

Decision Support System Berbasis Random Forest untuk Prioritisasi Pendayagunaan Aset Properti Idle PT PLN (Persero)

A Random Forest-Based Decision Support System for Prioritizing Idle Property Asset Utilization at PT PLN (Persero)

Muhammad Yasykur Roslan^{1,*}

¹⁾Program Studi Manajemen dan Penilaian Properti, Departemen Ekonomika dan Bisnis, Sekolah Vokasi, Universitas Gadjah Mada, Yogyakarta

Koresponden : *)muhammadyasykurroslan@mail.ugm.ac.id

ABSTRAK

Dari sekitar 5.457 lokasi properti idle PT PLN (Persero) yang tersebar di seluruh Indonesia, penyaringan dan penentuan prioritas pendayagunaannya selama ini hanya ditangani oleh empat orang di sub bidang pendayagunaan, karena PLN tidak memiliki perusahaan yang berada di bawah perusahaan induk properti seperti Pertamina atau KAI. Ketimpangan antara skala portofolio dan kapasitas pengelola ini membuat tahap Penyusunan Data Aset Properti Potensial dalam Perdir 0001 dan Perlak 0004 bertumpu pada penilaian subjektif yang tidak konsisten antar unit. Penelitian ini merancang dan memvalidasi Decision Support System berbasis Random Forest yang menghasilkan propensity score, yaitu indeks kemiripan setiap aset idle terhadap profil aset dayaguna, sebagai instrumen penyaringan pada tahap pra-Highest and Best Use (HBU). Model dilatih pada 3.876 observasi (3.142 aset idle dan 734 aset dayaguna) dengan 94 fitur yang merepresentasikan tiga pilar teoretis, yaitu karakteristik fisik lahan (Optimal Lot Theory), aksesibilitas lokasi (bid-rent theory), dan kondisi ekonomi wilayah, dengan fitur geospasial diekstraksi melalui GeoFeature Extractor yang dibangun khusus untuk penelitian ini. Skor dihitung melalui mekanisme out-of-bag untuk menghindari kontaminasi data latih, lalu disesuaikan dengan Skor Lokasi untuk mengoreksi dominasi variabel luas tanah pada feature importance (27,01%). Pada data uji, model mencapai AUC-ROC 0,9898 dan akurasi 95,62%, didukung validasi silang 5-fold (AUC 0,9925±0,0016) dan kalibrasi probabilitas yang baik (Brier score 0,0300). Dari 3.113 aset yang diranking, 445 aset (14,3%) teridentifikasi sebagai kandidat prioritas, memangkas sekitar 85,7% volume kandidat yang perlu dievaluasi sebelum analisis HBU dilaksanakan.

Kata Kunci : manajemen aset, aset idle, decision support system, propensity score, random forest

PENDAHULUAN

Fasilitas merupakan kebutuhan pokok bagi kehidupan manusia. Fasilitas bisa diartikan sebagai fasilitas bagi kehidupan dan fasilitas yang merupakan bagian dari infrastruktur. (Asih Aryani Soemitro, R., & Suprayitno, H. 2025). PT PLN (Persero) mengelola portofolio properti yang mencakup 133.961 catatan aset, terdiri atas 101.413 bidang tanah dan 32.548 unit bangunan yang tersebar di 34 provinsi. Dari jumlah tersebut, 117.691 record (87,9%) merupakan infrastruktur operasional ketenagalistrikan yang tidak dapat dikomersialisasikan, sementara sisanya, sekitar 16.257 record aset non-operasional berupa kantor, gudang, perumahan, tanah kosong, dan wisma, secara kategoris berpotensi menghasilkan pendapatan

melalui komersialisasi. Dari sekitar 6.280 lokasi non-operasional yang berpotensi dikomersialisasikan, baru 823 lokasi (13,1%) berstatus aktif dayaguna; sekitar 5.457 lokasi sisanya masih idle. Laporan keuangan konsolidasian PLN per 31 Desember 2025 mencatat nilai bruto aset tidak digunakan dalam operasi sebesar Rp14,24 triliun, dengan nilai buku bersih sekitar Rp6,97 triliun (PT PLN (Persero), 2026), sementara realisasi pelepasan aset tidak beroperasi hanya Rp554 miliar pada 2025 dan Rp1 triliun pada 2024, jauh di bawah skala portofolio yang perlu ditangani.

Gap antara potensi dan realisasi ini bersifat struktural. Berbeda dari BUMN lain yang memiliki entitas khusus pengelola properti, seperti Pertamina Patra Jasa atau KAI Properti, PLN tidak memiliki subholding properti tersendiri. Seluruh pengelolaan dan pendayagunaan aset properti PLN menjadi tanggung jawab satu divisi di kantor pusat dengan total 14 personel, dan sub bidang pendayagunaan yang secara langsung menangani komersialisasi hanya beranggotakan empat orang untuk menangani sekitar 5.457 lokasi aset idle dari Sabang sampai Merauke. Akibatnya, tahap Penyusunan Data Aset Properti Potensial dalam Peraturan Direksi PT PLN (Persero) Nomor 0001.KS/DIR/2025 tentang Kebijakan Strategis Manajemen Aset Properti (selanjutnya Perdir 0001) dan Peraturan Pelaksana Nomor 0004.SPO/DIR/2025 tentang Standar Prosedur Pendayagunaan Aset Properti (selanjutnya Perlak 0004) bertumpu pada penilaian subjektif yang berbeda-beda antar unit induk, tanpa referensi kuantitatif yang seragam (PT PLN (Persero), 2025a, 2025b). Direktorat Jenderal Kekayaan Negara (DJKN) Kementerian Keuangan turut mengidentifikasi pemetaan potensi pemanfaatan aset yang belum optimal sebagai hambatan utama pengelolaan aset publik di Indonesia, dengan target nasional menuju kondisi zero idle (Direktorat Jenderal Kekayaan Negara, 2019, 2021).

Perkembangan pembelajaran mesin membuka peluang untuk mengisi celah ini. Aset-aset PLN yang sudah berhasil dikomersialisasikan memiliki pola karakteristik fisik, aksesibilitas lokasi, dan kondisi ekonomi wilayah yang secara sistematis membedakannya dari aset yang masih idle, konsisten dengan bukti empiris keberhasilan machine learning dalam penilaian dan klasifikasi properti (Ho et al., 2020; Levantesi & Piscopo, 2020; Yilmazer & Kocaman, 2020) maupun prioritas pemeliharaan aset infrastruktur (Nithin et al., 2022). Model klasifikasi yang dilatih pada pola tersebut dapat menghasilkan propensity score, yaitu indeks kemiripan kuantitatif setiap aset idle terhadap profil aset dayaguna, sebagai dasar peringkat yang lebih terstruktur dan dapat dipertanggungjawabkan secara metodologis dibandingkan seleksi manual.

