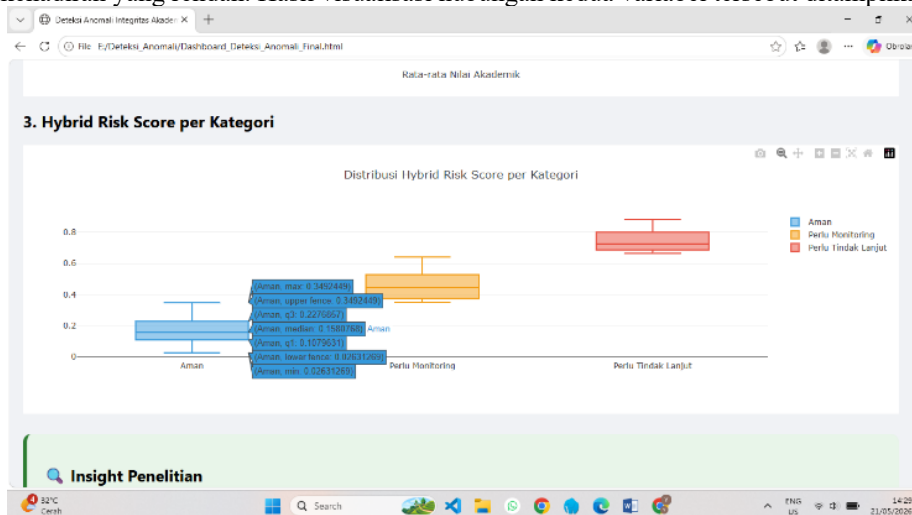
3.4 Implementasi Dashboard Monitoring

Hasil analisis yang telah diperoleh selanjutnya diimplementasikan dalam bentuk dashboard monitoring akademik berbasis Streamlit. Dashboard tersebut menyajikan informasi hasil deteksi secara interaktif, meliputi jumlah data siswa, distribusi kategori risiko, serta ringkasan hasil pemantauan akademik. Penyajian dalam bentuk dashboard bertujuan untuk mempermudah pihak sekolah dalam memahami hasil analisis dan memperoleh informasi mengenai kondisi akademik siswa secara lebih cepat. Tampilan dashboard hasil implementasi penelitian ditunjukkan pada Gambar 4.



Gambar 4. Dashboard Monitoring Akademik

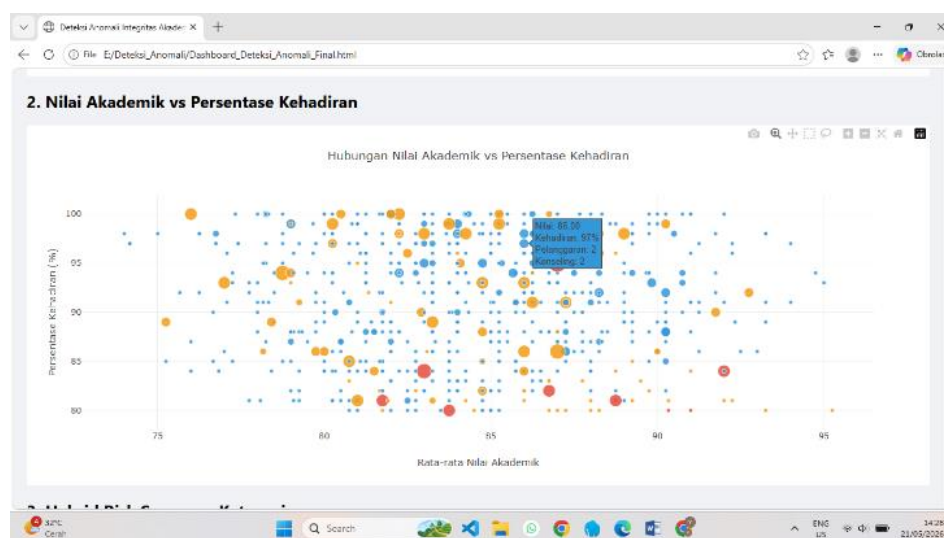
Untuk melihat hubungan antara karakteristik nilai akademik dan tingkat kehadiran siswa, dilakukan visualisasi menggunakan *scatter plot*. Visualisasi ini digunakan untuk mengamati pola persebaran siswa berdasarkan dua indikator tersebut. Hubungan antara nilai akademik dan kehadiran dapat memberikan gambaran mengenai kondisi siswa yang memiliki nilai relatif tinggi tetapi tingkat kehadiran yang rendah. Hasil visualisasi hubungan kedua variabel tersebut ditampilkan pada Gambar 5.



