

RESEARCH ARTICLE

Open Access

Penerapan *Ensemble Machine Learning Random Forest* dan *XGBoost* dengan *Explainable Artificial Intelligence* (XAI) untuk Prediksi *Urban Heat Island* dan *Land Surface Temperature* di DKI Jakarta

Hertanto Suryoprayogo^{1*}, Widang Muttaqin², Annisa Desianty³

^{1*,2,3} Program Studi Teknologi Informasi, Direktorat Kampus Jakarta, Universitas Telkom, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia.

*Correspondence email:
hertantosp@telkomuniversity.ac.id.

Received: 24 July 2026.
Revised: 20 August 2026.
Accepted: 28 August 2026.
Published: 30 August 2026.

Full list of author information is
available at the end of the article.

Abstract

The Urban Heat Island (UHI) effect in tropical urban settings arises from interactions among built surfaces, vegetation, water bodies, and urban energy dynamics. This study modeled Land Surface Temperature (LST) in DKI Jakarta using Random Forest and XGBoost optimized with RandomizedSearchCV and 5-fold cross-validation. The analysis used 5,821 grid points at approximately 300 m resolution and five predictors: road density, NDVI, NDBI, NDWI, and distance to green open space. XGBoost slightly outperformed Random Forest, achieving $R^2 = 0.507$ and $RMSE = 1.830^\circ\text{C}$ compared with $R^2 = 0.497$ and $RMSE = 1.849^\circ\text{C}$, although the difference was not statistically significant (Wilcoxon, $p = 0.352$). The RF-XGBoost ensemble did not improve performance due to very high residual correlation ($r = 0.988$) and a theoretical ensemble standard deviation reduction of only $\sim 0.3\%$. SHAP analysis identified NDBI as the dominant predictor ($\text{mean}|SHAP| = 0.986$), with the strongest interaction between NDBI and road density (0.101). Hyperparameter tuning changed model ranking, statistical significance, and the leading SHAP interaction pair.

Keywords: Urban Heat Island; Land Surface Temperature; Random Forest; XGBoost; Explainable AI; Remote Sensing.

Abstrak

Urban Heat Island (UHI) pada kawasan perkotaan tropis merupakan fenomena spasial yang terbentuk dari interaksi permukaan terbangun, vegetasi, air, dan energi perkotaan. Penelitian ini memodelkan Land Surface Temperature (LST) di DKI Jakarta dengan 5.821 titik grid beresolusi sekitar 300 m menggunakan Random Forest dan XGBoost yang dioptimasi melalui RandomizedSearchCV dengan 5-fold cross-validation. Model menggunakan lima prediktor, yaitu NDVI, NDBI, NDWI (Sentinel-2), jarak ke ruang terbuka hijau, dan kepadatan jalan. XGBoost sedikit lebih baik daripada Random Forest, dengan $R^2 = 0,507$ dan $RMSE = 1,830^\circ\text{C}$ dibandingkan $R^2 = 0,497$ dan $RMSE = 1,849^\circ\text{C}$, tetapi perbedaannya tidak signifikan secara statistik (Wilcoxon, $p = 0,352$). Ensemble RF-XGBoost tidak meningkatkan performa karena korelasi residual sangat tinggi ($r = 0,988$) dan reduksi simpangan baku ensemble teoretis hanya sekitar 0,3%. Analisis SHAP mengidentifikasi NDBI sebagai prediktor dominan ($\text{mean}|SHAP| = 0,986$), dengan interaksi terkuat antara NDBI dan kepadatan jalan (0,101). Hyperparameter tuning mengubah peringkat model, signifikansi statistik, dan pasangan interaksi SHAP yang dominan.

Kata Kunci: Urban Heat Island; Land Surface Temperature; Random Forest; XGBoost; Explainable AI; Remote Sensing.



1. Pendahuluan

Urban Heat Island (UHI) merupakan fenomena peningkatan suhu permukaan di kawasan urban yang berkaitan dengan perubahan keseimbangan energi akibat urbanisasi, termasuk dominasi permukaan terbangun, material aspal, dan bangunan yang menyimpan serta melepaskan panas, serta berkurangnya evapotranspirasi vegetasi (Masson *et al.*, 2021; Oke, 1982; Oke *et al.*, 2017). Dalam kerangka Oke (1982), energi radiasi netto yang diserap material perkotaan tersimpan sebagai panas sensibel akibat kapasitas panas dan konduktivitas termal yang tinggi, sementara berkurangnya evapotranspirasi mengurangi proses pendinginan permukaan. Kerangka tersebut dikembangkan lebih lanjut dalam kajian iklim perkotaan (Oke *et al.*, 2017; Masson *et al.*, 2021). Land Surface Temperature (LST) yang diturunkan dari citra satelit termal banyak digunakan untuk mengkaji kondisi termal permukaan perkotaan karena memiliki cakupan spasial yang luas dengan biaya relatif rendah (Li *et al.*, 2017). Perkembangan tata ruang dan kawasan di Jakarta berlangsung dengan perubahan volume dan morfologi bangunan yang pesat (Sarker *et al.*, 2024), sementara ruang terbuka hijau relatif terbatas dan tersebar tidak merata. Dalam Machine Learning (ML), hubungan nonlinier antara LST dan faktor yang memengaruhinya dapat dipelajari secara langsung tanpa mengasumsikan bentuk hubungan linear seperti pada regresi konvensional. Muzaky dan Jaelani (2019) menemukan hubungan negatif yang konsisten antara vegetasi dan LST di Jakarta. Siswanto *et al.* (2023) melaporkan intensitas Surface Urban Heat Island (SUHI) Jakarta metropolitan berkisar 3–6°C menggunakan pendekatan urban-rural gradient. Fajary *et al.* (2024) mengkaji SUHI intensity (SUHII) di beberapa kawasan di Jawa menggunakan citra multitemporal pada tahun 2000, 2009, dan 2019 serta menggunakan UTFVI sebagai metode standarisasi. Sarker *et al.* (2024) menunjukkan keterkaitan volume bangunan terbangun (urban built-up volume) dengan UHI di Jakarta, yang mendukung relevansi variabel morfologi tiga dimensi sebagai arah pengembangan analisis. Perkembangan Machine Learning dan Explainable AI (XAI) untuk pemodelan LST/UHI selanjutnya diteliti oleh Tahooni *et al.* (2025) melalui penerapan RF, SVR, dan XGBoost dengan SHAP untuk downscaling LST. Kondisi iklim semi-arid dalam penelitian tersebut berbeda secara fundamental dari Jakarta yang beriklim tropis, sehingga nilai R^2 tidak dapat dibandingkan secara langsung karena terdapat perbedaan task, resolusi, prediktor, dan desain validasi. Yang *et al.* (2026) dan Wei *et al.* (2025) menerapkan kerangka XGBoost-SHAP untuk menguji pengaruh bentuk kota terhadap LST pada skala grid atau blok, termasuk deteksi threshold dan interaction effects. Penelitian-penelitian tersebut menjadi rujukan metodologis utama bagi kerangka analisis penelitian ini.

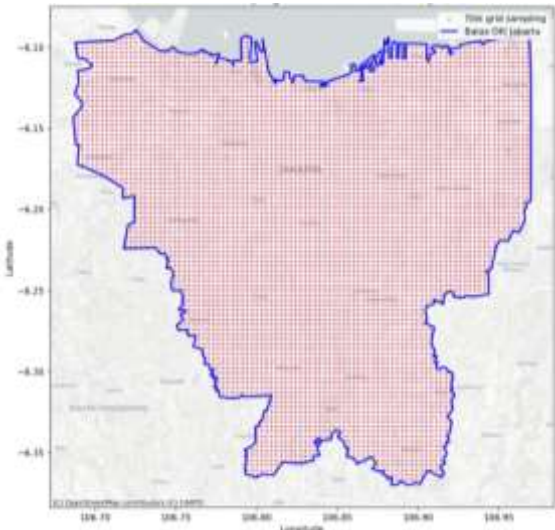
Di samping itu, perbedaan konseptual antara SUHI dan perubahan termal intra-urban juga perlu diperhatikan. Zhou *et al.* (2019) menjelaskan SUHI sebagai kontras LST antara permukaan urban dan non-urban (rural) yang dipadankan secara temporal dan atmosferik. UTFVI, yang secara umum dirumuskan sebagai $(T_s - T_m)/T_m$, juga banyak digunakan dalam studi Jawa dan Jakarta. Namun, ketika T_m dihitung sebagai rata-rata area kajian, ukuran tersebut secara ketat menggambarkan variasi relatif intra-urban, bukan kontras urban-rural absolut. Oleh karena itu, penelitian ini membedakan LST, anomali termal relatif intra-urban, UTFVI, dan SUHI absolut secara konseptual. Studi mengenai kondisi termal perkotaan di Jakarta telah memanfaatkan LST dan SUHI untuk menggambarkan pola spasial serta intensitas panas perkotaan (Fajary *et al.*, 2024; Muzaky & Jaelani, 2019; Siswanto *et al.*, 2023). Namun, masih terdapat ruang untuk mengembangkan pendekatan prediktif yang membandingkan Random Forest dan XGBoost secara sistematis pada dataset, prediktor, dan skema pengujian yang sama. Selain membandingkan metrik performa agregat, evaluasi dapat diperluas melalui pengujian statistik berpasangan terhadap error dan analisis korelasi residual untuk menilai apakah kombinasi kedua model memberikan manfaat melalui diversifikasi error. Pada sisi interpretabilitas, penelitian ini tidak hanya menggunakan SHAP untuk menentukan kepentingan prediktor, tetapi juga menganalisis hubungan nonlinier melalui threshold dan interaksi antarvariabel. Secara khusus, interaksi antara intensitas permukaan terbangun (NDBI) dan kepadatan jalan dianalisis untuk mengidentifikasi apakah kedua faktor tersebut memberikan pola kontribusi yang saling berkaitan dalam menjelaskan variasi LST. Dengan demikian, penelitian ini mengintegrasikan evaluasi performa model, diversifikasi error, dan interpretasi XAI dalam satu kerangka analisis, sekaligus mempertahankan pembedaan antara LST, anomali termal relatif intra-urban, UTFVI, dan SUHI absolut. Penelitian ini bertujuan untuk menguji dan membandingkan performa Random Forest dan XGBoost melalui hyperparameter tuning pada dataset dan pengujian yang sama. Dari sisi kombinasi model, penelitian ini mengevaluasi diversifikasi error RF dan XGBoost untuk menentukan apakah ensemble dapat meningkatkan performa prediksi. Selain itu, penelitian ini mengidentifikasi faktor spektral dan morfologi yang paling dominan dalam menjelaskan variasi LST di Jakarta, termasuk threshold dan interaksi antarvariabel, untuk memahami pola termal dan faktor yang berkontribusi terhadap heterogenitas spasial antarwilayah di Jakarta.