Penelitian ini merancang dan memvalidasi sebuah Decision Support System (DSS) berbasis Random Forest yang beroperasi sebagai instrumen penyaringan pada tahap Penyusunan Data Aset Properti Potensial dalam fase Perencanaan siklus pendayagunaan PLN. DSS ini secara sengaja diposisikan sebagai instrumen *pra-HBU*, bukan pengganti analisis Highest and Best Use (HBU) yang tetap memerlukan kajian lapangan, verifikasi legal, dan penilaian pasar oleh penilai properti. Tiga kontribusi utama penelitian ini adalah: (1) merancang skema prioritas berbasis data kuantitatif sebagai alternatif konkret terhadap seleksi subjektif yang selama ini berlaku; (2) membangun dan memvalidasi model Random Forest sebagai estimator propensity score, mengadaptasi konsep tersebut dari konteks kausal inferensi (Rosenbaum & Rubin, 1983; Lee et al., 2010) menjadi indeks kemiripan untuk keperluan ranking, sekaligus mengidentifikasi variabel yang paling membedakan profil aset idle dari aset dayaguna; dan (3) menghasilkan daftar peringkat prioritas aset idle PLN yang dapat langsung digunakan sebagai input kuantitatif pada siklus perencanaan pendayagunaan berikutnya.

STUDI PUSTAKA

Decision Support System dan Konteks Regulasi Pendayagunaan Aset PLN

Decision Support System (DSS) adalah sistem berbasis komputer yang membantu pengambil keputusan memecahkan masalah semi-terstruktur, menggabungkan basis data, basis model, dan antarmuka pengguna, tanpa menggantikan keputusan akhir manusia (Turban et al., 2011). Dalam konteks manajemen aset properti, DSS dapat berbentuk sistem pemeringkatan yang mengintegrasikan berbagai dimensi potensi aset ke dalam satu indeks prioritas.

Perdir 0001 mendefinisikan aset properti PLN dalam empat kategori resmi, yaitu aset tetap aktif beroperasi, aset tetap tidak beroperasi, aset tetap belum dimanfaatkan, dan properti investasi (PT PLN (Persero), 2025a), dengan pendayagunaan didefinisikan sebagai kegiatan mengoptimalkan pemanfaatan aset properti untuk mendukung pendapatan perusahaan. Perlak 0004 menetapkan siklus pendayagunaan dalam empat fase berurutan, yaitu Perencanaan, Pengembangan, Pelaksanaan, serta Monitoring, Evaluasi, dan Pelaporan (PT PLN (Persero), 2025b). Fase Perencanaan dibuka dengan tahap Penyusunan Data Aset Properti Potensial, titik masuk yang menentukan aset mana yang mendapat akses ke Kajian Highest and Best Use (HBU), sebuah analisis komprehensif yang mengkaji kelayakan fisik, peraturan, keuangan, dan produktivitas maksimal aset. Karena kajian HBU memerlukan waktu, biaya, dan keahlian yang signifikan, kualitas penyusunan daftar aset potensial menjadi penentu efisiensi keseluruhan siklus, sekaligus menjadi titik intervensi yang dituju DSS pada penelitian ini.

Tiga Pilar Teoretis Potensi Komersialisasi Aset

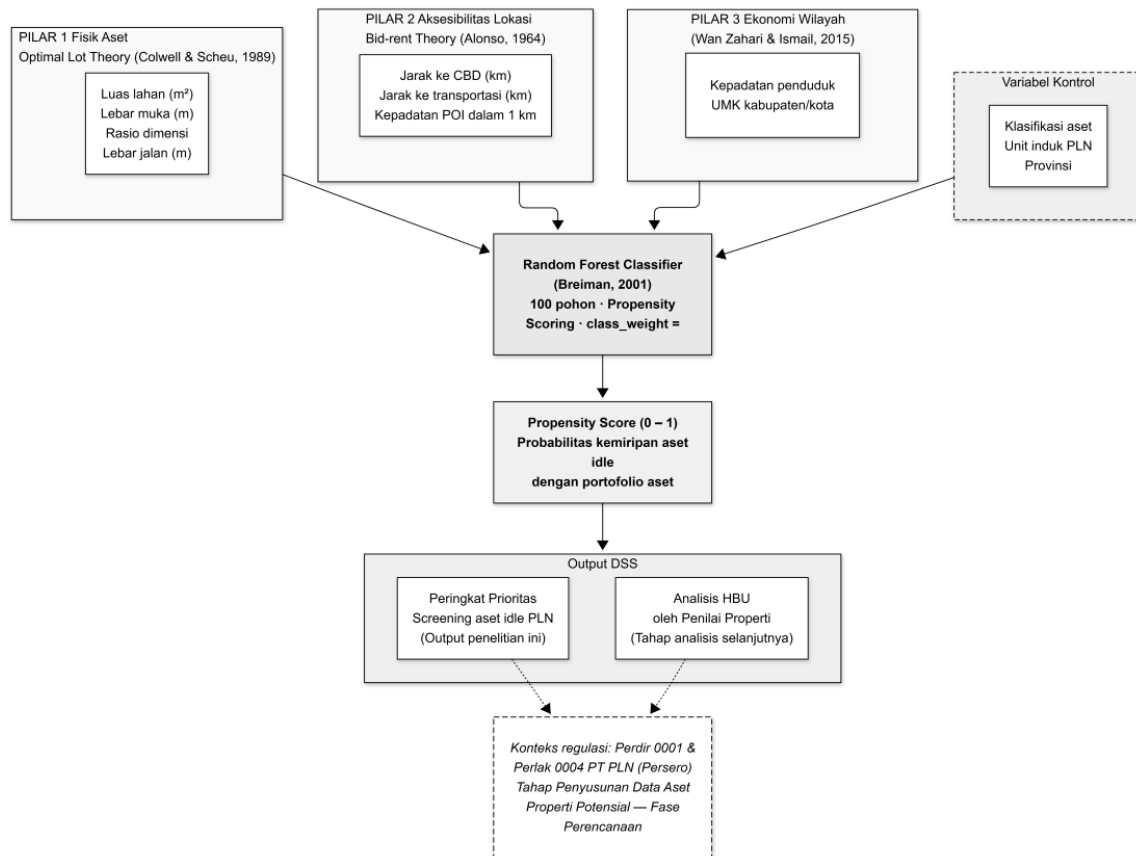
Potensi komersialisasi sebuah aset idle dioperasionalkan melalui tiga pilar teoretis yang bekerja secara simultan. Pilar pertama adalah karakteristik fisik lahan, berlandaskan *Optimal Lot Theory* (Colwell & Scheu, 1989), yang menyatakan bahwa kelayakan lahan untuk penggunaan komersial ditentukan oleh konfigurasi gabungan lebar muka (*frontage*) dan kedalaman lahan, bukan oleh luas total semata. Lahan dengan *frontage* lebar namun ukuran total tidak terlalu besar sering kali lebih layak secara komersial dibandingkan lahan luas dengan *frontage* sempit (Claurette & Li, 2019).

Pilar kedua adalah aksesibilitas lokasi, berlandaskan *bid-rent theory* (Alonso, 1964), yang menjelaskan bahwa harga lahan dan intensitas penggunaannya menurun seiring bertambahnya jarak dari pusat kegiatan ekonomi. Relevansi teori ini didukung bukti empiris lintas konteks, termasuk hubungan konsisten antara aksesibilitas dan nilai lahan di Australia (Rossini, 2020) serta pengaruh konfigurasi spasial jaringan jalan terhadap *bid-rent gradient* (Narvaez & Griffiths, 2013). Kepadatan Point of Interest (POI) komersial di sekitar lokasi turut digunakan sebagai proksi vibrancy kawasan (Yue et al., 2017; Cellmer, 2023), dengan data POI bersumber dari OpenStreetMap yang terbukti memiliki akurasi tinggi untuk kawasan perkotaan yang terdokumentasi baik (Haklay, 2010).

