 ISSN:2302-8149
e-ISSN:2540-9719

# Analisis *Outlier* pada Data Performa Akademik Mahasiswa menggunakan Metode *Local Outlier Factor (LOF)*

## *Outlier Analysis of Student Academic Performance Data using the Local Outlier Factor (LOF) Method*

**[1]Made Pasek Agus Ariawan*, [2]Ni Putu Eka Apriyanthi, [3]Ida Bagus Adisimakrisna Peling**
[1,3]Program Studi Sarjana Terapan Teknologi Rekayasa Perangkat lunak, Jurusan Teknologi Informasi, Politeknik Negeri Bali
[2]Program Studi Manajemen Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Bali
*e-mail: pasekagus@pnb.ac.id

(***received*:** 10 May 2025, ***revised*:** 25 October 2026, ***accepted*:** 28 October 2026)

**Abstrak**

Transformasi digital di sektor pendidikan tinggi saat ini menghasilkan akumulasi data akademik yang masif dan kompleks. Tantangan utama dalam pengelolaan data ini adalah adanya *outlier* atau anomali yang sering kali mencerminkan kasus khusus, seperti mahasiswa dengan potensi luar biasa atau yang berisiko mengalami kegagalan studi. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Local Outlier Factor* (LOF) dalam menganalisis data *outlier* pada profil akademik mahasiswa. Metodologi penelitian melibatkan tahap pengumpulan data 82 mahasiswa (IPK, Tes Programming, dan *Project*), pra-pemrosesan dengan *Min-Max Scaling*, penerapan algoritma LOF dengan ambang batas persentil ke-90, pengujian parameter pada delapan nilai K (10 hingga 80), serta visualisasi 3D. Hasil penelitian menunjukkan bahwa metode LOF efektif dalam mengidentifikasi pencilan lokal yang tidak terdeteksi oleh metode konvensional. Dari 82 mahasiswa yang dianalisis, LOF mendeteksi 9 *outlier* (10,98% dari total sampel) pada nilai K optimal = 20, dan 6 di antaranya tidak terdeteksi oleh Z-Score. Berdasarkan analisis, ditemukan pola anomali berupa ketimpangan kompetensi, seperti mahasiswa dengan IPK dan nilai tes tinggi namun memiliki nilai *project* rendah. Pengujian parameter menunjukkan nilai K (tetangga) yang optimal sangat krusial untuk menjaga keseimbangan antara sensitivitas deteksi dan stabilitas algoritma terhadap *noise* data. Secara praktis, hasil penelitian ini dapat dimanfaatkan program studi sebagai dasar sistem peringatan dini bagi mahasiswa berisiko, penentuan kandidat pembinaan atau beasiswa bagi mahasiswa berpotensi tinggi, serta evaluasi keselarasan antara capaian akademik formal dan kompetensi praktis.
**Kata kunci:** *3D visualization, academic profile, data anomaly, local outlier factor, student.*

***Abstract***

*The digital transformation of the higher education sector has led to the accumulation of massive and increasingly complex academic datasets. One of the main challenges in managing these data is the presence of outliers or anomalies, which may reflect exceptional cases, such as students with outstanding potential or those at risk of academic failure. This study aims to implement the Local Outlier Factor (LOF) algorithm to analyze outliers in student academic performance profiles. The research methodology involved collecting data from 82 students, comprising Grade Point Average (GPA), programming test scores, and project scores; preprocessing the data using Min-Max Scaling; applying the LOF algorithm with a 90th-percentile threshold; evaluating parameter sensitivity across eight values of K (ranging from 10 to 80); and performing three-dimensional (3D) visualization. The results demonstrate that the LOF method is effective in identifying local outliers that may not be detected by conventional methods. Among the 82 students analyzed, LOF identified 9 outliers (10.98% of the total sample) at the optimal K value of 20, six of which were not detected by the Z-score method. Further analysis revealed patterns of competency imbalance, including students with high GPAs and programming test scores but relatively low project scores. Parameter evaluation indicated that selecting an optimal K (number of neighbors) is crucial for maintaining a balance between detection sensitivity and algorithmic stability in the presence of data noise. In practical terms, the findings can support academic departments in developing early warning systems for*

 ISSN:2302-8149
e-ISSN:2540-9719

*students at risk of academic failure, identifying high-potential students for academic mentoring or scholarship opportunities, and evaluating the alignment between formal academic achievements and practical competencies.*
***Keywords:*** *3D visualization, academic profile, data anomaly, local outlier factor, student.*

## 1 Pendahuluan

Transformasi digital di sektor pendidikan tinggi telah menghasilkan akumulasi data akademik yang masif dan kompleks. Data ini mencakup berbagai dimensi kompetensi mahasiswa, mulai dari pencapaian teoretis hingga keterampilan praktis. Namun, tantangan utama dalam pengelolaan data ini adalah adanya outlier atau anomali data, yaitu observasi yang menyimpang secara signifikan dari pola umum kelompoknya. Dalam perspektif akademik, outlier sering kali mencerminkan kasus khusus, seperti mahasiswa dengan potensi luar biasa yang tersembunyi atau mahasiswa yang berisiko mengalami kegagalan studi [1]

Pengukuran performa mahasiswa di bidang teknologi informasi saat ini tidak lagi cukup hanya mengandalkan Indeks Prestasi Kumulatif (IPK). Variabel pendukung seperti kemampuan teknis melalui Tes Programming dan hasil karya nyata dalam Project Mahasiswa menjadi indikator krusial dalam menentukan kesiapan kerja lulusan. Ketimpangan antara nilai IPK yang tinggi dengan kemampuan praktis yang rendah (atau sebaliknya) sering kali membentuk pola data yang tidak beraturan. Penelitian terdahulu menunjukkan bahwa deteksi anomali pada multivariabel ini sangat penting untuk memberikan evaluasi objektif yang melampaui angka administratif semata [2].

Metode global seperti k-Means atau Z-Score cenderung mengabaikan kepadatan lokal, sehingga sering kali gagal mengidentifikasi outlier yang berada di dekat cluster tertentu [3]. Sebagai solusinya, metode Local Outlier Factor (LOF) menjadi pilihan unggul karena pendekatannya yang berbasis kepadatan (density-based) [4]. LOF bekerja dengan membandingkan kepadatan sebuah titik data terhadap kepadatan tetangganya, sehingga mampu mendeteksi pencilan lokal yang sangat spesifik pada kelompok mahasiswa tertentu [5].

Penelitian oleh Jailani & Erna,(2025) Penelitian ini bertujuan mendeteksi anomali pada pola manajemen waktu dan performa akademik 2.000 mahasiswa menggunakan metode Local Outlier Factor (LOF). Fokus utama analisis adalah hubungan antara rasio jam belajar-sosial dengan capaian IPK. Hasilnya, ditemukan 118 data anomali (5,9%). Mayoritas (109 orang) memiliki waktu belajar yang sangat dominan, sementara 9 lainnya lebih aktif secara sosial. Secara akademis, 75 mahasiswa anomali meraih IPK  3,00 dan 43 lainnya di bawah itu. Temuan ini mengungkap pola tidak lazim, seperti durasi belajar tinggi namun IPK rendah, yang menjadi wawasan strategis bagi institusi pendidikan dalam memberikan pendampingan mahasiswa.

Penelitian dari [7] menggunakan algoritma DBSCAN berbasis metode *Elbow* untuk mendeteksi anomali pada nilai akademik mahasiswa informatika. Dengan mengidentifikasi *outlier* sebagai informasi signifikan, studi ini memetakan keberhasilan studi berdasarkan titik klaster baik dan buruk. Hasilnya menjadi dasar evaluasi penting untuk memperbaiki sistem pengajaran dan keberagaman penilaian mahasiswa.

