

Penerapan Algoritma *Random Forest* Dalam Analisis Dan Prediksi Tingkat Kejahatan Berdasarkan Data Kriminalitas Polres Jember

Nurfarsyah Yoandito¹⁾, Ulya Anisatur Rosyidah²⁾, Ginanjar Abdurrahman³⁾

¹Mahasiswa Prodi Teknik Informatika, Universitas Muhammadiyah Jember
email: yoandito03@gmail.com

²Dosen Teknik Informatika, Universitas Muhammadiyah Jember
email: ulyaanisatur@unmuhjember.ac.id

³Dosen Teknik Informatika, Universitas Muhammadiyah Jember
email: abdurrahmanginanjar@unmuhjember.ac.id

Abstrak

Kejahatan merupakan permasalahan sosial yang dapat mengganggu stabilitas masyarakat, terutama di wilayah dengan intensitas aktivitas tinggi seperti Kota Jember. Penelitian ini bertujuan untuk menganalisis dan memprediksi tingkat kejahatan berdasarkan data kriminalitas dari Polres Jember tahun 2019–2023 menggunakan algoritma *Random Forest*. Data yang digunakan mencakup informasi jenis kejadian, waktu kejadian, tanggal, dan lokasi kejadian. Tahapan penelitian meliputi *Preprocessing Data*, pembentukan model klasifikasi kerawanan kejahatan (rawan dan tidak rawan), serta evaluasi performa model berdasarkan akurasi, presisi, recall, dan F1-score. Untuk mendukung visualisasi spasial dari hasil prediksi, digunakan perangkat lunak QGIS. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki akurasi sebesar 98% pada data uji, Recall 100%, dan F1-score 71,16%, dengan presisi sebesar 55,26%. Meskipun model memiliki sensitivitas tinggi terhadap kejadian rawan, presisi yang relatif rendah menunjukkan perlunya peningkatan dalam penanganan kesalahan klasifikasi positif palsu. Integrasi antara machine learning dan analisis spasial ini diharapkan dapat membantu aparat penegak hukum dalam merancang strategi pencegahan kejahatan yang lebih efektif dan berbasis data

Kata Kunci: Kejahatan, *Random Forest*, Prediksi, Klasifikasi, QGIS.

Abstract

Crime is a social issue that can disrupt community stability, especially in areas with high levels of activity, such as Jember City. This study aims to analyze and predict crime rates using the Random Forest algorithm based on crime data from the Jember Police Department from 2019 to 2023. The dataset includes information on the type, time, date, and location of each incident. The research stages include data preprocessing, classification model development for crime vulnerability (vulnerable and not vulnerable), and model performance evaluation using accuracy, precision, recall, and F1-score. Spatial visualization of prediction results was carried out using QGIS software. The results show that the Random Forest algorithm achieved 98% accuracy on the test data, with 100% recall, 71.16% F1-score, and 55.26% precision. Although the model shows high sensitivity in detecting vulnerable areas, the relatively low precision indicates the need for improvement in reducing false positives. The integration of machine learning and spatial analysis is expected to support law enforcement in designing more effective and data-driven crime prevention strategies..

Keywords: Crime, *Random Forest*, Prediction, Classification, QGIS

1. PENDAHULUAN

Kejahatan merupakan persoalan global yang terus dihadapi berbagai wilayah, terutama

di daerah perkotaan dengan aktivitas masyarakat yang padat. Ketimpangan pembangunan sering kali menjadi salah satu pemicunya. Kabupaten Jember sebagai salah

satu wilayah dengan populasi besar di Jawa Timur turut menghadapi tantangan serupa. Data Polres Jember menunjukkan adanya berbagai jenis tindak kriminal yang terjadi setiap tahun, seperti pencurian, kekerasan, dan penyalahgunaan narkoba (Kepolisian Negara Republik Indonesia Daerah Jawa Timur Resort Jember, 2024) dapat dilihat pada **Tabel 1**.

