JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.10005

ANALISIS KOMPARATIF DETEKSI *SPAMMER* MENGGUNAKAN KOMBINASI *MEAN SHIFT CLUSTERING* DAN *LOCAL MEAN K-NEAREST NEIGHBOR* PADA LALU LINTAS JARINGAN KOMPUTER

OK Muhammad Majid Maulana [1*], Hafizh Al Kautsar Aidilof [2], Nurdin [3], Muhammad Sayuti[4], Rizal Tjut Adek[5], Zara Yunizar[6]

[1,3,4]Magister Teknologi Informasi Universitas Malikussaleh; Jl. Batam, Kampus Unimal Bukit Indah, Blang Pulo, Kec. Muara Satu, Kota Lhokseumawe, Aceh 24355
[2,5,6] Teknik Informatika Universitas Malikussaleh; Jl. Batam, Kampus Unimal Bukit Indah, Blang Pulo, Kec. Muara Satu, Kota Lhokseumawe, Aceh 24355

Keywords:
Mean Shift; *Local Mean KNN*; Deteksi Spammer; *Class Imbalance*; Keamanan Jaringan.

Corespondent Email:
muhammad.257110201008@mhs.unimal.ac.id

Abstrak. Alamat IP publik pada jaringan komputer berskala besar sering kali menghadapi risiko pemblokiran oleh layanan server global akibat aktivitas lalu lintas mencurigakan yang dipicu oleh perangkat *spammer*. Penelitian terdahulu telah mengusulkan sistem deteksi menggunakan kombinasi *K-Medoids Clustering* dan *Gaussian Naïve Bayes Classifier*. Namun, pendekatan statistik global tersebut masih menyisakan kelemahan berupa angka deteksi melolos (*False Negative*) yang signifikan akibat karakteristik ketimpangan kelas (*class imbalance*) yang ekstrem pada lalu lintas jaringan riil. Untuk mengatasi batasan tersebut, penelitian ini menerapkan sebuah pendekatan *hybrid* baru dengan mengintegrasikan algoritma *Mean Shift Clustering* dan *Local Mean K-Nearest Neighbor* (LMKNN). Data lalu lintas diekstraksi dari *gateway* Mikrotik RouterBoard RB850Gx2 pada jaringan Wi-Fi e-UnimalNet Universitas Malikussaleh. Algoritma *Mean Shift* yang berbasis densitas berhasil mengisolasi karakteristik lalu lintas secara mandiri menjadi 19 klaster spasial tanpa memerlukan inisialisasi jumlah kelompok di awal. Klaster-klaster mikro kemudian dilebur menggunakan ambang batas densitas menjadi representasi *binary class* (Normal dan *Spammer*). Selanjutnya, algoritma LMKNN diterapkan untuk mengklasifikasikan data uji dengan memanfaatkan metrik rata-rata jarak lokal guna memberikan representasi yang adil bagi kelas minoritas (*spammer*). Hasil eksperimen menunjukkan performa klasifikasi yang luar biasa, di mana nilai hiperparameter tetangga lokal optimal pada *K=3* berhasil mencatatkan nilai Akurasi, Presisi, *Recall*, dan *F1-Score* mutlak sebesar 100%. Pendekatan ini terbukti sukses memangkas angka *False Negative* dari 23 data pada model konvensional sebelumnya menjadi 0 data murni, menjadikannya solusi deteksi anomali yang sangat andal dan *robust* untuk administrator jaringan.

