

Deteksi Anomali pada Rasio Jam Belajar dan Aktivitas Sosial terhadap Performa Akademis Mahasiswa menggunakan Metode *Local Outlier Factor* (LOF)

Abdul Kadir Jailani*¹, Angdy Erna²

¹Teknik Informatika, Universitas Dipa, Makassar

²Sistem Informasi, Universitas Dipa, Makassar

e-mail: *¹akjailani@undipa.ac.id, ²angdy@undipa.ac.id

Abstrak

Penelitian ini bertujuan mendeteksi anomali atau *outlier* pada data rasio jam belajar dan aktivitas sosial mahasiswa (*Rasio_Study_Social*) dihubungkan dengan performa akademis atau Indeks Prestasi Kumulatif (IPK) mahasiswa menggunakan metode *Local Outlier Factor* (LOF). *Dataset* yang digunakan terdiri dari 2000 data mahasiswa dengan atribut utama meliputi jam belajar per hari, jam aktivitas sosial per hari, dan IPK. Hasil analisis menunjukkan bahwa LOF berhasil mengidentifikasi 118 anomali (5,9% dari total data). Data Anomali tersebut terdiri dari mahasiswa dengan *Rasio_Study_Social* > 1 sebanyak 109 mahasiswa dan mahasiswa dengan *Rasio_Study_Social* < 1 sebanyak 9 mahasiswa, serta 75 mahasiswa memiliki IPK $\geq 3,00$ dan 43 mahasiswa memiliki IPK < 3,00. Umumnya Data anomali ini menunjukkan pola aktivitas yang tidak lazim, seperti jam belajar tinggi dengan IPK rendah atau sebaliknya. Diharapkan temuan ini dapat memberikan wawasan untuk mendukung pengambilan keputusan bagi institusi pendidikan.

Kata kunci : anomali, indeks prestasi kumulatif, *local outlier factor*, jam belajar, aktivitas sosial.

Abstract

*This study aims to detect anomalies or outliers in the ratio of study hours to social activity hours (*Rasio_Study_Social*) in relation to academic performance or Grade Point Average (GPA) of students using the *Local Outlier Factor* (LOF) method. The dataset comprises 2000 student records with key attributes including study hours per day, social activity hours per day, and CGPA. The analysis results show that the LOF method successfully identified 118 anomalies (5.9% of the total data). These anomalies consisted of 109 students with *Rasio_Study_Social* > 1 and 9 students with *Rasio_Study_Social* < 1, as well as 75 students having a GPA ≥ 3.00 , while 43 students having a GPA < 3.00. Generally, these anomalous data indicated unusual activity patterns, such as high study hours with low GPA or vice versa. It is hoped that these findings can provide insights to support decision-making for educational institutions.*

Keywords: anomaly, grade point average, *local outlier factor*, study hours, social activity.

1. Pendahuluan

Indeks Prestasi Kumulatif (IPK) secara implisit termasuk dalam atribut akademik yang sering digunakan dalam penelitian untuk memprediksi keberhasilan mahasiswa di perguruan tinggi [1].

IPK adalah indikator utama yang digunakan oleh perguruan tinggi untuk menilai performa akademik mahasiswa. IPK sering kali dijadikan acuan dalam evaluasi kelayakan beasiswa, seleksi pekerjaan, hingga pengakuan akademik.

Faktor-faktor yang memengaruhi IPK sangat beragam, mulai dari waktu yang dihabiskan untuk belajar, aktivitas sosial, kualitas tidur, hingga tingkat stres. Jam belajar, sebagai salah satu variabel yang paling umum diukur, diyakini memiliki korelasi positif dengan IPK. Namun, dalam beberapa kasus, mahasiswa dengan waktu belajar yang panjang tidak selalu memiliki IPK tinggi, yang menunjukkan adanya peran faktor lain, termasuk aktivitas sosial dan kesehatan mental.

Selain itu, mahasiswa sering kali harus menyeimbangkan aktivitas akademik dengan kegiatan sosial dan ekstrakurikuler. Jam sosial, yang mencakup interaksi dengan teman, partisipasi dalam organisasi, atau rekreasi, dapat memengaruhi kemampuan mahasiswa dalam mengelola waktu dan stres.

Beberapa penelitian menunjukkan bahwa interaksi sosial yang baik dapat meningkatkan performa akademik dengan menciptakan keseimbangan yang sehat. Menurut [2] menyoroti bahwa jaringan sosial, seperti hubungan pertemanan dan bantuan studi, berperan dalam menyebarkan pengaruh akademik. Temuan ini menunjukkan bahwa performa akademik mahasiswa cenderung menyesuaikan dengan teman-teman yang memiliki prestasi serupa, ini menandakan adanya difusi pengaruh melalui hubungan sosial dengan IPK, tetapi belum menyoroti secara implisit kecenderungan anomali antara hubungan sosial dengan performa akademik atau IPK. Misalnya, jika hubungan sosial menjadi terlalu dominan atau terlalu kecil akan mempengaruhi fokus belajar atau performa akademik. Sedangkan menurut [3] menemukan bahwa hubungan dengan teman sebaya secara signifikan meningkatkan motivasi belajar dan keterlibatan dalam pembelajaran, pada akhirnya memengaruhi pencapaian akademik. Namun, jika aktivitas sosial mengurangi waktu belajar, maka dampaknya dapat menjadi negatif terhadap pencapaian akademik, termasuk IPK.

