

ABSTRAK

Kriminalitas merupakan fenomena sosial yang memiliki pola spasial dan dipengaruhi oleh berbagai karakteristik demografis serta sosial-ekonomi wilayah. Namun, ketersediaan data kriminalitas pada level wilayah kecil sering kali terbatas sehingga menyulitkan proses estimasi risiko kriminalitas secara rinci. Penelitian ini bertujuan untuk mengembangkan pendekatan analisis kriminalitas berbasis *Small Area Estimation* (SAE), *Random Forest Regressor*, dan *Rough Set Theory* guna mengestimasi serta memetakan risiko kriminalitas pada level kapanewon di Daerah Istimewa Yogyakarta (DIY).

Penelitian menggunakan data kriminalitas administratif Polda DIY dan data karakteristik wilayah dari BPS serta Dukcapil periode 2022–2025. Data kriminalitas tingkat kabupaten/kota didisagregasi ke level kapanewon menggunakan pendekatan *population-weighted areal interpolation*. Estimasi area kecil dilakukan menggunakan model Fay–Herriot berbasis regresi sintesis berbobot untuk memperoleh estimasi *Empirical Best Linear Unbiased Prediction* (EBLUP). Selanjutnya, hasil estimasi digunakan sebagai target pada pemodelan *Random Forest Regressor* untuk prediksi risiko kriminalitas, sedangkan *Rough Set Theory* digunakan untuk mengekstraksi aturan keputusan berbasis karakteristik sosial-demografis wilayah.

Hasil penelitian menunjukkan bahwa penerapan Small Area Estimation meningkatkan kualitas estimasi kriminalitas ekonomi dengan nilai R^2 meningkat dari 0,806 menjadi 0,890 dan RMSE menurun dari 0,938 menjadi 0,680. Variabel sosial-demografis yang konsisten berasosiasi dengan risiko kriminalitas meliputi tingkat partisipasi sekolah, persentase pendidikan rendah, dependency ratio, dan sex ratio. Pemodelan Random Forest pada target kriminalitas ekonomi menghasilkan performa yang baik dengan nilai R^2 sebesar 0,9058, RMSE sebesar 0,630, dan MAE sebesar 0,510, melampaui ambang $R^2 > 0,70$ yang ditetapkan dalam hipotesis penelitian. Selanjutnya, Rough Set Theory menghasilkan 26 aturan keputusan dengan rata-rata confidence 0,942 yang mampu mengidentifikasi pola risiko kriminalitas berbasis karakteristik wilayah. Visualisasi spasial menunjukkan bahwa wilayah urban dan suburban DIY, khususnya sebagian wilayah Sleman dan Kota Yogyakarta, cenderung memiliki tingkat risiko kriminalitas ekonomi yang lebih tinggi dibandingkan wilayah perdesaan.

Penelitian ini memberikan empat kontribusi ilmiah: (i) kerangka *hybrid* SAE–RF–RST untuk pemetaan risiko kriminalitas pada resolusi sub-kabupaten yang sepanjang penelusuran literatur belum pernah diintegrasikan dalam satu *pipeline* koheren; (ii) penyelesaian *spatial misalignment* berbasis *mass-preserving aggregation* yang menjamin konsistensi agregat secara konstruktif antara estimasi kapanewon dan total kabupaten/kota; (iii) translasi prediksi Random Forest menjadi aturan keputusan IF–THEN melalui *Rough Set Theory* yang dapat divalidasi pakar domain dan dipakai sebagai input sistem pendukung keputusan; serta (iv) *baseline* estimasi risiko kriminalitas berdasarkan data administratif 78 kapanewon DIY periode 2022–2025 sebagai titik banding bagi studi keamanan publik di tingkat sub-kabupaten. Kerangka yang dihasilkan dirancang sebagai sistem pendukung keputusan (*decision support system*) bagi instansi penegak hukum di DIY, bukan sebagai sistem prediksi otomatis.

Kata kunci—kriminalitas, Small Area Estimation, Fay–Herriot, Random Forest, Rough Set Theory, pemetaan risiko kriminalitas.

ABSTRACT

Crime is a social phenomenon with spatial patterns influenced by demographic and socio-economic characteristics of regions. However, the availability of crime data at small-area levels is often limited, making detailed crime risk estimation difficult. This study aims to develop a crime analysis framework based on *Small Area Estimation* (SAE), *Random Forest Regressor*, and *Rough Set Theory* to estimate and map crime risk based on available administrative data at the kapanewon level in the Special Region of Yogyakarta, Indonesia.

This study uses administrative crime data from the Regional Police of Yogyakarta and regional socio-demographic data from Statistics Indonesia and the Civil Registry Office for the period 2022–2025. District-level crime data were spatially disaggregated into kapanewon-level observations using a *population-weighted areal interpolation* approach. Small area estimation was conducted using the Fay–Herriot model with weighted synthetic regression to obtain *Empirical Best Linear Unbiased Prediction* (EBLUP) estimates. The resulting estimates were subsequently used as targets in a *Random Forest Regressor* model for crime risk prediction, while *Rough Set Theory* was applied to extract decision rules based on socio-demographic characteristics.

The results indicate that the application of Small Area Estimation improved the quality of economic crime risk estimates, with the coefficient of determination (R^2) increasing from 0.806 to 0.890 and the root mean squared error (RMSE) decreasing from 0.938 to 0.680. Socio-demographic variables consistently associated with crime risk include school participation rate, percentage of population with low educational attainment, dependency ratio, and sex ratio. The Random Forest model achieved its best predictive performance for economic crime risk, yielding an R^2 of 0.9058, an RMSE of 0.630, and a mean absolute error of 0.510, exceeding the $R^2 > 0.70$ threshold established in the research hypothesis. Furthermore, Rough Set Theory generated 26 decision rules with an average confidence of 0.942, enabling the identification of crime risk patterns based on regional socio-demographic characteristics. Spatial visualization revealed that urban and suburban areas of the Special Region of Yogyakarta, particularly parts of Sleman Regency and Yogyakarta City, tend to exhibit higher levels of economic crime risk than rural areas.

This study makes four scientific contributions: (i) a hybrid SAE–RF–RST framework for estimated crime risk mapping based on available administrative data at sub-district resolution that, to the best of our knowledge, has not previously been integrated within a coherent pipeline; (ii) a mass-preserving aggregation approach that ensures aggregate consistency between sub-district estimates and district totals by construction, addressing the spatial misalignment problem; (iii) translation of Random Forest predictions into interpretable IF–THEN decision rules via Rough Set Theory that can be validated by domain experts and used as input to a decision support system; and (iv) a baseline dataset for crime risk estimation and mapping for 78 sub-districts (kapanewon) of the Special Region of Yogyakarta for the 2022–2025 period that can serve as a benchmark for sub-district public security studies. The resulting framework is designed as a decision support system for law enforcement agencies in Yogyakarta, not as an automated prediction system.

Keywords—crime, Small Area Estimation, Fay–Herriot, Random Forest, Rough Set Theory, crime risk mapping.