Meskipun demikian, penelitian-penelitian terdahulu tersebut masih menyisakan beberapa celah (*research gap*). Pertama, deteksi anomali pada data akademik umumnya bertumpu pada satu atau dua variabel, misalnya hubungan antara rasio jam belajar–aktivitas sosial dengan IPK [6] atau nilai akademik mahasiswa [7], sehingga ketimpangan antara capaian akademik formal, uji kompetensi langsung, dan hasil karya nyata belum terpetakan dalam satu ruang analisis. Kedua, kajian komparasi metode deteksi *outlier* yang ditelaah, seperti perbandingan LOF dan *Isolation Forest* [5] serta IQR dan *Isolation Forest* [2], dilakukan pada data kinerja dosen atau klaim asuransi, bukan pada profil kompetensi mahasiswa. Ketiga, dalam kajian yang ditelaah, penjelasan mengenai bagaimana nilai parameter tetangga (K) dipilih dan bagaimana ambang batas skor LOF ditetapkan belum diuraikan secara rinci, padahal keduanya menentukan jumlah dan jenis mahasiswa yang dinyatakan sebagai *outlier* [10]. Berdasarkan penelusuran terhadap literatur yang dikaji, belum ditemukan studi yang menggabungkan IPK, nilai *Project Based Learning* (PBL), dan skor Tes Programming dalam satu ruang tiga dimensi untuk deteksi *outlier* berbasis LOF disertai analisis sensitivitas parameter K dan perbandingan dengan metode lain. Celah inilah yang hendak dijawab oleh penelitian ini.

 ISSN:2302-8149
e-ISSN:2540-9719

Penerapan LOF pada variabel IPK, Tes Programming, dan Project memungkinkan institusi untuk melakukan segmentasi yang lebih presisi. Identifikasi ini tidak bertujuan untuk menyingkirkan data anomali, melainkan sebagai dasar pengambilan kebijakan, seperti pemberian beasiswa bagi outlier berprestasi atau program pendampingan intensif bagi mahasiswa yang terdeteksi memiliki pola anomali berisiko.

Berdasarkan urgensi tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma Local Outlier Factor dalam menganalisis data outlier pada profil akademik mahasiswa. Dengan memanfaatkan referensi mutakhir dari pangkalan data Garuda Kemendiktisaintek, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem pemantauan akademik yang lebih adaptif dan akurat.

## 2 Tinjauan Literatur

Sebagaimana dikemukakan oleh [8], identifikasi terhadap pola akademik mahasiswa kini menjadi instrumen krusial bagi institusi untuk memitigasi risiko *dropout*. Dalam studinya, penggunaan algoritma berbasis kepadatan terbukti mampu memberikan gambaran yang lebih objektif dibandingkan metode statistik konvensional dalam mengenali mahasiswa yang membutuhkan intervensi khusus.

Mereka menyoroti bahwa ketergantungan pada nilai teoretis semata sering kali menciptakan bias evaluasi. Dengan memasukkan variabel praktis seperti nilai praktikum, ditemukan bahwa banyak mahasiswa yang secara administratif "normal" justru merupakan outlier lokal yang memiliki ketimpangan antara pemahaman logika dan eksekusi teknis. Hal ini sejalan dengan arah penelitian ini yang mengintegrasikan variabel IPK, Tes Programming, dan Project sebagai satu kesatuan dimensi kompetensi yang utuh.

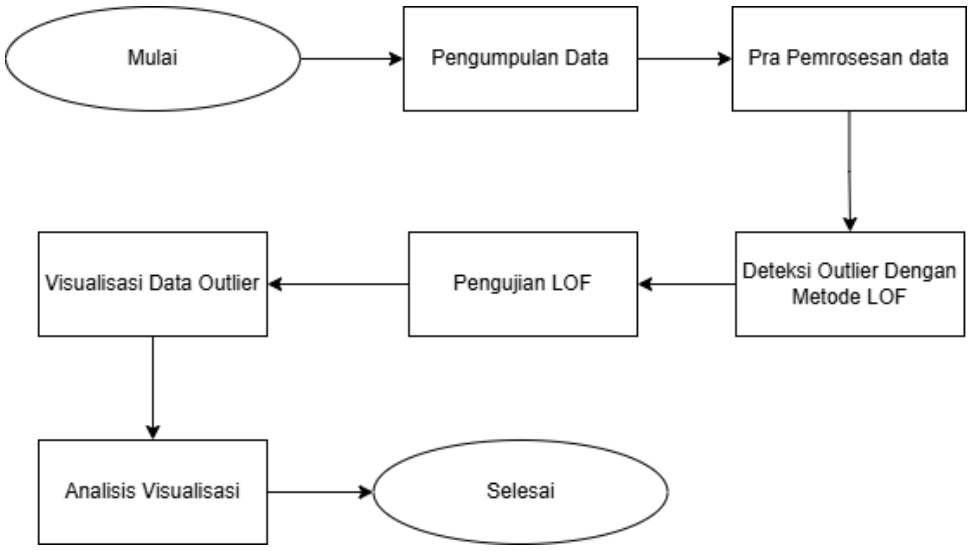
Secara metodologis, efektivitas *Local Outlier Factor* (LOF) telah banyak divalidasi sebagai algoritma yang unggul dalam menangani dataset akademik yang heterogen. [9] menjelaskan bahwa berbeda dengan algoritma klasterisasi seperti K-Means yang membagi data ke dalam kelompok besar, LOF memiliki sensitivitas tinggi terhadap pencilan di dalam sub-kelompok yang padat [10]. Keunggulan ini sangat relevan dalam data mahasiswa, di mana setiap angkatan atau konsentrasi studi memiliki standar kepadatan nilai yang berbeda-beda. Lebih lanjut, [10] menekankan pentingnya optimasi parameter tetangga (k) untuk memastikan bahwa outlier yang terdeteksi benar-benar merepresentasikan perilaku anomali, bukan sekadar derau (*noise*) data. Tatikonda et al., (2025) menjelaskan bahwa representasi data outlier dalam ruang tiga dimensi (3D) memudahkan para pengambil kebijakan di perguruan tinggi untuk memahami sebaran anomali secara intuitif.

Penilitian ini berbeda dengan penelitian-penelitian sebelumnya, studi ini tidak hanya berfokus pada deteksi anomali untuk kepentingan teknis pembersihan data, tetapi menggabungkan aspek prestasi akademik formal (IPK) dengan bukti keterampilan nyata (Project) dan uji kompetensi langsung (Tes Programming). Dengan merujuk pada kerangka pemikiran dari sepuluh jurnal terdahulu tersebut, penelitian ini memposisikan outlier sebagai subjek unik yang harus dipahami karakteristiknya guna menciptakan kebijakan akademik yang lebih inklusif dan responsif terhadap keberagaman potensi mahasiswa.

## 3 Metode Penelitian

Gambar 1. Memaparkan proses deteksi outlier pada data akademik mulai dari pengumpulan data sampai proses analisis dengan tahapan sebagai berikut

 ISSN:2302-8149
e-ISSN:2540-9719




**Gambar 1 Metodelogi penelitian**

**a. Pengumpulan Data**

Tahap awal penelitian dimulai dengan pengumpulan data akademik mahasiswa yang mencakup tiga variabel utama yaitu IPK, Tes Programming, dan Project. Data ini diperoleh dari pangkalan data institusi (atau file sumber data) yang kemudian disiapkan dalam format tabel untuk diproses lebih lanjut pada tahap berikutnya.

Data berasal dari mahasiswa semester 3 yang sudah mendapatkan matakuliah *programming* dan *project base learning*. Jumlah sampel yang dianalisis adalah $N$ = 82 mahasiswa setelah proses pembersihan data. Setiap mahasiswa direpresentasikan sebagai satu titik data pada ruang fitur tiga dimensi, yaitu IPK (skala 0–4), nilai *Project Based Learning* (PBL, skala 0–100), dan skor Tes Programming (TES, skala 0–1).

**b. Pra Pemrosesan data**

Tahap pembersihan data (*data cleaning*). Di sini data disiapkan agar siap diolah, seperti menangani data yang hilang (*missing values*), normalisasi, atau transformasi data.

Pemeriksaan data tidak menemukan nilai hilang (*missing values*) pada ketiga fitur, terdapat satu pasang observasi dengan nilai identik pada ketiga fitur yang tetap dipertahankan sehingga $N$ = 82. Karena ketiga fitur memiliki rentang nilai yang berbeda, seluruh fitur dinormalisasi ke rentang [0, 1] dengan *Min-Max Scaling* menurut Persamaan (1) [13], [14].

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

dengan $x_{ij}$ adalah nilai asli fitur ke-$j$ pada mahasiswa ke-$i$, sedangkan $\min(x_j)$ dan $\max(x_j)$ adalah nilai minimum dan maksimum fitur ke-$j$ pada seluruh sampel. Pada data penelitian ini, nilai minimum dan maksimum masing-masing adalah 3,32 dan 4,00 untuk IPK, 66,00 dan 93,00 untuk PBL, serta 0,46 dan 1,00 untuk TES.

