

## **Skripsi**

# **Klasifikasi Bangunan Permukiman dengan Algoritma XGBoost Berdasarkan Morfologi Urban dan Data Atribut OpenStreetMap (OSM)**

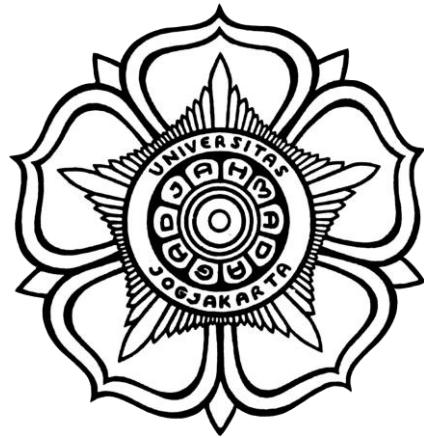


**Diajukan Oleh :**  
**Sultan Dzaki Zhulfalian**  
**NIM: 20/460278/TK/50867**

**Program Sarjana Program Studi Teknik Geodesi**  
**Departemen Teknik Geodesi**  
**Fakultas Teknik**  
**Universitas Gadjah Mada**  
**2024**

## **Final Project**

# **Classification of Residential Buildings with the XGBoost Algorithm Based on Urban Morphology and OpenStreetMap (OSM) Attribute Data**



**Submitted By :**

**Sultan Dzaki Zhulfaliant**

**NIM: 20/460278/TK/50867**

**Undergraduate Study Program Of Geodetic Engineering**

**Department Of Geodetic Engineering**

**Faculty Of Engineering**

**Universitas Gadjah Mada**

**2024**

**Klasifikasi Bangunan Permukiman dengan Algoritma XGBoost  
Berdasarkan Morfologi Urban dan Data Atribut  
OpenStreetMap (OSM)**

**Skripsi**

Untuk memenuhi persyaratan dalam mencapai derajat Sarjana  
Teknik Program Studi Teknik Geodesi



Diajukan Oleh :  
Sultan Dzaki Zhulfaliant  
NIM: 20/460278/TK/50867

Program Sarjana Program Studi Teknik Geodesi  
Departemen Teknik Geodesi  
Fakultas Teknik  
Universitas Gadjah Mada

2024

**Classification of Residential Buildings with the XGBoost  
Algorithm Based on Building Morphological Characteristics  
and OpenStreetMap (OSM) Attribute Data**

**Final Project**

Submitted as a partial fulfillment of requirements for obtaining the  
degree of Bachelor of Engineering in the Study Program of  
Geodetic Engineering



Submitted By :  
Sultan Dzaki Zhulfalian  
NIM: 20/460278/TK/50867

Undergraduate Study Program in Geodetic Engineering  
Department of Geodetic Engineering  
Faculty of Engineering  
Gadjah Mada University

2024

Skripsi

Klasifikasi Bangunan Permukiman dengan Algoritma XGBoost Berdasarkan  
Morfologi Urban dan Data Atribut OpenStreetMap (OSM)

Dipersiapkan dan disusun oleh :

Sultan Dzaki Zhulfalian  
NIM: 20/460278/TK/50867

telah dipertahankan di depan Dewan Pengaji  
pada tanggal 10 Juli 2024

Susunan Dewan Pengaji :

Ketua/Pembimbing

Prof. Ir. Trias Aditya Kurniawan M.  
S.T., M.Sc., Ph.D., IPU  
NIP. 197504222002121002

Anggota 1

Ir. Rochmad Muryamto, M.Eng. Sc.  
NIP. 196601111991031001

Anggota 2

Heri Sutanta S.T., M.Sc., Ph.D.  
NIP. 197605232002121002

Skripsi ini telah diterima sebagai salah satu persyaratan untuk mencapai derajat  
Sarjana Teknik pada Program Studi Teknik Geodesi

Tanggal ..31 JULI 2024.....

Ketua Program Studi Sarjana Teknik Geodesi

Heri Sutanta S.T., M.Sc., Ph.D.

NIP. 197605232002121002

Menyatakan, Ketua Departemen Teknik Geodesi

Prof. Ir. Trias Aditya Kurniawan M, S.T., M.Sc., Ph.D., IPU.  
NIP. 197504222002121002

## **HALAMAN PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : Sultan Dzaki Zhulfalian  
NIM : 20/460278/TK/50867  
Tahun Terdaftar : 2020  
Program Studi : Sarjana Teknik Geodesi  
Fakultas : Teknik

Menyatakan bahwa dalam dokumen ilmiah skripsi ini tidak terdaftar bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu Lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan sumbernya secara lengkap dalam daftar pustaka.

Dengan demikian saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi dan apabila dokumen ilmiah skripsi ini kemudian hari terbukti merupakan plagiasi dari karya penulis lain dan/atau dengan sengaja mengajukan karya atau pendapat yang merupakan hasil karya penulis lain, maka penulis bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Yogyakarta, 25 Juli 2024



Sultan Dzaki Zhulfalian

## **HALAMAN PERSEMBAHAN**

Skripsi ini saya persembahkan untuk diri saya di masa lalu, saat ini,  
dan di masa yang akan datang.

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan yang Maha Esa, atas segala limpahan rahmat, berkah, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "**Klasifikasi Bangunan Permukiman dengan Algoritma XGBoost Berdasarkan Morfologi Urban dan Data Atribut OpenStreetMap (OSM)**". Skripsi ini disusun untuk memenuhi persyaratan dalam memperoleh gelar sarjana di Program Studi Sarjana Teknik Geodesi, Fakultas Teknik, Universitas Gadjah Mada.

Selama penyusunan skripsi ini, penulis mendapatkan bimbingan, saran, bantuan, hingga dukungan dari berbagai pihak. Oleh karena itu, izinkan penulis mengucapkan rasa hormat dan terima kasih kepada Bapak Prof. Ir. Trias Aditya Kurniawan Muhammad, S.T., M.Sc., Ph.D., IPU selaku Ketua Departemen Teknik Geodesi dan dosen pembimbing skripsi yang telah memberikan bimbingan serta dukungan kepada penulis sehingga skripsi ini dapat terselesaikan dengan baik, serta Bapak dan Ibu dosen Program Studi Sarjana Teknik Geodesi yang telah memberi banyak ilmu pengetahuan, serta seluruh karyawan Departemen Teknik Geodesi yang telah memberikan bantuan selama penulis menempuh studi di bangku perkuliahan.

Penulis juga berterima kasih kepada berbagai pihak yang sudah mendukung penulis dalam menyelesaikan skripsi ini. Kepada keluarga penulis, Perwana Auliant, Yuni Nur Wijayanti, dan Syafila Balqis Putri Zhulfalian yang senantiasa mendukung dan memberikan penulis kepercayaan penuh untuk menjalani pendidikan di Yogyakarta. Kepada Cintia Lirifa Asmarani yang senantiasa mendukung dan memberikan masukan atas penulisan skripsi ini. Kepada teman-teman Inswapala dan SMA lainnya yang selalu menjadi tempat berpulang di Kota Bekasi. Kepada rekan-rekan Kowah, Kontrakan Dito, Trah Kumpulrejo, dan Geodesi 20 yang telah menjadi teman dan sahabat selama menjalani pendidikan di bangku perkuliahan ini. Dan kepada mentor-mentor di PT. Bhumi Vartha Technology yang telah menjadi sumber inspirasi dan pembimbing saya hingga saat ini.

Penulis menyadari ketidaksempurnaan dari naskah skripsi ini. Maka dari itu, dibutuhkan adanya kritik dan masukkan dari pembaca untuk terus mengembangkan ide dan konsep yang terkandung di dalam naskah ini. Penulis berharap, penelitian yang telah dituntaskan dapat menginspirasi pengaplikasian *machine learning* di dalam perencanaan wilayah perkotaan.

Yogyakarta, 3 Juli 2024

Penulis

## INTISARI

Pertumbuhan penduduk yang signifikan dan urbanisasi yang terus berlangsung di Indonesia telah menimbulkan berbagai tantangan perkotaan, seperti kemiskinan, peningkatan angka kriminalitas, dan munculnya permukiman kumuh. Untuk mengatasi tantangan ini, data akurat mengenai distribusi bangunan permukiman dan non-permukiman sangat penting untuk mendukung manajemen ruang kota yang lebih efektif dan terarah. Penelitian ini bertujuan mengembangkan model klasifikasi bangunan permukiman berdasarkan morfologi bangunan dan atribut OpenStreetMap (OSM) dengan menggunakan algoritma Extreme Gradient Boosting (XGBoost), untuk memfasilitasi perolehan data distribusi permukiman yang dapat diterapkan di berbagai wilayah di Indonesia.

Penelitian ini menggunakan data bangunan pada wilayah Kota Jakarta Selatan dalam pembuatan model dan wilayah Kota Jakarta Utara untuk menguji kapabilitas model yang telah dibuat. Pada penelitian ini, morfologi bangunan, seperti bentuk fisik, ukuran, dan struktur bangunan, digunakan sebagai fitur penting untuk mengidentifikasi pola distribusi permukiman yang berbeda. Selain itu, atribut OSM, seperti tipe jalan dan bangunan, memberikan konteks tambahan yang diperlukan untuk memperkaya pemodelan spasial. Pemilihan XGBoost sebagai algoritma *machine learning* didasarkan pada kecepatan pemrosesan dan efisiensinya yang tinggi, dengan akurasi yang sebanding dengan algoritma berbasis *Decision Tree* lainnya. Dalam pembuatan model ini, digunakan teknik *spatial cross-validation* untuk memastikan bahwa model dapat digunakan secara akurat pada data wilayah yang belum pernah dilihat sebelumnya.

Studi ini menunjukkan bahwa model XGBoost yang dioptimalkan dengan data morfologi urban dan atribut OSM mencapai akurasi klasifikasi sebesar 91% dan *F1-Score* rata-rata sebesar 78,6%. Penambahan data lokal sebesar 10% meningkatkan akurasi dan *F1-Score* rata-rata menjadi 91,8% dan 81,5%. Oleh karena itu, bahkan informasi tambahan yang minim mengenai wilayah yang akan diprediksi dapat meningkatkan kemampuan model untuk mengklasifikasi bangunan permukiman dan non-permukiman di wilayah tersebut.

**Kata Kunci:** Permukiman, XGBoost, Morfologi Urban, OpenStreetMap (OSM)

## **ABSTRACT**

*Significant population growth and ongoing urbanization in Indonesia have posed various urban challenges, such as poverty, increased crime rates, and the emergence of slums. To address these challenges, accurate data on the distribution of residential and non-residential buildings are crucial for supporting more effective and targeted urban space management. This research aims to develop a model for classifying residential buildings based on building morphology and OpenStreetMap (OSM) attributes using the Extreme Gradient Boosting (XGBoost) algorithm, to facilitate the acquisition of residential distribution data applicable to various regions in Indonesia.*

*This research uses building data from South Jakarta for model development and North Jakarta to test the model's capabilities. In this study, building morphology, such as physical shape, size, and structure, are used as key features to identify different residential distribution patterns. Additionally, OSM attributes, such as road and building types, provide the necessary context to enrich spatial modeling. The selection of XGBoost as the machine learning algorithm is based on its high processing speed and efficiency, with accuracy comparable to other Decision Tree-based algorithms. Spatial cross-validation techniques are used to ensure the model can accurately be applied to unseen regional data.*

*This study shows that the XGBoost model optimized with urban morphology data and OSM attributes achieves a classification accuracy of 91% and an average F1-Score of 78.6%. Adding 10% local data increases accuracy and the average F1-Score to 91.8% and 81.5%, respectively. Therefore, even minimal additional information about the predicted area can enhance the model's ability to classify residential and non-residential buildings in the region.*

**Keywords:** Residential, XGBoost, Urban Morphology, OpenStreetMap (OSM)

# DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN .....	v
HALAMAN PERNYATAAN.....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR.....	viii
INTISARI .....	x
<i>ABSTRACT</i> .....	xi
DAFTAR ISI .....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL .....	xiv
DAFTAR LAMPIRAN .....	xvii
BAB I.....	1
I.1. Latar Belakang .....	1
I.2. Rumusan Masalah.....	2
I.3. Tujuan Penelitian .....	3
I.4. Pertanyaan Penelitian.....	3
I.5. Ruang Lingkup .....	4
I.6. Manfaat Penelitian .....	4
I.7. Tinjauan Pustaka.....	5
BAB II .....	9
II.1. Permukiman.....	9
II.2. Informasi Geospasial .....	10
II.3. <i>Tobler First Law</i> (TFL) .....	11
II.4. OpenStreetMap (OSM).....	12
II.5. Morfologi Urban.....	14
II.5.1. Morfometrik Urban.....	15
II.5.2. Elemen-elemen Morfometrik.....	15
II.5.3. Karakter Morfometrik.....	17
II.6. <i>Machine Learning</i> .....	18
II.6.1. <i>Hyperparameter</i> .....	19
II.6.2. <i>Cross Validation</i> .....	20

II.6.3. <i>Spatial Cross Validation</i> .....	21
II.7. <i>Extreme Gradient Boosting</i> (XGBoost) .....	22
II.8. Evaluasi Model Klasifikasi.....	24
BAB III.....	27
III.1. Lokasi Penelitian .....	27
III.2. Peralatan dan Bahan .....	28
III.2.1. Peralatan Penelitian .....	28
III.2.2. Bahan Penelitian.....	28
III.3. Tahapan Penelitian .....	29
III.3.1. Studi Literatur .....	29
III.3.2. Persiapan Alat dan Bahan.....	32
III.3.3 Data <i>Pre-Processing</i> .....	34
III.3.4. Pengolahan Data.....	35
III.3.5. Pembuatan Model Klasifikasi .....	39
III.3.6. Analisis dan Evaluasi Model Klasifikasi.....	42
BAB IV HASIL DAN PEMBAHASAN.....	44
IV.1. Pengolahan Data .....	44
IV.1.1. Data yang Digunakan.....	44
IV.1.2 Perhitungan Karakter Morfometrik.....	47
IV.2. Hasil dan Evaluasi Model XGBoost .....	50
IV.2.1. Pembuatan Model XGBoost Standar .....	51
IV.2.2. Optimasi <i>Hyperparameter</i> .....	53
IV.2.3. Pengujian Model .....	56
IV.2.4. Hasil dan Evaluasi Model .....	59
IV.3. <i>Feature Importances</i> .....	65
IV.4. Pemanfaatan Model Lebih Lanjut.....	66
BAB V KESIMPULAN DAN SARAN .....	67
V.1. Kesimpulan.....	67
V.2. Saran.....	68
DAFTAR PUSTAKA.....	69
LAMPIRAN .....	75

## DAFTAR GAMBAR

Gambar II.1. Visualisasi data raster, vektor, dan dunia nyata (Sumber: Saab, 2003).....	11
Gambar II.2. Perbandingan plot tradisional (kiri) dan plot modern (kanan) di Glasgow, Skotlandia yang berasal dari sumber yang berbeda.....	16
Gambar II.3. Perbedaan geometri antara bangunan (kiri) dan sel teselasi (kanan) (Sumber: Fleischman, 2019). .....	16
Gambar II.4. Visualisasi strategi K-Fold Cross Validation (Berrar, 2019) .....	21
Gambar II.5. Ilustrasi dari Cross Validation standar dengan Cross Validation spasial (Sumber: Lovelace dkk., 2021). .....	21
Gambar II.6. Bagan alir dari algoritma XGBoost (Sumber: Nguyen dkk., 2021).....	22
Gambar II.7. Perbedaan algoritma sparsity aware dengan algoritma biasa (Sumber: Chen dan Guestrin, 2016). .....	24
Gambar III.1. Lokasi penelitian.....	27
Gambar III.2 Bagan alir penelitian. ....	30
Gambar III.3 Bagan alir penelitian (lanjutan 1). .....	31
Gambar III.4 Bagan alir penelitian (lanjutan 2). .....	32
Gambar III.5. Script yang digunakan untuk mengunduh data bangunan OSM. .....	33
Gambar III.6. Script yang digunakan untuk mengunduh data jaringan jalan OSM. .....	34
Gambar III.7. Potongan script yang digunakan pada tahap pre-processing data bangunan.	35
Gambar III.8. Script yang digunakan untuk membuat sel teselasi. .....	36
Gambar III.9. Potongan script yang digunakan untuk mengkalkulasi karakter morfometrik. ....	37
Gambar III.10. Tampilan kelas permukiman (kiri) dan non-permukiman (kanan) di Kecamatan Menteng, Jakarta Pusat. ....	38
Gambar III.11. Script yang digunakan untuk memberikan label pada bangunan. ....	39
Gambar III.12. Diagram alir pembuatan model XGBoost. ....	40
Gambar III.13. Script yang digunakan untuk membuat spatial CV.....	41
Gambar III.14. Script yang digunakan untuk membuat model XGBoost. ....	41
Gambar III.15. Script yang digunakan untuk membuat model XGBoost (Lanjutan). .....	42
Gambar IV.1. Distribusi jumlah bangunan permukiman dan non-permukiman di wilayah Jakarta Selatan. ....	45

Gambar IV.2. Distribusi jumlah bangunan permukiman dan non-permukiman di wilayah Jakarta Utara .....	45
Gambar IV.3. Contoh distribusi bangunan permukiman dan non-permukiman di Kecamatan Kelapa Gading, Jakarta Utara .....	46
Gambar IV.4. Contoh geometri bangunan yang tidak sesuai. ....	47
Gambar IV.5. Visualisasi karakter morfometrik .....	49
Gambar IV.6. Perbedaan <i>learning curve</i> untuk setiap nilai ‘ <i>learning_rate</i> ’ (Sumber: CS231n, 2024).....	55
Gambar IV.7. Hasil prediksi model XGBoost teroptimasi tanpa data lokal di Kecamatan Koja, Jakarta Utara .....	60
Gambar IV.8. Hasil prediksi model XGBoost teroptimasi dengan 10% data lokal di Kecamatan Koja, Jakarta Utara. ....	61
Gambar IV.9. Dampak penambahan data lokal pada bangunan non-permukiman. ....	62
Gambar IV.10. Dampak penambahan data lokal pada bangunan permukiman.....	63
Gambar IV.11. Perubahan metrik akurasi tanpa dan dengan data lokal. ....	63
Gambar IV.12. Perubahan metrik balanced accuracy tanpa dan dengan data lokal. ....	64
Gambar IV.13. Perubahan metrik F1-Score rata-rata tanpa dan dengan data lokal.....	64

## **DAFTAR TABEL**

Tabel II.1 Matriks konfusi .....	25
Tabel III.1. Daftar bahan penelitian.....	28
Tabel III.2. Daftar bahan penelitian (lanjutan). ....	29
Tabel III.2 Daftar kelas untuk setiap jenis penggunaan lahan.....	38
Tabel IV.1. Daftar kelompok karakter morfometrik yang digunakan.....	48
Tabel IV.2. Pembagian CV spasial .....	51
Tabel IV.3. Nilai standar dari hyperparameter XGBoost.....	52
Tabel IV.4. Hasil klasifikasi model XGBoost standar pada data validasi. ....	52
Tabel IV.5. Rentang parameter yang dicari. ....	53
Tabel IV.6. Kombinasi hyperparameter yang dipilih.....	54
Tabel IV.7. Hasil klasifikasi model XGBoost teroptimasi pada data validasi. ....	56
Tabel IV.8. Hasil klasifikasi model XGBoost pada data uji tanpa data lokal. ....	57
Tabel IV.9. Total bangunan dan jumlah bangunan yang diikutsertakan ke dalam data latih di setiap percobaan. ....	57
Tabel IV.10. Hasil klasifikasi model XGBoost pada data uji dengan data lokal. ....	58
Tabel IV.11. Confusion Matrix Koja tanpa data lokal. ....	62
Tabel IV.12. Confusion Matrix Koja dengan 10% data lokal.....	62
Tabel IV.11. Skor feature importance dari model XGBoost teroptimasi. ....	65

## **DAFTAR LAMPIRAN**

Lampiran A <i>Script</i> Pemrosesan Data dan Model Klasifikasi Bangunan Permukiman .....	76
Lampiran B Daftar Definisi Karakter Morfometrik yang Digunakan.....	77
Lampiran C Signifikansi Karakter Morfometrik .....	89
Lampiran D Visualisasi Hasil Prediksi Model dengan 10% Data Lokal .....	95

# **BAB I**

## **PENDAHULUAN**

### **I.1. Latar Belakang**

Indonesia, sebagai negara dengan populasi terbesar keempat di dunia, mengalami pertumbuhan penduduk yang signifikan. Jumlah penduduk tumbuh menjadi 270,20 juta jiwa dari 237,63 juta jiwa dalam kurun waktu 10 tahun (World Bank, 2022; BPS, 2021). Pertumbuhan ini disebabkan oleh beberapa faktor, seperti meningkatnya jumlah kelahiran, menurunnya angka kematian, dan migrasi penduduk karena faktor ekonomi (Suartha, 2016). Fenomena urbanisasi di Indonesia terjadi secara konstan, dengan tingkat urbanisasi rata-rata sebesar 56,7% di seluruh Indonesia dan 71,52% di Pulau Jawa (BPS Indonesia, 2021), artinya dari 100 orang yang tinggal di Indonesia, 57 orang berada di wilayah perkotaan. Tingkat urbanisasi dan pertumbuhan penduduk yang tinggi dapat berpengaruh pada perubahan struktur dan bentuk permukiman (Acheampong, 2018).

Menurut Prayojana dkk. (2020), pertumbuhan penduduk di wilayah perkotaan, yang disebabkan oleh tingkat urbanisasi dan peningkatan jumlah penduduk secara alami, dapat memicu berbagai permasalahan seperti kemiskinan, tingginya angka kriminalitas, dan munculnya permukiman kumuh di lahan kosong milik pemerintah. Pertumbuhan penduduk yang tidak terkendali dapat merusak strategi perencanaan kota, terutama dalam penyediaan perumahan yang layak (Harahap, 2013). Kondisi ini menjadi tantangan bagi pemerintah dalam manajemen dan pengelolaan ruang kota yang efisien.

Untuk mendukung manajemen dan pengelolaan kawasan permukiman yang lebih baik, dibutuhkan data akurat tentang persebaran bangunan permukiman dan non-permukiman. Sturrock dkk. (2018) menunjukkan bahwa data tersebut dapat dimanfaatkan untuk berbagai keperluan, seperti pelayanan kesehatan, mitigasi risiko bencana, pembangunan fasilitas umum, atau peningkatan kualitas hidup. Data persebaran permukiman juga dapat digunakan untuk melakukan disagregasi data populasi ke unit spasial yang lebih kecil, atau disebut *dasymetric mapping* (Kühnl dkk., 2023; Wright, 1936). Dengan mengenal distribusi populasi yang lebih baik, pemerintah dapat meningkatkan kualitas hidup masyarakat melalui program kerja yang lebih tepat sasaran.

Berkembangnya teknologi ekstraksi tapak bangunan dari citra resolusi tinggi dan ketersediaan data geospasial secara luas melalui lembaga resmi dan *Volunteered Geographic Information* (VGI) memudahkan pemanfaatan data bangunan untuk perencanaan perkotaan, terutama di negara berkembang (Atwal dkk., 2022; Sturrock dkk., 2018; Jochem dan Tatem, 2021). Momepy, *library* Python yang diperkenalkan oleh Fleischmann (2019), dapat digunakan untuk analisis bentuk perkotaan secara kuantitatif dengan menghitung nilai karakter morfologi bangunan. Nilai morfologi ini telah digunakan dalam berbagai penelitian untuk prediksi dan analisis dengan bantuan *machine learning*, seperti prediksi tinggi bangunan (Milojevic-Dupont dkk., 2020), identifikasi bangunan permukiman dan non-permukiman (Atwal dkk., 2022), atau prediksi usia bangunan (Nachtigall dkk., 2023).

Penelitian ini bertujuan untuk membuat sebuah model klasifikasi bangunan permukiman berdasarkan karakter morfologi bangunan dan atribut OpenStreetMap (OSM) menggunakan algoritma *Extreme Gradient Boosting* (XGBoost), sebagai upaya pengadaan data persebaran permukiman yang dapat digunakan di berbagai wilayah di Indonesia. Pemilihan XGBoost sebagai algoritma *machine learning* didasarkan pada kecepatan pemrosesan dan efisiensinya yang tinggi, dengan akurasi yang relatif setara dengan algoritma berbasis *Decision Tree* lainnya (Chen dan Guestrin, 2016), menjadikannya pilihan yang tepat untuk menganalisis pola spasial dari ratusan ribu data bangunan. Hasil klasifikasi yang diperoleh dapat dimanfaatkan oleh pemerintah pusat atau daerah sebagai bahan pertimbangan dalam pengambilan keputusan yang berkaitan dengan perumahan, seperti pembangunan fasilitas umum yang lebih mudah diakses dari lokasi permukiman, identifikasi permukiman kumuh yang membutuhkan intervensi atau bantuan sosial, serta analisis distribusi populasi di suatu wilayah menggunakan pemetaan dasimetrik.

## I.2. Rumusan Masalah

Pertumbuhan penduduk yang signifikan dan urbanisasi yang terus berlangsung di Indonesia, telah memunculkan berbagai tantangan perkotaan, misalnya kemiskinan, peningkatan angka kriminalitas, dan munculnya permukiman kumuh. Pertumbuhan penduduk yang tidak terkendali ini mengancam strategi perencanaan kota, terutama dalam penyediaan perumahan yang layak, dan sering menyebabkan pembangunan permukiman yang tidak sesuai dengan perencanaan tata guna lahan. Dalam mengatasi tantangan ini, data persebaran bangunan permukiman dan non-permukiman yang akurat menjadi sangat

penting untuk mendukung manajemen dan pengelolaan ruang kota yang lebih efektif dan tepat sasaran.

Sebagai upaya penyediaan data tersebut, telah berkembang berbagai teknologi untuk mendukung pengambilan keputusan dan perencanaan kawasan permukiman, salah satunya adalah penggunaan teknologi *remote sensing*, baik melalui citra satelit resolusi tinggi maupun foto udara. Namun, salah satu kelemahan dari teknologi ini adalah keterbatasan cakupan wilayah yang relatif kecil, serta ketersediaan data yang terbatas dan mahal. Keterbatasan-keterbatasan ini menjadi salah satu tantangan dalam upaya perencanaan dan pembangunan kawasan permukiman di berbagai wilayah di Indonesia. Oleh karena itu, diperlukan metode penelitian kawasan permukiman yang mudah direplikasi di seluruh wilayah di Indonesia.

### **I.3. Tujuan Penelitian**

Tujuan dari penelitian ini dibagi menjadi tujuan utama dan tujuan spesifik. Tujuan utama penelitian ini adalah untuk mengembangkan model klasifikasi bangunan permukiman yang akurat, menggunakan nilai morfologi bangunan, atribut OSM, dan algoritma XGBoost. Model ini bertujuan untuk memfasilitasi pengadaan data persebaran permukiman yang dapat diterapkan di berbagai wilayah di Indonesia. Adapun tujuan spesifik dari penelitian ini adalah:

1. Melakukan perhitungan nilai-nilai morfologi bangunan dengan bantuan *library* Momepy sebagai fitur utama dalam model klasifikasi yang dibuat.
2. Melakukan validasi dan evaluasi model yang dikembangkan untuk memastikan akurasi dan efektivitas klasifikasi bangunan permukiman dan non-permukiman.
3. Membuat model klasifikasi dengan algoritma XGBoost berdasarkan nilai morfologi bangunan untuk membedakan antara bangunan permukiman dan non-permukiman.

### **I.4. Pertanyaan Penelitian**

Adapun pertanyaan yang dicoba untuk dijawab dalam penelitian ini adalah sebagai berikut:

1. Bagaimana karakter morfologi bangunan dan atribut OSM dapat digunakan sebagai data latih dalam membuat model klasifikasi?
2. Bagaimana hasil dan akurasi dari model klasifikasi dengan algoritma XGBoost dalam memprediksi bangunan permukiman dan non-permukiman?

3. Apa fitur yang paling signifikan dalam membedakan antara bangunan permukiman dan non-permukiman?

### **I.5. Ruang Lingkup**

Ruang lingkup dari penelitian ini untuk memperjelas jangkauan penelitian adalah sebagai berikut:

1. Data yang digunakan merupakan data bangunan dan data jaringan jalan yang dirilis oleh OSM dan data penggunaan lahan dari Pemerintah Provinsi DKI Jakarta tahun 2020.
2. Pada penelitian ini, digunakan sel teselasi sebagai pengganti dari data bidang tanah. Penggunaan sel teselasi sendiri dipilih untuk mengatasi permasalahan ketersediaan data bidang tanah yang terbatas di berbagai wilayah di Indonesia.
3. Data penggunaan lahan digunakan sebagai *ground truth* dalam validasi dan evaluasi model klasifikasi serta acuan dalam pemberian label pada data bangunan.
4. Digunakan data bangunan Jakarta Selatan sebagai data latih dan data bangunan Jakarta Utara sebagai data uji.
5. Karakter morfologi bangunan yang digunakan adalah karakter yang berkaitan dengan dimensi, bentuk, distribusi, koneksi, intensitas, dan diversitas bangunan dan jaringan jalan.
6. Data atribut OSM yang digunakan adalah kelengkapan atribut OSM pada suatu bangunan.
7. Penelitian ini melakukan analisis perbandingan pengaturan parameter yang berbeda pada model algoritma XGBoost.

### **I.6. Manfaat Penelitian**

Bagi dunia ilmu pengetahuan, penelitian ini menguji aplikasi algoritma XGBoost dalam mengklasifikasikan bangunan permukiman berdasarkan nilai morfologi dari data geospasial. Ini merupakan kontribusi penting dalam bidang pengolahan data geospasial dan pengembangan model *machine learning*. Penelitian ini berpotensi memperluas pemahaman tentang cara menggunakan data morfologi bangunan dan atribut OSM untuk mengidentifikasi karakteristik permukiman, menawarkan metode baru atau memperbaiki metode yang sudah ada dalam analisis bentuk perkotaan. Ini juga bisa membantu dalam

menguji keefektifan dan keakuratan algoritma XGBoost dalam konteks yang belum banyak dijelajahi, seperti pemetaan dan klasifikasi bangunan di wilayah perkotaan Indonesia.

Untuk instansi pemerintah dan Dinas Pertanahan dan Tata Ruang, temuan ini dapat meningkatkan efisiensi dalam perencanaan dan pengelolaan perkotaan. Data yang dihasilkan mendukung identifikasi permukiman yang tidak layak dan membutuhkan intervensi, perencanaan infrastruktur dan layanan publik, serta mitigasi risiko bencana. Penggunaan data OSM untuk prediksi bangunan permukiman menawarkan solusi cepat dan efisien untuk berbagai daerah di Indonesia. Dengan menerapkan analisis karakter morfologi bangunan melalui data tapak bangunan OSM, penelitian ini dapat diaplikasikan di seluruh Indonesia, memanfaatkan ketersediaan data OSM yang luas dan metode yang cepat.

Untuk pelaku usaha, khususnya yang bergerak di bidang *real estate* dan pengembangan perkotaan, temuan ini dapat memberikan *insight* berharga mengenai distribusi dan karakteristik permukiman, mendukung keputusan investasi, dan pengembangan. Bagi masyarakat umum, peningkatan dalam perencanaan dan pengelolaan kota dapat berkontribusi pada peningkatan kualitas hidup, akses yang lebih baik ke fasilitas dan layanan, serta lingkungan yang lebih aman dan sehat. Secara keseluruhan, penelitian ini berpotensi memberikan kontribusi penting dalam upaya mengatasi tantangan perkotaan yang diakibatkan oleh pertumbuhan penduduk dan urbanisasi di Indonesia.

## I.7. Tinjauan Pustaka

Berkembangnya teknologi *remote sensing*, *machine learning*, dan meningkatnya ketersediaan data geospasial secara terbuka, telah memicu munculnya berbagai penelitian untuk analisis wilayah perkotaan, seperti identifikasi wilayah permukiman yang dilakukan dengan metode analisis citra berbasis objek (Putra dkk., 2023; Febrina dkk., 2019; Diyanah dan Bioresita, 2023); memanfaatkan citra satelit (Touzani dan Granderson, 2021); atau foto udara (Das dan Chand, 2021) dalam analisis klasifikasinya. Akan tetapi, penelitian-penelitian tersebut masih menggunakan data yang memiliki skala cukup kecil dan, umumnya, sulit untuk didapatkan serta digunakan pada wilayah lain.

Wang dkk. (2023) membahas mengenai pemanfaatan teknologi pengamatan bumi seperti citra satelit, yang memiliki berbagai keterbatasan dalam melakukan pemetaan pola perkotaan non-fisik, seperti kemiskinan, kesehatan, pendidikan, atau pendapatan, kecuali

pola dan informasi tersebut dapat terlihat dari penampilan fisik elemen perkotaan dalam citra yang digunakan. Dalam penelitian tersebut, Wang dkk. juga menyebutkan bahwa penggunaan citra satelit membutuhkan supervisi manusia untuk menentukan aturan klasifikasi atau pemilihan fitur, yang dapat berdampak pada sulitnya replikasi mengingat kondisi citra dan model yang bersifat *case-specific*.

Menurut Atwal dkk. (2022), untuk mengatasi masalah keterbatasan data dalam analisis perkotaan, dibutuhkan adanya sumber data yang dapat tersedia secara terbuka dan dapat diakses dengan mudah, seperti data OSM. Data OSM, yang bersifat *open source*, apabila dikolaborasikan dengan *machine learning*, dapat digunakan untuk berbagai keperluan seperti memetakan aksesibilitas dan jangkauan pendidikan dan kesehatan (Faiz, 2022), membuat indikator sosial-ekonomi (Feldmeyer dkk., 2020), pembuatan peta tipe hutan dengan OBIA dan data OSM (Brauchler dan Stoffels, 2020), atau untuk analisis permukiman informal di seluruh dunia (Soman dkk., 2020). Penelitian-penelitian tersebut menggunakan data bangunan OSM sebagai alternatif dari data bangunan yang diterbitkan oleh pemerintah, yang sulit untuk didapatkan.

