



KOMBINASI XGBOOST DAN RANDOM FOREST UNTUK PREDIKSI TINGKAT KRIMINALITAS BERDASARKAN DATA SOSIAL DAN EKONOMI

RICHARD PETRUS HAPOSAN SIAGIAN

NPM 21081010250

DOSEN PEMBIMBING

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI UNIVERSITAS PEMBANGUNAN NASIONAL
VETERAN JAWA TIMUR FAKULTAS ILMU KOMPUTER
PROGRAM STUDI
INFORMATIKA SURABAYA
2024**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tingkat kriminalitas merupakan salah satu indikator penting yang mencerminkan stabilitas sosial dan ekonomi suatu wilayah. Kriminalitas yang tinggi dapat berdampak buruk terhadap pertumbuhan ekonomi, investasi, dan kualitas hidup masyarakat. Oleh karena itu, memahami faktor-faktor yang memengaruhi tingkat kriminalitas dan memprediksi pola kriminalitas menjadi langkah krusial dalam perencanaan kebijakan pemerintah dan penegakan hukum.

Berbagai penelitian telah menunjukkan bahwa faktor-faktor sosial dan ekonomi, seperti tingkat pengangguran, kemiskinan, kepadatan penduduk, pendidikan, serta pendapatan masyarakat, memiliki pengaruh signifikan terhadap tingkat kriminalitas. Data ini, jika diolah dengan baik, dapat digunakan untuk membangun model prediksi yang mampu memetakan tingkat kejahatan pada wilayah tertentu. Dengan adanya prediksi yang akurat, pihak berwenang dapat merumuskan kebijakan preventif yang lebih tepat sasaran. Di era perkembangan teknologi saat ini, metode Machine Learning telah menjadi solusi efektif dalam menganalisis data dalam skala besar dan menghasilkan prediksi yang akurat. Dua metode populer dalam Machine Learning adalah XGBoost (Extreme Gradient Boosting) dan Random Forest. XGBoost dikenal karena performanya yang cepat dan akurat dalam menangani data kompleks, sedangkan Random Forest unggul dalam menangani variabel yang bersifat non-linear dan meminimalkan risiko overfitting. Kombinasi kedua metode ini berpotensi menghasilkan model yang lebih robust dan mampu memberikan prediksi tingkat kriminalitas dengan akurasi tinggi.

Penelitian ini bertujuan untuk memanfaatkan kombinasi XGBoost dan Random Forest dalam membangun model prediksi tingkat kriminalitas berdasarkan data sosial dan ekonomi. Studi ini diharapkan dapat membantu dalam identifikasi area rawan kriminalitas serta mendukung pemerintah dalam menyusun strategi pencegahan kriminal yang lebih efektif dan berbasis data.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Apa saja variabel sosial dan ekonomi yang memiliki pengaruh paling signifikan terhadap tingkat kriminalitas?
2. Bagaimana kinerja metode XGBoost dan Random Forest secara bersamaan dalam memprediksi tingkat kriminalitas?
3. Apakah kombinasi algoritma XGBoost dan Random Forest mampu meningkatkan akurasi prediksi tingkat kriminalitas dibandingkan dengan penggunaan metode tersebut secara individu?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Menganalisis urgensi penggunaan metode XGboost dan Random Forest dalam memprediksi tingkat kriminalitas.
2. Mengimplementasikan metode XGBoost dan Random Forest untuk mengidentifikasi faktor-faktor utama yang memengaruhi tingkat kriminalitas secara akurat.
3. Mengoptimalkan parameter model Random Forest menggunakan XGboost untuk meningkatkan performa prediksi tingkat kriminalitas.
4. Mengevaluasi efektivitas metode XGboost dan Random Forest dibandingkan dengan metode konvensional dalam memodelkan hubungan non-linear antar variabel daya saing.
5. Menganalisis pengaruh variabel-variabel sosial dan ekonomi terhadap tingkat kriminalitas.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1. Memperbanyak penerapan metode XGBoost dan Random Forest pada prediksi sebuah studi kasus.
2. Menjadi acuan bagi peneliti lain yang ingin menggunakan pendekatan XGboost dan Random Forest untuk prediksi sebuah studi kasus.
3. Memberikan informasi mengenai faktor-faktor utama yang memengaruhi tingkat kriminalitas.
4. Membantu dalam merancang kebijakan yang lebih efektif untuk pencegahan dan penanggulangan kriminalitas berdasarkan wilayah yang teridentifikasi berisiko tinggi.
5. Meningkatkan kesadaran akan kondisi lingkungan sosial dan ekonomi yang dapat berpotensi memicu tindak kriminalitas.