**c. Deteksi Outlier dengan Metode LOF**

Setelah data terkumpul dan dibersihkan (*pre-processing*), langkah inti adalah menerapkan algoritma Local Outlier Factor (LOF). Pada tahap ini, sistem menghitung kepadatan lokal dari setiap titik data mahasiswa relatif terhadap tetangga terdekatnya (k-distance). Mahasiswa yang memiliki kepadatan jauh lebih rendah dibandingkan kelompok tetangganya akan diidentifikasi sebagai data outlier.

Algoritma LOF [15] mengukur derajat keterasingan lokal suatu titik dengan membandingkan kepadatan lokalnya terhadap kepadatan lokal $k$ tetangga terdekatnya. Jarak antar-mahasiswa dihitung dengan jarak Euclidean pada ruang tiga fitur ternormalisasi (Persamaan 2). Untuk setiap titik $p$, $k$-distance($p$) adalah jarak ke tetangga terdekat ke-$k$, dan $N_k(p)$ adalah himpunan $k$ tetangga terdekat

 ISSN:2302-8149
e-ISSN:2540-9719

tersebut. Jarak keterjangkauan (*reachability distance*), kepadatan keterjangkauan lokal (*local reachability density*, lrd), dan skor LOF dirumuskan pada Persamaan (3) sampai dengan (5).

$$d(p,q) = \sqrt{\sum_{j=1}^{3}\left(p_j - q_j\right)^2} \tag{2}$$

$$\text{reach-dist}_k(p,o) = \max\{k\text{-distance}(o),\ d(p,o)\} \tag{3}$$

$$\text{lrd}_k(p) = \frac{|N_k(p)|}{\sum_{o\in N_k(p)} \text{reach-dist}_k\,(p,o)} \tag{4}$$

$$\text{LOF}_k(p) = \frac{1}{|N_k(p)|}\sum_{o\in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)} \tag{5}$$

Skor LOF yang mendekati 1 menunjukkan bahwa kepadatan sebuah titik sebanding dengan kepadatan tetangganya (*inlier*), skor di bawah 1 menunjukkan titik berada pada area yang lebih padat daripada tetangganya, sedangkan skor yang jauh lebih besar daripada 1 menunjukkan titik berada pada area yang lebih renggang daripada tetangganya sehingga diindikasikan sebagai *outlier* lokal.

Karena tidak ada batas skor LOF yang berlaku universal, ambang batas (*threshold*) $\theta_k$ ditetapkan secara adaptif sebagai persentil ke-90 dari skor LOF seluruh mahasiswa pada setiap nilai *K*, yaitu $\theta_k = P_{90}\{\text{LOF}_k(p_1), \ldots, \text{LOF}_k(p_N)\}$, sehingga sekitar 10% mahasiswa dengan skor tertinggi dinyatakan sebagai *outlier* menurut Persamaan (6). Persentil dipilih karena bersifat nonparametrik: sebaran skor LOF condong ke kanan dan berbatas bawah di sekitar 1 (pada *K* = 20, rerata 1,1925 lebih besar daripada median 1,0728) sehingga asumsi sebaran normal, misalnya rerata + 2 simpangan baku, kurang sesuai. Selain itu, ambang batas berbasis peringkat dapat dibandingkan antar-*K* dan menghasilkan daftar *outlier* yang jumlahnya dapat ditindaklanjuti oleh program studi. Persentil ke-90 merupakan keputusan desain penelitian yang dapat disesuaikan, konsekuensinya, jumlah *outlier* pada setiap *K* relatif tetap (sekitar 10% dari *N*), sehingga *K* memengaruhi siapa yang terdeteksi dan seberapa jelas keterpisahannya, bukan berapa banyaknya. Nilai rerata skor ($\mu_k$) dan ambang batas ($\theta_k$) untuk setiap *K* disajikan pada kolom Mean dan Threshold di Tabel 3.

$$\text{Status}(p) = \begin{cases}\text{Outlier}, & \text{LOF}_k(p) > \theta_k \\ \text{Inlier}, & \text{LOF}_k(p) \le \theta_k\end{cases} \tag{6}$$

**Spesifikasi perangkat lunak dan parameter.** Seluruh tahapan dijalankan pada *Google Colab* dengan bahasa pemrograman Python 3. Pustaka yang digunakan adalah pandas dan NumPy untuk pengolahan data, scikit-learn [16] (MinMaxScaler dan LocalOutlierFactor) untuk normalisasi dan perhitungan skor LOF, serta Matplotlib untuk visualisasi 3D. Parameter LOF yang digunakan adalah n_neighbors = *K* dengan $K \in \{10, 20, \ldots, 80\}$, metric = 'minkowski' dengan p = 2 (jarak Euclidean), algorithm = 'auto', contamination = 'auto', dan novelty = False (nilai bawaan scikit-learn selain n_neighbors). Skor LOF diperoleh dari nilai negative_outlier_factor_ yang dikalikan −1.

### d. Pengujian LOF

Tahap pengujian dilakukan untuk memastikan bahwa parameter yang digunakan (seperti nilai k atau *contamination rate*) telah optimal. Pengujian ini bertujuan untuk memvalidasi apakah skor LOF yang dihasilkan benar-benar merepresentasikan anomali yang signifikan, sehingga dapat dibedakan antara pencilan lokal yang bermakna dan *noise* data biasa.

Pengujian dilakukan pada delapan nilai *K* (10, 20, …, 80). Nilai *K* optimal dipilih berdasarkan empat pertimbangan: (1) *K* tidak kurang dari 10 untuk menghindari fluktuasi statistik [15] dan jauh lebih kecil daripada jumlah sampel *N*, (2) stabilitas ambang batas pada nilai *K* yang berurutan, (3) besarnya selisih antara ambang batas dan rata-rata skor LOF (*threshold* − *mean*) sebagai indikator keterpisahan *outlier* dari *inlier*, serta (4) keterulangan keanggotaan himpunan *outlier* pada nilai *K* yang bersebelahan, yaitu apakah mahasiswa yang ditandai pada suatu *K* juga ditandai pada *K* di sekitarnya.

Sebagai pembanding, hasil LOF pada *K* optimal dibandingkan dengan tiga metode deteksi *outlier* lain pada data yang sama, yaitu Z-Score (mahasiswa dengan $|z| > 3$ pada sedikitnya satu fitur),

 ISSN:2302-8149
e-ISSN:2540-9719

*Isolation Forest* [17] dengan n_estimators = 100, contamination = 0,1098 (setara proporsi *outlier* LOF), dan random_state = 42, serta DBSCAN [18] dengan eps = 0,1836 (titik siku grafik *k*-distance) dan min_samples = 6 (dua kali jumlah fitur), dengan titik *noise* pada DBSCAN dianggap sebagai *outlier*.

**e. Visualisasi Data Outlier**

Data yang telah diproses kemudian dipetakan ke dalam bentuk visualisasi, khususnya menggunakan Scatter Plot 3D. Visualisasi ini sangat penting untuk memberikan gambaran spasial mengenai posisi data outlier terhadap kelompok data mahasiswa lainnya (*inlier*), sehingga pola penyimpangan dapat terlihat secara intuitif. Visualisasi data merupakan teknik yang digunakan untuk menyajikan informasi kompleks dalam bentuk grafis sehingga lebih mudah dipahami oleh pengguna. Peran visualisasi tidak hanya sebatas menampilkan data, tetapi juga mengkomunikasikan pola, tren, dan hubungan di dalam data agar dapat digunakan dalam proses pengambilan Keputusan [12].

Pada *scatter plot* 3D, sumbu *x*, *y*, dan *z* masing-masing merepresentasikan IPK, PBL, dan TES ternormalisasi, *inlier* ditampilkan dengan warna biru, sedangkan *outlier* ditampilkan dengan warna merah, penanda yang lebih besar, dan label nomor urut sesuai Tabel 4.

**f. Analisis Visualisasi**

Tahap akhir adalah melakukan analisis mendalam terhadap hasil visualisasi. Pada fase ini, dilakukan interpretasi terhadap temuan outlier, misalnya mengidentifikasi karakteristik mahasiswa yang masuk dalam kategori outlier positif (prestasi praktis melampaui rata-rata) maupun outlier berisiko (penurunan performa signifikan). Hasil analisis ini menjadi dasar bagi penarikan kesimpulan dan rekomendasi kebijakan akademik.