Dalam bidang studi morfometrik urban, data bangunan dari OSM dapat digunakan untuk mendapatkan berbagai karakter morfometrik yang merepresentasikan bentuk dan pola bangunan di dalam suatu kawasan. Karakter morfometrik tersebut dapat digunakan untuk membuat model klasifikasi atau regresi untuk berbagai studi perkotaan. Sebagai contoh, Milojevic-Dupont dkk. (2020) memprediksi tinggi bangunan berdasarkan karakter morfologi urban. Dalam penelitian tersebut, berbagai karakter morfometrik dari bangunan, jalan, blok, dan kota digunakan sebagai data latih dalam pembuatan model regresi dengan algoritma XGBoost. Pada penelitian tersebut, model XGBoost dibandingkan dengan beberapa model lainnya dan ditemukan bahwa XGBoost memiliki kapabilitas dan kemampuan prediksi yang lebih tinggi. Selain itu, model ini juga cocok untuk digunakan pada kasus data tidak seimbang dan masif.

Penelitian terkait klasifikasi bangunan permukiman dengan menggunakan karakter morfometrik telah dilakukan oleh Sturrock dkk. (2018), Atwal dkk. (2022), dan Hartmann dkk. (2024). Sturrock dkk. menggunakan data bangunan dan jalan OSM di Botswana dan Swaziland untuk keperluan penyemprotan insektisida malaria. Karakter morfometrik yang dianalisis pada penelitian ini meliputi aspek geometri bangunan, jarak dengan jalan terdekat, kondisi persebaran bangunan tetangga, luas dari bangunan tetangga,

jumlah sisi dari bangunan tetangga, dan sub-poligon. Penelitian ini menggunakan model *super learner* yang memberikan bobot pada hasil penilaian dari berbagai algoritma *machine learning* yang berbeda. Berdasarkan penelitian ini, didapatkan akurasi dan *recall* sebesar 0,89 dan 0,86.

Dalam penelitian oleh Atwal dkk. (2022), data atribut OSM dan karakter morfometrik bangunan digunakan untuk mengidentifikasi bangunan permukiman di Virginia, Carolina Utara, dan Colorado. Data atribut OSM menjadi data tambahan dari penelitian sebelumnya, dan mencakup jenis penggunaan bangunan, ketersediaan informasi tambahan, dan atribut OSM lainnya yang membantu membedakan bangunan permukiman. Karakter morfologi yang dianalisis meliputi bentuk bangunan, kedekatannya dengan objek-objek seperti jalan, tempat parkir umum, dan data penggunaan lahan. Penelitian ini memanfaatkan algoritma *Decision Tree Classification*, yang dikenal memiliki kemampuan kuat dalam menginterpretasi signifikansi dari fitur-fitur tertentu. Hasil penelitian ini menunjukkan akurasi sebesar 0,98 dan rata-rata *F1-Score* sebesar 0,89, menandakan bahwa dengan menambahkan informasi atribut OSM sebagai fitur tambahan dapat meningkatkan akurasi dari model yang digunakan.

Pada tahun 2024, Hartmann dkk. melakukan penelitian yang sama, tetapi menggunakan data kadaster permukiman resmi yang dirilis oleh pemerintah Jerman. Penelitian ini melakukan klasifikasi bangunan dengan karakter morfometrik yang meliputi geometri bangunan, morfologi urban, dan ketetanggaan dari bangunan tersebut dan menggunakan algoritma *Random Forest Classifier*. Penelitian ini juga bereksperimen dengan proporsi jumlah bangunan non-permukiman dan permukiman 25:75, 50:50, dan 75:25. Dari penelitian yang dilakukan, ditemukan bahwa proporsi awal (25:75) memberikan hasil klasifikasi yang cukup seimbang dengan akurasi 0,86 dan *F1-Score* 0,91.

Sebagai alternatif penggunaan citra satelit untuk analisis perkotaan yang sulit direplikasi, dapat digunakan data bangunan *open source* seperti OSM ataupun data resmi dari pemerintah. Kemudian, untuk dapat mengambil *insight* dan pola yang terdapat pada data bangunan, dapat dilakukan analisis karakter morfometrik bangunan. Adapun, penelitian ini bertujuan untuk mengaplikasikan pendekatan klasifikasi permukiman dengan memanfaatkan karakter morfometrik bangunan dan data OSM beserta atributnya. Metode klasifikasi yang digunakan adalah algoritma XGBoost yang merupakan hasil

pengembangan dari model berbasis pohon seperti *Random Forest* dan *Gradient Boosting*. Penelitian ini, juga menggunakan *library* Momepy, yang diciptakan oleh Fleischmann (2019), yang dapat digunakan untuk melakukan perhitungan karakter morfologi secara lebih komprehensif. Dengan menggunakan *library* Momepy, perhitungan karakter morfometrik dapat dengan lebih mudah direplikasi di berbagai studi yang berbeda, sehingga penelitian ini dapat direplikasi di berbagai wilayah di Indonesia selama tersedia data tapak bangunan yang lengkap.

## **BAB II**

### **DASAR TEORI**

#### **II.1. Permukiman**

Berdasarkan Undang-undang Republik Indonesia No. 1 Tahun 2011 tentang Perumahan dan Kawasan Permukiman, dijelaskan bahwa permukiman merupakan bagian dari lingkungan hidup di luar kawasan lindung, baik itu kawasan perkotaan maupun perdesaan, yang berfungsi sebagai lingkungan tempat tinggal atau hunian serta tempat kegiatan yang mendukung perikehidupan dan penghidupan. Sementara itu, Peraturan Pemerintah Nomor 21 Tahun 2021 tentang Penyelenggaraan Penataan Ruang mendefinisikan permukiman sebagai bagian dari lingkungan hunian yang terdiri atas lebih dari satu satuan perumahan yang memiliki prasarana, sarana, utilitas umum, serta penunjang kegiatan fungsi lain, baik di kawasan perdesaan maupun perkotaan. Dari dua definisi tersebut, dapat disimpulkan bahwa permukiman tidak hanya sekedar tempat tinggal, tetapi juga mencakup tempat kegiatan yang mendukung kehidupan sehari-hari dengan ketersediaan prasarana, sarana, dan utilitas umum, menjadikannya sebuah ekosistem yang komprehensif, terintegrasi, dan mampu mendukung penghidupan penduduknya.

Berdasarkan klasifikasi oleh Prayitno (2014), permukiman dibagi menjadi dua jenis berdasarkan legalitas dan formalitasnya: permukiman formal dan permukiman informal. Permukiman formal adalah jenis permukiman yang dibangun dengan mengikuti semua aturan dan prosedur hukum yang berlaku, dikenal juga sebagai permukiman terencana karena perencanaannya yang matang. Sebaliknya, permukiman informal adalah permukiman yang tumbuh tanpa mengikuti prosedur hukum yang formal, seringkali karena keterbatasan sumber daya atau kebutuhan mendesak akan perumahan, sehingga disebut sebagai permukiman tidak terencana. Permukiman formal biasanya dikembangkan oleh pengembang properti, sementara permukiman informal muncul dan berkembang sesuai dengan kebutuhan dan kemampuan masyarakat dalam menyediakan rumah untuk diri mereka sendiri.

Berdasarkan Wesnawa (2015), permukiman dapat dibagi ke dalam tiga level: makro, meso, dan mikro. Tujuan dari pembagian skala ini adalah untuk memberikan

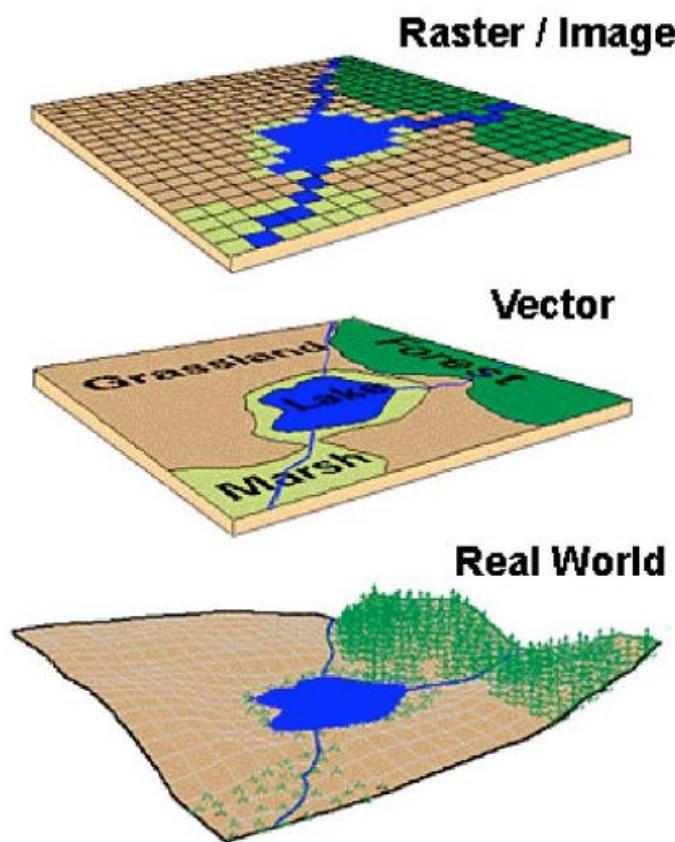
penjelasan yang lebih terarah dan efisien mengenai isu-isu permukiman serta memfasilitasi penelitian dengan lingkup yang lebih terfokus. Pada skala makro, permukiman dilihat sebagai kumpulan beberapa permukiman yang menyatu menjadi satu wilayah yang terbangun, memiliki karakteristik regional dan tidak dapat dianalisis sebagai entitas terpisah. Pada skala meso, permukiman dianggap sebagai bagian dari wilayah terbangun yang secara eksklusif digunakan untuk permukiman, yang juga dapat dianggap sebagai bagian dari skala makro. Sementara itu, analisis skala mikro berpusat pada tingkat unit rumah atau individu dalam permukiman, dengan fokus utama pada aspek ketetanggaan. Pendekatan ini memungkinkan pengamatan yang lebih detail dan spesifik terhadap dinamika dalam permukiman (Wesnawa, 2015).

## **II.2. Informasi Geospasial**

Informasi geospasial (IG) merupakan hasil pengolahan data geospasial yang digunakan sebagai alat bantu dalam penyusunan kebijakan, pengambilan keputusan, dan/atau pelaksanaan kegiatan yang berhubungan dengan keantarkasaan (Undang-Undang No. 4 Tahun 2011). IG sangat penting dalam banyak hal, seperti pembangunan infrastruktur, pemanfaatan sumber daya alam, ketahanan nasional, mitigasi bencana, dan bidang lainnya yang penting, seperti pertanian. Dengan bantuan data geospasial, prinsip-prinsip umum dapat diterapkan pada kondisi spesifik setiap lokasi. Ini memungkinkan kita melacak peristiwa yang terjadi di suatu lokasi dan membantu kita memahami perbedaan antara lokasi yang berbeda. IG sangat penting untuk perencanaan dan pengambilan keputusan modern (Adams dkk., 2014). Data geospasial yang telah diproses sehingga dapat digunakan sebagai alat bantu dalam perumusan kebijakan, pengambilan keputusan, dan/atau pelaksanaan kegiatan yang berkaitan dengan ruang bumi didefinisikan sebagai IG menurut UU tersebut.

Kebutuhan akan IG, terutama di daerah padat penduduk, sangat krusial karena daerah tersebut merupakan pusat kegiatan dengan kepadatan bangunan serta sarana dan prasarana yang kompleks, ditandai dengan karakteristik, struktur, dan jenis data geospasial yang heterogen. Informasi geospasial yang divisualisasikan dapat digunakan sebagai dasar dalam pengambilan keputusan terkait keberlanjutan perencanaan, pembangunan, dan pengoperasian infrastruktur di kota-kota.

Data merupakan aspek penting dalam penyediaan IG. Data geospasial umumnya dapat disimpan dalam format vektor, raster, dan citra (Saab, 2003). Data vektor terdiri dari koordinat titik yang digunakan untuk menampilkan, memposisikan, dan menyimpan informasi spasial dengan memanfaatkan titik, garis, dan poligon. Sementara itu, data raster merepresentasikan objek geografis dalam bentuk piksel-piksel, yang biasanya diperoleh melalui teknologi penginderaan jauh. Visualisasi dari format data vektor dan raster ditunjukkan pada gambar II.1.



**Gambar II.1.** Visualisasi data raster, vektor, dan dunia nyata (Sumber: Saab, 2003).

### **II.3. Tobler First Law (TFL)**

*Tobler First Law* (TFL) atau Hukum Pertama Tobler, yang berbunyi "segala sesuatu terhubung dengan segala sesuatu yang lain, tetapi hal-hal yang lebih dekat secara spasial cenderung lebih terkait daripada hal-hal yang lebih jauh secara spasial." (Tobler, 1970), diperkenalkan saat Tobler membahas model pertumbuhan penduduk di Detroit menggunakan simulasi 3D dengan perhitungan yang mendekati perkiraan. TFL telah menjadi sangat berpengaruh, menyoroti pentingnya kedekatan dan keterkaitan geografis

dalam analisis spasial. TFL dapat dioperasionalkan melalui teknik interpolasi spasial, yang mengasumsikan bahwa nilai pada permukaan yang tidak diukur dalam suatu wilayah dapat ditentukan berdasarkan set nilai yang diukur di lokasi lain, menunjukkan autokorelasi spasial. Teknik seperti interpolasi berbasis jarak dan *kriging* memungkinkan objektivitas dalam mengukur keterkaitan spasial. TFL relevan dalam berbagai model teoretis dalam geografi manusia dan perkotaan, menunjukkan pengaruh jarak terhadap distribusi populasi, aktivitas ekonomi, dan pola permukiman (Miller, 2004).

Meskipun TFL memiliki aplikasi luas, pada konteks dan situasi tertentu, hukum ini tidak memberikan penjelasan yang memadai. Ini terutama terjadi di perbatasan dengan penghalang alami atau politik, "hal-hal yang dekat" tidak lagi "lebih berkaitan" dibandingkan dengan "hal-hal yang jauh". Selain itu, proses difusi dan migrasi sering berlangsung secara hierarkis dan tidak selalu mengikuti prinsip kedekatan yang diusulkan oleh TFL (Waters, 2021).

Menurut Waters (2021), TFL memainkan peran penting dalam analisis spasial dan geografi dengan menekankan pada keterkaitan antara kedekatan dan keterkaitan geografis. Meskipun ada situasi TFL mungkin tidak berlaku, hukum ini tetap merupakan prinsip fundamental dalam memahami distribusi fenomena geografis dan menjadi landasan yang kuat untuk penelitian tentang dinamika ruang dan perencanaan perkotaan karena entitas geografis yang berdekatan biasanya memiliki korelasi dan kemiripan yang tinggi.

#### **II.4. OpenStreetMap (OSM)**

OpenStreetMap (OSM) merupakan inisiatif komunitas geospasial yang menyediakan akses bebas dan terbuka ke data spasial global (OpenStreetMap, 2024). OSM menyajikan basisdata spasial yang mencakup hampir seluruh dunia, menjadikannya salah satu proyek berbasis *Volunteered Geographic Information* (VGI) paling berhasil (Arsanjani dkk., 2015; Faiz, 2022). Karena OSM dikelola di bawah lisensi yang memungkinkan penggandaan, modifikasi, dan distribusi data, penggunaannya menjadi terbuka luas (Bennett, 2010). Gratisnya akses informasi geospasial memfasilitasi berbagai organisasi untuk mengintegrasikan atau melengkapi informasi geospasial yang mereka miliki (Chintya dkk., 2022). Berdasarkan data terkini dari situs resmi OSM per 10 Maret 2024, OSM telah memiliki lebih dari 10 juta pengguna, 28 miliar titik GPS, dan 9 miliar *nodes* (OpenStreetMap, 2024).

Namun, penggunaan OSM juga dihadapkan pada beberapa tantangan signifikan, terutama terkait kualitas dan akurasi data. Variabilitas kualitas data yang disumbangkan oleh sukarelawan menjadi salah satu permasalahan utama. Tidak semua kontributor memiliki keahlian geospasial yang sama, yang dapat menyebabkan data menjadi tidak seragam. Meski ada usaha komunitas untuk validasi dan pembaruan data, tetapi terdapat risiko keberadaan data yang tidak akurat di beberapa wilayah. Neis dan Zipf (2012) menemukan bahwa meski cakupan dan detail data OSM meningkat seiring waktu, upaya mempertahankan akurasi dan kekinian data masih menemui tantangan, terutama di daerah dengan sedikit kontributor. Hal ini menegaskan pentingnya validasi yang lebih ketat dan partisipasi komunitas yang lebih aktif untuk meningkatkan kualitas data OSM secara umum.

Pada tahun 2014, Fan dkk. melakukan studi mengenai kualitas bangunan dan kelengkapan tematik data OSM. Studi ini menemukan bahwa data OSM memiliki kelengkapan bangunan yang lebih baik dibandingkan dengan data ATKIS milik pemerintah Jerman, terutama untuk bangunan yang relatif baru. Kelebihan ini menunjukkan frekuensi penambahan data baru yang lebih tinggi pada data OSM. Selain itu, data OSM menunjukkan akurasi tematik yang sempurna, sebesar 100%, untuk wilayah Munich. Namun, data tapak bangunan OSM, yang berasal dari digitasi citra Bing, umumnya memiliki bentuk yang lebih sederhana dibandingkan dengan data ATKIS, yang dihasilkan dari pengukuran kadastral. Di Indonesia, sebuah studi oleh Zainab dan Jung-Hong (2021) membandingkan data OSM dengan data Jakarta Satu, sebuah inisiatif dari Pemerintah Provinsi DKI Jakarta untuk menyediakan data terbuka. Penelitian ini berfokus pada data jaringan jalan dan menemukan bahwa OSM lebih unggul dalam hal kelengkapan dan akurasi data, khususnya dalam aspek atribut dan tematik, dibandingkan dengan data Jakarta Satu.

Atribut atau 'tags' merupakan pasangan antara kata kunci dan nilai yang memberikan informasi terkait objek yang direpresentasikan oleh suatu elemen pada OSM (Bennett, 2010). Keberadaan atribut pada data OSM memudahkan pengguna untuk mengambil hanya objek yang relevan atau melakukan analisis lebih lanjut dengan memanfaatkan detail informasi tersebut. Per Februari 2024, OSM memiliki lebih dari 96 ribu kata kunci dan lebih dari 155 juta atribut yang beragam (OpenStreetMap, 2024). Atribut ini melengkapi informasi spasial dari data OSM dengan informasi tematik, berisi

berbagai informasi seperti penggunaan bangunan, jenis atap, tinggi bangunan, jumlah lantai, jenis ubin, dan informasi lainnya yang berkaitan dengan kondisi bangunan tersebut.

## **II.5. Morfologi Urban**

Definisi dasar dari *urban morphology* atau morfologi urban adalah studi mengenai bentuk perkotaan, agen, serta proses yang bertanggung jawab terhadap perubahan perkotaan yang berkaitan dengan elemen fisik dari struktur bangunan dan kota, seperti *urban tissues*, jaringan jalan, plot, dan bangunan (Oliveira, 2016). Berdasarkan Kropf (2017), morfologi urban merupakan bidang ilmu perkotaan yang mempelajari mengenai wujud fisik dari perkotaan di dalam ruang yang berubah terhadap waktu. Morfologi urban dapat didefinisikan sebagai kajian proses pembentukan dan kecenderungan yang melandasi bentuk perkotaan secara nyata (Batty, 2008; Boeing, 2021), serta pada implikasi dari aspek sosial, ekonomi, dan fisik lingkungan (Berghauser Pont, 2018; de Castro dkk., 2019; Moudon, 1997). Domain morfologi urban juga mengacu pada substansi mengenai bentuk perkotaan (*urban form*), yang berfokus pada proses pembentukan dan transformasinya (Cortes, 2005; Whitehand dkk., 1996); pola spasial pada skala yang bervariasi (Batty, 2008; Boeing, 2018); dan karakteristik fisik ruang untuk mempromosikan pembangunan perkotaan yang berkelanjutan (Jenks dkk., 2008; Lai dkk., 2018).

Bidang ini berlandaskan pada analisis ruang, yang secara tradisional masih melalui pengamatan visual dan kualitatif. Menurut Dibble dkk. (2015), bidang penelitian mengenai morfologi urban masih memiliki kekurangan pada metode analisis kuantitatif yang dapat diaplikasikan secara universal untuk menganalisis bentuk perkotaan. Sifat penelitian di bidang morfologi urban yang cenderung bersifat kualitatif menjadikannya sulit untuk direplikasi. Dibble dkk. juga menemukan bahwa penelitian mengenai morfologi urban masih belum memiliki metode yang sistematis dan dapat dikuantifikasi untuk menilai bentuk perkotaan dan menarik kesimpulan melalui analisis data yang mendalam. Selain itu, menurut Fleischmann dan Arribas-Bel (2022), terdapat kendala dalam studi morfologi urban untuk melakukan studi yang melengkapi wilayah yang besar dengan menggunakan pendekatan yang bersifat subjektif. Sehingga, dibutuhkan pendekatan lain yang bersifat objektif dan dapat dikembangkan di skala yang luas tanpa mengorbankan detil.

### **II.5.1. Morfometrik Urban**

*Urban morphometric* atau morfometrik urban, merupakan sebuah pendekatan analisis bentuk perkotaan melalui pengukuran yang komprehensif dan sistematis dari karakter morfologi yang diperkenalkan oleh Dibble dkk. (2015). Morfometrik sendiri berasal dari bidang keilmuan biologi yang diartikan sebagai metode karakterisasi, analisis, dan perbandingan bentuk biologis secara kuantitatif (Roth dan Mercer, 2000). Kemudian, istilah morfometrik perkotaan disebutkan pertama kali oleh Carneiro dkk. (2010) yang mengambil penggunaan kata morfometrik yang mengacu kepada studi mengenai variasi dan perubahan wujud pada objek bangunan. Pendekatan tersebut yang dicoba oleh Dibble dkk. untuk diterapkan pada bidang morfologi urban untuk membuat sebuah pendekatan yang berfokus kepada penilaian kuantitatif yang bersifat objektif dan dapat direplikasi. Morfometrik urban yang diperkenalkan oleh Dibble dkk. berkembang menjadi metode kuantitatif untuk analisis bentuk perkotaan yang dapat digunakan untuk menganalisis morfologi urban secara konsisten di tingkat regional dan nasional (Fleischmann dan Arribas-Bel, 2022).

### **II.5.2. Elemen-elemen Morfometrik**

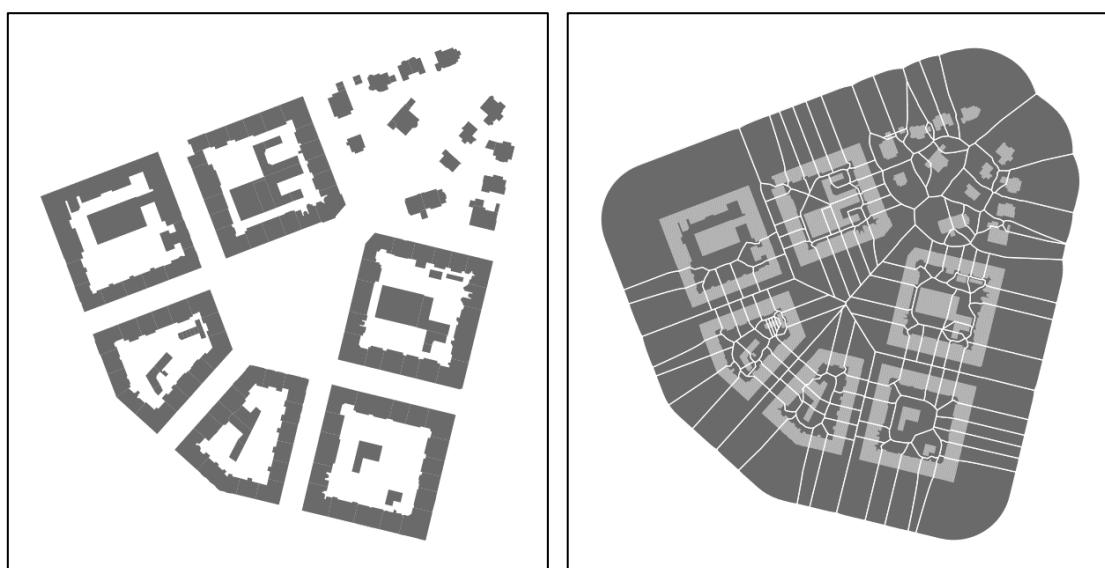
*Morphometric elements* atau elemen-elemen morfometrik merupakan komponen dasar yang digunakan untuk menganalisis dan mengukur bentuk serta struktur fisik perkotaan secara kuantitatif dan sistematis. Secara umum, terdapat tiga elemen fundamental dalam morfologi urban, yaitu bangunan, jalan, dan plot (Oliveira, 2016; Moudon, 1997). Dari sudut pandang morfometrik, bangunan dan jalan memiliki definisi yang cukup jelas. Bangunan direpresentasikan oleh tapak bangunan berupa poligon, sedangkan jalan direpresentasikan sebagai garis tengah jaringan jalan dan fungsinya sebagai penghubung dan pembagi ruang dalam kota (Fleischmann dkk., 2022).

Akan tetapi, plot umumnya lebih rumit untuk didefinisikan karena sifatnya yang polisemik (bermakna banyak) (Kropf, 2018) dan memiliki peran struktural yang ambigu (Levy, 1999). Selain itu, cara pemetaan plot umumnya juga tidak terdokumentasi secara jelas, dan apabila terdokumentasi, tidak semuanya mendefinisikan atau merepresentasikan plot dengan cara yang sama. Hal tersebut menjadikannya kurang cocok untuk menjadi komponen dasar dalam aplikasi morfometrik urban (Fleischmann dkk., 2020). Ketidakkonsistensiplot diilustrasikan melalui Gambar II.2. Gambar tersebut menampilkan dua plot dengan perbedaan signifikan mulai dari bentuk, ukuran, hingga definisinya.



**Gambar II.2.** Perbandingan plot tradisional (kiri) dan plot modern (kanan) di Glasgow, Skotlandia yang berasal dari sumber yang berbeda. Terlihat bahwa dari kedua gambar, batas antara area privat dan publik lebih terlihat jelas pada plot tradisional dibandingkan plot modern yang memiliki batas antara area privat dan publik yang relatif kabur

(Sumber: Fleischmann dkk., 2020)



**Gambar II.3.** Perbedaan geometri antara bangunan (kiri) dan sel teselasi (kanan).  
(Sumber: Fleischman, 2019).

Untuk mengantisipasi ketidakakuratan plot, Fleischmann dkk. (2020) mengusulkan penggunaan *morphological tessellation* (MT) atau teselasi morfologi, sebuah metode untuk mendapatkan satuan analisis *morphological cell* atau *tessellation cell* yang dapat digunakan sebagai alternatif daripada penggunaan plot. Teselasi sendiri merupakan

metode pembuatan poligon yang berdasarkan pada diagram Voronoi, didapatkan dari tapak bangunan (Hamaina dkk., 2012). Sel teselasi menjadi alternatif batas tanah untuk setiap bangunan yang berdekatan menggantikan peran plot tanah. Perbandingan antara geometri bangunan dan sel teselasi dapat dilihat pada Gambar II.2..

### **II.5.3. Karakter Morfometrik**

*Morphometric characters* atau karakter morfometrik merupakan aspek-aspek yang dapat diukur dari setiap elemen morfometrik yang dapat digunakan untuk membedakan antara satu bentuk perkotaan dengan lainnya (Fleischmann dkk., 2022; Fleischmann dkk., 2021). Berbagai penelitian mengenai morfologi urban umumnya melakukan analisis dengan karakter morfometrik yang berbeda-beda. Fleischmann dkk. (2022), melakukan penyederhanaan karakter morfologi menjadi enam kategori ontologis dan tiga skala konseptual. Karakter morfometrik dapat dibagi menjadi enam kategori, yaitu:

a. Dimensi (*Dimension*)

Karakter yang mewakili bentuk dimensi geometri dasar dari setiap objek individual.

b. Bentuk (*Shape*)

Karakter yang mewakili fitur matematis dari dimensi geometris dari setiap objek individual.

c. Distribusi Spasial (*Spatial Distribution*)

Karakter yang mewakili distribusi objek di dalam ruang dan posisi relatif.

d. Intensitas (*Intensity*)

Karakter yang mewakili intensitas dari pemakaian ruang merujuk kepada densitas dari elemen di dalam suatu batasan wilayah.

e. Konektivitas (*Connectivity*)

Karakter yang mewakili hubungan spasial antara segmen dari suatu jaringan jalan.

f. Diversitas (*Diversity*)

Karakter yang mewakili variasi dan keragaman dari elemen dan karakternya di dalam kawasan studi.

Kemudian, karakter morfologi dapat dibagi menjadi tiga skala konseptual, yaitu:

- a. Kecil (*Small*)  
Karakter yang memiliki cakupan bangunan, plot, jalan, atau blok.
- b. Sedang (*Medium*)  
Karakter yang memiliki cakupan lingkungan, jarak berjalan kaki (5–10 menit), atau kawasan.
- c. Besar (*Large*)  
Karakter yang memiliki cakupan kota/kabupaten, kawasan perkotaan, ataupun kawasan metropolitan.

## **II.6. Machine Learning**

*Machine learning* dapat diartikan sebagai disiplin ilmu yang memungkinkan komputer untuk belajar tanpa perlu diprogram secara eksplisit (Sharma dan Kumar, 2017). *Machine learning* juga dapat dijelaskan sebagai serangkaian program komputer yang dapat memperbaiki performanya berdasarkan pengalaman yang diukur menggunakan metrik tertentu dalam menyelesaikan tugas yang diberikan (Das dkk., 2015). *Machine learning* hadir untuk menjawab keterbatasan yang terdapat pada metode tradisional dalam perancangan algoritma estimasi. Pada metode tradisional, manusia atau operator perlu menyusun model matematis secara manual berdasarkan pemahaman mendalam mengenai topik yang sedang dikerjakan. *Machine learning* memperkenalkan langkah yang lebih efisien, yaitu pembentukan data latih (Simeone, 2018), yang merupakan kumpulan data berisi contoh pola perilaku terkait subjek yang diteliti. Data ini nantinya akan diproses oleh komputer dengan menggunakan algoritma pembelajaran yang spesifik.

Algoritma *machine learning* yang telah mempelajari pola dan *trend* dari data latih, dapat digunakan sebagai sebuah model untuk melakukan suatu pekerjaan spesifik yang berkaitan dengan data latih. Kinerja suatu model *machine learning* dapat dievaluasi menggunakan data uji, yaitu kumpulan data yang digunakan untuk menguji efektivitas model *machine learning* yang telah dikembangkan. Secara keseluruhan, *machine learning* beroperasi dengan cara mengidentifikasi dan memanfaatkan pola yang ada dalam data (Angra dan Ahuja, 2017). Han dkk. (2011) membagi *machine learning* menjadi empat jenis utama, yaitu:

- a. Pembelajaran Terawasi (*Supervised Learning*)

Metode ini memanfaatkan data terlabel untuk mengawasi proses pembuatan model, yang kemudian digunakan untuk memprediksi label dari data baru dalam pengujian. Label ini bisa berbentuk angka atau kategori, dengan analisis klasifikasi dan regresi sebagai contoh aplikasinya.

b. Pembelajaran Tidak Terawasi (*Unsupervised Learning*)

Berbeda dengan pembelajaran terawasi, metode ini tidak membutuhkan data terlabel dalam pelatihan mesin. Model yang dihasilkan menentukan kelas data melalui klasterisasi, dengan analisis kluster sebagai salah satu contoh metodenya.

c. Pembelajaran Semi-terawasi (*Semi-supervised Learning*)

Metode ini menggabungkan data terlabel dan tidak terlabel dalam proses pelatihan. Digunakan ketika data latih hanya sebagian yang memiliki label, terutama jika label tersebut hilang. Data terlabel membantu menentukan kelas data, sedangkan data tidak terlabel membantu mempertajam pemisahan antar kelas. *Long Short Term Memory* (LSTM) adalah salah satu contoh penggunaannya.

d. Pembelajaran Aktif (*Active Learning*)

Dalam metode ini, pengguna memainkan peran aktif dalam pelatihan mesin, seperti dalam menentukan label data tidak terlabel. Tujuan dari pembelajaran aktif adalah untuk meningkatkan kualitas model dengan menyertakan *input* dari pengguna, namun harus diperhatikan jumlah data yang dimasukkan untuk menghindari terjadinya *overfitting*.

#### **II.6.1. Hyperparameter**

Parameter pada *machine learning* dapat dibedakan menjadi dua jenis: parameter model dan *hyperparameter*. Parameter model adalah variabel yang nilainya ditentukan melalui proses pembelajaran dari data latih oleh algoritma, sedangkan *hyperparameter* adalah variabel yang dapat diatur oleh pengembang untuk mempengaruhi proses pembelajaran *machine learning* (Hillel dkk., 2018). Setiap model *machine learning* umumnya memiliki *hyperparameters* yang dapat diatur untuk meningkatkan performa model terhadap suatu dataset karena setiap dataset dapat memiliki kombinasi *hyperparameter* ideal yang berbeda-beda (Feurer dan Hutter, 2019). Sehingga, dalam pembuatan model *machine learning* dapat dilakukan optimasi nilai *hyperparameter* untuk dapat meningkatkan performa model yang dibuat.

