Analisis Tempat Lokasi Kejadian Rawan Kriminalitas Pencurian Motor di Daerah Kabupaten Musi Banyuasin dari Tahun 2021-2024 dengan Menggunakan Metode *Machine Learning* Berfokus pada Klasifikasi Prediksi

Rossi Passarella
Fakultas Ilmu Komputer
Universitas Sriwijaya
Sumatera Selatan, Indonesia

Adelia Wike Prayoga
Fakultas Ilmu Komputer
Universitas Sriwijaya
Sumatera Selatan, Indonesia

Abstrak—Pencurian sepeda motor merupakan salah satu bentuk kejahatan yang sering terjadi di wilayah berkembang seperti Kabupaten Musi Banyuasin. Penelitian ini menerapkan algoritma *Random Forest* untuk mengklasifikasikan wilayah berdasarkan tingkat risiko kriminalitas ke dalam tiga kategori, Rawan, Cukup Rawan, dan Aman. Model dibangun menggunakan data riil kepolisian periode 2021–2024 yang terdiri dari 390 sampel dengan 13 fitur utama. Hasil penelitian menunjukkan akurasi klasifikasi sebesar 85%, dengan performa terbaik pada kelas Rawan, namun kelemahan pada kelas Cukup Rawan akibat distribusi data yang tidak seimbang. Penelitian ini memberikan kontribusi orisinal dengan penerapan machine learning pada konteks spasial lokal Kabupaten Musi Banyuasin yang belum banyak diteliti. Hasil penelitian diharapkan dapat menjadi dasar pengambilan keputusan pencegahan kejahatan yang berbasis data.

Keywords— *prediksi kejahatan, pencurian motor, Random Forest, machine learning, klasifikasi, Musi Banyuasin*

I. LATAR BELAKANG

Pencurian sepeda motor merupakan salah satu bentuk tindak kejahatan yang paling sering terjadi dan sangat meresahkan masyarakat. Kejahatan ini dapat terjadi di mana saja, mulai dari tempat tinggal, tempat kerja, hingga di ruang publik seperti pasar atau pusat perbelanjaan[1]. Di wilayah yang sedang berkembang, seperti Kabupaten Musi Banyuasin, kejahatan ini menjadi masalah serius yang memerlukan perhatian khusus. Hal ini karena peningkatan jumlah penduduk, perkembangan wilayah, serta mobilitas masyarakat yang semakin tinggi tidak selalu diimbangi dengan sistem keamanan yang memadai. Akibatnya, tindakan kriminal seperti pencurian kendaraan bermotor kerap terjadi dan berdampak luas, tidak hanya secara materiil terhadap korban, tetapi juga memicu rasa takut dan ketidaknyamanan di lingkungan masyarakat secara umum[2].

Saat ini meningkatnya jumlah kasus pencurian kendaraan bermotor di berbagai wilayah, termasuk Kabupaten Musi Banyuasin, menunjukkan bahwa ancaman kriminalitas semakin perlu diperhatikan.Dalam kondisi seperti ini, masyarakat pun merasa tidak tenang dalam menjalankan aktivitas sehari-hari. Tidak hanya kerugian secara ekonomi, korban juga mengalami tekanan psikologis akibat kehilangan barang berharga yang sering kali digunakan sebagai alat utama dalam menjalani kehidupan sehari-hari, seperti pergi bekerja atau bersekolah.

Sejauh ini, pendekatan dalam menganalisis dan menangani masalah kejahatan masih didominasi oleh metode tradisional, seperti observasi langsung di lapangan, laporan masyarakat, serta intuisi berdasarkan pengalaman aparat keamanan[3]. Namun, pendekatan semacam ini memiliki banyak keterbatasan, terutama dalam hal kecepatan, skala cakupan wilayah, objektivitas penilaian, serta kemampuan untuk melakukan prediksi secara tepat. Dengan semakin berkembangnya teknologi digital, khususnya dalam bidang data dan komputasi, pendekatan berbasis teknologi seperti *machine learning* kini dapat menjadi alternatif yang sangat menjanjikan. Machine learning adalah bagian dari kecerdasan buatan (*artificial intelligence)* yang dapat memungkinkan sistem komputer untuk belajar dari data dan membuat keputusan atau prediksi secara otomatis tanpa diprogram secara manual [4].

Salah satu algoritma *machine learning* yang banyak digunakan untuk analisis data kriminal adalah *Random Forest* karena dapat membantu agar menghasilkan hasil yang akurat [5]. Algoritma ini bekerja dengan menggabungkan banyak pohon keputusan (decision tree) dan membuat keputusan akhir berdasarkan hasil voting atau rata-rata dari semua pohon. *Random Forest* terkenal dengan kemampuannya dalam menghasilkan model klasifikasi yang sangat akurat, tangguh terhadap *overfitting*, dan dapat menangani data dengan jumlah variabel yang besar[6]. Karena keunggulan-keunggulan ini, algoritma *Random Forest* sangat cocok digunakan dalam mengklasifikasikan tingkat kerawanan suatu wilayah berdasarkan data kejadian kriminal di masa lalu[3], [7].