Tabel 1. Kejahatan Kab.Jember Thn. 2019-2023

KECAMATAN	TAHUN				
	2019	2020	2021	2022	2023
Ajung	68	46	73	123	78
Ambulu	80	45	73	131	74
Arjasa	88	47	80	106	98
Balung	77	48	72	129	79
Bangsalsari	88	55	60	135	80
Jenggawah	79	43	68	107	103
Kalisat	88	53	71	139	94
Kaliwates	95	47	72	107	80
Kencong	74	36	46	132	91
Mayang	75	55	60	129	94
Mumbulsari	84	35	73	132	97
Pakusari	81	44	77	120	94
Panti	108	51	65	131	100
Patrang	81	44	79	134	120
Rambipuji	83	40	63	149	88
Silo	84	56	71	123	93
Sukorambi	72	64	73	134	93
Sumbersari	96	48	86	115	97
Tanggul	77	54	82	109	99
Tempurejo	61	46	75	139	82
Umbulsari	59	49	70	137	86
Wuluhan	78	55	85	112	80

Sumber. Badan Pusat Statistik Jember, 2025.

Upaya penanggulangan kejahatan tidak cukup hanya dengan pendekatan reaktif. Diperlukan strategi prediktif berbasis data untuk mendeteksi potensi kejahatan secara dini. Pemanfaatan teknologi seperti *Data Mining* dan *Machine Learning* menjadi solusi, salah satunya melalui algoritma *Random Forest* yang dikenal mampu menangani data kompleks dengan akurasi tinggi(Alma Zuhairah, 2022).

Beberapa penelitian sebelumnya telah membuktikan efektivitas *Random Forest* dalam memprediksi kasus-kasus kejahatan, meskipun sebagian besar studi masih berfokus pada wilayah metropolitan atau kasus spesifik (Khatun et al., 2021). Oleh karena itu, penelitian ini berupaya mengisi celah tersebut dengan mengaplikasikan *Random Forest* untuk

menganalisis dan memprediksi pola kejahatan di Kota Jember berdasarkan data kriminalitas tahun 2019–2023. Selain prediksi, penelitian ini juga menyajikan visualisasi spasial menggunakan QGIS untuk mendukung strategi keamanan yang lebih tepat sasaran. Dengan pendekatan berbasis lokal dan integrasi analisis spasial, penelitian ini diharapkan dapat menjadi kontribusi nyata dalam pengelolaan keamanan daerah melalui teknologi prediktif yang aplikatif (shodiq & papilaya, 2019). Berdasarkan latar belakang tersebut, dapat dirumuskan beberapa permasalahan yang akan dibahas dalam penelitian ini, yaitu:

- 1) Bagaimana hasil analisis dan prediksi tingkat kejahatan ini dapat digunakan untuk membantu aparat penegak hukum dalam merumuskan strategi pencegahan kejahatan?
- 2) Seberapa besar tingkat akurat (*Accuracy*), presisi (*Precision*) dan recall algoritma *Random Forest* dalam memprediksi jenis kejahatan dan lokasi rawan kejahatan berdasarkan data historis?

Dari rumusan tersebut tujuan penelitian ini yaitu:

- (1) Menghasilkan model prediksi lokasi rawan kejahatan yang dapat digunakan sebagai dasar dalam merumuskan strategi pencegahan oleh aparat penegak hukum.
- (2) Menganalisis tingkat kejahatan di Kota Jember berdasarkan data kriminalitas dari Polres Jember serta mengukur kinerja model *Random Forest* dalam memprediksi tingkat kerawanan (ditinjau dari akurasi, presisi, dan *Recall*).

1.5. Batasan Penelitian

Untuk memfokuskan penelitian ini, maka perlu menetapkan batasan masalah sebagai berikut:

- 1) Penelitian ini menggunakan data sekunder berupa laporan kejadian kriminal dari Polres Jember yang dihimpun oleh Badan Pusat Statistik Jember tahun 2019 – 2023 dengan jumlah data sebanyak 9184 data.
- 2) Fokus penelitian ini dibatasi pada empat fitur utama yang digunakan sebagai variabel input dalam model prediksi, yaitu waktu kejadian, tanggal kejadian, jenis kejahatan, dan lokasi kejadiannya. Batasan