Menurut Yang dan Shami (2020), terdapat berbagai metode untuk mencari nilai *hyperparameter* yang ideal, seperti:

a. *Babysitting*

Dalam metode ini, nilai *hyperparameter* akan dicari secara manual. Metode ini dilakukan dengan menguji berbagai kombinasi *hyperparameter* secara satu persatu berdasarkan pengalaman, perkiraan, dan analisis dari hasil sebelumnya.

b. *Grid Search*

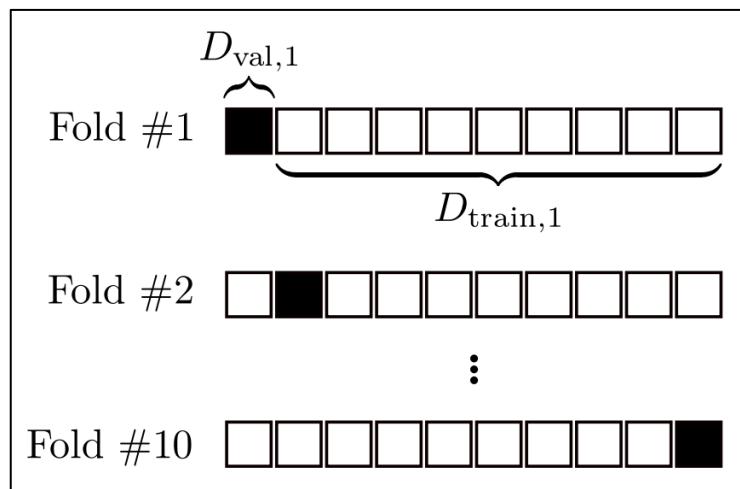
Pada metode ini, akan dibuat kombinasi nilai *hyperparameter* yang akan dievaluasi seluruhnya untuk mencari kombinasi nilai *hyperparameter* terbaik. Metode ini membutuhkan waktu dan tenaga komputasi yang tinggi karena metode ini akan menguji setiap kombinasi untuk mencari kombinasi terbaik.

c. *Random Search*

Pada metode ini, sama seperti *grid search*, akan dibuat kombinasi nilai *hyperparameter* yang kemudian dipilih secara acak untuk dievaluasi. Metode ini lebih efisien dibandingkan dengan *grid search* karena tidak menguji keseluruhan kombinasi, melainkan metode ini mengambil sampel secara acak dari seluruh kombinasi *hyperparameter*.

### **II.6.2. Cross Validation**

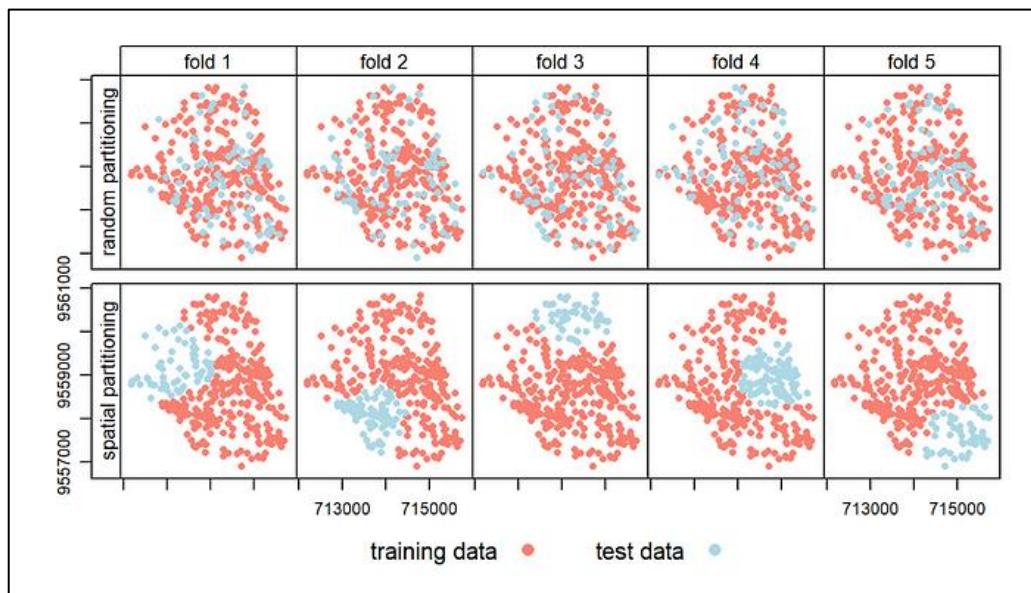
*Cross validation* (CV) atau validasi silang merupakan sebuah metode untuk seleksi dan evaluasi model *machine learning* (Berrar, 2019). Menurut Berrar (2019), metode ini dapat digunakan untuk mengoptimasi nilai *hyperparameter*, membandingkan hasil dari beberapa model berbeda, dan mengevaluasi performa model. Metode CV memiliki beberapa strategi yang berbeda, salah satu strategi yang paling sering digunakan adalah *k-fold cross validation* (Berrar, 2019). Pada strategi tersebut, data akan dibagi menjadi beberapa lipatan berjumlah k secara acak. Model *machine learning* kemudian akan dilatih dengan subset data k-1 yang merepresentasikan data latih dan sisanya menjadi data validasi. Ilustrasi dari strategi ini dapat dilihat pada Gambar II.3.. Prosedur ini diulangi sampai setiap lipatan telah digunakan sebagai data validasi dan performa model pada tiap lipatan diukur. Rata-rata dari performa k percobaan akan menjadi hasil performa dari CV.



**Gambar II.4.** Visualisasi strategi *K-Fold Cross Validation* (Berrar, 2019)

### II.6.3. Spatial Cross Validation

*Spatial cross validation* atau CV spasial merupakan metode CV yang mempertimbangkan kehadiran *spatial autocorrelation* antara setiap *data points* (Meyer dkk., 2019). CV spasial memiliki peran penting untuk mencegah performa model yang *overfitting* terhadap data latih sehingga model gagal dalam memprediksi data lain yang belum pernah dilihat oleh model. CV spasial bekerja dengan membagi data latih ke dalam beberapa lipatan berdasarkan kedekatan lokasinya, berbeda dengan *k-fold* CV yang membagi data secara acak. Perbedaan antara CV dengan CV spasial dapat dilihat pada Gambar II.4.



**Gambar II.5.** Ilustrasi dari *Cross Validation* standar dengan *Cross Validation* spasial  
(Sumber: Lovelace dkk., 2021).

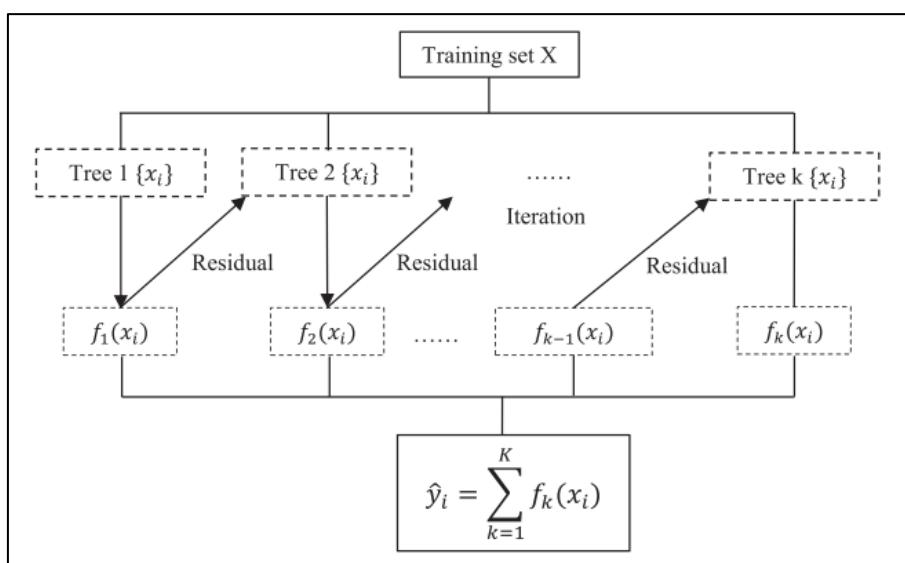
## II.7. Extreme Gradient Boosting (XGBoost)

*Extreme Gradient Boosting*, atau XGBoost, adalah pengembangan dari algoritma *Gradient Boosting* (GB) yang pertama kali diperkenalkan oleh Chen dan Guestrin pada tahun 2016. GB bekerja dengan menerapkan konsep *ensemble learning* dan *boosting* untuk menemukan solusi optimal bagi permasalahan regresi, klasifikasi, dan *ranking*. *Ensemble learning* adalah konsep yang menggunakan beberapa *Decision Tree* (DT) yang dilatih secara individual. Kemudian hasil dari setiap DT akan digabungkan dengan beberapa cara, seperti pengambilan rata-rata atau pemungutan suara terbanyak, untuk menentukan prediksi akhir (Ho, 1995). Penggunaan *ensemble learning* ditunjukkan pada persamaan II.1 (Chen dan Gusestrin, 2016). Pada *ensemble learning*, masing-masing *learner* atau pohon digunakan dalam memperhitungkan hasil keluaran.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \dots \dots \dots \quad (\text{II.1})$$

Sedangkan, *boosting* merupakan konsep pada model berbasis DT. DT akan dilatih secara berurutan dengan melakukan perbaikan pada DT selanjutnya untuk mengurangi nilai *loss* (Breiman, 1996), yang diaplikasikan pada algoritma XGBoost melalui persamaan II.2 dan gambar II.5.  $f_t(x_i)$  merupakan perbaikan berdasarkan pohon atau iterasi sebelumnya (Chen dan Guestrin, 2016).

$$\mathcal{L}^{(t)} = \sum_{i=1} l(\hat{y}_i^{(k-1)}, y_i) + f_k(x_i) \dots \dots \dots \quad (\text{II.2})$$



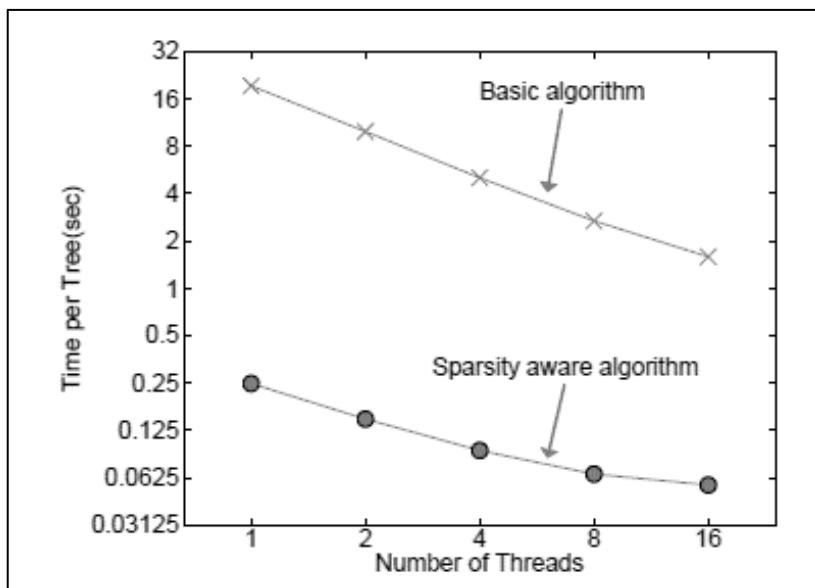
**Gambar II.6.** Bagan alir dari algoritma XGBoost (Sumber: Nguyen dkk., 2021).

XGBoost meningkatkan GB dengan memasukkan prinsip regularisasi ke dalam fungsi kerugian (*loss function*) untuk mengontrol kompleksitas model dan mengurangi risiko *overfitting*, yaitu kondisi ketika model memperlihatkan akurasi tinggi pada data latih tetapi gagal memprediksi data baru yang belum pernah dilihat sebelumnya (Hawkins, 2004). Fungsi tujuan (*objective function*) ini digunakan untuk mengevaluasi performa model selama fase pelatihan. Persamaan II.3 memaparkan fungsi kerugian yang digunakan dalam algoritma XGBoost, yang dihitung dengan menjumlahkan kerugian dengan bias dari regularisasi seperti yang ditunjukkan di persamaan II.4, untuk menghindari pembobotan berlebihan saat model menjadi terlalu kompleks (Chen & Guestrin, 2016). Dalam persamaan II.4, terdapat dua *hyperparameter* yang bisa disesuaikan:  $\gamma$ , yang meningkatkan bias dengan bertambahnya jumlah *node* dalam pohon, dan  $\lambda$ , yang meningkatkan bias dengan bertambahnya bobot pada *node* tertentu.

$$\mathcal{L}^{(t)} = \sum_{i=1} l(\hat{y}_i^{(t-1)}, y_i) + f_t(x_i) + \Omega(f_t) \dots \dots \dots \text{(II.3)}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda |\omega|^2 \dots \dots \dots \text{(II.4)}$$

Selain prinsip regularisasi, XGBoost juga menambahkan parameter *shrinkage* untuk menghindari *overfitting*, sebuah konsep yang diperkenalkan oleh Friedman pada tahun 2002. *Shrinkage* mengatur besar bobot dari setiap DT, dan memberikan ruang bagi DT selanjutnya untuk mengembangkan model. Kelebihan lain dari XGBoost adalah kemampuannya dalam menghadapi kondisi data yang tidak seimbang (*imbalanced data*) dan data yang hilang (*missing values*). Hal ini menjadi keunggulan karena sebagian besar kasus memiliki kondisi data yang tidak seimbang, seperti dalam kasus deteksi penipuan (*fraud detection*) dan deteksi spam (*spam detection*). Selain itu, XGBoost memiliki kemampuan dalam mengatasi *missing values* karena algoritma ini bersifat *sparsity-aware*. Model dengan sifat *sparsity-aware* dapat menangani data yang hilang dengan membuat kompleksitas komputasi menjadi linear terhadap jumlah entri yang tidak hilang dari fitur yang dimasukkan (Chen dan Guestrin, 2016). Hal ini dapat meningkatkan performa model menjadi lebih cepat dibandingkan dengan model lainnya, seperti yang dibuktikan oleh Chen dan Guestrin pada gambar II.6.



**Gambar II.7.** Perbedaan algoritma sparsity aware dengan algoritma biasa

(Sumber: Chen dan Guestrin, 2016).

Hal ini menjadikan XGBoost sebagai model yang tepat untuk menangani kasus yang dengan data yang bersifat tidak seimbang dan memiliki informasi yang hilang (*missing values*).

## II.8. Evaluasi Model Klasifikasi

Dalam analisis klasifikasi, evaluasi performa menggunakan metrik tertentu untuk menilai kinerja suatu model klasifikasi dalam memprediksi label kelas pada data. Penting untuk melakukan evaluasi performa pada data uji, bukan pada data latih. Hal ini karena penggunaan data latih untuk mengukur performa klasifikasi cenderung menghasilkan penilaian yang terlalu optimis (Powers, 2020). Data latih berfungsi untuk melatih model klasifikasi agar dapat mengenali pola dan karakteristik data tersebut. Selama proses pelatihan, model klasifikasi bisa terlalu terfokus pada pola dan karakteristik yang ada dalam data latih, sehingga kemampuan model untuk mengenali data baru atau data yang belum pernah dilihat sebelumnya dapat berkurang. Data uji digunakan untuk mengevaluasi kemampuan model klasifikasi dalam mengenali data baru secara objektif, tanpa bias dari data latih. Oleh karena itu, sangat penting untuk menggunakan data uji sebagai basis evaluasi, agar hasil evaluasi yang diperoleh lebih akurat dan dapat dijadikan dasar dalam memilih model klasifikasi yang paling sesuai.

*Confusion matrix* atau matriks konfusi merupakan salah satu alat evaluasi yang umum digunakan dalam data mining untuk mengukur kinerja model klasifikasi (Fawcett, 2006). Matriks konfusi terdiri dari sel  $m \times m$ , dengan  $m$  merupakan jumlah kelas dalam data. Setiap baris dalam tabel mewakili kelas asli (aktual) dari data, dan setiap kolom mewakili prediksi model untuk kelas tersebut. Dalam evaluasi performa klasifikasi biner, hanya terdapat dua kelas yang perlu diidentifikasi: kelas positif dan kelas negatif. Kelas positif adalah kelas yang dianggap penting dan ingin diidentifikasi dengan benar oleh model klasifikasi, sedangkan kelas negatif adalah kelas yang dianggap kurang penting. Tabel II.1 merupakan contoh matriks konfusi untuk klasifikasi biner, yang hanya memiliki sel  $2 \times 2$ .

**Tabel II.1** Matriks konfusi

		Hasil Prediksi	
		Positif	Negatif
Data Aktual	P	Positif Benar (TP)	Negatif Salah (FN)
	N	Positif Salah (FP)	Negatif Benar (TN)

Matriks konfusi, dapat diturunkan ke dalam metrik-metrik lain seperti:

a. *Accuracy*

*Accuracy* atau akurasi, merupakan rasio prediksi benar (positif dan negatif) dibandingkan dengan keseluruhan data. Akurasi mengukur sejauh mana model klasifikasi benar dalam memprediksi kelas data.

$$Akurasi = \frac{(TP+TN)}{(TP+FP+FN+TN)} \dots \dots \dots \quad (\text{II.5})$$

b. *Precision*

*Precision* atau presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Presisi mengukur sejauh mana prediksi positif yang dilakukan oleh model adalah benar.

$$Precision = \frac{(TP)}{(TP+FP)} \dots \dots \dots \quad (\text{II.6})$$

c. *Recall*

*Recall* atau *sensitivity* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* mengukur sejauh mana model mampu mengidentifikasi kelas positif secara akurat.

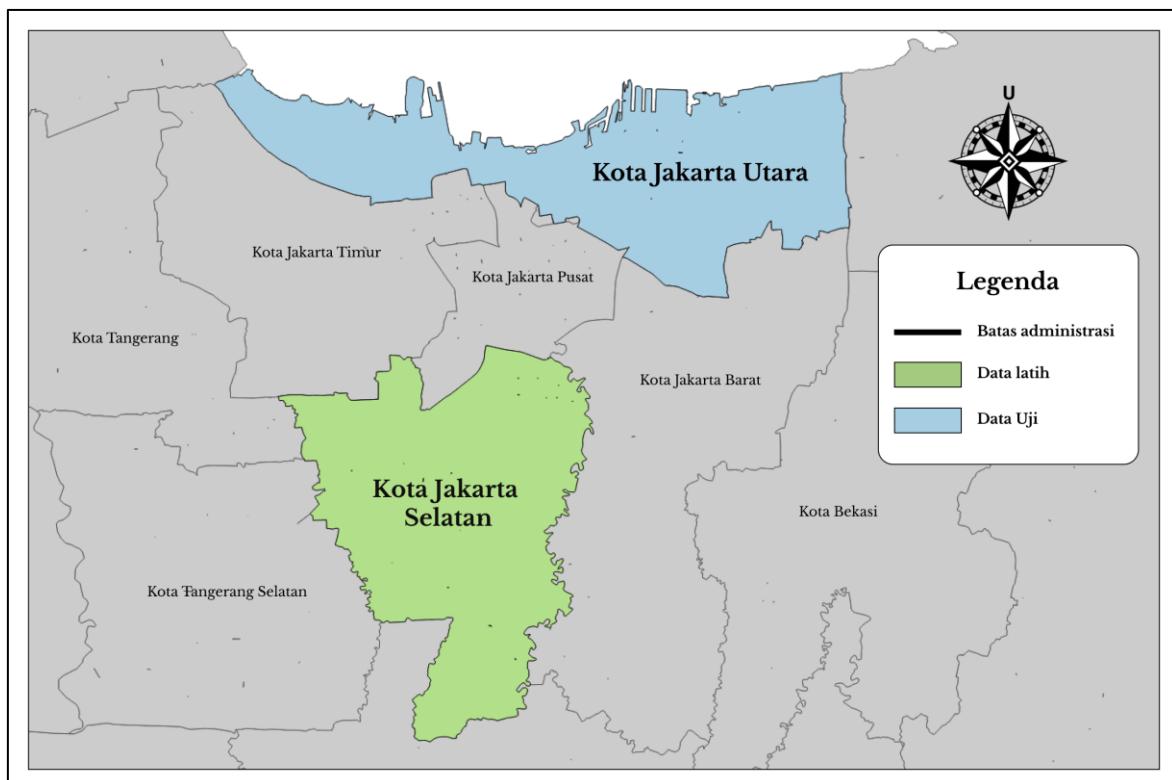


## BAB III

### PELAKSANAAN

#### III.1. Lokasi Penelitian

Penelitian ini dilaksanakan menggunakan data bangunan di wilayah Kota Administrasi Jakarta Selatan dan Kota Administrasi Jakarta Utara. Kota Administrasi Jakarta Selatan terletak antara  $106.31^\circ$  sampai  $106,97^\circ$  Bujur Timur (BT) dan  $5,16^\circ$  sampai  $6,37^\circ$  Lintang Selatan (LS) dengan luas sebesar  $154,00 \text{ km}^2$ . Sedangkan Kota Administrasi Jakarta Utara terletak antara  $106.71^\circ$  sampai  $106,97^\circ$  BT dan  $6,07^\circ$  sampai  $6,18^\circ$  LS dengan luas sebesar  $146,66 \text{ km}^2$ . Lokasi penelitian dapat dilihat pada gambar III.1 berikut:



**Gambar III.1.** Lokasi penelitian.

Pemilihan Kota Jakarta Selatan dan Jakarta Utara sebagai lokasi penelitian karena ketersediaan data penggunaan bangunan yang relatif lengkap dan memiliki kualitas data bangunan yang lebih baik dibandingkan wilayah lain di Indonesia. Kedua wilayah ini juga memiliki permukiman formal dan informal serta kawasan komersil yang cukup padat, sehingga sesuai dengan tujuan penelitian untuk memisahkan bangunan permukiman dan non-permukiman.

Sebagaimana yang terlihat pada Gambar III.1, pada penelitian ini penggunaan data bangunan dibagi menjadi data latih dan data uji. Kota Jakarta Selatan sebagai data latih dan Kota Jakarta Utara sebagai data uji. Pada penelitian ini, model dilatih dengan menggunakan data bangunan di Kota Jakarta Selatan untuk kemudian diuji kapabilitasnya dengan menggunakan data bangunan di Kota Jakarta Utara. Pembagian ini dilakukan untuk mengetahui apakah model mampu untuk menggeneralisasi pola bangunan di Kota Jakarta Selatan untuk wilayah perkotaan lainnya. Dengan pengujian pada wilayah yang berbeda, performa model untuk digunakan pada wilayah-wilayah lainnya di Indonesia akan diuji.

## III.2. Peralatan dan Bahan

### III.2.1. Peralatan Penelitian

Pada penelitian ini, peralatan yang digunakan dalam pembuatan model klasifikasi bangunan permukiman adalah sebagai berikut:

- a. Laptop Lenovo Ideapad Slim 5i, digunakan untuk persiapan data, pengolahan data, visualisasi data, dan penulisan laporan. Perangkat keras ini memiliki sistem Intel Core i5 generasi 11 dan RAM 8GB.
- b. VSCode versi 1.87.1, digunakan sebagai *compiler* dan perangkat lunak utama dalam pengoperasian Jupyter Notebook dengan bahasa pemrograman Python versi 3.8.
- c. QGIS, digunakan sebagai perangkat lunak utama dalam visualisasi data spasial.
- d. Microsoft Word, digunakan sebagai perangkat lunak utama dalam penulisan laporan penelitian.

### III.2.2. Bahan Penelitian

Pada penelitian ini, bahan yang digunakan dalam pembuatan model klasifikasi bangunan permukiman disajikan di dalam Tabel III.1. dan Tabel III.2..

**Tabel III.1.** Daftar bahan penelitian.

No	Data	Sumber	Keterangan
1	Data bangunan DKI Jakarta	OpenStreetMap	<ol style="list-style-type: none"><li>a. CRS : UTM 48 S (EPSG: 32748)</li><li>b. Format data vektor</li><li>c. Diunduh pada Maret 2024</li><li>d. Atribut : Kategori penggunaan bangunan</li></ol>

**Tabel III.2.** Daftar bahan penelitian (lanjutan).

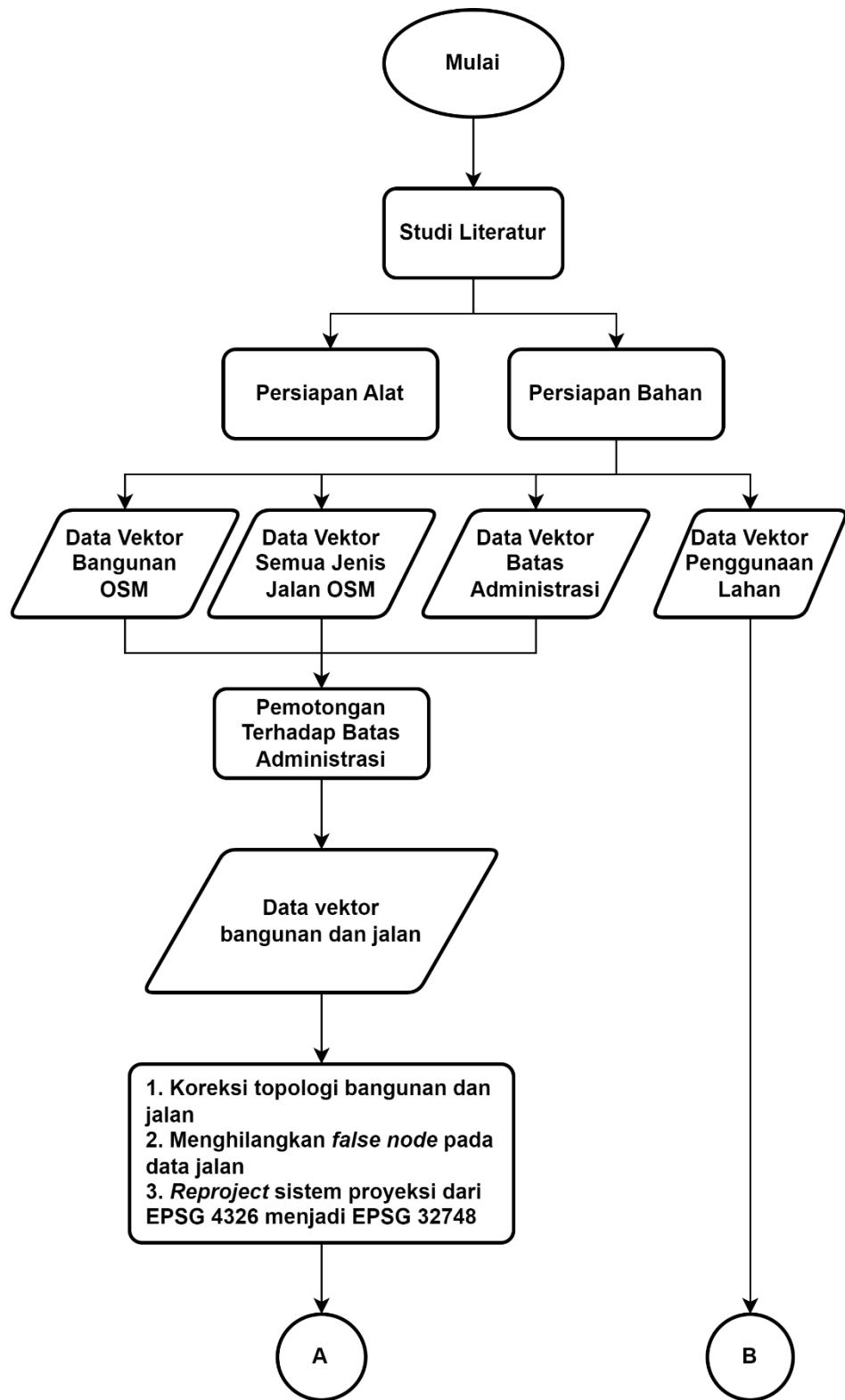
No	Data	Sumber	Keterangan
2	Data jaringan jalan DKI Jakarta	OpenStreetMap	a. CRS : UTM 48 S (EPSG: 32748) b. Format vektor c. Diunduh pada Maret 2024 d. Atribut : Jenis jalan
3	Data penggunaan lahan DKI Jakarta	Jakarta Satu	a. CRS : UTM 48 S (EPSG: 32748) b. Data merupakan hasil survei lapangan c. Format data vektor d. Diterbitkan tahun 2020 e. Atribut : Data penggunaan lahan
4	Peta batas administrasi DKI Jakarta	Jakarta Satu	a. CRS : UTM 48 S (EPSG: 32748) b. Format data vektor c. Diterbitkan tahun 2022 oleh Dinas Cipta Karya, Tata Ruang, dan Pertanahan Provinsi DKI Jakarta

### **III.3. Tahapan Penelitian**

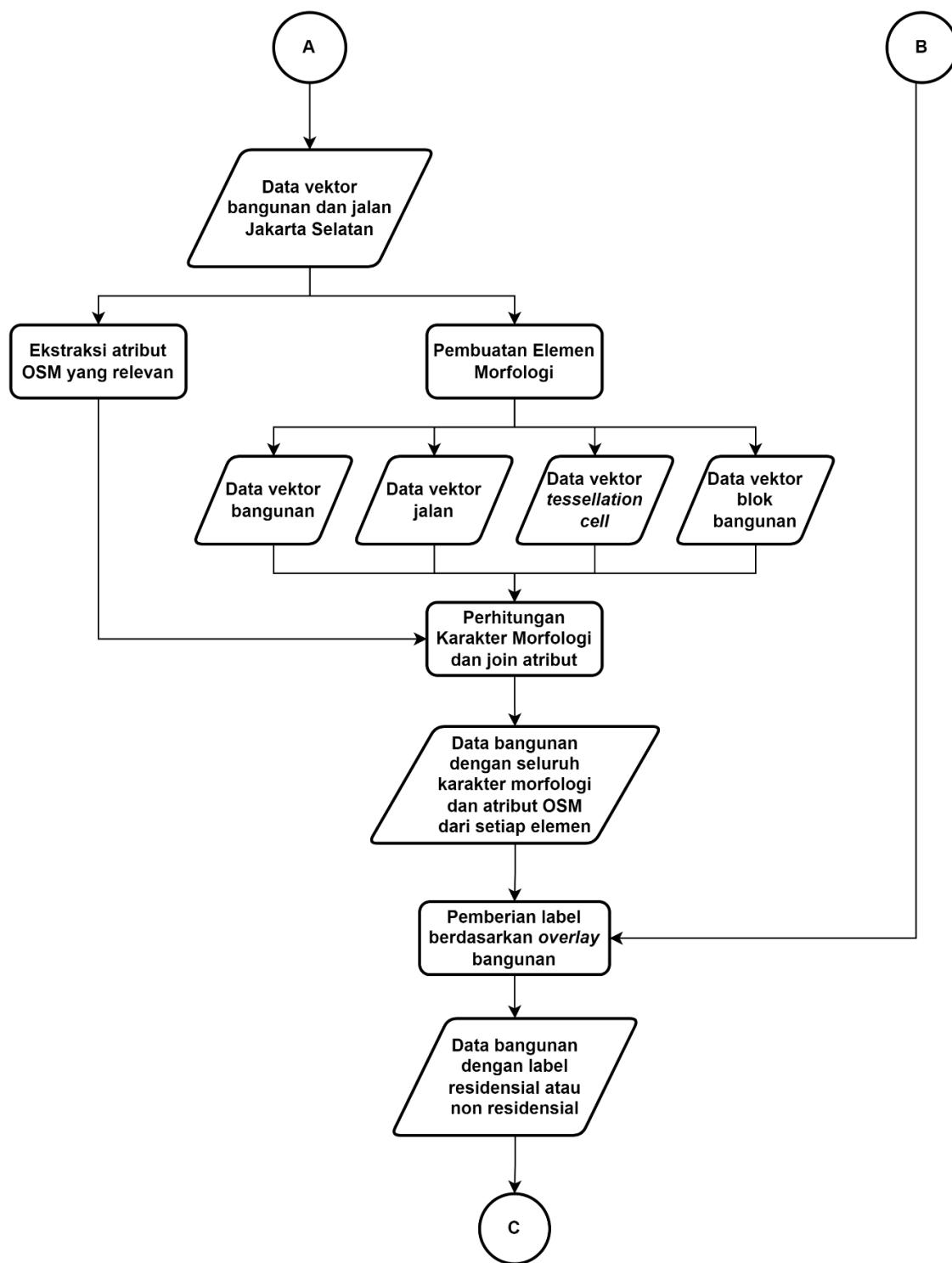
Penelitian dilakukan melalui beberapa tahapan. Secara umum tahapan penelitian terdiri dari pengumpulan data, kontrol kualitas data, pengolahan data, pembuatan model klasifikasi, evaluasi model klasifikasi. Tahapan dalam penelitian ditunjukkan pada Gambar III.2 sampai dengan Gambar III.4.

#### **III.3.1. Studi Literatur**

Tahapan awal dari penelitian ini adalah studi literatur, yang bertujuan untuk mengumpulkan dan menganalisis informasi yang relevan terhadap permasalahan yang ingin dipecahkan dari penelitian ini. Penggunaan literatur meliputi jurnal, artikel, dan buku yang memuat informasi seputar pertumbuhan penduduk, urbanisasi, permukiman, sistem informasi geografis, morfologi urban, *machine learning*, XGBoost, hingga evaluasi hasil klasifikasi. Literatur yang digunakan berasal dari sumber nasional dan internasional.