2. Data dan Studi Area

Studi area yang digunakan adalah DKI Jakarta (106,68–106,97°BT dan –6,08 hingga –6,37°LS). Akuisisi data citra satelit dan data vektor dilakukan melalui pemilihan batas administrasi tingkat provinsi dari GADM dan intersect dengan *bounding box* citra satelit, serta mengecualikan Kepulauan Seribu yang secara morfologis tidak relevan dengan UHI perkotaan daratan. Data satelit diakses melalui STAC API Microsoft Planetary Computer menggunakan 21 *scene* Landsat 8/9 dan 51 *scene* Sentinel-2 pada rentang Juni–September 2024. Seluruh *scene* digabungkan melalui *median mosaicking* untuk memastikan cakupan spasial dapat memenuhi kebutuhan data serta mengatasi kelemahan pendekatan satu *scene* tunggal yang rentan terhadap tutupan awan. OpenStreetMap digunakan sebagai sumber data terbuka untuk dua proksi morfologi, yaitu jarak titik grid ke RTH terdekat menggunakan fungsi berbasis poligon dan kepadatan jaringan jalan dalam radius *buffer* 250 m. Lingkungan komputasi penelitian menggunakan Google Colab dengan pustaka *pystac-client*, *planetary-computer*, *rioxarray*, *geopandas*, *scikit-learn*, *xgboost*, *shap*, *osmnx*, *contextily*, dan *seaborn*. Selanjutnya, *preprocessing* dilakukan untuk menangani tutupan awan dan piksel rusak, menggabungkan beberapa *scene* menjadi satu mosaik, serta menyamakan sistem koordinat seluruh citra. *Preprocessing* band Sentinel-2 dilakukan melalui *masking*, *median mosaicking* multi-*scene*, dan reproyeksi ke EPSG:4326 sebagai dasar perhitungan indeks spektral pada tahap berikutnya. *Preprocessing* band termal Landsat dilakukan melalui *masking* piksel, *median mosaicking* multi-*scene*, konversi radiometrik ke Celsius, dan reproyeksi ke EPSG:4326 sehingga menghasilkan lapisan LST. Band termal Landsat Collection 2 Level-2 dikonversi dengan emisivitas berbasis ASTER Global Emissivity Dataset, sedangkan piksel *nodata* (DN = 0) dieksklusi sebelum agregasi *median* multi-*scene*.

$$LST\ (^{\circ}C) = [DN \times 0,00341802 + 149,0] - 273,15 \tag{1}$$

Grid sistematis dengan interval ≈ 300 m dibentuk dan di-*clip* terhadap area kajian. Setelah penghapusan observasi tidak valid (*nodata* pada salah satu variabel), diperoleh 5.821 titik. Kepadatan jalan dihitung sebagai total panjang segmen jalan dalam *buffer* lingkaran 250 m dibagi luas *buffer* (satuan m/m²), sedangkan jarak ke RTH dihitung sebagai jarak Euclidean minimum ke poligon RTH terdekat dalam proyeksi UTM 48S untuk memperoleh akurasi jarak metrik. Distribusi grid sampling dan batas administrasi yang digunakan dalam penelitian ditunjukkan pada Gambar 1.



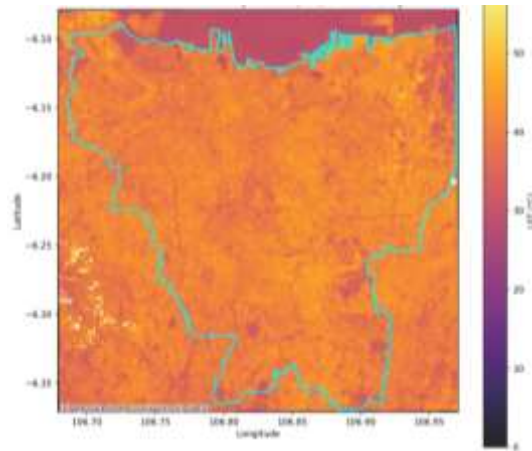
Gambar 1. Studi area, grid sampling 300 m, dan batas administrasi DKI Jakarta.

Sumber data yang digunakan dalam penelitian meliputi data penginderaan jauh dan data vektor. Rincian sumber, resolusi, dan fungsi masing-masing data disajikan pada Tabel 1.

Tabel 1. Sumber Data Penelitian			
Data	Sumber	Resolusi	Keterangan
LST	Landsat Collection 2 Level-2 (Microsoft Planetary Computer STAC API)	30 m	Variabel target

NDVI, NDBI, NDWI	Sentinel-2 Level-2A	10–20 m	Prediktor spektral
Jarak ke RTH, Kepadatan Jalan	OpenStreetMap (OSMnx)	Vektor	Prediktor morfologi
Batas Administrasi	GADM level 1 & 2	Vektor	Area kajian & breakdown wilayah

Hasil mosaik LST yang digunakan sebagai variabel target dalam penelitian ditunjukkan pada Gambar 2. Mosaik tersebut memberikan gambaran distribusi spasial suhu permukaan di wilayah studi setelah proses *preprocessing* dan penggabungan data multi-*scene*. Informasi spasial LST ini selanjutnya digunakan sebagai dasar dalam pembentukan dataset dan analisis prediksi menggunakan variabel prediktor yang telah ditentukan.



Gambar 2. Mosaik Land Surface Temperature (LST).

3. Metode

Dalam tahapan *Feature Engineering*, dilakukan beberapa proses untuk menurunkan variabel prediktor berupa indeks spektral dan variabel morfologi urban dari citra, kemudian mengonversinya ke titik *grid* agar dapat digunakan sebagai tabel data untuk model *machine learning*. Proses *grid sampling* dan ekstraksi fitur dilakukan dengan membentuk grid sistematis berjarak ≈ 300 m yang menghasilkan 5.821 titik valid, masing-masing berisi nilai LST dan ketiga indeks spektral. Band SWIR (B11) memiliki resolusi *native* 20 m sehingga dilakukan penyesuaian resolusi (*resampling*) sebelum perhitungan NDBI. Secara fisis, NDVI memanfaatkan penyerapan klorofil pada band merah dan pantulan tinggi mesofil daun pada NIR (Rouse *et al.*, 1974). NDBI memanfaatkan pantulan material terbangun yang relatif tinggi pada SWIR dibandingkan NIR (Zha *et al.*, 2003), sedangkan NDWI memanfaatkan perbedaan karakteristik pantulan dan penyerapan air pada band hijau dan NIR (McFeeters, 1996). Perhitungan indeks spektral NDVI, NDBI, dan NDWI dari band Sentinel-2 dilakukan menggunakan persamaan berikut:

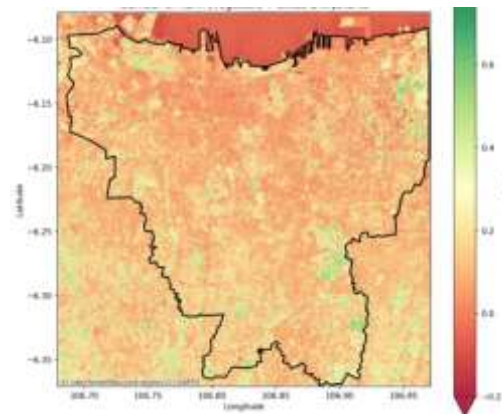
$$NDVI = (NIR - Red) / (NIR + Red) \quad (2)$$

$$NDBI = (SWIR - NIR) / (SWIR + NIR) \quad (3)$$

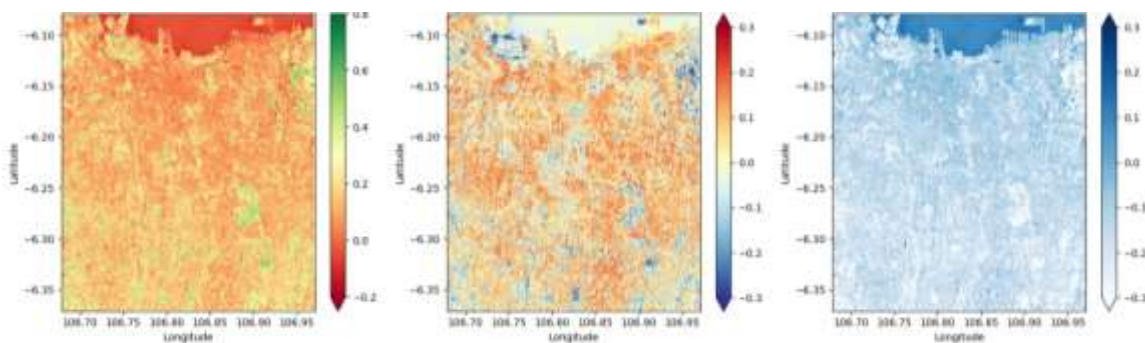
$$NDWI = (Green - NIR) / (Green + NIR) \quad (4)$$

Hasil perhitungan indeks spektral ditunjukkan pada Gambar 3 dan Gambar 3b. Gambar 3 menunjukkan mosaik NDVI, sedangkan Gambar 3b memperlihatkan perbandingan spasial NDVI, NDBI, dan NDWI. Ketiga indeks menunjukkan pola spasial yang berbeda dan saling melengkapi dalam merepresentasikan karakteristik tutupan lahan. NDVI dengan nilai tinggi (hijau) terkonsentrasi pada RTH dan wilayah selatan, sedangkan NDBI dengan nilai tinggi (merah pada panel tengah Gambar 3b) terkonsentrasi pada kawasan padat bangunan di bagian pusat-barat kota. Perbedaan pola tersebut menunjukkan pemisahan spasial yang jelas antara karakteristik vegetasi dan area terbangun. NDWI pada panel kanan Gambar 3b didominasi nilai

negatif di sebagian besar area kajian, yang menunjukkan terbatasnya area badan air terbuka yang terdeteksi pada resolusi 10 m. Pola tersebut menjadi informasi yang relevan dalam interpretasi hubungan NDWI dengan LST.



Gambar 3. Mosaik NDVI Jakarta dari Sentinel-2, di-*overlay* dengan *basemap* dan batas administrasi



Gambar 3b. Perbandingan spasial NDVI, NDBI, dan NDWI yang menunjukkan pola vegetasi dibandingkan dengan area terbangun.

Tahapan *Data Split* dilakukan untuk memisahkan data menjadi bagian latih dan bagian uji, sehingga performa model dapat diukur menggunakan data yang belum pernah dilihat model selama proses pelatihan. Data dibagi menjadi 80% data latih dan 20% data uji menggunakan *random_state* = 42. Selanjutnya, *Hyperparameter Tuning* dilakukan secara terpisah untuk Random Forest (RF) dan XGBoost menggunakan *RandomizedSearchCV* dengan *5-fold cross-validation* dan 25 kombinasi *hyperparameter* acak. Optimasi dilakukan terhadap skor R^2 untuk memperoleh konfigurasi *hyperparameter* terbaik pada masing-masing model. Hasil *Hyperparameter Tuning* menunjukkan bahwa Random Forest menggunakan 800 *tree* dengan kedalaman 10 serta regularisasi tambahan berupa *min_samples_split* = 10 dan *min_samples_leaf* = 2 untuk membantu mencegah *overfitting*. XGBoost memilih konfigurasi dengan kedalaman yang lebih dangkal (*max_depth* = 3), *learning rate* rendah sebesar 0,01, dan 800 *estimator* sebagai strategi *gradient boosting* untuk memperoleh generalisasi yang optimal pada data bervolume menengah. Rincian *hyperparameter* terbaik disajikan pada Tabel 2.

Tabel 2. Hyperparameter terbaik hasil *RandomizedSearchCV*

Model	n_estimators	max_depth	Parameter Lain	Best CV R2 (saat tuning)
Random Forest	800	10	min_samples_split=10, min_samples_leaf=2, max_features=sqrt	0.445
XGBoost	800	3	learning_rate=0.01, subsample=0.6, colsample_bytree=0.9	0.448

Setelah proses *tuning*, model final Random Forest dan XGBoost dilatih menggunakan *hyperparameter* terbaik yang diperoleh dari proses pencarian. Random Forest menerapkan mekanisme *bagging*, sedangkan XGBoost menerapkan mekanisme *boosting*. Kedua model kemudian digunakan untuk menghasilkan prediksi pada data uji. *Ensemble RF-XGBoost Modeling* dilakukan untuk menguji apakah penggabungan prediksi Random Forest dan XGBoost dapat meningkatkan akurasi dibandingkan model individual. Ensemble menggunakan dua pendekatan, yaitu *simple averaging* dan *weighted averaging*:

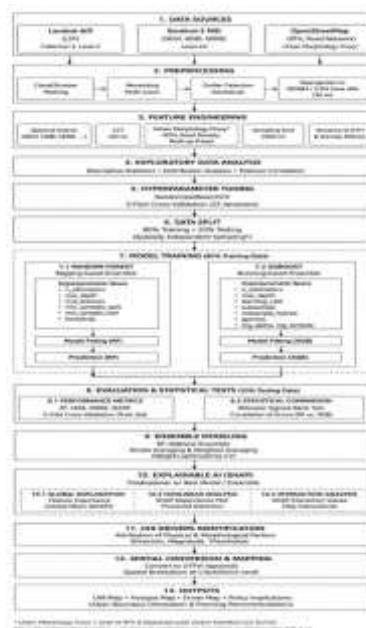
$$\hat{y}_e = (\hat{y}_{RF} + \hat{y}_{XGB}) / 2 \quad (5)$$

$$w_i = (1/RMSE_i) / \sum (1/RMSE_i) \hat{y}_e, w = w_{RF} \hat{y}_{RF} + w_{XGB} \hat{y}_{XGB} \text{ (weighted averaging)} \quad (6)$$

Bobot pada *weighted averaging* ditentukan secara berbanding terbalik dengan RMSE. Korelasi residual antar-model digunakan sebagai dasar kuantitatif untuk menjelaskan keberhasilan atau kegagalan ensemble dalam meningkatkan performa. *Explainable Artificial Intelligence* (SHAP) digunakan untuk menjelaskan kontribusi setiap variabel terhadap prediksi LST secara global maupun lokal sehingga model dapat diinterpretasikan. *TreeExplainer* diterapkan pada XGBoost sebagai model dengan performa *test-set* terbaik. SHAP menguraikan setiap prediksi menjadi nilai ekspektasi dan kontribusi masing-masing fitur:

$$f(x) = E[f(X)] + \sum \phi_j \quad (7)$$

Interpretasi SHAP secara global dan lokal dilakukan menggunakan *summary plot*, *dependence plot*, dan *waterfall plot* pada satu titik contoh. Selanjutnya, perhitungan *SHAP interaction values* dilakukan untuk mengidentifikasi pasangan fitur dengan efek gabungan terkuat. Deteksi *threshold* berbasis data pada hubungan setiap fitur dengan nilai SHAP dilakukan menggunakan *decision tree* yang memodelkan nilai fitur terhadap nilai SHAP-nya. *UHI Drivers Identification* dilakukan berdasarkan hasil SHAP untuk merangkum arah, besaran, dan ambang batas pengaruh setiap faktor sebagai *driver* UHI. Selanjutnya, dilakukan *Spatial Conversion & Mapping* untuk mengonversi LST menjadi indeks yang lebih mudah diinterpretasikan sebagai ukuran tekanan panas, yaitu deviasi termal dan UTFVI, serta untuk menguji variasi pola tersebut antarwilayah di Kota Jakarta. Evaluasi performa dilakukan pada data uji menggunakan R^2 , MAE, dan RMSE. Analisis residual dilakukan menggunakan histogram dan *scatter plot* residual terhadap prediksi untuk memeriksa ada atau tidaknya bias sistematis. Pemetaan spasial residual digunakan untuk mengidentifikasi lokasi geografis tempat model cenderung menghasilkan kesalahan. Uji Wilcoxon *signed-rank* terhadap error absolut kedua model digunakan untuk menguji signifikansi statistik perbedaan performa. Selain akurasi, waktu komputasi yang mencakup proses *tuning* dan *training* dibandingkan sebagai pertimbangan efisiensi model.



Gambar 4. Diagram Alir Metodologi.