## 4 Hasil dan Pembahasan

### a. Pengumpulan Data

**Tabel 1 Pengumpulan data awal**

| NO | NIM | IPK | PBL | Tes |
|---|---|---|---|---|
| **1** | 23xxxxxx01 | 3.95 | 85.975 | 0.96 |
| **2** | 23xxxxxx04 | 3.68 | 93 | 0.96 |
| **3** | 23xxxxxx05 | 4 | 85.975 | 0.88 |
| **4** | 23xxxxxx09 | 4 | 93 | 0.92 |
| **5** | 23xxxxxx13 | 3.95 | 93 | 0.8 |
| **6** | 23xxxxxx16 | 3.89 | 82.6 | 0.88 |
| **7** | 23xxxxxx17 | 3.68 | 85.875 | 0.92 |
| **8** | 23xxxxxx21 | 3.58 | 85 | 0.76 |
| **9** | 23xxxxxx25 | 3.58 | 73.225 | 0.69 |
| **10** | 23xxxxxx28 | 3.95 | 93 | 0.96 |
| **11** | 23xxxxxx29 | 3.53 | 72.4375 | 0.5 |
| **12** | 23xxxxxx33 | 3.65 | 72.7875 | 0.69 |
| **13** | 23xxxxxx37 | 3.95 | 82 | 0.92 |
| **14** | 23xxxxxx40 | 3.47 | 72.625 | 0.77 |
| **15** | 23xxxxxx41 | 3.71 | 82 | 0.92 |

Tabel 1 merupakan data yang didapatkan dari berbagai sumber. IPK didapatkan dari sistem internal yang ada di perguruan tinggi. PBL merupakan nilai dari Project yang dilakukan mahasiswa selama satu semester. Tes merupakan hasil tes secara mandiri yang dilakukan oleh peneliti ke mahasiswa nilai tes merupakan kombinasi dari tes teori, praktek programming dan waktu pengerjaan tes.

Setelah proses pembersihan, jumlah mahasiswa yang dianalisis adalah $N$ = 82 mahasiswa, Tabel 1 hanya menampilkan 15 data sebagai contoh. Pada sampel tersebut IPK berkisar antara 3,32 dan 4,00

 ISSN:2302-8149
e-ISSN:2540-9719

(rerata 3,73, simpangan baku 0,17), nilai PBL antara 66,00 dan 93,00 (rerata 83,79, simpangan baku 5,77), serta skor TES antara 0,46 dan 1,00 (rerata 0,84, simpangan baku 0,12).

**b. Pra Pemrosesan Data**

**Tabel 2 Pra pemrosesan data**

| No | NIM | scaled_IPK | scaled_PBL | scaled_TES |
|---|---|---|---|---|
| **1** | 23xxxxxx01 | 0.926470588 | 0.739814815 | 0.925925926 |
| **2** | 23xxxxxx04 | 0.529411765 | 1 | 0.925925926 |
| **3** | 23xxxxxx05 | 1 | 0.739814815 | 0.777777778 |
| **4** | 23xxxxxx09 | 1 | 1 | 0.851851852 |
| **5** | 23xxxxxx13 | 0.926470588 | 1 | 0.62962963 |
| **6** | 23xxxxxx16 | 0.838235294 | 0.614814815 | 0.777777778 |
| **7** | 23xxxxxx17 | 0.529411765 | 0.736111111 | 0.851851852 |
| **8** | 23xxxxxx21 | 0.382352941 | 0.703703704 | 0.555555556 |
| **9** | 23xxxxxx25 | 0.382352941 | 0.267592593 | 0.425925926 |
| **10** | 23xxxxxx28 | 0.926470588 | 1 | 0.925925926 |
| **11** | 23xxxxxx29 | 0.308823529 | 0.238425926 | 0.074074074 |
| **12** | 23xxxxxx33 | 0.485294118 | 0.251388889 | 0.425925926 |
| **13** | 23xxxxxx37 | 0.926470588 | 0.592592593 | 0.851851852 |
| **14** | 23xxxxxx40 | 0.220588235 | 0.24537037 | 0.574074074 |
| **15** | 23xxxxxx41 | 0.573529412 | 0.592592593 | 0.851851852 |

Setelah data mahasiswa yang meliputi Indeks Prestasi Kumulatif (IPK), nilai *Project Based Learning* (PBL), dan skor TES dikumpulkan, data tersebut tidak langsung diolah. Mengingat setiap variabel memiliki rentang nilai yang berbeda, misalnya IPK dalam skala 0–4 sementara TES mungkin dalam skala 0–100, maka dilakukanlah tahap Normalisasi menggunakan metode Min-Max Scaling.
Dalam Tabel 2, seluruh nilai telah dikonversi ke dalam rentang 0 hingga 1. Angka 1 merepresentasikan nilai tertinggi yang ada pada dataset asli, sedangkan angka yang mendekati 0 merepresentasikan nilai terendah. Sebagai contoh, pada kolom scaled_PBL, mahasiswa dengan NIM akhiran 09 dan 13 memiliki nilai 1, yang berarti mereka adalah pemilik skor PBL tertinggi di antara seluruh sampel. Sebaliknya, pada kolom scaled_TES, mahasiswa dengan NIM akhiran 29 memiliki nilai sangat rendah, yaitu 0.074, yang menunjukkan posisi skornya berada di batas bawah populasi.

Sebagai contoh perhitungan Persamaan (1), IPK mahasiswa No. 1 sebesar 3,95 dinormalisasi menjadi (3,95 − 3,32)/(4,00 − 3,32) = 0,9265, sedangkan nilai PBL sebesar 85,975 menjadi (85,975 − 66,00)/(93,00 − 66,00) = 0,7398, sesuai dengan nilai pada Tabel 2.
Tujuan utama dari pembuatan skala yang seragam ini adalah untuk mempersiapkan data sebelum masuk ke tahap Deteksi Outlier menggunakan metode Local Outlier Factor (LOF), dengan mengubah semua fitur ke dalam skala yang sama, algoritma dapat menghitung jarak antar-mahasiswa secara adil dan akurat [13]. Tanpa tahap ini, variabel dengan rentang angka yang besar akan mendominasi perhitungan, sehingga hasil deteksi outlier menjadi tidak valid [14].

**c. Implementasi**

Tahap inti dari penelitian ini adalah penerapan algoritma Local Outlier Factor (LOF) untuk mengidentifikasi mahasiswa dengan pola akademik yang anomali. Setelah melalui tahap pra-pemrosesan menggunakan *Min-Max Scaling*, seluruh fitur (IPK, PBL, dan TES) berada pada skala yang setara, sehingga perhitungan jarak antar-titik data menjadi objektif.
Proses deteksi dimulai dengan menentukan k-distance, yaitu jarak antara suatu titik data mahasiswa terhadap tetangga terdekat ke-k. Dalam analisis ini, setiap data mahasiswa dibandingkan dengan kepadatan lokal kelompok di sekitarnya. Algoritma LOF bekerja dengan prinsip bahwa data normal (*inlier*) cenderung memiliki kepadatan yang serupa dengan tetangganya, sementara data pencilan (*outlier*) akan berada di wilayah yang jauh lebih renggang.

 ISSN:2302-8149
e-ISSN:2540-9719

Hasil perhitungan menunjukkan bahwa sistem berhasil memberikan skor LOF pada setiap NIM. Mahasiswa yang memiliki skor LOF tinggi (melebihi ambang batas tertentu) dikategorikan sebagai pencilan. Hal ini terjadi karena terdapat perbedaan mencolok pada kombinasi nilai mereka, misalnya, seorang mahasiswa mungkin memiliki skor TES yang sangat tinggi namun nilai PBL yang sangat rendah secara ekstrem dibandingkan kelompok mahasiswa dengan profil serupa.
Interpretasi dari tahap ini memberikan wawasan penting untuk terdeteksi bukan sekadar "data yang salah", melainkan representasi dari mahasiswa yang membutuhkan perhatian khusus atau memiliki karakteristik belajar yang berbeda secara signifikan dari pola umum populasi mahasiswa lainnya.

Pada konfigurasi optimal $K$ = 20 dengan ambang batas θ = 1,5458, dari $N$ = 82 mahasiswa terdapat 9 mahasiswa yang skor LOF-nya melampaui ambang batas, sehingga proporsi *outlier* yang terdeteksi adalah 10,98% dari total sampel (9/82), sedangkan 73 mahasiswa lainnya dikategorikan sebagai *inlier*. Skor LOF tertinggi, yaitu sekitar 2,13, dimiliki oleh mahasiswa No. 57 (Tabel 4).