Pilar ketiga adalah kondisi ekonomi wilayah. Tinjauan Wan Zahari and Ismail (2015) menunjukkan bahwa kepadatan penduduk dan tingkat pendapatan wilayah secara konsisten mempengaruhi intensitas pemanfaatan lahan; penelitian ini menggunakan kepadatan penduduk sebagai proksi ukuran pasar dan Upah Minimum Kabupaten/Kota (UMK) sebagai proksi daya beli.

Ketiga pilar ini dirancang untuk saling mengoreksi kelemahan satu sama lain: pilar fisik tanpa pilar lokasi buta terhadap posisi pasar, pilar lokasi tanpa pilar fisik mengabaikan kelayakan fungsional lahan, dan kedua pilar tersebut tanpa pilar ekonomi wilayah mengasumsikan daya serap pasar yang seragam di seluruh Indonesia, asumsi yang tidak realistis mengingat rentang wilayah operasi PLN. Argumen inilah yang mendasari keputusan metodologis untuk menggunakan model Random Forest, yang mampu menangkap interaksi non-linear antar ketiga dimensi secara simultan, alih-alih skema skoring sederhana berbasis

bobot tetap. Kerangka konseptual yang menghubungkan ketiga pilar ini dengan model klasifikasi dan output DSS disajikan pada Gambar 1.



Gambar 1. Kerangka Konseptual Penelitian

Sumber: Diolah oleh peneliti, 2026

Random Forest sebagai Estimator dan Adaptasi Propensity Score

Random Forest adalah algoritma model pembelajaran yang diperkenalkan Breiman (2001), bekerja dengan membangun sejumlah besar pohon keputusan pada sampel bootstrap data pelatihan, dengan hanya sebagian acak fitur dipertimbangkan pada setiap titik percabangan (feature bagging). Keunggulan utamanya dibandingkan model parametrik seperti regresi logistik terletak pada kemampuan menangkap hubungan non-linear dan interaksi antarvariabel tanpa spesifikasi bentuk fungsional a priori, yang relevan untuk data aset properti yang heterogen. Dalam konteks properti, Random Forest terbukti unggul dibandingkan Support Vector Machine dalam prediksi harga di Hong Kong (Ho et al., 2020) dan regresi berganda dalam mass appraisal (Yilmazer & Kocaman, 2020), sementara Levantesi and Piscopo (2020) menunjukkan kontribusi signifikan variabel ekonomi wilayah di luar variabel fisik konvensional pada pasar London.

Konsep skor kecenderungan, diperkenalkan Rosenbaum and Rubin (1983) untuk kausal inferensi sebagai probabilitas suatu unit menerima perlakuan tertentu dikondisikan pada kovariat yang diamati, diadaptasi dalam penelitian ini mengikuti pendekatan Lee et al. (2010) dan Zhao et al. (2016) yang menggunakan machine learning sebagai estimatornya. Berbeda dari konteks aslinya, propensity score dalam penelitian ini bukan untuk kausal inferensi, melainkan indeks kemiripan kuantitatif yang mengukur probabilitas prediktif bahwa sebuah aset idle memiliki profil mirip dengan aset yang sudah berhasil dikomersialisasikan.

Posisi Penelitian

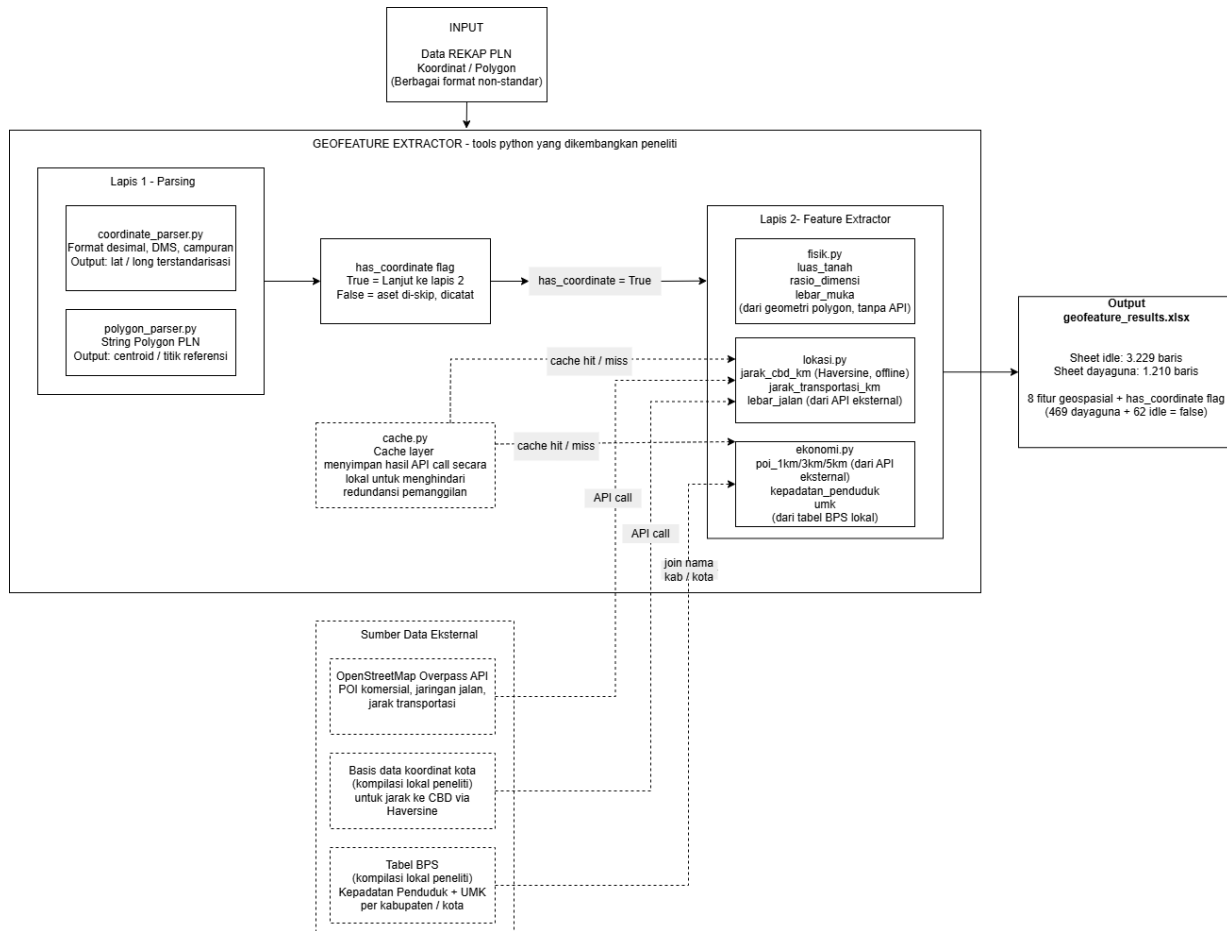
Studi machine learning untuk properti umumnya berfokus pada prediksi nilai pasar atau harga transaksi (Ho et al., 2020; Levantesi & Piscopo, 2020; Yilmazer & Kocaman, 2020), bukan pada klasifikasi status pemanfaatan atau scoring potensi komersialisasi aset yang tidak aktif diperdagangkan. Studi propensity score berbasis Random Forest sejauh ini terbatas pada domain uji klinis (Lee et al., 2010; Zhao et al., 2016), sementara studi prioritas berbasis machine learning pada aset infrastruktur terbatas pada konteks maintenance, bukan komersialisasi (Nithin et al., 2022). Penelitian ini mengisi ketiga celah tersebut secara simultan: menggunakan klasifikasi biner status pemanfaatan sebagai basis pembelajaran, mengadaptasi propensity score sebagai indeks kemiripan untuk ranking, dan menerapkannya pada konteks aset properti idle BUMN Indonesia yang belum pernah dikaji dalam literatur machine learning maupun manajemen aset.

