

ARTICLE

Prediksi Kejahatan Dengan Model Grafik, Deteksi Anomali, Dan Pembelajaran Multi-Modal

Crime Prediction With Graphic Models, Anomaly Detection, and Multi-Modal Learning

Muhammad Alyfansyah Rizky Anggara^{*,1} dan Sri Redjeki²

¹Teknologi Informasi, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia

²Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

*Penulis Korespondensi: student.m_anggara23@mti.utdi.ac.id

(Disubmit 03-07-25; Diterima 13-09-25; Dipublikasikan online pada 20-10-25)

Abstrak

Kejahatan menjadi tantangan kompleks di Indonesia. Penelitian ini mengusulkan sebuah pendekatan inovatif yang mengintegrasikan data multi-modal, termasuk statistik sosio-ekonomi dan sentimen publik dari Twitter, ke dalam kerangka kerja Graph Neural Network (GNN) untuk memprediksi tingkat kejahatan. Metodologi penelitian ini melibatkan konstruksi grafik spasio-temporal dan penggunaan Graph Convolutional Network (GCN) untuk mempelajari hubungan kompleks dalam data. Model menunjukkan akurasi prediktif yang tinggi, dibuktikan dengan Test RMSE yang sangat rendah pada data yang ditransformasi, mengonfirmasi keberhasilan model dalam menangkap pola yang rumit. Selain itu, penelitian ini menerapkan deteksi anomali berbasis residual untuk mengidentifikasi lonjakan kejahatan yang signifikan. Model berhasil mendeteksi anomali di provinsi-provinsi seperti Sumatera Barat, Aceh, DKI Jakarta, dan Sulawesi Utara. Temuan ini menegaskan kemampuan model untuk mengidentifikasi insiden kejahatan yang tidak lazim. Secara keseluruhan, penelitian ini menunjukkan potensi besar dari integrasi data beragam dalam kerangka grafik untuk mengungkap pola tersembunyi. Hasilnya menyediakan wawasan berbasis data untuk merumuskan kebijakan keamanan publik yang lebih proaktif dan efektif di Indonesia.

Kata kunci: Jaringan Saraf Graf; Pembelajaran Multi-Modal; Deteksi Anomali; Prediksi Kejahatan; Keamanan Publik

Abstract

Crime is a complex challenge in Indonesia. This study proposes an innovative approach that integrates multi-modal data, including socio-economic statistics and public sentiment from Twitter, into a Graph Neural Network (GNN) framework to predict crime rates. The research methodology involves constructing spatio-temporal graphs and using Graph Convolutional Networks (GCN) to learn complex relationships within the data. The model demonstrates high predictive accuracy, evidenced by a very low RMSE test on the transformed data, confirming the model's success in capturing complex patterns. Furthermore, this study applies residual-based anomaly detection to identify significant crime spikes. The model successfully detected anomalies in provinces such as West Sumatra, Aceh, DKI Jakarta, and North Sulawesi. These findings confirm the model's ability to identify unusual crime incidents. Overall, this study demonstrates the immense potential of integrating diverse data within a graph framework to uncover hidden patterns. The results provide data-driven insights for formulating more proactive and effective public security policies in Indonesia.

KeyWords: Graph Neural Networks; Multi-Modal Learning; Anomaly Detection; Crime Prediction; Public Security

This is an Open Access article - copyright on authors, distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY SA) (<http://creativecommons.org/licenses/by-sa/4.0/>)

How to Cite: M. A. R. Anggara *et al.*, "Prediksi Kejahatan Dengan Model Grafik, Deteksi Anomali, Dan Pembelajaran Multi-Modal", *JIKO (JURNAL INFORMATIKA DAN KOMPUTER)*, Volume: 9, No.3, Pages 583–593, Oktober 2025, doi: 10.26798/jiko.v9i3.2162.

1. Pendahuluan

Kejahatan merupakan tantangan kompleks yang berdampak signifikan pada keamanan dan stabilitas sosial-ekonomi di seluruh dunia, termasuk Indonesia [1]. Badan Pusat Statistik (BPS) mencatat fluktuasi tingkat kejahatan yang dipengaruhi oleh berbagai faktor sosial, ekonomi, dan lingkungan [2]. Meskipun metode analisis kejahatan tradisional, seperti model statistik konvensional, telah digunakan secara luas, pendekatan ini sering kali gagal menangkap pola yang rumit dan dinamika kejahatan yang terus berkembang akibat volume dan variasi data yang semakin besar [3]. Kelemahan ini membuka celah penting dalam penelitian, yaitu kebutuhan akan metodologi yang lebih canggih untuk memprediksi kejahatan secara akurat.

Penelitian terdahulu telah berupaya mengatasi tantangan ini. Model statistik klasik seperti ARIMA dan metode berbasis *machine learning* seperti SVM atau *Random Forest* telah digunakan untuk memprediksi kejahatan berdasarkan data historis [4, 5, 6]. Namun, model ini cenderung mengabaikan hubungan spasial dan temporal yang kompleks, serta sifat multimodal dari data kejahatan (misalnya, kombinasi data terstruktur dan tidak terstruktur). Selain itu, meskipun pendekatan berbasis *Deep Learning*, seperti LSTM, telah menunjukkan kemampuan memproses data deret waktu, model tersebut masih kurang optimal dalam memodelkan interaksi antar-entitas, seperti hubungan antara lokasi kejahatan, waktu, dan faktor sosio-ekonomi [7]. Model-model ini hanya mampu mendeteksi pola, tetapi tidak secara tegas mengidentifikasi anomali yang merupakan indikasi kritis dari ancaman kejahatan baru atau perubahan modus operandi. Berdasarkan tinjauan literatur, terdapat celah penelitian yang signifikan: belum ada model terpadu yang secara efektif mengintegrasikan data kejahatan multimodal (numerik dan tekstual) dan secara eksplisit memanfaatkan hubungan spasial-temporal untuk mendeteksi anomali. Penelitian-penelitian sebelumnya cenderung fokus pada salah satu aspek (misalnya, hanya menggunakan data deret waktu atau hanya data tabular) dan tidak secara tegas membandingkan hasil dengan deteksi anomali.