Berdasarkan penelitian-penelitian di atas, maka akan dihubungkan antara rasio jam belajar yang paling sering diukur dengan aktivitas sosial dalam performa belajar yang ditandai dengan pencapaian IPK. Fokus penelitian ini adalah menemukan anomali atau *outlier* yakni data yang secara signifikan berbeda dari pola umum dalam *dataset*.

Dataset yang digunakan bersumber dari kaggle [4] yang terdiri dari 2000 *record* dan 8 atribut, tetapi peneliti hanya berfokus ke atribut jam belajar per hari, aktivitas sosial per hari, dan IPK. Dalam analisis *dataset* yang kompleks seperti ini, kemungkinan anomali sering kali muncul. *Outlier* adalah data yang secara signifikan berbeda dari pola umum yang ditemukan dalam *dataset*. Kehadiran *outlier* dapat memengaruhi hasil analisis statistik dan prediksi model. Oleh karena itu, deteksi dan analisis *outlier* menjadi penting, terutama dalam konteks pendidikan. Dalam penelitian ini, metode Local *Outlier* Factor (LOF) digunakan untuk mendeteksi mahasiswa dengan pola aktivitas belajar dan sosial yang tidak lazim. LOF adalah metode berbasis densitas yang membandingkan kepadatan lokal suatu data dengan tetangganya. Metode ini memungkinkan deteksi *outlier* dengan cara yang lebih sensitif dan adaptif dibandingkan metode tradisional seperti Z-score [5], [6], [7].

Penelitian ini diharapkan memberikan kontribusi untuk memahami pola unik yang dimiliki oleh mahasiswa dengan perilaku belajar dan sosial yang berbeda. Dengan memahami *outlier*, institusi pendidikan dapat mengembangkan program dukungan akademik dan sosial yang lebih personal untuk memenuhi kebutuhan mahasiswa secara lebih efektif.

2. Metode Penelitian

2.1 *Dataset*

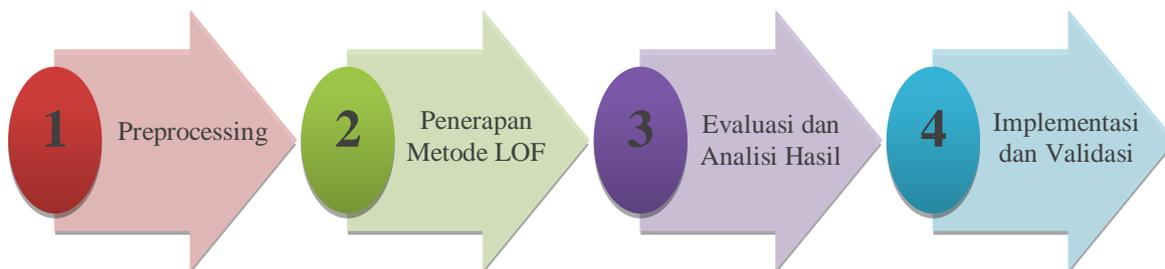
Dataset penelitian terdiri dari 2000 *record* data mahasiswa dan terdiri dari 8 variabel, yaitu:

- a. *Student_ID*: Identitas mahasiswa atau Nomor Induk Mahasiswa (NIM).
- b. *Study_Hours_Per_Day*: Jam belajar per hari.
- c. *Extracurricular_Hours_Per_Day*: Jam aktivitas ekstrakurikuler per hari.
- d. *Sleep_Hours_Per_Day*: Jam tidur per hari.
- e. *Social_Hours_Per_Day*: Jam aktivitas sosial per hari.
- f. *Physical_Activity_Hours_Per_Day*: Jam aktivitas fisik per hari.
- g. *GPA*: Indeks Prestasi Kumulatif (IPK).
- h. *Stress_Level*: Kategori tingkat stres.

Penelitian ini bersifat kuantitatif dengan pendekatan eksplorasi untuk mendeteksi *outlier* pada rasio waktu belajar harian (*Study_Hours_Per_Day*) dan aktivitas sosial (*Social_Hours_Per_Day*) terhadap Indeks Prestasi Kumulatif (*GPA*) mahasiswa. Data diperoleh dari hasil observasi dengan berbagai variabel aktivitas mahasiswa.

2.2 Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian mendeteksi *outlier* dengan menerapkan metode Local *Outlier* Factor.