1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah sebagai berikut :

1. Penelitian ini berfokus pada tingkat kriminalitas.
2. Dataset yang diambil dari BPS berupa data soisal dan ekonomi, berupa data pengangguran , data kemiskinan, kepadatan penduduk, populasi, dan lain lain
3. Metode yang digunakan dalam penelitian ini adalah kombinasi dari XGboost untuk memprediksi tingkat kriminalitas , dan Random Forest untuk mengklasifikasi tingkat kriminalitas nya.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Beberapa penelitian terkait prediksi tingkat kriminalitas telah dilakukan menggunakan metode dan pendekatan yang beragam, terutama berbasis Machine Learning. Penelitian-penelitian terdahulu ini memberikan landasan penting untuk mengembangkan metode yang lebih akurat dalam memprediksi tingkat kriminalitas. Beberapa penelitian telah dilakukan terkait analisis dan prediksi tingkat kriminalitas dengan menggunakan pendekatan berbasis data. Sema Nabillah Dewi, Imam Cholissodin, dan Edy Santoso (2018) dalam penelitian berjudul "Prediksi Jumlah Kriminalitas Menggunakan Metode Extreme Learning Machine (Studi Kasus di Kabupaten Probolinggo)" memanfaatkan metode Extreme Learning Machine (ELM) untuk memprediksi jumlah kriminalitas pada periode 2012 hingga 2017. Penelitian ini menunjukkan bahwa ELM mampu memberikan hasil prediksi yang cukup baik dengan nilai error terkecil Mean Square Error (MSE) sebesar 0,037662, menggunakan konfigurasi 7 fitur input dan fungsi aktivasi sigmoid biner. Namun, penelitian ini hanya menggunakan metode tunggal dan belum mempertimbangkan algoritma ensemble learning, seperti Random Forest atau XGBoost, yang mampu menangkap pola data yang lebih kompleks dan variabel independen yang lebih beragam.

Penelitian lain yang dilakukan oleh Fardani et al. (2015) menggunakan metode Extreme Learning Machine untuk memprediksi jumlah kunjungan pasien di rumah sakit. Dengan menggunakan fungsi aktivasi sigmoid dan konfigurasi neuron yang optimal, penelitian ini berhasil mencapai nilai error rendah sebesar 0,027. Meskipun penelitian ini tidak berfokus pada prediksi kriminalitas, pendekatan yang digunakan menunjukkan bahwa metode Machine Learning seperti ELM memiliki potensi dalam memprediksi data yang bersifat dinamis dan time series. Namun, dalam konteks kriminalitas yang melibatkan lebih banyak faktor seperti data sosial dan ekonomi, diperlukan metode yang lebih canggih untuk mengatasi kerumitan data tersebut.

Selanjutnya, penelitian oleh Mendome et al. (2016) menerapkan model ARIMA (AutoRegressive Integrated Moving Average) untuk memprediksi tingkat kriminalitas di wilayah Polresta Manado, Sulawesi Utara. Penelitian ini memanfaatkan data deret waktu untuk menangkap pola historis kejahatan dan menunjukkan bahwa ARIMA mampu memodelkan tren

jangka pendek dengan baik. Namun, metode ini memiliki keterbatasan dalam mengolah data dengan variabel independen yang banyak dan kompleks, seperti faktor pengangguran, kemiskinan, tingkat pendidikan, dan kepadatan penduduk, yang dapat memengaruhi tingkat kriminalitas.

Penelitian lain yang relevan adalah Sahu dan Dwivedi (2017) yang mengimplementasikan Random Forest untuk memprediksi pola kejahatan berdasarkan dataset kriminalitas di India. Hasil penelitian ini menunjukkan bahwa Random Forest memiliki performa yang baik dalam memproses variabel dengan skala besar dan mampu memberikan interpretasi penting mengenai variabel dominan yang memengaruhi tingkat kejahatan. Namun, penelitian ini belum melakukan perbandingan dengan metode Machine Learning lainnya seperti XGBoost, yang dikenal memiliki akurasi lebih tinggi pada data yang kompleks.

Sementara itu, Chen et al. (2019) menggunakan metode XGBoost untuk memprediksi kejahatan di area perkotaan dengan mempertimbangkan faktor lingkungan dan demografi. Penelitian ini menunjukkan bahwa XGBoost dapat memberikan prediksi yang akurat dan efisien dengan waktu komputasi yang cepat. Keunggulan XGBoost terletak pada kemampuannya menangani missing values, mencegah overfitting, serta bekerja optimal pada data dengan banyak fitur. Namun, penelitian ini juga dilakukan hanya menggunakan metode tunggal tanpa mempertimbangkan kombinasi dengan metode lainnya untuk meningkatkan performa.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa metode Machine Learning, seperti Extreme Learning Machine (ELM), ARIMA, Random Forest, dan XGBoost, telah menunjukkan hasil yang baik dalam memprediksi data bersifat dinamis, termasuk tingkat kriminalitas. Namun, penelitian sebelumnya cenderung menggunakan metode secara individual dan belum banyak mengeksplorasi pendekatan kombinasi metode (hybrid) untuk meningkatkan akurasi dan performa model. Selain itu, penelitian terdahulu belum sepenuhnya memanfaatkan data sosial dan ekonomi secara menyeluruh sebagai faktor utama dalam memprediksi tingkat kriminalitas.