Karena penelitian ini tidak memiliki referensi rural secara temporal dan atmosferik pada setiap *scene*, penelitian ini secara eksplisit membedakan dua kelompok metrik. Anomali LST dihitung sebagai berikut:

$$\text{Anomali } LST_i = LST_i - LST_{\text{Jakarta}} \quad (8)$$

Sedangkan UTFVI dihitung menggunakan persamaan:

$$UTFVI_i = (LST_i - LST_{\text{Jakarta}}) / LST_{\text{Jakarta}} \quad (9)$$

Nilai anomali LST kemudian diklasifikasikan ke dalam enam kelas menggunakan batas -1 , 0 , 1 , 2 , dan 3 °C dari rata-rata. Batas -1 °C dipilih sebagai keputusan analitis untuk klasifikasi yang mencerminkan tekanan panas kota padat secara lebih realistis dibandingkan *split* matematis 50:50 menggunakan nilai rata-rata.

4. Hasil dan Pembahasan

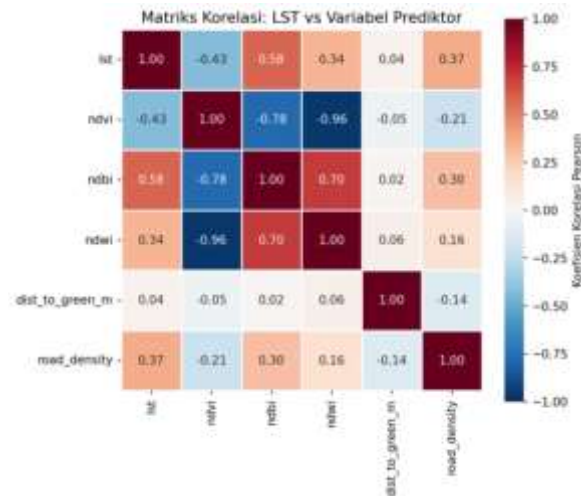
4.1 Hasil

Sebanyak 5.821 dataset diobservasi dalam penelitian ini. Hasil statistik deskriptif menunjukkan bahwa Land Surface Temperature (LST) memiliki nilai rata-rata sebesar 40,77°C dengan standar deviasi 2,65°C, serta rentang nilai 30,20–51,32°C. Selisih antara nilai minimum dan maksimum mencapai 21,12°C, yang menunjukkan adanya heterogenitas suhu permukaan yang cukup besar di wilayah perkotaan yang dikaji. Statistik deskriptif seluruh variabel penelitian disajikan pada Tabel 3.

Tabel 3. Statistik deskriptif variabel penelitian

Variabel	N	Mean	Std Dev	Min	Median	Max
LST (°C)	5821	40.77	2.65	30.20	41.05	51.32
NDVI	5821	0.150	0.130	-0.082	0.104	0.627
NDBI	5821	0.035	0.108	-0.367	0.051	0.319
NDWI	5821	-0.164	0.110	-0.550	-0.140	0.210
Jarak ke RTH (m)	5821	377.75	296.00	0.00	316.25	1873.92
Kepadatan Jalan (m/m ²)	5821	0.024	0.013	0.000	0.024	0.076

Hasil matriks korelasi Pearson menunjukkan bahwa NDBI memiliki korelasi positif terkuat dengan LST, yaitu sebesar $r = 0,58$. Hubungan tersebut menunjukkan bahwa peningkatan intensitas permukaan terbangun berkaitan dengan peningkatan LST dalam dataset penelitian. Sebaliknya, NDVI memiliki korelasi negatif dengan LST sebesar $r = -0,43$, yang sejalan dengan kecenderungan vegetasi memberikan efek pendinginan melalui proses evapotranspirasi. NDWI menunjukkan korelasi positif dengan LST sebesar $r = 0,34$. Hubungan ini tampak berbeda dari ekspektasi umum mengenai peran air sebagai elemen pendingin. Namun, dalam konteks dataset penelitian ini, hubungan tersebut perlu dipahami dengan mempertimbangkan karakteristik spasial variabel dan hubungan antarprediktor. Korelasi NDVI dan NDWI yang sangat kuat dan negatif ($r = -0,96$) menunjukkan adanya hubungan yang erat antara kedua indeks tersebut, sehingga interpretasi NDWI terhadap LST perlu dilakukan secara hati-hati dan tidak semata-mata dianggap sebagai representasi keberadaan badan air. Secara lebih spesifik, korelasi NDVI–NDWI yang tinggi mengindikasikan bahwa kedua indeks tersebut menangkap variasi kondisi permukaan yang saling bertaut (misalnya perubahan komposisi vegetasi dan karakter spektral permukaan yang turut memengaruhi respons NDWI). Dengan demikian, pola korelasi NDWI yang tampak positif terhadap LST kemungkinan tidak hanya mencerminkan keberadaan air terbuka, tetapi juga dapat merefleksikan keterkaitan spektral dengan vegetasi dan/atau tutupan lahan lain yang berasosiasi dengan variasi NDVI. Implikasi dari temuan ini adalah evaluasi kontribusi setiap fitur tidak cukup hanya berdasarkan korelasi linear, tetapi perlu dilengkapi dengan analisis model berbasis nonlinier serta interpretasi menggunakan XAI (SHAP) agar kontribusi tiap fitur terhadap prediksi LST dapat dipahami lebih terarah.



Gambar 5. Matriks korelasi Pearson antar-variabel penelitian

Hasil 5-Fold Cross-Validation setelah proses *hyperparameter tuning* menunjukkan bahwa XGBoost memiliki kinerja sedikit lebih baik dibandingkan Random Forest. Nilai rata-rata R^2 XGBoost sebesar $0,460 \pm 0,035$, sedangkan Random Forest sebesar $0,454 \pm 0,032$. Pada metrik RMSE, XGBoost menghasilkan $1,944 \pm 0,089^\circ\text{C}$, lebih rendah dibandingkan Random Forest sebesar $1,954 \pm 0,080^\circ\text{C}$. Demikian pula, nilai MAE XGBoost sebesar $1,511 \pm 0,054^\circ\text{C}$ sedikit lebih rendah daripada Random Forest sebesar $1,518 \pm 0,052^\circ\text{C}$. Hasil tersebut menunjukkan bahwa berdasarkan validasi silang, XGBoost memberikan performa rata-rata yang sedikit lebih baik, meskipun selisih antarmodel relatif kecil.

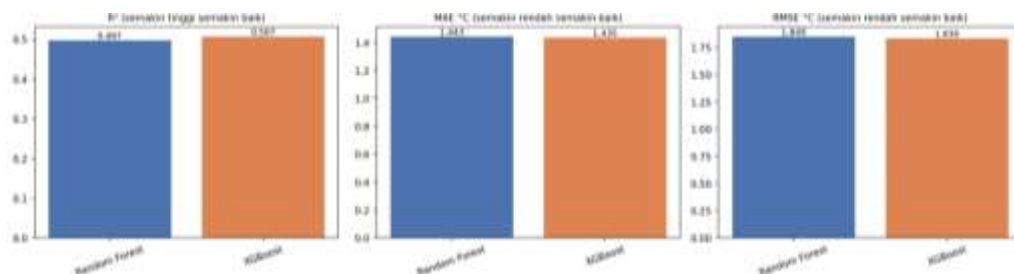
Tabel 4. Hasil 5-Fold Cross-Validation (hyperparameter ter-tuning)

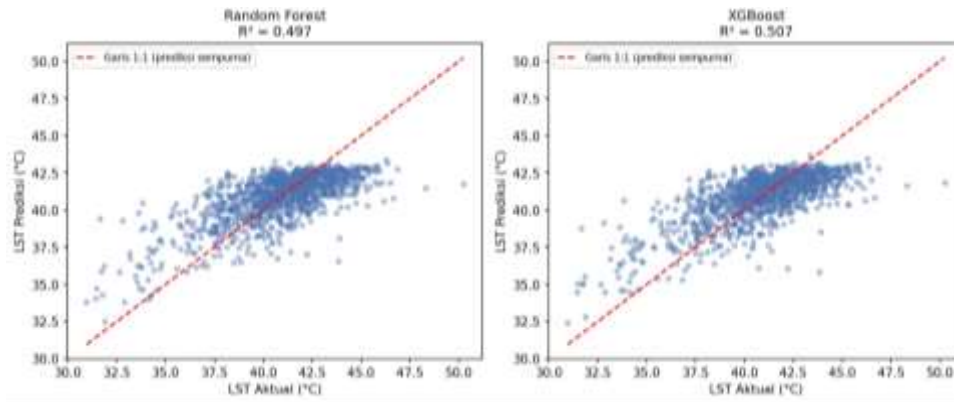
Model	R2 (mean+-std)	RMSE (mean+-std)	MAE (mean+-std)
Random Forest	0.454 +- 0.032	1.954 +- 0.080	1.518 +- 0.052
XGBoost	0.460 +- 0.035	1.944 +- 0.089	1.511 +- 0.054

Pada data uji dengan menggunakan *split* tunggal, XGBoost kembali menunjukkan kinerja yang sedikit lebih baik dibandingkan Random Forest. XGBoost memperoleh R^2 sebesar 0,507, MAE sebesar $1,435^\circ\text{C}$, dan RMSE sebesar $1,830^\circ\text{C}$. Sementara itu, Random Forest memperoleh R^2 sebesar 0,497, MAE sebesar $1,443^\circ\text{C}$, dan RMSE sebesar $1,849^\circ\text{C}$. Perbandingan tersebut menunjukkan bahwa XGBoost menghasilkan nilai R^2 lebih tinggi serta nilai MAE dan RMSE lebih rendah dibandingkan Random Forest pada data uji.