**Gambar III.2** Bagan alir penelitian.



**Gambar III.3** Bagan alir penelitian (lanjutan 1).

C

Pembuatan model  
dengan algoritma  
XGBoost

Hasil  
klasifikasi

Analisis dan  
evaluasi hasil  
klasifikasi

Accuracy > 0,8  
F1-Score > 0,8

Tidak

Penulisan Laporan

Selesai

**Gambar III.4** Bagan alir penelitian (lanjutan 2).

### **III.3.2. Persiapan Alat dan Bahan**

Dalam tahap ini, dipersiapkan perangkat lunak beserta *library* Python yang digunakan selama penelitian, yaitu:

- a. Visual Studio Code 1.89.1 : <https://code.visualstudio.com/download>
- b. QGIS 3.28.11 : <https://www.qgis.org/en/site/forusers/download.html>

- c. Python 3.12.2 : <https://www.python.org/downloads/>
- d. XGBoost 2.0.3 : Python Package Index (PyPI)
- e. Geopandas 0.14.3 : Python Package Index (PyPI)
- f. Momepy 0.7.0 : Python Package Index (PyPI)
- g. Scikit-Learn 1.4.1.post1. : Python Package Index (PyPI)
- h. OSMNX 1.9.1 : Python Package Index (PyPI)

Kemudian, perlu dipersiapkan juga data bangunan, jaringan jalan, dan *landuse* di Kota Jakarta Selatan. Data bangunan dan data jaringan jalan yang digunakan berasal dari *platform* OpenStreetMap yang diunduh dengan menggunakan *library* OSMNX, sebuah *library* yang digunakan untuk mengunduh, mengolah, menganalisis, dan memvisualisasikan data jaringan jalan dan fitur geospasial dari OpenStreetMap (Boeing, 2024). Untuk data penggunaan lahan, penelitian ini menggunakan data dari laman <https://jakartasatu.jakarta.go.id/portal/apps/sites/#/public>. Karena kondisi data yang besar dan membutuhkan kapasitas RAM yang tinggi, data kemudian dibagi berdasarkan kecamatan untuk setiap tahapan pemrosesan data. *Script* yang digunakan untuk mengunduh data bangunan dan jaringan jalan dapat dilihat pada Gambar III.5. dan Gambar III.6..

```
def download_building(place):
    with warnings.catch_warnings():
        warnings.simplefilter("ignore")
        kec = place.replace(" Jakarta Utara, Indonesia", "")
        print(f"{kec} Start!")
        osm.config(timeout=300)
    try:
        buildings = osm.features_from_place(place, {"building": True, "type": "polygon"})
        print(f"Successfully downloaded {kec}")
    except Exception as e:
        print(f"Failed to download {kec}: {e}")

    # Ensure directory exists
    output_dir = r"Data Collection\building\validation" # Adjusted to use forward slashes
    os.makedirs(output_dir, exist_ok=True) # Create the directory if it doesn't exist

    # Sanitize file name
    file_name = kec
    file_path = os.path.join(output_dir, f"building_{file_name}.csv")

    try:
        buildings = buildings[
            (buildings.geom_type == "Polygon") | (buildings.geom_type == "MultiPolygon")]
        buildings.reset_index(drop=True)
        buildings = buildings[["building", "geometry"]]
        for column in buildings.columns:
            for index, row in buildings.iterrows():
                if isinstance(row[column], list):
                    buildings.loc[index, column] = ""
        print(buildings.shape[0])
        buildings.to_csv(file_path)
        print(f"Successfully exported to {file_path}")
    except Exception as e:
        print(f"Failed to export {kec}: {e}")
```

Mengunduh data bangunan dari OSM

Menyeleksi atribut bangunan osm yang diunduh

**Gambar III.5.** *Script* yang digunakan untuk mengunduh data bangunan OSM.

```

for kec in [
    "Cilincing",
    "Kelapa Gading",
    "Koja",
    "Penjaringan",
    "Pademangan",
    "Tanjung Priok",
]:
    graphs = osm.graph_from_place(f"{kec}, Jakarta Utara, Indonesia")
    output_dir = r"Data Collection\street\validation"
    file_path = os.path.join(output_dir, f"streets_{kec}_drive.geojson")

try:
    streets = osm.graph_to_gdfs(
        graphs,
        nodes=False,
        edges=True,
        node_geometry=False,
        fill_edge_geometry=True,
    )
    streets = streets[~streets.highway.isna()].reset_index(drop=True)
    streets = streets[["geometry", "highway"]]
    for column in streets.columns:
        for index, row in streets.iterrows():
            if isinstance(row[column], list):
                streets.loc[index, column] = ""
    streets.to_file(file_path, driver="GeoJSON")
except Exception as e:
    print(f"Failed to export {e}")

```

Mengunduh jalan dalam format *graph* dari OSM

Mengubah format *graph* ke dalam format Geopandas jalan dalam format *graph* dari OSM

**Gambar III.6.** Script yang digunakan untuk mengunduh data jaringan jalan OSM.

### III.3.3 Data Pre-Processing

Pada tahapan ini, dilakukan beberapa proses dan kontrol kualitas untuk memastikan bahwa data yang digunakan memiliki kualitas yang cukup baik dan dapat digunakan di dalam penelitian. Adapun proses yang dilakukan, yaitu:

1. Mengatur sistem proyeksi ke dalam UTM 48S (EPSG:32748)
2. *Dissolve* data penggunaan lahan berdasarkan label bangunan
3. Menambahkan *unique id*
4. Menghilangkan *false nodes* atau *nodes* dengan derajat 2 atau lebih rendah dari data jaringan jalan dengan menggunakan fungsi ‘Momepy.remove\_false\_nodes( )’ pada Momepy

Menurut Chintya dkk. (2022), kualitas data spasial menjadi faktor penting dalam memanfaatkan data VGI. Data tapak bangunan dari OpenStreetMap merupakan hasil kontribusi komunitas yang seringkali memiliki kualitas yang beragam. Oleh karena itu, data tapak bangunan perlu dipastikan terlebih dahulu tidak memiliki *overlap* dan *containment*. Setelah melalui tahapan ini, data sudah tidak memiliki *overlap* dan juga

*containment*. Adapun potongan *script* yang digunakan dalam tahapan *pre-processing* terdapat pada Gambar III.7..

```
for file in files:
    gdf, kec = read_gdf(file)
    Mengecek keberadaan overlap
    # Check for overlap errors
    logging.info(f"{kec} : Overlap...")
    overlap_error = utils.check_overlap(gdf, "bID", 0.5)
    if overlap_error is not None:
        logging.info(f"{kec} : There is {len(overlap_error)} overlap...")
        overlap_path = os.path.join(error_path, f"{kec}_overlap.csv")
        overlap_error.to_csv(overlap_path)
    else:
        logging.info(f"{kec} : There is no overlap...")

    # Check for containment errors
    logging.info(f"{kec} : Containment...")
    contain_error = utils.checkContainment(gdf, "bID", 0.5)
    if contain_error is not None:
        logging.info(f"{kec} : There is {len(contain_error)} containment...")
        contain_path = os.path.join(error_path, f"{kec}_contain.csv")
        contain_error.to_csv(contain_path)
    else:
        logging.info(f"{kec} : There is no contain...")
```

Mengecek keberadaan *containment* pada data bangunan

Gambar III.7. Potongan *script* yang digunakan pada tahap *pre-processing* data bangunan.

### III.3.4. Pengolahan Data

Tahapan ini, merupakan tahap pengolahan data tapak bangunan dan jaringan jalan DKI Jakarta untuk mendapatkan karakter morfometrik suatu bangunan dan bangunan-bangunan di sekitarnya. Pada tahapan ini, atribut OSM juga diolah untuk digunakan sebagai salah satu fitur dalam klasifikasi bangunan permukiman.

#### III.3.4.1. Pembuatan Sel Teselasi

Tahapan ini merupakan tahapan pembuatan elemen morfologi tambahan, yaitu sel teselasi dari data tapak bangunan. Sel teselasi dibuat dengan menggunakan *library* Momepy yang ditulis oleh Fleischmann (2019). Setelah proses ini, setiap bangunan memiliki sel teselasi yang dapat menunjukkan hubungan dan interaksi antara satu bangunan dengan lainnya, seperti ketetanggaan bangunan, jarak antarbangunan, dan lain-lain. Adapun *script* yang digunakan untuk membuat sel teselasi dapat dilihat pada Gambar III.8..

```

def create_tess(gdf, name): # type: ignore
    logging.info(
        f"\n{name} Start...",
    )

    if gdf.crs != crs:
        gdf = gdf.set_crs(crs, allow_override=True)

    limit = mm.buffered_limit(gdf, 100)

    logging.info("Starting tessellation : %s", name)
    try:
        tessellation = mm.Tessellation(
            gdf, "bID_kec", limit=limit, verbose=True, use_dask=True
        )
        logging.info("Finish tessellation : %s", name)
    except Exception as e:
        logging.error("Failed in tessellation : %s - %s", name, e)
        raise

    tess = tessellation.tessellation
    bID_list = tess.bID_kec.to_list()

    bID_before = gdf.shape[0]
    gdf = gdf[gdf.bID_kec.isin(bID_list)]
    bID_after = gdf.shape[0]

    logging.info(f"\n{name} : bID before : {bID_before}, bID after : {bID_after}")

    gdf.to_csv(file.replace("_clean.csv", "_final.csv"), index=False) # update the bID
    tess.to_csv(os.path.join(folder_out, f"\n{name}_tessel.csv"), index=False)

    logging.info("Succeed: %s", name)

```

Membuat sel teselasi dengan bantuan *library* Momepy

**Gambar III.8.** *Script* yang digunakan untuk membuat sel teselasi.

#### III.3.4.2. Kalkulasi Karakter Morfometrik Bangunan dan Penggunaan Atribut OSM

Tahapan ini bertujuan untuk melakukan perhitungan karakter morfometrik bangunan dengan bantuan *library* Momepy. Adapun karakter yang digunakan merupakan karakter yang berkaitan dengan dimensi, bentuk, distribusi spasial, intensitas, dan diversitas bangunan. Daftar dan definisi dari seluruh karakter morfometrik yang dihitung dapat dilihat pada Lampiran B. Kemudian, atribut OSM juga menjadi fitur tambahan dalam model klasifikasi bangunan permukiman.

Umumnya, terdapat perbedaan yang signifikan antara bangunan permukiman dan non-permukiman karena probabilitas bangunan non-permukiman untuk memiliki atribut tambahan lebih tinggi daripada bangunan permukiman. Sehingga ketersediaan atribut tertentu dapat menjadi informasi tambahan bagi model dalam menentukan apakah suatu bangunan merupakan bangunan permukiman atau non-permukiman. Adapun potongan *script* yang digunakan untuk mengkalkulasi karakter morfometrik dapat dilihat pada Gambar III.9..

```

def ft_building_tess_geom(building, tessel):
    logging.info(f"{name}: Building geometry Feature Group...")
    building_gdf = building
    tessel_gdf = tessel

    logging.info(f"{name}: Building...")
    building_gdf["b_area"] = building_gdf.area
    building_gdf["b_perimeter"] = mm.Perimeter(building_gdf).series
    building_gdf["b_convexity"] = mm.Convexity(building_gdf).series
    building_gdf["b_circular_comp"] = mm.CircularCompactness(
        building_gdf
    ).series
    ccd = mm.CentroidCorners(building_gdf, verbose=False)
    building_gdf["b_ccd_means"] = ccd.mean
    building_gdf["b_ccd_std"] = ccd.std
    del gc(ccd)
    building_gdf["b_corners"] = mm.Corners(building_gdf, verbose=False).series
    building_gdf["b_elong"] = mm.Elongation(building_gdf).series
    building_gdf["b_lal"] = mm.LongestAxis
    building_gdf["b_eri"] = mm.EquivalentR
    building_gdf, "b_area", "b_perimet"
).series
    building_gdf["b_orientation"] = mm.Ori
    building_gdf, verbose=False
).series

```

Contoh perhitungan karakter jumlah sudut dan elongasi bangunan dengan bantuan *library* Momepy

**Gambar III.9.** Potongan *script* yang digunakan untuk mengkalkulasi karakter morfometrik.

#### III.3.4.3. Pemberian Label Bangunan

Tahapan ini bertujuan untuk memberikan label bangunan sesuai dengan data penggunaan lahan yang dirilis oleh Pemerintah Provinsi DKI Jakarta melalui laman Jakarta Satu. Data tersebut berisikan informasi penggunaan petak-petak lahan yang digunakan sebagai *ground truth* dalam penelitian ini. Bangunan OSM kemudian di-*overlay* di atas data penggunaan lahan, dan kemudian diidentifikasi apakah bangunan tersebut berada di dalam penggunaan lahan permukiman atau non-permukiman. Data penggunaan lahan yang dirilis oleh Pemerintah Provinsi DKI Jakarta pada tahun 2020 silam, merupakan data hasil survey lapangan dan berisi berbagai macam penggunaan lahan. Tabel III.2 menunjukkan pembagian kelas dari setiap jenis penggunaan lahan.

Pembagian antara permukiman dan non-permukiman berfokus pada perbedaan bangunan tempat tinggal yang berada di kawasan permukiman. Oleh karena itu, rumah susun atau apartemen tidak dikategorikan sebagai bagian dari permukiman, mengingat lokasi dan bentuk bangunannya yang umumnya sangat berbeda dari rumah-rumah biasa. Seperti yang ditunjukkan pada Gambar III.10, bangunan-bangunan dalam permukiman

biasanya berdekatan satu sama lain, sedangkan bangunan non-permukiman lebih tersebar dan terpencar.

**Tabel III.2** Daftar kelas untuk setiap jenis penggunaan lahan.

Kelas	Penggunaan Lahan
Permukiman	Hunian (Kecuali Rumah Susun)
Non-Permukiman	Budaya, Campuran, Industri, Jasa, Keagamaan, Khusus, Pelayanan Kesehatan, Pelayanan Pendidikan, Pelayanan Umum, Pemerintahan, Penggunaan Lain, Penyimpanan, Perhotelan, Perkantoran dan Perdagangan, Terminal, Wisata, dan Hunian (Rumah Susun)

Dalam pemberian label kepada setiap bangunan, dilakukan *overlay* antara data bangunan dan penggunaan lahan sehingga setiap bangunan memiliki label sesuai dengan jenis penggunaan lahan yang memiliki luas perpotongan yang paling besar. Bangunan OSM yang tidak berpotongan dengan data penggunaan lahan kemudian dihapuskan dari penelitian. Sebagai contoh di Kecamatan Setiabudi, Jakarta Selatan awalnya memiliki total 37.484 bangunan menjadi 37.348 setelah proses pemberian label karena terdapat bangunan yang tidak berpotongan dengan penggunaan lahan apapun. Setelah proses ini, setiap data sudah memiliki karakteristik morfometrik dan label bangunan. *Script* yang digunakan untuk memberikan *label* pada data dapat dilihat pada Gambar III.11..



**Gambar III.10.** Tampilan kelas permukiman (kiri) dan non-permukiman (kanan) di Kecamatan Menteng, Jakarta Pusat.

```

def add_label(gdf, label, name):
    logging.info("Add label start!")

    intersections = gdf.overlay(label, how="intersection")
    logging.info(f"{name}: Intersections done...")
```

Melakukan *overlay* antara data penggunaan lahan dengan data bangunan

```

df = intersections[["bID_kec", "label", "geometry"]]
df = df.dissolve(["bID_kec", "label"])
logging.info(f"{name}: Dissolve done...")
```

```

df.reset_index(inplace=True)
df["inter_area"] = df.geometry.area
```

```

result = df.groupby("bID_kec", group_keys=False).apply(
    lambda x: x.loc[x["inter_area"].idxmax()], include_groups=True
)
```

Memilih penggunaan lahan dengan luas *overlap* yang terbesar sebagai label

```

result = result.drop(columns="bID_kec").reset_index()
final_gdf = gdf.merge(
    result.drop(columns=["inter_area", "geometry"]),
    left_on="bID_kec",
    right_on="bID_kec",
)
```

```

final_gdf = gpd.GeoDataFrame(final_gdf)

return final_gdf

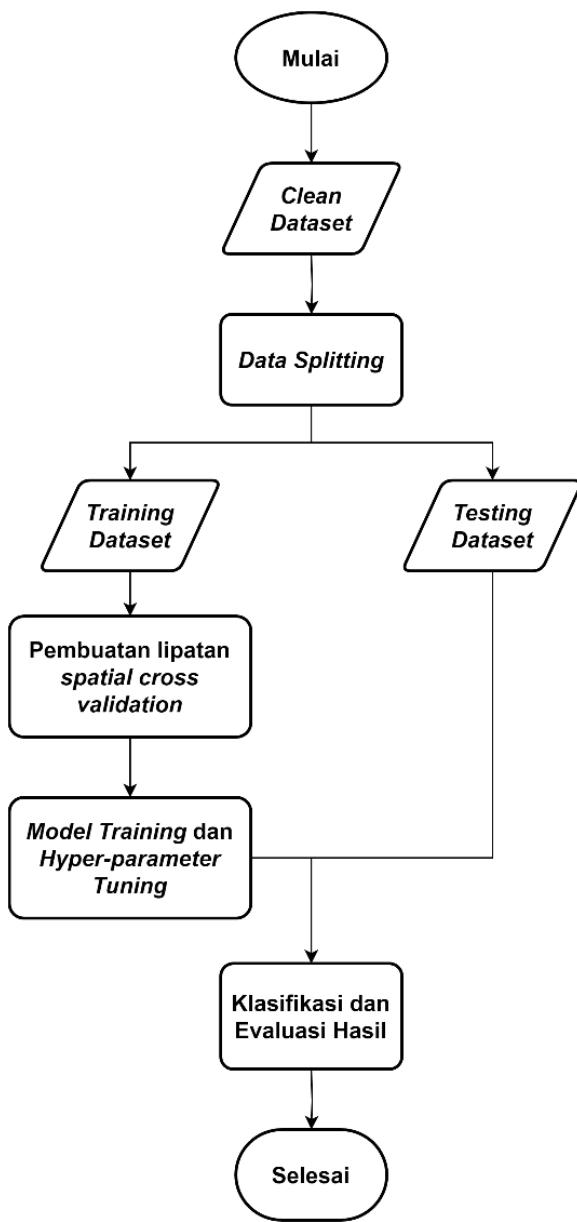
```

**Gambar III.11.** Script yang digunakan untuk memberikan label pada bangunan.

### III.3.5. Pembuatan Model Klasifikasi

Tahapan ini merupakan tahapan untuk membuat model klasifikasi dengan algoritma XGBoost. Model dibuat dengan menggunakan *library* XGBoost dan SkicitLearn pada perangkat lunak VSCode. Adapun tahapan terperinci dari pembuatan model klasifikasi disajikan pada Gambar III.12..

Data yang telah bersih dan dilabeli sebagai permukiman atau non-permukiman kemudian akan dibagi menjadi lima lipatan dengan metode validasi silang spasial atau *spatial cross validation* (CV). Script yang digunakan untuk *spatial* CV dapat dilihat pada Gambar III.13.. Dengan menggunakan kombinasi *spatial* CV yang telah dibuat, model standar akan divalidasi untuk mengetahui kemampuan model tanpa adanya konfigurasi *hyperparameter*.



**Gambar III.12.** Diagram alir pembuatan model XGBoost.

Dengan mengetahui performa model standar terhadap data validasi atau *spatial CV*, selanjutnya dilakukan optimasi *hyperparameter* untuk meningkatkan performa model dalam memprediksi penggunaan bangunan permukiman. Optimasi *hyperparameter* dilakukan dengan menggunakan teknik *Randomized Grid Search*,

Setelah melalui tahapan optimasi *hyperparameter*, model yang telah dioptimasi selanjutnya digunakan untuk mengklasifikasi data bangunan di setiap kecamatan di wilayah Kota Administrasi Jakarta Utara. Data bangunan di wilayah Jakarta Utara berperan sebagai data uji untuk mengetahui kapabilitas prediksi model. Adapun *script* yang digunakan untuk membuat model dapat dilihat pada Gambar III.14. dan III.15..

```

files = glob.glob(os.path.join(folder_in, "*.csv"))
for key, items in cross_val.items():
    logging.info(f"{key} start...")
    val_files = []
    train_files = []
    logging.info(f"{key}: start splitting...")
    # Split the training and validation in cv
    for file in files:
        if any(item in file for item in items):
            val_files.append(file) # Validation
        else:
            train_files.append(file) # Training

logging.info(f"{key}: start read file...")
cv_val_list = []
for file in val_files:
    df = pd.read_csv(file)
    cv_val_list.append(df)

cv_train_list = []
for file in train_files:
    df = pd.read_csv(file)
    cv_train_list.append(df)

logging.info(f"{key}: start concat...")
val_df = pd.concat(cv_val_list, ignore_index=True)
train_df = pd.concat(cv_train_list, ignore_index=True)

```

Potongan *script* untuk membagi model menjadi latih dan validasi untuk *cross validation*

**Gambar III.13.** *Script* yang digunakan untuk membuat spatial CV.

```

for train, val in zip(training_files, validation_files):
    cv = os.path.basename(train).replace("_train.csv", "")
    logging.info(f"{cv}: Start...")

    train_df = pd.read_csv(train)
    val_df = pd.read_csv(val)

    try:
        train_df = train_df.drop(columns=dropped_columns, axis=1)
    except:
        val_df = val_df.drop(columns=dropped_columns, axis=1)

    val_df = val_df[train_df.columns] # to match the training columns

    X_train = train_df.drop(columns=["bID_kec", "label", "geometry"], axis=1).copy()
    y_train = train_df[["label"]].copy()

    X_val = val_df.drop(columns=["bID_kec", "label", "geometry"], axis=1).copy()
    y_val = val_df[["label"]].copy()

    logging.info(f"{cv}: Start DMatrix...")
    # Creating DMatrix for training and validation sets
    dtrain = xgb.DMatrix(data=X_train, label=y_train)
    dval = xgb.DMatrix(data=X_val, label=y_val)

```

Transformasi data latih dan validasi ke dalam format DMatrix, format yang digunakan dalam pembuatan model dengan XGBoost

**Gambar III.14.** *Script* yang digunakan untuk membuat model XGBoost.

```

logging.info(f"{cv}: Start training...")

eval_metric = "auc"

params = {
    "objective": "binary:logistic",
    "eval_metric": eval_metric,
}

evals = [(dtrain, "train"), (dval, "eval")]
evals_result = {}

# Training the model
clf_xgb = xgb.train(
    params,
    dtrain,
    num_boost_round=500,
    evals=evals,
    evals_result=evals_result,
    early_stopping_rounds=30,
    verbose_eval=False,
)

```

Potongan *script* untuk melatih model XGBoost

**Gambar III.15.** *Script* yang digunakan untuk membuat model XGBoost (Lanjutan).

Dalam menguji performa model terhadap data bangunan Jakarta Utara, dilakukan beberapa eksperimen dengan menambahkan sebagian data lokal ke dalam data latih untuk mengetahui signifikansi penambahan sebagian informasi morfologi bangunan dari wilayah yang diuji terhadap performa model. Data lokal yang dimaksud merupakan sebagian data bangunan dari wilayah yang akan diuji. Sebelum memprediksi penggunaan bangunan dari suatu kecamatan di Jakarta Utara, sebagian bangunan dari kecamatan tersebut akan dimasukkan ke dalam data latih sehingga model memiliki sedikit informasi terkait kondisi morfologi dari kecamatan tersebut. Dalam eksperimen ini, akan dilakukan tiga kali percobaan dengan persentase bangunan lokal sebesar 2,5%, 5%, dan 10%.

### III.3.6. Analisis dan Evaluasi Model Klasifikasi

Tahap ini adalah tahapan untuk menganalisis dan mengevaluasi hasil prediksi dari model klasifikasi yang menggunakan algoritma XGBoost. Tahapan ini krusial untuk menentukan apakah model memiliki tingkat akurasi yang memadai ketika dibandingkan dengan data penggunaan lahan di DKI Jakarta. Proses ini mengandalkan penggunaan

metrik dan teknik evaluasi yang spesifik untuk memastikan bahwa hasil analisis akurat, dapat diandalkan, dan relevan dengan tujuan analisis.

Evaluasi model klasifikasi dilakukan dengan matriks konfusi, yang menjadi dasar perhitungan beberapa metrik, seperti akurasi, *recall*, *balanced accuracy*, dan *F1-Score*. Dalam penelitian ini, metrik akurasi, dan *F1-Score* dijadikan indikator utama untuk menilai kepuasan terhadap hasil klasifikasi. Berdasarkan penelitian Sturrock dkk. (2018) dan Atwal dkk. (2022), angka 0,8 ditetapkan sebagai batas minimum untuk akurasi dan *F1-Score*.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Model klasifikasi permukiman telah dibuat oleh beberapa peneliti sebelumnya. Konsep pemanfaatan data morfologi bangunan yang didapatkan dari data tapak bangunan, jaringan jalan, dan atribut OSM digunakan sebagai variabel tambahan, dengan tujuan menciptakan sebuah model klasifikasi yang dapat memprediksi penggunaan bangunan tanpa perlu melakukan survei lapangan ke setiap bangunan di suatu wilayah.

Penelitian ini mencoba untuk melakukan hal serupa di Indonesia, khususnya di DKI Jakarta, sebagai *proof of concept* bahwa metodologi ini bisa diterapkan pada kondisi permukiman di Indonesia yang relatif padat dan tidak teratur dengan memanfaatkan data OSM yang tersedia di seluruh wilayah di Indonesia. Karena keterbatasan data penggunaan lahan di wilayah lainnya, penelitian ini menggunakan data DKI Jakarta, khususnya Kota Jakarta Selatan, untuk membuat model yang diuji kapabilitasnya dengan data bangunan di Kota Jakarta Utara. Pada penelitian ini, dilakukan dua percobaan untuk mengevaluasi kemampuan model untuk memprediksi bangunan permukiman dari data yang belum pernah dilihat sebelumnya, dan untuk menguji apakah performa model dapat ditingkatkan dengan kehadiran data lokal.

Pada bab ini, dibahas seluruh proses dan tantangan dalam pembuatan model klasifikasi permukiman dengan algoritma XGBoost, mulai dari proses akuisisi data sampai dengan interpretasi fitur signifikan.

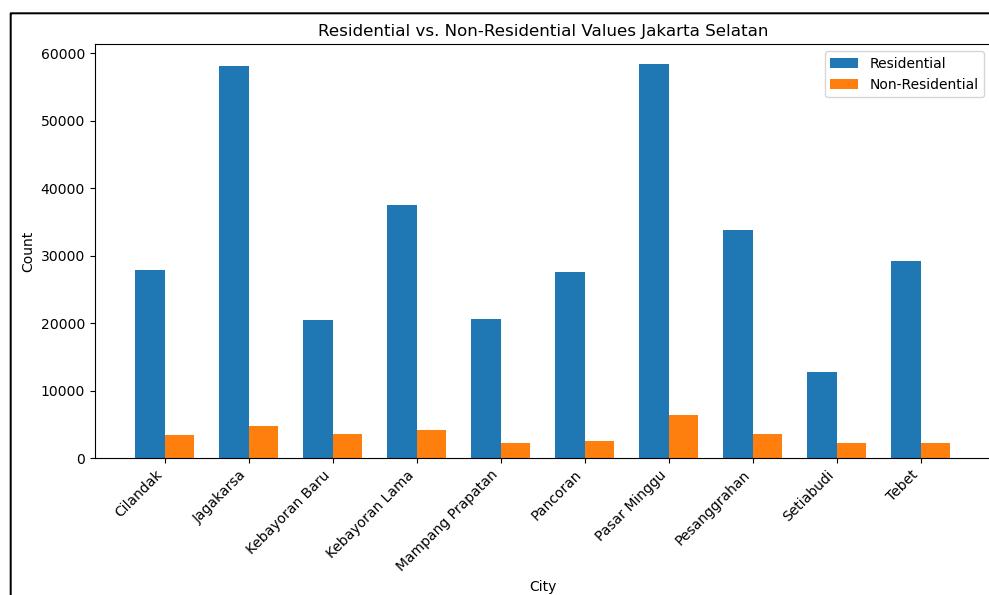
#### **IV.1. Pengolahan Data**

Penelitian ini menggunakan data tapak bangunan dan jaringan jalan dari OSM untuk membuat model klasifikasi permukiman. Sebelum data dapat digunakan, data perlu melalui proses kontrol kualitas terlebih dahulu untuk menghilangkan kesalahan pada data. Kemudian, dilakukan perhitungan karakteristik morfologi yang digunakan untuk membuat model klasifikasi.

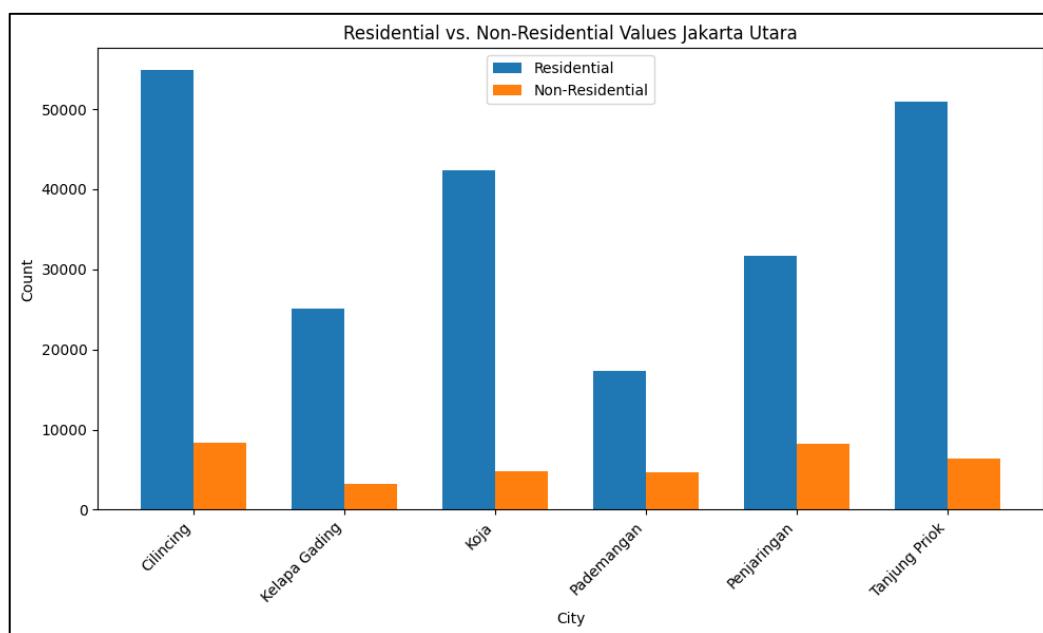
##### **IV.1.1. Data yang Digunakan**

Data bangunan yang digunakan dalam penelitian ini meliputi seluruh bangunan yang terletak di wilayah Kota Jakarta Selatan dan sebagian bangunan di wilayah Kota

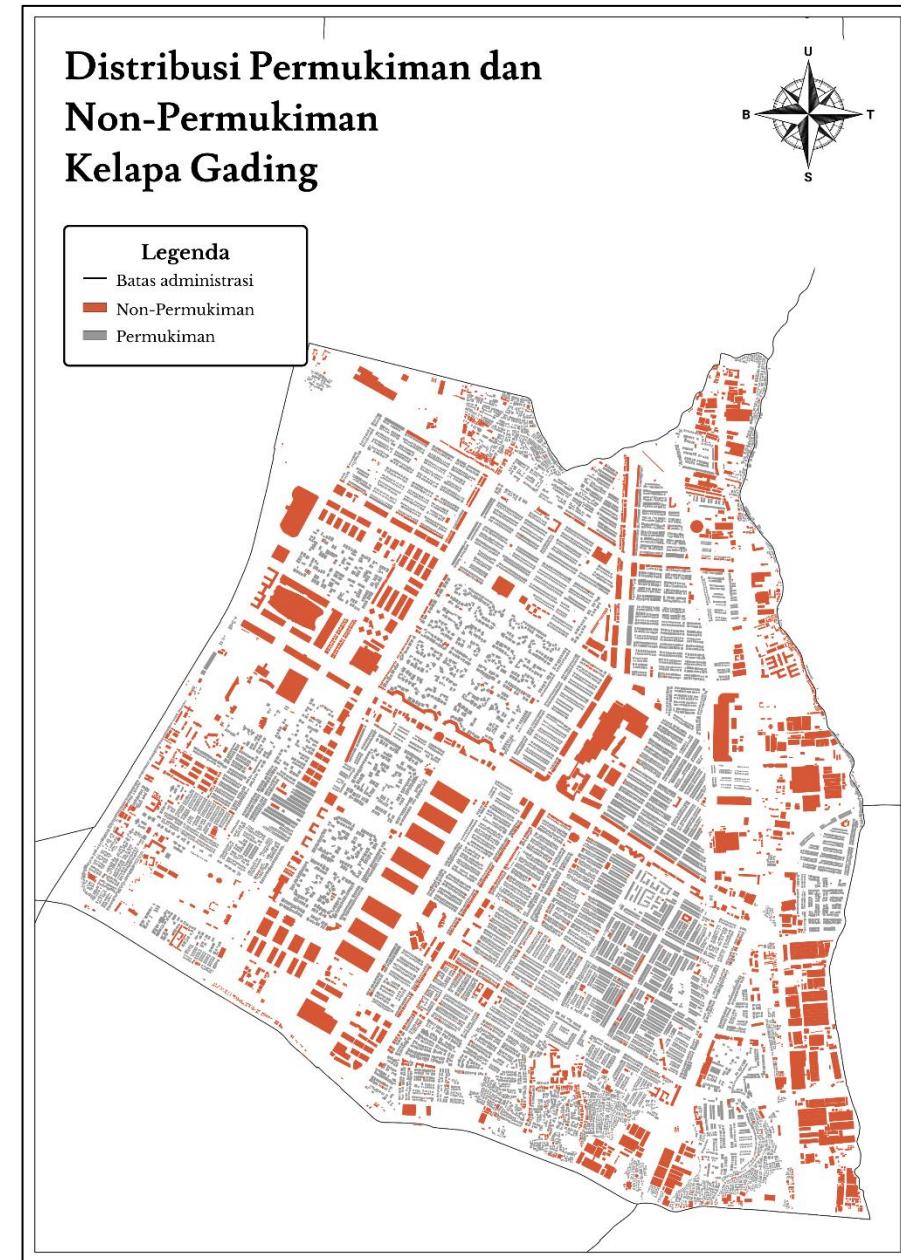
Jakarta Utara, dengan total sekitar 360 ribu bangunan di wilayah Kota Jakarta Selatan dan sekitar 260 ribu bangunan di wilayah Kota Jakarta Utara. Data yang telah diunduh dari laman OSM kemudian diproyeksikan ke dalam sistem proyeksi UTM 48S (EPSG:32748), yang merupakan sistem proyeksi UTM di wilayah DKI Jakarta dan sekitarnya. Sedangkan, data penggunaan lahan diambil dari laman resmi Pemerintah Provinsi DKI Jakarta.