2.2 Landasan Teori

2.2.1 Data sosial dan ekonomi

Kriminalitas didefinisikan sebagai suatu tindakan yang melanggar hukum dan norma sosial yang berlaku di suatu wilayah. Tingkat kriminalitas sering kali digunakan sebagai indikator kestabilan sosial dan keamanan suatu daerah. Menurut Soekanto (2002), kriminalitas dipengaruhi oleh berbagai faktor, baik yang bersifat internal seperti psikologis individu, maupun eksternal seperti kondisi sosial, ekonomi, dan lingkungan.

38 Provinsi	Tingkat Pengangguran Terbuka Menurut Provinsi (Persen)												
	2015		2016		2017		2018		2019		2020		2021
	Februari	Agustus	Februari	Agustus	Februari	Agustus	Februari	Agustus	Februari	Agustus	Februari	Agustus	Februari
ACEH	7,73	9,93	8,13	7,57	7,39	6,57	6,54	6,34	5,48	6,17	5,4	6,59	6,3
SUMATERA UTARA	6,39	6,71	6,49	5,84	6,41	5,6	5,61	5,55	5,57	5,39	4,71	6,91	6,01
SUMATERA BARAT	5,99	6,89	5,81	5,09	5,8	5,58	5,68	5,66	5,38	5,38	5,25	6,88	6,67
RIAU	6,72	7,83	5,94	7,43	5,76	6,22	5,55	5,98	5,36	5,76	4,92	6,32	4,96
JAMBI	2,73	4,34	4,66	4	3,67	3,87	3,56	3,73	3,52	4,06	4,26	5,13	4,76
SUMATERA SELATAN	5,03	6,07	3,94	4,31	3,8	4,39	4,08	4,27	4,02	4,53	3,9	5,51	5,17
BENGKULU	3,21	4,91	3,84	3,3	2,81	3,74	2,63	3,35	2,41	3,26	3,08	4,07	3,72
LAMPUNG	3,44	5,14	4,54	4,62	4,43	4,33	4,32	4,04	3,95	4,03	4,26	4,67	4,54
KEP. BANGKA RIU LINGG	3,35	6,29	6,17	2,6	4,46	3,78	3,59	3,61	3,32	3,58	3,35	5,25	5,04

Gambar 1 Data Tingkat Pengangguran dari website Badan Pusat Statistik (BPS)

Faktor-faktor utama yang sering dikaitkan dengan kriminalitas antara lain:

1. Tingkat Pengangguran: Rendahnya akses terhadap lapangan pekerjaan meningkatkan risiko kriminalitas akibat tekanan ekonomi.
2. Tingkat Kemiskinan: Kemiskinan sering kali dikaitkan dengan peningkatan tindak kriminalitas karena keterbatasan sumber daya ekonomi.
3. Kepadatan Penduduk: Kepadatan yang tinggi cenderung memicu konflik sosial dan interaksi negatif.
4. Tingkat Pendidikan: Pendidikan yang rendah berdampak pada minimnya peluang kerja dan keterampilan, yang mendorong individu untuk melakukan tindak kriminal.

Dengan memahami faktor-faktor ini, pemodelan prediksi kriminalitas dapat dilakukan dengan memanfaatkan data sosial dan ekonomi yang tersedia.

2.2.2 Machine Learning

Machine Learning adalah cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa pemrograman eksplisit. Dalam Machine Learning, komputer belajar dari pola yang terdapat dalam data menggunakan algoritma tertentu dan mampu melakukan generalisasi untuk memprediksi data yang belum pernah dilihat sebelumnya. Secara umum, Machine Learning dibagi menjadi tiga jenis, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning digunakan ketika data yang digunakan untuk melatih model memiliki label atau target, sehingga cocok untuk masalah klasifikasi dan regresi. Sementara itu, unsupervised learning digunakan untuk menemukan pola dalam data yang tidak memiliki label, seperti dalam klusterisasi. Reinforcement learning bekerja dengan memberikan umpan balik berupa reward atau penalty berdasarkan tindakan yang diambil model. Dalam penelitian ini, metode supervised learning digunakan karena memiliki kemampuan untuk memprediksi tingkat kriminalitas berdasarkan faktor-faktor sosial dan ekonomi yang telah dikategorikan sebagai variabel independen

2.2.3 Gradient Boosting

2.2.3 Klasifikasi

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Metode Penelitian

3.2 Desain Sistem

3.2.1 Preprocessing Data

3.2.2 Pendeteksian Menggunakan CNN

3.2.3 Klasifikasi Kerusakan Menggunakan Gradient Boosting

3.2.4 Evaluasi Model

DAFTAR PUSTAKA