KERANGKA KERJA HIBRIDA K-MEANS–PCA DALAM PREDIKSI KEJAHATAN GEOSPASIAL DENGAN MODEL PEMBELAJARAN MESIN BERBASIS ATTENTION

LAPORAN NON-REGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

Disusun oleh :

Muhammad Hilmy Hafizh

22.11.5173

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

KERANGKA KERJA HIBRIDA K-MEANS–PCA DALAM PREDIKSI KEJAHATAN GEOSPASIAL DENGAN MODEL PEMBELAJARAN MESIN BERBASIS ATTENTION

LAPORAN NON-REGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

Disusun oleh :

Muhammad Hilmy Hafizh

22.11.5173

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

JALUR NON-REGULER

KERANGKA KERJA HIBRIDA K-MEANS–PCA DALAM PREDIKSI

yang disusun dan diajukan oleh

Muhammad Hilmy Hafizh

22.11.5173

telah disetujui oleh Dosen Pembimbing

pada tanggal 26-02-2026

Dosen Pembimbing,

Uyock Anggoro Saputro, S.Kom., M.Kom.,

NIK. 190302419

HALAMAN PENGESAHAN

JALUR NON-REGULER

KERANGKA KERJA HIBRIDA K-MEANS–PCA DALAM PREDIKSI KEJAHATAN GEOSPASIAL DENGAN MODEL PEMBELAJARAN MESIN BERBASIS ATTENTION

yang disusun dan diajukan oleh

Muhammad Hilmy Hafizh

22.11.5173

Telah dipertahankan di depan Dewan Penguji

pada tanggal 26-02-2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Eli Pujastuti, S.Kom., M.Kom. **NIK. 190302227**	
Arifiyanto Hadinegoro, S.Kom., M.T. **NIK. 190302289**	
Uyock Anggoro Saputro, S.Kom., M.Kom. **NIK. 190302419**	

Laporan ini telah diterima sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Komputer

Tanggal 26-02-2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof.Dr.Kusrini, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah in.

Nama mahasiswa : Muhammad Hilmy Hafizh

NIM : 22.11.5173

Menyatakan bahwa Laporan dengan judul berikut:

Kerangka Kerja Hibrida K-Means–PCA dalam Prediksi Kejahatan Geospasial dengan Model Pembelajaran Mesin Berbasis Attention

Dosen Pembimbing : Uyock Anggoro Saputro, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik. baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Pebruari 2026

Muhammad Hilmy Hafizh

HALAMAN PERSEMBAHAN

Dengan segala kerendahan dan ketulusan hati penulis mempersembahkan karya ini kepada:

1. Program Studi Informatika, Universitas Amikom Yogyakarta, yang telah memberikan kesempatan bagi saya untuk menempuh dan memperoleh ilmu.
2. Ayah dan ibuku tercinta yang selalu memberikan doanya dan semangat.
3. Kakak dan adikku tercinta yang selalu mendukung, dan memberikan semangat dan doanya.
4. Seluruh rekan kuliah khususnya Angkatan 2022 atas dukungan dan sharing ilmunya hingga terselesaikannya proposal ini.
5. Semua pihak yang telah memberikan dukungan, membantu penulis dan tidak dapat disebutkan satu persatu.

Semoga Tuhan Yang Maha Esa senantiasa memberikan limpahan rahmat dan karunia atas segala kebaikannya. Akhirnya penulis berharap semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan segenap pembaca.

Yogyakarta, 26-02-2026

Muhammad Hilmy Hafizh

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah Swt. yang telah memberikan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan proposal skripsi yang berjudul "**Kerangka Kerja Hibrida K-Means–PCA dalam Prediksi Kejahatan Geospasial dengan Model Pembelajaran Mesin Berbasis Attention**". Dengan segala kerendahan dan ketulusan hati penulis menyampaikan penghargaan dan ucapan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr.Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi S1 Informatika, Universitas Amikom Yogyakarta.
4. Bapak Uyock Anggoro Saputro, S.Kom., M.Kom., selaku Pembimbing Akademik, Universitas Amikom Yogyakarta, dan pembimbing Tugas Akhir yang telah memberikan bimbingan dan arahan selama penulis menempuh Pendidikan.

Semoga Tuhan Yang Maha Esa senantiasa memberikan limpahan rahmat dan karunia atas segala kebaikannya. Untuk mengembangkan karya ini, dengan segala kerendahan hati penulis mengharapkan kritik dan saran untuk hasil yang lebih baik. Akhirnya penulis berharap semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan segenap pembaca.

Yogyakarta,26-02-2026

Muhammad Hilmy Hafizh

DAFTAR ISI

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

X	=	Koordinat longitude (bujur)
y	=	Koordinat latitude (lintang)
μ	=	Nilai rata-rata (mean) dari suatu fitur
σ	=	Simpangan baku (standard deviation) dari suatu fitur
n	=	Jumlah data atau observasi
k	=	Jumlah cluster pada algoritma K-Means
Ci	=	Cluster ke-i
φ	=	Fungsi transformasi PCA yang memetakan data asli ke ruang komponen utama
$\varphi(x)$	=	Proyeksi data x ke ruang fitur berdimensi lebih rendah menggunakan PCA
Z	=	Matriks fitur hasil transformasi PCA
m	=	Jumlah komponen utama (principal components) yang dipilih
λ	=	Nilai eigen (eigenvalue) dari matriks kovarians
v	=	Vektor eigen (eigenvector)
B	=	Jumlah pohon keputusan pada algoritma Random Forest
Y	=	Nilai prediksi yang dihasilkan model
y	=	Nilai aktual atau label sebenarnya
P(c)	=	Probabilitas suatu data termasuk ke dalam kelas c
TP	=	True Positive
FP	=	False Positive
FN	=	False Negative
TN	=	True Negative
ML	=	Machine Learning
DL	=	Deep Learning
PCA	=	Principal Component Analysis
XGBoost	=	Extreme Gradient Boosting
RF	=	Random Forest
LR	=	Logistic Regression
XAI	=	Explainable Artificial Intelligence
IDS	=	Intrusion Detection System
IoT	=	Internet of Things
GIS	=	Geographic Information System
BiGRU	=	Bidirectional Gated Recurrent Unit
PC	=	Principal Component