Tabel 5. Performa model pada data uji (split tunggal)

Model	R2	MAE (C)	RMSE (C)
Random Forest	0.497	1.443	1.849
XGBoost	0.507	1.435	1.830

Gambar 6. Perbandingan R^2 , MAE, dan RMSE RF vs. XGBoost pada data uji (pasca-tuning)



Gambar 7. Hubungan LST aktual dan LST prediksi untuk kedua model

Untuk mengetahui apakah perbedaan performa kedua model bersifat signifikan secara statistik, dilakukan uji Wilcoxon *signed-rank* terhadap error absolut pada observasi uji yang sama. Hasil pengujian pada Tabel 6 menghasilkan statistik sebesar 328.913 dan $p = 0,352$. Dengan $\alpha = 0,05$, nilai tersebut menunjukkan bahwa tidak terdapat bukti yang cukup untuk menyatakan adanya perbedaan distribusi error absolut yang signifikan antara Random Forest dan XGBoost. Dengan demikian, meskipun XGBoost menghasilkan performa numerik yang sedikit lebih baik pada Tabel 5, perbedaan tersebut belum terbukti signifikan secara statistik. Sebagai pembandingan, dokumentasi eksperimen sebelum proses *hyperparameter tuning* menunjukkan nilai $p \approx 0,0003$ dengan Random Forest memiliki performa yang lebih baik secara signifikan. Perbedaan antara hasil sebelum dan sesudah *tuning* menunjukkan bahwa proses optimasi *hyperparameter* dapat memengaruhi perbandingan relatif kedua model. Oleh karena itu, hasil pasca-*tuning* digunakan sebagai dasar evaluasi utama dalam penelitian ini.

Tabel 6. Uji Wilcoxon signed-rank terhadap error absolut

Model A	Model B	Statistik	p-value	Kesimpulan
Random Forest	XGBoost	328913.0	0.352	TIDAK Signifikan ($p > 0.05$)

Pengujian komparatif model menggunakan observasi uji yang identik untuk Random Forest dan XGBoost. Pendekatan ini memungkinkan error kedua model dibandingkan secara berpasangan, sehingga evaluasi tidak hanya didasarkan pada perbedaan nilai R^2 , MAE, atau RMSE dari satu *split* data. Hasil uji Wilcoxon menjadi dasar statistik untuk menilai apakah perbedaan performa kedua model cukup kuat untuk dianggap berbeda secara signifikan. Analisis ensemble dilakukan untuk mengevaluasi apakah penggabungan Random Forest dan XGBoost dapat memberikan peningkatan kinerja melalui diversitas error. Kelayakan ensemble tidak hanya ditentukan oleh apakah kedua algoritma memiliki performa individual yang baik, tetapi juga oleh sejauh mana kedua model menghasilkan kesalahan yang berbeda. Pada penelitian ini, korelasi residual antara Random Forest dan XGBoost sebesar $r_{\text{error}} = 0,988$, yang mendekati nilai maksimum $r = 1$. Nilai tersebut menunjukkan bahwa residual kedua model sangat berkorelasi dan cenderung memiliki pola kesalahan yang sama. Secara matematis, varians error ensemble dapat dinyatakan sebagai:

$$\text{Var}(w_1 e_1 + w_2 e_2) = w_1^2 \text{Var}(e_1) + w_2^2 \text{Var}(e_2) + 2w_1 w_2 \text{Cov}(e_1, e_2) \quad (10)$$

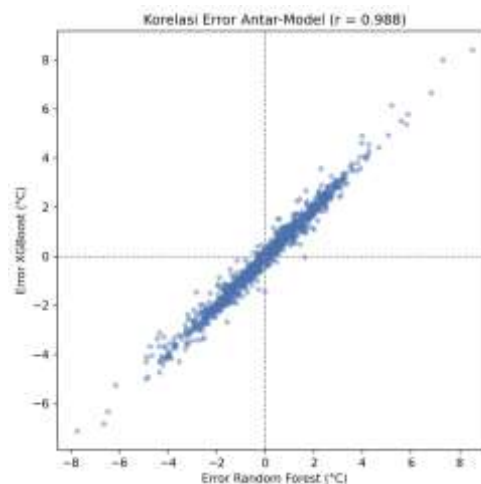
Ketika residual kedua model memiliki korelasi positif yang sangat tinggi, kovarians menjadi besar sehingga potensi pengurangan varians melalui penggabungan prediksi menjadi terbatas. Hasil empiris menunjukkan bahwa *simple averaging* menghasilkan RMSE sebesar $1,834^\circ\text{C}$, sedangkan *weighted averaging* juga menghasilkan RMSE sebesar $1,834^\circ\text{C}$. Kedua nilai tersebut sedikit lebih tinggi dibandingkan XGBoost individual yang menghasilkan RMSE sebesar $1,830^\circ\text{C}$. Bobot *weighted averaging* yang diperoleh adalah 0,497 untuk Random Forest dan 0,503 untuk XGBoost, yang menunjukkan bahwa kedua model memperoleh bobot yang hampir seimbang karena performa individual keduanya relatif berdekatan. Karena Random Forest dan XGBoost menghasilkan prediksi pada titik uji yang identik, error absolut kedua model dibandingkan secara berpasangan menggunakan uji Wilcoxon *signed-rank*. Korelasi residual sebagai indikator diversitas error dihitung dengan:

$$r_e = \text{corr}(e_{\text{RF}}, e_{\text{XGB}}) \quad (11)$$

Korelasi residual sebesar $r = 0,988$ menunjukkan bahwa kedua model cenderung menghasilkan kesalahan pada lokasi yang sama. Dengan demikian, kedua model memiliki tingkat diversitas error yang rendah sehingga manfaat penggabungan prediksi menjadi terbatas. Secara teoritis, untuk dua error dengan varians yang hampir sama dan bobot yang mendekati 0,5:0,5, reduksi simpangan baku ensemble terhadap model individual terbaik hanya sekitar $-0,3\%$ pada $r = 0,988$. Hasil tersebut konsisten dengan perbedaan RMSE empiris, yaitu $1,834^{\circ}\text{C}$ pada ensemble dibandingkan $1,830^{\circ}\text{C}$ pada XGBoost individual. Korelasi prediksi kedua model juga tinggi, yaitu $r = 0,986$, sedangkan korelasi error mencapai $r = 0,988$. Hal tersebut mengindikasikan bahwa Random Forest dan XGBoost menangkap pola spasial yang sangat serupa dan cenderung kurang saling melengkapi dalam konteks dataset penelitian ini.

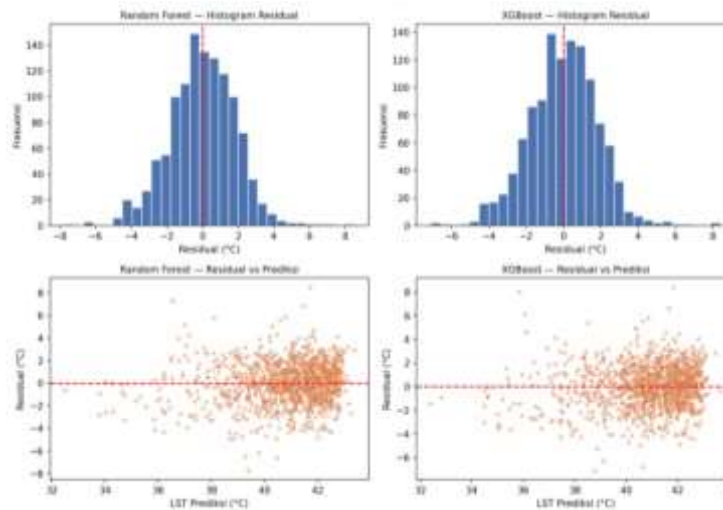
Tabel 7. Kinerja ensemble RF–XGBoost

Model	R2	MAE	RMSE
Random Forest	0.497	1.443	1.849
XGBoost	0.507	1.435	1.830
Ensemble (Simple Avg)	0.505	1.436	1.834
Ensemble (Weighted Avg)	0.505	1.436	1.834



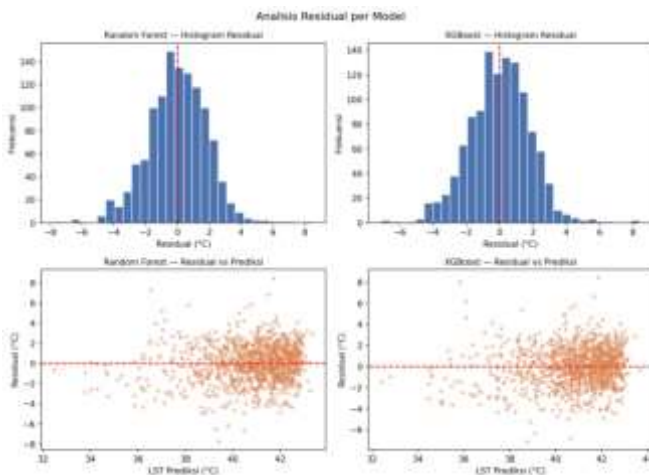
Gambar 8. Korelasi residual (error) RF dan XGBoost pada data uji