**Gambar IV.1.** Distribusi jumlah bangunan permukiman dan non-permukiman di wilayah Jakarta Selatan.



**Gambar IV.2.** Distribusi jumlah bangunan permukiman dan non-permukiman di wilayah Jakarta Utara

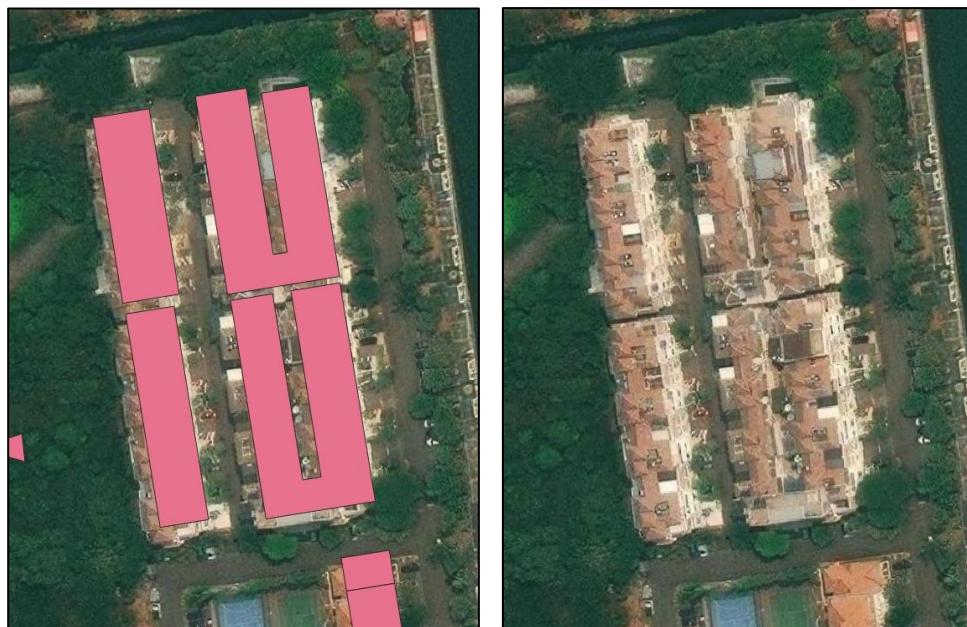


**Gambar IV.3.** Contoh distribusi bangunan permukiman dan non-permukiman di Kecamatan Kelapa Gading, Jakarta Utara

Berdasarkan visualisasi distribusi data bangunan pada Gambar IV.1 dan Gambar IV.2 terlihat bahwa terdapat kondisi data tidak seimbang antara data permukiman dan non-permukiman di dalam data Jakarta Selatan dan Jakarta Utara. Kondisi ini dapat menyebabkan model klasifikasi mengalami bias untuk memprediksi suatu bangunan sebagai bangunan permukiman daripada bangunan non-permukiman.

Selain ketidakseimbangan data, permasalahan lain yang muncul adalah kualitas data OSM yang kurang baik. Permasalahan ini meliputi geometri bangunan yang tumpang

tindih (*overlap*) dan adanya beberapa bangunan yang terdigitasi menjadi satu. Sehingga, dibutuhkan pengecekan kualitas untuk memastikan bahwa data bangunan tidak memiliki kesalahan geometri dan dapat digunakan untuk pembuatan model klasifikasi permukiman dan non-permukiman.



**Gambar IV.4.** Contoh geometri bangunan yang tidak sesuai.

Pada Gambar IV.4. ditemukan bahwa terdapat geometri bangunan OSM yang tidak sesuai dengan bentuk bangunan di lapangan, terlihat dari bagaimana beberapa bangunan permukiman yang tergabung menjadi satu bangunan. Hal ini dapat berdampak terhadap kemampuan model untuk memprediksi penggunaan bangunan. Di dalam data uji dan latih, ditemukan beberapa bangunan permukiman yang terdigigitasi menjadi satu bangunan besar. Hal ini dapat berdampak pada kekeliruan model untuk mengidentifikasi bangunan dan mempelajari karakteristik bangunan itu sendiri. Pada penelitian ini, data dengan kondisi tersebut, tidak dikoreksi.

#### IV.1.2 Perhitungan Karakter Morfometrik

Setelah data tapak bangunan dan jaringan jalan tersedia, tahapan selanjutnya adalah membuat sel teselasi dan melakukan perhitungan karakter morfometrik dengan menggunakan *library* Momepy. Karakter-karakter tersebut dihitung dari data bangunan, sel teselasi, jaringan jalan, dan node jalan dengan menggunakan *library* Momepy. Dengan menggunakan karakter morfometrik, model mempelajari pola dan tren karakter

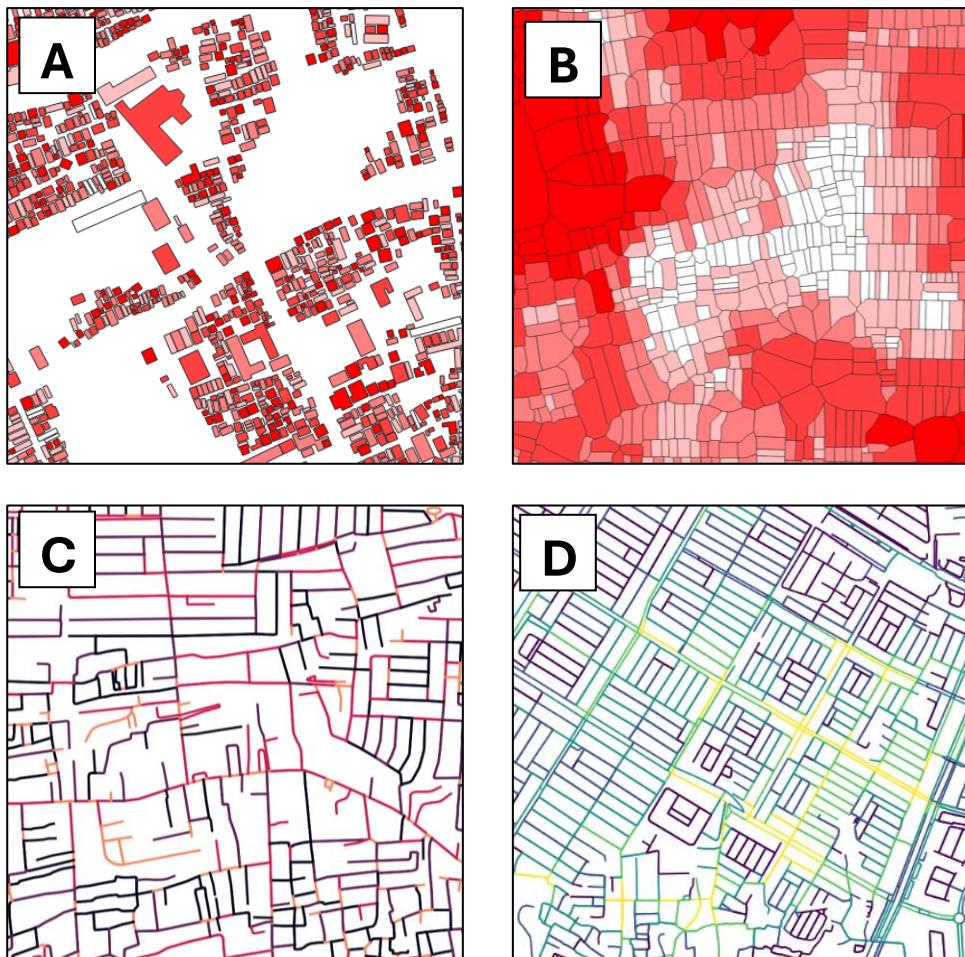
morfometrik yang terdapat pada bangunan permukiman dan non-permukiman untuk membedakan keduanya.

Karakter morfometrik yang digunakan adalah karakter yang berkaitan dengan kondisi geometri bangunan dan sel teselasi, kondisi geometri bangunan dan sel teselasi dari bangunan tetangga, dan geometri dari serta sentralitas jaringan jalan di sekitar bangunan. Daftar kelompok karakter morfometrik yang digunakan tertera pada Tabel III.2. Daftar lengkap karakter yang digunakan dalam penelitian ini tertera pada Lampiran B.

**Tabel IV.1.** Daftar kelompok karakter morfometrik yang digunakan.

<b>Kelompok</b>	<b>Sub-Kelompok</b>	<b>Jumlah</b>	<b>Contoh</b>
Bangunan	Geometri bangunan	12	Luas bangunan
	Geometri bangunan terdekat	3	Kedekatan antarbangunan tetangga
	Bangunan dalam jarak 50, 150, dan 300 m	$23 \times 3$	Jumlah bangunan di dalam <i>buffer</i> 300 m
	Atribut OSM	1	Penggunaan bangunan OSM
Sel Teselasi	Geometri sel teselasi	8	Keliling sel teselasi
	Geometri sel teselasi terdekat	2	Luas area yang ditutupi oleh sel teselasi dan sel teselasi tetangganya
	Sel teselasi dalam jarak 50, 150, dan 300 m	$21 \times 3$	Rata-rata luas dari sel teselasi di dalam <i>buffer</i> 50 m
Jalan	Jalan terdekat	15	Panjang jalan terdekat
	Jalan dalam jarak 50, 150, 300 m	$13 \times 3$	Standar deviasi panjang jalan di dalam <i>buffer</i> 150 m
	Persimpangan terdekat	2	Jarak persimpangan bangunan terdekat dengan bangunan
	Jumlah persimpangan dalam jarak 50, 150, 300 m	$1 \times 3$	Jumlah persimpangan terdekat di dalam <i>buffer</i> 300 m
	Atribut OSM	1	Jenis jalan OSM

Pada penelitian ini, digunakan total 218 karakter yang digunakan sebagai *input* dari model klasifikasi permukiman. Fitur-fitur tersebut dipilih berdasarkan beberapa studi mengenai morfologi urban dan studi *spatial network* (Lobo dkk., 2020; Boeing, 2020; Louf dan Barthelemy, 2014; Fleishmann dkk., 2021; Milojevic-Dupont dkk., 2020).



**Gambar IV.5.** Visualisasi karakter morfometrik. (A) Karakter *elongation* setiap bangunan dengan rentang 0-1. (B) Karakter *covered area* sel teselasi dan sekitarnya dalam satuan  $m^2$ . (C) Karakter keterbukaan jalan dengan rentang 0-1. (D) Karakter *street meshedness* dengan rentang 0-1.

#### IV.1.2.1. Geometri bangunan dan sel teselasi

Fitur geometri bangunan memiliki informasi yang berkaitan dengan penggunaan bangunan tersebut (Lloyd dkk., 2020). Fitur spesifik seperti luas bangunan sudah memiliki kemampuan prediktif yang cukup baik untuk mengidentifikasi apakah suatu bangunan merupakan bangunan permukiman atau non-permukiman. Pada penelitian ini, ditambahkan juga beberapa karakteristik lainnya seperti keliling bangunan dan indeks *convexity*

bangunan. Karakteristik tambahan tersebut dapat membantu model untuk membedakan topologi bangunan berdasarkan karakteristik selain luas bangunan.

Fitur dari sel teselasi juga ditambahkan untuk merepresentasikan relasi suatu bangunan dengan kondisi bangunan di sekitarnya (Fleischmann dkk., 2020). Dengan menambahkan bentuk dari sel teselasi, model dapat melihat bagaimana kedekatan suatu bangunan dengan bangunan di sekitarnya.

#### IV.1.2.2. Kondisi sekitar bangunan

Kondisi di sekitar bangunan memiliki informasi penting yang berkaitan dengan penggunaan bangunan. Kondisi tersebut direpresentasikan dengan fitur yang mewakili objek di sekitarnya, seperti bangunan terdekat, jalan terdekat, keselarasan orientasi bangunan terhadap bangunan di sekitarnya, dan jumlah bangunan dalam *buffer* 50, 150, dan 300 meter. Fitur-fitur tersebut dapat menunjukkan kondisi dari *spatial auto-correlation* di dalam data.

Dalam perhitungan karakter morfometrik untuk fitur yang merepresentasikan bangunan, digunakan *library* Pysal untuk mendapatkan bobot spasial yang menunjukkan relasi dari bangunan dengan objek-objek di sekitarnya. Pada penelitian ini digunakan bobot spasial *distance band* dan juga *queen contiguity* untuk menemukan relasi antarbangunan.

#### IV.1.2.3. Posisi bangunan relatif terhadap kecamatannya

Posisi bangunan relatif terhadap kecamatannya dapat dilihat dengan karakteristik *meshedness*, *closeness*, dan *betweenness* dari jalan yang berada paling dekat dengan bangunan. Karakteristik tersebut memperhitungkan aspek posisi jalan di dalam suatu wilayah.

## **IV.2. Hasil dan Evaluasi Model XGBoost**

Pada bagian ini, dibahas mengenai hasil dan evaluasi dari model klasifikasi permukiman yang dibuat. Pembahasan dimulai dari evaluasi model awal tanpa *hyperparameter tuning*, dilanjutkan dengan proses *hyperparameter tuning*, evaluasi model setelah *hyperparameter tuning*, dan terakhir menguji kemampuan model untuk mengklasifikasikan data uji.

Dalam penelitian ini, digunakan beberapa metrik seperti *recall*, *balanced accuracy*, dan *F1-Score* untuk melihat kemampuan model dalam memprediksi bangunan permukiman dan non-permukiman.

Selain itu, diterapkan *spatial cross validation* (CV) atau validasi silang spasial pada pelatihan dengan membagi data latih berdasarkan kecamatan dengan jarak yang saling berjauhan. Pendekatan ini bertujuan untuk menguji kemampuan model dalam memprediksi data bangunan dari wilayah yang belum pernah dilihat sebelumnya (Meyer dkk., 2019). Data latih dibagi menjadi lima kelipatan yang berbeda dengan jumlah yang relatif sama. Pembagian *spatial CV* dapat dilihat pada Tabel IV.2.

CV spasial digunakan pada tahapan pembuatan model awal dan *hyperparameter tuning*. Untuk setiap kombinasi CV, setiap kecamatan digunakan sebagai data validasi, sementara kecamatan lainnya digunakan sebagai data latih.

**Tabel IV.2.** Pembagian CV spasial

CV	Kecamatan	Jumlah bangunan
1	Jagakarsa dan Setiabudi	77.888
2	Pasar Minggu dan Pancoran	94.979
3	Pesanggrahan dan Kebayoran Lama	79.043
4	Cilandak, Mampang Prapatan, dan Tebet	85.636
5	Kebayoran Baru, Tebet, dan Cilandak	86.766

Model yang telah melalui proses *hyperparameter tuning* kemudian diuji dengan data bangunan di setiap kecamatan di Kota Jakarta Utara. Pada penelitian ini, dilakukan beberapa percobaan, yaitu pengujian tanpa ada informasi bangunan lokal, pengujian dengan 2.5%, 5%, dan 10% informasi bangunan lokal.

#### IV.2.1. Pembuatan Model XGBoost Standar

Pengujian performa model XGBoost dilakukan dengan parameter *standar* tanpa ada optimasi *hyperparameter* untuk setiap lipatan dari CV spasial. Adapun parameter standar dari XGBoost ditampilkan pada Tabel IV.3.

**Tabel IV.3.** Nilai standar dari *hyperparameter* XGBoost.

<i>Hyperparameter</i>	Nilai standar
n_estimators	100
learning_rate (eta)	0,3
Gamma	0
max_depth	6
min_child_weight	1
Subsample	1
colsample_bytree	1
reg_lambda	1
reg_alpha	0

Hasil percobaan model standar di setiap lipatan CV dapat dilihat pada Tabel IV.4. Dari hasil percobaan model standar untuk setiap CV, ditemukan bahwa model memiliki akurasi rata-rata sebesar 0,921 dan F1-Score rata-rata sebesar 0,688. Hal tersebut dapat terjadi karena kondisi data yang tidak seimbang antara bangunan permukiman dan non-permukiman, yang berakibat pada model yang mampu untuk mengidentifikasi bangunan permukiman jauh lebih baik dibandingkan bangunan non-permukiman. Kondisi tersebut ditunjukkan dari nilai *recall* dan *specificity* yang jauh berbeda, dimana *recall* dapat diartikan sebagai kemampuan model untuk mengidentifikasi bangunan permukiman dan *specificity* sebagai kemampuan model untuk mengidentifikasi bangunan non-permukiman.

**Tabel IV.4.** Hasil klasifikasi model XGBoost standar pada data validasi. Permukiman memiliki label 1 dan non-permukiman memiliki label 0.

Lipatan CV	Label	Recall	F1-Score	Akurasi	Balanced Accuracy	F1-Score Rata-rata
1	1	0,985	0,959	0,923	0,641	0,685
	0	0,297	0,411			
2	1	0,995	0,960	0,925	0,629	0,680
	0	0,263	0,400			
3	1	0,990	0,956	0,918	0,629	0,675
	0	0,267	0,393			
4	1	0,987	0,961	0,927	0,665	0,713
	0	0,343	0,466			
5	1	0,982	0,955	0,917	0,677	0,723
	0	0,373	0,491			

Untuk mengatasi permasalahan ketidakseimbangan model dalam memprediksi permukiman dan non-permukiman, digunakan metrik *balanced accuracy* dan *F1-Score* rata-rata untuk memberikan akurasi yang lebih representatif untuk bangunan permukiman

dan non-permukiman. Dari hasil model XGBoost standar, dapat dilihat bahwa rata-rata *balanced accuracy* dan *F1-Score* rata-rata hanya sebesar 0,648 dan 0,695. Nilai tersebut menunjukkan bahwa memang terdapat ketidakseimbangan model dalam memprediksi bangunan permukiman dan non-permukiman.

Berdasarkan hasil ini, langkah selanjutnya adalah mengoptimasi *hyperparameter* untuk meningkatkan *balanced accuracy* dan *F1-Score* rata-rata di setiap lipatan CV.

#### **IV.2.2. Optimasi Hyperparameter**

Dalam mengoptimasi hyperparameter, digunakan metode pencarian acak (*randomized grid-search*) dengan CV spasial. Metode ini dipilih karena mampu mempersingkat pencarian *hyperparameter* terbaik dibandingkan metode lainnya (Bergstra dan Bengio, 2012). Pada tahap ini, dicari kombinasi terbaik dari 10 *hyperparameter* yang berbeda untuk menghasilkan kombinasi *balanced accuracy* dan *F1-Score* terbaik. Pencarian acak dilakukan dengan menyusun 150 kombinasi acak dari 8 parameter, kemudian dikalikan dengan 5 lipatan CV, menghasilkan total 750 percobaan. Pada tahap ini, jumlah *boosted trees* (*n\_estimators*) dibatasi pada angka 100 dan ‘*learning\_rate*’ menggunakan nilai standar (0,3). Berikut ini adalah parameter yang dipilih beserta rentang nilai yang dicari dengan pencarian acak:

**Tabel IV.5.** Rentang parameter yang dicari.

Parameter	Keterangan	Rentang Nilai
gamma	Nilai <i>loss</i> minimum yang dibutuhkan untuk membuat partisi	0 - 0,05 - 0,1 - 0,15 - 2
max_depth	Kedalaman maksimum pohon	3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 – 11 – 12
subsample	Rasio sampel baris data dari data latih	0,5 - 0,6 - 0,7 - 0,8 - 0,85 - 0,9 - 0,95 - 1
colsample_bytree	Rasio sampel fitur dari data latih	0,5 - 0,6 - 0,7 - 0,8 - 0,9 - 1
min_child_weight	Jumlah minimum sampel untuk membuat <i>node</i> baru	1 – 2 – 3 – 4 – 5
reg_lambda	Parameter regularisasi lambda	1 – 1,5 – 2 – 2,5
reg_alpha	Parameter regularisasi alpha	0 – 0,1 – 0,5
scale_pos_weight	Bobot penyeimbang nilai positif dan negatif	0,1 – 0,2 – 0,3 – 0,4 – 0,5

Dari hasil pencarian hyperparameter dengan pencarian acak, didapatkan kombinasi terbaik yang tertera pada Tabel IV.6.

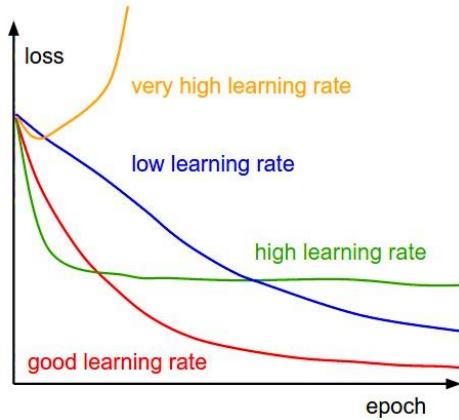
**Tabel IV.6.** Kombinasi *hyperparameter* yang dipilih.

<b><i>Hyperparameter</i></b>	<b>Nilai standar</b>
gamma	0,1
max_depth	5
min_child_weight	5
subsample	0,95
colsample_bytree	0,7
reg_lambda	1,5
reg_alpha	0,1
scale_pos_weight	0,4

Selanjutnya, adalah menentukan nilai dari parameter ‘n\_estimators’ dan ‘learning\_rate’ yang mampu meningkatkan metrik *balanced accuracy* dan *f1-score* rata-rata. Pencarian parameter ‘learning\_rate’ dan ‘n\_estimators’ dilakukan secara terpisah karena nilai ‘learning\_rate’ bergantung pada jumlah ‘n\_estimators’. Untuk mencari ‘learning\_rate’ terbaik, perlu digunakan nilai n\_estimators yang tinggi. Parameter ‘learning\_rate’ mengatur kontribusi dari setiap pohon dan mengontrol kecepatan algoritma dalam menemukan minimum gradien. Semakin kecil nilai ‘learning\_rate’, semakin lama waktu yang dibutuhkan untuk mencapai nilai gradien terkecil. Di sisi lain, semakin besar nilai ‘learning\_rate’, semakin besar kemungkinan terjadinya *overfitting* karena model belajar terlalu cepat dan terlalu banyak dari data latih. Oleh karena itu, perlu dicari nilai yang ideal untuk mencapai hasil validasi terbaik, seperti yang ditunjukkan pada Gambar IV.6.

Pada percobaan ini, nilai ‘n\_estimators’ diatur pada angka 500. Pencarian parameter ini dilakukan dengan menggunakan kombinasi *hyperparameter* terpilih dan penggunaan teknik *early stopping*. *Early stopping* merupakan teknik untuk menghentikan pelatihan model ketika nilai akurasi terbaik didapatkan, sebelum menurun karena *overfitting*. Rentang nilai ‘learning\_rate’ yang dicari adalah 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, dan 0.15. Setelah mendapatkan nilai ‘learning\_rate’ terbaik,

selanjutnya adalah mencari nilai ‘n\_estimators’ dari rentang 100, 150, 200, 250, 300, 350, 400, 450, dan 500. Sehingga, didapatkan nilai ‘learning\_rate’ sebesar 0,05 dan nilai ‘n\_estimators’ sebesar 350.



**Gambar IV.6.** Perbedaan *learning curve* untuk setiap nilai ‘learning\_rate’

(Sumber: CS231n, 2024).

Tahapan terakhir adalah menentukan nilai *threshold* dalam pengklasifikasian data. Berdasarkan Zou et al. (2019), nilai standar untuk *threshold* dalam kasus klasifikasi biner adalah 0,5, namun nilai ini tidak selalu optimal untuk data yang tidak seimbang. Oleh karena itu, perlu menentukan nilai *threshold* yang dapat memaksimalkan nilai *F1-Score* rata-rata, yang merupakan hasil dari perhitungan *recall* dan *precision* dari kelas positif dan negatif (Zou et al., 2019).

Pencarian nilai *threshold* dilakukan menggunakan model yang telah dioptimasi dengan CV spasial. Untuk setiap lipatan CV, dilakukan pembuatan kurva *precision-recall* menggunakan *library* Scikit-learn untuk mendapatkan nilai *precision* dan *recall* pada setiap nilai *threshold*. Berdasarkan grafik kurva *precision-recall*, dipilih nilai *threshold* terbaik yang menghasilkan kombinasi *precision* dan *recall* optimal untuk mencapai nilai *F1-Score* tertinggi. Nilai *threshold* terbaik dari setiap CV kemudian dirata-ratakan untuk mendapatkan nilai *threshold* final yang digunakan pada model akhir. Dari proses ini, diperoleh nilai *threshold* terbaik sebesar 0,518.

Dengan kombinasi *hyperparameter* yang didapatkan, dihasilkan nilai *balanced accuracy* dan F1-Score rata-rata sebesar 0,729 dan 0,742 untuk data validasi. Berdasarkan hasil validasi dari model XGBoost teroptimasi pada Tabel IV.7, telah terdapat beberapa peningkatan pada nilai *balanced accuracy* dan F1-Score dari model XGBoost standar.

Nilai rata-rata dari metrik *balanced accuracy* meningkat sebesar 0,081 dengan nilai sebesar 0,729 dan metrik *F1-Score* rata-rata meningkat 0,047 dengan nilai sebesar 0,742. Akan tetapi, peningkatan kemampuan model untuk mengklasifikasi bangunan non-permukiman, berdampak pada menurunnya kemampuan model dalam mengklasifikasi bangunan permukiman yang berdampak pada menurunnya akurasi model dan *recall* permukiman.

**Tabel IV.7.** Hasil klasifikasi model XGBoost teroptimasi pada data validasi.

Lipatan CV	Label	Recall	F1-Score	Akurasi	Balanced Accuracy	F1-Score Rata-rata
1	1	0,956	0,951	0,911	0,708	0,717
	0	0,460	0,484			
2	1	0,964	0,957	0,922	0,740	0,756
	0	0,517	0,555			
3	1	0,959	0,952	0,913	0,730	0,743
	0	0,501	0,534			
4	1	0,963	0,956	0,920	0,729	0,746
	0	0,495	0,535			
5	1	0,955	0,945	0,908	0,736	0,747
	0	0,517	0,545			

Setelah mendapatkan model terbaik untuk data validasi, tahapan selanjutnya adalah menggunakan model yang telah teroptimasi pada data uji.

#### IV.2.3. Pengujian Model

Proses pengujian dilakukan dengan menguji setiap kecamatan di wilayah Kota Jakarta Utara dengan model yang teroptimasi dan *threshold* yang telah ditentukan. Untuk tahapan pengujian ini, data latih yang digunakan adalah seluruh data dari wilayah Kota Jakarta Selatan. Pada tahapan ini, dilakukan pengujian klasifikasi tanpa data lokal dan sedikit data lokal.

Hasil pengujian tanpa data lokal tertera pada Tabel IV.8. Dari hasil pengujian, dapat terlihat bahwa model XGBoost dapat memprediksi bangunan permukiman dan non-permukiman dengan baik. Dari hasil pengujian ini, didapatkan rata-rata *balanced accuracy* sebesar 0,756 dan F1-Score rata-rata sebesar 0,786.

**Tabel IV.8.** Hasil klasifikasi model XGBoost pada data uji tanpa data lokal.

Kecamatan	Label	Recall	F1-Score	Akurasi	Balanced Accuracy	F1-Score Rata-rata
Cilincing	1	0,976	0,949	0,909	0,721	0,761
	0	0,465	0,574			
Kelapa Gading	1	0,964	0,956	0,921	0,783	0,799
	0	0,602	0,643			
Koja	1	0,976	0,957	0,922	0,711	0,748
	0	0,446	0,539			
Pademangan	1	0,970	0,931	0,888	0,779	0,812
	0	0,588	0,692			
Penjaringan	1	0,955	0,934	0,891	0,789	0,811
	0	0,624	0,688			
Tanjung Priok	1	0,983	0,961	0,929	0,741	0,786
	0	0,500	0,611			

Dalam pengujian model XGBoost teroptimasi dengan data lokal, dilakukan tiga kali percobaan dengan persentase data lokal yang berbeda, yaitu 2,5%, 5% dan 10%. Data lokal diambil secara acak dengan menggunakan fungsi *train\_test\_split* dari *library* Scikit-learn. Untuk memitigasi dampak dari *sampling effect*, pengujian ini dilakukan sebanyak lima kali dan hasil yang digunakan merupakan hasil rata-rata dari setiap pengujian. Informasi terkait jumlah bangunan di setiap kecamatan dapat dilihat pada Tabel IV.9.

**Tabel IV.9.** Total bangunan dan jumlah bangunan yang diikutsertakan ke dalam data latih di setiap percobaan.

Kecamatan	Total Bangunan	2,5 % Bangunan	5% Bangunan	10 % Bangunan
Cilincing	63.103	1.577	3.155	6.310
Kelapa Gading	28.306	707	1.415	2.830
Koja	47.170	1.179	2.358	4.717
Pademangan	21.973	549	1.098	2.197
Penjaringan	39.914	997	1.995	3.991
Tanjung Priok	57.300	1.432	2.865	5.730

Penambahan data lokal, menghasilkan nilai *balanced accuracy* dan *F1-Score* rata-rata sebesar 0,796 dan 0,815. Hasil dari rata-rata percobaan dengan data lokal untuk setiap persentase tertera pada Tabel IV.10.

**Tabel IV.10.** Hasil klasifikasi model XGBoost pada data uji dengan data lokal.

Kecamatan	Data Lokal	Recall (positif)	Recall (negatif)	F1-Score (positif)	F1-Score (negatif)	Akurasi	Balanced Accuracy	F1-Score Rata-rata
Cilincing	2,5 %	0,971	0,516	0,951	0,605	0,912	0,744	0,778
	5 %	0,970	0,535	0,951	0,617	0,914	0,753	0,784
	10 %	0,968	0,571	0,953	0,639	0,916	0,769	0,796
Kelapa Gading	2,5 %	0,963	0,621	0,957	0,654	0,923	0,792	0,805
	5 %	0,965	0,638	0,958	0,670	0,926	0,801	0,814
	10 %	0,963	0,657	0,959	0,680	0,927	0,810	0,820
Koja	2,5 %	0,974	0,486	0,958	0,566	0,924	0,730	0,762
	5 %	0,972	0,509	0,959	0,580	0,925	0,740	0,769
	10 %	0,971	0,537	0,959	0,599	0,926	0,754	0,779
Pademangan	2,5 %	0,960	0,625	0,931	0,706	0,888	0,793	0,819
	5 %	0,955	0,653	0,932	0,718	0,890	0,804	0,825
	10 %	0,949	0,692	0,934	0,737	0,894	0,821	0,835
Penjaringan	2,5 %	0,951	0,661	0,936	0,709	0,895	0,806	0,823
	5 %	0,952	0,691	0,940	0,730	0,901	0,821	0,835
	10 %	0,950	0,743	0,945	0,761	0,910	0,847	0,853
Tanjung Priok	2,5 %	0,982	0,524	0,962	0,628	0,931	0,753	0,795
	5 %	0,981	0,539	0,963	0,639	0,932	0,760	0,801
	10 %	0,979	0,570	0,963	0,655	0,933	0,774	0,809

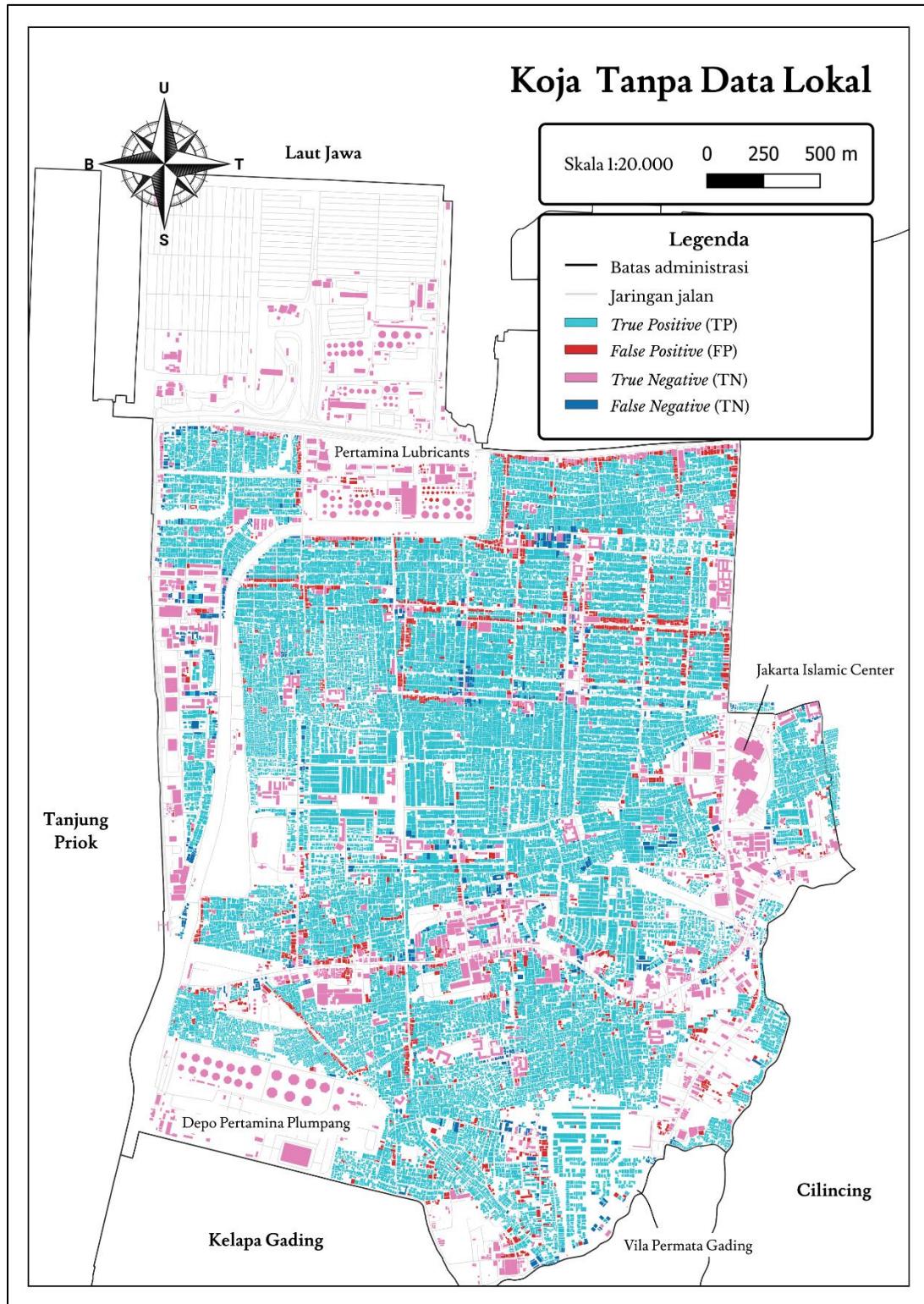
Berdasarkan hasil percobaan dengan tiga persentase data lokal, ditemukan bahwa menambahkan data lokal ke dalam data latih tidak meningkatkan akurasi secara signifikan, tetapi meningkatkan nilai *balanced score* dan *F1-Score* rata-rata. Hal ini menunjukkan bahwa menambahkan data lokal membantu model untuk memahami karakteristik dari bangunan non-permukiman yang umumnya berbeda-beda dari setiap wilayah. Akan tetapi, peningkatan kemampuan model untuk memprediksi bangunan non-permukiman

mengakibatkan menurunnya kemampuan model untuk memprediksi kemampuan permukiman.