Gambar 1. Tahapan Penelitian

Deskripsi tahapan penelitian adalah sebagai berikut:

1. Preprocessing (Pengolahan Awal Data): Memuat *dataset* dan memastikan semua atribut yang akan digunakan, membersihkan data, melakukan transformasi rasio jam belajar dengan aktivitas sosial dan melakukan normalisasi data.
2. Penerapan Metode Local *Outlier* Factor (LOF): Menerapkan metode LOF untuk mendeteksi *outlier* atau anomali berdasarkan rasio waktu belajar dan aktivitas sosial terhadap IPK, menentukan *Number of Neighbors* (k) untuk menemukan hasil terbaik, dan mengidentifikasi nilai *outlier* LOF $>1,5$.
3. Evaluasi dan Analisis Hasil: Membuat visualisasi hasil *outlier*, menggunakan teknik evaluasi *Receiver Operating Characteristic* (ROC) dan *Correlation Matrix* untuk memahami hubungan statistiknya.
 - a. Visualisasi Hasil:
Gunakan operator *Scatter* atau *Plot View* di RapidMiner untuk membuat scatter plot antara rasio waktu dan GPA. Berikan label untuk menandai *outlier*.
Tambahkan visualisasi histogram LOF untuk memetakan distribusi nilai LOF dan menentukan ambang batas deteksi.
 - b. Rekomendasi Teknik Evaluasi:
Gunakan *Receiver Operating Characteristic* (ROC) curve untuk mengevaluasi kinerja deteksi LOF jika tersedia label ground truth.
Lakukan analisis korelasi antara rasio dan IPK menggunakan operator *Correlation Matrix* untuk memahami hubungan statistik.
 - c. Interpretasi Data:
Tinjau pola pada *outlier*: Apakah memiliki rasio sangat tinggi/rendah tetapi tetap menghasilkan IPK yang ekstrem?
Lakukan analisis kelompok (clustering) tambahan pada data tanpa *outlier* menggunakan metode seperti *k-Means* untuk mengidentifikasi pola kelompok mahasiswa reguler.
4. Implementasi dan Validasi:
 - a. Eksperimen dilakukan dengan berbagai nilai parameter LOF untuk memastikan robustnya hasil deteksi.
 - b. Hasil deteksi dievaluasi melalui diskusi dan diamati bersama dosen di bidang pendidikan untuk interpretasi lebih lanjut.
 - c. Alat dan Software adalah RapidMiner Studio, digunakan untuk semua proses mulai dari preprocessing, deteksi *outlier* dengan LOF, hingga visualisasi hasil dan analisis korelasi.

2.3 Konsep Anomali

Anomali atau *Outlier* adalah data yang menyimpang secara signifikan dari pola yang umum dalam *dataset*. anomali bisa terjadi karena berbagai alasan, seperti kesalahan pengukuran, kerusakan sistem, atau kejadian yang jarang terjadi [8].

Anomali dalam analisis data pendidikan telah menjadi topik penting, karena dapat mengungkapkan fenomena unik yang terlewatkan dalam analisis rata-rata. Dalam konteks pendidikan, penelitian sebelumnya menunjukkan bahwa faktor-faktor seperti jam belajar, aktivitas sosial, dan kesehatan fisik berkontribusi terhadap IPK [9].

Berdasarkan [10] faktor-faktor yang mempengaruhi IPK:

- a. Durasi Belajar: Mahasiswa yang mengalokasikan waktu lebih banyak untuk belajar cenderung memiliki IPK lebih tinggi.

- b. Kehadiran di Kelas: Kehadiran yang konsisten berkorelasi positif dengan prestasi akademik.
- c. Interaksi Sosial: Jumlah interaksi percakapan dan pola aktivitas sosial seperti pesta juga memengaruhi hasil akademik.
- d. Mobilitas dan Aktivitas Fisik: Aktivitas harian yang terstruktur atau tidak juga diukur dan dikaitkan dengan performa akademik.
- e. Pola Waktu: Analisis data berbasis deret waktu, seperti jam aktivitas malam hari, memberikan wawasan tambahan.

Faktor sosial penting dipertimbangkan dalam mendukung keberhasilan akademik mahasiswa, serta membandingkan kontribusi antara faktor individu dan jaringan sosial terhadap prestasi mahasiswa [11].

2.4 Metode Local Outlier Factor (LOF)

Menurut [12] LOF merupakan metode yang mempertimbangkan kepadatan lokal dalam mendeteksi *outlier* atau anomali. Algoritma ini menghitung rasio kepadatan lokal data dengan tetangganya untuk menentukan apakah data tersebut merupakan anomali.

Nilai LOF dihitung berdasarkan rasio kepadatan lokal data terhadap kepadatan tetangga terdekatnya. Nilai LOF yang lebih besar dari 1 menunjukkan anomali, sementara nilai mendekati 1 berarti data tersebut konsisten dengan lingkungannya [13]. LOF sering digunakan dalam analisis data multivariabel untuk mendeteksi *outlier* pada *dataset* dengan berbagai atribut, seperti dalam pemantauan jaringan atau analisis perilaku konsumen [14].