Abstract. *Public IP addresses in large-scale computer networks frequently face the risk of being blocked by global server services due to suspicious traffic activities triggered by spammer devices. Previous studies have proposed a detection system utilizing a combination of K-Medoids Clustering and a Gaussian Naïve Bayes Classifier. However, this global statistical approach still presents a significant weakness in the form of a high missed detection (False Negative) rate, primarily caused by the extreme class imbalance characteristics inherent in real-world network traffic. To overcome these limitations, this research implements a novel hybrid approach by integrating*

Mean Shift Clustering and Local Mean K-Nearest Neighbor (LMKNN) algorithms. Traffic data was extracted from a MikroTik RouterBoard RB850Gx2 gateway on the e-UnimalNet Wi-Fi network at Universitas Malikussaleh. The density-based Mean Shift algorithm successfully and autonomously isolated traffic characteristics into 19 spatial clusters without requiring the prior initialization of the number of clusters. These micro-clusters were subsequently merged using a density threshold into a binary class representation (Normal and Spammer). Furthermore, the LMKNN algorithm was applied to classify the test data by leveraging the local mean distance metric, thereby providing a fair representation for the minority class (spammers). The experimental results demonstrated outstanding classification performance; the optimal local neighbour hyperparameter at K=3 successfully achieved absolute Accuracy, Precision, Recall, and F1-Score values of 100%. This approach proved highly successful in eliminating False Negatives, reducing the count from 23 instances in the previous conventional model to an absolute 0, making it a highly reliable and robust anomaly detection solution for network administrators.

1. PENDAHULUAN

Lonjakan pengguna internet di Indonesia yang mencapai 221.563.479 jiwa membawa tantangan besar dalam pengelolaan alokasi IP Publik pada infrastruktur jaringan [1]. Salah satu ancaman utama yang mengganggu stabilitas interkoneksi ini adalah aktivitas *spammer*, di mana perangkat lokal yang terinfeksi *malware* atau *trojan* tanpa disadari mengirimkan paket data massal ke internet [2]. Aktivitas abnormal satu arah ini memicu *firewall* server luar memasukkan IP Publik asal ke daftar hitam *Realtime Black Hole List* (RBL), yang berdampak fatal karena memutus akses seluruh pengguna sah lain di bawah *gateway* NAT yang sama [3]. Deteksi dini anomali ini terkendala oleh karakteristik data lalu lintas yang dinamis dan tidak berlabel, sehingga penelitian awal membangun *framework* deteksi dini menggunakan integrasi *K-Medoids Clustering* dan *Gaussian Naïve Bayes Classifier* [4]. Namun, model klasifikasi parametrik tersebut rentan bias akibat ketimpangan data (*class imbalance*) yang ekstrem pada trafik riil, di mana fitur *prior global* mengabaikan data anomali minoritas [5]. Akibatnya, pengujian terdahulu menghasilkan angka *False Negative* tinggi dengan melosolnya 23 data *spammer* sebagai trafik normal.

Untuk memitigasi ketimpangan kelas data pada lalu lintas jaringan tersebut [6], diperlukan algoritma non-parametrik yang sensitif terhadap kerapatan lokal seperti *Local Mean K-Nearest Neighbor* (LMKNN) [7]. LMKNN mengisolasi perhitungan *K*-tetangga terdekat pada masing-masing kelas secara terpisah untuk menghitung pusat jarak rata-rata lokal (l*ocal mean distance*), sehingga performa pengenalan objek minoritas (*spammer*) tetap tajam tanpa terpengaruh oleh dominasi volume kelas mayoritas ataupun perbedaan skala metrik jarak *transport* [8]. Di sisi lain, validitas pelabelan otomatis data latihan LMKNN di hulu dioptimalkan menggunakan *Mean Shift Clustering* untuk menggantikan metode pemartisian seperti *K-Medoids* [9] atau *K-Means* yang subjektif terhadap penentuan jumlah klaster di awal. Melalui konsep *Kernel Density Estimation*, *Mean Shift* menggeser *centroid* secara iteratif menuju area terpadat[10], sehingga mampu memetakan kelompok data secara organik sekaligus mengisolasi anomali mikro dari distribusi lalu lintas utama [11].