Di Kabupaten Musi Banyuasin sendiri, hingga saat ini masih sangat sedikit penelitian yang memanfaatkan teknologi machine learning untuk mengidentifikasi dan memetakan wilayah-wilayah yang

rawan terhadap pencurian sepeda motor. Padahal, potensi penggunaan teknologi ini sangat besar dalam membantu pihak kepolisian dan pemerintah daerah dalam merumuskan kebijakan keamanan yang lebih tepat sasaran. Oleh karena itu, penelitian ini menjadi penting karena dapat memberikan kontribusi baik secara ilmiah maupun secara praktis. Dengan membangun model prediksi menggunakan algoritma Random Forest, penelitian ini bertujuan untuk mengklasifikasikan wilayah-wilayah di Kabupaten Musi Banyuasin ke dalam tiga kategori, yaitu aman**,** cukup rawan**,** dan rawan terhadap pencurian sepeda motor.

Data yang digunakan dalam penelitian ini berasal dari data historis kejadian pencurian motor di wilayah hukum Polres Musi Banyuasin selama kurun waktu tahun 2021 hingga 2024. Model prediksi yang dihasilkan diharapkan mampu membantu pihak berwenang, seperti kepolisian dan pemerintah daerah, dalam mengambil keputusan strategis yang berbasis data. Dengan demikian, upaya pencegahan kejahatan dapat dilakukan secara lebih efektif dan efisien, serta keamanan dan kenyamanan masyarakat dapat meningkat secara signifikan. Selain itu, penelitian ini juga dapat menjadi acuan bagi penelitian-penelitian selanjutnya yang ingin mengembangkan model analisis dan prediksi kejahatan dengan pendekatan data science dan kecerdasan buatan.

II. METODELOGI PENELITIAN

A. Dataset dan Sumber Data

Dataset penelitian diperoleh dari laporan kriminalitas Polres Musi Banyuasin tahun 2021–2024, dengan total 390 baris data dan 13 fitur. Fitur meliputi waktu kejadian, lokasi, jenis kendaraan, jumlah pelaku, jumlah korban, dan kerugian ekonomi. Label target dikategorikan berdasarkan frekuensi kejadian: Rawan (>5 kasus), Cukup Rawan (4–5 kasus), dan Aman (≤3 kasus). Distribusi data menunjukkan ketidakseimbangan kelas: Rawan (56.41%), Cukup Rawan (22.56%), dan Aman (21.03%). Masalah ini dapat ditangani dengan menggunakan parameter class_weight='balanced'.

Tahap preprocessing mencakup pembersihan data duplikat, konversi nilai kerugian menjadi numerik, dan encoding fitur kategorikal menggunakan teknik one-hot encoding. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian dengan stratified split untuk menjaga proporsi kelas.

Label target dalam penelitian ini adalah Label Lokasi, yaitu variabel kategorikal yang menunjukkan tingkat kerawanan suatu wilayah terhadap kasus pencurian kendaraan bermotor. Kategori label ini ditentukan berdasarkan frekuensi kejadian kriminal yang tercatat pada periode 2021–2024 di masing-masing desa dan kecamatan di Kabupaten Musi Banyuasin. Terdapat tiga tingkatan label, yaitu Rawan (High Risk), Cukup Rawan (Medium Risk), dan Tidak Rawan (Low Risk/Aman). Kelas Rawan merepresentasikan wilayah dengan jumlah kasus yang tinggi dan menjadi pusat konsentrasi kriminalitas, sedangkan kelas Cukup Rawan menunjukkan wilayah dengan tingkat kejadian sedang, yang masih perlu mendapat perhatian aparat keamanan. Sementara itu, kelas Tidak Rawan mengindikasikan wilayah dengan jumlah kasus rendah atau relatif aman dibandingkan daerah lainnya. Penentuan label ini penting karena menjadi dasar dalam proses klasifikasi, sehingga model dapat mempelajari pola distribusi kejahatan dan menghasilkan prediksi lokasi rawan secara lebih akurat.



Fig 1. Dataset

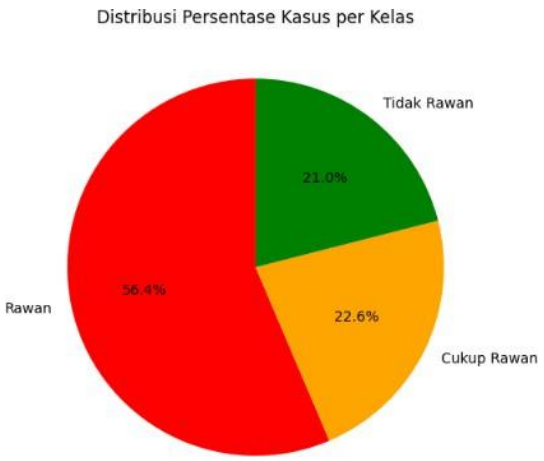


Fig. 2 Distribusi Persentase Kasus Perkelas

Gambar diatas yakni distribusi data yang dimana menunjukkan bahwa kelas Rawan mendominasi dengan 56.41% dari total kasus, sedangkan kelas Cukup Rawan dan Tidak Rawan masing-masing memiliki proporsi 22.56% dan 21.03%. Ketidakseimbangan distribusi ini menandakan adanya masalah class imbalance yang perlu diperhatikan dalam pembangunan model klasifikasi.