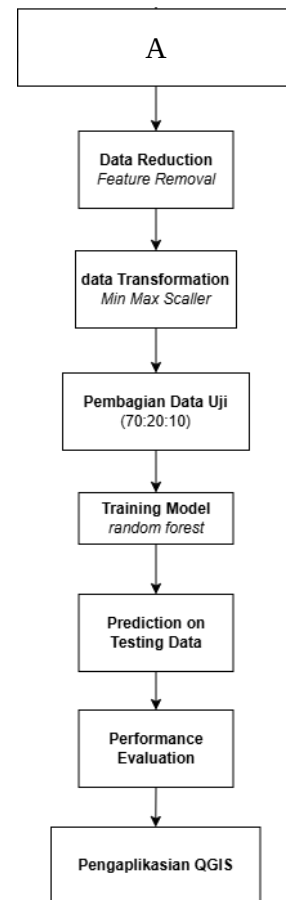
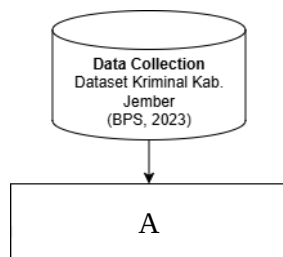
ini ditetapkan untuk menjaga kejelasan analisis dan menghindari *Noise* dari variabel yang kurang stabil.

- 3) Menggunakan algoritma *Random Forest* sebagai model dasar, dengan dilakukan *Tuning Parameter* untuk meningkatkan akurasi model.
- 4) Kinerja model dievaluasi menggunakan metrik seperti *Accuracy*, *Precision*, *Recall* dan *F1-Score* untuk mengukur kemampuan model dalam memprediksi kejadian kejahatan.
- 5) Analisis yang akan digunakan yaitu, analisis deskriptif untuk menggambarkan karakteristik umum dari data kejahatan, seperti distribusi frekuensi jenis kejahatan, lokasi kejadian, dan waktu kejadian. Analisis spasial akan dilakukan untuk mengidentifikasi hotspot kejahatan atau area dengan risiko kejahatan tinggi menggunakan peta yang divisualisasikan menggunakan *Tools QGIS*

2. METODE PENELITIAN

A. Diagram Alur Penelitian

Penelitian dilakukan menggunakan metode kuantitatif dengan pendekatan *Supervised Learning*. Dengan diagram alur penelitian seperti pada bagan dibawah.



B. Fitur yang Digunakan

Fitur yang digunakan antara lain waktu kejadian, tanggal, jenis kejadian, dan lokasi. Proses penelitian meliputi:

1. Pra-pemrosesan data (pembersihan, *Encoding*, transformasi)
2. Pembagian data latih, validasi dan uji (70:20:10)
3. Penerapan algoritma *Random Forest* dengan *Tuning Parameter*
4. Penyeimbangan data menggunakan SMOTE
5. Evaluasi model menggunakan *Confusion Matrix*, serta metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*
6. Visualisasi spasial hasil prediksi menggunakan perangkat lunak QGIS

3. HASIL DAN PEMBAHASAN

A. Deskripsi Dataset

Dataset yang digunakan berasal dari Polres Jember dengan total sebanyak 9.184 data kriminalitas selama tahun 2019–2023 (Badan Pusat Statistik Jember, 2023). Pada tahap ini, data-data yang relevan dikumpulkan dari sumber yang dapat dipercaya. Data ini mencakup laporan kriminalitas dari tahun 2019 hingga 2023. Dari data yang tersedia terdapat 7 variabel yaitu entri yang berisi informasi jenis kejahatan, waktu kejadian, tanggal kejadian, cuaca dan lokasi kejadian (berdasarkan kecamatan). Sebelum dilakukan pelatihan model, data mengalami proses pembersihan dimana data yang tidak digunakan akan dihapus dan data kosong akan diisi (Cahyo & Nilogiri, 2021), dilanjutkan dengan transformasi data dimana merupakan langkah utama dalam proses ini adalah membentuk variabel target baru yang disebut status_kerawanan, yaitu klasifikasi lokasi kejadian menjadi dua kategori: "Rawan" dan "Tidak Rawan". Penentuan label ini dilakukan hanya 2 kategori karena berdasarkan nilai pada kolom jumlah_kejahatan yang tidak lebih dari 3 kasus.

Adapun ketentuannya adalah sebagai berikut:

- Jika jumlah_kejahatan lebih dari 1, maka data diklasifikasikan sebagai "Rawan"
- Jika jumlah_kejahatan kurang dari atau sama dengan 1, maka diklasifikasikan sebagai "Tidak Rawan",

sehingga rincian data menjadi seperti pada

Tabel 2.