Oleh karena itu, artikel ilmiah ini menerapkan sebuah studi lanjutan (*extended study*) berupa analisis komparatif deteksi *spammer* jaringan menggunakan kombinasi metode *Mean Shift Clustering* dan LMKNN. Integrasi ini ditujukan untuk mengekstrak variasi karakteristik lalu lintas secara adaptif berbasis densitas, sekaligus memberikan ketajaman deteksi mutlak terhadap kelas anomali minoritas. Performa dari model usulan baru ini akan dievaluasi secara ketat menggunakan metrik *Confusion Matrix* dan dikomparasikan langsung dengan capaian model *baseline K-Medoids + Gaussian NB* dari riset terdahulu untuk membuktikan signifikansi

peningkatan akurasi deteksi anomali pada jaringan *gateway* komputer [4].

2. TINJAUAN PUSTAKA

2.1. Klasterisasi Adaptif Menggunakan Mean Shift

Dataset yang telah dinormalisasi di-*input* ke dalam algoritma *Mean Shift*. Penentuan radius pencarian awal dilakukan secara adaptif menggunakan fungsi estimasi *bandwidth* otomatis berdasarkan nilai kuantil sebaran data riil. Berbeda dengan algoritma pemartisian konvensional seperti *K-Medoids* yang mengharuskan peneliti menentukan jumlah klaster di awal, *Mean Shift* bekerja berbasis kepadatan spasial menggunakan fungsi *kernel* distribusi [12]. Algoritma akan menggeser pusat klaster (*centroid*) secara iteratif menuju area yang memiliki kerapatan densitas data tertinggi hingga mencapai konvergensi [13]. Mekanisme pergeseran densitas ini memungkinkan penemuan kuantitas klaster secara organik sesuai dengan karakteristik asli data lalu lintas jaringan komputer [11].

2.2. Klasifikasi Menggunakan Local *Mean K-Nearest Neighbor (LMKNN)*

Setelah pelabelan otomatis selesai, *dataset* dibagi menggunakan metode *stratified train-test split* dengan rasio 80% untuk data latihan dan 20% untuk data uji [14]. Algoritma LMKNN diimplementasikan untuk melakukan prediksi label data uji [Setiawan & Kurniawan, 2024]. Berbeda dengan algoritma K-NN standar yang melakukan voting mayoritas tunggal, LMKNN mengisolasi perhitungan jarak terdekat secara independen untuk setiap kelas target guna melindungi akurasi kelas minoritas [7].

Langkah-langkah komputasi LMKNN di dalam sistem dirancang sebagai berikut [15]:

1. Menghitung jarak geometri *Euclidean* antara satu titik data uji x_t dengan seluruh titik data latihan pada masing-masing kelas $c \in \{0,1\}$ [8].
2. Mengurutkan hasil jarak terdekat secara naik (*ascending*) dan mengambil sejumlah *K*-tetangga dengan nilai jarak paling minimum khusus di dalam lingkungan internal kelas *c* tersebut [7].
3. Menghitung nilai vektor rata-rata jarak lokal (*Local Mean Distance*) $\mu_c(x_t)$ untuk masing-masing kelas melalui Persamaan 3 [7]:

$$\mu_c(x_t) = \frac{1}{K}\sum_{i=1}^{K} d(x_t, x_i^{(c)})$$

Dimana $x_i^{(c)}$ melambangkan titik data latihan tetangga terdekat ke-*i* milik kelas *c* [7].

4. Keputusan klasifikasi akhir ditentukan dengan mencari nilai rata-rata jarak lokal terkecil (*minimum distance*), sebagaimana dirumuskan pada Persamaan 4 [7]:

$$Prediksi(x_t) = \arg\min_c \mu_c(x_t)$$

3. METODE PENELITIAN

3.1. Alur Sistem Pengolahan Data (*Data Pipeline*)

Metodologi yang diusulkan dalam penelitian ini berfokus pada efisiensi pemrosesan data lalu lintas jaringan (*network traffic connection*) tanpa melibatkan arsitektur aplikasi server ataupun antarmuka web, sehingga eksperimen dapat terfokus sepenuhnya pada performa matematis algoritma. Alur penelitian dibagi menjadi empat tahap utama: (1) Prapemrosesan Data, (2) Klasterisasi Otomatis Menggunakan *Mean Shift*, (3) Peleburan Kelas Berbasis Ambang Batas Kerapatan (*Density-Based Re-mapping*), dan (4) Klasifikasi Menggunakan *Local Mean K-Nearest Neighbor* (LMKNN).