Tahap-tahap LOF adalah sebagai berikut:

1. Hitung Jarak Tetangga Terdekat (*k*-distance):
 - a. Untuk setiap titik *P*, tentukan jarak ke tetangga *k*-terdekatnya, yang disebut sebagai *k*-distance(*P*).
 - b. Tetangga *k*-terdekat ($N_k(P)$) mencakup semua titik dalam jarak *k*-distance(*P*).
2. Hitung Reachability Distance:

Untuk setiap pasangan titik *P* dan *O* dalam $N_k(P)$, hitung reachability distance:

$$reachability_dist_k(P, O) = \max(k - distance(O), distance(P, O)) \quad (1)$$

distance(*P*,*O*): jarak langsung antara titik *P* dan *O*.

Jika *P* sangat dekat dengan *O*, reachability distance minimum adalah *k*-distance(*O*).

3. Hitung Local Reachability Density (LRD):

LRD adalah ukuran kepadatan lokal *P* relatif terhadap tetangganya. Dirumuskan sebagai kebalikan dari rata-rata reachability distance, sebagai berikut:

$$LRD_k(P) = \frac{1}{\frac{1}{|N_k(P)|} \sum_{O \in N_k(P)} reachability_dist_k(P, O)} \quad (2)$$

$|N_k P|$: Jumlah tetangga *P* dalam $N_k P$.

Semakin kecil nilai LRD, semakin jarang *P* ditemukan di daerah tersebut.

4. Hitung Local Outlier Factor (LOF):

LOF mengukur rasio antara LRD titik *P* dan rata-rata LRD tetangganya dengan rumus sebagai berikut:

$$LOF_k(P) = \frac{\sum_{O \in N_k(P)} \frac{LRD_k(O)}{LRD_k(P)}}{|N_k(P)|} \quad (3)$$

Jika $LOF_k \approx 1$: Titik *P* memiliki kepadatan lokal yang serupa dengan tetangganya (tidak *outlier*).

Jika $LOF_k > 1$: Titik *P* lebih jarang dibanding dengan tetangganya (potensi *outlier*).

Nilai $LOF_k > 2$: sering dindikasikan kuat bahwa *P* adalah *outlier*.

Menurut [12]interpretasi nilai LOF bersifat fleksibel dan bergantung pada aplikasi spesifik. Nilai ambang atau *threshold_value* seperti 1,5 sering digunakan dalam kasus dimana data memiliki distribusi yang cukup beragam.

3. Hasil dan Pembahasan

3.1 Pengolahan Awal Data (Preprocessing)

Memuat Data: *Dataset* dimuat ke RapidMiner untuk semua atribut yang digunakan, seperti waktu belajar (*Study_Hours_Per_Day*), aktivitas sosial (*Social_Hours_Per_Day*), dan IPK (GPA), terdefinisi dengan benar.

Tabel 1. *Dataset*student_lifestyle [4]

<i>Student_ID</i>	<i>Study_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	<i>GPA</i>
1	6,9	2,8	2,99
2	5,3	4,2	2,75
3	5,1	1,2	2,67
4	6,5	1,7	2,88
5	8,1	2,2	3,51
6	6	0,3	2,85
7	8	5,7	3,08
8	8,4	3	3,2
9	5,2	4	2,82
10	7,7	4,5	2,76
...
1996	6,5	2,1	3,32
1997	6,3	1,5	2,65
1998	6,2	0,8	3,14
1999	8,1	3,5	3,04
2000	9	3,1	3,58
1999	8,1	3,5	3,04
2000	9	3,1	3,58

Pembersihan Data:Periksa keberadaan nilai kosong atau tidak valid pada data menggunakan operator *Replace Missing Values*dengan metode *Average of Neighbors*. Algoritma ini akan mengganti nilai yang hilang dengan rata-rata nilai tetangganya (sebelum dan sesudahnya dalam *dataset*).

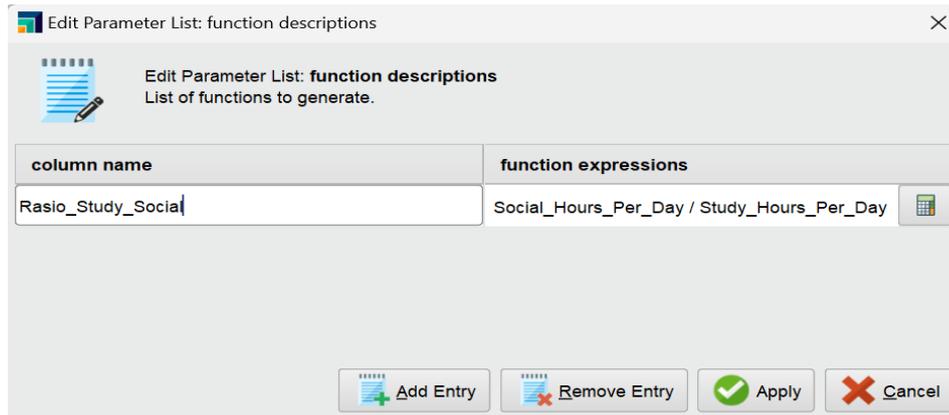
Implementasinya misalnya pada atribut *Study_Hours_Per_Day* yang nilainya *np.nan* (nilai kosong) akan mengubah nilai *missing* tersebut dengan metode *average of neighbors* dengan cara mengambil rata-rata dinilai tetangganya misalnya nilai tetangga adalah 2,0 dan 3,5, maka nilai yang diberikan adalah 2,75 ((2,0+3,5)/2). Begitu seterusnya apabila ditemukan nilai *missing* akan diberi nilai rata-rata tetangga.