Tabel 2 Transformasi Data

No.	Tanggal kejadian	Lokasi	cuaca	Jumlah Kejadian	Tahun	Bulan	Hari ke	Akhir Pekan
1	02-01-2019	Ajung	Hujan	2	2019	1	2	0
2	08-01-2019	Ajung	Mendung	1	2019	1	8	0
3	10-01-2019	Ajung	Cerah	1	2019	1	10	0
4	17-01-2019	Ajung	Berawan	1	2019	1	17	0
5	20-01-2019	Ajung	Berawan	1	2019	1	20	1
6	21-01-2019	Ajung	Hujan	1	2019	1	21	0
...	...	...	...	...	...	...	...	...
8970	18-12-2019	Wuluhan	Cerah	1	2023	12	18	0
8971	25-12-2019	Wuluhan	Badai	2	2023	12	28	0

Sumber : Hasil Pengujian, 2025.

Selanjutnya penyeimbangan kelas menggunakan metode SMOTE atau *Synthetic Minority Oversampling Technique*. Hal ini

dilakukan untuk mengatasi ketidakseimbangan antara jumlah data kategori “rawan” (minoritas) dan “tidak rawan” (mayoritas) (Khatun et al., 2021). Dengan hasil distribusi label hasil *Resampling* pada **Tabel 3**.

Tabel 3. Sebelum dan sesudah SMOTE

Kelas	Distribusi Label Training	
	Sebelum SMOTE	Setelah SMOTE
Tidak Rawan	6133	6133
Rawan	146	6133

Sumber : Hasil Pengujian, 2025.

Selain itu ntuk mengatasi sifat data yang tidak seimbang lanjutan, dilakukan proses berlapis setelah SMOTE yaitu *Tuning Hyperparameter*, proses ini dilakukan menggunakan GridSearchCV dengan 3-fold cross-validation, serta menggunakan metrik f1_macro sebagai dasar evaluasi karena. Parameter yang diuji meliputi jumlah *Estimator* (n_estimators), kedalaman pohon (max_depth), jumlah minimum sampel untuk pemisahan (min_samples_split), dan penyesuaian bobot kelas (class_weight). Model terbaik dipilih berdasarkan kombinasi parameter yang menghasilkan nilai f1_macro tertinggi dan selanjutnya digunakan dalam proses prediksi kerawanan kejahatan (Khairunnisa, 2023). Sehingga model model menghasilkan konfigurasi terbaik dengan hasil pada **Tabel 4**.

Tabel 4. Hasil *Hyperparameter*

n_esti mators	max_ depth	min_ samples _split	mean_ test_ score	Rank Test Score
200	10	5	0.989401	1
200	None	5	0.989319	2
200	10	2	0.989319	3

Sumber : Hasil Pengujian, 2025.

B. Implementasi Random Forest

Random Forest digunakan karena keandalannya dalam klasifikasi dan kemampuannya menangani data kompleks serta non-linear. Model dilatih menjadi 70:20:10 (Wijoyo et al., 2024). 70 untuk data latih, 20 untuk data validasi dan 10 untuk data ujicoba. Untuk menentukan akar pertama pada penghitungan manual yaitu dengan penghitungan nilai *Gini Impurity*. *Gini Impurity* adalah matrik yang digunakan untuk mengukur kemurnian suatu akar pada pohon keputusan

semakin rendah nilai gini maka data pada *node* dinilai semakin homogen atau murni. Nilai gini berada antara 0 dan 1, Dimana nilai 0 berarti data yang sangat murni dan 1 berarti tidak murni (Trivusi, 2022).

Adapun rumus dari nilai Gini dapat dilihat dalam persamaan (1) dibawah

$$Gini = \sum (pi)^2 \quad (1)$$

Setelah melakukan perhitungan dengan sistem. Hasil prediksi dapat dilihat pada **Tabel 5** dan **Tabel 6**.

Tabel 5 Hasil Validation Test Set 20%

No	Lokasi	Tahu	Bula	Hari	Hari	Jumlah	Rolling	Akhir	Cuaca	Y	Y
	Enc	n	n	Ke	sblmnya	3hari	Pekan	Enc	True	Pred	
1	20	2019	9	28	5	10	10	1	0	0	0
2	11	2019	3	14	3	10	10	0	1	0	0
3	4	2019	8	20	1	10	10	0	3	0	0
4	1	2019	3	4	0	10	10	0	4	0	0
5	2	2019	9	25	2	10	10	0	4	0	0
...	...	...	...	...	...	...	...	...	...	...	...
1793	10	2023	12	1	4	10	10	0	2	0	0
1794	7	2023	10	4	2	10	10	0	1	0	0

Sumber : Hasil Pengujian, 2025.

