

ANALISIS PREDIKSI TERHADAP PENINGKATAN TINDAK PIDANA DENGAN METODE NAIVE BAYES BERDASARKAN LAPORAN KRIMINALITAS

Budi Dwi Kurniawan, Rudi Heriansyah, Zaid Romegar Mair

Teknik Informatika, Universitas Indo Global Mandiri

Jalan Jend. Sudirman Km.4 No. 62, Kota Palembang, Indonesia

2022110026p@students.uigm.ac.id

ABSTRAK

Kriminalitas merupakan permasalahan sosial yang terus berkembang dan memerlukan analisis mendalam untuk memahami pola serta faktor yang memengaruhinya. Polres PALI Polda Sumsel menghadapi tantangan dalam mengidentifikasi tren kriminalitas guna merancang strategi pencegahan yang lebih efektif. Penelitian ini menerapkan metode *Naïve Bayes*, yang dikenal dalam klasifikasi berbasis probabilitas, untuk menganalisis serta memprediksi peningkatan tindak pidana. Data laporan kriminalitas dari Polres PALI digunakan untuk mengklasifikasikan angka kriminalitas berdasarkan faktor utama, seperti jumlah kasus dan tingkat pengangguran. Evaluasi model dilakukan menggunakan *classification report*, mencakup akurasi, *precision*, *recall*, dan *F1-score*. Hasil analisis menunjukkan bahwa wilayah dengan tingkat pengangguran tinggi cenderung mengalami peningkatan kriminalitas. Model *Naïve Bayes* terbukti efektif dalam mendeteksi pola kriminalitas, terutama dalam kategori pencurian berat, dengan akurasi mencapai 90%. Model ini memiliki *recall* 1.00 dalam mendeteksi peningkatan kriminalitas, tetapi kurang akurat dalam mengklasifikasikan penurunan angka kejahatan (*recall* 0.67). Metode ini dapat digunakan sebagai alat bantu dalam analisis prediksi angka kriminalitas, karena memiliki kecepatan dan efisiensi tinggi.

Kata kunci : Kriminalitas, Prediksi, *Naïve Bayes*, Polres PALI

1. PENDAHULUAN

Kriminalitas merupakan masalah sosial yang terus berkembang dan memberikan dampak luas pada berbagai aspek kehidupan masyarakat, baik dari sisi ekonomi, psikologis, maupun sosial. Fenomena kejahatan ini tidak hanya mempengaruhi korban langsung, tetapi juga menciptakan rasa ketidakamanan dalam masyarakat, yang dapat mengganggu kestabilan sosial dan perekonomian [1].

Di Indonesia, tingkat kejahatan yang cenderung fluktuatif menuntut aparat penegak hukum untuk lebih responsif dan proaktif dalam menganalisis faktor-faktor yang memengaruhi timbulnya tindak pidana. Polres PALI Polda Sumsel, sebagai salah satu institusi yang memiliki tugas dalam menjaga keamanan dan ketertiban di wilayah hukumnya, menghadapi tantangan besar dalam mengidentifikasi pola-pola kejahatan yang muncul. Perubahan sosial yang cepat, ditambah dengan kemajuan teknologi, turut memberikan dampak pada jenis-jenis kejahatan yang terjadi, dengan kejahatan siber menjadi salah satu fenomena yang semakin kompleks [2].

Dalam menghadapi masalah ini, penting bagi aparat penegak hukum untuk memiliki alat yang tepat guna mengidentifikasi dan menganalisis tren kriminalitas secara lebih efektif dan efisien. Dengan kemajuan teknologi, analisis berbasis data, khususnya yang menggunakan model prediktif seperti *Naive Bayes*, menjadi salah satu pendekatan yang relevan dan semakin sering digunakan dalam menilai dan memprediksi tren kriminalitas di wilayah tertentu [1], [3], [4].

Meskipun teknologi analisis berbasis data menawarkan potensi yang besar, tantangan utama yang dihadapi oleh Polres PALI Polda Sumsel adalah bagaimana mengklasifikasikan dan memprediksi tingkat peningkatan atau penurunan tindak pidana secara tepat dan efisien [5].

Salah satu jenis kejahatan yang membutuhkan perhatian serius adalah pencurian berat, yang memiliki frekuensi tinggi dan dampak yang cukup besar pada masyarakat. Keberhasilan dalam mengidentifikasi tren kriminalitas, khususnya dalam kategori pencurian berat, sangat bergantung pada kemampuan model analisis dalam memprediksi pola-pola yang ada dengan akurat. Sebuah model analisis yang dapat mengidentifikasi potensi terjadinya kejahatan dengan tingkat keakuratan tinggi akan sangat membantu aparat penegak hukum dalam merumuskan kebijakan keamanan yang tepat, serta mengalokasikan sumber daya secara lebih efisien. Di sisi lain, pemahaman yang lebih baik terhadap tren kriminalitas memungkinkan pihak berwenang untuk melakukan pencegahan lebih dini, yang pada akhirnya dapat menurunkan angka kejahatan dan meningkatkan rasa aman di kalangan masyarakat. Keakuratan dalam memprediksi tren kejahatan menjadi krusial, karena kebijakan yang tidak didasarkan pada analisis yang tepat berisiko tidak efektif atau bahkan dapat memperburuk situasi [6], [7].