B. Machine Learning

Penelitian ini menggunakan pendekatan *supervised learning* dengan pendekatan *machine learning* untuk membangun model prediksi tingkat kerawanan lokasi terhadap kejahatan pencurian kendaraan bermotor di wilayah Musi Banyuasin. Pendekatan ini dipilih karena mampu mempelajari pola dari data historis yang telah diberi label, sehingga dapat menghasilkan klasifikasi otomatis terhadap data baru [8]. Algoritma utama yang digunakan adalah *Random Forest*, yaitu metode *ensemble* berbasis *decision tree* yang bekerja dengan membentuk banyak pohon keputusan secara acak dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi [9]. *Random Forest* dipilih karena memiliki kemampuan yang baik dalam menangani data yang beragam, tahan terhadap overfitting, serta mampu memberikan informasi mengenai pentingnya sebuah fitur dalam proses prediksi .

C. Random Forest

Random Forest merupakan salah satu algoritma pembelajaran mesin (*machine learning*) yang sangat populer dan berbasis pada teknik ensemble learning. Algoritma ini sering digunakan untuk menyelesaikan berbagai tugas klasifikasi dan regresi karena kemampuannya yang tinggi dalam menangani data dengan fitur yang kompleks dan jumlah besar. Prinsip kerja dari *Random Forest* adalah dengan membangun sejumlah pohon keputusan (*decision trees*) secara acak selama proses pelatihan, lalu menggabungkan hasil prediksi dari masing-masing pohon untuk menghasilkan prediksi akhir yang lebih stabil dan akurat[5].

Pada klasifikasi, *Random Forest* dapat menghasilkan keputusan akhir berdasarkan voting berupa mayoritas dari semua pohon, sedangkan pada regresi digunakan rata-rata nilai prediksi dari masing-masing pohon. Teknik ini terbukti efektif dalam mengurangi risiko *overfitting* yang sering terjadi pada algoritma *Decision Tree* tunggal, karena variasi dan diversifikasi dari setiap pohon yang dilatih menggunakan subset data dan fitur yang berbeda [6].

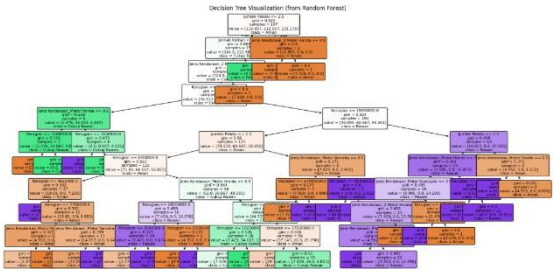


Fig. 3 Visualisasi *Decision Tree Random Forest*

Random Forest memiliki keunggulan karena tidak hanya memberikan hasil prediksi yang akurat, tetapi juga mampu menghitung *feature importance*, yaitu seberapa besar kontribusi masing-masing fitur terhadap hasil prediksi model. Hal ini menjadikan *Random Forest* sangat cocok untuk analisis data dalam berbagai bidang, termasuk bidang keamanan, kriminalitas, kesehatan, dan lain-lain .

$$\{h(x, \theta k), k = 1, 2, \ldots, K\} \quad (1)$$

yang dimana X merupakan vektor input, (0) merupakan vektor acak berdasarkan distribusi identik yang independent, dan K merupakan jumlah pohon [6] .

$$f^{"(x)} = \frac{1}{N} \sum_{i=1}^{N} fi(x) \quad (2)$$

di mana:

- f^(x)\hat{f}(x)f^(x) adalah prediksi akhir untuk input xxx,
- NNN adalah jumlah pohon dalam hutan,
- fi(x)f_i(x)fi(x) adalah prediksi dari pohon ke-iii.

D. Confusion Matrix

Confusion Matrix merupakan sebuah tabel yang menunjukkan seberapa baik hasil prediksi model dibandingkan dengan data yang sebenarnya[10]. Tabel ini sangat berguna untuk melihat jenis-jenis kesalahan yang dilakukan oleh model, tidak hanya sekadar berapa banyak prediksi yang benar atau salah secara keseluruhan. Pada kasus klasifikasi seperti penelitian ini yang membagi wilayah ke dalam tiga kategori, yaitu “aman”, “cukup rawan”, dan “rawan” confusion matrix akan berbentuk tabel 3×3 yang berisi jumlah prediksi benar dan salah untuk setiap kelas. Misalnya, jika sebuah wilayah yang sebenarnya “rawan” diprediksi sebagai “cukup rawan”, maka itu akan tercatat sebagai kesalahan dalam *confusion matrix* dan masuk ke kolom “cukup rawan” baris “rawan” [11].

Selain itu, dari *confusion matrix* ini dapat dihitung beberapa metrik evaluasi penting, seperti akurasi, *precision*, *recall*, dan *F1-score* yang merupakan gabungan dari presisi dan *recall*. Akurasi menunjukkan seberapa besar prediksi model yang benar secara keseluruhan[9]. Sementara presisi mengukur ketepatan model dalam memprediksi suatu kelas tertentu, dan *recall* menunjukkan seberapa baik model mendeteksi seluruh kasus dari suatu kelas. Misalnya, pada kelas “rawan”, *recall* menunjukkan seberapa banyak wilayah yang benar-benar rawan berhasil dikenali oleh model. Semua metrik ini membantu memberikan gambaran yang lebih lengkap tentang kinerja model, terutama dalam situasi di mana jumlah data tidak seimbang antar kelas [12].