METODA

Desain Penelitian dan Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif non-eksperimental dengan skema supervised learning, karena label kelas yang dibutuhkan sudah terbentuk secara alami dari rekaman operasional PLN: aset berstatus dayaguna menjadi label positif (1) dan aset yang masih idle menjadi label negatif (0). Seluruh data bersumber dari rekap aset internal PLN yang terdesentralisasi per unit induk, dikompilasi peneliti menjadi satu dataset terintegrasi mencakup lebih dari 130.000 catatan aset, kemudian difilter menjadi 3.229 baris data idle dan 1.210 baris data dayaguna (total 4.439 baris sebelum pembersihan).

Karena data rekap tidak memuat fitur geospasial yang dibutuhkan model, peneliti mengembangkan GeoFeature Extractor, sebuah program Python yang mengekstrak delapan variabel geospasial dan ekonomi secara otomatis dari koordinat aset melalui tiga sumber eksternal: OpenStreetMap melalui Overpass API untuk Point of Interest dan jarak ke fasilitas transportasi; formula Haversine terhadap basis data koordinat 100 kota yang dikompilasi lokal untuk jarak ke CBD; serta data Badan Pusat Statistik (BPS) yang dikompilasi lokal untuk kepadatan penduduk dan UMK per kabupaten/kota. Program ini bekerja dalam dua lapis, yaitu parsing koordinat dan polygon yang menstandarisasi berbagai format koordinat PLN yang tidak seragam, dan ekstraksi fitur melalui tiga sub-modul (Fisik, Lokasi, Ekonomi) yang berjalan paralel, dilengkapi mekanisme caching untuk menghindari redundansi pemanggilan API. Arsitektur lengkap disajikan pada Gambar 2.

**Gambar 2.** Arsitektur GeoFeature Extractor

Sumber: Diolah oleh peneliti, 2026

Variabel Penelitian

Variabel dependen adalah status pemanfaatan aset yang dikodekan biner (1 = dayaguna, 0 = idle). Variabel independen terdiri atas sembilan variabel numerik dari tiga pilar teoretis, disertai tiga variabel kontrol kategorikal (klasifikasi aset, unit induk, provinsi) yang di-encode melalui one-hot encoding, menghasilkan 85 kolom kategorikal yang digabung dengan 9 fitur numerik menjadi 94 fitur input model. Rincian variabel disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Pilar	Variabel	Sumber Teoretis / Justifikasi
Fisik Aset	Luas lahan (m ²)	Optimal Lot Theory (Colwell & Scheu, 1989)
Fisik Aset	Lebar muka (m)	Optimal Lot Theory; frontage sebagai penentu utama kelaikan komersial
Fisik Aset	Rasio dimensi	Optimal Lot Theory; rasio lebar muka terhadap kedalaman lahan
Fisik Aset	Lebar jalan (m)	Optimal Lot Theory; kelaikan fisik-fungsional akses
Lokasi	Jarak ke CBD (km)	Bid-Rent Theory (Alonso, 1964)
Lokasi	Jarak ke transportasi (km)	Bid-Rent Theory; aksesibilitas simpul transportasi
Lokasi	POI 1 km	Proksi vibrancy kawasan (Yue et al., 2017)
Ekonomi Wilayah	Kepadatan penduduk	Wan Zahari & Ismail (2015); proksi ukuran pasar
Ekonomi Wilayah	UMK	Wan Zahari & Ismail (2015); proksi daya beli
Kontrol	Klasifikasi aset, unit induk, provinsi	One-hot encoding (85 kolom); mengendalikan variasi tipe, tata kelola, dan lokasi

Sumber: diolah dari data internal PLN, BPS, dan OpenStreetMap, 2026

Variabel status kepemilikan tidak dimasukkan sebagai fitur prediktor meskipun secara konseptual relevan. Seluruh aset dayaguna secara struktural telah berstatus legal *clean and clear* sebagai prasyarat komersialisasi menurut Perlak 0004, sehingga kolom ini tidak tersedia pada dataset dayaguna; memasukkannya berpotensi menciptakan data leakage karena status bersih merupakan konsekuensi dari proses komersialisasi, bukan prediktor independen. Variabel ini alih-alih diposisikan sebagai gate filtering pra-model yang mendefinisikan batas populasi yang dianalisis, bukan sebagai fitur di dalam model.

Persiapan dan Pembersihan Data

Seluruh 4.439 observasi diproses melalui tiga langkah pembersihan berurutan. Pertama, observasi tanpa koordinat geografis valid dikeluarkan (469 dari 1.210 baris dayaguna, atau 38,7%; dan 62 dari 3.229 baris idle). Kedua, observasi dengan luas lahan tidak valid (nol, atau tidak tersedia karena tidak memiliki data polygon geometri) dikeluarkan, mengurangi 25 baris idle dan 7 baris dayaguna. Ketiga, nilai lebar muka di atas 500 meter, diidentifikasi sebagai kesalahan geometri polygon, diubah menjadi missing value untuk ditangani melalui imputasi median yang dihitung eksklusif dari data pelatihan guna mencegah data leakage. Proses ini menghasilkan dataset bersih 3.876 observasi (3.142 idle, 734 dayaguna); uji chi-square menunjukkan pengeluaran data bersifat non-random ($\chi^2 = 26,78$, $p < 0,001$) dengan underrepresentasi kategori Bangunan Komersial pada data yang dipertahankan, dibahas lebih lanjut sebagai keterbatasan penelitian.

Ketidakseimbangan kelas (81,1% idle berbanding 18,9% dayaguna) ditangani melalui parameter *class_weight='balanced'* pada Random Forest, yang memberi bobot lebih tinggi pada kelas minoritas secara proporsional terbalik dengan frekuensinya, dipilih karena secara metodologis lebih bersih dibandingkan oversampling seperti SMOTE (He & Garcia, 2009).

Spesifikasi Model dan Adaptasi Propensity Score

Random Forest adalah algoritma ensemble learning yang membangun B pohon keputusan independen dan menggabungkan prediksinya melalui majority voting untuk klasifikasi (Breiman, 2001), dengan kriteria pemisahan pada setiap node berbasis Gini impurity. Konfigurasi final model, hasil eksperimen komparatif enam nilai *max_depth* (5, 10, 15, 20, 30, dan tanpa batas), disajikan pada Tabel 2.