Sejumlah penelitian telah dilakukan untuk memahami pola dan faktor yang memengaruhi tingkat kriminalitas di berbagai wilayah. Resiana & Aditya (2023) menganalisis tingkat kriminalitas selama pandemi COVID-19 di Daerah Istimewa Yogyakarta dan menemukan peningkatan kasus sebesar 15% dari tahun 2019 ke 2020 [4].

Studi lain oleh Ghiffari et al. (2021) menggunakan metode *Canonical Correspondence Analysis (CCA)* untuk memetakan pola tindak kriminal berdasarkan karakteristik wilayah [8]. Dari berbagai penelitian ini, penggunaan analisis data statistik terbukti mampu mengidentifikasi tren kriminalitas serta memberikan dasar yang kuat dalam perencanaan strategi pencegahan.

Penelitian ini menggunakan metode *Naïve Bayes*, yang merupakan salah satu teknik klasifikasi berbasis probabilitas yang efektif dalam mengelompokkan data kriminalitas [9].

Data laporan kriminalitas dari Polres PALI digunakan untuk mengklasifikasikan angka kejahatan berdasarkan faktor utama, seperti jumlah kasus dan tingkat pengangguran [10].

Evaluasi model dilakukan menggunakan *classification report*, yang mencakup metrik akurasi, *precision*, *recall*, dan *F1-score* guna mengukur efektivitas prediksi model dalam mendeteksi tren kriminalitas [11].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang dapat menganalisis peningkatan atau penurunan angka kriminalitas secara efektif, khususnya dalam kategori pencurian berat di wilayah Polres PALI Polda Sumsel. Selain itu, penelitian ini juga berupaya untuk mengukur akurasi model *Naïve Bayes* dalam memprediksi tren kriminalitas. Dengan hasil penelitian ini, diharapkan aparat kepolisian dapat memiliki alat bantu analisis yang lebih akurat guna meningkatkan efektivitas strategi pencegahan kejahatan.

Hasil analisis menunjukkan bahwa wilayah dengan tingkat pengangguran tinggi cenderung mengalami peningkatan angka kriminalitas. Model *Naïve Bayes* terbukti efektif dalam mendeteksi pola kriminalitas, terutama dalam kategori pencurian berat, dengan tingkat akurasi mencapai 90%. Model ini memiliki *recall* 1.00 dalam mengidentifikasi peningkatan kriminalitas, tetapi kurang akurat dalam mengklasifikasikan penurunan kejahatan (*recall* 0.67). Kecepatan dan efisiensi metode *Naïve Bayes* menjadikannya alat yang potensial untuk analisis kriminalitas di Polres PALI, meskipun masih diperlukan penyempurnaan untuk meningkatkan akurasi pada kategori tertentu. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam bagi aparat penegak hukum dalam menyusun strategi pencegahan kejahatan yang lebih efektif dan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Analisis dan Prediksi

Analisis dan prediksi merupakan dua pendekatan utama dalam memahami pola kriminalitas serta memproyeksikan tren kejahatan di masa mendatang. Analisis statistik digunakan sebagai alat untuk mengidentifikasi hubungan antara variabel yang berkontribusi terhadap kriminalitas, seperti tingkat pengangguran dan jumlah kasus kejahatan. Dalam

konteks ini, analisis deskriptif berfungsi untuk menggambarkan karakteristik data kriminalitas, sedangkan analisis prediktif dimanfaatkan guna memperkirakan pola kejahatan berdasarkan data historis yang tersedia [12].

Dalam penelitian ini, metode *Naïve Bayes* digunakan sebagai pendekatan utama dalam prediksi kriminalitas. Metode ini bekerja berdasarkan prinsip probabilitas untuk mengklasifikasikan tingkat kriminalitas berdasarkan faktor-faktor utama yang relevan. Keunggulan dari metode ini terletak pada kemampuannya dalam mengolah data dengan cepat serta memberikan hasil prediksi yang dapat digunakan sebagai dasar bagi aparat kepolisian dalam mengantisipasi potensi peningkatan angka kriminalitas [13].

Oleh karena itu, penerapan metode prediktif berbasis *Naïve Bayes* diharapkan dapat mendukung efektivitas strategi pencegahan kejahatan yang lebih terstruktur dan berbasis data.

2.2. Kriminalitas

Kriminalitas merupakan tindakan yang bertentangan dengan hukum dan norma sosial yang berlaku dalam suatu masyarakat. Berbagai faktor dapat memengaruhi tingkat kriminalitas, termasuk aspek sosial, ekonomi, lingkungan, serta kebijakan hukum yang diterapkan. Dalam penelitian ini, faktor ekonomi, khususnya tingkat pengangguran, menjadi variabel utama dalam menganalisis pola kejahatan di wilayah hukum Polres PALI Polda Sumsel [1], [3], [12].

Salah satu bentuk kejahatan yang menjadi fokus dalam penelitian ini adalah pencurian dengan pemberatan. Berdasarkan data laporan kriminalitas di Polres PALI, jumlah kasus pencurian dengan pemberatan mengalami peningkatan signifikan dalam empat tahun terakhir, dari 33 kasus pada tahun 2020 menjadi 127 kasus pada tahun 2023. Peningkatan ini menunjukkan adanya kebutuhan mendesak untuk mengembangkan strategi pencegahan yang lebih efektif berbasis analisis data yang akurat. Dengan memahami pola kriminalitas berdasarkan data historis, aparat kepolisian dapat menyusun kebijakan yang lebih tepat dalam menekan angka kejahatan, khususnya pada kasus pencurian dengan pemberatan yang menjadi kejahatan paling dominan di wilayah tersebut.