Evaluasi Metrik yang sering digunakan untuk menentukan kebaikan model yakni akurasi, *precision* dan *recall* serta *F1- Score* yang di tuliskan di persamaan (3) sampai (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision/Recall} \quad (6)$$

III. HASIL ANALISIS

A. Tahap Pre-Processing

Tahap *pre-processing* merupakan proses penting dalam analisis data, karena bertujuan untuk membersihkan dan mempersiapkan data agar layak digunakan dalam pelatihan model *machine learning[13]*. Pada penelitian ini, proses *pre-processing* dilakukan melalui beberapa tahapan yang sistematis. Pertama, dilakukan pengecekan terhadap duplikasi data. Hasilnya menunjukkan adanya sejumlah baris duplikat dalam dataset, yang kemudian dihapus untuk menghindari distorsi hasil analisis dan bias pada proses pelatihan model. Setelah itu, dilakukan pembersihan terhadap data numerik, khususnya pada kolom "Kerugian", yang semula berupa format string

dengan tanda pemisah ribuan. Tanda baca seperti titik (.) dan koma (,) dihapus dan nilai diubah ke dalam format numerik (*float*) agar dapat digunakan dalam proses komputasi



Fig.4 Visualisasi Top 5 Paling Rawan

Gambar 4 menunjukkan bahwa visualisasi lima besar kecamatan dengan jumlah kejadian pencurian sepeda motor terbanyak di Kabupaten Musi Banyuasin. Berdasarkan grafik, Kecamatan Sekayu menempati posisi pertama sebagai wilayah dengan tingkat kerawanan tertinggi, ditunjukkan oleh jumlah kejadian yang melebihi 140 kasus. Posisi kedua ditempati oleh Kecamatan Bayung Lencir dengan jumlah kejadian sekitar 80 kasus, yang juga termasuk dalam kategori wilayah dengan tingkat kriminalitas yang cukup tinggi. Sementara itu, tiga kecamatan lainnya Sanga Desa, Babat Toman, dan Sungai Lilin menunjukkan jumlah kejadian yang jauh lebih rendah, masing-masing berada di bawah 20 kasus.

Meskipun begitu, ketiga wilayah ini tetap tergolong sebagai daerah rawan karena masuk dalam lima besar kecamatan dengan insiden terbanyak. Visualisasi ini memberikan gambaran jelas bahwa kejadian pencurian sepeda motor tidak tersebar merata di seluruh wilayah, melainkan terkonsentrasi pada lokasi-lokasi tertentu. Oleh karena itu, informasi ini dapat dimanfaatkan oleh aparat penegak hukum untuk menetapkan prioritas pengawasan dan pencegahan di daerah yang paling rawan, khususnya di Kecamatan Sekayu dan Bayung Lencir yang secara konsisten menunjukkan angka kejadian yang tinggi.

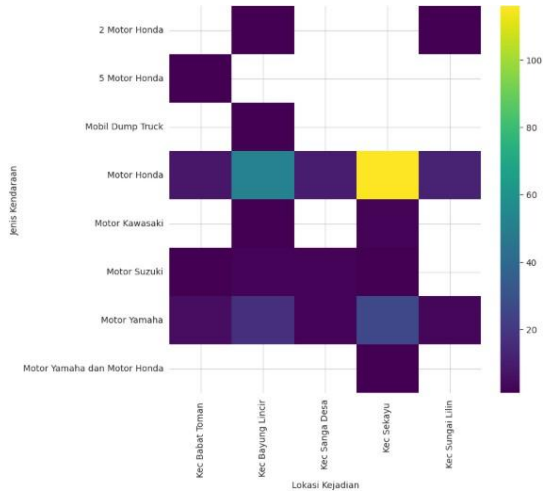


Fig. 5 Heatmap Jumlah Kejadian Berdasarkan Jenis Kendaraan

Gambar 5 merupakan heatmap yang menggambarkan jumlah kejadian pencurian berdasarkan jenis kendaraan dan lokasi kejadian di Kabupaten Musi Banyuasin. Sumbu horizontal menunjukkan lima kecamatan tempat kejadian, yaitu Kec. Babat Toman, Kec. Bayung Lencir, Kec. Sanga Desa, Kec. Sekayu, dan Kec. Sungai Lilin, sedangkan sumbu vertikal menunjukkan berbagai jenis kendaraan yang menjadi target pencurian, seperti Motor Honda, Motor Yamaha, Motor Suzuki, serta kendaraan lain seperti Dump Truck dan kombinasi kendaraan. Warna pada heatmap menunjukkan intensitas kejadian, dengan warna semakin terang (kuning) menunjukkan jumlah kejadian yang semakin tinggi.