Tabel 6 Hasil Test Set 10%

No	Lokasi	Tahu	Bula	Hari	Hari	Jumlah	Rolling	Akhir	Cuaca	Y	Y
	Enc	n	n	Ke	sblmnya	3hari	Pekan	Enc	True	Pred	
1	6	2019	5	13	0	10	10	0	1	0	0
2	7	2019	4	2	1	10	10	0	2	0	0
3	10	2019	2	26	1	10	10	0	2	0	0
4	15	2019	7	5	4	10	10	0	4	0	0
5	10	2019	7	24	2	10	10	0	4	0	0
...	...	...	...	...	...	...	...	...	...	...	...
897	8	2023	12	25	0	10	10	0	2	0	0
898	17	2023	1	6	4	10	10	0	4	0	1

Sumber : Hasil Pengujian, 2025.

C. Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan *Confusion Matrix* dan metrik seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score* (Moerdyanto et al., 2023). Dengan Validation Set model yang diuji terhadap 1.794 data yang terdiri atas dua kelas, yaitu kelas 0 (Tidak Rawan) dan kelas 1 (Rawan). Dengan tabel *Confusion Matrix* yang dapat dilihat pada **Tabel 7**.

Tabel 7 *Confusion Matrix Validation Set*

Kelas	Prediksi Tidak Rawan (0)	Prediksi Rawan (1)
-------	--------------------------	--------------------

Aktual 0	1703 (True Negative)	50 (False Positive)
Aktual 1	0 (False Negative)	41 (True Positive)

Sumber: Hasil Pengujian, 2025.

Total Data : 1.794

Distribusi aktual :

(1) Tidak Rawan (0) : 1.753

(2) Rawan (1) : 41

Perhitungan *Validation Set*

(1) *Accuracy* (Akurasi):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{41+1703}{41+1703+50+0} = \frac{1744}{1794} = 0,9721 = 97,21\%$$

(2) *Precision* (Presisi):

$$Precision = \frac{TP}{FP+TP} = \frac{41}{41+50} = \frac{41}{91} = 0,48 = 48\%$$

(3) *Recall*:

$$Recall = \frac{TP}{FP+FN} = \frac{41}{41+0} = 1,0 = 100\%$$

(4) *F1-Score*:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,4505 \times 1,0}{0,4505 + 1,0} = 0,6513 = 65,13\%$$

Dengan rangkuman pada **Tabel 8**.

Tabel 8. Hasil Uji *Validation Set*

Metrik	Nilai
Akurasi	97,21%
Presisi	48%
Recall	100%
F1-Score	65,13%

Sumber: Hasil Pengujian, 2025.

Selanjutnya yaitu perhitungan evaluasi untuk *Data Test Set* dengan tabel *Confusion Matrix* yang dapat dilihat pada **Tabel 9**.

Tabel 9 *Confusion Matrix Test Set*

Kelas	Prediksi Tidak Rawan (0)	Prediksi Rawan (1)
Aktual 0	860 (True Negative)	17 (False Positive)
Aktual 1	0 (False Negative)	21 (True Positive)

Sumber: Hasil Pengujian, 2025.

Total Data : 898

Distribusi actual :

(1) Tidak Rawan (0) : 877

(2) Rawan (1) : 21

Perhitungan *Validation Set*

(1) *Accuracy* (Akurasi):

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{21+860}{21+860+17+0} \\ &= \frac{881}{898} = 0,980 \\ &= 98,0\% \end{aligned}$$

(2) *Precision* (Presisi):

$$\begin{aligned} \text{Precision} &= \frac{TP}{FP+TP} = \frac{21}{21+17} \\ &= \frac{21}{38} = 0,5526 \\ &= 55,26\% \end{aligned}$$

(3) *Recall*:

$$\begin{aligned} \text{Recall} &= \frac{TP}{FP+FN} = \frac{21}{21+0} \\ &= 1,0 \\ &= 100\% \end{aligned}$$

(4) *F1-Score*:

$$\begin{aligned} \text{F1-Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \\ &2 \times \frac{0,5526 \times 1,0}{0,5526 + 1,0} \\ &= 0,7116 \\ &= 71,16\% \end{aligned}$$

Dengan rangkuman pada **Tabel 10**.