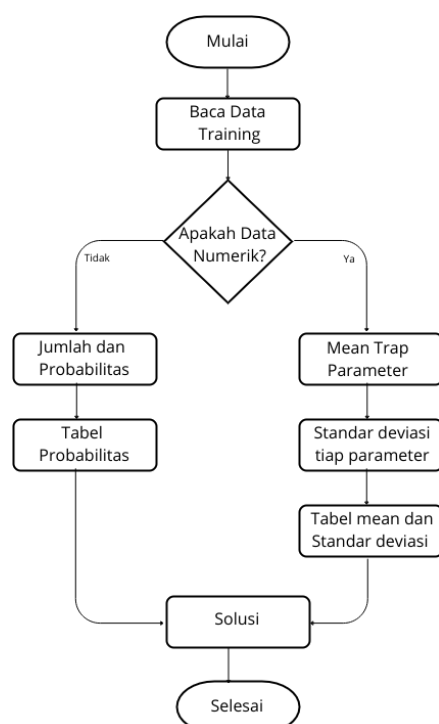
2.3. Data Mining

Data mining merupakan suatu teknik dalam analisis data yang digunakan untuk menggali informasi tersembunyi dari kumpulan data berskala besar. Teknik ini bertujuan untuk menemukan pola dan hubungan yang dapat digunakan dalam pengambilan keputusan strategis, termasuk dalam bidang kriminalitas [14]–[16].

Dalam penelitian ini, *data mining* diterapkan guna mengidentifikasi tren kejahatan dengan beberapa metode utama, seperti klasifikasi dan prediksi.

Metode klasifikasi digunakan untuk mengelompokkan data kriminalitas ke dalam kategori tertentu berdasarkan karakteristiknya, sedangkan metode prediksi bertujuan untuk memperkirakan kemungkinan tren kejahatan di masa mendatang berdasarkan pola yang telah ditemukan dalam data historis. Pendekatan yang digunakan dalam penelitian ini adalah metode *Naïve Bayes*, yang dikenal sebagai algoritma klasifikasi berbasis probabilitas yang cepat dan efisien dalam menangani data kategori. Dengan metode ini, hasil analisis dapat membantu dalam memetakan wilayah dengan tingkat kriminalitas tinggi serta mendukung perancangan kebijakan pencegahan yang lebih efektif [17].

2.4. Metode Naïve Bayes



Gambar 1. Flowchart Naïve Bayes

Metode *Naïve Bayes* merupakan salah satu teknik klasifikasi yang berbasis pada *Teorema Bayes*, yang menghitung probabilitas suatu kejadian berdasarkan informasi sebelumnya. Metode ini banyak digunakan dalam analisis data karena kemampuannya dalam melakukan klasifikasi dengan cepat dan efisien, meskipun memiliki asumsi dasar bahwa setiap variabel bersifat independen satu sama lain. Dalam praktiknya, metode ini tetap dapat menghasilkan klasifikasi yang akurat meskipun asumsi independensi tersebut tidak sepenuhnya terpenuhi [2], [12].

Metode *Naïve Bayes* memiliki beberapa keunggulan yang menjadikannya relevan dalam penelitian ini, di antaranya adalah kemudahan implementasi, kecepatan pemrosesan yang tinggi, serta efektivitas dalam klasifikasi data kriminalitas berbasis kategori. [17].

Dalam penelitian ini, metode *Naïve Bayes* digunakan untuk mengklasifikasikan tingkat kriminalitas berdasarkan data historis di Polres PALI Polda Sumsel. Tujuan utama penerapan metode ini adalah untuk menghasilkan prediksi yang dapat dijadikan referensi dalam penyusunan strategi pencegahan kriminalitas yang lebih berbasis data. Dibawah ini merupakan gambar *Flowchart Naïve Bayes* secara umum:

Secara matematis, Teorema Bayes dinyatakan dalam Rumus berikut:

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)} \quad (1) [18]$$

Penjelasan:

- h : Hipotesis bahwa data termasuk dalam suatu kelas tertentu
- D : Data yang belum memiliki kelas
- $P(h)$: Probabilitas hipotesis (prior probability)
- $P(D)$: Probabilitas dari data D
- $P(h|D)$: Probabilitas hipotesis h berdasarkan kondisi data D (Posterior Probability)
- $P(D|h)$: Probabilitas data D berdasarkan kondisi hipotesis h

2.5. Review Hasil Penelitian Terdahulu

Beberapa penelitian terdahulu telah mengkaji penggunaan metode *Naïve Bayes* dalam berbagai konteks analisis data, termasuk prediksi kriminalitas. Penelitian yang dilakukan oleh Djalal et al. (2023) membandingkan beberapa algoritma *machine learning*, seperti *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine (SVM)* dalam memprediksi kejahatan. Hasilnya menunjukkan bahwa *Naïve Bayes* memiliki akurasi yang cukup baik untuk data yang terstruktur dan unggul dalam kecepatan pelatihan dibandingkan algoritma lainnya [19].