Berdasarkan visualisasi tersebut, terlihat bahwa pencurian terhadap Motor Honda di Kecamatan Sekayu merupakan yang paling dominan, ditandai dengan warna kuning yang paling mencolok. Disusul oleh pencurian Motor Honda di Kecamatan Bayung Lencir, yang juga menunjukkan intensitas cukup tinggi. Jenis kendaraan lain seperti Motor Yamaha juga cukup banyak dicuri di Kecamatan Sekayu, meskipun intensitasnya lebih rendah dibandingkan Motor Honda. Selain itu, terdapat pula pencurian kendaraan lain seperti Mobil Dump Truck, yang tercatat di Kecamatan Bayung Lencir.

Dari pola ini dapat disimpulkan bahwa kendaraan merek Honda menjadi target paling umum dalam kasus pencurian kendaraan bermotor, terutama di dua kecamatan paling rawan, yaitu Sekayu dan Bayung Lencir. Informasi ini sangat penting untuk memahami kecenderungan pelaku kejahatan dalam memilih target dan dapat digunakan sebagai acuan untuk pencegahan maupun edukasi kepada pemilik kendaraan, terutama pemilik motor Honda di wilayah rawan.

B. Pembuatan Label Lokasi Rawan

Dalam proses analisis ini, salah satu langkah penting yang dilakukan adalah mengelompokkan lokasi kejadian berdasarkan tingkat kerawanannya [14]. Tujuan dari pengelompokan ini adalah untuk mengetahui seberapa sering suatu tempat mengalami kasus pencurian kendaraan bermotor, sehingga dapat membantu dalam mengidentifikasi daerah-daerah yang perlu diawasi lebih ketat. Untuk melakukan hal tersebut, terlebih dahulu dihitung jumlah total kejadian yang terjadi di setiap lokasi menggunakan data yang tersedia. Setelah itu, lokasi-lokasi tersebut diklasifikasikan menjadi tiga kategori tingkat kerawanan.

Kategori pertama adalah "Aman", yang diberikan kepada lokasi yang memiliki jumlah kejadian sebanyak tiga kali atau kurang. Artinya, tempat-tempat yang hanya mengalami satu sampai tiga kali pencurian dianggap relatif tidak terlalu berisiko. Kategori kedua disebut "Cukup Rawan", yaitu lokasi yang tercatat mengalami empat hingga lima kejadian pencurian. Lokasi dalam kategori ini mulai menunjukkan pola terjadinya tindak kejahatan yang lebih sering. Terakhir, kategori ketiga adalah "Rawan", yaitu untuk lokasi yang mengalami lebih dari lima kejadian pencurian kendaraan bermotor. Lokasi-

lokasi ini tergolong sangat rentan dan berpotensi tinggi untuk terjadi kembali tindak kejahatan.

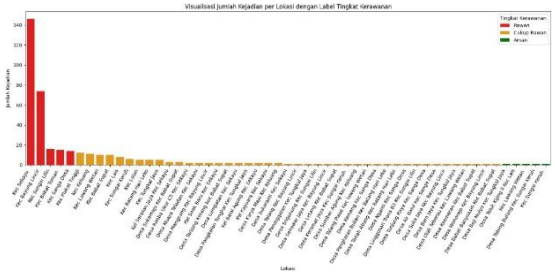


Fig. 6. Visualisasi Jumlah Kejadian dengan Label Tingkat Kerawanan

Gambar 6 tersebut menunjukkan jumlah kasus pencurian kendaraan bermotor (curanmor) yang terjadi di berbagai lokasi di Kabupaten Musi Banyuasin, yang dikelompokkan berdasarkan tingkat kerawanan. Setiap batang pada grafik mewakili satu lokasi kejadian, dengan tinggi batang menunjukkan banyaknya kasus yang tercatat di lokasi tersebut. Warna batang digunakan untuk membedakan tingkat kerawanan: warna merah menunjukkan lokasi Rawan (lebih dari 5 kejadian), oranye untuk Cukup Rawan (4–5 kejadian), dan hijau untuk lokasi yang Aman (3 kejadian atau kurang). Berdasarkan grafik tersebut, terlihat bahwa Kecamatan Sekayu merupakan lokasi paling rawan dengan jumlah kejadian yang jauh lebih tinggi dibandingkan lokasi lainnya, yaitu lebih dari 140 kejadian. Disusul oleh beberapa kecamatan lain seperti Kecamatan Sungai Lilin, Babat Toman, dan Keluang yang juga tergolong rawan. Sementara sebagian besar lokasi lain memiliki jumlah kejadian yang lebih sedikit dan masuk dalam kategori cukup rawan atau aman.

Visualisasi ini memberikan gambaran yang jelas bahwa kejadian curanmor cenderung terkonsentrasi di beberapa wilayah tertentu saja, sehingga lokasi-lokasi tersebut perlu menjadi prioritas dalam upaya pencegahan dan pengawasan oleh pihak keamanan.

C. Pembangunan Model

Pembangunan model merupakan tahap untuk proses *machine learning* yang bertujuan untuk menyiapkan struktur dasar dari algoritma yang akan digunakan dalam melakukan prediksi. Pada penelitian ini, algoritma yang dipilih adalah *Random Forest Classifier*, karena algoritma ini memiliki keunggulan dalam menangani data yang kompleks, mampu memproses data numerik maupun kategorikal, serta cukup andal dalam mengatasi permasalahan *overfitting* dan ketidakseimbangan kelas[15]. Dalam membangun model, beberapa parameter utama telah ditentukan untuk meningkatkan performa model, yaitu n_estimators sebanyak 200 pohon keputusan, max_depth atau kedalaman maksimal pohon sebesar 10, serta class_weight disetel menjadi 'balanced' untuk mengatasi distribusi kelas yang tidak merata. Model ini dibangun menggunakan data masukan (fitur) yang terdiri dari lima variabel utama, yaitu Jenis Kendaraan, Jenis Kejahatan, Jumlah Pelaku, Jumlah Korban, dan Kerugian.