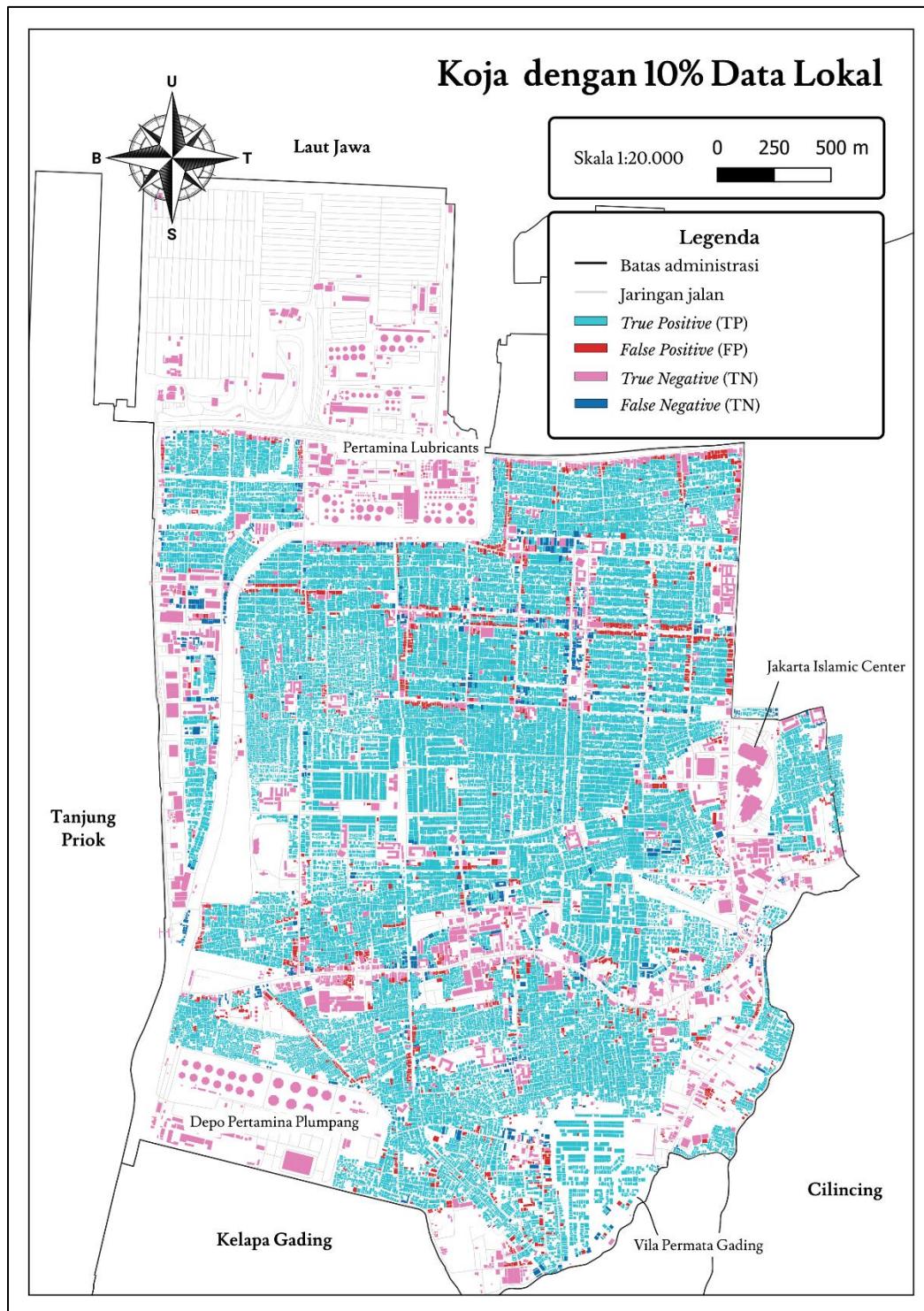
#### **IV.2.4. Hasil dan Evaluasi Model**

Berdasarkan hasil dari penelitian yang telah dilakukan, model XGBoost terbukti dapat digunakan dalam melakukan klasifikasi bangunan permukiman dan non-permukiman menggunakan karakter morfologi bangunan pada kondisi bangunan di wilayah Jakarta yang padat. Hal tersebut ditunjukkan dengan akurasi model pada wilayah Jakarta Utara tanpa data lokal yang dapat mencapai angka rata-rata 91% di setiap kecamatan. Akan tetapi, model mengalami kesulitan untuk mengklasifikasi bangunan non-permukiman yang berada di tengah-tengah permukiman seperti, masjid, ruko, bangunan serbaguna, dan bangunan lainnya yang memiliki geometri serupa dan dekat dengan permukiman. Hal ini berdampak kepada nilai *recall* dari bangunan non-permukiman yang cukup rendah, sekitar 53,7% di setiap kecamatan. Contoh hasil prediksi model XGBoost teroptimasi tanpa data lokal dapat dilihat pada Gambar IV.6 untuk wilayah Koja, Jakarta Utara. Pada visualisasi tersebut, bangunan permukiman yang diklasifikasi dengan benar ditunjukkan dengan label ‘Positif benar’ dan bangunan non-permukiman dengan label ‘Negatif Benar’. Untuk bangunan non-permukiman yang diprediksi sebagai bangunan permukiman ditunjukkan dengan label ‘Positif Salah’ dan sebaliknya, ditunjukkan dengan label ‘Negatif Salah’.

Pada penelitian ini, dilakukan beberapa eksperimen untuk memasukkan sebagian kecil data lokal ke dalam data latih dengan harapan dapat meningkatkan akurasi dan kemampuan model dalam mengklasifikasi bangunan di wilayah tersebut. Berdasarkan hasil dari beberapa eksperimen, ditemukan bahwa menambahkan sebagian kecil data lokal (2,5%, 5%, dan 10%) tidak meningkatkan akurasi model secara signifikan. Akan tetapi, penambahan data lokal membantu model untuk membedakan kondisi permukiman dan non-permukiman dengan lebih baik yang ditunjukkan dengan meningkatnya nilai *balanced accuracy* dan *F1-Score* rata-rata. Hal ini menunjukkan bahwa ketersediaan data penggunaan bangunan dapat meningkatkan kemampuan model untuk mengidentifikasi bangunan non-permukiman di suatu wilayah yang belum pernah dilihat sebelumnya.



**Gambar IV.7.** Hasil prediksi model XGBoost teroptimasi tanpa data lokal di Kecamatan Koja, Jakarta Utara.



**Gambar IV.8.** Hasil prediksi model XGBoost teroptimasi dengan 10% data lokal di Kecamatan Koja, Jakarta Utara.

Berdasarkan visualisasi hasil prediksi tanpa dan dengan data lokal pada Gambar IV.7 dan IV.8, terdapat peningkatan kemampuan model untuk memprediksi kelas negatif atau bangunan non-permukiman seperti yang terlihat pada *confusion matrix* sebelum penambahan data lokal di Tabel IV.11. dan setelah penambahan data lokal di Tabel IV.12..

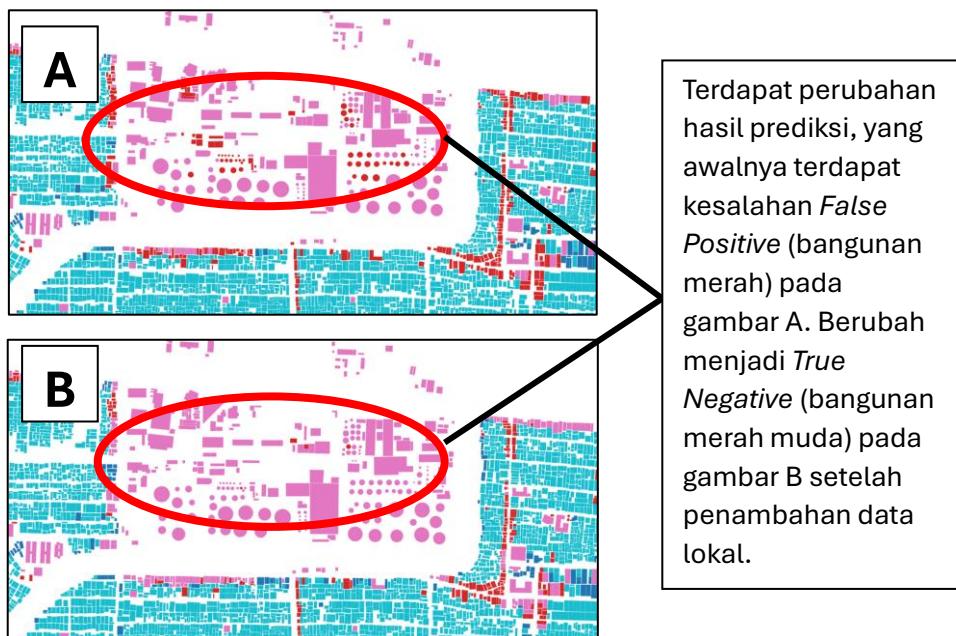
**Tabel IV.11.** Confusion Matrix Koja tanpa data lokal.

		Hasil Prediksi	
		Positif	Negatif
Data Aktual	P	41.319	1.022
	N	2.673	2.156

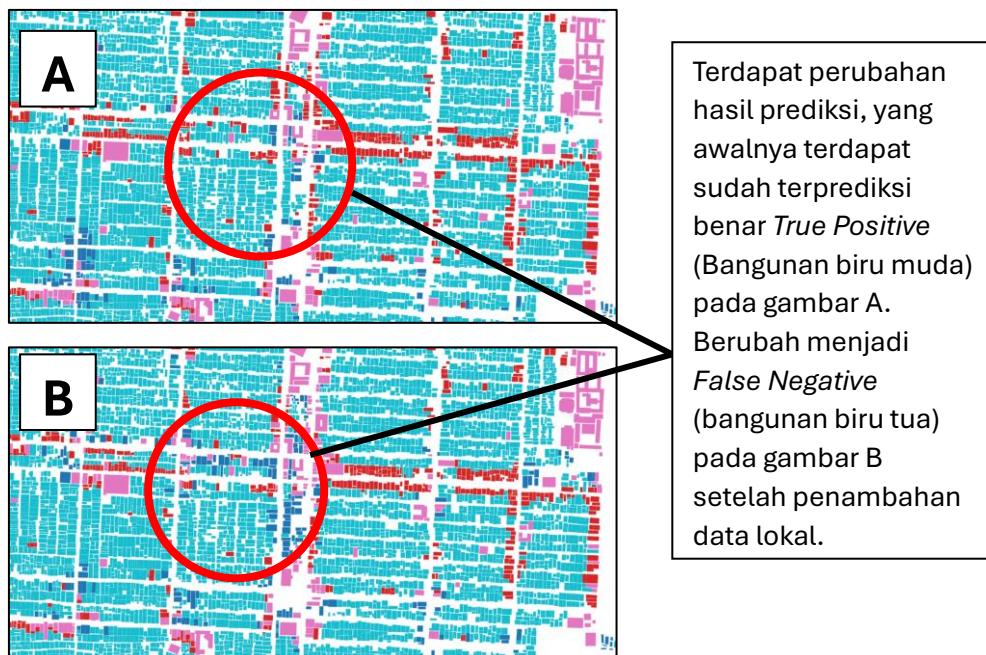
**Tabel IV.12.** Confusion Matrix Koja dengan 10% data lokal.

		Hasil Prediksi	
		Positif	Negatif
Data Aktual	P	41.199	1.142
	N	2.514	2.514

Berdasarkan *confusion matrix* pada Tabel IV.11. dan Tabel IV.12. tersebut, ditemukan bahwa terdapat peningkatan pada kelas Negatif Benar dan penurunan pada kelas Positif Benar. Penambahan data lokal meningkatkan sensitivitas model terhadap data bangunan non-permukiman, tetapi berdampak pada menurunnya sensitivitas model terhadap data bangunan permukiman seperti yang terlihat pada Gambar IV.10.

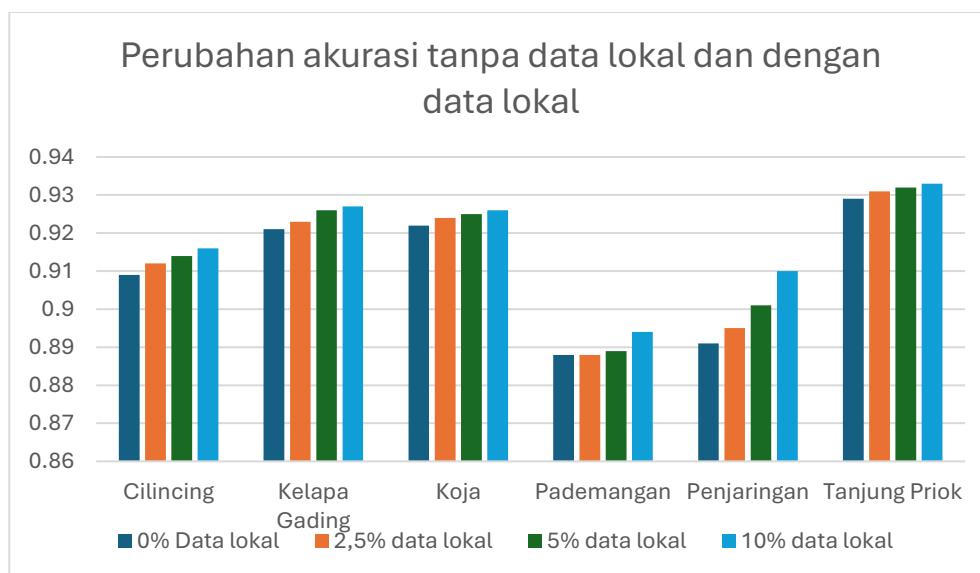


**Gambar IV.9.** Dampak penambahan data lokal pada bangunan non-permukiman. (A) Hasil tanpa data lokal. (B) Hasil dengan 10% data lokal.

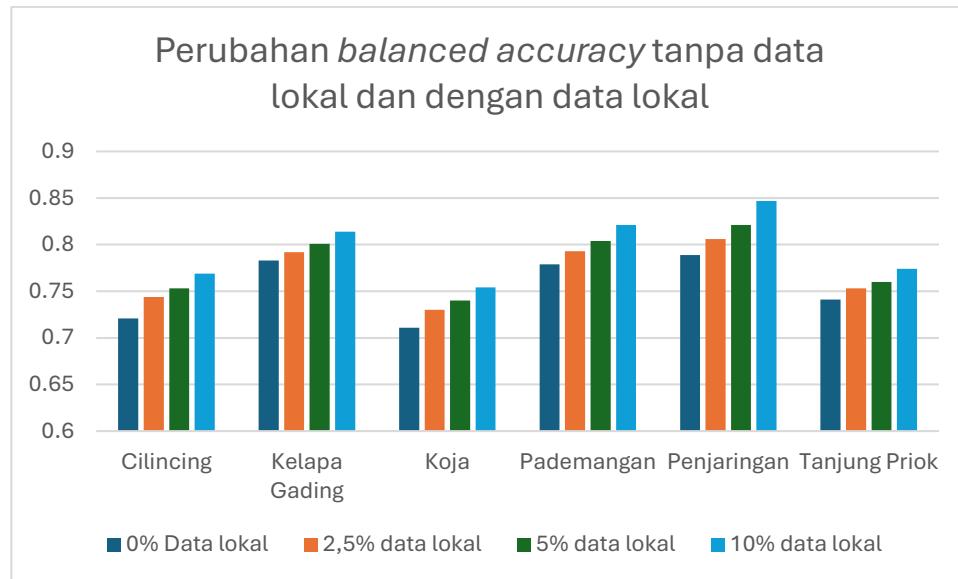


**Gambar IV.10.** Dampak penambahan data lokal pada bangunan permukiman. (A) Hasil tanpa data lokal. (B) Hasil dengan 10% data lokal.

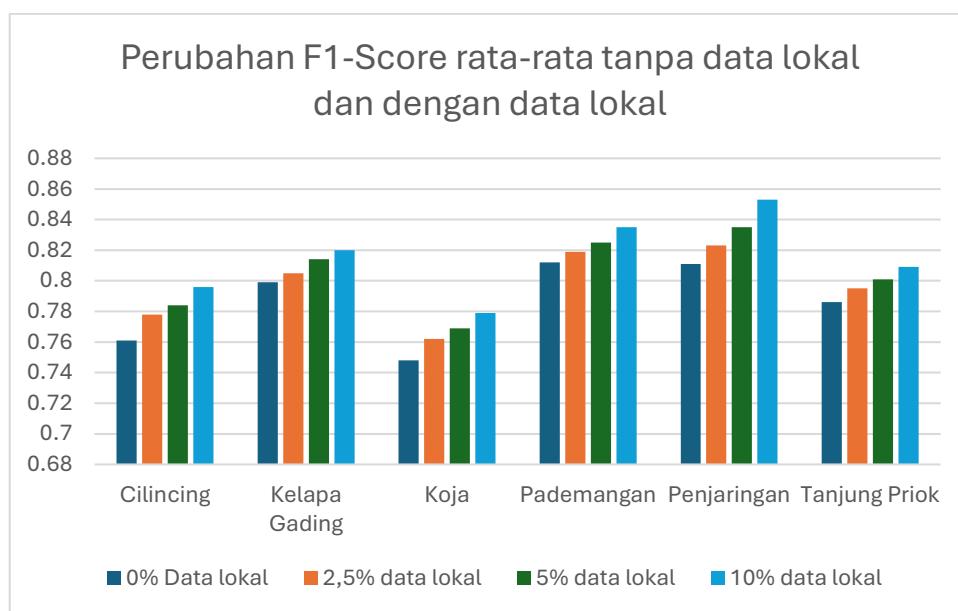
Walaupun terdapat pertukaran antara sensitivitas bangunan permukiman dan bangunan non-permukiman, penambahan data lokal terbukti meningkatkan nilai dari metrik akurasi, *balanced accuracy*, dan *F1-Score* rata-rata. Grafik peningkatan dari metrik-metrik tersebut dapat dilihat pada Gambar IV.11 sampai dengan Gambar IV.13.



**Gambar IV.11.** Perubahan metrik akurasi tanpa dan dengan data lokal.



**Gambar IV.12.** Perubahan metrik *balanced accuracy* tanpa dan dengan data lokal.



**Gambar IV.13.** Perubahan metrik *F1-Score* rata-rata tanpa dan dengan data lokal.

Adapun kendala dalam pengaplikasian metodologi ini berada pada kondisi morfologi bangunan yang berbeda-beda di setiap wilayah di Indonesia. Pada penelitian ini, digunakan data bangunan dari Jakarta Selatan sebagai data latih, sehingga memungkinkan model digunakan untuk klasifikasi bangunan di wilayah Jakarta Utara mengingat kondisi morfologi bangunan yang tidak berbeda jauh. Untuk dapat diaplikasikan di berbagai wilayah di Indonesia, dibutuhkan data bangunan yang sesuai dan dapat merepresentasikan kondisi morfologi dari wilayah yang ingin diklasifikasi.

### IV.3. Feature Importances

Berdasarkan nilai *feature importance* dari model XGBoost teroptimasi tanpa data lokal, ditemukan bahwa setiap kelompok karakter memiliki kontribusi untuk meningkatkan kemampuan klasifikasi model. Tingkat kepentingan dari setiap kelompok dan skala karakter dapat dilihat pada Tabel IV.11.

**Tabel IV.11.** Skor *feature importance* dari model XGBoost teroptimasi. Metrik *importance* yang digunakan adalah nilai *gain* relatif terhadap jumlah *gain* dari seluruh karakter.

Karakter	Jumlah	Total Tingkat Kepentingan	Rata-rata Tingkat Kepentingan	Karakter Individual Terbaik	Gain Terbaik
<b>Per-kelompok</b>					
Bangunan	85	36,23%	4,16‰	Jumlah bangunan tetangga dalam <i>buffer</i> 50m	4,62%
Sel Teselasi	73	36,87%	5,05‰	Area terlingkupi oleh sel teselasi ( <i>covered area</i> )	8,57%
Jalan	60	26,9%	4,2‰	Lebar jalan rata-rata dalam <i>buffer</i> 50 m	1,96%
<b>Per-skala</b>					
Objek dan sekitarnya	50	41,08%	8,21‰	Area terlingkupi oleh sel teselasi ( <i>covered area</i> )	8,57%
50 m	58	29,33%	5,06‰	Jumlah bangunan tetangga dalam <i>buffer</i> 50 m	4,62%
150 m	58	15,82%	2,73‰	Standar deviasi sumbu terpanjang bangunan	2,47%
300 m	58	13,77%	2,37‰	Median keliling sel teselasi	0,65%

Kelompok karakter dengan rata-rata *gain* terbesar adalah kelompok Sel Teselasi dengan rata-rata *gain* sebesar 0,5%. Di antara seluruh karakter morfometrik, luas area yang terlingkupi oleh sel teselasi merupakan karakter yang paling signifikan untuk memisahkan antara bangunan permukiman dan non-permukiman dengan nilai *gain* sebesar 8,57%.

Pada penelitian ini, skala dari setiap karakter morfometrik juga dibandingkan. Berdasarkan nilai *gain*, skala yang memiliki pengaruh yang signifikan adalah skala objek dan sekitarnya. Nilai rata-rata *gain* dari setiap karakter semakin kecil dengan meningkatnya ukuran skala. Hal ini menunjukkan bahwa informasi mengenai geometri bangunan dan lingkungan terdekat menjadi penentu utama apakah suatu bangunan merupakan permukiman atau non-permukiman.

Berdasarkan nilai *gain* dari setiap karakter, ditemukan bahwa karakter yang merepresentasikan bangunan dan lingkungan di sekitarnya memiliki kemampuan prediktif yang tinggi. Hal ini mengindikasikan bahwa karakter-karakter tersebut, mungkin menyimpan informasi mengenai autokorelasi spasial dari penggunaan bangunan, menjadikannya karakter yang memiliki pengaruh signifikan terhadap kemampuan prediktif model XGBoost.

#### **IV.4. Pemanfaatan Model Lebih Lanjut**

Objektif dari penelitian ini adalah untuk membuat sebuah model klasifikasi permukiman yang dapat digunakan untuk memperkaya data persebaran permukiman di berbagai wilayah di Indonesia. Berdasarkan penelitian ini, ditemukan bahwa karakter morfologi bangunan dan atribut OSM dapat digunakan untuk membuat model klasifikasi bangunan permukiman dan non-permukiman dengan algoritma XGBoost. Model ini dapat dimanfaatkan untuk mengetahui persebaran permukiman dan non-permukiman di wilayah yang tidak memiliki data penggunaan bangunan yang lengkap. Untuk dapat meningkatkan kemampuan model, penelitian ini juga mencoba untuk menambahkan sedikit data lokal ke dalam data latih. Percobaan ini membuktikan bahwa dengan menambahkan sedikit data lokal mampu untuk meningkatkan kemampuan model dalam memprediksi penggunaan bangunan di wilayah yang belum pernah dilihat model sebelumnya.

Model ini dilatih dan diuji dengan data bangunan di wilayah Kota Jakarta Selatan dan Kota Jakarta Utara yang mungkin memiliki kondisi morfologi urban yang berbeda dengan wilayah lain di Indonesia. Untuk pemanfaatan model di wilayah lainnya, dibutuhkan sampel data bangunan dari wilayah yang ingin diprediksi guna meningkatkan kemampuan model dalam memprediksi bangunan permukiman di wilayah tersebut.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **V.1. Kesimpulan**

Dari hasil penelitian yang dilakukan, berikut ini merupakan jawaban dari setiap pertanyaan penelitian:

- a. Model XGBoost yang memanfaatkan data morfologi urban dan atribut OSM terbukti mampu mengklasifikasikan bangunan permukiman dan non-permukiman dengan akurasi yang tinggi dan hasil yang cukup representatif di wilayah yang belum pernah dilihat model sebelumnya.
- b. Tanpa data lokal, model ini mencapai akurasi sebesar 91% dan *F1-Score* rata-rata sebesar 78,6%. Dengan penambahan data lokal sebesar 10%, dapat meningkatkan akurasi dan *F1-Score* rata-rata menjadi 91,8% dan 81,5%. Penambahan data lokal sebesar 2,5%, 5%, dan 10% tidak meningkatkan akurasi secara signifikan, tetapi membantu model dalam membedakan kondisi permukiman dan non-permukiman dengan lebih baik. Hal ini ditunjukkan dengan meningkatnya nilai *balanced accuracy* dan *F1-Score* rata-rata sebesar .
- c. Dalam mengklasifikasi bangunan permukiman dan non-permukiman, karakter morfometrik yang memiliki dampak signifikan terhadap kemampuan model adalah karakter yang merepresentasikan bangunan dan lingkungan di sekitarnya. Dalam penelitian ini, karakter individual dengan *gain* tertinggi merupakan luas area yang tertutupi oleh sel teselasi dan sel teselasi di sekitarnya (*tessellation covered area*).

Selain dari hasil penelitian di atas, dapat disimpulkan beberapa temuan lainnya seperti:

- a. Data OSM memiliki berbagai kelemahan yang berkaitan dengan kualitas data. Akan tetapi, mengingat ketersediaan data bangunan yang cukup masif di berbagai wilayah di Indonesia, penggunaan data OSM dapat membantu proses pengadaan data persebaran permukiman di berbagai wilayah yang tidak memiliki akses terhadap data bangunan resmi.
- b. Berdasarkan penelitian ini, model dapat dimanfaatkan untuk wilayah lainnya dengan karakteristik bangunan yang serupa dengan bangunan di wilayah DKI

Jakarta. Untuk wilayah dengan karakteristik yang berbeda, seperti wilayah pedesaan atau kota lainnya, diperlukan data penggunaan bangunan lokal agar model dapat mempelajari karakteristik wilayah tersebut dengan lebih baik.

## V.2. Saran

Terdapat beberapa saran yang dapat dijadikan acuan untuk penelitian lebih lanjut terkait pemanfaatan data morfologi bangunan dan algoritma XGBoost untuk mengklasifikasi bangunan permukiman dan non-permukiman.

- a. Menambah jumlah dan variasi dari data latih dari berbagai wilayah dengan karakteristik yang berbeda dapat membantu model dalam menggeneralisasi klasifikasi dengan lebih baik pada skala nasional.
- b. Penggunaan data bangunan dengan kualitas yang lebih baik dapat meningkatkan keakuratan model.
- c. Menerapkan dan membandingkan hasil model XGBoost dengan model *machine learning* lainnya dapat dilakukan untuk menemukan metode terbaik untuk melakukan klasifikasi bangunan permukiman dan non-permukiman.
- d. Mengintegrasikan karakteristik lain seperti data demografi, data jenis atap, atau data ekonomi yang mungkin memiliki pengaruh terhadap penggunaan bangunan.

## DAFTAR PUSTAKA

- Acheampong, R.A. (2018). Spatial Planning In Ghana: Origins, Contemporary Reforms And Practices, And New Perspectives. Springer: Ghana.
- Adams, B., Gahegan, M., Gupta, P., & Hosking, R. (2014). Geographic Information Observatories For Supporting Science. Ceur Workshop Proceedings.
- Amri, L. H. A., & Wijayanti, R. A. (2019). Pemanfaatan Sistem Informasi Geospasial Online Untuk Mendukung Pengambilan Keputusan Pemanfaatan Tata Ruang. *Jurnal Teknologi Informasi Dan Terapan*, 6(2), 62-66.
- Angra, S., & Ahuja, S. (2017, March). Machine Learning And Its Applications: A Review. In 2017 International Conference On Big Data Analytics And Computational Intelligence (Icbdac) (Pp. 57-60). Ieee.
- Aronoff, S. (1989). Geographic Information Systems: A Management Perspective.
- Arsanjani, Jj., A. Zipf, P. Mooney, Dan M. Helbich. (2015). Openstreetmap In Giscience: Experiences, Research, And Applications. London: Springer. Doi:10.1007/978-3-319-14280-7
- Atwal, K. S., Anderson, T., Pfoser, D., & Züfle, A. (2022). Predicting Building Types Using Openstreetmap. *Scientific Reports*, 12(1), 19976.
- Badan Pusat Statistik. (2021). Hasil Sensus Penduduk 2020.
- Basuki, B., Apriyeni, B. A. R., Purnamasari, I., Rachman, H. A., Rahman, F. A., & Mubarokah, N. (2023). Pengantar Informasi Geospasial. Penerbit Tahta Media.
- Batty, M. (2008). Cities As Complex Systems: Scaling, Interactions, Networks, Dynamics, And Urban Morphologies. In The Encyclopedia Of Complexity & System Science (Vol. 1, Issue 131, Pp. 1–62). Springer Berlin.
- Bennett, J. (2010). Openstreetmap (1st Ed.). Packt Publishing Ltd. Diakses Dari <Https://Www.Perlego.Com/Book/389235/Openstreet Map-Pdf> (Original Work Published 2010). [7 Oktober 2020].
- Berghauer Pont, M. (2018). An Analytical Approach To Urban Form. In Urban Book Series. Springer International Publishing. [Https://Doi.Org/10.1007/978-3-319-76126-8\\_7](Https://Doi.Org/10.1007/978-3-319-76126-8_7).
- Bergstra, J., & Bengio, Y. (2012). Random Search For Hyper-Parameter Optimization. *Journal Of Machine Learning Research*, 13(2).
- Berrar, D. (2019). Cross-Validation.
- Boeing G. (2020). A Multi-Scale Analysis Of 27,000 Urban Street Networks: Every Us City, Town, Urbanized Area, And Zillow Neighborhood. *Environment And Planning B: Urban Analytics And City Science*. 2020; 47 (4):590–608.
- Boeing, G. (2021). Spatial Information And The Legibility Of Urban Form: Big Data In Urban Morphology. *International Journal Of Information Management*, 56(September 2019), 102013. <Https://Doi.Org/10.1016/J.Ijinfomgt.2019.09.009>
- Brauchler, M., & Stoffels, J. (2020). Leveraging Osm And Geobia To Create And Update Forest Type Maps. *Isprs International Journal Of Geo-Information*, 9(9), 499.
- Carneiro, C., Morello, E., Voegtle, T., & Golay, F. (2010). Digital Urban Morphometrics: Automatic Extraction And Assessment Of Morphological Properties Of Buildings. *Transactions In Gis*, 14(4), 497-531.

- Chen, T., & Guestrin, C. (2016, August). Xgboost: A Scalable Tree Boosting System. In Proceedings Of The 22nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining (Pp. 785-794).
- Chintya, N. P. P., Mutiarasari, W. M., Prasetyo, D. Y., & Mayangsari, W. R. (2022). Evaluasi Kualitas Tematik Peta Penggunaan Lahan Yang Diperoleh Dari Data Volunteered Geographic Information Studi Kasus Di Kota Yogyakarta:(Thematic Quality Assessment Of Land-Use Map Derived From Volunteered Geographic Information Data, Case Study In Yogyakarta City). Geomatika, 28(1), 51-60.
- Cortes, P. (2005). Morphologies Of Fragmentation And Continuity. In M. Barke (Ed.), Approaches In Urban Morphology (Pp. 44–58). University Of Northumbria.
- Cs231n: Convolutional Neural Networks For Visual Recognition. (2024). Neural Networks Part 3: Learning And Evaluation. Retrieved June 18, 2024, From <Https://Cs231n.Github.Io/Neural-Networks-3/>.
- Das, P., & Chand, S. (2021, February). Attentionbuildnet For Building Extraction From Aerial Imagery. In 2021 International Conference On Computing, Communication, And Intelligent Systems (Icccis) (Pp. 576-580). Ieee.
- Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications Of Artificial Intelligence In Machine Learning: Review And Prospect. International Journal Of Computer Applications, 115(9).
- De Castro, K. B., Roig, H. L., Neumann, M. R. B., Rossi, M. S., Seraphim, A. P. A. C. C., Réquia, W. J., Costa, A. B. B. Da, & Höfer, R. (2019). New Perspectives In Land Use Mapping Based On Urban Morphology: A Case Study Of The Federal District, Brazil. Land Use Policy, 87(June 2018), 104032. <Https://Doi.Org/10.1016/J.Landusepol.2019.104032>
- Dibble, J., Prelorendjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., & Porta, S. (2015). Urban Morphometrics: Towards A Science Of Urban Evolution. Arxiv.Org. Retrieved From <Http://Arxiv.Org/Abs/1506.04875v2>
- Dianyah, I., & Bioresita, F. (2023). Analisa Kawasan Permukiman Kumuh Di Kecamatan Kenjeran Surabaya Menggunakan Metode Ndbi Dan Obia Serta Data Citra Sentinel-2 Tahun 2022. Geoid, 19(1), 180-191.
- Faiz, A. (2022). Data Openstreetmap Untuk Memetakan Aksesibilitas Spasial Dan Jangkauan Pelayanan Dasar Pendidikan Dan Kesehatan Di Kota Semarang. Jurnal Litbang Provinsi Jawa Tengah, 20(1), 79-96.
- Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality Assessment For Building Footprints Data On Openstreetmap. International Journal Of Geographical Information Science, 28(4), 700-719.
- Fathi, S., Sajadzadeh, H., Sheshkal, F. M., Aram, F., Pinter, G., Felde, I., & Mosavi, A. (2020). The Role Of Urban Morphology Design On Enhancing Physical Activity And Public Health. International Journal Of Environmental Research And Public Health, 17(7), 1–29. <Https://Doi.Org/10.3390/Ijerph17072359>
- Fawcett, T. (2006). An Introduction To Roc Analysis. Pattern Recognition Letters, 27(8), 861-874.
- Febrina, W. A., Yanuarsyah, I., & Hudjimartsu, S. (2019, October). Kombinasi Obia (Object-Based Image Analysis) Untuk Identifikasi Wilayah Permukiman. In Seminar Nasional Teknologi Informasi (Vol. 2, Pp. 354-358).