Tabel 2. Parameter Model Random Forest

Parameter	Nilai	Justifikasi
n_estimators	100	Jumlah pohon standar; penambahan pohon tidak meningkatkan AUC secara signifikan
max_depth	20	Menghasilkan AUC tertinggi pada data uji dengan overfitting gap terkendali (4,38%)
class_weight	balanced	Menyeimbangkan bobot kelas minoritas secara otomatis
random_state	42	Reprodusibilitas; uji sensitivitas mengonfirmasi hasil tidak bergantung pada seed
max_features	$\sqrt{n}$	Default klasifikasi; akar kuadrat dari jumlah fitur pada tiap node
oob_score	True	Mengaktifkan out-of-bag prediction agar propensity score bersifat out-of-sample

Sumber: Diolah oleh peneliti, 2026

Propensity score untuk aset i dengan vektor fitur X_i mengikuti adaptasi Lee et al. (2010) atas kerangka Rosenbaum and Rubin (1983):

$$e(X_i) = P(Y_i = 1 | X_i) = (1/B) \sum P_b(Y = 1 | X_i) \quad \dots (1)$$

Di mana $P_b(Y=1|X_i)$ adalah probabilitas kelas positif dari pohon ke- b dan B adalah jumlah total pohon. Untuk memastikan skor bersifat out-of-sample bagi seluruh aset, scoring digabungkan dari dua mekanisme: aset dalam data pelatihan (≈ 2.514 aset) di-score

menggunakan out-of-bag prediction, dari pohon yang tidak menyertakan aset tersebut dalam sampel bootstrap-nya, sedangkan aset dalam data uji (≈ 628 aset) di-score menggunakan predict_proba standar.

Strategi Evaluasi Model

Evaluasi model dirancang dalam lima lapis:

1. evaluasi hold-out pada pembagian data stratified 80:20 (3.100 pelatihan, 776 pengujian), menggunakan akurasi dan AUC-ROC (Hanley & McNeil, 1982);
2. validasi silang stratifikasi 5-fold (Kohavi, 1995) untuk menguji stabilitas dan generalisasi;
3. ablation test yang menghapus variabel luas tanah untuk memverifikasi bahwa model tidak bergantung secara trivial pada satu variabel (Meyes et al., 2019);
4. kalibrasi probabilitas melalui Brier score (Brier, 1950) dan kurva kalibrasi (Niculescu-Mizil & Caruana, 2005); dan
5. uji sensitivitas nilai awal menggunakan 10 nilai random_state berbeda untuk memastikan stabilitas hasil tidak bergantung pada kebetulan inisialisasi, dengan ambang batas simpangan baku AUC di bawah 0,003.

Penyesuaian Skor Berbasis Viabilitas Lokasi

Karena variabel luas tanah mendominasi feature importance (27,01%), berisiko memberi skor tinggi pada aset kecil di lokasi terpencil dengan aktivitas ekonomi minim, diterapkan penyesuaian berbasis viabilitas lokasi. Location score menggabungkan dua variabel yang merepresentasikan intensitas komersial dan daya beli pasar, yaitu jumlah POI dalam radius 5 km dan kepadatan penduduk kabupaten/kota, keduanya ditransformasi logaritma natural dan dinormalisasi Min-Max ke rentang [0,1]:

$$LS(i) = 0,5 \times \text{MinMax}[\ln(1+poi_5km)] + 0,5 \times \text{MinMax}[\ln(1+kepadatan_penduduk)] \dots (2)$$

Skor akhir yang menjadi dasar peringkat adalah adjusted score, dihitung sebagai rata-rata geometris antara propensity score dan location score:

$$AS(i) = \sqrt{PS(i) \times LS(i)} \dots (3)$$

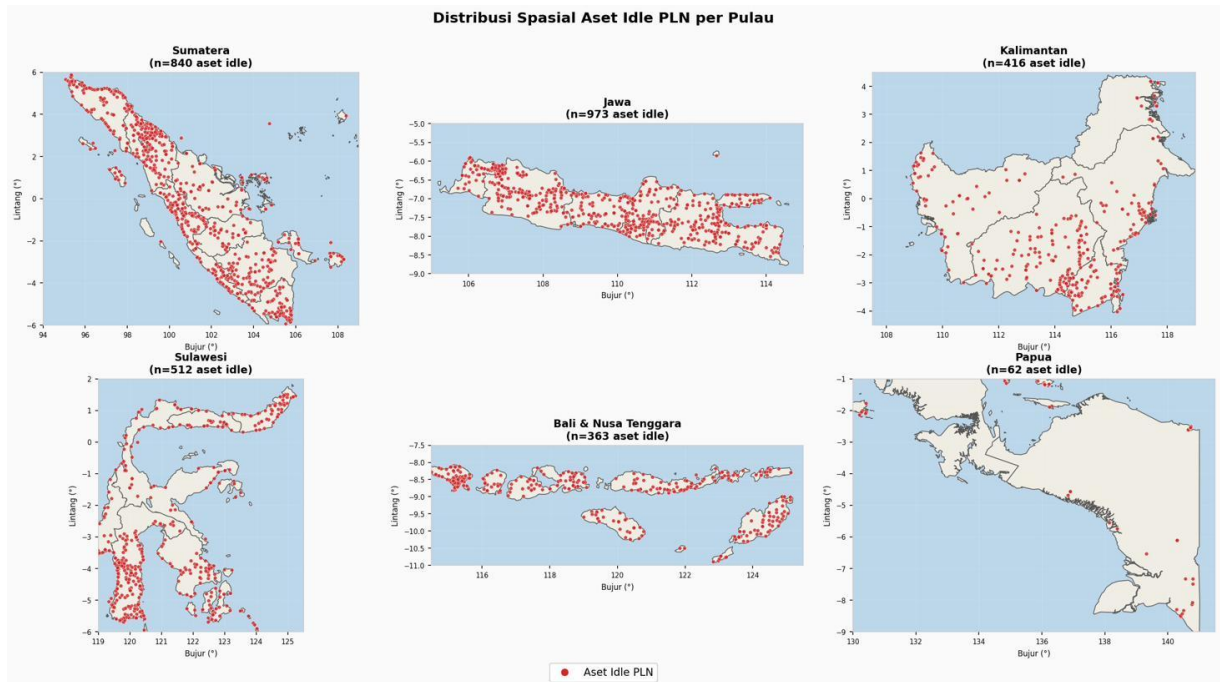
Rata-rata geometris dipilih karena sifatnya yang penalizing: sebuah aset hanya dapat meraih peringkat tinggi jika sekaligus memiliki profil fisik mirip aset dayaguna (propensity score tinggi) dan berada di kawasan dengan aktivitas ekonomi memadai (location score tinggi). Penyesuaian ini bekerja sebagai lapisan pasca-model yang memodifikasi urutan peringkat tanpa memengaruhi model klasifikasi maupun metrik evaluasinya. Seluruh tahapan diimplementasikan menggunakan Python dengan scikit-learn untuk pemodelan (Pedregosa et al., 2011).