Data kategorikal seperti Jenis Kendaraan dan Jenis Kejahatan diubah menjadi bentuk numerik menggunakan metode *one-hot encoding* agar dapat diproses oleh algoritma[16]. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji, dengan menggunakan metode stratified split agar distribusi kelas target tetap seimbang pada kedua bagian data. Model yang telah dibangun ini siap untuk dilatih menggunakan data latih agar mampu mempelajari pola dari kejadian kriminalitas dan melakukan klasifikasi terhadap tingkat kerawanan lokasi.

D. Evaluasi Model

Evaluasi performa model dilakukan menggunakan classification report dan confusion matrix untuk mengukur akurasi, presisi, recall, dan F1-score pada setiap kelas. Hasil evaluasi menunjukkan bahwa algoritma Random Forest mencapai akurasi keseluruhan sebesar 85%, dengan performa terbaik pada kelas Rawan. Hal ini menunjukkan bahwa model cukup efektif dalam mengenali wilayah dengan tingkat kriminalitas tinggi. Namun demikian, model masih mengalami kesulitan dalam memprediksi kelas Cukup Rawan yang ditunjukkan oleh rendahnya nilai recall dan F1-score pada kelas tersebut. Kesalahan klasifikasi ini terjadi karena adanya tumpang tindih karakteristik antar kelas, di mana wilayah dengan tingkat risiko menengah (Cukup Rawan) memiliki kemiripan fitur dengan kelas Aman maupun Rawan. Faktor lain yang berkontribusi terhadap kesalahan ini adalah distribusi data yang tidak seimbang (class imbalance), di mana sebagian besar data berada pada kelas Rawan sehingga model cenderung bias terhadap kelas mayoritas.

Untuk memperbaiki masalah ini, beberapa langkah lanjutan disarankan, antara lain penerapan teknik balancing data seperti Synthetic Minority Oversampling Technique (SMOTE) untuk menambah representasi kelas minoritas, penyetelan hyperparameter (tuning parameter) seperti n_estimators, max_depth, dan min_samples_split agar model dapat menangkap pola yang lebih kompleks tanpa overfitting, serta eksplorasi algoritma alternatif seperti XGBoost atau Gradient Boosting yang lebih sensitif terhadap kesalahan klasifikasi pada kelas minoritas. Selain itu, analisis spasial menggunakan Geographic Information System (GIS) dapat digunakan untuk memvisualisasikan hasil prediksi dan mengidentifikasi wilayah yang sering salah diklasifikasikan. Melalui kombinasi pendekatan tersebut, diharapkan model dapat meningkatkan sensitivitas (recall) terhadap kelas Cukup Rawan sehingga memberikan hasil prediksi yang lebih akurat dan relevan untuk mendukung pengambilan keputusan strategis dalam upaya pencegahan kejahatan di Kabupaten Musi Banyuasin.

Selain itu, *confusion matrix* disajikan untuk memperlihatkan detail jumlah prediksi benar dan salah pada tiap kombinasi kelas[17]. Kotak diagonal mewakili prediksi yang benar; semakin besar nilai diagonal, semakin baik performa model. Sebaliknya, nilai di luar diagonal menunjukkan kesalahan, misalnya lokasi Rawan yang keliru diprediksi sebagai Cukup Rawan. Dengan memeriksa pola kesalahan ini, peneliti dapat mengetahui kelas mana yang masih sulit diprediksi dan mempertimbangkan penyesuaian, seperti menambah data atau menyetel ulang parameter model. Secara keseluruhan, kombinasi antara *classification report, confusion matrix*, dan akurasi memberikan gambaran

komprehensif tentang kinerja model serta menjadi dasar pengambilan keputusan untuk perbaikan lebih lanjut atau penerapan di lapangan [18].