Penelitian lain oleh Safitri et al. (2021) menyoroti keunggulan *Naïve Bayes* dalam efisiensi pemrosesan data besar, meskipun algoritma *Random Forest* memberikan akurasi yang lebih tinggi [1].

Namun, kesederhanaan *Naïve Bayes* menjadikannya pilihan yang menarik, terutama dalam kasus yang memerlukan perhitungan cepat dan pemrosesan data yang luas. Selain itu, penelitian oleh Khan et al. (2022) membandingkan *Naïve Bayes*, *Random Forest*, dan *Neural Networks* dalam analisis kriminalitas. Meskipun *Neural Networks* memiliki akurasi terbaik, *Naïve Bayes* tetap kompetitif dalam menangani dataset berdimensi tinggi dengan kecepatan yang lebih unggul [20].

Penelitian ini menunjukkan bahwa meskipun *Naïve Bayes* bukan algoritma dengan akurasi tertinggi, ia tetap menjadi pilihan yang efektif karena kestabilan, efisiensi, serta kemudahan implementasinya dalam analisis prediksi kriminalitas. Oleh karena itu, penelitian ini menggunakan pendekatan *Naïve Bayes* untuk mengklasifikasikan tren kriminalitas di wilayah Polres PALI Polda Sumsel dengan mempertimbangkan kecepatan, kemudahan

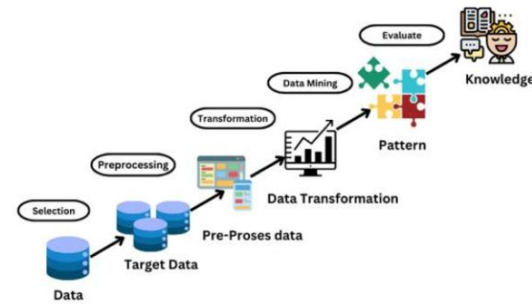
implementasi, serta kemampuan algoritma dalam menangani data dengan banyak variabel.

3. METODE PENELITIAN

Pada Penelitian ini menggunakan metode *Knowledge Discovery in Databases* (KDD) dalam analisis dan prediksi kriminalitas di wilayah Polres PALI Polda Sumsel. KDD adalah pendekatan sistematis yang terdiri dari beberapa tahapan utama dalam mengolah data hingga menghasilkan informasi yang bermanfaat [21]. Tahapan-tahapan dalam metode KDD yang digunakan dalam penelitian ini adalah sebagai berikut:

- Selection (Pemilihan Data):** Data yang digunakan dalam penelitian ini diperoleh dari laporan kriminalitas Polres PALI Polda Sumsel selama periode 2020 hingga 2023. Data ini mencakup informasi seperti jenis tindak pidana, lokasi kejadian, waktu kejadian, dan faktor lainnya yang relevan dengan prediksi kriminalitas.
- Pre-processing (Pembersihan Data):** Pada tahap ini, dilakukan pembersihan data untuk menghilangkan duplikasi, menangani data yang hilang, serta mengonversi variabel ke dalam format yang sesuai untuk analisis. Data yang tidak lengkap atau tidak relevan akan dihapus atau disesuaikan agar dapat diproses dengan lebih akurat.
- Transformation (Transformasi Data):** Data yang telah dibersihkan kemudian dikonversi ke dalam bentuk yang lebih sesuai untuk analisis. Proses ini mencakup normalisasi data, encoding variabel kategorikal, serta agregasi data jika diperlukan untuk meningkatkan kualitas prediksi.
- Data Mining (Penerapan Model Naïve Bayes):** Pada tahap ini, algoritma *Naïve Bayes* diterapkan untuk mengidentifikasi pola kriminalitas berdasarkan data historis. Model ini dikembangkan dengan membagi data menjadi dua bagian, yaitu data latih (training set) dan data uji (test set). Data latih digunakan untuk membangun model prediksi, sedangkan data uji digunakan untuk mengukur akurasi dan efektivitas model.
- Interpretation and Evaluation (Interpretasi dan Evaluasi Hasil):** Setelah model diterapkan, dilakukan evaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja prediksi. Interpretasi hasil dilakukan dengan menganalisis pola kriminalitas yang ditemukan serta memberikan rekomendasi berbasis data bagi kepolisian dalam upaya pencegahan kejahatan.

Dengan menggunakan pendekatan KDD, penelitian ini diharapkan dapat menghasilkan model prediksi kriminalitas yang akurat dan bermanfaat dalam pengambilan keputusan oleh aparat kepolisian.



Gambar 2. Proses *Knowledge Discovery in Database Process* (KDD) [22]

4. HASIL DAN PEMBAHASAN

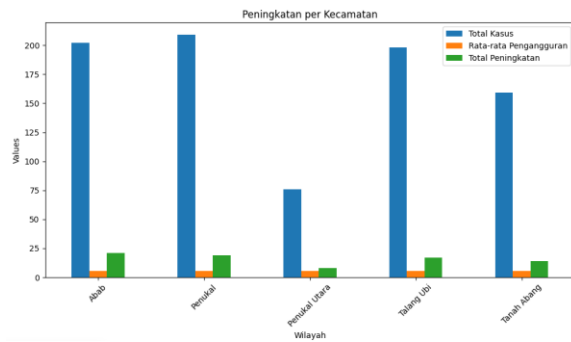
Memuat Bagian ini menyajikan hasil penelitian yang diperoleh dari analisis data laporan kriminalitas di wilayah Polres PALI, POLDA Sumatera Selatan. Penelitian ini bertujuan untuk mengidentifikasi angka peningkatan atau penurunan kasus pencurian berat serta mengevaluasi model analisis menggunakan metode *Naïve Bayes*. Hasil penelitian mencakup proses seleksi data, analisis angka, serta evaluasi performa model analisis yang digunakan. Penjelasan setiap tahap disajikan secara sistematis untuk menjawab rumusan masalah yang telah dirumuskan sebelumnya.