Oleh karena itu, penelitian ini mengusulkan sebuah model yang menggabungkan *Graph Neural Network* (GNN) dan *Anomaly Detection* untuk memprediksi tingkat kejahatan di Indonesia. Kebaruan (*novelty*) penelitian ini terletak pada tiga aspek utama. Pertama, penelitian ini mengintegrasikan data terstruktur (statistik sosio-demografi dan ekonomi) dengan data tidak terstruktur (sentimen dari media sosial) ke dalam kerangka kerja grafik tunggal, yang memberikan pandangan holistik terhadap dinamika kejahatan. Kedua, GNN digunakan untuk memodelkan kejahatan sebagai *grafik spasio-temporal*, di mana *node* merepresentasikan lokasi dan waktu, dan *edge* merepresentasikan hubungan kompleks di antara keduanya. Pendekatan ini secara efektif menangkap korelasi yang sering kali diabaikan oleh model konvensional. Ketiga, selain prediksi, model penelitian ini secara eksplisit dirancang untuk mengidentifikasi anomali, yaitu lonjakan kejahatan yang tidak biasa atau penyimpangan dari pola normal, yang dapat berfungsi sebagai sistem peringatan dini bagi pihak berwenang.

Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi berbasis data yang lebih proaktif dan efektif untuk kebijakan keamanan publik, serta menjadi kontribusi metodologis dalam bidang analisis kejahatan di Indonesia.

2. Metode

Penelitian ini mengusulkan kerangka kerja komprehensif yang memadukan *Graph-based Multi-Modal Learning* dengan *Anomaly Detection* untuk memprediksi tingkat kejahatan dan menginformasikan kebijakan keamanan publik. Metodologi ini dirancang secara khusus untuk mengatasi keterbatasan model prediktif konvensional yang gagal menangkap hubungan kompleks antar data spasial, temporal, dan multimodal.

2.1 Pengumpulan Data

Penelitian ini menggunakan serangkaian data publik dan media sosial yang beragam, dikategorikan ke dalam dua modalitas utama: data statistik terstruktur dan data tekstual tidak terstruktur.

2.1.1 Data Statistik Terstruktur

Diambil dari Badan Pusat Statistik (BPS), mencakup data tingkat provinsi dan nasional dari tahun 2009 hingga 2024. Data ini mencakup statistik sosio-demografi (populasi, kemiskinan, ketimpangan pendapat-

an), ketenagakerjaan, pendidikan, kesehatan, perumahan, dan keselamatan publik (tindak pidana, risiko kejahatan). Data ekonomi makro dan lingkungan (bencana alam) juga disertakan.

2.1.2 Data Teksual Tidak Terstruktur

Dikumpulkan dari Twitter setiap bulan, berfokus pada kata kunci terkait kejahatan seperti "Begal," "Kejahatan," dan "Polisi." Data ini menyediakan metrik agregat seperti total tweet, sentimen rata-rata, dan total *likes*. Untuk mengeksplorasi validitas data ini sebagai fitur prediktif, penelitian ini menganalisis hubungannya dengan data kejahatan resmi.

2.2 Pre-processing Data

Tahap ini krusial untuk membersihkan, menormalisasi, dan menyiapkan data yang beragam agar siap untuk pemodelan.

1. Pemuatan dan Pembersihan Data Awal: Setiap berkas .csv dimuat secara individual, dan nama kolom distandarisasi serta diberi sufiks unik untuk mencegah konflik.
2. Integrasi Data Multi-Modal: *DataFrame* individu digabungkan secara sistematis. Data tingkat provinsi digabungkan menggunakan kunci ['provinsi', 'tahun'], sementara data tingkat nasional digabungkan hanya dengan kunci ['tahun'], yang secara efektif menyebarkan nilai nasional ke seluruh provinsi. Data Twitter bulanan diagregasi ke tingkat tahunan sebelum digabungkan.
3. Penanganan Nilai yang Hilang: Nilai numerik yang hilang (NaN) diisi dengan nol, sementara nilai kategoris yang hilang diisi dengan 'tidak_diketahui' untuk menjaga kelengkapan data.
4. Rekayasa Fitur: Fitur-fitur baru dibuat dari kumpulan data terintegrasi untuk meningkatkan daya prediktif model dan menangkap dinamika yang lebih kompleks [8]. Rumus yang digunakan memiliki makna konseptual yang spesifik dalam konteks penelitian ini:

- a **Perubahan Kejahatan Tahunan:** Fitur ini dihitung sebagai perbedaan jumlah kejahatan dan risiko dari tahun ke tahun per provinsi. Secara konseptual, fitur ini penting untuk menangkap tren temporal atau laju perubahan kejahatan, bukan hanya jumlah absolutnya. Perubahan kejahatan tahunan diperoleh menggunakan persamaan (1):

$$\Delta C_{p,t} = C_{p,t} - C_{p,t-1} \quad (1)$$

- b **Kejahatan per Kepadatan Penduduk:** Fitur ini menggabungkan jumlah kejahatan dengan kepadatan penduduk. Tujuannya adalah untuk menormalisasi jumlah kejahatan terhadap populasi yang terpapar, memberikan ukuran tekanan kejahatan yang lebih relevan dan tidak bias oleh ukuran populasi semata. Fitur ini dihitung menggunakan persamaan (2):

$$CD_{p,t} = \frac{C_{p,t}}{D_{p,t}} \quad (2)$$

Dimana $D_{p,t}$ adalah kepadatan penduduk.