Tabel 10. Hasil Uji Data Test

Metrik	Nilai
Akurasi	98%
Presisi	55,26%
Recall	100%
F1-Score	71,16%

Sumber: Hasil Pengujian, 2025.

Penerapan algoritma *Random Forest* dalam penelitian ini menunjukkan kemampuan yang cukup baik dalam menganalisis dan memprediksi tingkat kejahatan. Model ini berhasil mencapai akurasi sebesar 98%, dengan *Recall* sempurna sebesar 100%. Artinya, hampir seluruh kasus kejahatan yang tergolong rawan berhasil dikenali. Namun, nilai *Precision* yang

masih berada di angka 55,26% menandakan bahwa model cukup sering memberikan peringatan pada wilayah yang sebenarnya tidak tergolong rawan. Meskipun begitu, *F1-Score* sebesar 71,16% menunjukkan bahwa secara keseluruhan model sudah bekerja cukup seimbang dalam mengenali dan memprediksi kasus kejahatan.

D. Pengaplikasian QGIS

Untuk mendukung interpretasi hasil, data prediksi divisualisasikan dalam bentuk peta menggunakan QGIS (Kepolisian Negara Republik Indonesia Daerah Jawa Timur Resort Jember, 2024). Peta ini menunjukkan klasifikasi kerawanan kejahatan berdasarkan lokasi, sebelum mengaplikasikan dengan aplikasi QGIS data hasil prediksi wajib berisi koordinat masing-masing daerah yang akan ditampilkan. Dengan ketentuan kolom yang mencakup:

- Tanggal kejadian
- Lokasi kejadian
- Jumlah kejadian
- Hasil prediksi (rawan/tidak rawan)
- Koordinat lokasi (*Latitude* dan *Longitude*).

Langkah-langkah visualisasi hasil prediksi dalam QGIS adalah sebagai berikut:

- Buka QGIS.
- Pilih menu Layer > Add Layer > Add Delimited Text Layer.
- Pilih file .csv, kemudian tetapkan kolom latitude dan longitude sebagai referensi koordinat (Reyvvan Maulid, 2024).
- Menambahkan peta dasar (base map) agar visual lebih informatif.
- Gunakan plugin seperti OpenStreetMap untuk latar belakang wilayah Kabupaten Jember.
- Mengatur simbolisasi berdasarkan klasifikasi dengan memberi warna berbeda untuk kategori hasil prediksi, seperti merah untuk kategori rawan dan hijau untuk kategori tidak rawan.
- Ukuran titik dapat disesuaikan berdasarkan jumlah kejadian (qgis.org, 2025).

Dengan hasil visualisasi pada **Gambar 1**.



Gambar 1 Visualisasi QGIS
Sumber: Hasil Pengujian, 2025.

Kecamatan seperti Summersari, Patrang, Kaliwates, dan Panti muncul sebagai wilayah dengan tingkat kerawanan tinggi. Visualisasi ini sangat bermanfaat bagi aparat keamanan untuk fokus pengawasan di wilayah prioritas. eterangan:

- Letak Polres Jember
- Letak Polsek Terdekat
- Titik Rawan Kejahatan
- Titik Tidak Rawan Kejahatan

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma *Random Forest* dalam menganalisis dan memprediksi pola kejahatan di Kota Jember, dapat disimpulkan beberapa hal penting yaitu penerapan SMOTE sangat membantu dalam mengatasi ketidakseimbangan kelas dan meningkatkan sensitivitas model terhadap kelas minoritas (rawan kejahatan). Serta visualisasi prediksi kerawanan menggunakan QGIS berhasil menunjukkan daerah-daerah rawan kejahatan secara geografis sehingga dapat membantu pihak yang berwenang dalam melaksanakan upaya preventif untuk mengekan angka kejahatan di kabupaten Jember. Dan model ini berhasil mencapai akurasi sebesar 98%, dengan *Recall* sempurna sebesar 100%. Artinya, hampir seluruh kasus kejahatan yang tergolong rawan berhasil dikenali. Namun, nilai *Precision* yang masih berada di angka 55,26% menandakan bahwa model cukup sering memberikan peringatan pada wilayah yang sebenarnya tidak tergolong rawan. Meskipun begitu, skor F1 sebesar 71,16% menunjukkan bahwa secara keseluruhan model sudah bekerja

cukup seimbang dalam mengenali dan memprediksi kasus kejahatan.