ANALISIS PENELITIAN

Deskripsi Data

Data set akhir terdiri atas 3.876 observasi (81,1% idle, 18,9% dayaguna), tersebar di 34 provinsi dengan konsentrasi terbesar di Jawa dan Sumatera, sejalan dengan distribusi historis operasional PLN (Gambar 3). Perbedaan paling mencolok antar kelas terlihat pada empat variabel: median luas tanah aset dayaguna 90 m² berbanding 1.165 m² pada aset idle (selisih 12,9 kali); median jarak ke CBD 8,26 km berbanding 37 km (4,5 kali lebih jauh); median POI dalam radius 1 km sebesar 9 berbanding hanya 1; sementara lebar muka justru menunjukkan pola sebaliknya, aset dayaguna memiliki median lebih lebar (41,70 m berbanding 23,13 m), konsisten dengan Optimal Lot Theory yang menekankan pentingnya frontage di atas luas total. Dua variabel, yakni rasio dimensi dan lebar jalan, menunjukkan distribusi yang nyaris

berhimpit antar kelas, mengindikasikan kontribusi individual yang terbatas meski tetap dipertahankan karena berpotensi berkontribusi melalui interaksi dalam struktur pohon.



Gambar 3. Distribusi Spasial Aset Idle PLN per Pulau
Sumber: Data diolah, 2026

Matriks korelasi Pearson menunjukkan korelasi tertinggi antar variabel independen pada pasangan kepadatan penduduk-UMK ($r=0,58$) dan POI 1 km-kepadatan penduduk ($r=0,49$), keduanya di bawah ambang multikolinearitas berat 0,7. Temuan yang lebih menarik adalah bahwa variabel yang paling berkorelasi dengan label kelas justru bukan luas tanah ($r=-0,06$) melainkan lebar muka ($r=0,28$), mengindikasikan bahwa hubungan luas tanah dengan status komersialisasi bersifat non-linear, sebuah pola yang tidak tertangkap korelasi Pearson namun terungkap melalui Partial Dependence Plot pada bagian berikut.

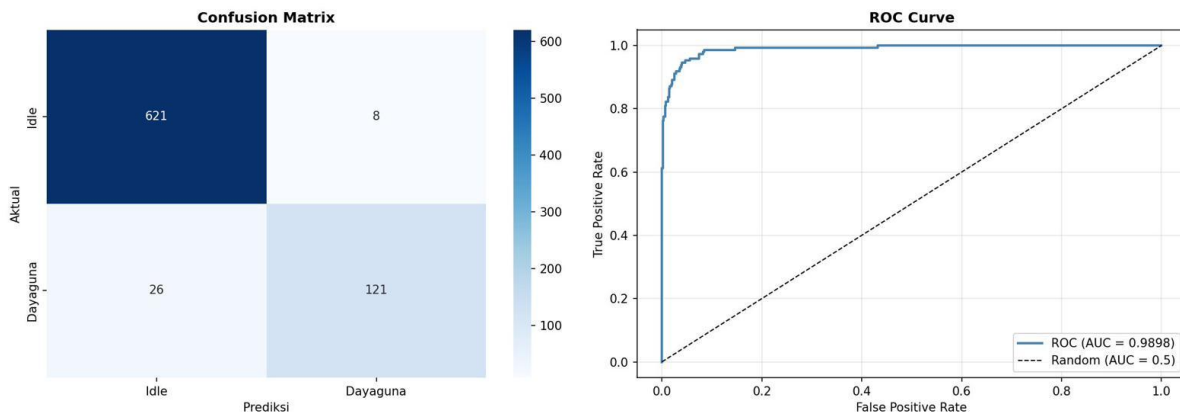
Evaluasi Kinerja Model

Model dengan konfigurasi $n_estimators=100$, $max_depth=20$, $class_weight='balanced'$ mencapai akurasi uji 95,62% dan AUC-ROC 0,9898 (Gambar 4), didukung validasi silang 5-fold dengan AUC rata-rata $0,9925 \pm 0,0016$ tanpa satu fold pun di bawah 0,99. Ablation test menunjukkan penghapusan variabel luas tanah, meski memiliki feature importance tertinggi, hanya menurunkan AUC sebesar 0,0079 (dari 0,9898 menjadi 0,9819), mengonfirmasi bahwa model tidak berfungsi sebagai penyortir variabel tunggal melainkan menangkap pola multidimensi. Kalibrasi probabilitas diukur melalui Brier score 0,0300, atau peningkatan 80,5% dibandingkan baseline no-skill (0,1535), mengindikasikan bahwa skor dapat dimaknai sebagai indikator kemiripan kuantitatif yang stabil. Uji sensitivitas terhadap 10 nilai $random_state$ menghasilkan simpangan baku AUC hanya 0,0013, jauh di bawah ambang batas 0,003, mengonfirmasi bahwa peringkat yang dihasilkan tidak bergantung pada kebetulan teknis inisialisasi model. Ringkasan seluruh hasil evaluasi disajikan pada Tabel 3.

Tabel 3. Ringkasan Evaluasi Model

Aspek Evaluasi	Metrik	Nilai
Hold-out (776 data uji)	AUC-ROC	0,9898
Hold-out (776 data uji)	Akurasi	95,62%
Validasi silang 5-fold	AUC rata-rata $\pm$ std	0,9925 $\pm$ 0,0016
Ablation test (tanpa luas_tanah)	Penurunan AUC	0,0079
Kalibrasi probabilitas	Brier score (vs. baseline 0,1535)	0,0300 (–80,5%)
Sensitivitas 10 seed	Simpangan baku AUC	0,0013

Sumber: Data diolah, 2026

**Gambar 4.** Confusion Matrix dan ROC Curve pada Data Uji

Sumber: Data diolah, 2026

Feature Importance dan Partial Dependence Plot

Tabel 4 dan Gambar 5 menyajikan 15 fitur dengan feature importance tertinggi. Variabel luas tanah menduduki peringkat pertama (0,2701), jauh di atas variabel lain, karena perbedaan median yang tajam antar kelas menjadikannya titik pemisah pertama yang paling efisien dalam struktur pohon. klas_Bangunan Komersial di peringkat kedua (0,0924) bekerja secara eksklusif, bukan inklusif: dari 2.238 aset idle berlabel ini, hanya 11,0% masuk kuintil teratas propensity score (skor median 0,0019), menunjukkan bahwa label ini justru berkorelasi negatif dengan potensi komersialisasi. Dari 15 fitur teratas, ketiga pilar teoretis terwakili proporsional: pilar fisik oleh luas tanah, lebar muka, lebar jalan, dan rasio dimensi; pilar lokasi oleh POI 1 km, jarak CBD, dan jarak transportasi; pilar ekonomi wilayah oleh kepadatan penduduk dan UMK.