DAFTAR ISTILAH

Crime Hotspot	= Wilayah geografis dengan konsentrasi kejadian kriminal yang tinggi
Geospatial Data	= Data yang memiliki referensi lokasi geografis
Spatio-Temporal Analysis	= Analisis data yang mempertimbangkan dimensi ruang dan waktu
Clustering	= Proses pengelompokan data berdasarkan kemiripan karakteristik
K-Means Clustering	= Algoritma unsupervised learning untuk mengelompokkan data ke dalam k cluster
Dimensionality Reduction	= Proses pengurangan jumlah fitur tanpa menghilangkan informasi penting
Principal Component Analysis (PCA)	= Metode statistik untuk mentransformasikan fitur ke komponen baru
Overfitting	= Kondisi ketika model terlalu menyesuaikan data latih
Generalization	= Kemampuan model dalam memprediksi data baru
Classification	= Proses penentuan kelas data
Confusion Matrix	= Tabel evaluasi performa model klasifikasi
Accuracy	= Rasio prediksi benar terhadap seluruh data
Precision	= Ketepatan prediksi kelas positif
Recall	= Kemampuan mendeteksi seluruh data kelas positif
F1-Score	= Rata-rata harmonik precision dan recall
Predictive Policing	= Pendekatan kepolisian berbasis prediksi berbasis data
Smart City	= Konsep kota berbasis teknologi untuk layanan publik

INTISARI

Kejahatan perkotaan di daerah padat penduduk menimbulkan tantangan kritis bagi manajemen keselamatan publik. Masalah ini semakin diperparah dengan kesulitan mendeteksi pola spasial tersembunyi dan menangani data geospasial dimensi tinggi, yang sering membatasi efektivitas model prediktif tradisional. Studi ini bertujuan untuk mengembangkan kerangka prediksi kejahatan geospasial yang kuat yang mampu mengklasifikasikan daerah berisiko tinggi, sedang, dan rendah di seluruh Pulau Jawa, Indonesia. Model yang diusulkan mengintegrasikan pengelompokan K-Means untuk mengungkap hotspot spasial laten, Analisis Komponen Utama (PCA) untuk mengurangi dimensi fitur, dan tiga pengklasifikasi yang diawasi *Logistic Regression*, *Random Forest*, dan XGBoost untuk mengevaluasi kinerja prediktif. Himpunan data terdiri dari insiden kejahatan geokode yang diperkaya dengan atribut demografis dan kontekstual, dibagi menjadi 80% pelatihan dan 20% subset pengujian. Hasilnya menunjukkan bahwa K-Means secara efektif mengidentifikasi hotspot spasial, dan PCA mempertahankan lebih dari 98% varians sambil meningkatkan efisiensi komputasi. Regresi Logistik mencapai generalisasi yang lebih stabil (Akurasi, pelatihan 0,63, tes 0,52), sedangkan Random Forest dan XGBoost, meskipun akurasi pelatihan sempurna (1,00), menderita overfitting pada data pengujian. Analisis matriks kebingungan lebih lanjut menegaskan bahwa pengoptimalan fitur hanya menghasilkan peningkatan marjinal. Pekerjaan di masa depan akan berfokus pada mitigasi *overfitting* dalam model ansambel, menggabungkan fitur sosial-ekonomi dan temporal yang lebih kaya, memvalidasi pada kumpulan data dunia nyata yang lebih besar, dan mengintegrasikan AI yang Dapat Dijelaskan (XAI) untuk memastikan interpretabilitas dan keadilan dalam pemolisian prediktif dan aplikasi kota pintar.

Kata kunci: Deteksi Titik Rawan Kejahatan; Pengurangan Dimensi; Pembelajaran Mesin untuk Keamanan Publik; Kepolisian Prediktif; Analisis Spasio-Temporal;

ABSTRACT

Urban crime in densely populated regions poses critical challenges for public safety management. This problem is further compounded by the difficulty of detecting hidden spatial patterns and handling high dimensional geospatial data, which often limits the effectiveness of traditional predictive models. This study aims to develop a robust geospatial crime prediction framework capable of classifying high, medium, and low risk areas across Java Island, Indonesia. The proposed model integrates K-Means clustering to uncover latent spatial hotspots, Principal Component Analysis (PCA) to reduce feature dimensionality, and three supervised classifiers Logistic Regression, Random Forest, and XGBoost to evaluate predictive performance. The dataset consists of geocoded crime incidents enriched with demographic and contextual attributes, partitioned into 80% training and 20% testing subsets. The results indicate that K-Means effectively identified spatial hotspots, and PCA preserved over 98% of variance while improving computational efficiency. Logistic Regression achieved more stable generalization (Accuracy, 0.63 training, 0.52 test), while Random Forest and XGBoost, despite perfect training accuracy (1.00), suffered from overfitting on test data. Confusion matrix analysis further confirmed that feature optimization yielded only marginal improvements. Future work will focus on mitigating overfitting in ensemble models, incorporating richer socio-economic and temporal features, validating on larger real-world datasets, and integrating Explainable AI (XAI) to ensure interpretability and fairness in predictive policing and smart city applications.

Keyword: Crime Hotspot Detection; Dimensionality Reduction; Machine Learning for Public Safety; Predictive Policing; Spatio-Temporal Analytics;