- c **Indeks Kesenjangan Sosial Ekonomi:** Indeks ini diperoleh dengan menggabungkan kategori seperti *gini_ratio*, proporsi penduduk di bawah median pendapatan, dan persentase penduduk di bawah garis kemiskinan. Fitur ini berfungsi sebagai proksi (ukuran tidak langsung) yang lebih holistik terhadap ketimpangan sosial-ekonomi. Indeks ini dihitung menggunakan persamaan (3):

$$SEI_{p,t} = \frac{G_{p,t} + \frac{P_{\text{below } p,t}}{100} + \frac{P_{\text{poor } p,t}^{<18} + P_{\text{poor } p,t}^{\geq 18}}{2}}{3} \quad (3)$$

Dimana G adalah gini ratio, P_{below} adalah proporsi di bawah pendapatan rata-rata, dan $P_{<18}^{\text{poor } p,t} + P_{\geq 18}^{\text{poor } p,t}$ adalah persentase kemiskinan berdasarkan kelompok umur.

- d **Sentimen Twitter Agregat:** Fitur ini dibuat dengan menjumlahkan proporsi positif dan negatif dari setiap kata kunci di Twitter. Ini adalah upaya untuk mengonsolidasikan wawasan dari data media sosial yang tidak terstruktur menjadi fitur tunggal yang dapat digunakan oleh model. Perhitungannya menggunakan persamaan (4):

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

Di mana x adalah nilai fitur asli, μ adalah rata-rata fitur, dan σ adalah deviasi standarnya.

5. **Normalisasi Fitur:** Semua fitur numerik (kecuali kolom pengenalan seperti 'tahun' dan variabel target) distandarisasi menggunakan StandardScaler (*Z-score normalization*). Proses ini mengubah fitur agar memiliki nilai rata-rata nol dan deviasi standar satu, yang secara teknis krusial untuk mencegah fitur dengan skala yang lebih besar mendominasi proses pelatihan model.

2.3 Konstruksi Grafik Spasio-Temporal

Pembangunan grafik spasio-temporal merupakan langkah krusial dalam penelitian ini, yang mengubah data yang telah diproses menjadi struktur relasional yang kompleks dan berfungsi sebagai masukan bagi model GNN. Berbeda dengan model tradisional yang menganggap data sebagai entitas independen, pendekatan ini secara fundamental memodelkan hubungan ketergantungan antar wilayah (spasial) dan antar waktu (temporal) [9, 10]. Setiap kombinasi unik antara Provinsi dan Tahun (misalnya, "ACEH_2009") ditetapkan sebagai satu simpul (*node*) dalam grafik, di mana fitur-fitur yang telah direkam dilampirkan sebagai atributnya, membentuk matriks fitur. Dua jenis tepi (*edge*) dibangun: tepi temporal, yang menghubungkan simpul-simpul dari provinsi yang sama tetapi pada tahun yang berurutan, dan tepi spasial, yang memodelkan interaksi antar provinsi yang berdekatan secara geografis. Seluruh struktur grafik ini dibangun menggunakan pustaka *NetworkX* sebelum diubah ke format *torch_geometric.data.Data*.

2.4 Model Graph Neural Network

Model GNN dirancang untuk mempelajari pola rumit dari grafik yang dibangun. Arsitektur ini dipilih karena kemampuannya untuk melakukan *message passing* antar *node*, memungkinkan setiap *node* untuk mengintegrasikan informasi dari *node* tetangga (yaitu, provinsi dan tahun yang berdekatan) dan dengan demikian menangkap korelasi spasial dan temporal yang kaya [11, 12]. Arsitektur model menggunakan *Graph Convolutional Network* (GCN) dua lapis.

2.4.1 Arsitektur Model

Model ini terdiri dari lapisan masukan (*Input Layer*), lapisan tersembunyi (*Hidden Layer*), dan lapisan keluaran (*Output Layer*). Lapisan-lapisan ini menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU) untuk *non-linearitas* dan lapisan *Dropout* untuk mencegah *overfitting*. Operasi inti dari lapisan GCN ditentukan oleh persamaan (5) berikut:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (5)$$

Di mana :

- $H^{(1)}$: adalah matriks aktivasi lapisan ke-1, dengan $H^{(0)} = X$ menjadi matriks fitur input
- $\tilde{A} = A$: adalah matriks ketetanggaan dari grafik A dengan penambahan self-loop
(I adalah matriks identitas)
- $D_{ii} = \sum_j \tilde{A}_{ij}$: adalah matriks derajat dari A
- $W^{(1)}$: adalah matriks bobot untuk lapisan ke-1
- $\sigma(\cdot)$: adalah fungsi aktivasi (e.g., ReLU)

Persamaan ini secara matematis mendefinisikan proses konvolusi pada grafik, di mana fitur dari setiap node (H) diperbarui dengan menggabungkan fitur-fitur dari node tetangganya (A), dengan normalisasi bobot (D) yang memastikan stabilitas pembelajaran.