a. Transformasi Rasio:

Menggunakan operator *Generate Attributes* untuk membuat atribut baru berupa rasio (*Rasio_Study_Social*) waktu belajar terhadap aktivitas sosial, yaitu:

$$Rasio_Study_Social = \frac{Study_Hours_Per_Day}{Social_Hours_Per_Day} \tag{4}$$

Pada operator ini, dilakukan edit parameter seperti gambar 2 berikut:



Gambar 2. Pembuatan atribut Rasio_Study_Social

Kemudia akan diperoleh hasil dari atribut baru atau Rasio_Study_Social (R) sebagai berikut:

Tabel 2. Hasil atribut baru

<i>Student_ID</i>	<i>Rasio_Study_Social</i>	<i>GPA</i>
1	2,46	2,99
2	1,26	2,75
3	4,25	2,67
4	3,82	2,88
5	3,68	3,51
6	20,00	2,85
7	1,40	3,08
8	2,80	3,2
9	1,30	2,82
10	1,71	2,76
...
1995	15,80	3,08
1996	3,10	3,32
1997	4,20	2,65
1998	7,75	3,14
1999	2,31	3,04
2000	2,90	3,58

Berdasarkan tabel 2. Nilai dari atribut Rasio_Study_Social (R) disimpulkan sebagai berikut:

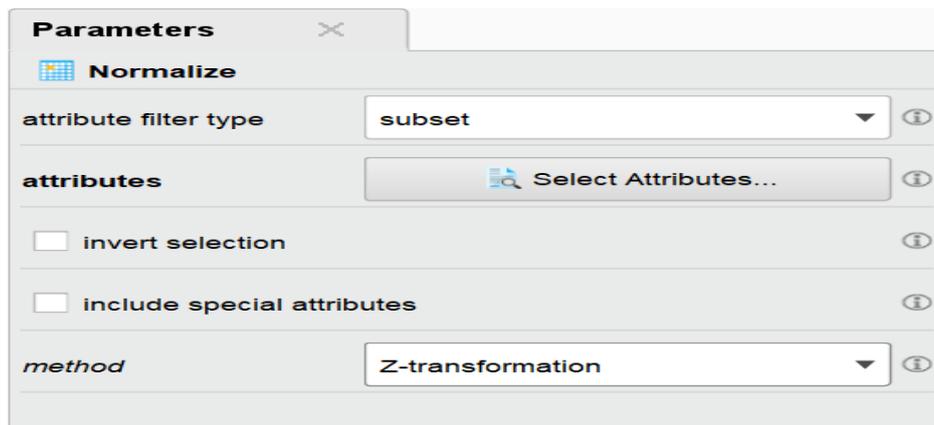
Jika $R > 1$, berarti jumlah jam belajar mahasiswa per hari lebih besar dari jumlah jam beraktivitas sosial perhari.

Jika $R = 1$, berarti jumlah jam belajar mahasiswa per hari sama dengan jumlah jam beraktivitas sosial perhari.

Jika $R < 1$, berarti jumlah jam belajar mahasiswa per hari lebih sedikit dari jumlah jam beraktivitas sosial perhari.

Dari data di atas diperoleh $R > 1$ sebesar 98,8%, $R = 1$ sebesar 0,2%, dan $R < 1$ sebesar 1%. Dalam penelitian ini penulis akan mencari data yang anomali atau *outlier* dari data ini dan menentukan termasuk ke kelompok $R > 1$ atau $R = 1$ atau $R < 1$.

- b. Normalisasi Data: Menggunakan operator *Normalize* atau Normalisasi sangat penting dalam preprocessing data untuk memastikan bahwa atribut dengan skala besar tidak mendominasi atribut dengan skala kecil. Penelitian ini menggunakan metode z-transformasi pada operator *Normalize* seperti gambar 3.



Gambar 3. Metode Z-transformasi

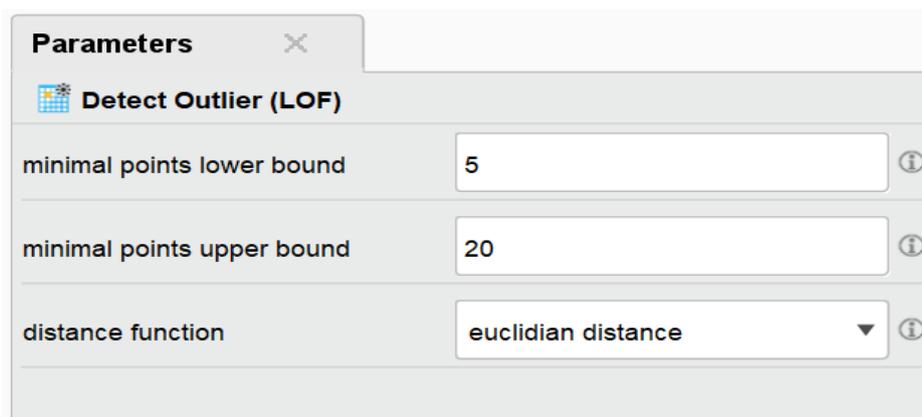
3.2 Penerapan Metode LOF (Local Outlier Factor)

- a. Menggunakan operator *Anomaly Detection with LOF* (gambar 4) untuk mendeteksi *outlier* berdasarkan rasio waktu belajar dan aktivitas sosial terhadap IPK.