4.1. Seleksi Data

Dataset yang digunakan dalam penelitian ini berasal dari laporan resmi Polres PALI, POLDA Sumatera Selatan, yang mencakup periode Januari 2020 hingga Desember 2022. Data ini terdiri dari 400 laporan kriminalitas, dengan 100 laporan di antaranya berkaitan langsung dengan kasus pencurian berat. Dataset ini mencakup beberapa variabel utama yang digunakan dalam analisis, yaitu wilayah kejadian, jumlah kasus, tingkat pengangguran, serta status peningkatan angka kriminalitas berdasarkan data dari periode sebelumnya.

Wilayah kejadian mencakup beberapa kecamatan, yaitu Talang Ubi, Penukul, Abab, Tanah Abang, dan Penukul Utara. Variabel jumlah kasus merepresentasikan total tindak pidana yang dilaporkan di setiap wilayah dalam periode tertentu. Tingkat pengangguran dihitung berdasarkan persentase pengangguran di masing-masing kecamatan dan diasumsikan memiliki hubungan dengan angka kriminalitas. Variabel peningkatan berfungsi sebagai label klasifikasi untuk menunjukkan apakah jumlah kasus mengalami peningkatan dibandingkan dengan periode sebelumnya.

Untuk mendukung analisis, data yang digunakan telah diseleksi berdasarkan keterkaitan dengan fokus penelitian. Visualisasi data dalam bentuk diagram batang digunakan untuk memberikan gambaran mengenai distribusi peningkatan kasus berdasarkan wilayah.



Gambar 3. Diagram Peningkatan Per Kecamatan

Gambar diatas menunjukkan distribusi jumlah kasus, rata-rata tingkat pengangguran, dan total peningkatan kasus kriminalitas di setiap kecamatan wilayah Polres PALI.

4.2. Praproses Data

Langkah praproses data dilakukan untuk memastikan bahwa dataset yang digunakan dalam penelitian ini berkualitas tinggi dan sesuai untuk analisis. Tahapan praproses data meliputi pembersihan data, transformasi data, dan pengelompokan data sesuai kebutuhan penelitian.

4.2.1. Pembersihan Data

Pada tahap ini, berbagai masalah dalam dataset telah diidentifikasi dan diselesaikan untuk memastikan kualitas data yang optimal sebelum dilakukan analisis. Data yang hilang pada atribut penting, seperti jumlah kasus atau tingkat pengangguran, telah ditangani dengan metode *mean imputation* atau dihapus jika jumlahnya kecil dan tidak memengaruhi analisis secara signifikan. Selain itu, data tidak valid, seperti tingkat pengangguran di atas 100% atau jumlah kasus negatif, telah diperbaiki atau dihapus untuk menjaga konsistensi dataset.

Tabel dibawah ini memberikan ringkasan statistik dari dataset setelah dilakukan pembersihan data.

Tabel 1. Statistical Summary

Statistik Deskriptif	Jumlah Kasus	Tingkat Pengangguran
Count	100	100
Mean	8,44	5,477
Std	4,45	0,297
Min	1	5
25%	4,75	5,2
50%	9	5,5
75%	12	5,7
Max	15	6

4.2.2. Transformasi Data

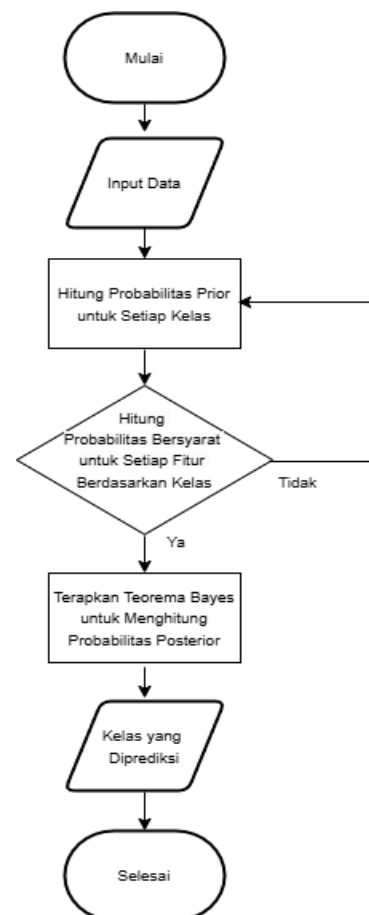
Transformasi data dilakukan untuk mengubah beberapa atribut menjadi format yang lebih sesuai untuk proses analisis. Dalam hal ini, atribut Wilayah dan Peningkatan telah dikonversi ke dalam format numerik menggunakan *label encoding*.