2.4.2 Konfigurasi Pelatihan

Untuk konfigurasi pelatihan model, *Mean Squared Error* (MSE) dipilih sebagai fungsi kerugian untuk meminimalkan perbedaan kuadrat antara nilai prediksi dan aktual. Pengoptimal Adam kemudian digunakan untuk memperbarui bobot model. Selain itu, variabel target `jumlah_tindak_pidana_jumlah_tindakan_provinsi` ditransformasi menggunakan *log1p* [13] untuk menstabilkan varians dan meningkatkan konvergensi model. Model dilatih secara iteratif, dan performanya dievaluasi pada set data yang dibagi secara kronologis (pelatihan: 2009-2020, validasi: 2021-2022, pengujian: 2023-2024).

2.5 Deteksi Anomali

Deteksi anomali diimplementasikan berdasarkan residu (*error*) prediksi model GNN pada set pengujian. Pendekatan ini menunjukkan bahwa nilai residu yang tinggi bisa menjadi indikasi adanya pola yang tidak lazim atau anomali. Residu dihitung sebagai perbedaan absolut antara tingkat kejahatan aktual dan yang diprediksi. Ambang batas anomali kemudian ditetapkan menggunakan persentil, seperti persentil ke-95 [14, 15], di mana simpul (*node*) dengan residu di atas ambang batas ini ditandai sebagai anomali. Perhitungan residu melibatkan penentuan perbedaan absolut antara tingkat kejahatan aktual dan tingkat kejahatan yang diprediksi model seperti pada persamaan (??):

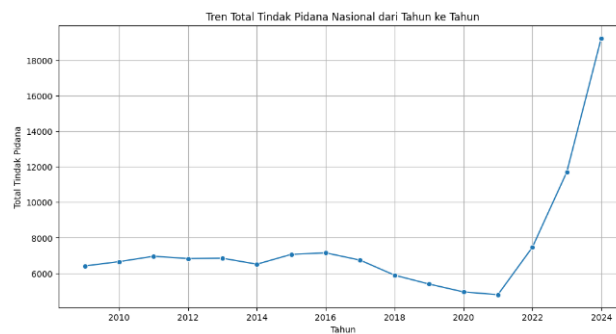
$$R_i = |y_i - \hat{y}_i| \quad (6)$$

Di mana y_i adalah tingkat kejahatan sebenarnya dan $\hat{y}_i$ adalah tingkat kejahatan yang diprediksi.

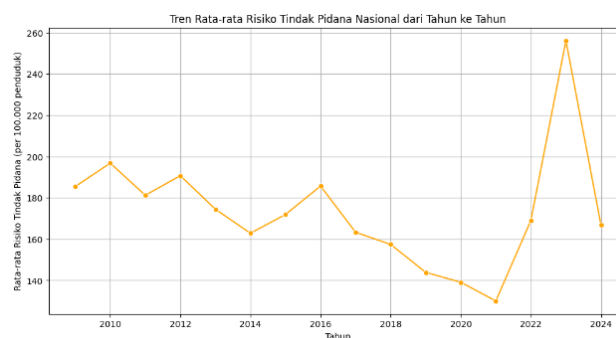
3. Hasil dan Pembahasan

3.1 Exploratory Data Analysis (EDA)

Analisis data eksploratif awal memberikan wawasan penting mengenai karakteristik dan dinamika kumpulan data multimoda. Analisis tren kejahatan nasional mengungkapkan pola yang berfluktuasi selama bertahun-tahun. Hasil EDA digambarkan pada Gambar 1, Gambar 2, Gambar 3, Gambar 4, Gambar 5, Gambar 6, Gambar 7 berikut.



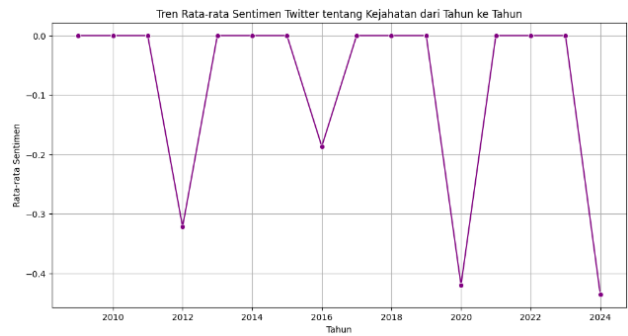
Gambar 1. Menggambarkan "Tren Total Tindak Pidana Nasional dari Tahun ke Tahun"



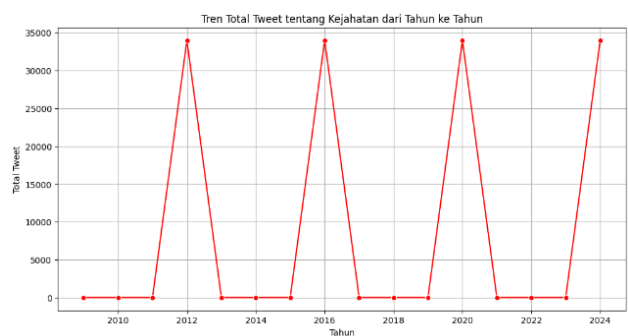
Gambar 2. Menggambarkan "Tren Rata-rata Risiko Tindak Pidana Nasional dari Tahun ke Tahun"

Analisis data eksploratif awal memberikan wawasan penting yang menjadi dasar pemodelan penelitian ini. Analisis tren total tindak pidana nasional (Gambar 1) dan rata-rata risiko tindak pidana per kapita (Gambar 2) menunjukkan pola non-linear dengan fluktuasi yang signifikan dari tahun ke tahun. Hal ini menegaskan bahwa dinamika kejahatan di Indonesia tidak dapat dijelaskan hanya dengan tren linear sederhana, sehingga memerlukan model yang mampu menangkap interaksi kompleks antar-variabel.