**d. Pengujian**

Setelah tahap pra-pemrosesan data selesai, dilakukan tahap Pengujian LOF untuk menentukan konfigurasi algoritma yang paling optimal. Parameter utama yang diuji adalah nilai K (Neighbors), yang merepresentasikan jumlah tetangga terdekat yang digunakan untuk menghitung kepadatan lokal setiap data mahasiswa. Pengujian dilakukan sebanyak delapan kali dengan rentang nilai K antara 10 hingga 80.

Berdasarkan hasil pengujian yang tersaji pada tabel, terlihat adanya korelasi yang signifikan antara jumlah tetangga (K) dengan rata-rata skor LOF (Mean) serta nilai ambang batas (Threshold). Pada pengujian awal dengan K=10, diperoleh nilai *threshold* sebesar 1.5399. Tingginya nilai *threshold* pada K yang rendah menunjukkan bahwa algoritma sangat sensitif terhadap perbedaan kepadatan dalam lingkup lokal yang sempit. Dalam kondisi ini, hanya mahasiswa dengan profil nilai yang benar-benar ekstrem yang akan terdeteksi sebagai pencilan.

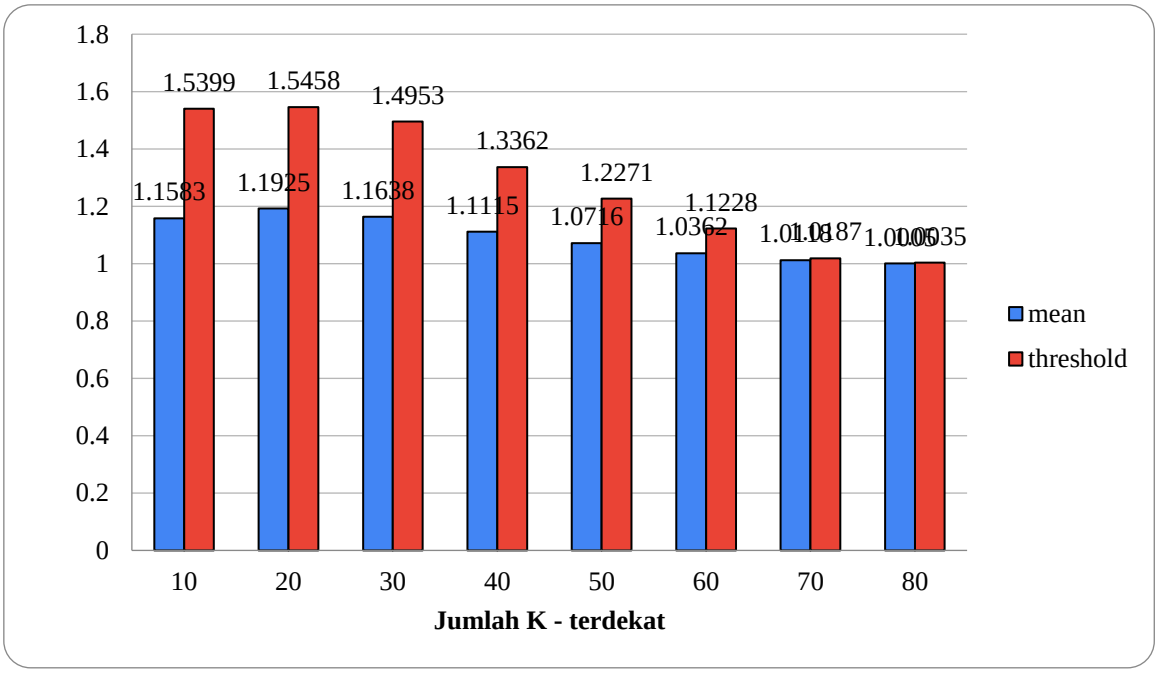
Seiring dengan meningkatnya nilai K, ditemukan tren penurunan pada nilai *threshold*. Pada pengujian ke-7 dan ke-8 dengan K=70 dan K=80, nilai *threshold* menyusut hingga ke angka 1.0187 dan 1.0035. Hal ini mengindikasikan bahwa ketika cakupan tetangga diperluas, perbedaan kepadatan antar-titik data mahasiswa menjadi lebih halus atau seragam secara global. Penurunan *threshold* ini perlu diwaspadai dalam analisis, karena jika nilai batas terlalu rendah, terdapat risiko data mahasiswa normal akan ikut terklasifikasi sebagai *outlier* (*false positive*).

Melalui rangkaian pengujian ini, dapat disimpulkan bahwa pemilihan nilai K yang tepat sangat bergantung pada distribusi data mahasiswa di lapangan. Penentuan parameter ini bertujuan untuk menemukan titik keseimbangan (*optimum point*) di mana algoritma cukup sensitif untuk menangkap anomali akademik (seperti ketimpangan antara nilai IPK dan skor TES), namun tetap stabil sehingga tidak salah dalam mengategorikan mahasiswa reguler sebagai pencilan.

**Tabel 3 Pengujian metode LOF**

| Pengujian | K | Mean | Threshold |
|---|---|---|---|
| **1** | 10 | 1.1583 | 1.5399 |
| **2** | 20 | 1.1925 | 1.5458 |
| **3** | 30 | 1.1638 | 1.4953 |
| **4** | 40 | 1.1115 | 1.3362 |
| **5** | 50 | 1.0716 | 1.2271 |
| **6** | 60 | 1.0362 | 1.1228 |
| **7** | 70 | 1.0118 | 1.0187 |
| **8** | 80 | 1.0005 | 1.0035 |

 ISSN:2302-8149
e-ISSN:2540-9719




**Gambar 2 Metodelogi penelitian**

Untuk memperkuat analisis kuantitatif, hasil pengujian parameter K divisualisasikan ke dalam bentuk grafik batang yang membandingkan nilai rata-rata skor LOF (*Mean*) dengan nilai ambang batas (*Threshold*). Grafik ini memberikan gambaran komprehensif mengenai stabilitas algoritma Local Outlier Factor dalam merespons perubahan jumlah tetangga terdekat yang digunakan.

Berdasarkan grafik tersebut, terlihat bahwa nilai Mean (batang biru) menunjukkan tren penurunan yang sangat konsisten. Dimulai dari angka 1,1583 pada K=10 dan terus melandai hingga mencapai titik terendahnya sebesar 1,0005 pada K=80. Secara teknis, penurunan ini mengindikasikan bahwa semakin besar cakupan tetangga yang dilibatkan, maka perbedaan kepadatan lokal antar-mahasiswa akan terlihat semakin seragam. Titik data yang semula dianggap jauh dari lingkungannya, secara perlahan mulai "terserap" ke dalam kepadatan populasi yang lebih luas, sehingga rata-rata skor anomali mendekati angka 1 (angka ideal untuk data normal).

Di sisi lain, nilai Threshold (batang merah) menunjukkan dinamika yang lebih fluktuatif namun cenderung menurun tajam setelah melewati K=20. Puncak *threshold* tertinggi berada pada pengujian kedua (1,5458), yang menandakan bahwa pada lingkungan lokal yang kecil, model memiliki standar yang sangat ketat untuk menetapkan sebuah data sebagai pencilan. Namun, pengamatan paling krusial terlihat pada K=70 dan K=80, di mana selisih antara *Mean* dan *Threshold* berada pada titik paling rapat (masing-masing 0,0069 dan 0,0030). Kondisi ini menunjukkan bahwa pada K ≥ 70, batas pemisah antara mahasiswa reguler dan mahasiswa pencilan menjadi sangat tipis, yang dapat menurunkan tingkat selektivitas algoritma.

Menariknya, pada K=80, nilai *threshold* (1,0035) praktis berimpit dengan rata-rata skor (1,0005), sehingga skor LOF hampir seluruh mahasiswa mendekati 1. Hal ini terjadi karena cakupan 80 tetangga hampir meliputi seluruh sampel ($N$ = 82), sehingga setiap mahasiswa dibandingkan dengan hampir seluruh data dan perbedaan kepadatan lokal menghilang. Melalui visualisasi ini, dapat disimpulkan bahwa pemilihan nilai K yang ideal berada pada rentang di mana selisih antara batang biru dan merah cukup signifikan, guna menjamin bahwa mahasiswa yang terdeteksi sebagai *outlier* benar-benar memiliki profil akademik yang unik dan signifikan secara statistik.