Gambar 5. Scatter Plot Nilai Akademik dan Kehadiran

Berdasarkan Gambar 5, sebagian besar siswa dengan tingkat kehadiran yang tinggi memiliki pola akademik yang relatif stabil. Namun, terdapat beberapa siswa dengan karakteristik yang berbeda, seperti nilai akademik tinggi tetapi tingkat kehadiran rendah, sehingga teridentifikasi sebagai data anomali.

Selain melihat hubungan antara nilai akademik dan kehadiran, hasil penelitian juga divisualisasikan berdasarkan distribusi *Hybrid Risk Score* pada setiap kategori risiko. Visualisasi ini digunakan untuk mengetahui perbedaan tingkat risiko antara kategori Aman, Perlu Monitoring, dan Perlu Tindak Lanjut. Perbedaan distribusi skor tersebut dapat memberikan gambaran mengenai tingkat risiko akademik pada masing-masing kelompok siswa. Distribusi *Hybrid Risk Score* berdasarkan kategori risiko ditampilkan pada Gambar 6.



Gambar 6. Scatter Plot Nilai Akademik dan Kehadiran

Gambar 6 menunjukkan distribusi *Hybrid Risk Score* pada setiap kategori risiko. Kategori **Perlu Tindak Lanjut** memiliki nilai risiko yang lebih tinggi dibandingkan kategori **Aman** dan **Perlu Monitoring**, sehingga memerlukan evaluasi lebih lanjut oleh pihak sekolah.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Isolation Forest* yang dipadukan dengan *hybrid validation* mampu mengidentifikasi pola akademik siswa yang berbeda dari mayoritas data. Sebagian besar siswa berada pada kategori **Aman** (83,69%), sedangkan **14,92%** termasuk kategori **Perlu Monitoring** dan **1,39%** berada pada kategori **Perlu Tindak Lanjut**. Hasil tersebut menunjukkan bahwa sebagian besar siswa memiliki pola akademik yang masih normal, sementara sebagian kecil memerlukan perhatian lebih lanjut berdasarkan kombinasi indikator akademik dan perilaku.

Siswa yang berada pada kategori **Perlu Monitoring** dan **Perlu Tindak Lanjut** umumnya menunjukkan karakteristik seperti ketidaksesuaian antara nilai akademik dan tingkat kehadiran, peningkatan jumlah pelanggaran disiplin, atau frekuensi konseling BK yang lebih tinggi. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa *Isolation Forest* efektif untuk mendeteksi anomali pada data pendidikan tanpa memerlukan data berlabel. Selain itu, penerapan *hybrid validation* membantu meningkatkan interpretasi hasil melalui penggabungan model *machine learning* dengan aturan akademik yang relevan [15].

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan **analisis data mining** menggunakan *Isolation Forest* dan *rule-based validation* dapat mendukung proses monitoring akademik secara lebih objektif dan berbasis data. Implementasi dashboard berbasis Streamlit juga memudahkan penyajian hasil analisis sehingga dapat membantu pihak sekolah dalam proses pemantauan akademik.