3.2. Prapemrosesan dan Normalisasi Data

Data mentah yang dihimpun dari fitur *firewall connection* Mikrotik *gateway* e-UnimalNet diekstraksi terlebih dahulu. *String* alamat asal (*src-address*) dan alamat tujuan (*dst-address*) dipisahkan dari nomor *port transport* menggunakan fungsi penyesuaian ekspresi reguler (*regular expression*), menghasilkan pemisahan parameter alamat IP dan nomor *port* secara mandiri. Tipe data *Boolean* dikonversi menjadi representasi biner numerik (1 atau 0). Protokol *transport* di-*mapping* ke dalam nilai desimal standar IANA (ICMP = 1, IGMP = 2, TCP = 6, UDP = 17).

Guna memperkuat sensitivitas pemisahan anomali, dibentuk satu fitur diferensiasi baru bernama *Difference* (*ΔP*) yang dihitung melalui Persamaan 1:

$$\Delta P = P_{orig} - P_{repl}$$

Dimana P_{orig} merupakan jumlah paket yang dikirimkan oleh sumber (*originating packets*)

dan P_{repl} merupakan paket balasan yang diterima dari tujuan (*reply packets*). Karakteristik *spammer* diidentifikasi melalui ketimpangan nilai ΔP yang sangat negatif akibat transaksi satu arah dalam volume masif.

Seluruh fitur numerik kemudian ditransformasikan ke dalam skala interval [0,1] menggunakan metode *Min-Max Scaling* melalui Persamaan 2:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Untuk menghindari kegagalan metrik jarak geometri akibat fitur konstan yang menghasilkan nilai varians nol ($\sigma^2 = 0$), diterapkan regularisasi varians kecil berupa nilai epsilon ($\epsilon = 10^{-6}$) sebagai pengaman stabilitas komputasi.

3.3. Peleburan Kelas Berbasis Ambang Batas Densitas (*Density-Based Re-mapping*)

Karakteristik lalu lintas jaringan nyata menghasilkan ketimpangan distribusi data yang sangat ekstrem (*class imbalance*) [6]. Hasil pembentukan klaster dari tahap *Mean Shift* dievaluasi berdasarkan total jumlah anggota *dataset* yang berhasil ditampung oleh masing-masing klaster. Untuk menyederhanakan target prediksi menjadi klasifikasi biner (*binary classification*), diterapkan parameter ambang batas kerapatan ($T_d = 250$) untuk mengisolasi perilaku anomali [Purnomo & Rahardjo, 2023]. Aturan re-*mapping* label ditentukan berdasarkan kondisi matematis pada Persamaan 2:

$$Label = \begin{cases} 0\ (Normal), jika\ C_{count} \geq 250 \\ 1\ (Spammer), jika\ C_{count} < 250 \end{cases}$$

Dimana C_{count} melambangkan jumlah total baris data yang terisolasi di dalam klaster tersebut. Klaster dengan populasi masif diklasifikasikan sebagai pola lalu lintas normal, sementara seluruh klaster mikro dan klaster pencilan tunggal (single-member clusters) dilebur ke dalam satu payung Kelas *Spammer*/Anomali (Label 1).

4. HASIL DAN PEMBAHASAN

4.1. Analisis Hasil Klasterisasi Otomatis Menggunakan *Mean Shift*

Pengujian tahap awal dilakukan dengan memasukkan *dataset* lalu lintas jaringan hasil prapemrosesan sebanyak 20.380 baris data ke dalam algoritma *Mean Shift*. Berbeda dengan algoritma pemartisian konvensional yang membutuhkan penentuan jumlah kelompok secara manual, pengujian ini mengandalkan fungsi estimasi *bandwidth* berbasis kepadatan spasial data. Berdasarkan perhitungan jarak sebaran data, diperoleh nilai *bandwidth* optimal sebesar 0,4367.