**Penentuan nilai K optimal.** Berdasarkan Tabel 3, Gambar 2, dan keanggotaan himpunan *outlier* pada setiap $K$, $K$ = 20 dipilih sebagai nilai optimal dengan empat alasan. Pertama, $K$ = 20 berada pada wilayah ambang batas yang stabil (*plateau*), yaitu $K$ = 10 sampai 30, tempat ambang batas hanya berubah dari 1,5399 menjadi 1,5458 (naik 0,4%) dan kemudian 1,4953 (turun 3,3%). Di luar wilayah ini ambang batas turun sebesar 8–11% pada setiap kenaikan $K$ sebesar 10 (dari 1,4953 pada $K$ = 30 menjadi 1,0187 pada $K$ = 70) dan hanya 1,0035 pada $K$ = 80, yang menandakan perbedaan kepadatan lokal tersamarkan karena cakupan tetangga meluas ke kelompok mahasiswa lain.

 ISSN:2302-8149
e-ISSN:2540-9719

Kedua, pada $K$ = 20 rata-rata skor LOF (1,1925) dan ambang batas (1,5458) mencapai nilai tertinggi dari seluruh pengujian, sehingga variasi kepadatan lokal antar-mahasiswa paling terekspresikan dan *outlier* lokal lebih mudah dipisahkan dari kelompoknya. Ketiga, meskipun selisih ambang batas dan rata-rata pada $K$ = 10 (0,3816) sedikit lebih besar daripada $K$ = 20 (0,3533), $K$ = 10 berada tepat pada batas bawah yang disarankan [15] sehingga estimasi kepadatan lokalnya bertumpu pada sedikit tetangga dan lebih rentan terhadap fluktuasi acak, $K$ = 20 menghasilkan ambang batas yang hampir identik (selisih 0,0059) dengan estimasi kepadatan yang lebih stabil. Keempat, ditinjau dari keanggotaan himpunan *outlier*, $K$ = 20 menjembatani dua rezim: lima dari sembilan *outlier* pada $K$ = 20 (No. 26, 28, 31, 37, dan 57) juga terdeteksi pada $K$ = 10, dan lima lainnya (No. 20, 47, 50, 56, dan 57) juga terdeteksi pada $K$ = 30 sampai 60, sehingga setiap *outlier* pada $K$ = 20 didukung oleh sedikitnya satu nilai $K$ di sekitarnya, mahasiswa No. 57 bahkan terdeteksi pada seluruh $K$ = 10 sampai 70. Pada $K \geq 70$ irisan dengan himpunan $K$ = 20 menyusut menjadi 4 dan 2 mahasiswa.

Seluruh nilai $K$ menghasilkan jumlah *outlier* yang sama (9 mahasiswa) karena ambang batas ditetapkan pada persentil ke-90, sehingga $K$ memengaruhi siapa yang terdeteksi, bukan berapa banyaknya. Pada $K$ = 20 terdeteksi 9 *outlier* (10,98% dari $N$ = 82). Tiga di antaranya (No. 31, 47, dan 37 pada Tabel 4) memiliki skor 1,5471–1,5600 yang sangat dekat dengan ambang batas, dan mahasiswa peringkat ke-10 (skor 1,5336) hanya berselisih 0,0122 dari ambang batas, sehingga batas antara *outlier* dan *inlier* bersifat landai dan mahasiswa marginal perlu diverifikasi, misalnya melalui konfirmasi dosen wali. Ketiganya tetap memiliki dukungan dari $K$ lain (No. 31 dan 37 pada $K$ = 10, No. 47 pada $K$ = 30 sampai 60). Enam *outlier* lainnya memiliki skor 1,6780–2,1295 yang jelas terpisah dari ambang batas.

**e. Analisis Visualisasi**

**Tabel 4 Visualisasi outlier**

| No | NIM | IPK | PBL | TES | LOF_Score | Status |
|---|---|---|---|---|---|---|
| **57** | 2315354007 | 3,63 | 66 | 0,96 | 2,129492955 | Outlier |
| **50** | 2315354066 | 3,32 | 83 | 0,46 | 2,036158486 | Outlier |
| **56** | 2315354003 | 3,79 | 66 | 0,8 | 1,95035907 | Outlier |
| **20** | 2315354057 | 3,89 | 73 | 0,54 | 1,721328291 | Outlier |
| **26** | 2315354073 | 3,74 | 72,375 | 0,88 | 1,710293302 | Outlier |
| **28** | 2315354081 | 3,39 | 81 | 0,96 | 1,678027416 | Outlier |
| **31** | 2315354010 | 3,79 | 82,7 | 0,61 | 1,559991249 | Outlier |
| **47** | 2315354056 | 3,47 | 83,25 | 0,61 | 1,548697822 | Outlier |
| **37** | 2315354026 | 3,47 | 81,3 | 0,69 | 1,547148064 | Outlier |

**1. Karakteristik Skor LOF**

Seluruh mahasiswa dalam daftar ini memiliki LOF_Score di atas nilai ambang batas (*threshold*). Mahasiswa dengan nomor urut 57 (NIM akhir 007) memiliki skor tertinggi yaitu 2,1294. Dalam logika LOF, semakin tinggi skor tersebut di atas angka 1, semakin rendah kepadatan lokal mahasiswa tersebut dibandingkan tetangga terdekatnya. Hal ini menunjukkan bahwa mahasiswa tersebut memiliki profil akademik yang sangat unik atau berbeda signifikan dibandingkan rekan-rekan dengan latar belakang nilai serupa.

**2. Pola Anomali Akademik**

Jika dianalisis secara manual, kita dapat melihat alasan mengapa data-data ini dikategorikan sebagai pencilan antara lain

a. Ketidakseimbangan antar-variabel yang telihat dari data mahasiswa pada baris pertama (NIM akhir 007). Mahasiswa ini memiliki IPK yang cukup baik (3,63) dan TES yang sangat tinggi (0,96), namun nilai PBL-nya relatif rendah (66) dibandingkan mahasiswa lain dengan IPK setara.
b. Variasi Nilai TES dari data Mahasiswa pada nomor urut 50 memiliki nilai TES yang sangat rendah (0,46) padahal memiliki nilai PBL yang cukup tinggi (83). Perbedaan

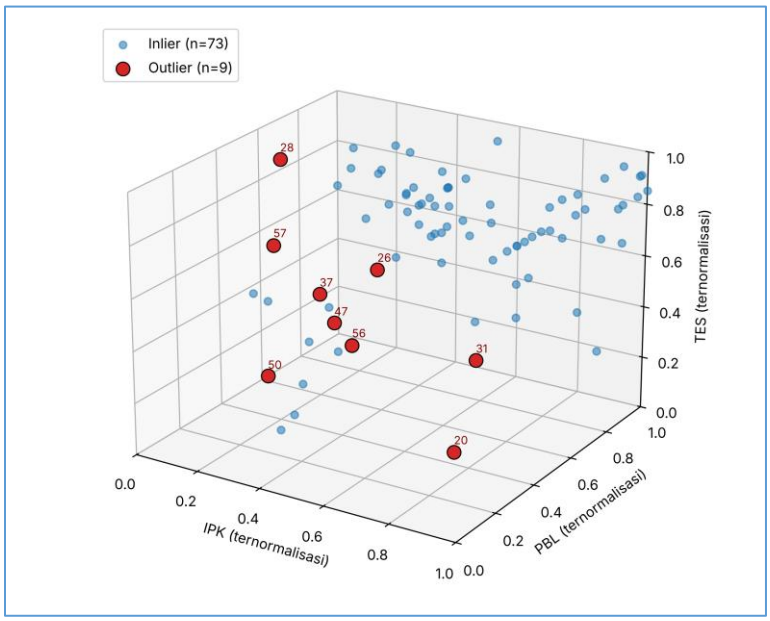
 ISSN:2302-8149
e-ISSN:2540-9719

kontras antara kemampuan di kelas (PBL) dan hasil ujian (TES) inilah yang dideteksi oleh algoritma sebagai sebuah anomali lokal.

Secara keseluruhan, sembilan *outlier* pada Tabel 4 dapat dikelompokkan ke dalam tiga pola dominan: (i) PBL rendah dibandingkan TES dan IPK (No. 57, 56, dan 26), (ii) TES rendah dibandingkan IPK dan/atau PBL (No. 50, 20, dan 31), serta (iii) IPK rendah dibandingkan TES dan/atau PBL (No. 28, 47, dan 37).