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode CRISP-DM untuk membangun sistem deteksi anomali integritas akademik siswa kelas XII berbasis algoritma *Isolation Forest*. Dari 650 data siswa, model berhasil mengidentifikasi 52 siswa (8%) sebagai data anomali. Melalui pendekatan *hybrid validation* yang menggabungkan hasil *Isolation Forest* dan *rule-based validation*, diperoleh 544 siswa (83,69%) berada pada kategori **Aman**, 97 siswa (14,92%) pada kategori **Perlu Monitoring**, dan 9 siswa (1,39%) pada kategori **Perlu Tindak Lanjut**. Hasil deteksi anomali divisualisasikan melalui dashboard berbasis Streamlit yang dapat dimanfaatkan sebagai pendukung pengambilan keputusan oleh pihak sekolah dalam melakukan pemantauan akademik secara objektif dan berbasis data. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan mencakup lebih banyak jenjang kelas, menambahkan atribut akademik yang lebih beragam, membandingkan algoritma *Isolation Forest* dengan metode deteksi anomali lain seperti *Local Outlier Factor* atau *One-Class SVM*, serta melakukan evaluasi performa model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* apabila tersedia data berlabel (*ground truth*).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak sekolah yang telah memberikan izin dan dukungan data dalam pelaksanaan penelitian ini, serta seluruh pihak yang telah membantu proses penyusunan naskah ini.

REFERENSI

- [1] and V. A. K. O. L. Holden, M. E. Norris, “‘Academic Integrity in Online Assessment: A Research Review.’ *Frontiers in Education*,” 2021, [Online]. Available: <https://www.frontiersin.org/journals/education/articles/10.3389/educ.2021.639752/full>
- [2] R. Kaliisa and J. Arild, “CADA : a teacher-facing learning analytics dashboard to foster teachers ’ awareness of students ’ participation

- and discourse patterns in online discussions,” pp. 937–958, 2023.
- [3] A. Pratama, “Sistem Monitoring Akademik Siswa Berbasis Dashboard.” [Online]. Available: <https://journal.unnes.ac.id/>
- [4] T. Guo, X. Bai, X. Tian, S. Firmin, and F. Xia, “Educational Anomaly Analytics : Features , Methods , and Challenges,” vol. 4, no. January, pp. 1–16, 2022, doi: 10.3389/fdata.2021.811840.
- [5] M. Yağci, “Educational data mining : prediction of students ’ academic performance using machine learning algorithms,” 2022, doi: 10.1186/s40561-022-00192-z.
- [6] R. Foorthuis, *On the nature and types of anomalies : a review of deviations in data*, vol. 12, no. 4. Springer International Publishing, 2021. doi: 10.1007/s41060-021-00265-1.
- [7] and Z. H. Z. F. T. Liu, K. M. Ting, “Scikit-learn Developers, 2024. ‘Isolation Forest Example.’ Scikit-learn Documentation.,” 2024, [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>
- [8] C. Schröer, F. Kruse, J. Marx, F. Kruse, and J. Marx, “ScienceDirect ScienceDirect A Systematic Literature Review A Systematic Literature Review on Applying Process Model on Applying CRISP-DM Process Model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [9] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, “IMPLEMENTASI DATA MINING UNTUK MEMREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4 . 5,” vol. 11, no. 2, pp. 130–138, 2015.
- [10] B. Feng and L. Zhang, “Optimizing the Isolation Forest Algorithm for Identifying Abnormal Behaviors of Students in Education Management Big Data,” pp. 31–39, 2024.
- [11] M. W. Sunarto, D. Kurniawan, E. Siswanto, and H. I. Huda, “Deteksi Anomali Menggunakan Extended Isolation Forest (Eif),” vol. 1, no. 2, pp. 96–111, 2021.
- [12] L. Ying, “Dynamic Detection and Analysis Strategy of Students’ Abnormal Behaviors Based on LOF-GMM Algorithm,” 2025, [Online]. Available: <https://ieeexplore.ieee.org/document/11010681>
- [13] H. B. and S. Pannilage, “‘Early Warning System for Academic Integrity Violations Using Machine Learning.’ Education and Information Technologies.,” 2025, [Online]. Available: <https://link.springer.com/article/10.1007/s10639-024-12999-5>
- [14] L. G. Aldawood, “A Hybrid Anomaly–Rule–Pattern Detection Framework for Streaming-Based Persistent Intrusion Detection. Informatica, 49(36).,” 2025, [Online]. Available: <https://doi.org/10.31449/inf.v49i36.12171%0A>
- [15] D. A. Stavrositu, A., Barbu, C., Muntean, M., Cristea, D.S., & Sarpe, “Hybrid Detection of Anomalies in Financial Transactions: A Rule-Based and Machine Learning Approach. Economics and Applied Informatics, (3), 162-170.,” 2025.