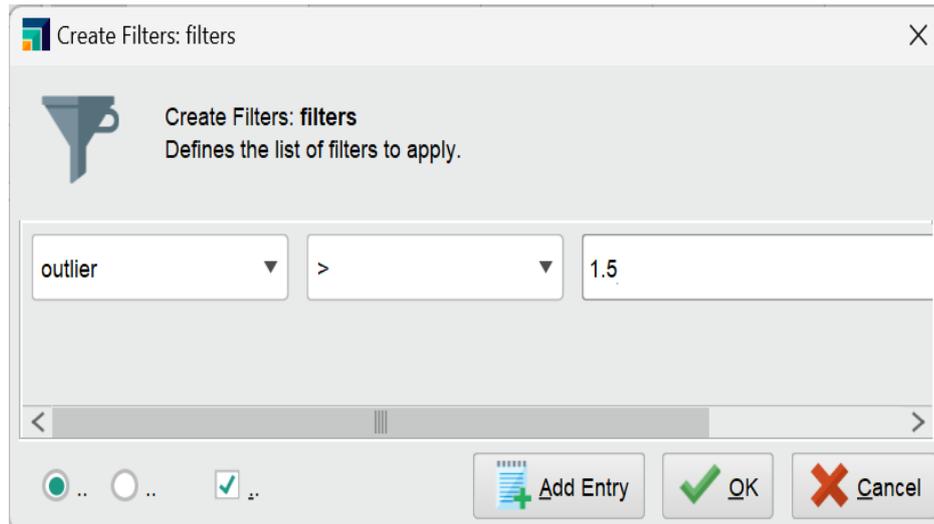
Gambar 4. Operator detect outlier (LOF)

- b. Pengaturan Parameter:
Number of Neighbors (k): Melakukan eksperimen dengan $k = \{5, 10, 20\}$ untuk mengevaluasi hasil terbaik.
 Menentukan metrik jarak (*distance measure*) menggunakan *Euclidean distance*, untuk menghitung kedekatan antar data dapat dilihat pada gambar 5.



Gambar 5. Pengaturan parameter

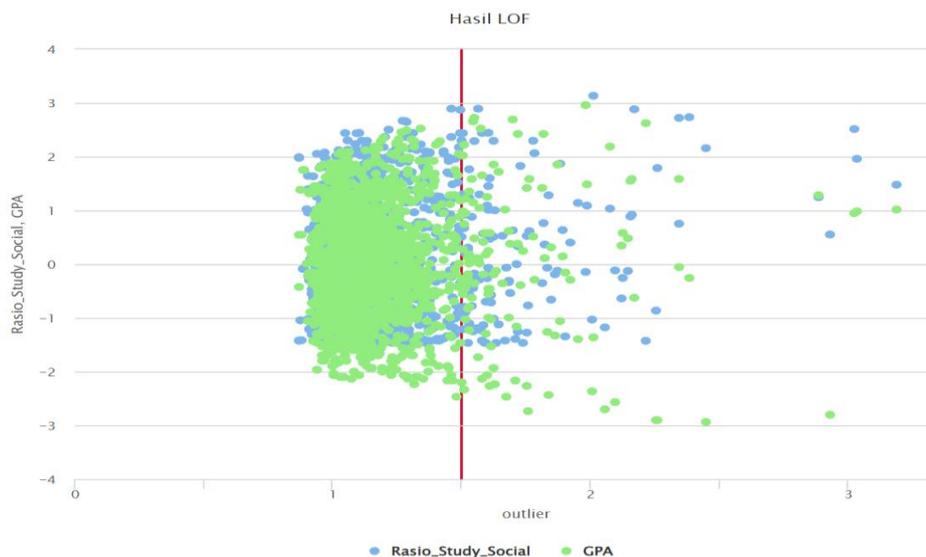
- c. Identifikasi *Outlier* (gambar 6): Mahasiswa dengan nilai LOF tinggi (di atas ambang batas tertentu atau *threshold_value*, yakni > 1.5) dianggap sebagai *outlier*.