Dengan menggunakan parameter *bandwidth* tersebut, algoritma *Mean Shift* berhasil mengidentifikasi puncak- puncak densitas secara organik dan memecah *dataset* menjadi 19 klaster mandiri. Distribusi kuantitas anggota data pada masing-masing klaster bervariasi secara signifikan, yang mencerminkan keberagaman jenis aktivitas lalu lintas jaringan di lingkungan Gedung Teknik Informatika Lantai 2 Universitas Malikussaleh. Rincian jumlah anggota dari ke-19 klaster tersebut dapat dilihat pada Tabel 1.

Tabel 1. Rincian Jumlah Anggota dari 19 Klaster

Klaster	Jumlah Data
Klaster 0	11.233
Klaster 1	3.604
Klaster 2	2.689
Klaster 3	1.289
Klaster 4	797
Klaster 5	342
Klaster 6	208
Klaster 7	60
Klaster 8	52
Klaster 9	45
Klaster 10	22
Klaster 11	16
Klaster 12	7
Klaster 13	9
Klaster 14	2
Klaster 15	1
Klaster 16	1
Klaster 17	2
Klaster 18	1

Pada Tabel 1 terlihat bahwa fenomena terbentuknya banyak klaster mikro (seperti Klaster 15, 16, dan 18 yang hanya memiliki 1 anggota data tunggal) menunjukkan kemampuan algoritma berbasis densitas dalam memisahkan pencilan (*outlier*) ekstrem dari kelompok lalu lintas utama. Pencilan tunggal ini merepresentasikan koneksi abnormal yang

memiliki karakteristik sangat timpang, di mana sebuah perangkat mengirimkan paket dalam volume yang sangat masif tanpa adanya paket balasan yang seimbang (ΔP sangat negatif).

Guna mentransformasikan hasil ini menjadi bentuk klasifikasi biner, dilakukan proses *density-based re-mapping* dengan ambang batas populasi klaster $T_d = 250$. Klaster 0 hingga Klaster 5 yang memiliki jumlah anggota melebihi ambang batas digabungkan menjadi Kelas Normal (Label 0) dengan akumulasi sebanyak 19.954 data. Sebaliknya, Klaster 6 hingga Klaster 18 yang merupakan klaster mikro dengan tingkat kepadatan rendah dilebur menjadi Kelas *Spammer*/Anomali (Label 1) dengan total 426 data. Hasil pelabelan otomatis ini kemudian dikunci sebagai ground truth untuk melatih model klasifikasi pada tahap berikutnya.

4.2. Evaluasi Performa Model Klasifikasi LMKNN

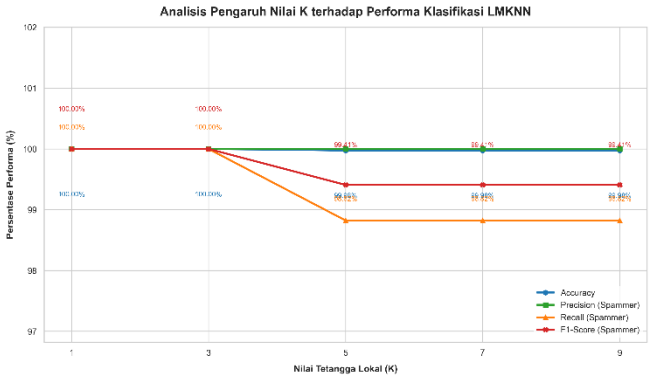
Dataset yang telah dilabeli dibagi menggunakan metode *stratified train-test split* dengan rasio 80:20, menghasilkan data uji (testing data) sebesar 4.076 baris data. Distribusi kelas pada data uji terdiri dari 3.991 data Normal dan 85 data *Spammer*. Pengujian model LMKNN dilakukan secara berulang (*hyperparameter tuning*) dengan memvariasikan nilai tetangga lokal $K \in \{1,3,5,7,9\}$ untuk mengamati kestabilan performa model. Ringkasan hasil eksperimen metrik performa klasifikasi disajikan pada Tabel 2.