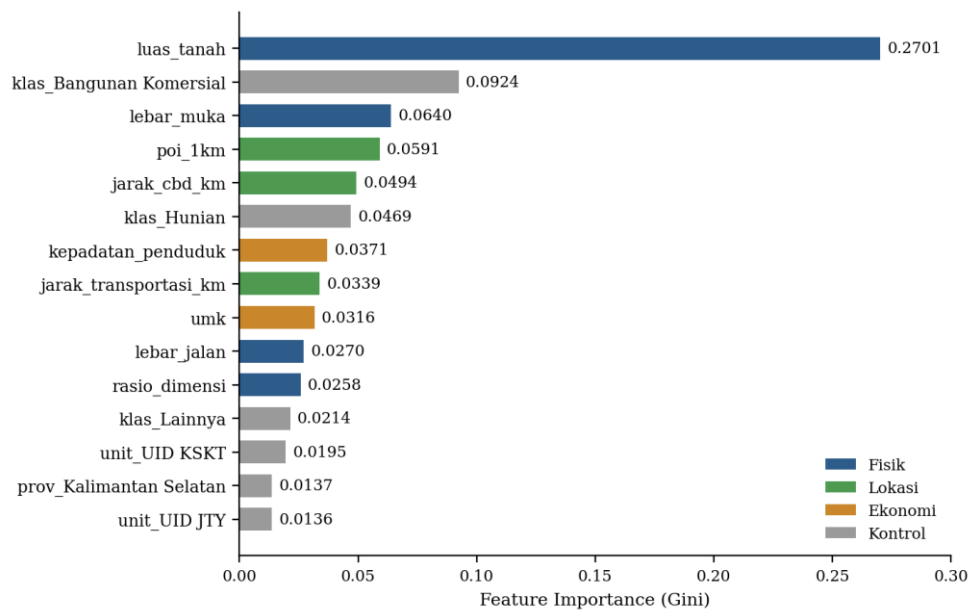
Tabel 4a. 15 Fitur Teratas Berdasarkan Feature Importance

No.	Variabel	Importance	Pilar
1	luas_tanah	0,2701	Fisik
2	klas_Bangunan Komersial	0,0924	Kontrol
3	lebar_muka	0,0640	Fisik
4	poi_1km	0,0591	Lokasi
5	jarak_cbd_km	0,0494	Lokasi
6	klas_Hunian	0,0469	Kontrol
7	kepadatan_penduduk	0,0371	Ekonomi
8	jarak_transportasi_km	0,0339	Lokasi
9	umk	0,0316	Ekonomi
10	lebar_jalan	0,0270	Fisik
11	rasio_dimensi	0,0258	Fisik
12	klas_Lainnya	0,0214	Kontrol

Tabel 4b. 15 Fitur Teratas Berdasarkan Feature Importance

No.	Variabel	Importance	Pilar
1	luas_tanah	0,2701	Fisik
2	klas_Bangunan Komersial	0,0924	Kontrol
3	lebar_muka	0,0640	Fisik
4	poi_1km	0,0591	Lokasi
5	jarak_cbd_km	0,0494	Lokasi
6	klas_Hunian	0,0469	Kontrol
7	kepadatan_penduduk	0,0371	Ekonomi
8	jarak_transportasi_km	0,0339	Lokasi
9	umk	0,0316	Ekonomi
10	lebar_jalan	0,0270	Fisik
11	rasio_dimensi	0,0258	Fisik
12	klas_Lainnya	0,0214	Kontrol
13	unit_UID KSKT	0,0195	Kontrol
14	prov_Kalimantan Selatan	0,0137	Kontrol
15	unit_UID JTY	0,0136	Kontrol

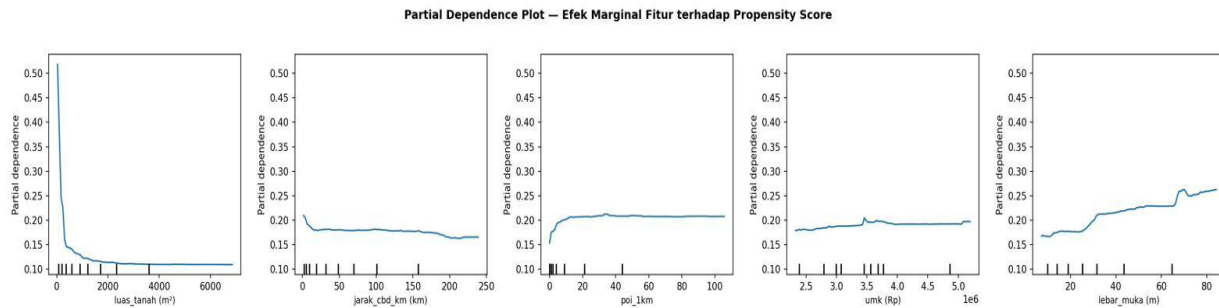
Sumber: Data diolah, 2026



Gambar 5. Ranking Feature Importance (15 Fitur Teratas)

Sumber: Data diolah, 2026

Partial Dependence Plot (Gambar 6) mengungkap pola non-linear yang tidak tertangkap korelasi Pearson. Luas tanah menunjukkan pola threshold tajam: partial dependence turun drastis dari $\approx 0,52$ ke $\approx 0,15$ pada rentang 0–200 m², kemudian mendatar setelah 1.000 m², menjelaskan mengapa korelasinya mendekati nol namun feature importance-nya dominan. Jarak ke CBD dan POI 1 km menunjukkan pola serupa dengan titik saturasi, konsisten dengan bid-rent theory dan konsep neighbourhood vibrancy (Yue et al., 2017). Lebar muka menunjukkan hubungan monoton bertahap dari $\approx 0,17$ pada 5 m menjadi $\approx 0,26$ pada 80 m, sejalan dengan Optimal Lot Theory, sementara UMK hanya bergerak dalam rentang sempit ($\approx 0,025$), mengindikasikan efek nyata namun lemah secara besaran sebagai faktor kontekstual penguat, bukan driver utama.

**Gambar 6.** Partial Dependence Plot Variabel Utama terhadap Propensity Score

Sumber: Data diolah, 2026

Peringkat Prioritas Aset

Propensity score dihitung untuk seluruh 3.142 aset idle; setelah pengecualian 29 aset dengan luas di bawah 50 m² (di bawah persentil ke-25 luas lahan aset dayaguna), 3.113 aset diranking berdasarkan adjusted score. Distribusi skor sangat condong ke kiri, median propensity score keseluruhan hanya 0,0052, mencerminkan bahwa mayoritas aset idle memiliki profil yang berbeda fundamental dari portofolio dayaguna, bukan kegagalan model. Sebanyak 445 aset (14,3%) memperoleh propensity score di atas 0,1 dan menjadi kandidat prioritas, memangkas sekitar 85,7% volume kandidat yang perlu dievaluasi sebelum analisis HBU. Tabel 5 menyajikan 10 aset dengan adjusted score tertinggi.

Tabel 5. Top 10 Aset Idle dengan Adjusted Score Tertinggi

Rank	Klasifikasi	Provinsi	Luas (m ²)	Prop. Score	Loc. Score	Adj. Score
1	Hunian	Jawa Timur	497	0,8683	0,8902	0,8792
2	Hunian	Jawa Timur	126	0,7367	0,8780	0,8043
3	Bang. Komersial	DKI Jakarta	254	0,5970	0,9925	0,7698
4	Hunian	Jawa Timur	253	0,5950	0,8902	0,7278
5	Hunian	Jawa Timur	417	0,5925	0,8923	0,7271
6	Tanah	DKI Jakarta	254	0,5062	0,9929	0,7090
7	Lainnya	Jawa Barat	355	0,5140	0,9668	0,7049
8	Hunian	Sumatera Utara	198	0,5552	0,8687	0,6945
9	Lainnya	Kalimantan Selatan	104	0,5136	0,8510	0,6611
10	Bang. Komersial	Jawa Tengah	64.594	0,5490	0,7714	0,6508

Sumber: Data diolah, 2026

Dari daftar ini, klasifikasi Hunian mendominasi, sejalan dengan komposisi portofolio dayaguna historis PLN, namun aset berklasifikasi Bangunan Komersial atau Tanah yang tetap lolos ke peringkat atas justru layak mendapat perhatian khusus karena berhasil bersaing meski bukan tipe dominan dalam data referensi. Sebanyak 641 aset (20,6% dari total yang diranking) memiliki kolom `is_placeholder=True`, yaitu data polygon geometri berupa kotak default alih-alih bentuk lahan sesungguhnya; variabel fisik pada aset ini berasal dari imputasi median sehingga propensity score-nya lebih mencerminkan potensi lokasi dan ekonomi wilayah, bukan konfirmasi fisik lahan, dan memerlukan verifikasi lapangan sebelum dilanjutkan ke analisis HBU.