Analisis residual tidak cukup dievaluasi hanya berdasarkan ukuran agregat, tetapi juga perlu dilihat berdasarkan distribusi dan pola spasialnya. Dengan mempertimbangkan korelasi residual yang sangat tinggi pada kedua model, evaluasi residual lanjutan menjadi penting untuk memastikan bahwa kesamaan pola error tersebut juga tercermin pada sebaran residual dan karakteristik error relatif terhadap nilai prediksi. Histogram residual pada Gambar 9 menunjukkan bahwa residual kedua model relatif simetris dan berpusat di sekitar nol, sehingga tidak terlihat adanya bias sistematis yang besar. Selanjutnya, plot residual terhadap nilai prediksi juga tidak menunjukkan pola corong yang jelas, sehingga tidak terdapat indikasi visual heteroskedastisitas yang kuat. Ketiadaan pola corong tersebut mengindikasikan varians error relatif stabil pada berbagai tingkat nilai prediksi, yang berarti ketidakpastian model tidak meningkat secara tajam ketika prediksi berada pada rentang yang lebih tinggi atau lebih rendah. Dengan demikian, evaluasi residual menguatkan bahwa kualitas kesalahan model lebih bersifat acak (random error) dan tidak didominasi oleh gangguan dengan karakter tertentu pada skala prediksi tertentu. Temuan ini memperkuat kesimpulan bahwa perbedaan performa yang kecil antara RF dan XGBoost bukan berasal dari adanya ketidakaturan error, melainkan lebih dipengaruhi oleh kesamaan pola yang ditangkap oleh keduanya.

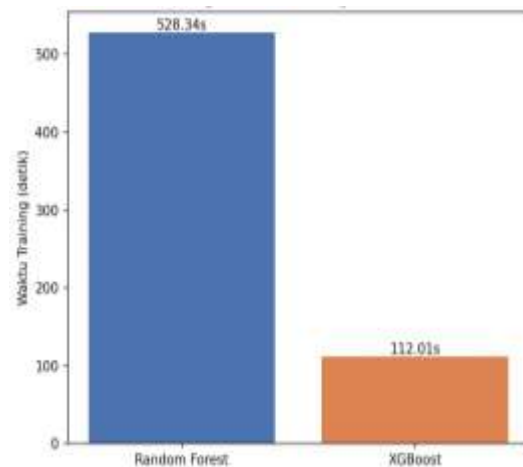


Gambar 9. Distribusi residual (histogram) dan residual terhadap nilai prediksi untuk RF dan XGBoost

Peta residual digunakan untuk mengidentifikasi lokasi *under-prediction* (residual positif) dan *over-prediction* (residual negatif) secara geografis. Distribusi spasial residual Random Forest dan XGBoost pada titik uji menunjukkan pola sebaran yang sangat mirip antarmodel. Kesamaan pola tersebut menjadi bukti visual yang mendukung tingginya korelasi residual sebesar $r = 0,988$. Kedua model menunjukkan pola sebaran error yang serupa, dengan beberapa klaster *over-prediction* pada area transisi tutupan lahan, seperti tepi RTH-permukiman dan tepi sungai atau kanal. Kondisi tersebut diduga berkaitan dengan karakteristik heterogenitas tutupan lahan yang lebih sulit direpresentasikan pada resolusi grid 300 m. Klaster residual secara spasial juga menunjukkan bahwa sebagian variasi LST belum sepenuhnya terepresentasikan oleh lima prediktor yang digunakan dalam penelitian ini. Oleh karena itu, kemungkinan adanya autokorelasi spasial pada residual perlu diuji lebih lanjut.



Gambar 10. Distribusi spasial residual RF dan XGBoost



Gambar 10b. Perbandingan waktu komputasi (tuning + training) RF vs. XGBoost

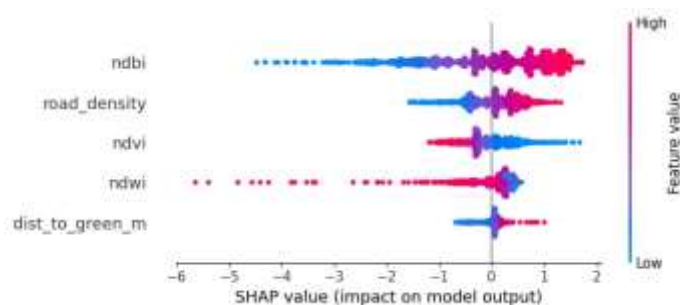
Perbandingan waktu komputasi *tuning* dan *training* Random Forest dan XGBoost juga digunakan sebagai pertimbangan efisiensi selain aspek akurasi. Untuk mengidentifikasi kontribusi masing-masing prediktor terhadap prediksi LST, dilakukan analisis SHAP (*SH*apley *A*dditive *eX*planations) pada model XGBoost terpilih. Hasil *mean absolute SHAP value* pada Tabel 8 menunjukkan bahwa NDBI memiliki nilai tertinggi, yaitu 0,986, diikuti oleh Kepadatan Jalan sebesar 0,408, NDVI sebesar 0,362, NDWI sebesar 0,361, dan Jarak ke RTH sebesar 0,152.

Tabel 8. Mean absolute SHAP value pada model XGBoost terpilih

Fitur	Mean SHAP
NDBI	0.986

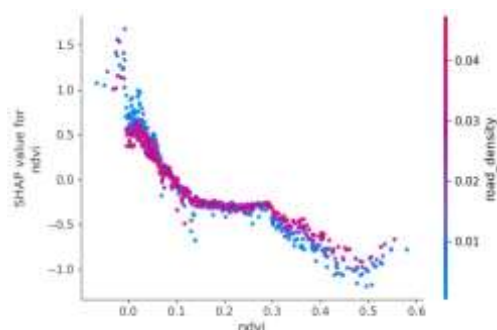
Kepadatan Jalan	0.408
NDVI	0.362
NDWI	0.361
Jarak ke RTH	0.152

NDBI memiliki nilai *mean absolute SHAP* sebesar 0,986, jauh lebih tinggi dibandingkan prediktor lainnya. Dalam skala output LST (°C), nilai tersebut menunjukkan bahwa kontribusi absolut NDBI terhadap variasi prediksi model secara rata-rata mencapai hampir 1°C. Dengan demikian, NDBI merupakan fitur dengan kontribusi individual terbesar dalam model XGBoost. Dominasi tersebut menunjukkan bahwa intensitas permukaan terbangun merupakan prediktor paling informatif bagi variasi LST dalam dataset penelitian ini.



Gambar 11. SHAP summary plot model XGBoost teroptimasi. Setiap titik mewakili satu observasi, warna menunjukkan nilai fitur (merah = tinggi, biru = rendah), sedangkan posisi horizontal menunjukkan arah dan besar kontribusi terhadap prediksi LST.

Hubungan nonlinear antara fitur dan prediksi LST dianalisis melalui *SHAP dependence plot*. Pada Gambar 12, hubungan antara NDVI dan nilai SHAP menunjukkan pola nonlinear. Kontribusi NDVI terhadap peningkatan prediksi LST (SHAP positif) terutama terkonsentrasi pada nilai NDVI yang rendah. Seiring dengan meningkatnya NDVI, kontribusinya terhadap prediksi LST bergeser ke arah negatif yang menunjukkan kecenderungan efek pendinginan.



Gambar 12. SHAP dependence plot untuk NDVI

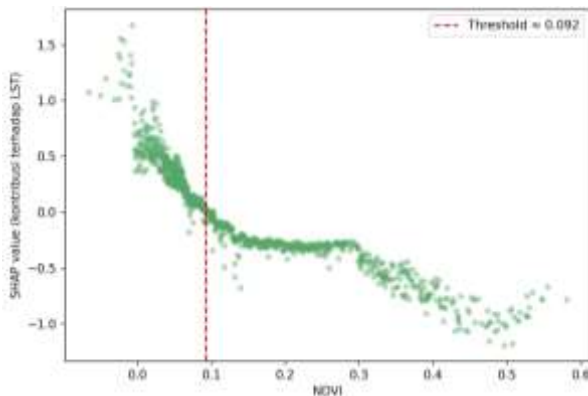
Breakpoint empiris dari hubungan fitur–SHAP ditunjukkan pada Tabel 9. Nilai breakpoint diperoleh dari *decision tree* dengan kedalaman 1 dan digunakan untuk menggambarkan titik perubahan empiris dalam hubungan antara fitur dan kontribusinya terhadap prediksi model.