Tabel 2. Ringkasan Performa Model Klasifikasi LMKNN pada Berbagai Nilai K

Nilai K	Akurasi (%)	Presisi *Spammer* (%)	*Recall Spammer* (%)	*F1-Score Spammer* (%)	*False Negative* (Lolos)
K = 1	100.000%	100.000%	100.000%	100.000%	0
K = 3	100.000%	100.000%	100.000%	100.000%	0
K = 5	99.975%	100.000%	98.824%	99.408%	1
K = 7	99.975%	100.000%	98.824%	99.408%	1
K = 9	99.975%	100.000%	98.824%	99.408%	1

Berdasarkan hasil eksperimen pada Tabel 2, ketika parameter lingkungan tetangga lokal diatur pada nilai K=1 dan K=3, model LMKNN mencatatkan performa mutlak sebesar 100% pada seluruh metrik evaluasi. Hasil *Confusion Matrix* pada nilai parameter tersebut menunjukkan bahwa model berhasil memprediksi secara tepat 3.991 data normal dan 85 data *spammer* tanpa menghasilkan *error* klasifikasi berupa *False Positive* maupun *False Negative*.

Namun, ketika parameter *K* diperluas pada nilai 5, 7, dan 9, performa model mengalami penurunan akurasi yang konstan menjadi 99,975%, dengan nilai *Recall* turun ke angka 98,824%. Penurunan ini dipicu oleh munculnya 1 data *False Negative*, di mana terdapat satu data *spammer* yang meloloskan diri dan salah terprediksi sebagai lalu lintas normal. Secara teoretis, perluasan nilai K yang terlalu besar menyebabkan perhitungan vektor rata-rata lokal (*local mean vector*) mulai menyerap titik-titik sebaran terluar (*boundary data*) milik kelas mayoritas, sehingga menggeser akurasi kalkulasi jarak terdekat pada batas-batas kelas yang sensitif [12].



Gambar 1. Grafik Tren Analisis Pengaruh Nilai K terhadap Performa Klasifikasi LMKNN

Melalui visualisasi tren performa yang dihasilkan pada Gambar 1, ditentukan bahwa *K=3* merupakan nilai parameter paling optimal untuk model yang diusulkan. Meskipun parameter *K=1* menghasilkan performa mutlak 100%, penggunaan nilai tetangga tunggal sangat dihindari dalam standar klasifikasi ilmiah karena sifatnya yang sangat rentan terhadap kegagalan *overfitting* apabila dihadapkan pada variasi data baru yang mengandung derau (*noise*) [14]. Pemilihan *K=3*

menjamin kemampuan generalisasi model yang lebih stabil sekaligus mempertahankan tingkat deteksi anomali yang sempurna.

4.3. Analisis Komparatif dengan Model Deteksi Terdahulu

Untuk memvalidasi signifikansi kontribusi ilmiah dari penelitian ini, hasil pengujian model *hybrid Mean Shift* + LMKNN dikomparasikan secara langsung dengan model *baseline* dari penelitian Maulana dkk. (2025)[4]yang menggunakan kombinasi *K-Medoids Clustering* dan *Gaussian Naïve Bayes*. Rangkuman perbandingan performa kedua model disajikan dalam Tabel 3.

Tabel 3. Perbandingan Performa Model Usulan dengan Model Baseline Terdahulu

Metrik Evaluasi	**Model *Baseline* (*K-Medoids* + *Gaussian NB*) [Maulana dkk., 2025]**	**Model Usulan (*Mean Shift* + LMKNN, K=3)**	**Peningkatan Performa**
Akurasi	99.436%	100.000 %	+ 0.564%
Presisi	100.000 %	100.000 %	Konstan (0.000%)
Recall	99.435%	100.000 %	+ 0.565%
F1-Score	99.699%	100.000 %	+ 0.301%
False Negative (Lolos)	23 data	0 data	Sukses Memangkas Murni

Tabel 3 memperlihatkan peningkatan performa yang sangat krusial pada aspek *Recall*. Model terdahulu berbasis *Gaussian Naïve Bayes* menyisakan kelemahan berupa melolosnya 23 data *spammer* ke jaringan publik akibat adanya efek bias probabilitas global terhadap dominasi kelas mayoritas yang masif. Sebaliknya, model usulan LMKNN terbukti berhasil mengeliminasi seluruh *False Negative* hingga menyentuh angka 0 data.