B. Saran

Adapun beberapa saran yang dapat disampaikan dari hasil penelitian ini antara lain:

1. Penelitian ini masih memiliki ruang untuk dikembangkan, seperti pembuatan sistem aplikasi untuk mempermudah pengguna, terutama dalam hal penambahan jumlah data, penyempurnaan *Preprocessing*, serta eksplorasi algoritma lain seperti *Gradient Boosting* atau *XGBoost*. Diharapkan penelitian selanjutnya dapat membandingkan lebih banyak model pembelajaran mesin untuk memperoleh performa yang lebih optimal.
2. Dataset dilengkapi dengan fitur tambahan seperti jumlah penduduk, kondisi cuaca, atau tingkat pendidikan sebagai faktor pendukung dalam prediksi.
3. Kendala yang dihadapi dalam penelitian ini adalah keterbatasan data spasial yang akurat dan minimnya informasi koordinat yang presisi dari laporan kejadian. Harapannya, untuk penelitian selanjutnya data dapat diperoleh secara lebih lengkap dan real-time dari instansi terkait.
4. Ketersediaan data yang lebih terstruktur, konsisten, dan terbuka akan sangat mendukung pengembangan model analitik dan prediksi yang lebih baik. Oleh karena itu, kolaborasi antara lembaga pemerintah, kepolisian, dan akademisi perlu terus didorong untuk membangun ekosistem data terbuka yang mendukung inovasi di bidang keamanan publik.

C. DAFTAR PUSTAKA

Alma Zuhairah. (2022). Penerapan Algoritma *Random Forest*, *Support Vector Machine* (SVM) dan *Gradient Boosted Tree*(GBT) untuk deteksi penipuan (*Fraud Detection*) padatransaksikartu kredit.

- Badan Pusat Statistik Kabupaten Jember. (2023). *Laporan Kriminalitas Kabupaten Jember*
- Cahyo, A. E., & Nilogiri, A. (2021). Klasifikasi Gangguan Autisme Pada Anak Menggunakan Algoritma C4.5 Dengan Teknik *Random Forest*.
- Kepolisian Negara Republik Indonesia Daerah Jawa Timur Resort Jember. (2024, September). POLRES JEMBER. <https://Jember.Jatim.Polri.Go.Id/>.
- Khairunnisa, A. (2023). Perbandingan Model *Random Forest* Dan Xgboost Untuk Prediksi Kejahatan Kesusilaan Di Provinsi Jawa Barat. JIKO (Jurnal Informatika Dan Komputer), 7(2), 202. <https://doi.org/10.26798/jiko.v7i2.799>
- Khatun, R., Ayon, S. I., Hossain, R., & Alam, J. (2021). *Data mining technique to analyse and predict crime using crime categories and arrest records. Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 1052–1060. <https://doi.org/10.11591/ijeecs.v22.i2.pp1052-1060>
- Moerdyanto, O. P., Kadek, I., & Nuryana, D. (2023). Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma *Decision Tree*. *Journal of Informatics and Computer Science*, 05.
- qgis.org. (2025, June 20). QGIS. [Www.Qgis.Org](http://www.Qgis.Org).
- Reyvan Maulid. (2024, August 21). Pentingnya File CSV dalam *Python Data Science*. Belajar *Data Science* Di Rumah.
- Sari, L. (2024). Optimasi Algoritma *Machine Learning* untuk Prediksi Kejahatan Berbasis Temporal.
- shodiq, nur, & papilaya, samuel. (2019). Analisis Pola Spasial Kriminalitas Wilayah Hukum Kepolisian Resort Kota Salatiga.
- Surya Ningsih, K., Jamilah Aruan, N., & Taufik Al Afkari Siahaan, A. (2022). Yayasan Insan Cipta Medan Aplikasi Buku Tamu Menggunakan Fitur Kamera Dan Ajax Berbasis Website Pada Kantor Dispora Kota Medan.
- Trivusi. (2022, September 17). Algoritma *Random Forest*: Pengertian dan Kegunaannya. <https://Www.Trivusi.Web.Id/2022/08/Algoritma-Random-Forest.Html>.
- Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'ban, R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran *Machine Learning*.