4.2.3. Pengelompokan Data

Dataset dikelompokkan berdasarkan kecamatan dan kategori kasus untuk memahami distribusi data dan tren kriminalitas pada setiap wilayah.

4.3. Penerapan Model Naïve Bayes

Penelitian ini menggunakan metode *Naïve Bayes* untuk memprediksi angka peningkatan atau penurunan kasus pencurian berat di wilayah Polres PALI. Implementasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka *Scikit-learn* dalam notebook interaktif untuk mempermudah analisis dan dokumentasi hasil. Dataset yang telah melalui tahap pembersihan dan transformasi dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan pembagian sebesar 80% untuk data latih dan 20% untuk data uji.



Gambar 4. Implementasi Model Naïve Bayes

Gambar tersebut menunjukkan alur kerja dalam implementasi model Naïve Bayes menggunakan dataset yang telah diproses.

4.4. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja metode *Naïve Bayes* dalam memprediksi angka kriminalitas. Evaluasi mencakup analisis akurasi model, matriks evaluasi, serta visualisasi hasil.

Tabel 2. Tingkat Akurasi Model *Naïve Bayes* berdasarkan pembagian dataset

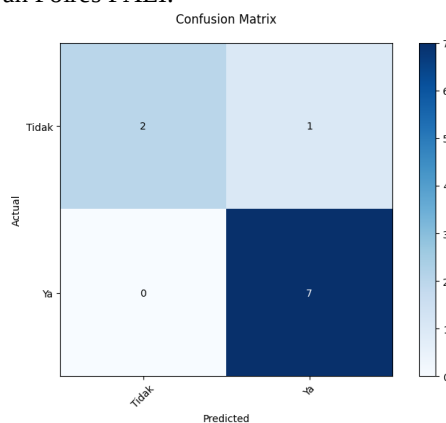
Pembagian Data	Data Latih (%)	Data Validasi (%)	Data Uji (%)	Akurasi Data Latih	Akurasi Data Validasi	Akurasi Data Uji
70:20:10	70	20	10	97.5%	100%	90%
70:10:20	70	10	20	97.14%	93.33%	90%
60:20:20	60	20	20	96.67%	95%	90%

Tabel 3. Classification Report model terbaik pembagian data (70:20:10)

Metric	Precision	Recall	F1-Score	Support
Class 0 (tidak)	01.00	0.67	0.80	3
Class 1 (ya)	0.88	01.00	0.93	7
Accuracy	-	-	0.90	10
Macro Avg	0.94	0.83	0.87	10
Weighted Avg	0.91	0.90	0.89	10

Hasil evaluasi model dengan pembagian data 70:20:10 menunjukkan bahwa performa klasifikasi untuk kelas "Tidak" (Class 0) memiliki *precision* sebesar 1.00, *recall* sebesar 0.67, dan *F1-score* sebesar 0.80, dengan jumlah *support* sebanyak 3 data. Sementara itu, untuk kelas "Ya" (Class 1), model mencapai *precision* sebesar 0.88, *recall* sebesar 1.00, dan *F1-score* sebesar 0.93, dengan jumlah *support* sebanyak 7 data. Akurasi keseluruhan model tercatat sebesar 90% berdasarkan total 10 data uji.

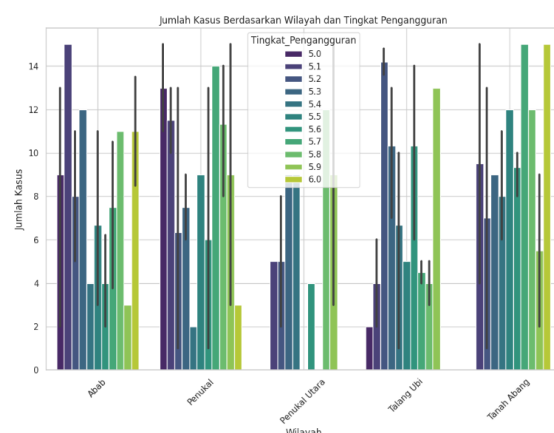
Selain itu, nilai rata-rata makro (*macro average*) dari model menunjukkan *precision* sebesar 0.94, *recall* sebesar 0.83, dan *F1-score* sebesar 0.87, sedangkan rata-rata berbobot (*weighted average*) menunjukkan *precision* sebesar 0.91, *recall* sebesar 0.90, dan *F1-score* sebesar 0.89. Secara keseluruhan, model memiliki kinerja yang sangat baik dalam mendeteksi kategori "Ya" dengan *F1-score* mencapai 0.93, menunjukkan bahwa model dapat mengidentifikasi peningkatan angka kriminalitas dengan baik. Namun, performa pada kategori "Tidak" mengalami kelemahan, khususnya pada *recall* yang hanya sebesar 0.67, sehingga beberapa kasus yang seharusnya diklasifikasikan sebagai "Tidak" tidak terdeteksi dengan baik oleh model. Meskipun demikian, akurasi keseluruhan sebesar 90% menunjukkan bahwa model memiliki performa yang solid dalam memprediksi peningkatan atau penurunan angka kriminalitas di wilayah Polres PALI.

Gambar 5 *Confusion Matrix* model terbaik pembagian data (70:20:10)

Gambar dibawah 5 menunjukkan hasil evaluasi model *Naïve Bayes* dengan confusion matrix yang menggambarkan jumlah prediksi benar dan salah untuk setiap kelas.