- Feldmeyer, D., Meisch, C., Sauter, H., & Birkmann, J. (2020). Using Openstreetmap Data And Machine Learning To Generate Socio-Economic Indicators. *Isprs International Journal Of Geo-Information*, 9(9), 498.
- Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, 3-33.
- Fleischmann, M. (2019). Momepy: Urban Morphology Measuring Toolkit. *Journal Of Open Source Software*, 4(43), 1807.
- Fleischmann, M., & Arribas-Bel, D. (2022). Classifying Urban Form At National Scale: The British Morphosignatures.
- Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2020). Morphological Tessellation As A Way Of Partitioning Space: Improving Consistency In Urban Morphology At The Plot Scale. *Computers, Environment And Urban Systems*, 80, 101441.
- Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2022). Methodological Foundation Of A Numerical Taxonomy Of Urban Form. *Environment And Planning B: Urban Analytics And City Science*, 49(4), 1283-1299.
- Fleischmann, M., Romice, O., & Porta, S. (2021). Measuring Urban Form: Overcoming Terminological Inconsistencies For A Quantitative And Comprehensive Morphologic Analysis Of Cities. *Environment And Planning B: Urban Analytics And City Science*, 48(8), 2133-2150.
- Forget, Y., Linard, C. & Gilbert, M. Supervised Classification Of Built-Up Areas In Sub-Saharan African Cities Using Landsat Imagery And Openstreetmap. *Remote Sens.* 10, 1145 (2018).
- Friedman, J. H. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.
- Hamaina, R., Leduc, T., & Moreau, G. (2012). Towards Urban Fabrics Characterization Based On Buildings Footprints. In *Bridging The Geographic Information Sciences: International Agile'2012 Conference*, Avignon (France), April, 24-27, 2012 (Pp. 327-346). Springer Berlin Heidelberg.
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts And Techniques*. Morgan Kaufmann.
- Harahap, F. R. (2013). Dampak Urbanisasi Bagi Perkembangan Kota Di Indonesia. *Society*, 1(1), 35-45.
- Harmon, J. E., & Anderson, S. J. (2003). *The Design And Implementation Of Geographic Information Systems*. John Wiley & Sons.
- Hartmann, A., Behnisch, M., Hecht, R., & Meinel, G. (2024). Prediction Of Residential And Non-Residential Building Usage In Germany Based On A Novel Nationwide Reference Data Set. *Environment And Planning B: Urban Analytics And City Science*, 51(1), 216-233.
- Hawkins, D. M. (2004). The Problem Of Overfitting. *Journal Of Chemical Information And Computer Sciences*, 44(1), 1-12.
- Hillel, T., Elshafie, M. Z., & Jin, Y. (2018). Recreating Passenger Mode Choice-Sets For Transport Simulation: A Case Study Of London, Uk. *Proceedings Of The Institution Of Civil Engineers-Smart Infrastructure And Construction*, 171(1), 29-42.

- Indonesia. Peraturan Pemerintah Nomor 21 Tahun 2021 Tentang Penyelenggaraan Penataan Ruang. Lembaran Negara Ri Tahun 2021 Nomor 31, Tambahan Lembaran Ri Nomor 6633. Sekretariat Negara. Jakarta
- Indonesia. Undang-Undang Nomor 4 Tahun 2011 Tentang Informasi Geospasial. Lembaran Negara Ri Tahun 2011 Nomor 49 Tambahan Lembaran Ri Nomor 5124. Sekretariat Negara. Jakarta.
- Indonesia. Undang-Undang Republik Indonesia No. 1 Tahun 2011 Tentang Perumahan Dan Kawasan Permukiman. Lembaran Negara Ri Tahun 2011 Nomor 7, Tambahan Lembaran Ri Nomor 5188. Sekretariat Negara. Jakarta.
- Jenks, M., Jones, C., Irvine, K. N., Fuller, R. A., Devine-Wright, P., Tratalos, J., Payne, S. R., Warren, P. H., Lomas, K. J., & Gaston, K. J. (2008). Sustainable City Form. In Media (Vol. 2). <Https://Doi.Org/10.1007/978-1-4020-8647-2>
- Jochem, W. C., & Tatem, A. J. (2021). Tools For Mapping Multi-Scale Settlement Patterns Of Building Footprints: An Introduction To The R Package Foot. Plos One, 16(2), E0247535.
- Kropf K (2018) Plots, Property And Behaviour. *Urban Morphology* 22(1): 5–14.
- Kropf, K. (2017, July). Bridging Configurational And Urban Tissue Analysis. In Proceedings Of 11th Space Syntax Symposium, Lisbon (Pp. 165-1).
- Kühnl, M., Sapena, M., Wurm, M., Geiß, C., & Taubenböck, H. (2023). Multitemporal Landslide Exposure And Vulnerability Assessment In Medellín, Colombia. *Natural Hazards*, 119(2), 883-906.
- Lai, P. C., Chen, S., Low, C. T., Cerin, E., Stimson, R., & Wong, P. Y. P. (2018). Neighborhood Variation Of Sustainable Urban Morphological Characteristics. *International Journal Of Environmental Research And Public Health*, 15(3). <Https://Doi.Org/10.3390/Ijerph15030465>
- Levy A (1999) Urban Morphology And The Problem Of The Modern Urban Fabric: Some Questions For Research. *Urban Morphology* 3: 79–85.
- Lloyd, C. T., Sturrock, H. J., Leisure, D. R., Jochem, W. C., Lázár, A. N., & Tatem, A. J. (2020). Using Gis And Machine Learning To Classify Residential Status Of Urban Buildings In Low And Middle Income Settings. *Remote Sensing*, 12(23), 3847.
- Lobo J, Alberti M, Allen-Dumas M, Arcaute E, Barthelemy M, Bojorquez Tapia La, Et Al. *Urban Science: Integrated Theory From The First Cities To Sustainable Metropolises*. 2020.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science And Systems*. John Wiley & Sons.
- Louf R, Barthelemy M. (2014). A Typology Of Street Patterns. *Journal Of The Royal Society Interface*; 11 (101):20140924. <Https://Doi.Org/10.1098/Rsif.2014.0924> Pmid: 25297318.
- Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance Of Spatial Predictor Variable Selection In Machine Learning Applications—Moving From Data Reproduction To Spatial Prediction. *Ecological Modelling*, 411, 108815.
- Miller, H. J. (2004). Tobler's First Law And Spatial Analysis. *Annals Of The Association Of American Geographers*, 94(2), 284-289.
- Milojevic-Dupont N, Hans N, Kaack Lh, Zumwald M, Andrieux F, De Barros Soares D, Et Al. (2020) Learning From Urban Form To Predict Building Heights. *Plos One* 15(12): E0242010. <Https://Doi.Org/10.1371/Journal.Pone.0242010>

- Moudon, A. V. (1997). Urban Morphology As An Emerging Interdisciplinary Field Urban Morphology, 1(1), 3–10. <Https://Doi.Org/10.1111/J.1469-8137.1956.Tb05265>.
- Nachtigall, F., Milojevic-Dupont, N., Wagner, F., & Creutzig, F. (2023). Predicting Building Age From Urban Form At Large Scale. Computers, Environment And Urban Systems, 105, 102010.
- Neis, P., & Zipf, A. (2012). Analyzing The Contributor Activity Of A Volunteered Geographic Information Project — The Case Of Openstreetmap. Isprs International Journal Of Geo-Information, 1(2), 146-165
- Nguyen, H. D., Dao, N. D., & Shin, M. (2021). Prediction Of Seismic Drift Responses Of Planar Steel Moment Frames Using Artificial Neural Network And Extreme Gradient Boosting. Engineering Structures, 242, 112518.
- Oliveira, V. (2016). Urban Morphology: An Introduction To The Study Of The Physical Form Of Cities. Springer.
- Openstreetmap. (2024). Data Statistics. Diakses Pada 10 Maret 2024, Dari [Https://Planet.Openstreetmap.Org/Statistics/Data\\_Stats.Html](Https://Planet.Openstreetmap.Org/Statistics/Data_Stats.Html)
- Powers, D. M. (2020). Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness And Correlation. Arxiv Preprint Arxiv:2010.16061.
- Prayitno, B. (2014). Skema Inovatif Penanganan Permukiman Kumuh. Yogyakarta: Gajah Mada University Press.
- Prayojana, T. W., Mardhatil, M., Fazri, A. N., & Saputra, B. (2020). Dampak Urbanisasi Terhadap Pemukiman Kumuh (Slum Area). Jurnal Kependudukan Dan Pembangunan Lingkungan, 1(2), 60-69.
- Putra, A. C. P., Arimbawa, W., Widhaningtyas, T. U., & Fariz, T. R. (2023). Pemanfaatan Citra Sentinel-2 Untuk Deteksi Awal Permukiman Kumuh Di Sebagian Kota Denpasar Menggunakan Object-Based Image Analysis Dan Algoritma Machine Learning. Pranatacara Bhumiandala: Jurnal Riset Planologi, 4(2), 117-128.
- Putra, M. A., & Whardana, A. K. (2017). Humanitarian Openstreetmap Team Role Towards Mapping In Indonesia. Ijtb (International Journal Of Technology And Business), 1(1).
- Roth, V. L., & Mercer, J. M. (2000). Morphometrics In Development And Evolution. American Zoologist, 40(5), 801-810.
- Saab, David. (2003). Conceptualizing Space: Mapping Schemas as Meaningful Representations. <10.13140/2.1.3030.1767>.
- Sharma, D., & Kumar, N. (2017). A Review On Machine Learning Algorithms, Tasks And Applications. International Journal Of Advanced Research In Computer Engineering & Technology (Ijaracet), 6(10), 2278-1323.
- Simeone, O. (2018). A Very Brief Introduction To Machine Learning With Applications To Communication Systems. Ieee Transactions On Cognitive Communications And Networking, 4(4), 648-664.
- Soman, S., Beukes, A., Nederhood, C., Marchio, N., & Bettencourt, L. M. (2020). Worldwide Detection Of Informal Settlements Via Topological Analysis Of Crowdsourced Digital Maps. Isprs International Journal Of Geo-Information, 9(11), 685.
- Star, J., And Estes, J.E. (1990). Geographic Information System: An Introduction. Prentice Hall. Englewood Liff.

- Sturrock Hjw, Woolheater K, Bennett Af, Andrade-Pacheco R, Midekisa A (2018) Predicting Residential Structures From Open Source Remotely Enumerated Data Using Machine Learning. *Plos One* 13(9): E0204399. <Https://Doi.Org/10.1371/Journal.Pone.0204399>
- Suartha, N. (2016). Faktor-Faktor Yang Mempengaruhi Tingginya Laju Pertumbuhan Dan Implementasi Kebijakan Penduduk Di Provinsi Bali. *Piramida*, 12(1), 1-7.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth In The Detroit Region. *Economic Geography* 46: 234–40
- Touzani, S., & Granderson, J. (2021). Open Data And Deep Semantic Segmentation For Automated Extraction Of Building Footprints. *Remote Sensing*, 13(13), 2578.
- Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., & Porta, S. (2023). Eo+ Morphometrics: Understanding Cities Through Urban Morphology At Large Scale. *Landscape And Urban Planning*, 233, 104691.
- Waters, N. (2017). Tobler's First Law Of Geography. *The International Encyclopedia Of Geography*, 1-13.
- Wesnawa, I.G.A. (2015). Geografi Permukiman. Graha Ilmu: Yogyakarta
- Whitehand, J. W. R., Batty, M., & Longley, P. (1996). Fractal Cities: A Geometry Of Form And Function. *The Geographical Journal*, 162(1), 113. <Https://Doi.Org/10.2307/3060277>
- World Bank. (2022). Indonesia Overview. Diakses Pada 2 Maret 2024, Dari <Https://Www.Worldbank.Org/In/Country/Indonesia/Overview>
- Wright, J. K. (1936). A Method Of Mapping Densities Of Population: With Cape Cod As An Example. *Geographical Review*, 26(1), 103-110.
- Yang, L., & Shami, A. (2020). On Hyperparameter Optimization Of Machine Learning Algorithms: Theory And Practice. *Neurocomputing*, 415, 295-316.
- Zainab, R., & Jung-Hong, H. (2021). A Comparative Study Of Openstreetmap And Jakarta Satu Road Datasets.
- Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding The Best Classification Threshold In Imbalanced Classification. *Big Data Research*, 5, 2-8.

## **LAMPIRAN**

## Lampiran A

### **Script Pemrosesan Data dan Model Klasifikasi Bangunan Permukiman**

Lampiran ini merujuk ke pada sebuah GitHub Repository yang berisikan seluruh *script* yang digunakan dalam pemrosesan data dan pembuatan model klasifikasi.

Utas	QR Code
<u><a href="https://github.com/zhulfalian01/Residential-Classification">https://github.com/zhulfalian01/Residential-Classification</a></u>	

## Lampiran B

### Daftar Definisi Karakter Morfometrik yang Digunakan

Lampiran ini berisikan seluruh karakter individual beserta nama, satuan, nama variabel, dan definisi.

#### B.1. Bangunan dan Bangunan di Sekitarnya

1. Luas bangunan. Satuan: meter persegi. Variabel: b\_area.
2. Keliling bangunan. Satuan: meter. Variabel: b\_perimeter.
3. Konveksitas tapak bangunan. Definisi: Luas tapak bangunan dibagi dengan luas dari *convex hull* yang mengitari tapak. Satuan:  $x \in [0, 1]$ . Variabel: b\_convexity.
4. Kemiripan tapak bangunan dengan lingkaran (*circular compactness*). Definisi: Luas tapak bangunan dibagi dengan luas lingkaran terkecil yang dapat menutupi tapak bangunan. Satuan: Satuan:  $x \in [0, 1]$ . Variabel: b\_circular\_comp.
5. Rata-rata jarak sudut bangunan dengan *centroid*. Satuan: meter. Variabel: b\_ccd\_means.
6. Standar deviasi jarak sudut bangunan dengan *centroid*. Satuan: meter. Variabel: b\_ccd\_std.
7. Jumlah sudut bangunan. Satuan: jumlah. Variabel: b\_corners.
8. Rasio panjang dan lebar dari *bounding box* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_elongation.
9. Panjang sumbu diagonal terpanjang dari tapak bangunan. Satuan: meter. Variabel: b\_lal.
10. Kemiripan tapak bangunan dengan persegi panjang (*equivalent rectangular*). Satuan:  $x \in [0, 1]$ . Variabel: b\_eri.
11. Orientasi sumbu terpanjang *bounding box* terhadap sumbu mata angin. Satuan: derajat. Variabel: b\_orientation.
12. Rasio luas bangunan dan luas sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: b\_area\_ratio.
13. Rata-rata jarak bangunan dengan bangunan di sekitarnya. Satuan: meter. Variabel: b\_neigh\_dis.
14. Kerapatan bangunan dengan bangunan di sekitarnya. Satuan:  $x \in [0, 1]$ . Variabel: b\_adjacency.
15. Keselarasan orientasi bangunan dengan orientasi bangunan di sekitarnya. Satuan: derajat. Variabel: b\_alignment.
16. Penggunaan bangunan berdasarkan data OSM. Informasi penggunaan bangunan diturunkan menjadi tiga kategori, *residential*, *non-residential*, dan *yes* (tidak ada informasi penggunaan bangunan). Satuan: kategori. Variabel: building\_residential, building\_non\_residential, building\_yes.

#### B.2. Bangunan dengan *Buffer 50 m*

17. Jumlah bangunan di sekitar bangunan dalam *buffer 50 m* dari *centroid* bangunan. Satuan: jumlah. Variabel: b\_neighor\_50.
18. Rata-rata luas seluruh bangunan dalam *buffer 50 m* dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_avg\_area\_50.

19. Median luas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_median\_area\_50.
20. Standar deviasi luas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_std\_area\_50.
21. Rata-rata keliling seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_avg\_perimeter\_50.
22. Median keliling seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_median\_perimeter\_50.
23. Standar deviasi keliling seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_perimeter\_50.
24. Rata-rata konveksitas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_convexity\_50.
25. Median konveksitas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_convexity\_50.
26. Standar deviasi konveksitas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_convexity\_50.
27. Rata-rata *circular compactness* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_circular\_comp\_50.
28. Median *circular compactness* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_circular\_comp\_50.
29. Standar deviasi *circular compactness* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_circular\_comp\_50.
30. Rata-rata rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_elong\_50.
31. Median rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_elong\_50.
32. Standar deviasi rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_elong\_50.
33. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_avg\_lal\_50.
34. Median rasio lebar dan panjang dari *bounding box* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_median\_lal\_50.
35. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_lal\_50.
36. Rata-rata *equivalent rectangular* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_eri\_50.
37. Median *equivalent rectangular* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_50.
38. Standar deviasi *equivalent rectangular* seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_eri\_50.
39. Total luas seluruh bangunan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_total\_area\_50.

### B.3. Bangunan dengan *Buffer* 150 m

40. Jumlah bangunan di sekitar bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan: jumlah. Variabel: b\_neighbor\_150.
41. Rata-rata luas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan: meter persegi. Variabel: b\_avg\_area\_150.
42. Median luas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_median\_area\_150.
43. Standar deviasi luas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan: meter persegi. Variabel: b\_std\_area\_150.
44. Rata-rata keliling seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan: meter. Variabel: b\_avg\_perimeter\_150.
45. Median keliling seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan: meter. Variabel: b\_median\_perimeter\_150.
46. Standar deviasi keliling seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_perimeter\_150.
47. Rata-rata konveksitas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_convexity\_150.
48. Median konveksitas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan.  
Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_convexity\_150.
49. Standar deviasi konveksitas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_convexity\_150.
50. Rata-rata *circular compactness* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_circular\_comp\_150.
51. Median *circular compactness* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_circular\_comp\_150.
52. Standar deviasi *circular compactness* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_circular\_comp\_150.
53. Rata-rata rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_elong\_150.
54. Median rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_elong\_150.
55. Standar deviasi rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_elong\_150.
56. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_avg\_lal\_150.
57. Median rasio lebar dan panjang dari *bounding box* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_median\_lal\_150.
58. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_lal\_150.
59. Rata-rata *equivalent rectangular* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_eri\_150.

60. Median *equivalent rectangular* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_150.
61. Standar deviasi *equivalent rectangular* seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_eri\_150.
62. Total luas seluruh bangunan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_total\_area\_150.

#### **B.4. Bangunan dengan Buffer 300 m**

63. Jumlah bangunan di sekitar bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: jumlah. Variabel: b\_neighbor\_300.
64. Rata-rata luas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_avg\_area\_300.
65. Median luas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_median\_area\_300.
66. Standar deviasi luas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_std\_area\_300.
67. Rata-rata keliling seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_avg\_perimeter\_300.
68. Median keliling seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_median\_perimeter\_300.
69. Standar deviasi keliling seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_perimeter\_300.
70. Rata-rata konveksitas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_convexity\_300.
71. Median konveksitas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_convexity\_300.
72. Standar deviasi konveksitas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_convexity\_300.
73. Rata-rata *circular compactness* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_circular\_comp\_300.
74. Median *circular compactness* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_circular\_comp\_300.
75. Standar deviasi *circular compactness* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_circular\_comp\_300.
76. Rata-rata rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_elong\_300.
77. Median rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_elong\_300.
78. Standar deviasi rasio panjang dan lebar *bounding box* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_elong\_300.
79. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_avg\_lal\_300.

80. Median rasio lebar dan panjang dari *bounding* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_median\_lal\_300.
81. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: b\_std\_lal\_300.
82. Rata-rata *equivalent rectangular* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_avg\_eri\_300.
83. Median *equivalent rectangular* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_300.
84. Standar deviasi *equivalent rectangular* seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: b\_std\_eri\_300.
85. Total luas seluruh bangunan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter persegi. Variabel: b\_total\_area\_300.

#### **B.5. Sel Teselasi dan Sel Teselasi di Sekitarnya**

86. Luas sel teselasi. Satuan: meter persegi. Variabel: t\_area.
87. Keliling sel teselasi. Satuan: meter. Variabel: t\_perimeter.
88. Konveksitas sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_convexity.
89. Kemiripan sel teselasi dengan lingkaran (*circular compactness*). Definisi: Luas tapak sel teselasi dibagi dengan luas lingkaran terkecil yang dapat menutupi tapak sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_circular\_comp.
90. Rasio panjang dan lebar *bounding box* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_elong.
91. Panjang sumbu diagonal terpanjang dari sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_lal.
92. Kemiripan tapak sel teselasi dengan persegi panjang (*equivalent rectangular*). Satuan:  $x \in [0, 1]$ . Variabel: t\_eri.
93. Orientasi sumbu terpanjang *bounding box* sel teselasi terhadap sumbu mata angin. Satuan: derajat. Variabel: t\_orientation.
94. Keselarasan orientasi sel teselasi dengan orientasi bangunan. Satuan: derajat. Variabel: t\_alignment.
95. Luas area yang tertutupi oleh sel teselasi dan sel teselasi di sekitarnya. Satuan: meter persegi. Variabel: t\_cov\_area.

#### **B.6. Sel Teselasi dengan *buffer* 50 m**

96. Rata-rata luas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_avg\_area\_50.
97. Median luas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_median\_area\_50.
98. Standar deviasi luas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_std\_area\_50.
99. Rata-rata keliling seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_perimeter\_50.

100. Median keliling seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_perimeter\_50.
101. Standar deviasi keliling seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_perimeter\_50.
102. Rata-rata konveksitas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_convexity\_50.
103. Median konveksitas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_convexity\_50.
104. Standar deviasi konveksitas seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_convexity\_50.
105. Rata-rata *circular compactness* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_circular\_comp\_50.
106. Median *circular compactness* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_circular\_comp\_50.
107. Standar deviasi *circular compactness* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_circular\_comp\_50.
108. Rata-rata rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_elong\_50.
109. Median rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_elong\_50.
110. Standar deviasi rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_elong\_50.
111. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_lal\_50.
112. Median rasio lebar dan panjang dari *bounding box* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_lal\_50.
113. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_lal\_50.
114. Rata-rata *equivalent rectangular* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_eri\_50.
115. Median *equivalent rectangular* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_50.
116. Standar deviasi *equivalent rectangular* seluruh sel teselasi dalam *buffer* 50 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_eri\_50.

## B.7. Sel Teselasi dengan *buffer* 150 m

117. Rata-rata luas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_avg\_area\_150.
118. Median luas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_median\_area\_150.
119. Standar deviasi luas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_std\_area\_150.

120. Rata-rata keliling seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_perimeter\_150.
121. Median keliling seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_perimeter\_150.
122. Standar deviasi keliling seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_perimeter\_150.
123. Rata-rata konveksitas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_convexity\_150.
124. Median konveksitas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_convexity\_150.
125. Standar deviasi konveksitas seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_convexity\_150.
126. Rata-rata *circular compactness* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_circular\_comp\_150.
127. Median *circular compactness* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_circular\_comp\_150.
128. Standar deviasi *circular compactness* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_circular\_comp\_150.
129. Rata-rata rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_elong\_150.
130. Median rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_elong\_150.
131. Standar deviasi rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_elong\_150.
132. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_lal\_150.
133. Median rasio lebar dan panjang dari *bounding box* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_lal\_150.
134. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_lal\_150.
135. Rata-rata *equivalent rectangular* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_eri\_150.
136. Median *equivalent rectangular* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_150.
137. Standar deviasi *equivalent rectangular* seluruh sel teselasi dalam *buffer* 150 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_eri\_150.

## B.8. Sel Teselasi dengan *buffer* 300 m

138. Rata-rata luas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_avg\_area\_300.

139. Median luas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_median\_area\_300.
140. Standar deviasi luas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter persegi. Variabel: t\_std\_area\_300.
141. Rata-rata keliling seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_perimeter\_300.
142. Median keliling seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_perimeter\_300.
143. Standar deviasi keliling seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_perimeter\_300.
144. Rata-rata konveksitas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_convexity\_300.
145. Median konveksitas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_convexity\_300.
146. Standar deviasi konveksitas seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_convexity\_300.
147. Rata-rata *circular compactness* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_circular\_comp\_300.
148. Median *circular compactness* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_circular\_comp\_300.
149. Standar deviasi *circular compactness* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_circular\_comp\_300.
150. Rata-rata rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_elong\_300.
151. Median rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_median\_elong\_300.
152. Standar deviasi rasio panjang dan lebar *bounding box* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_elong\_300.
153. Rata-rata sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_avg\_lal\_300.
154. Median rasio lebar dan panjang dari *bounding box* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_median\_lal\_300.
155. Standar deviasi sumbu diagonal *bounding box* terpanjang seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan: meter. Variabel: t\_std\_lal\_300.
156. Rata-rata *equivalent rectangular* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_avg\_eri\_300.
157. Median *equivalent rectangular* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: b\_median\_eri\_300.
158. Standar deviasi *equivalent rectangular* seluruh sel teselasi dalam *buffer* 300 m dari *centroid* sel teselasi. Satuan:  $x \in [0, 1]$ . Variabel: t\_std\_eri\_300.

## B.9. Jalan dan Persimpangan Terdekat

159. Derajat cabang persimpangan. Definisi: derajat ditentukan dengan jumlah cabang dari suatu persimpangan. Satuan: jumlah. Variabel: n\_degree.
160. Jarak antara bangunan dengan jalan terdekat. Satuan: meter. Variabel: s\_closest\_street.
161. Jarak antara bangunan dengan persimpangan terdekat. Satuan: meter. Variabel: s\_closest\_intersect\_dist.
162. Rata-rata lebar segmen jalan terdekat dengan bangunan. Definisi: lebar didefinisikan dengan rata-rata jarak antara bangunan di kedua sisi . Satuan: meter. Variabel: s\_width.
163. Standar deviasi lebar jalan terdekat dengan bangunan. Satuan: meter. Variabel: s\_width\_def.
164. Keterbukaan jalan terdekat dengan bangunan. Definisi: proporsi jalan tanpa ada bangunan di sisi jalan. Satuan:  $x \in [0, 1]$ . Variabel: s\_openness.
165. Jumlah panjang seluruh segmen pada jalan terdekat dengan bangunan. Satuan: meter. Variabel: s\_sum\_length.
166. Keselarasan orientasi jalan terdekat dengan orientasi bangunan. Satuan: derajat. Variabel: s\_alignment.
167. Panjang segmen jalan terdekat dengan bangunan. Satuan: meter. Variabel: s\_length.
168. Indeks *betweenness* lokal (500 m) pada segmen jalan terdekat dengan bangunan. Definisi: Seberapa sering suatu jalan menjadi penghubung dengan jaringan jalan lainnya. Satuan:  $x \in [0, 1]$ . Variabel: s\_betweenness\_500.
169. Indeks *betweenness* global pada jalan terdekat dengan bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_betweenness\_global.
170. Indeks *closeness* lokal (500 m) pada segmen jalan terdekat dengan bangunan. Definisi: Seberapa dekat suatu segmen jalan relatif terhadap jaringan jalan di sekitarnya. Satuan:  $x \in [0, 1]$ . Variabel: s\_closeness\_500.
171. Indeks *closeness* global pada segmen jalan terdekat dengan bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_closeness\_global.
172. Indeks *straightness* lokal (500 m) pada segmen jalan terdekat dengan bangunan. Definisi: Seberapa lurus suatu segmen jalan relatif terhadap jaringan jalan di sekitarnya. Satuan:  $x \in [0, 1]$ . Variabel: s\_straightness\_500.
173. Indeks *straightness* global pada segmen jalan terdekat dengan bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_straightness\_global.
174. Indeks *meshedness* lokal (500 m) pada segmen jalan terdekat dengan bangunan. Definisi: Seberapa terpusat suatu segmen jalan relatif terhadap jaringan jalan di sekitarnya. Satuan:  $x \in [0, 1]$ . Variabel: s\_meshedness\_500.
175. Indeks *meshedness* global pada segmen jalan terdekat dengan bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_meshedness\_global.
176. Jenis jalan berdasarkan OSM. Informasi jenis jalan dikelompokkan menjadi lima: *major roads, service roads, residential, pedestrian, dan others*. Satuan: kategori.

Variabel: highway\_Major Roads, highway\_Service Roads, highway\_Residential, highway\_Pedestrian, dan highway\_Others.

#### B.10. Jalan dan Persimpangan dalam *buffer* 50 m

177. Rata-rata indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_betweenness\_500\_street\_50.
178. Nilai maksimum indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_betweenness\_500\_street\_50.
179. Rata-rata indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_closeness\_500\_street\_50.
180. Nilai maksimum indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_closeness\_500\_street\_50.
181. Rata-rata indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_straightness\_500\_street\_50.
182. Nilai maksimum indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_straightness\_500\_street\_50.
183. Rata-rata indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_meshedness\_500\_street\_50.
184. Nilai maksimum indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_meshedness\_500\_street\_50.
185. Rata-rata lebar jalan dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_width\_street\_50.
186. Standar deviasi lebar jalan dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_std\_width\_street\_50.
187. Rata-rata panjang segmen jalan dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_length\_street\_50.
188. Total panjang jalan dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_total\_street\_50.
189. Standar deviasi panjang jalan dari setiap jalan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_std\_total\_street\_50.
190. Jumlah persimpangan dalam *buffer* 50 m dari *centroid* bangunan. Satuan: jumlah. Variabel: s\_intersection\_count\_50.

## **B.11. Jalan dan Persimpangan dengan *buffer* 150 m**

191. Rata-rata indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_betweenness\_500\_street\_150.
192. Nilai maksimum indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_betweenness\_500\_street\_150.
193. Rata-rata indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_closeness\_500\_street\_150.
194. Nilai maksimum indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_closeness\_500\_street\_150.
195. Rata-rata indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_straightness\_500\_street\_150.
196. Nilai maksimum indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_straightness\_500\_street\_150.
197. Rata-rata indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_meshedness\_500\_street\_150.
198. Nilai maksimum indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_max\_meshedness\_500\_street\_150.
199. Rata-rata lebar jalan dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_width\_street\_150.
200. Standar deviasi lebar jalan dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_std\_width\_street\_150.
201. Rata-rata panjang segmen jalan dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_length\_street\_150.
202. Total panjang jalan dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_mean\_total\_street\_150.
203. Standar deviasi panjang jalan dari setiap jalan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: meter. Variabel: s\_std\_total\_street\_150.
204. Jumlah persimpangan dalam *buffer* 150 m dari *centroid* bangunan. Satuan: jumlah. Variabel: s\_intersection\_count\_150.

## **B.12. Jalan dan Persimpangan dengan *buffer* 300 m**

205. Rata-rata indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: s\_av\_betweenness\_500\_street\_300.

206. Nilai maksimum indeks *betweenness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_max\_betweenness\_500\_street\_300*.
207. Rata-rata indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_av\_closeness\_500\_street\_300*.
208. Nilai maksimum indeks *closeness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_max\_closeness\_500\_street\_300*
209. Rata-rata indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_av\_straightness\_500\_street\_300*.
210. Nilai maksimum indeks *straightness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_max\_straightness\_500\_street\_300*.
211. Rata-rata indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_av\_meshedness\_500\_street\_300*.
212. Nilai maksimum indeks *meshedness* lokal (500 m) dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan:  $x \in [0, 1]$ . Variabel: *s\_max\_meshedness\_500\_street\_300*.
213. Rata-rata lebar jalan dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: *s\_mean\_width\_street\_300*.
214. Standar deviasi lebar jalan dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: *s\_std\_width\_street\_300*.
215. Rata-rata panjang segmen jalan dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: *s\_mean\_length\_street\_300*.
216. Total panjang jalan dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: *s\_mean\_total\_street\_300*.
217. Standar deviasi panjang jalan dari setiap jalan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: meter. Variabel: *s\_std\_total\_street\_300*.
218. Jumlah persimpangan dalam *buffer* 300 m dari *centroid* bangunan. Satuan: jumlah. Variabel: *s\_intersection\_count\_300*.

## Referensi:

1. Fleischmann, M. (2019). Momepy: Urban Morphology Measuring Toolkit. *Journal Of Open Source Software*, 4(43), 1807.
2. Milojevic-Dupont N, Hans N, Kaack Lh, Zumwald M, Andrieux F, De Barros Soares D, Et Al. (2020) Learning From Urban Form To Predict Building Heights. *Plos One* 15(12): E0242010. [Https://Doi.Org/10.1371/Journal.Pone.0242010](https://doi.org/10.1371/journal.pone.0242010).

## Lampiran C

### Signifikansi Karakter Morfometrik

Lampiran ini berisikan informasi *gain* atau nilai yang merepresentasikan signifikansi dampak dari kehadiran tiap karakter morfometrik.