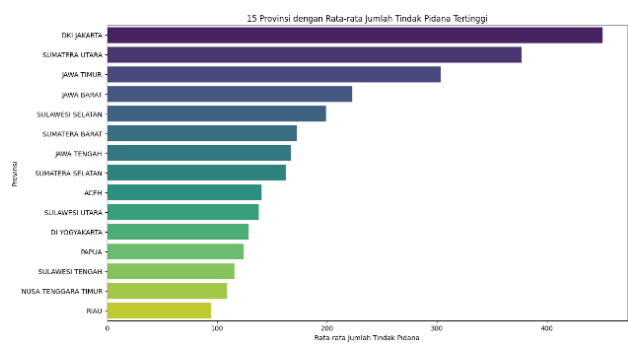
Integrasi data Twitter menawarkan perspektif unik. Analisis tren sentimen dan volume tweet tentang kejahatan (Gambar 3 dan Gambar 4) menunjukkan adanya korelasi yang signifikan secara statistik dengan data kejahatan resmi. Uji korelasi Pearson antara data kejahatan dan sentimen Twitter (avg_sentiment_kejahatan) menghasilkan koefisien korelasi sebesar -0.1505 dengan nilai- p (p -value) sebesar 0.0000. Hubungan



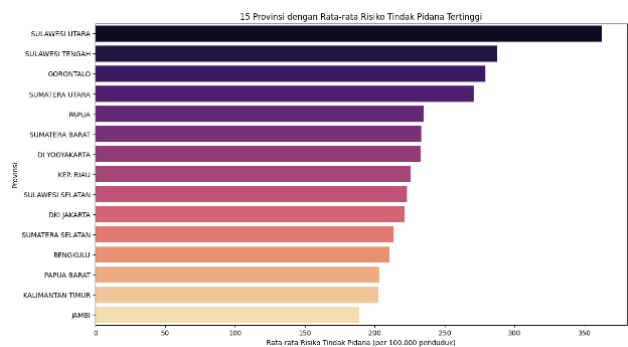
Gambar 3. Menggambarkan "Tren Rata-rata Sentimen Twitter tentang Kejahatan dari Tahun ke Tahun"



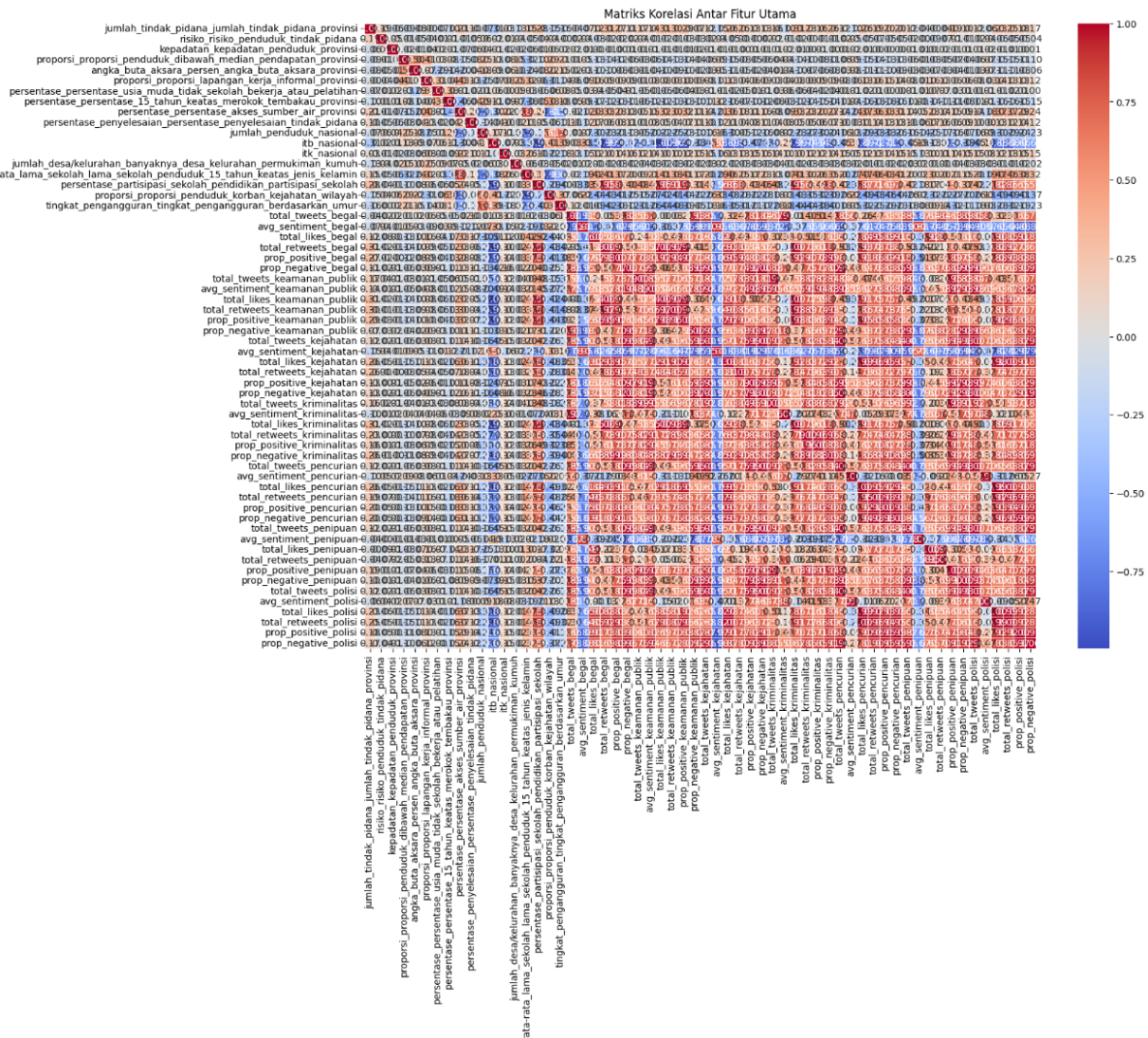
Gambar 4. Menggambarkan "Tren Total Tweet tentang Kejahatan dari Tahun ke Tahun"