Tabel 9. Breakpoint empiris dari hubungan fitur–SHAP

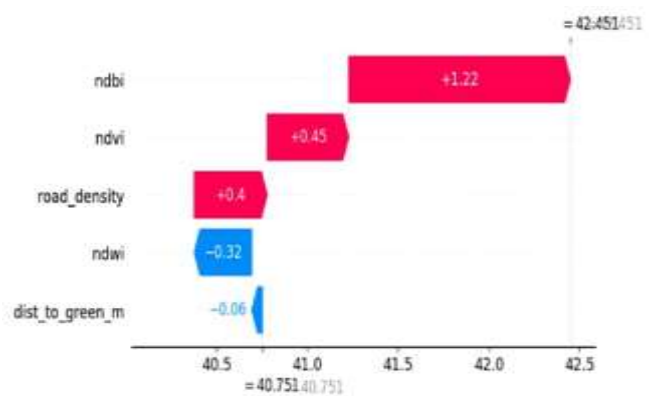
Fitur	Threshold (Breakpoint Empiris)
NDVI	0.092
NDBI	-0.018
NDWI	0.002
Jarak ke RTH (m)	91.62
Kepadatan Jalan	0.019

Breakpoint tersebut tidak dimaksudkan sebagai ambang batas universal, melainkan sebagai breakpoint empiris yang diperoleh dari dataset penelitian. Nilai NDVI $\approx 0,092$ berada sedikit di bawah median NDVI

area kajian, yaitu 0,104. Hal tersebut menunjukkan bahwa lebih dari separuh titik kajian berada pada nilai NDVI di sekitar atau di atas breakpoint tersebut. Temuan ini dapat menjadi indikasi awal yang relevan untuk analisis kebijakan penghijauan, tetapi penerapannya sebagai ambang kebijakan tetap memerlukan validasi lebih lanjut.



Gambar 13. Contoh breakpoint empiris hubungan NDVI dengan kontribusi SHAP, dengan garis threshold terdeteksi.



Gambar 13b. SHAP waterfall plot pada satu titik lokasi contoh yang merinci kontribusi tiap fitur dari nilai basis menuju prediksi akhir.

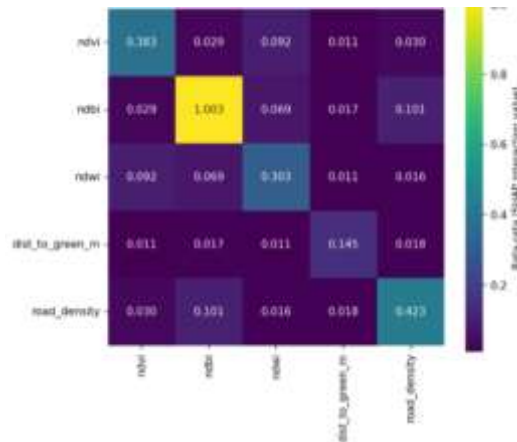
Interpretasi pada tingkat titik individual ditunjukkan melalui *SHAP waterfall plot* pada Gambar 13b. Plot tersebut memperlihatkan bagaimana setiap fitur memberikan kontribusi terhadap prediksi pada satu titik grid, dimulai dari nilai basis (*expected value*) dan kemudian mengalami pergeseran positif atau negatif akibat kontribusi masing-masing fitur hingga mencapai nilai prediksi akhir. Interaksi antar-faktor dianalisis menggunakan *mean absolute SHAP interaction values*. Hasil pada Tabel 10 menunjukkan bahwa nilai diagonal terbesar terdapat pada NDBI, yaitu 1,003, sedangkan untuk interaksi non-diagonal, nilai terbesar terdapat pada pasangan NDBI × Kepadatan Jalan sebesar 0,101. Nilai tersebut lebih tinggi dibandingkan interaksi NDBI × NDWI sebesar 0,069 dan NDVI × NDWI sebesar 0,092.

Tabel 10. Mean absolute SHAP interaction values

Matriks	NDVI	NDBI	NDWI	Jarak RTH	Kepadatan Jalan
NDVI	0.383	0.029	0.092	0.011	0.030
NDBI	0.029	1.003	0.069	0.017	0.101
NDWI	0.092	0.069	0.303	0.011	0.016
Jarak RTH	0.011	0.017	0.011	0.145	0.018
Kepadatan Jalan	0.030	0.101	0.016	0.018	0.423

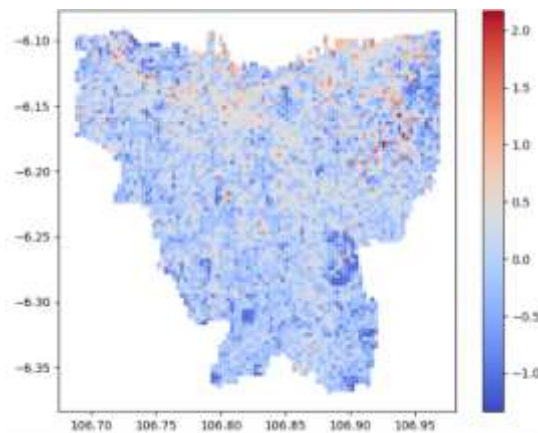
Hasil tersebut menunjukkan bahwa pasangan NDBI × Kepadatan Jalan merupakan interaksi non-diagonal terkuat dalam model. Dengan kata lain, kontribusi gabungan antara intensitas permukaan terbangun (yang direpresentasikan oleh NDBI) dan kerapatan infrastruktur jalan berperan lebih dominan dibanding pasangan fitur lainnya dalam menjelaskan variasi prediksi LST. Temuan ini menunjukkan adanya keterkaitan antara intensitas permukaan terbangun dan kepadatan jalan dalam menjelaskan variasi prediksi LST. Secara konseptual, kondisi tersebut selaras dengan karakteristik lingkungan perkotaan yang padat, di mana permukaan terbangun dan jaringan jalan dapat berkontribusi terhadap akumulasi panas melalui peningkatan penyerapan radiasi oleh material keras, sekaligus memodifikasi kondisi iklim perkotaan melalui konfigurasi "urban canyon". Konsep urban canyon yang dikemukakan oleh Oke (1982) relevan karena ruang jalan yang lebih rapat dan didominasi permukaan terbangun cenderung meningkatkan keterbatasan pertukaran panas dan menekan proses pendinginan yang efektif, sehingga respons termal permukaan menjadi lebih kuat dan terstruktur mengikuti kombinasi tutupan lahan serta infrastruktur jalan. Namun demikian, nilai SHAP interaction harus dipahami sebagai besarnya interaksi dalam kerangka model prediktif, yaitu seberapa jauh kontribusi satu fitur bergantung pada tingkat fitur lainnya untuk memengaruhi output prediksi. Dengan demikian, interpretasi nilai interaksi tidak dapat langsung disetarakan dengan pembuktian mekanisme fisik secara kausal bahwa efek gabungan kedua faktor melebihi jumlah efek individualnya di dunia nyata. Eksperimen pada penelitian ini menunjukkan bahwa setelah proses tuning, pasangan interaksi

yang paling menonjol dalam menjelaskan LST adalah NDBI $\times$ Kepadatan Jalan, sehingga kombinasi kedua faktor tersebut dapat diprioritaskan sebagai dasar pemaknaan driver termal dalam konteks perencanaan tata ruang dan mitigasi panas perkotaan.



Gambar 14. Matriks interaksi SHAP antar-faktor (pasca-tuning).

Pemetaan SHAP secara spasial digunakan untuk melihat variasi kontribusi fitur terhadap LST pada lokasi yang berbeda. Peta spasial kontribusi SHAP untuk NDVI menunjukkan adanya variasi geografis dalam kontribusi vegetasi terhadap prediksi LST. Efek pendinginan, yang ditunjukkan oleh nilai SHAP negatif, paling kuat terdapat pada titik-titik yang berdekatan dengan RTH berukuran besar. Sebaliknya, pada area dengan kepadatan bangunan yang tinggi, kontribusi NDVI cenderung mendekati nol, yang menunjukkan bahwa keberadaan vegetasi pada lokasi tersebut relatif terbatas untuk memberikan kontribusi termal yang besar menurut model.



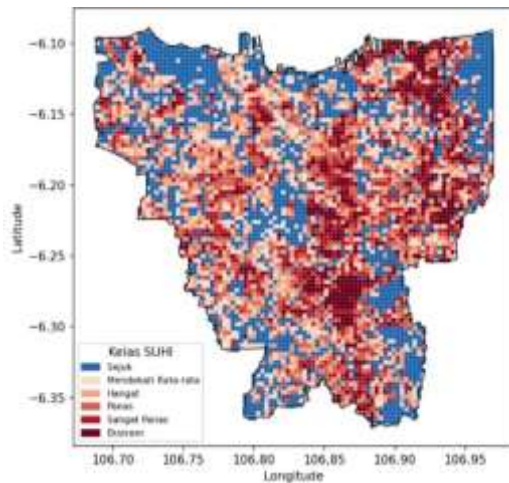
Gambar 15. Peta spasial kontribusi SHAP untuk NDVI—variasi geografis efek pendinginan vegetasi.

Rata-rata LST Jakarta pada dataset penelitian adalah 40,77°C. Dengan menggunakan nilai tersebut sebagai baseline, sebanyak 1.732 titik (29,8%) berada pada kelas *Sejuk*, sedangkan 4.089 titik (70,2%) berada pada kelas *Mendekati Rata-rata* hingga *Ekstrem*. Distribusi tersebut menunjukkan bahwa sebagian besar titik pengamatan memiliki kondisi termal yang berada pada kategori relatif lebih tinggi terhadap rata-rata LST kota, sesuai dengan pendekatan anomali termal intra-urban yang digunakan dalam penelitian.