Gambar 6. Filter nilai outlier

3.3 Evaluasi dan Analisis Hasil

Visualisasi Hasil dari semua perhitungan LOF dari *dataset* diperoleh *outlier* dengan nilai *threshold_value* di atas 1,5 dapat dilihat pada gambar 7 berikut:

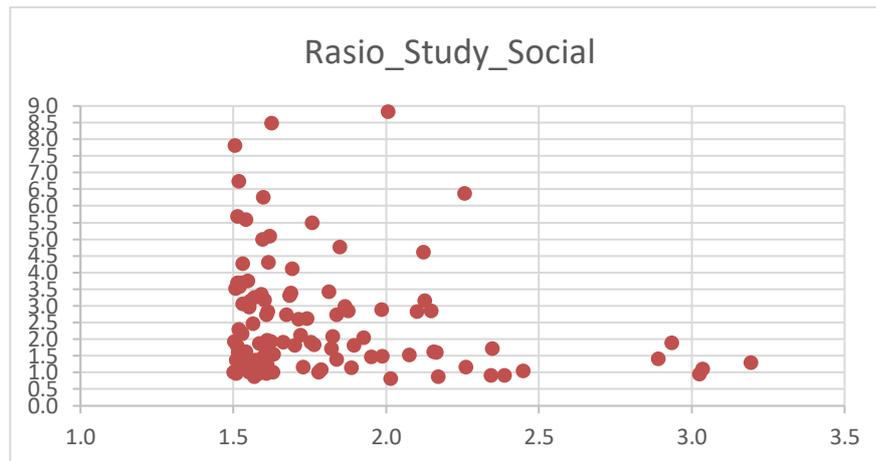


Gambar 7. Hasil semua LOF

Berdasarkan gambar 7, data *outlier* berada di sebelah kanan garis vertikal merah sebanyak 118 data *outlier* dari 2000 record data atau ditemukan data yang anomali sebanyak 5,9% (118 record). Pada gambar tersebut dapat lihat *threshold_value* lebih besar dari 1,5.

Pada gambar 8 merupakan rincian dari 118 data *outlier* berdasarkan R (Rasio_Study_Social) dengan penjelasan sebagai berikut:

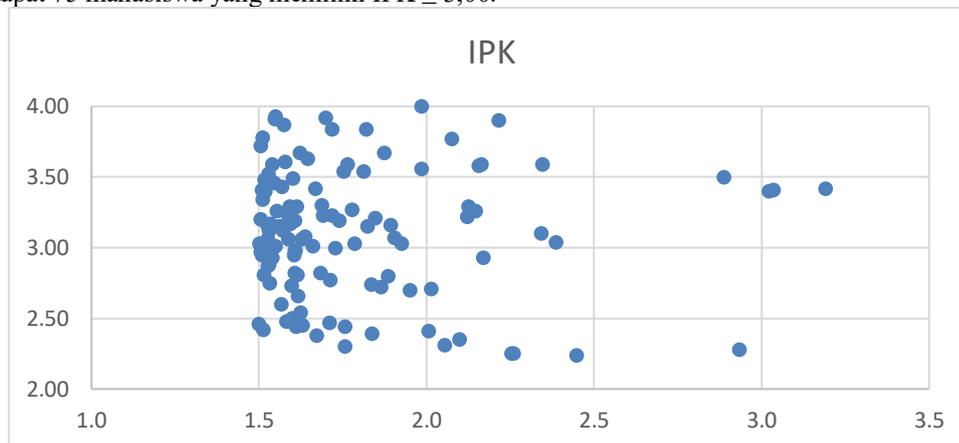
1. Terdapat 9 mahasiswa yang $R < 1$ atau jumlah jam belajar mahasiswa per hari lebih sedikit dari jumlah jam beraktivitas sosial perhari.
2. Terdapat 109 mahasiswa yang $R > 1$ atau jumlah jam belajar mahasiswa per hari lebih banyak dari jumlah jam beraktivitas sosial perhari.



Gambar 8. *Outlier* rasio_study_social

Pada gambar 9 merupakan rincian dari 118 data *outlier* berdasarkan IPK (GPA) dengan penjelasan sebagai berikut:

1. Terdapat 43 mahasiswa yang memiliki IPK $< 3,00$.
2. Terdapat 75 mahasiswa yang memiliki IPK $\geq 3,00$.



Gambar 9. *Outlier* GPA (IPK)

3.4 Implementasi dan Validasi

- a. Eksperimen dilakukan dengan berbagai nilai parameter *LOFminimal points lower bound* sama dengan 5 dan *minimal points upper bounds* sama dengan 20 untuk memastikan robustnya hasil deteksi dengan pendekatan *euclidian distance*.
- b. Hasil deteksi dievaluasi melalui diskusi dan pengamatan bersama peneliti atau dosen bidang pendidikan tinggi untuk interpretasi lebih lanjut.
- c. Alat dan Software yang digunakan dalam penelitian ini sebagai besar menggunakan RapidMiner Studio, penggunaanya untuk semua proses mulai dari preprocessing, deteksi *outlier* dengan LOF, hingga visualisasi hasil dan analisis korelasi.
- d. Temuan data *outlier* sebanyak 118 data memiliki Rasio_Study_Social sebagai besar atau 109 data di atas 1 atau memiliki jumlah jam belajar lebih besar dibanding dengan jumlah jam beraktivitas sosial, tetapi dari data ini ditemukan pula bahwa IPK lebih besar atau sama dengan 3,00 hanya berjumlah 75 mahasiswa. Ini berarti ada selisih maksimal 34 mahasiswa, meskipun jam belajarnya tinggi performa akademik atau IPKnya lebih kecil dari 3,00.
- e. Berdasarkan temuan data *outlier* ini, sekiranya dapat menjadi bahan informasi bagi *stackholder* perguruan tinggi dalam membuat kebijakan untuk melihat sisi jumlah jam belajar dan aktivitas sosial mahasiswa.