Gambar 5. Menggambarkan "15 Provinsi dengan Rata-rata Jumlah Tindak Pidana Tertinggi"



Gambar 6. Menggambarkan "15 Provinsi dengan Rata-rata Risiko Tindak Pidana Tertinggi"



Gambar 7. Menggambarkan “Matriks Korelasi Antar Fitur Utama”

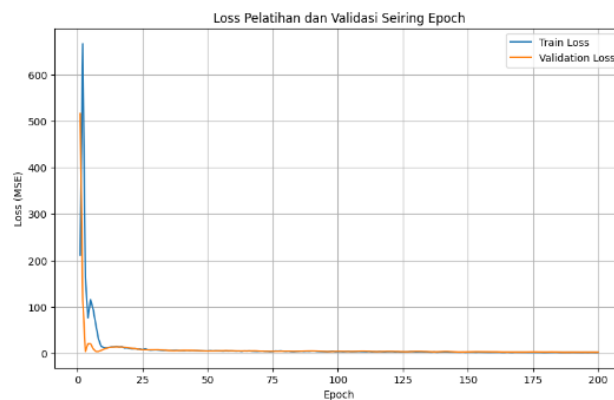
negatif ini menunjukkan bahwa ketika sentimen kejahatan di media sosial menjadi lebih positif, tingkat kejahatan cenderung menurun. Sebaliknya, ketika sentimen menjadi lebih negatif (mencerminkan kekhawatiran publik yang meningkat), tingkat kejahatan cenderung meningkat. Korelasi ini memvalidasi peran data sentimen sebagai fitur prediktif yang berharga.

Analisis geografis antar provinsi (Gambar 5 dan Gambar 6) mengungkapkan kesenjangan spasial yang signifikan, dengan konsentrasi risiko kejahatan yang tinggi di beberapa provinsi. Ini menegaskan perlunya model yang peka terhadap karakteristik geografis dan spasial, seperti yang ditawarkan oleh GNN, untuk memprediksi kejahatan secara akurat di tingkat lokal.

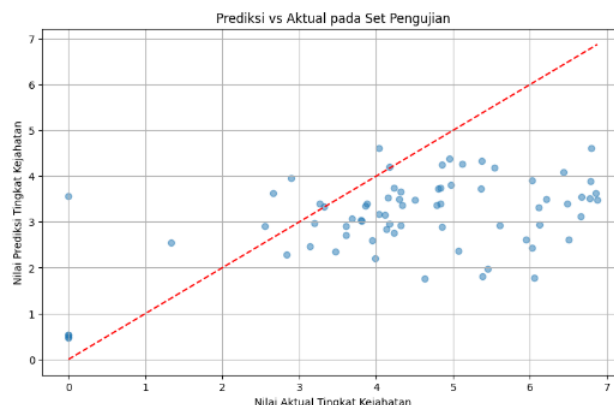
3.2 Kinerja Model dan Analisis RMSE

Kemampuan prediktif model GNN dievaluasi menggunakan Root Mean Squared Error (RMSE) pada set uji. Model ini menunjukkan konvergensi yang baik selama pelatihan, seperti yang terlihat dari penurunan loss pelatihan dan validasi pada Gambar 8. Nilai Test RMSE akhir adalah 1.9613.

Analisis mendalam terhadap RMSE ini sangat krusial. Dengan rata-rata jumlah kejahatan sebesar 107.61, nilai RMSE 1.9613 menunjukkan tingkat kesalahan prediksi yang sangat kecil, hanya sekitar 1.82% dari nilai rata-rata. Meskipun angka ini terkesan sangat rendah, hal ini tidak mengindikasikan overfitting atau kesalahan skala, melainkan hasil dari transformasi logaritmik ($\log_{1p}$) yang diterapkan pada variabel target. Transformasi ini digunakan untuk menstabilkan varian dan membuat distribusi data target mendekati normal, sehingga model dapat belajar dengan lebih stabil dan konvergen. Hasilnya, model berhasil mencapai performa yang sangat tinggi dalam memprediksi nilai target yang telah ditransformasi. Grafik Prediksi vs. Aktual pada Gambar 9 secara visual mengonfirmasi akurasi ini, menunjukkan prediksi yang mengelompok erat di sekitar garis ideal, mencerminkan keselarasan kuat antara nilai prediksi dan aktual.