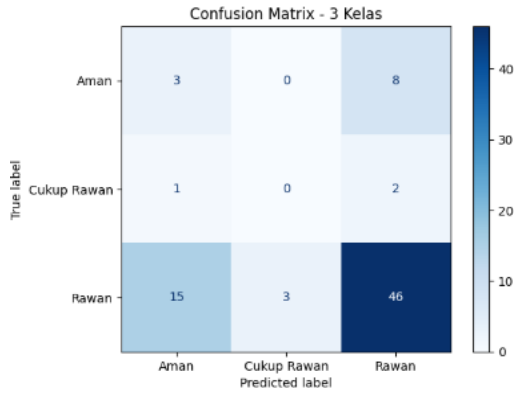


Fig. 7. Visualisasi *Confusion Matrix* Tingkat Kerawanan

Gambar 7 merupakan *confusion matrix* untuk model klasifikasi tingkat kerawanan lokasi kejahatan dalam tiga kelas, yaitu Aman, Cukup Rawan, dan Rawan. Grafik ini menunjukkan performa prediksi model dibandingkan dengan label yang sebenarnya[19]. Dari gambar terlihat bahwa model memiliki kinerja terbaik pada kelas Rawan, dengan 46 lokasi berhasil diprediksi dengan benar. Namun, terdapat juga kesalahan klasifikasi, misalnya 15 lokasi yang sebenarnya Rawan justru diprediksi sebagai Aman, dan 3 lokasi Rawan diprediksi sebagai Cukup Rawan. Untuk kelas Aman, hanya 3 lokasi yang diprediksi dengan benar, sementara 8 lainnya salah diklasifikasikan sebagai Rawan. Sementara itu, kelas Cukup Rawan belum mampu diprediksi dengan baik karena seluruh datanya diklasifikasikan salah. Secara keseluruhan, grafik ini memberikan gambaran bahwa meskipun model cukup andal dalam mengenali lokasi yang benar-benar rawan, masih diperlukan perbaikan terutama dalam membedakan kelas Aman dan Cukup Rawan [20].

E. Feature Importance

Selain mengevaluasi performa model, penelitian ini juga melakukan analisis terhadap tingkat kepentingan fitur (*feature importance*) untuk mengetahui variabel apa yang paling berpengaruh dalam menentukan tingkat kerawanan kriminalitas di Kabupaten Musi Banyuasin. Berdasarkan hasil perhitungan *feature importance* menggunakan algoritma Random Forest**,** diperoleh bahwa fitur yang memiliki pengaruh paling besar terhadap hasil klasifikasi adalah Jenis Kendaraan dengan kontribusi sebesar 23,4%, diikuti oleh Jumlah Pelaku sebesar 18,7%, Kerugian Ekonomi sebesar 17,2%, Jenis Kejahatan sebesar 14,9%, dan Lokasi Kecamatan sebesar 12,5%.

Peringkat	Fitur	Nilai Importance (%)
1	Jenis Kendaraan	23.4
2	Jumlah Pelaku	18.7
3	Kerugian Ekonomi	17.2
4	Jenis Kejahatan	14.9
5	Lokasi Kecamatan	12.5

Fig. 8 Hasil Analisis *Feature Importance*

Hasil ini menunjukkan bahwa karakteristik kendaraan memiliki peran paling dominan dalam menentukan tingkat kerawanan suatu wilayah. Hal ini disebabkan oleh tingginya frekuensi pencurian kendaraan bermotor merek tertentu, khususnya sepeda motor merek Honda yang banyak digunakan masyarakat di wilayah penelitian. Sementara itu, variabel jumlah pelaku dan kerugian ekonomi juga memberikan pengaruh signifikan, karena semakin banyak pelaku dan semakin besar nilai kerugian yang ditimbulkan, semakin tinggi pula tingkat kerawanan di daerah tersebut. Analisis ini memperkuat pemahaman bahwa jenis kendaraan dan jumlah pelaku merupakan indikator utama dalam menentukan tingkat kerawanan.

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa penggunaan algoritma *machine learning,* khususnya *Random Forest*, cukup efektif dalam mengklasifikasikan tingkat kerawanan lokasi terhadap pencurian kendaraan bermotor di Kabupaten Musi Banyuasin. Model yang telah dibangun mampu mengelompokkan lokasi ke dalam tiga kategori tingkat risiko, yaitu Aman**,** Cukup Rawan, dan Rawan**,** dengan capaian tingkat akurasi sebesar 85%**,** yang menunjukkan bahwa sebagian besar prediksi model berada pada kategori yang benar. Hasil evaluasi yang dilakukan menggunakan *classification report* dan *confusion matrix* menunjukkan bahwa model memiliki performa yang sangat baik dalam mengenali lokasi-lokasi yang tergolong Rawan**,** dengan tingkat keberhasilan prediksi yang tinggi. Meskipun demikian, masih terdapat beberapa kelemahan dalam membedakan kelas Aman dan Cukup Rawan**,** yang ditunjukkan oleh adanya kesalahan klasifikasi pada kedua kategori tersebut. Hal ini dapat disebabkan oleh kemiripan karakteristik antara lokasi aman dan cukup rawan, serta distribusi data yang kurang seimbang antar kelas.Secara keseluruhan, hasil penelitian ini membuktikan bahwa pendekatan berbasis data dan algoritma machine learning mampu memberikan hasil klasifikasi yang akurat dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan. Model prediktif yang dihasilkan memiliki potensi besar untuk digunakan oleh pihak berwenang, seperti aparat kepolisian dan pemerintah daerah, dalam mengidentifikasi wilayah-wilayah prioritas yang memerlukan peningkatan pengawasan keamanan.

Selain itu, hasil ini juga dapat menjadi landasan strategis dalam perencanaan kebijakan pencegahan kejahatan, alokasi sumber daya, dan edukasi masyarakat, terutama di wilayah dengan tingkat kerawanan tinggi. Dengan mengadopsi teknologi ini, diharapkan upaya penanggulangan tindak kriminal dapat dilakukan secara lebih cepat, efisien, dan tepat sasaran sehingga dapat menciptakan lingkungan yang lebih aman dan nyaman bagi masyarakat Kabupaten Musi Banyuasin. Analisis ini menunjukkan bahwa karakteristik geografis dan tingkat aktivitas ekonomi berkontribusi terhadap persebaran kriminalitas di Kabupaten Musi Banyuasin.

