

ABSTRACT

Archipelagic nations such as Indonesia comprise extensive coastal regions. These areas are highly vulnerable to marine threats, particularly those exacerbated by climate change (e.g., sea-level rise, erosion, and coastal flooding), which can potentially lead to substantial losses. Existing vulnerability information tends to be generalized and lacks spatial detail, while the conventional Coastal Vulnerability Index (CVI) model has limitations in its informational presentation and inherent subjectivity. Moreover, machine learning models recently utilized for predicting CVI values are often "black box" in nature.

To present a coastal vulnerability assessment that is objective, accurate, transparent, and spatially detailed, this study implements a framework that calculates vulnerability using a raster-based CVI model with nine coastal landscape parameters (i.e., elevation, slope, geomorphology, land cover, distance to shoreline, shoreline change rate, significant wave height, tidal range, and relative sea-level rise), leveraging the conventional CVI output as ground truth. Three machine learning algorithms: Random Forest, XGBoost, and LightGBM, are employed for classification and to learn the nonlinear relationships between the variables and their output, optimized via the RandomizedSearchCV method. Subsequently, the SHapley Additive exPlanations (SHAP) method from Explainable AI (XAI) is applied to interpret the machine learning models' logic and determine the contribution of each variable.

The results demonstrate that the raster-based CVI visualization is effective in identifying vulnerability hotspots; each machine learning model achieves exceptionally high classification performance (F1-score and Cohen's Kappa >0.99) with minimal error; and the SHAP analysis successfully deconstructs the transparent parameter contributions. This analysis confirms that distance to shoreline, significant wave height, shoreline change rate, and tidal range are the most influential factors, while also revealing the unique predictive strategy of each machine learning model. Moreover, Friedman, Iman-Davenport, and Sequential Rank Agreement Test suggests that LightGBM model produces the best performance and the most stable interpretation.

Keywords -- Coastal Area, Machine Learning, CVI, Coastal Vulnerability, Coastal Change, Raster-based Visualization, SHAP

INTISARI

Negara kepulauan seperti Indonesia terdiri dari banyak wilayah pesisir. Wilayah ini sangat rentan terhadap ancaman dari lautan, seperti yang diakibatkan oleh perubahan iklim (kenaikan muka air laut, erosi, dan banjir pesisir), yang berpotensi menimbulkan kerugian besar. Informasi kerentanan yang ada cenderung umum dan kurang detail spasial, sementara pemanfaatan model *Coastal Vulnerability Index* (CVI) untuk menilai kerentanan pesisir memiliki keterbatasan dalam presentasi informasi dan subjektivitas. Sedangkan, model pembelajaran mesin yang baru-baru ini sering dimanfaatkan untuk memprediksi nilai CVI sering kali bersifat "kotak hitam".

Untuk menyajikan penilaian kerentanan pesisir yang objektif, akurat, transparan, dan terperinci secara spasial, penelitian ini mengimplementasikan kerangka kerja yang menghitung kerentanan menggunakan model CVI berbasis raster dengan sembilan parameter bentang alam pesisir (seperti elevasi, kemiringan, geomorfologi, tutupan lahan, jarak garis pantai, laju perubahan garis pantai, tinggi gelombang, pasang surut, dan kenaikan muka air laut relatif), memanfaatkan luaran CVI konvensional sebagai *ground truth*. Tiga algoritma pembelajaran mesin, yaitu Random Forest, XGBoost, dan LightGBM, dimanfaatkan untuk klasifikasi dan mempelajari hubungan nonlinear antar parameter dengan luarannya yang dioptimasi dengan metode RandomizedSearchCV, lalu metode *SHapley Additive exPlanations* (SHAP) dari *Explainable AI* (XAI) diterapkan untuk menginterpretasi logika model pembelajaran mesin untuk mengetahui kontribusi tiap parameter.

Hasil menunjukkan visualisasi CVI berbasis raster efektif mengidentifikasi *hotspot* kerentanan; tiap model pembelajaran mesin mencapai kinerja klasifikasi sangat tinggi (*F1-score* dan *Cohen's Kappa* >0.99) dengan kesalahan minimal; dan analisis SHAP berhasil menjabarkan kontribusi transparan parameter, mengkonfirmasi bahwa jarak ke garis pantai, ketinggian gelombang signifikan, laju perubahan garis pantai, dan kisaran pasang surut adalah yang paling berpengaruh, sekaligus mengungkap strategi prediktif unik tiap model pembelajaran mesin. Selain itu, uji statistik Friedman, Iman-Davenport, dan *Sequential Rank Agreement* (SRA) menghasilkan konklusi bahwa model LightGBM menghasilkan performa paling baik dan paling stabil interpretasinya.

Kata kunci –Wilayah Pesisir, Pembelajaran Mesin, CVI, Kerentanan Pesisir, Perubahan Pesisir, Visualisasi berbasis raster, SHAP