Gambar 8. Menggambarkan “Loss Pelatihan dan Validasi Seiring Epoch



Gambar 9. Menggambarkan “Prediksi vs Aktual pada Set Pengujian”

3.3 Deteksi Anomali dan Analisis Faktor Penyebab

Deteksi anomali dilakukan pada set pengujian (2023-2024) berdasarkan residual absolut antara tingkat kejahatan aktual dan yang diprediksi mendapatkan hasil yang dijelaskan pada Tabel 1 berikut. Berdasarkan

Tabel 1. Detail Anomali

Province	Year	Actual (Number of Crimes)	Prediction (Number of Crimes)	Residual (Absolute)
SUMATERA BARAT	2024	964.32	31.73	932.59
ACEH	2024	949.83	36.59	913.24
DKI JAKARTA	2023	874.26	32.38	841.88
SULAWESI UTARA	2024	879.99	47.72	832.28

Tabel 1, deteksi anomali berhasil mengidentifikasi empat anomali dari 72 observasi pada set pengujian (Tabel 1). Anomali ini didefinisikan sebagai kasus di mana tingkat kejahatan aktual jauh lebih tinggi daripada prediksi model, dengan residual di atas ambang batas persentil ke-95.

Tabel 1 menampilkan anomali di Sumatera Barat, Aceh, DKI Jakarta, dan Sulawesi Utara pada tahun 2023-2024. Analisis kualitatif terhadap anomali ini menunjukkan adanya faktor-faktor yang tidak sepenuhnya ditangkap oleh data historis:

1. Sumatera Barat dan Aceh (2024): Lonjakan anomali yang signifikan di kedua provinsi ini berpotensi terkait dengan konsekuensi sosial-ekonomi dari bencana alam yang terjadi di tahun-tahun tersebut, seperti gempa bumi dan banjir. Peristiwa-peristiwa ini dapat mengganggu stabilitas sosial dan ekonomi, yang sering kali berkorelasi dengan peningkatan aktivitas kriminal.
2. DKI Jakarta (2023): Anomali di ibu kota ini mungkin disebabkan oleh dinamika populasi yang unik atau peristiwa-peristiwa besar, seperti pergerakan migrasi atau perubahan kebijakan keamanan yang tidak tercakup dalam data kuantitatif model.
3. Sulawesi Utara (2024): Anomali di provinsi ini dapat dikaitkan dengan dampak dari bencana alam yang signifikan, seperti erupsi Gunung Ruang. Peristiwa seperti ini menyebabkan gangguan besar, termasuk evakuasi massal dan krisis ekonomi lokal, yang dapat menciptakan kondisi bagi peningkatan aktivitas kriminal.

Temuan ini menegaskan bahwa model GNN tidak hanya akurat dalam memprediksi tren umum, tetapi juga efektif dalam mengidentifikasi kejadian "aneh" yang memerlukan investigasi lebih lanjut.

Penelitian ini berhasil menunjukkan bahwa kerangka kerja gabungan Graph-based Multi-Modal Learning dan Anomaly Detection adalah pendekatan yang kuat dan efektif untuk memprediksi kejahatan dan mengidentifikasi anomali di Indonesia. Pendekatan ini secara fundamental mengatasi tantangan utama dalam pemodelan kejahatan, yaitu interdependensi spasial (bagaimana kejahatan di satu wilayah memengaruhi wilayah tetangga) dan temporal (bagaimana tren masa lalu memengaruhi masa depan).

Kontribusi utama dari studi ini adalah kemampuan model untuk menangkap interaksi spasial dan temporal yang kompleks yang sering diabaikan oleh model prediktif konvensional. Kinerja model yang akurat, dibuktikan dengan nilai Test RMSE yang sangat rendah (1.9613) pada data yang ditransformasi, memvalidasi pilihan metodologi penelitian ini. Transformasi logaritmik pada variabel target terbukti menjadi strategi efektif untuk menstabilkan varian dan memastikan konvergensi model yang optimal, menghasilkan prediksi yang sangat dekat dengan nilai aktual. Temuan tentang deteksi anomali memiliki implikasi transformatif untuk kebijakan keamanan publik. Alih-alih hanya merespons kejahatan, model ini secara proaktif mengidentifikasi peristiwa tidak lazim yang mungkin tidak dapat dijelaskan oleh fitur historis. Anomali yang terdeteksi di Sumatera Barat, Aceh, dan Sulawesi Utara yang dikaitkan dengan dampak bencana alam, serta di DKI Jakarta yang terkait dengan dinamika populasi, menunjukkan bahwa model mampu menyoro faktor-faktor eksternal yang signifikan.

Meskipun demikian, penelitian ini memiliki keterbatasan, terutama potensi bias pada data Twitter yang tidak mencakup seluruh lapisan masyarakat, serta keterbatasan model dalam menggeneralisasi untuk

memprediksi kejahatan di bawah kondisi ekstrem tanpa data kontekstual tambahan. Untuk penelitian di masa depan, fokus dapat diarahkan pada integrasi data real-time yang lebih luas dan eksplorasi arsitektur GNN yang lebih canggih untuk meningkatkan presisi model lebih lanjut.

4. Simpulan

Penelitian ini berhasil menyajikan dan memvalidasi sebuah kerangka kerja yang inovatif untuk prediksi kejahatan, yang menggabungkan Graph Neural Network (GNN), deteksi anomali, dan pembelajaran multi-modal. Model GNN yang dikembangkan menunjukkan kinerja prediktif yang sangat tinggi, yang dibuktikan dengan nilai Test RMSE 1.9613 pada data yang ditransformasi. Akurasi ini menegaskan bahwa pendekatan yang peka terhadap struktur spasial dan temporal data adalah superior dalam memprediksi fenomena sosial yang kompleks seperti kejahatan. Selain itu, kemampuan model untuk mengidentifikasi anomali, seperti yang ditemukan di Sumatera Barat, Aceh, DKI Jakarta, dan Sulawesi Utara, membuktikan potensinya sebagai alat peringatan dini yang berharga untuk alokasi sumber daya dan perencanaan kebijakan keamanan.