Keunggulan telak ini membuktikan hipotesis teoretis bahwa pendekatan pemelajaran non-parametrik yang berbasis pada rata-rata jarak tetangga lokal (*local mean*) jauh lebih tangguh dalam menangani masalah ketimpangan data ekstrem (*class imbalance problem*) pada lalu lintas jaringan komputer dibandingkan dengan model probabilitas parametrik global konvensional [7].

5. KESIMPULAN

5.1. Kesimpulan

Berdasarkan hasil eksperimen dan pembahasan pada studi lanjutan deteksi *spammer* jaringan ini, dapat ditarik beberapa kesimpulan utama terkait performa integrasi model yang dikembangkan.

1. penerapan algoritma *Mean Shift Clustering* yang berbasis densitas terbukti sangat efektif dalam memetakan karakteristik lalu lintas jaringan secara adaptif. Tanpa memerlukan *input* jumlah kelompok di awal perhitungan, algoritma ini berhasil mengisolasi sebaran data secara organik menjadi 19 klaster mandiri serta memisahkan klaster anomali mikro dari kelompok lalu lintas utama.
2. penggunaan algoritma *Local Mean K-Nearest Neighbor* (LMKNN) sukses mengatasi kendala ketimpangan kelas data (*class imbalance*) yang menjadi kelemahan mendasar pada penelitian berbasis probabilitas global terdahulu. Melalui perhitungan metrik rata-rata jarak lokal, model mampu memberikan ruang representasi yang adil dan akurat bagi kelas minoritas (*spammer*).
3. hasil evaluasi *hiperparameter* menunjukkan bahwa nilai lingkungan tetangga lokal *K=3* merupakan parameter paling optimal untuk model usulan dengan capaian nilai Akurasi, Presisi, *Recall*, dan *F1-Score* murni sebesar 100%. Integrasi model baru ini terbukti secara signifikan berhasil meningkatkan nilai sensitivitas *Recall* sekaligus memangkas angka kebocoran deteksi (*False Negative*) dari 23 data pada penelitian terdahulu menjadi 0 data murni.

5.2. Saran

Meskipun sistem telah mencatatkan performa deteksi yang sempurna, terdapat beberapa aspek keterbatasan yang dapat dipertimbangkan sebagai saran pengembangan untuk penelitian mendatang. Administrator jaringan dan peneliti selanjutnya disarankan

untuk melakukan eksplorasi penggunaan metrik jarak alternatif selain *Euclidean Distance*, seperti *Manhattan Distance* atau *Mahalanobis Distance*, guna mengamati fluktuasi tingkat efisiensi komputasi model LMKNN apabila dihadapkan pada *dataset* lalu lintas yang berskala jauh lebih masif. Selain itu, arah pengembangan riset berikutnya juga dapat difokuskan pada optimasi algoritma pembatasan radius *bandwidth* pada *Mean Shift* dengan mengintegrasikan metode-metode optimasi berbasis *heuristik*. Penambahan algoritma optimasi tersebut diharapkan dapat mempercepat waktu konvergensi pencarian pusat densitas lalu lintas jaringan secara *real-time* sebelum data diklasifikasikan oleh LMKNN.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Unit Pelaksana Teknis Pusat Komputer Universitas Malikussaleh dan Program Studi Teknik Informatika Universitas Malikussaleh atas bantuan, fasilitasi, dan dukungan yang diberikan sepanjang penelitian ini, mulai dari pengumpulan data, pengembangan sistem, hingga penyelesaian tulisan ini.