4. Kesimpulan

Penelitian ini menemukan data yang anomali atau *outlier* sebanyak 118 (5,9%) dari 2000 record data mahasiswa, dimana data *outlier* tersebut ditemukan dengan menerapkan metode LOF. Data *outlier* ini, memiliki berbagai variasi yakni; ada yang Rasio_Studi_Social lebih besar dari 1 (109 data mahasiswa) dan ada juga yang nilainya lebih kecil dari 1 (9 data mahasiswa), ada mahasiswa yang memiliki IPK lebih besar dari 3,00 (75 data mahasiswa) dan ada juga yang IPK lebih kecil dari 3,00 (43 data mahasiswa), tetapi tidak semua yang Rasio_Studi_Socialnya tinggi atau lebih besar dari 1 memiliki performa akademik atau IPK lebih dari 3,00. Diharapkan temuan ini dapat memberikan wawasan untuk mendukung pengambilan keputusan bagi institusi pendidikan.

Daftar Pustaka

- [1] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, 2020, doi: 10.1186/s41239-020-0177-7.
- [2] S. Dokuka, D. Valeeva, and M. Yudkevich, "How academic achievement spreads: The role of distinct social networks in academic performance diffusion," *PLoS One*, vol. 15, no. 7 July, pp. 1–16, 2020, doi: 10.1371/journal.pone.0236737.
- [3] Y. Shao, S. Kang, Q. Lu, C. Zhang, and R. Li, "How peer relationships affect academic achievement among junior high school students: The chain mediating roles of learning motivation and learning engagement," *BMC Psychol.*, vol. 12, no. 1, pp. 1–12, 2024, doi: 10.1186/s40359-024-01780-z.
- [4] Sumit Kumar, "https://www.kaggle.com/datasets/steve1215rogg/student-lifestyle-dataset/data," Kaggle. Accessed: Dec. 03, 2024. [Online]. Available: <https://www.kaggle.com/datasets/steve1215rogg/student-lifestyle-dataset/data>
- [5] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A review of local *outlier* factor algorithms for *outlier* detection in big data streams," *Big Data Cogn. Comput.*, vol. 5, no. 1, pp. 1–24, 2021, doi: 10.3390/bdcc5010001.
- [6] Z. Li and L. Zhang, "An Ensemble *Outlier* Detection Method Based on Information Entropy-Weighted Subspaces for High-Dimensional Data," *Entropy*, vol. 25, no. 8, 2023, doi: 10.3390/e25081185.
- [7] A. Mudinas *et al.*, "Oxford Scholarship Online India Revisited : Conversations on Continuity and Caste in Modern India," *Mach. Learn. Appl. An Int. J.*, vol. 9, no. 2, pp. 4–23, 2020, [Online]. Available: <http://allman.rhon.itam.mx/~mendoza/Foresight.pdf%0Ahttps://pdfs.semanticscholar.org/e390/c5d56ddcc8e9f6f27264ee7196539d0e7f78.pdf%0Ahttps://doi.org/10.1016/j.ijforecast.2018.07.006%0Ahttp://arxiv.org/abs/1903.05440>
- [8] F. M. Ghamry, G. M. El-Banby, A. S. El-Fishawy, F. E. A. El-Samie, and M. I. Dessouky, "A survey of anomaly detection techniques," *J. Opt.*, vol. 53, no. 2, pp. 756–774, 2024, doi: 10.1007/s12596-023-01147-4.
- [9] P. Aoun, J. E., Smith, R. T., & Clarke, "Predicting Academic Performance: The Role of Study Hours, Social Activities, and Physical Health," *Int. J. Educ. Res.*, vol. 92, pp. 45–56, 2018, doi: <https://doi.org/10.1007/s12596-023-01147-4>.
- [10] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell, "SmartGPA: How smartphones can assess and predict academic performance of college students," *UbiComp 2015 - Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 295–306, 2015, doi: 10.1145/2750858.2804251.
- [11] V. Kassarnig, E. Mones, A. Bjerre-Nielsen, P. Sapiezynski, D. D. Lassen, and S. Lehmann, "Academic performance and behavioral patterns," *EPJ Data Sci.*, vol. 7, no. 1, pp. 1–16, 2018, doi: 10.1140/epjds/s13688-018-0138-8.
- [12] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local *outliers*," *SIGMOD Rec. (ACM Spec. Interes. Gr. Manag. Data)*, vol. 29, no. 2, pp. 93–104, 2000, doi: 10.1145/335191.335388.
- [13] J. Han, M. Kamber, and J. Pei, "Third Edition : Data Mining Concepts and Techniques," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2012, [Online]. Available: <http://library.books24x7.com/toc.aspx?bkid=44712>
- [14] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015, doi: 10.1007/s10618-014-0365-y.