Confusion matrix ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan kategori peningkatan dengan recall sebesar 1.00, namun masih mengalami sedikit kesalahan dalam mengklasifikasikan kategori penurunan dengan recall 0.67. Model ini menunjukkan akurasi keseluruhan sebesar 90%, yang mencerminkan performa yang solid dalam memprediksi angka kriminalitas.

Hasil klasifikasi model menunjukkan bahwa jumlah kasus menjadi faktor dominan yang berkontribusi terhadap tingkat kriminalitas, sementara tingkat pengangguran memberikan pengaruh yang lebih kecil dalam penentuan angka kriminalitas. Hasil utama dari analisis ini menunjukkan bahwa model mampu mengklasifikasikan angka kriminalitas berdasarkan data historis dengan cukup baik. Meskipun tingkat pengangguran memiliki beberapa pengaruh, Jumlah Kasus adalah faktor yang lebih dominan dalam mempengaruhi tingkat kriminalitas di wilayah tertentu.



Gambar 6. Jumlah Kasus Berdasarkan Wilayah dan Tingkat Pengangguran

Hasil grafik di atas menunjukkan bahwa wilayah dengan jumlah kasus yang lebih tinggi cenderung memiliki angka kriminalitas yang lebih tinggi,

sementara tingkat pengangguran tidak memiliki pengaruh yang sangat kuat terhadap perubahan jumlah kasus. Oleh karena itu, hasil klasifikasi ini memberikan wawasan penting bagi pihak berwenang dalam menentukan strategi pencegahan yang lebih tepat sasaran, khususnya di wilayah dengan jumlah kasus yang tinggi dan risiko kriminalitas yang besar.

4.5. Interpretasi Hasil

Model *Naïve Bayes* yang telah dilatih digunakan untuk memprediksi angka kriminalitas berdasarkan jumlah kasus dan tingkat pengangguran. Data uji dinormalisasi untuk memastikan kompatibilitas dengan data pelatihan sebelum dimasukkan ke dalam model. Prediksi yang dihasilkan berupa label "Ya" (terjadi peningkatan kriminalitas) atau "Tidak" (tidak terjadi peningkatan), yang kemudian dikonversi kembali ke format aslinya menggunakan *LabelEncoder* untuk mempermudah interpretasi.

Tabel 4. Data Uji dan Hasil Prediksi Model

No	Wilayah	Jumlah Kasus	Tingkat Pengangguran	Prediksi
1	Abab	12	5	Ya
2	Penukal Utara	5	5.1	Tidak
3	Abab	7	5.6	Ya
4	Abab	11	5.8	Ya
5	Tanah Abang	11	5.4	Ya
6	Talang Ubi	6	5.1	Tidak
7	Tanah Abang	8	5.4	Ya
8	Tanah Abang	15	6.0	Ya
9	Penukal	9	5.5	Ya
10	Abab	8	6	Ya

Hasil prediksi menunjukkan bahwa wilayah dengan jumlah kasus lebih dari 5 dan tingkat pengangguran yang tinggi cenderung mengalami peningkatan kriminalitas. Faktor utama yang memengaruhi prediksi adalah jumlah kasus yang tinggi, dengan tingkat pengangguran sebagai variabel pendukung. Misalnya, wilayah Abab dan Tanah Abang dengan jumlah kasus lebih dari 7 dan tingkat pengangguran ≥ 5.4 diprediksi mengalami peningkatan kriminalitas. Sebaliknya, wilayah seperti Penukal Utara dan Talang Ubi dengan jumlah kasus rendah dan tingkat pengangguran moderat cenderung tidak mengalami peningkatan.

Temuan ini dapat menjadi acuan bagi aparat kepolisian dalam menentukan strategi pencegahan kejahatan, seperti meningkatkan pengawasan di wilayah dengan jumlah kasus tinggi serta menurunkan tingkat pengangguran sebagai langkah preventif.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian model prediksi, angka kriminalitas di wilayah Polres

PALI dapat diklasifikasikan dengan baik menggunakan metode *Naïve Bayes*, dengan akurasi 90% dan kemampuan sangat baik dalam mendeteksi peningkatan kriminalitas (*recall* 1.00). Namun, model ini masih perlu perbaikan dalam mengklasifikasikan penurunan kriminalitas (*recall* 0.67). Untuk meningkatkan akurasi, disarankan penggunaan dataset yang lebih besar dan seimbang serta penambahan variabel seperti tingkat pendidikan dan status ekonomi. Selain itu, metode analitik lebih kompleks seperti *Random Forest* atau *Gradient Boosting* dapat meningkatkan prediksi, dan pengembangan sistem prediksi berbasis *real-time* akan membantu respons yang lebih cepat. Hasil penelitian ini dapat menjadi dasar dalam merumuskan strategi pencegahan berbasis data di wilayah dengan tingkat pengangguran tinggi seperti Penukal, Abab, dan Tanah Abang, dengan fokus pada pemberdayaan masyarakat dan pengentasan kemiskinan.