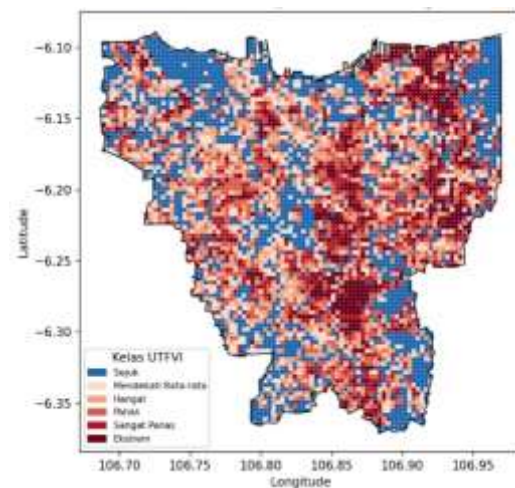
Tabel 11. Distribusi kelas anomali termal relatif/UTFVI

Kelas	Jumlah Titik	Persentase
Sejuk	1732	29.8%
Mendekati Rata-rata	918	15.8%
Hangat	984	16.9%
Panas	925	15.9%

Sangat Panas	661	11.4%
Ekstrem	601	10.3%



Gambar 16. Peta anomali termal relatif terhadap rata-rata LST Jakarta (40,8°C).



Gambar 17. Peta UTFVI berbasis mean-referenced LST.

4.2 Pembahasan

Hasil penelitian menunjukkan bahwa tuning hyperparameter memengaruhi performa dan interpretasi model. Sebelum tuning, Random Forest (RF) lebih unggul, sedangkan setelah tuning XGBoost sedikit lebih baik. Namun, hasil uji Wilcoxon menunjukkan perbedaan keduanya tidak signifikan ($p=0,352$). Perubahan pasangan interaksi SHAP juga menunjukkan bahwa interpretasi model dapat sensitif terhadap konfigurasi hyperparameter. XGBoost hanya sedikit lebih unggul karena penelitian menggunakan lima prediktor, dengan korelasi tinggi antara NDVI dan NDWI. Kondisi tersebut membatasi tambahan informasi yang dapat dimanfaatkan oleh boosting. Hal ini juga terlihat pada ensemble, ketika RF dan XGBoost memiliki korelasi residual yang sangat tinggi ($r=0,988$), sehingga ensemble tidak memberikan peningkatan dibandingkan XGBoost individual. Berdasarkan analisis SHAP, NDBI merupakan prediktor paling dominan dengan mean $|SHAP|$ sebesar 0,986. Temuan ini menunjukkan bahwa intensitas permukaan terbangun merupakan faktor penting dalam variasi LST. Interaksi NDBI dan kepadatan jalan juga menjadi interaksi paling kuat, sehingga kawasan dengan kepadatan bangunan dan jalan tinggi berpotensi mengalami pemanasan yang lebih besar. Korelasi NDVI–NDWI yang sangat tinggi ($r=-0,961$) menunjukkan adanya tumpang tindih informasi. Oleh karena itu, kontribusi kedua variabel berdasarkan SHAP perlu ditafsirkan secara hati-hati karena model dapat membagi kontribusi antarfitur yang saling berkorelasi. Hasil penelitian tidak dapat dibandingkan langsung dengan nilai SUHI urban–rural karena penelitian ini menggunakan anomali termal intra-urban. Perbedaan performa model dengan penelitian lain juga dapat dipengaruhi oleh karakteristik wilayah dan jumlah prediktor yang digunakan. Secara keseluruhan, penggunaan SHAP memberikan nilai tambah karena tidak hanya menunjukkan lokasi panas, tetapi juga membantu mengidentifikasi faktor yang berkontribusi terhadap variasi LST sehingga lebih relevan untuk mendukung perencanaan dan mitigasi panas perkotaan.

5. Kesimpulan

Penelitian menunjukkan bahwa XGBoost memiliki performa sedikit lebih baik daripada Random Forest ($R^2=0,507$ vs $0,497$), tetapi perbedaannya tidak signifikan secara statistik (Wilcoxon $p=0,352$). Ensemble RF–XGBoost juga tidak memberikan peningkatan berarti karena korelasi residual yang sangat tinggi ($0,988$), sehingga kedua model menghasilkan pola kesalahan yang hampir sama. Hasil SHAP menunjukkan bahwa NDBI merupakan prediktor LST paling dominan (mean $|SHAP|=0,986$), diikuti kepadatan jalan, NDVI, NDWI, dan jarak ke RTH. Interaksi NDBI×kepadatan jalan menjadi temuan penting untuk interpretasi tata ruang. Hyperparameter tuning juga terbukti memengaruhi peringkat model, signifikansi statistik, dan pola interaksi SHAP. Namun, hasil penelitian belum dapat digeneralisasikan sebagai performa spasial independen atau

SUHI karena keterbatasan skenario train-test split, cakupan periode, dan jumlah prediktor. Anomali termal dan UTFVI yang digunakan bersifat relatif dalam konteks intra-urban, bukan ukuran kontras urban-rural.

Referensi

- Fajary, F. R., Lee, H. S., Kubota, T., Bhanage, V., Pradana, R. P., Nimiya, H., & Putra, I. D. G. A. (2024). Comprehensive spatiotemporal evaluation of urban growth, surface urban heat island, and urban thermal conditions on Java island of Indonesia and implications for urban planning. *Heliyon*, 10(13), e33708. <https://doi.org/10.1016/j.heliyon.2024.e33708>
- Li, X., Zhou, Y., Asrar, G., Imhoff, M., & Xuecao, L. (2017). The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States. *Science of the Total Environment*, 605–606, 426–435. <https://doi.org/10.1016/j.scitotenv.2017.06.229>
- Masson, V., Lemonsu, A., Hidalgo, J., & Voogt, J. (2021). Urban climates and climate change. *[Informasi jurnal/penerbit perlu diverifikasi]*, 45(1), 411–444.
- McFeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Muzaky, H., & Jaelani, L. M. (2019). *Analisis pengaruh tutupan lahan terhadap distribusi suhu permukaan: Kajian urban heat island di Jakarta, Bandung dan Surabaya* [PDF].
- Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455), 1–24. <https://doi.org/10.1002/qj.49710845502>
- Oke, T. R., Mills, G., Christen, A., & Voogt, J. A. (2017). *Urban climates*. Cambridge University Press.
- Sarker, T., Fan, P., Messina, J. P., Mujahid, N., Aldrian, E., & Chen, J. (2024). Impact of urban built-up volume on urban environment: A case of Jakarta. *Sustainable Cities and Society*, 105, 105346. <https://doi.org/10.1016/j.scs.2024.105346>
- Siswanto, S., Nuryanto, D. E., Ferdiansyah, M. R., Prastiwi, A. D., Dewi, O. C., Gamal, A., & Dimiyati, M. (2023). Spatio-temporal characteristics of urban heat island of Jakarta metropolitan. *Remote Sensing Applications: Society and Environment*, 32, 101062. <https://doi.org/10.1016/j.rsase.2023.101062>
- Tahooni, A., Kakroodi, A. A., Kiavarz, M., & Mansourian, H. (2025). High-resolution urban LST downscaling via machine learning and SHAP: A case study in a rapidly urbanizing semi-arid region. *Sustainable Cities and Society*, 134, 106897. <https://doi.org/10.1016/j.scs.2025.106897>
- Wei, J., Li, Y., Jia, L., Liu, B., & Jiang, Y. (2025). The impact of spatiotemporal effect and relevant factors on the urban thermal environment through the XGBoost-SHAP model. *Land*, 14(2). <https://doi.org/10.3390/land14020394>
- Yang, J., Li, H., Xin, J., Yu, W., Ren, J., Yu, H., Xiao, X., Xia, C., Li, J., & Xiao, H. (2026). Investigating the effect of urban form on land surface temperature at block and grid scales based on XGBoost-SHAP. *Environmental Modelling & Software*. <https://doi.org/10.1016/j.envsoft.2025.106738>
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24, 583–594. <https://doi.org/10.1080/01431160304987>

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z., & Sobrino, J. A. (2019). Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives. *Remote Sensing*, 11(1), 48. <https://doi.org/10.3390/rs11010048>

How Cites

Suryoprayogo, H., Muttaqin, W., & Desianty, A. (2026). Penerapan Ensemble Machine Learning Random Forest dan XGBoost dengan Explainable Artificial Intelligence (XAI) untuk Prediksi Urban Heat Island dan Land Surface Temperature di DKI Jakarta. *Computer Journal*, 4(2), 275-291. <https://doi.org/10.58477/cj.v4i2.501>.

Publisher's Note

Yayasan Pendidikan Mitra Mandiri Aceh (YPPMA) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Submit your manuscript to YPMMA Journal and *benefit* from: <https://journal.ypmma.org/index.php/cj>.