DAFTAR PUSTAKA

[1] APJII, "Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang," Feb. 2024. [Online]. Available: https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang

[2] F. P. E. Putra, A. Zulfikri, M. A. Huda, Hasbullah, Mahendra, and M. Surur, "Analisis Keamanan Jaringan Dari Serangan Malware Menggunakan Firewall Filtering Dengan Port Blocking," *Digit. Transform. Technol.*, vol. 3, no. 2, pp. 857–863, Sep. 2023.

[3] E. S. Panggabean, C. I. Angkat, D. Sartika, D. Riswan, A. P. Utama, and M. Iqbal, "Memahami Spam Terhadap Digitalisasi Masyarakat Desa Perkebunan Sei Balai Kecamatan Sei Balai Kabupaten Batubara," *J. Pengabdi. Masy.*, vol. 2, no. 2, pp. 270–278, Dec. 2023.

[4] O. K. M. M. M. Majid, R. T. Adek, and Z. Yunizar, "Spammer Detection On Computer Networks Using Gaussian Naive Bayes Classifier And K-Medoids As Acquisition Training Data," in *Proceedings of International Conference on Multidisciplinary Engineering (ICOMDEN)*, 2024, p. 11.

[5] DQLab, "Gunakan Algoritma Data Science untuk Cek Spam Email," Mar. 2023.

[6] U. M. Fadil, D. S. Hasibuan, K. E. W. Siregar, W. Supi, H. F. Ramadani, and I. Rusydi, "PENERAPAN TEKNIK SMOTE UNTUK MENDETEKSI PERILAKU JARINGAN BERBASIS TRAFIK ENKRIPSI".

[7] A. Z. Putri, M. Daud, and H. A. K. Aidilof, "Klasifikasi Penyakit Jantung Berbasis Data Rekam Medis Menggunakan Algoritma Local Mean K-Nearest Neighbor," *J. Nas. Teknol. dan Sist. Inf.*, vol. 12, no. 1, pp. 134–143, 2026.

[8] A. Suarisman, A. Nazir, F. Syafria, and L. Afriyanti, "Perbandingan jarak metrik pada klasifikasi jamur beracun menggunakan algoritma K-Nearest Neighbor (K-NN)," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 10–19, 2023.

[9] F. Zahra, A. Khalif, and B. N. Sari, "PENGELOMPOKAN TINGKAT KEMISKINAN DI SETIAP PROVINSI DI INDONESIA MENGGUNAKAN ALGORITMA K-MEDOIDS," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1243–1249, 2024, doi: 10.23960/jitet.v12i2.4199.

[10] M. R. F. Damopolii, L. Zahrotun, and A. H. S. Jones, "Analisis Perbandingan Metode DBSCAN dan Mean Shift Dalam Mengelompokkan Data IPM Kabupaten/Kota se-Indonesia," *J. FASILKOM*, vol. 15, no. 3, pp. 562–571, 2025.

[11] R. F. Utari, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1017–1025, 2024.

[12] A. K. Wardhani, R. I. Sudra, E. Nugraha, and A. N. Putri, "Optimasi Nilai K pada Algoritma K-Nearest Neighbor untuk Klasifikasi Kesehatan Janin," *J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 44–51, 2025.

[13] N. Imanda, M. Teknologi, I. Universitas, and K. Lhokseumawe, "Penerapan Algorima Naive Bayes Pada," vol. 12, no. 3, 2024.

[14] J. K. Nainggolan, F. Sinaga, A. M. Sitorus, A. Khairia, and B. A. Wijaya, "Analisa Komparasi dengan Algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) untuk Prediksi Penyakit Jantung," *Dinamik*, vol. 30, no. 2, pp. 297–306, 2025.

[15] M. E. Siregar, S. Dermawan, and A. A. Hisyam, "Perbandingan Kinerja Naive Bayes Dan Knn Dalam Analisis Sentimen Komentar X Dengan Dan Tanpa Text Preprocessing (Studi Kasus: Danantara)," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.6612.