DAFTAR PUSTAKA

- [1] R. Safitri, N. Alfira, D. Tamitiadini, W. W. A. Dewi, and N. Febriani, *Analisis sentimen: Metode alternatif penelitian big data*, Universitas Brawijaya Press, 2021.
- [2] I. R. T. Yunardi, N. Kom, and M. Kom, *Data Mining dan Machine Learning dengan Orange3 Tutorial dan Aplikasinya*, Universitas Indonesia Press, 2022.
- [3] A. Aini, A. Zakir, and D. Irwan, "Penerapan Algoritma C4.5 Dalam Penggolompokan Penanganan Kasus Kejahatan Yang Sering Terjadi Di Polsek Percut Sei Tuan," *Djtechno J. Teknol. Inf.*, vol. 3, no. 2, pp. 244–250, 2022.
- [4] Z. Resiana and T. Aditya, "Analitik Geovisual Pengaruh Pandemi COVID-19 Terhadap Pola Dan Kecenderungan Kriminalitas Di Daerah Istimewa Yogyakarta," *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 6, no. 1, p. 24, 2023.
- [5] E. Pratama, *Analisis Metode Clustering dengan Algoritma Spectral Clustering dalam Pengelompokan Tingkat Tindak Kriminalitas*, Universitas Medan Area Press, 2023.
- [6] S. I. Usman, *Analisis Perilaku dan Dampak Malware Ransomware Locky Metode Dinamis Framework FlareVM*, Universitas Nusaputra Press, 2024.
- [7] P. Muthmainnah and R. Saputra, "Analisis Sentimen Layanan Jasa Ekspedisi J&T Express Dengan Menggunakan Metode Naïve Bayes Optimasi Adaboost," unpublished.
- [8] L. Ghiffari, N. Gusriani, and K. Parmikanti, "Pemetaan Jenis Tindak Kriminal di Indonesia Berdasarkan Karakteristik Wilayah Menggunakan Canonical Correspondence Analysis (CCA)," *J. Stat. dan Apl.*, vol. 5, no. 2, pp. 133–145, 2021.
- [9] M. Z. Zufar, *Aplikasi Pendeteksi Situs Phising Berbasis Website Menggunakan Metode Naïve Bayes*, Universitas Politeknik Tegal Press, 2023.

- [10] A. R. S. Harahap, *Aplikasi Scoping Review “Klasifikasi Algoritma Deteksi Serangan Phishing” Berbasis Web Menggunakan Metode Web Development Life Cycle*, Universitas Islam Negeri Jakarta Press, 2023.
- [11] B. Budiman, “Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru,” *Nuansa Inform.*, vol. 15, no. 2, pp. 37–52, 2021.
- [12] M. R. Supriadi and R. Andarsyah, *Deteksi Halaman Website Phishing Menggunakan Algoritma Machine Learning Gradient Boosting Classifier*, Universitas Indonesia Press, 2023.
- [13] N. Hafidhigbal, *Implementasi Model Indobert untuk Klasifikasi Jenis Tindak Pidana Korupsi Berbasis Web Application*, Universitas Islam Negeri Jakarta Press, 2024.
- [14] R. A. Rival, R. M. R. Martiansah, L. Sari, et al., “Penerapan Algoritma Apriori untuk Menentukan Pola Transaksi Pembelian Minuman dan Makanan di Coffee Gubuk,” *SENTIMAS Semin.*, 2022.
- [15] S. A. Fadhilah, “Analisis Perilaku Pembelian Konsumen Menggunakan Metode Association Rule-Market Basket Analysis Dan Clustering Analysis (Studi Kasus: Jore Coffee & Eatery),” *Dspace.uui.ac.id*, 2022.
- [16] P. N. Safitri, “Analisis Strategi Penjualan Produk Menggunakan Association Rule dengan Algoritma Apriori (Studi Kasus: Penjualan Produk Pamella Satu Supermarket),” *Dspace.uui.ac.id*, 2022.
- [17] M. L. Hakim, “Analisis Sentimen Opini Masyarakat Terhadap Pengesahan RKUHP Pada Twitter Menggunakan Algoritma Support Vector Machine (SVM) dan Decision Tree,” *Repository.uinjkt.ac.id*.
- [18] D. R. Nuruzzulfa, “Perbandingan Adaboost C4.5 dan Adaboost Naive Bayes Pada Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati,” *Repository.unimus.ac.id*, 2020.
- [19] M. R. Djalal, M. Ali, H. Nurohmah, and D. Ajiatmo, “Aplikasi Algoritma Differential Evolution untuk Desain Optimal Load Frequency Control pada Pembangkit Listrik Tenaga Hibrid Angin dan Diesel,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 511–516, 2018.
- [20] M. Khan, A. Ali, and Y. Alharbi, “Predicting and Preventing Crime: A Crime Prediction Model Using San Francisco Crime Data by Classification Techniques,” *Complexity*, vol. 2022, 2022.
- [21] Y. Aqsho Ramadhan, A. Faqih, and G. Dwilestari, “Prediksi Penjualan Handphone di Toko X menggunakan Algoritma Regresi Linear,” *J. Inform. Terpadu*, vol. 9, no. 1, pp. 40–44, 2023.
- [22] F. Ardiansyah, F. Hamdan, S. Sugiyanto, et al., “Klasifikasi Customer Relationship Management Menggunakan Dataset KDD Cup 2009 dengan Teknik Reduksi Dimensi,” *J. Sist. Komput.*, 2022.