### 3. Visualisasi Scatter Plot 3D

Gambar 3 menampilkan *scatter plot* 3D hasil deteksi LOF pada $K$ = 20. Mayoritas *inlier* (73 mahasiswa) membentuk gugus yang padat pada rentang antarkuartil IPK ternormalisasi 0,46–0,84, PBL 0,64–0,77, dan TES 0,63–0,85, dengan pusat gugus (median) di sekitar (0,57, 0,70, 0,78). Sembilan *outlier* berada di luar gugus inti tersebut pada jarak Euclidean 0,47–0,97 dari pusat gugus (rerata 0,65), sedangkan median jarak *inlier* hanya 0,29. Mahasiswa No. 57 dan 56 (PBL ternormalisasi 0) dan No. 50 (IPK dan TES ternormalisasi 0) berada di sudut ruang fitur, sedangkan No. 26, 28, 31, dan 37 berada paling dekat dengan gugus (jarak 0,47–0,52) namun pada wilayah dengan kepadatan lokal yang lebih renggang. Perlu dicatat bahwa terdapat *inlier* yang berjarak hingga 0,88 dari pusat gugus namun tidak ditandai, sebab status *outlier* pada LOF ditentukan oleh kepadatan lokal terhadap tetangganya, bukan oleh jarak ke pusat data.




**Gambar 3 Scatter plot 3D hasil deteksi outlier LOF (K = 20)**

### f. Perbandingan LOF dengan Metode Lain

Untuk menguji klaim bahwa LOF mampu mengidentifikasi pencilan lokal yang tidak terdeteksi oleh metode konvensional, hasil LOF pada $K$ = 20 dibandingkan dengan Z-Score, *Isolation Forest*, dan DBSCAN pada data yang sama ($N$ = 82). Hasil perbandingan disajikan pada Tabel 5.

**Tabel 5 Perbandingan hasil deteksi outlier antar-metode**

| Metode | Parameter | Outlier terdeteksi | Persentase (%) | Irisan dengan LOF |
|---|---|---|---|---|
| **LOF** | $K$ = 20, θ = 1,5458 (persentil ke-90) | 9 | 10,98 | – |
| **Z-Score** | $\|z\| > 3$ | 3 | 3,66 | 3 dari 9 |
| ***Isolation Forest*** | n_estimators = 100, contamination = 0,1098 | 9 | 10,98 | 5 dari 9 |

 ISSN:2302-8149
e-ISSN:2540-9719

| | | | | |
|---|---|---|---|---|
| **DBSCAN** | eps = 0,1836, min_samples = 6 | 19 | 23,17 | 8 dari 9 |

Perbedaan hasil antar-metode berakar pada prinsip kerjanya. Z-Score bersifat univariat dan global sehingga hanya menandai nilai yang ekstrem pada satu fitur dan tidak menangkap kombinasi nilai yang janggal antar-fitur, misalnya IPK dan TES tinggi dengan PBL rendah. *Isolation Forest* mengisolasi titik melalui partisi acak sehingga cenderung peka terhadap anomali global, sedangkan DBSCAN menggunakan satu radius ε untuk seluruh data sehingga kurang peka apabila kepadatan antar-kelompok mahasiswa berbeda [18]. LOF, sebaliknya, membandingkan kepadatan setiap mahasiswa dengan kepadatan tetangga terdekatnya sendiri [15] sehingga dapat mengungkap *outlier* lokal yang berada dekat dengan gugus data yang padat.

Hasil pada Tabel 5 sejalan dengan uraian tersebut. Z-Score hanya mendeteksi 3 mahasiswa (No. 50, 56, dan 57), seluruhnya berada di antara tiga skor LOF tertinggi dan memiliki nilai minimum pada sedikitnya satu fitur, dengan demikian 6 dari 9 *outlier* LOF (No. 20, 26, 28, 31, 37, dan 47) tidak terdeteksi oleh Z-Score karena nilai setiap fiturnya masih wajar secara univariat sedangkan kombinasinya janggal. *Isolation Forest* menandai 9 mahasiswa dan beririsan pada 5 mahasiswa dengan LOF (No. 20, 28, 50, 56, dan 57, irisan ini bervariasi antara 4 dan 6 pada 100 nilai random_state yang berbeda sehingga kesesuaiannya dengan LOF bersifat sedang. DBSCAN menandai 19 mahasiswa (23,17%) sebagai *noise* dan mencakup 8 dari 9 *outlier* LOF (kecuali No. 28), tetapi jumlah *noise* ini lebih dari dua kali lipat dan berubah antara 14 dan 24 ketika eps divariasikan 0,8–1,5 kali nilai siku, sehingga DBSCAN dengan satu nilai eps kurang selektif pada data yang kepadatannya tidak seragam. Pada data ini tidak ada *outlier* LOF yang luput dari seluruh metode lain secara bersamaan, keunggulan LOF terletak pada selektivitasnya, yaitu menghasilkan daftar ringkas 9 mahasiswa yang memuat seluruh kasus ekstrem hasil Z-Score sekaligus kasus lokal yang tidak terdeteksi Z-Score (6 mahasiswa) maupun *Isolation Forest* (4 mahasiswa: No. 26, 31, 37, dan 47), tanpa menandai hampir seperempat sampel seperti DBSCAN. Karena tidak tersedia label *outlier* sebagai acuan (*ground truth*), perbandingan ini menunjukkan tingkat kesesuaian antar-metode (irisan hasil), bukan nilai akurasi.

**g. Pembahasan terhadap Penelitian Terdahulu**

Proporsi *outlier* pada penelitian ini adalah 10,98% (9 dari 82 mahasiswa), sekitar 1,9 kali proporsi 5,9% (118 dari 2.000 mahasiswa) pada penelitian Jailani dan Erna [6] yang juga memakai LOF. Perbandingan ini bersifat indikatif karena [6] memakai sampel yang jauh lebih besar dan hanya dua dimensi (rasio jam belajar–aktivitas sosial dan IPK), sedangkan penelitian ini memakai tiga dimensi kompetensi dengan sampel yang lebih kecil, dan proporsi *outlier* pada penelitian ini sebagian ditentukan oleh ambang batas persentil ke-90. Kesamaan keduanya terletak pada temuan bahwa *outlier* umumnya berupa ketimpangan antar-variabel, yaitu durasi belajar tinggi namun IPK rendah pada [6] serta IPK dan TES tinggi namun PBL rendah (atau sebaliknya) pada penelitian ini, yang sama-sama bernilai sebagai dasar pendampingan mahasiswa.

Dibandingkan dengan penelitian [7] yang memakai DBSCAN dengan metode *Elbow* pada nilai akademik mahasiswa informatika, penelitian ini menghasilkan skor LOF yang kontinu sehingga tingkat keanehan setiap mahasiswa dapat diurutkan (Tabel 4), tidak hanya dibagi menjadi anggota gugus dan *noise*. Jika [7] menentukan parameter melalui metode *Elbow*, penelitian ini menentukan *K* melalui analisis sensitivitas delapan nilai *K*, sejalan dengan [10] yang menekankan pentingnya optimasi parameter tetangga. Studi komparasi LOF dan *Isolation Forest* pada data kinerja dosen [5] juga menegaskan manfaat membandingkan beberapa metode, penelitian ini memperluasnya ke profil kompetensi mahasiswa dengan menambahkan Z-Score dan DBSCAN. Temuan ini juga sejalan dengan [8] yang memandang identifikasi pola akademik sebagai instrumen penting untuk memitigasi risiko *dropout* mahasiswa dengan IPK dan TES yang sama-sama rendah (misalnya No. 50) dapat menjadi kandidat intervensi dini.

Implikasi praktis dari ketiga pola pada Tabel 4 adalah sebagai berikut. Mahasiswa dengan PBL rendah tetapi TES dan IPK tinggi dapat diarahkan pada pendampingan berbasis proyek, mahasiswa dengan TES rendah tetapi PBL tinggi dapat diberi penguatan dasar pemrograman, dan mahasiswa dengan IPK rendah tetapi TES tinggi dapat ditelusuri oleh dosen wali untuk mengetahui faktor yang memengaruhi capaian akademik formalnya. Dengan demikian, daftar *outlier* berfungsi sebagai sistem

peringatan dini dan sebagai dasar identifikasi mahasiswa berpotensi tinggi, bukan sebagai label negatif terhadap mahasiswa.