<b>Variabel Karakter</b>	<b>Kelompok</b>	<b>Skala</b>	<b>Gain</b>
b_neighbor_50	bangunan	50 m	613,9412231
b_avg_area_50	bangunan	50 m	16,60886765
b_median_area_50	bangunan	50 m	22,90306664
b_std_area_50	bangunan	50 m	155,0571442
b_avg_perimeter_50	bangunan	50 m	30,46068764
b_median_perimeter_50	bangunan	50 m	19,41451263
b_std_perimeter_50	bangunan	50 m	85,34621429
b_avg_convexity_50	bangunan	50 m	14,72685623
b_median_convexity_50	bangunan	50 m	24,77246094
b_std_convexity_50	bangunan	50 m	29,96344566
b_avg_circular_comp_50	bangunan	50 m	36,57291031
b_median_circular_comp_50	bangunan	50 m	11,36078453
b_std_circular_comp_50	bangunan	50 m	33,01060104
b_avg_elong_50	bangunan	50 m	84,09265137
b_median_elong_50	bangunan	50 m	24,86128616
b_std_elong_50	bangunan	50 m	22,9515419
b_avg_lal_50	bangunan	50 m	15,57551384
b_median_lal_50	bangunan	50 m	21,43028831
b_std_lal_50	bangunan	50 m	392,5307922
b_avg_eri_50	bangunan	50 m	21,87806511
b_median_eri_50	bangunan	50 m	24,53138733
b_std_eri_50	bangunan	50 m	20,07085037
b_total_area_50	bangunan	50 m	96,50524902
b_neighbor_150	bangunan	150 m	19,92190361
b_avg_area_150	bangunan	150 m	25,24506187
b_median_area_150	bangunan	150 m	30,51528168
b_std_area_150	bangunan	150 m	116,9955368
b_avg_perimeter_150	bangunan	150 m	28,98542595
b_median_perimeter_150	bangunan	150 m	64,939888
b_std_perimeter_150	bangunan	150 m	16,30671883
b_avg_convexity_150	bangunan	150 m	36,903862
b_median_convexity_150	bangunan	150 m	11,95582199
b_std_convexity_150	bangunan	150 m	30,56795883
b_avg_circular_comp_150	bangunan	150 m	21,06927681
b_median_circular_comp_150	bangunan	150 m	29,09333801
b_std_circular_comp_150	bangunan	150 m	69,58799744
b_avg_elong_150	bangunan	150 m	60,63422775

b_median_elong_150	bangunan	150 m	21,83555603
b_std_elong_150	bangunan	150 m	29,37328529
b_avg_lal_150	bangunan	150 m	24,17714691
b_median_lal_150	bangunan	150 m	35,24161911
b_std_lal_150	bangunan	150 m	328,1348267
b_avg_eri_150	bangunan	150 m	26,83508492
b_median_eri_150	bangunan	150 m	25,36627579
b_std_eri_150	bangunan	150 m	26,75452042
b_total_area_150	bangunan	150 m	27,72348785
b_neighbor_300	bangunan	300 m	29,72100258
b_avg_area_300	bangunan	300 m	26,99151039
b_median_area_300	bangunan	300 m	30,56696129
b_std_area_300	bangunan	300 m	29,81433487
b_avg_perimeter_300	bangunan	300 m	23,89245033
b_median_perimeter_300	bangunan	300 m	44,85174561
b_std_perimeter_300	bangunan	300 m	23,51257706
b_avg_convexity_300	bangunan	300 m	20,98301697
b_std_convexity_300	bangunan	300 m	24,74589539
b_avg_circular_comp_300	bangunan	300 m	37,94234848
b_median_circular_comp_300	bangunan	300 m	28,88127708
b_std_circular_comp_300	bangunan	300 m	45,87437439
b_avg_elong_300	bangunan	300 m	42,12701797
b_median_elong_300	bangunan	300 m	30,81214142
b_std_elong_300	bangunan	300 m	26,15207481
b_avg_lal_300	bangunan	300 m	26,24660873
b_median_lal_300	bangunan	300 m	56,86575317
b_std_lal_300	bangunan	300 m	19,62453842
b_avg_eri_300	bangunan	300 m	28,049366
b_median_eri_300	bangunan	300 m	35,41049957
b_std_eri_300	bangunan	300 m	29,81323624
b_total_area_300	bangunan	300 m	36,82967758
b_median_convexity_300	bangunan	300 m	0
b_area	bangunan	Objek dan sekitarnya	18,83686638
b_perimeter	bangunan	Objek dan sekitarnya	23,92859268
b_convexity	bangunan	Objek dan sekitarnya	64,06705475
b_circular_comp	bangunan	Objek dan sekitarnya	9,33370018
b_ccd_means	bangunan	Objek dan sekitarnya	70,06504059
b_ccd_std	bangunan	Objek dan sekitarnya	20,81796837
b_corners	bangunan	Objek dan sekitarnya	47,41192245
b_elong	bangunan	Objek dan sekitarnya	16,64999199
b_lal	bangunan	Objek dan sekitarnya	32,27160263
b_eri	bangunan	Objek dan sekitarnya	7,637419701
b_orientation	bangunan	Objek dan sekitarnya	21,83316231
b_area_ratio	bangunan	Objek dan sekitarnya	3,576167583

b_neigh_dis	bangunan	Objek dan sekitarnya	232,0460663
b_alignment	bangunan	Objek dan sekitarnya	23,81115532
b_adjacency	bangunan	Objek dan sekitarnya	22,91813469
building_non_residential	bangunan	Objek dan sekitarnya	288,9790344
building_residential	bangunan	Objek dan sekitarnya	15,87153912
building_yes	bangunan	Objek dan sekitarnya	265,4560852
s_av_betweenness_500_street_50	jalan	50 m	64,32975006
s_max_betweenness_500_street_50	jalan	50 m	148,3747559
s_av_closeness_500_street_50	jalan	50 m	40,95786667
s_max_closeness_500_street_50	jalan	50 m	105,7446289
s_av_straightness_500_street_50	jalan	50 m	35,27920914
s_max_straightness_500_street_50	jalan	50 m	137,6222382
s_av_meshedness_500_street_50	jalan	50 m	45,79326248
s_max_meshedness_500_street_50	jalan	50 m	32,08342743
s_mean_width_street_50	jalan	50 m	260,762085
s_std_width_street_50	jalan	50 m	41,60715103
s_mean_length_street_50	jalan	50 m	25,03194046
s_mean_total_street_50	jalan	50 m	23,73696518
s_std_total_street_50	jalan	50 m	52,21480942
s_intersection_count_50	jalan	50 m	66,82021332
s_av_betweenness_500_street_150	jalan	150 m	26,73467636
s_max_betweenness_500_street_150	jalan	150 m	48,58120728
s_av_closeness_500_street_150	jalan	150 m	32,86569977
s_max_closeness_500_street_150	jalan	150 m	33,84966278
s_av_straightness_500_street_150	jalan	150 m	25,07902336
s_max_straightness_500_street_150	jalan	150 m	29,39976311
s_av_meshedness_500_street_150	jalan	150 m	25,71966743
s_max_meshedness_500_street_150	jalan	150 m	29,13081169
s_mean_width_street_150	jalan	150 m	40,56703949
s_std_width_street_150	jalan	150 m	47,25502014
s_mean_length_street_150	jalan	150 m	15,62164593
s_mean_total_street_150	jalan	150 m	17,50924873
s_std_total_street_150	jalan	150 m	26,15512848
s_intersection_count_150	jalan	150 m	53,51880264
s_av_betweenness_500_street_300	jalan	300 m	27,92173386
s_max_betweenness_500_street_300	jalan	300 m	45,73891068
s_av_closeness_500_street_300	jalan	300 m	27,5062294
s_max_closeness_500_street_300	jalan	300 m	35,35330582
s_av_straightness_500_street_300	jalan	300 m	25,4251442
s_max_straightness_500_street_300	jalan	300 m	35,18997192
s_av_meshedness_500_street_300	jalan	300 m	31,48968506
s_max_meshedness_500_street_300	jalan	300 m	25,49291992
s_mean_width_street_300	jalan	300 m	31,06741524
s_std_width_street_300	jalan	300 m	33,74745941

s_mean_length_street_300	jalan	300 m	21,52046013
s_mean_total_street_300	jalan	300 m	17,88650322
s_std_total_street_300	jalan	300 m	27,2406292
s_intersection_count_300	jalan	300 m	29,04560089
s_closest_street	jalan	Objek dan sekitarnya	24,32599831
n_degree	jalan	Objek dan sekitarnya	203,8095856
s_closest_intersect_dist	jalan	Objek dan sekitarnya	24,49687195
s_width	jalan	Objek dan sekitarnya	51,39806366
s_width_def	jalan	Objek dan sekitarnya	22,14625549
s_openness	jalan	Objek dan sekitarnya	27,43435097
s_sum_length	jalan	Objek dan sekitarnya	18,56926346
s_alignment	jalan	Objek dan sekitarnya	12,89126682
s_length	jalan	Objek dan sekitarnya	44,90603256
s_betweenness_500	jalan	Objek dan sekitarnya	65,16465759
s_betweenness_global	jalan	Objek dan sekitarnya	137,6688538
s_closeness_500	jalan	Objek dan sekitarnya	70,00543976
s_closeness_global	jalan	Objek dan sekitarnya	29,70680618
s_straightness_500	jalan	Objek dan sekitarnya	40,87596512
s_straightness_global	jalan	Objek dan sekitarnya	38,44382477
s_meshedness_500	jalan	Objek dan sekitarnya	46,35113525
s_meshedness_global	jalan	Objek dan sekitarnya	14,65375519
highway_Major Roads	jalan	Objek dan sekitarnya	176,177536
highway_Other	jalan	Objek dan sekitarnya	259,1522217
highway_Residential	jalan	Objek dan sekitarnya	242,13797
highway_Service Roads	jalan	Objek dan sekitarnya	75,97158813
highway_Pedestrian	jalan	Objek dan sekitarnya	0
t_avg_area_50	sel teselasi	50 m	51,165905
t_median_area_50	sel teselasi	50 m	33,22733307
t_std_area_50	sel teselasi	50 m	25,14842987
t_avg_perimeter_50	sel teselasi	50 m	53,3002739
t_median_perimeter_50	sel teselasi	50 m	24,08240128
t_std_perimeter_50	sel teselasi	50 m	357,1034241
t_avg_convexity_50	sel teselasi	50 m	12,72476196
t_median_convexity_50	sel teselasi	50 m	11,28301907
t_std_convexity_50	sel teselasi	50 m	10,14648247
t_avg_circular_comp_50	sel teselasi	50 m	29,39930725
t_median_circular_comp_50	sel teselasi	50 m	21,51333427
t_std_circular_comp_50	sel teselasi	50 m	13,58448124
t_avg_elong_50	sel teselasi	50 m	15,64741135
t_median_elong_50	sel teselasi	50 m	14,48975086
t_std_elong_50	sel teselasi	50 m	16,95100403
t_avg_lal_50	sel teselasi	50 m	103,7403336
t_median_lal_50	sel teselasi	50 m	27,913517
t_std_lal_50	sel teselasi	50 m	136,9020233

t_avg_eri_50	sel teselasi	50 m	10,00191402
t_median_eri_50	sel teselasi	50 m	13,73424721
t_std_eri_50	sel teselasi	50 m	15,72633648
t_avg_area_150	sel teselasi	150 m	26,61592293
t_median_area_150	sel teselasi	150 m	29,18608284
t_std_area_150	sel teselasi	150 m	17,85234451
t_avg_perimeter_150	sel teselasi	150 m	37,72530746
t_median_perimeter_150	sel teselasi	150 m	34,58987808
t_std_perimeter_150	sel teselasi	150 m	38,21432877
t_avg_convexity_150	sel teselasi	150 m	15,98663616
t_median_convexity_150	sel teselasi	150 m	18,78831673
t_std_convexity_150	sel teselasi	150 m	17,87027359
t_avg_circular_comp_150	sel teselasi	150 m	21,36687279
t_median_circular_comp_150	sel teselasi	150 m	19,69046021
t_std_circular_comp_150	sel teselasi	150 m	23,67832756
t_avg_elong_150	sel teselasi	150 m	41,91444397
t_median_elong_150	sel teselasi	150 m	25,27956009
t_std_elong_150	sel teselasi	150 m	26,62550735
t_avg_lal_150	sel teselasi	150 m	21,51426315
t_median_lal_150	sel teselasi	150 m	43,00323868
t_std_lal_150	sel teselasi	150 m	25,82460785
t_avg_eri_150	sel teselasi	150 m	25,26465225
t_median_eri_150	sel teselasi	150 m	17,00137711
t_std_eri_150	sel teselasi	150 m	13,39375496
t_avg_area_300	sel teselasi	300 m	32,82465744
t_median_area_300	sel teselasi	300 m	63,21641159
t_std_area_300	sel teselasi	300 m	25,99839592
t_avg_perimeter_300	sel teselasi	300 m	38,61968994
t_median_perimeter_300	sel teselasi	300 m	86,53562164
t_std_perimeter_300	sel teselasi	300 m	29,14257622
t_avg_convexity_300	sel teselasi	300 m	24,93153381
t_median_convexity_300	sel teselasi	300 m	20,19214058
t_std_convexity_300	sel teselasi	300 m	30,43877411
t_avg_circular_comp_300	sel teselasi	300 m	34,3968544
t_median_circular_comp_300	sel teselasi	300 m	35,28639603
t_std_circular_comp_300	sel teselasi	300 m	22,04109764
t_avg_elong_300	sel teselasi	300 m	28,5816555
t_median_elong_300	sel teselasi	300 m	28,42639351
t_std_elong_300	sel teselasi	300 m	27,65711403
t_avg_lal_300	sel teselasi	300 m	36,57650375
t_median_lal_300	sel teselasi	300 m	46,40536118
t_std_lal_300	sel teselasi	300 m	21,86539459
t_avg_eri_300	sel teselasi	300 m	28,97998238
t_median_eri_300	sel teselasi	300 m	14,75779915

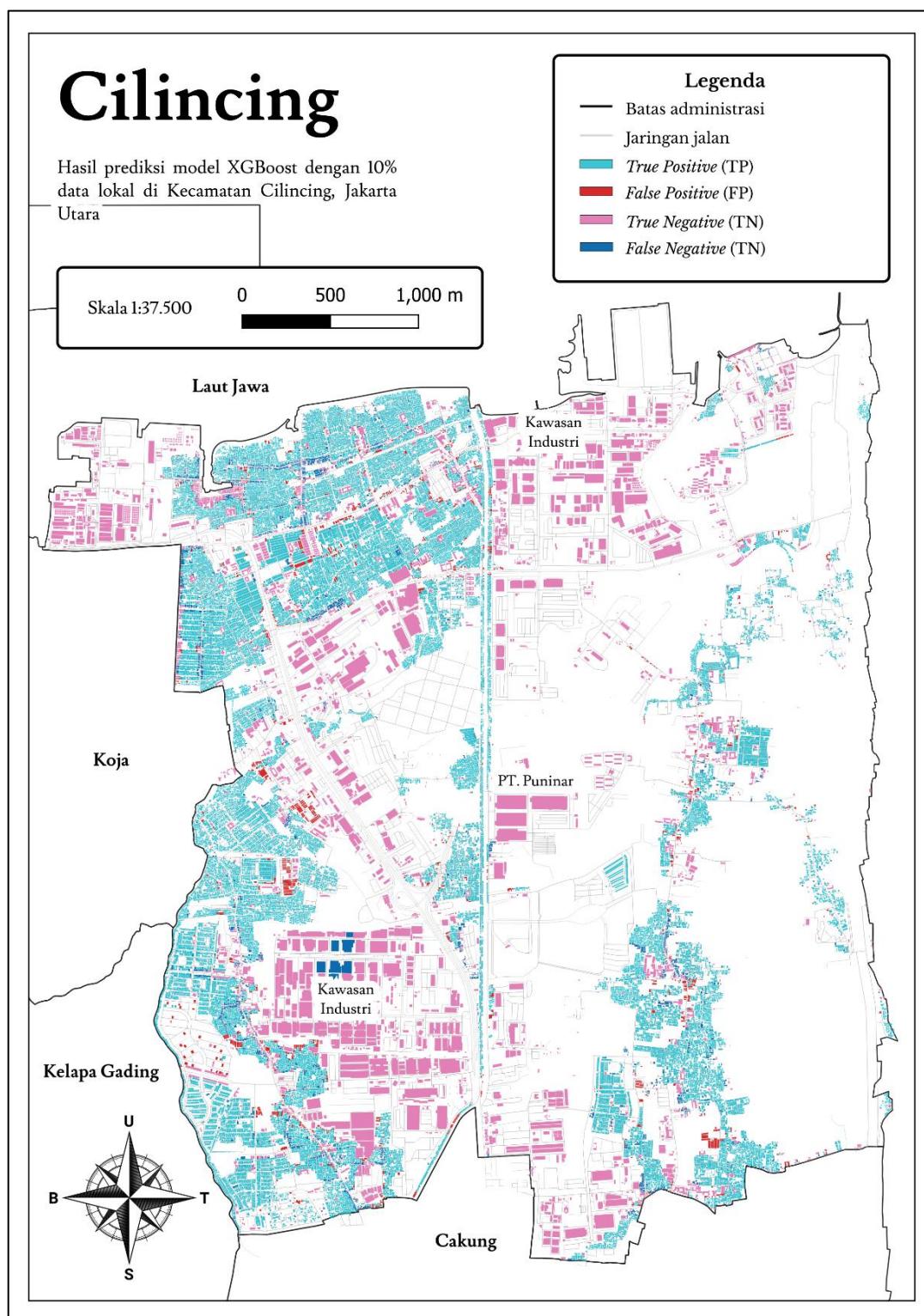
t_std_eri_300	sel teselasi	300 m	37,25948334
t_area	sel teselasi	Objek dan Sekitarnya	196,2753448
t_perimeter	sel teselasi	Objek dan Sekitarnya	551,4871826
t_convexity	sel teselasi	Objek dan Sekitarnya	25,1702919
t_circular_comp	sel teselasi	Objek dan Sekitarnya	6,939546108
t_elong	sel teselasi	Objek dan Sekitarnya	13,98955059
t_lal	sel teselasi	Objek dan Sekitarnya	679,1129761
t_eri	sel teselasi	Objek dan Sekitarnya	11,98602581
t_orientation	sel teselasi	Objek dan Sekitarnya	17,95364571
t_alignment	sel teselasi	Objek dan Sekitarnya	3,433274984
t_cov_area	sel teselasi	Objek dan Sekitarnya	1138,256958

## Lampiran D

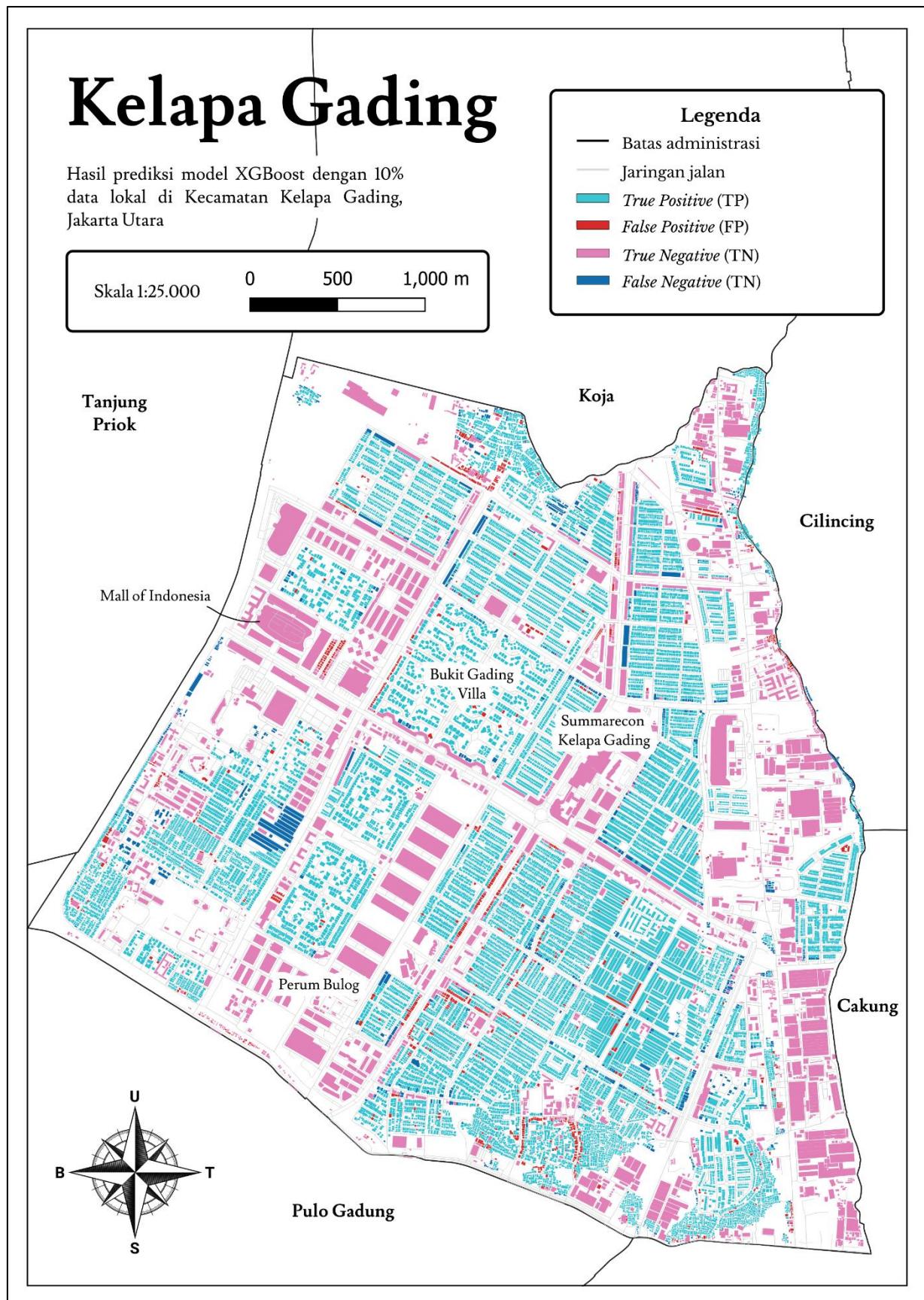
### Visualisasi Hasil Prediksi Model dengan 10% Data Lokal di Kota Jakarta Utara

Lampiran ini berisikan peta hasil prediksi Model XGBoost dengan penambahan 10% data lokal pada setiap kecamatan di Kota Jakarta Utara. Pada peta ini, permukiman ditunjukkan dengan kelas positif dan non-permukiman ditunjukkan dengan kelas negatif.

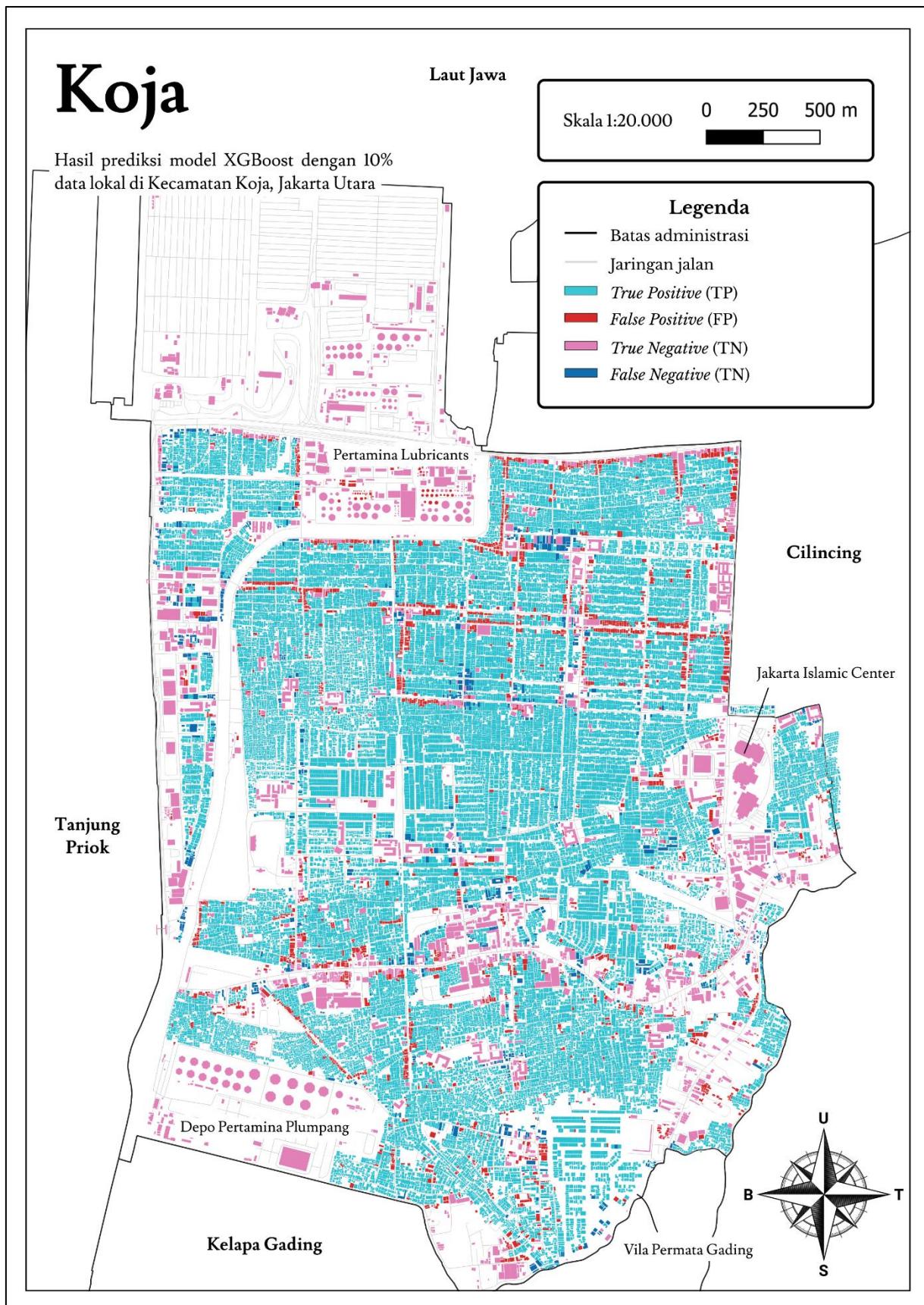
#### D.1. Kecamatan Cilincing



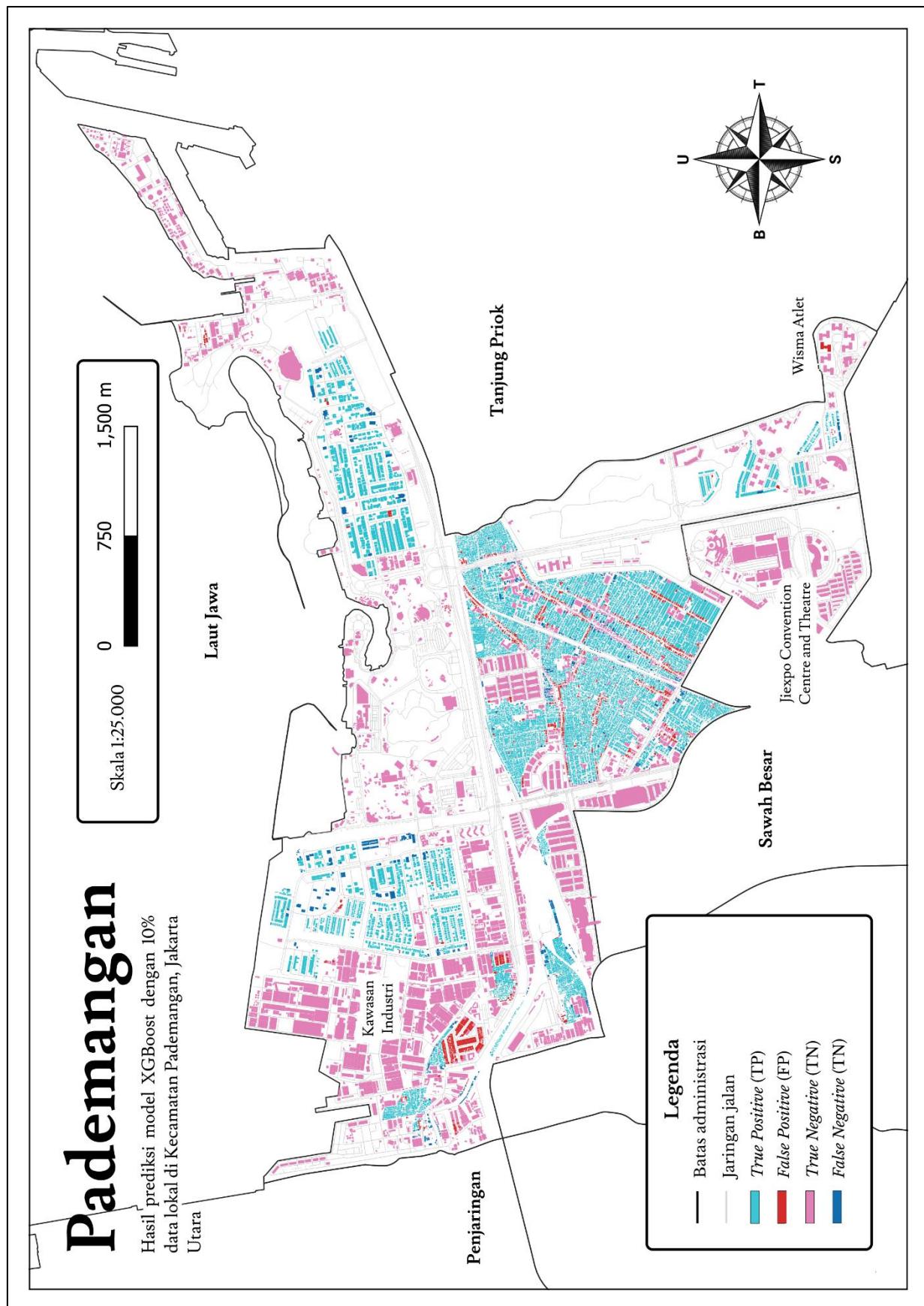
## D.2. Kecamatan Kelapa Gading



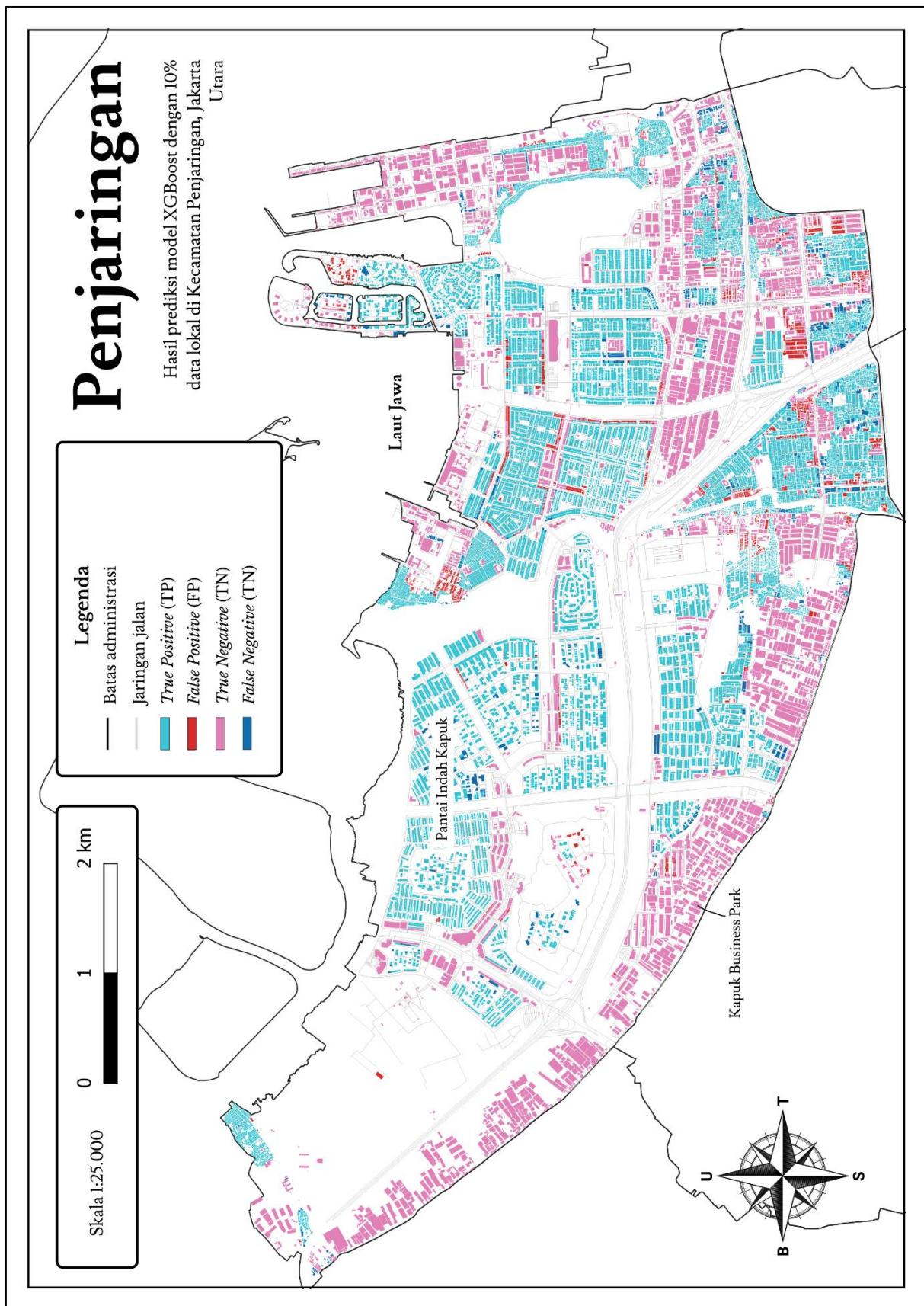
### D.3. Kecamatan Koja



#### D.4. Kecamatan Pademangan



## D.5. Kecamatan Penjaringan



## D.6. Kecamatan Tanjung Priok