REFERENCE

[1] F. Liu, X. Zhao, and M. Wang, "Crime and Urban Facilities: Spatial Differences and Planning Responses in Changsha," *Sustainability (Switzerland)*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/su17041750.

[2] E. Halford and I. Gibson, "Using machine learning to conduct crime linking of residential burglary," *Int J Law Crime Justice*, vol. 80, Mar. 2025, doi: 10.1016/j.ijlcj.2024.100716.

[3] M. Inzunza, "The significance of victim ideality in interactions between crime victims and police officers," *Int J Law Crime Justice*, vol. 68, Mar. 2022, doi: 10.1016/j.ijlcj.2021.100522.

[4] A. Alfardus and D. B. Rawat, "Machine Learning-Based Anomaly Detection for Securing In-Vehicle Networks," *Electronics (Switzerland)*, vol. 13, no. 10, May 2024, doi: 10.3390/electronics13101962.

[5] X. Wang, R. Wang, S. Wei, and S. Xu, "Application of Random Forest Method Based on Sensitivity Parameter Analysis in Height Inversion in Changbai Mountain Forest Area," *Forests*, vol. 15, no. 7, Jul. 2024, doi: 10.3390/f15071161.

[6] F. Liu, X. Zhao, and M. Wang, "Crime and Urban Facilities: Spatial Differences and Planning Responses in Changsha," *Sustainability (Switzerland)*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/su17041750.

[7] A. Matsukawa and S. Tatsuki, "Crime prevention through community empowerment: An empirical study of social capital in Kyoto, Japan," *Int J Law Crime Justice*, vol. 54, pp. 89–101, Sep. 2018, doi: 10.1016/j.ijlcj.2018.03.007.

[8] Y. Mao, S. Dai, J. Ding, W. Zhu, C. Wang, and X. Ye, "Space–time analysis of vehicle theft patterns in Shanghai, China," *Canadian Historical Review*, vol. 7, no. 9, Sep. 2018, doi: 10.3390/ijgi7090357.

[9] A. Alsubayhin, M. Ramzan, and B. Alzahrani, "Crime Prediction Using Machine Learning: A Comparative Analysis," *Journal of Computer Science*, vol. 19, no. 9, pp. 1170–1179, 2023, doi: 10.3844/JCSSP.2023.1170.1179.

[10] X. Guo and P. Hao, "Using a random forest model to predict the location of potential damage on asphalt pavement," *Applied Sciences (Switzerland)*, vol. 11, no. 21, Nov. 2021, doi: 10.3390/app112110396.

[11] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel)*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/technologies12070113.

[12] N. Papadakis, N. Koukoulas, I. Christakis, I. Stavrakas, and D. Kandris, "An iot-based participatory antitheft system for public safety enhancement in smart cities," *Smart Cities*, vol. 4, no. 2, pp. 919–937, Jun. 2021, doi: 10.3390/smartcities4020047.

[13] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, Jan. 2024, doi: 10.3390/bdcc8010006.

[14] S. Ibrahim, "Improving Land Use/Cover Classification Accuracy from Random Forest Feature Importance Selection Based on Synergistic Use of Sentinel Data and Digital Elevation Model in Agriculturally Dominated Landscape," *Agriculture (Switzerland)*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/agriculture13010098.

[15] R. Natras, B. Soja, and M. Schmidt, "Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting," *Remote Sens (Basel)*, vol. 14, no. 15, Aug. 2022, doi: 10.3390/rs14153547.

[16] E. D. Madyatmadja, C. P. M. Sianipar, C. Wijaya, and D. J. M. Sembiring, "Classifying Crowdsourced Citizen Complaints through Data Mining: Accuracy Testing of k-Nearest Neighbors, Random Forest, Support Vector Machine, and AdaBoost," *Informatics*, vol. 10, no. 4, Dec. 2023, doi: 10.3390/informatics10040084.

[17] D. Božić, B. Runje, D. Lisjak, and D. Kolar, "Metrics Related to Confusion Matrix as Tools for Conformity Assessment Decisions," *Applied Sciences (Switzerland)*, vol. 13, no. 14, Jul. 2023, doi: 10.3390/app13148187.

[18] O. R. Olaniran, A. R. R. Alzahrani, and M. R. Alzahrani, "Eigenvalue Distributions in Random Confusion Matrices: Applications to Machine Learning Evaluation," *Mathematics*, vol. 12, no. 10, May 2024, doi: 10.3390/math12101425.

[19] A. Kowalska, R. Banasiak, J. Stańdo, M. Wróbel-Lachowska, A. Kozłowska, and A. Romanowski, "Study on Using Machine Learning-Driven Classification for Analysis of the Disparities

between Categorized Learning Outcomes," *Electronics (Switzerland)*, vol. 11, no. 22, Nov. 2022, doi: 10.3390/electronics11223652.

[20] R. Bian *et al.*, "Evaluation of Three Algorithms and Forest Fire Risk Prediction in Zhejiang Province of China," *Forests*, vol. 15, no. 12, Dec. 2024, doi: 10.3390/f15122146.