Penelitian ini memiliki beberapa keterbatasan. Pertama, sampel ($N$ = 82) berasal dari satu institusi sehingga generalisasi hasil masih terbatas. Kedua, ambang batas ditetapkan pada persentil ke-90 sehingga jumlah *outlier* (sekitar 10%) merupakan keputusan desain, bukan temuan, analisis sensitivitas terhadap persentil lain, misalnya ke-85 dan ke-95, disarankan. Ketiga, tidak adanya label *outlier* sebagai acuan membuat evaluasi bertumpu pada kesesuaian antar-metode dan interpretasi pola, dan hasil *Isolation Forest* serta DBSCAN bergantung pada parameternya. Keempat, mahasiswa marginal perlu diverifikasi. Penelitian selanjutnya dapat menambah jumlah sampel dan angkatan, melibatkan validasi pakar (dosen wali) sebagai acuan, serta menambahkan variabel lain seperti kehadiran dan aktivitas pembelajaran.

## 5 Kesimpulan

(Penelitian ini menyimpulkan bahwa transformasi digital di sektor pendidikan telah menghasilkan data akademik yang kompleks, sehingga memerlukan metode evaluasi yang lebih mendalam daripada sekadar memantau Indeks Prestasi Kumulatif (IPK). Implementasi algoritma *Local Outlier Factor* (LOF) terbukti efektif dalam memetakan profil akademik mahasiswa dengan mengintegrasikan tiga variabel utama, yaitu IPK, Tes Programming, dan nilai *Project*. Melalui pendekatan berbasis kepadatan (*density-based*), metode ini mampu mengidentifikasi mahasiswa yang memiliki pola anomali atau "pencilan lokal" yang sering kali tidak terdeteksi oleh metode statistik konvensional. Hasil pengujian menunjukkan bahwa keakuratan deteksi anomali sangat bergantung pada pengaturan parameter tetangga (K). Penggunaan nilai K yang optimal memungkinkan algoritma untuk tetap stabil dalam membedakan antara mahasiswa dengan karakteristik unik yang signifikan secara statistik dan derau (*noise*) data biasa. Berdasarkan analisis visualisasi, terdeteksi adanya ketimpangan kompetensi pada sejumlah mahasiswa, seperti tingginya kemampuan teoretis yang tidak dibarengi dengan hasil karya nyata (*project*), atau sebaliknya. Dari 82 mahasiswa yang dianalisis, nilai K optimal sebesar 20 dengan ambang batas persentil ke-90 mendeteksi 9 mahasiswa *outlier* (10,98%), enam di antaranya tidak terdeteksi oleh Z-Score, *Isolation Forest* sepakat pada 5 dan DBSCAN pada 8 dari 9 *outlier* tersebut, namun DBSCAN menandai 19 mahasiswa sehingga kurang selektif. Secara praktis, hasil ini dapat dipakai program studi sebagai dasar sistem peringatan dini bagi mahasiswa berisiko, penentuan kandidat pembinaan atau beasiswa, serta pendampingan yang disesuaikan dengan pola ketimpangan kompetensi. Penelitian selanjutnya disarankan memperbesar sampel, melibatkan beberapa angkatan atau program studi, dan menambahkan validasi pakar sebagai acuan.

## Referensi

[1] Vrantika Br Samosir, Agung Mulyo Widodo, Nizirwan Anwar, Binastya Anggara Sekti, and Nixon Erzed, “Identifikasi *Outlier* menggunakan Teknik Data Mining *Clustering* untuk Analisis Data *Tracer Study* pada Fakultas Ilmu Komputer Universitas Esa Unggul,” *IKRA-ITH Informatika : Jurnal Komputer dan Informatika*, Vol. 8, No. 1, pp. 162–174, Mar. 2024, DOI: 10.37817/ikraith-informatika.v8i1.3211.

[2] Jovita and S. Susanto, “Perbandingan Metode *IQR vs Isolation Forest* dalam Deteksi *Outlier* pada Data Klaim Asuransi di PT Askrindo Cabang Semarang *Comparison of IQR and Isolation Forest Methods for Outlier Detection in Insurance Claim Data at PT Askrindo Semarang Branch*,” *Jurnal Transformatika*, Vol. 23, pp. 268–277, Jan. 2026, DOI: 10.26623/transformatika.v23i2.13726.

[3] M. Khoiron, “Pemanfaatan *K-Means Clustering* dalam Melihat *Innovation Mindset* Mahasiswa STIMIK ESQ menggunakan Model *Berkeley Innovation Index*,” 2017. DOI: 10.13140/RG.2.2.26337.81763.

[4] P. A. Ariawan, “Optimasi Pengelompokan Data pada Metode *K-Means* dengan Analisis *Outlier*,” *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 5, No. 2, pp. 88–95, Sep. 2019, DOI: 10.25077/teknosi.v5i2.2019.88-95.

[5] W. Yustanti, "Studi Komparasi *Local Outlier Factor (LOF)* dan *Isolation Forest (IF)* pada Analisis Anomali Kinerja Dosen," *Journal of Informatics and Computer Science*, Vol. 06, 2024.

[6] A. K. Jailani and A. Erna, "Deteksi Anomali pada Rasio Jam Belajar dan Aktivitas Sosial terhadap Performa Akademis Mahasiswa menggunakan Metode *Local Outlier Factor (LOF)*," *SISITI : Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, Vol. 14, No. 1, pp. 79–88, Mar. 2025, DOI: 10.36774/sisiti.v14i1.1678.

[7] estannisa asfarina fadlilah, "Identifikasi Anomali Data Akademik menggunakan *Dbscan Outlier Detection*," *Prosiding Seminar Nasional Sains dan Teknologi*, Vol. 12, No. 1, pp. 336–342, Nov. 2022, DOI: 10.36499/psnst.v12i1.7012.

[8] P. Satya Saputra, "Analisis Prediktif *Dropout* Mahasiswa berdasarkan Kinerja Akademik Semester Awal menggunakan *Machine Learning*," *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, Vol. 07, 2026.

[9] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa menggunakan Metode *K-Means Clustering* (Studi Kasus : Matakuliah Pemrograman Dasar)," *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 9, No. 2, 2023, DOI: 10.25077/teknosi.v9i2.2023.122-131.

[10] Y. Zhou, H. Xia, D. Yu, J. Cheng, and J. Li, "*Outlier Detection Method based on High-Density Iteration*," *Inf. SCI. (N. Y).*, Vol. 662, p. 120286, 2024, DOI: https://doi.org/10.1016/j.ins.2024.120286.

[11] R. Tatikonda, R. Thatikonda, S. M. Potluri, R. Thota, V. S. Kalluri, and A. Bhuvanesh, "Data-*Driven Store Design: Floor Visualization for Informed Decision Making*," in *2025 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, 2025, pp. 1–6. DOI: 10.1109/APSIT63993.2025.11085151.

[12] N. Sa'adah, T. Andiansyah, T. Hidayat, C. Amni, and S. Sunyanti, "Analisis Distribusi dan Tren Nilai Mahasiswa dengan Pendekatan Statistik Deskriptif dan Visualisasi Data," *sudo Jurnal Teknik Informatika*, Vol. 4, No. 3, pp. 203–211, Sep. 2025, DOI: 10.56211/sudo.v4i3.1233.

[13] A. Pily, Susanti, U. Rio, and Tashid, "Komparasi *K-Means Clustering* dengan *Euclidean* dan *Cosine Similarity* untuk Segmentasi dan Rekomendasi Produk pada Data *E-Commerce*," *The Indonesian Journal of Computer Science*, Vol. 14, Apr. 2025, DOI: 10.33022/ijcs.v14i2.4713.

[14] Nathanael Kenneth Lay, Desi Arisandi, and Henoch Juli Christanto, "Analisa Perbandingan Algoritma *K-Means* dan *DBSCAN* untuk Klastering Wilayah Rawan Bencana," in *Building of Informatics, Technology and Science (BITS)*, Dec. 2025, p. 18751886.

[15] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "*LOF: Identifying Density-based Local Outliers*," *ACM SIGMOD Record*, Vol. 29, No. 2, pp. 93–104, Jun. 2000, DOI: 10.1145/335191.335388.

[16] F. Pedregosa *et al.*, "*Scikit-Learn: Machine Learning in Python*," *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.

[17] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "*Isolation Forest*," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422, DOI: 10.1109/ICDM.2008.17.

[18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "*A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, pp. 226–231.