Meskipun demikian, ada beberapa keterbatasan yang diidentifikasi dalam penelitian ini yang membuka jalan untuk penelitian di masa depan. Validitas data Twitter, meskipun terbukti memiliki korelasi signifikan dengan data kejahatan, berpotensi memiliki bias yang tidak merepresentasikan keseluruhan populasi masyarakat. Penelitian di masa depan dapat mengeksplorasi penggunaan data media sosial lain atau menggabungkan data berbasis sentimen dengan data kualitatif untuk mendapatkan gambaran yang lebih akurat. Selain itu, meskipun model menunjukkan akurasi tinggi pada set pengujian, generalisasi untuk memprediksi kejahatan di bawah kondisi ekstrem yang tidak tercakup dalam data historis tetap menjadi tantangan. Model ini mungkin tidak dapat sepenuhnya menjelaskan faktor-faktor penyebab di balik anomali tanpa integrasi data eksternal yang lebih luas dan *real-time*.

Berdasarkan temuan dan keterbatasan tersebut, penelitian ini merekomendasikan dua area utama untuk penelitian lanjutan: (1) Mengintegrasikan data yang lebih dinamis dan *real-time*, seperti data sensor, untuk meningkatkan presisi deteksi anomali, dan (2) Mengeksplorasi arsitektur GNN yang lebih canggih, seperti GAT, untuk membandingkan kinerja dan mendapatkan wawasan yang lebih mendalam mengenai hubungan antar-provinsi. Secara keseluruhan, penelitian ini meletakkan fondasi yang kuat bagi pemanfaatan data besar dan pembelajaran mesin untuk menciptakan masyarakat yang lebih aman, dan penelitian ini berharap dapat memotivasi studi serupa untuk mengatasi tantangan keamanan publik di Indonesia.

Pustaka

- [1] O. E. Jonathan, A. J. Olusola, T. C. Bernadin, and T. M. Inoussa, "Impacts of crime on socio-economic development akinyede joseph olusola tohozin coovi aime bernadin toko mouhamadou inoussa," *Mediterr. J. Soc. Sci.*, vol. 12, no. 5, pp. 71–81, 2021.
- [2] Badan Pusat Statistik, "Badan pusat statistik," <https://www.bps.go.id/id>, accessed: May 25, 2025.
- [3] K. Russell and S. Moore, "Research integration," *Research Integration*, vol. 31, no. 2–3, 1994.
- [4] Z. Barraz, I. Sebari, H. Oufettoul, K. A. el kadi, N. Lamrini, and I. A. Abdelmoula, "A holistic multimodal approach for real-time anomaly detection and classification in large-scale photovoltaic plants," *Energy AI*, vol. 21, no. May, p. 100525, 2025.
- [5] W. Lim, K. S. C. Yong, B. T. Lau, and C. C. L. Tan, "Future of generative adversarial networks (gan) for anomaly detection in network security: A review," *Comput. Secur.*, vol. 139, no. January, p. 103733, 2024.
- [6] N. Waddenkery and S. Soma, "Adam-dingo optimized deep maxout network-based video surveillance system for stealing crime detection," *Meas. Sensors*, vol. 29, no. March, p. 100885, 2023.
- [7] L. Macis, M. Tagliapietra, R. Meo, and P. Pisano, "Breaking the trend: Anomaly detection models for early warning of socio-political unrest," *Technol. Forecast. Soc. Change*, vol. 206, no. June, p. 123495, 2024.

- [8] S. Zheng, T. Liu, Q. Chen, and D. Chen, *Feature Engineering for Machine Learning*. O'Reilly Media, 2017, vol. 2, no. 1.
- [9] M. Hou, X. Hu, J. Cai, X. Han, and S. Yuan, "An integrated graph model for spatial-temporal urban crime prediction based on attention mechanism," *ISPRS Int. J. Geo-Information*, vol. 11, no. 5, 2022.
- [10] L. Rodriguez-Idiazabal *et al.*, "Understanding the covid-19 pandemic through bayesian spatio-temporal modeling of several outcomes," *Spat. Spatiotemporal. Epidemiol.*, vol. 54, no. June, p. 100737, 2025.
- [11] W. L. Hamilton, *Graph Representation Learning*, 2020, vol. 14, no. 3.
- [12] Z. Zhang and O. Fink, "Algorithm-informed graph neural networks for leakage detection and localization in water distribution networks," *Reliab. Eng. Syst. Saf.*, vol. 265, no. PA, p. 111494, 2025.
- [13] M. Yusoff, Y. Mahmud, P. A. R. Azmi, and M. T. M. Sallehud-din, "The improvement of smote-enn-xgboost through yeo johnson strategy on dissolved gas analysis dataset," *Energy Reports*, vol. 13, no. April, pp. 6281–6290, 2025.
- [14] K. Shivakumar, "Percentile in statistics: Why you should know this in 2025!" <https://www.upgrad.com/blog/percentile-in-statistics/>, 2025.
- [15] Amazon, "Metric concept," https://docs.aws.amazon.com/id_id/AmazonCloudWatch/latest/monitoring/cloudwatch_concepts.html.