Robustness: Analisis Terstratifikasi Berdasarkan Tipe Aset

Karena kelompok idle maupun dayaguna mencampur aset tanah kosong (klasifikasi Tanah: 11,1% dari idle, 6,0% dari dayaguna) dengan aset berbangunan (klasifikasi Tanah dan Bangunan), dilakukan analisis lanjutan untuk memverifikasi bahwa gap ukuran lahan 12,9 kali lipat dan dominasi luas tanah pada feature importance bukan artefak pencampuran kedua

tipe, menggunakan model yang identik dengan yang dilaporkan sebelumnya tanpa pelatihan ulang. Rasio gap luas tanah antara idle dan dayaguna ternyata hampir identik pada kedua tipe aset, yaitu 13,89 kali lipat pada kelompok Tanah dan 13,10 kali lipat pada kelompok Tanah dan Bangunan, sementara AUC-ROC tetap tinggi pada kedua kelompok (Tabel 6).

Tabel 6. Evaluasi Kinerja Model per Tipe Aset pada Data Uji

Tipe Aset	N Data Uji (Idle/Dayaguna)	AUC-ROC	Akurasi
Tanah	76 (66/10)	0,9985	0,9737
Tanah dan Bangunan	700 (563/137)	0,9892	0,9543
Keseluruhan	776 (629/147)	0,9898	0,9562

Sumber: Data diolah, 2026. Kelompok Tanah pada data uji hanya berisi 76 observasi (10 dayaguna), sehingga AUC 0,9985 rentan fluktuasi akibat ukuran sampel kecil.

Analisis SHAP (SHapley Additive exPlanations) mengonfirmasi luas tanah tetap menjadi fitur dengan kontribusi tertinggi pada kedua tipe aset (mean|SHAP| 0,158 pada Tanah dan 0,155 pada Tanah dan Bangunan), sementara POI 1 km menunjukkan pembeda yang jauh lebih tajam pada aset bertipe tanah kosong (peringkat SHAP ke-2) dibandingkan aset berbangunan (peringkat ke-5). Hasil ini memperkuat keandalan DSS yang dibangun, karena logika model terbukti konsisten dan tidak bergantung secara artifisial pada satu kombinasi tipe aset tertentu.

Implikasi bagi Praktik Manajemen Aset PLN

Makna efisiensi yang dihasilkan DSS ini baru tampak penuh ketika dikembalikan ke konteks kapasitas kelembagaan yang membatasinya: empat personel sub bidang pendayagunaan yang menangani sekitar 5.457 lokasi aset idle secara nasional. Pengurangan volume kandidat sebesar 85,7%, dari 3.113 menjadi 445, berarti kapasitas yang sama kini dapat diarahkan secara terstruktur ke aset-aset dengan profil paling menjanjikan, bukan tersebar tanpa panduan di atas ribuan kandidat yang tidak berurutan. DSS ini tidak menggantikan analisis HBU, kajian kelayakan fisik, legal, finansial, dan produktivitas maksimum tetap harus dilakukan penilai properti pada tiap aset prioritas, namun mengubah titik awal proses tersebut dari lebih dari 3.000 aset tanpa urutan menjadi 445 kandidat yang telah tersaring dan terurut berdasarkan kemiripan profil dengan portofolio dayaguna yang terbukti secara historis.

KESIMPULAN

Penelitian ini berhasil merancang, membangun, dan memvalidasi Decision Support System berbasis Random Forest yang dapat dioperasionalkan pada tahap Penyusunan Data Aset Properti Potensial dalam fase Perencanaan siklus pendayagunaan aset PLN sebagaimana diatur dalam Perdir 0001 dan Perlak 0004. Model yang dibangun menunjukkan performa kuat dan stabil pada seluruh dimensi evaluasi: AUC-ROC 0,9898 pada data uji, dikonfirmasi validasi silang 5-fold ($0,9925 \pm 0,0016$), kalibrasi probabilitas yang baik (Brier score 0,0300, peningkatan 80,5% dari baseline), dan stabilitas terhadap variasi inisialisasi (simpangan baku AUC antar-seed 0,0013). Analisis feature importance dan Partial Dependence Plot mengonfirmasi bahwa ketiga pilar teoretis (Optimal Lot Theory, bid-rent theory, dan kondisi ekonomi wilayah) terwakili proporsional dalam prediksi model, dengan temuan kunci berupa pola threshold non-linear pada luas tanah yang menjelaskan mengapa korelasi Pearson-nya mendekati nol namun feature importance-nya dominan, sekaligus menjadi justifikasi empiris penggunaan algoritma non-parametrik seperti Random Forest.

Secara praktis, DSS ini menghasilkan peringkat 3.113 aset idle PLN, mengidentifikasi 445 aset (14,3%) sebagai kandidat prioritas dan memangkas sekitar 85,7% volume kandidat yang perlu dievaluasi sebelum analisis HBU. Profil aset potensial yang teridentifikasi (ukuran

kecil hingga sedang di bawah 500 m², lokasi dekat pusat kota, lingkungan komersial aktif, dan wilayah dengan daya beli relatif tinggi) konsisten dengan prediksi ketiga pilar teoretis. Analisis robustness lanjutan mengonfirmasi bahwa temuan ini bertahan konsisten pada aset berklasifikasi Tanah maupun Tanah dan Bangunan, memperkuat keandalan model sebagai instrumen screening yang tidak bergantung secara artifisial pada komposisi tipe aset tertentu.

Penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan dalam penggunaan hasilnya: non-random dropout pada data dayaguna akibat data koordinat yang tidak lengkap (mengondisikan portofolio referensi model condong ke aset bertipe hunian berukuran kecil), eksklusif variabel status kepemilikan sebagai gate filtering pra-model, asimetri taksonomi pada label klasifikasi aset antar dataset idle dan dayaguna, keberadaan 641 aset (20,6%) dengan geometri placeholder yang memerlukan verifikasi lapangan, serta granularitas data ekonomi wilayah pada tingkat kabupaten/kota. Bagi PLN, DSS ini paling efektif digunakan sebagai titik awal yang terstruktur, bukan keputusan akhir, dengan verifikasi administratif dan lapangan tetap diperlukan sebelum penugasan analisis HBU. Penelitian selanjutnya dapat mengarah pada perbandingan algoritma ensemble lain (XGBoost, LightGBM), generalisasi kerangka DSS ke BUMN lain dengan tantangan aset idle serupa, evaluasi berbasis outcome prospektif terhadap implementasi rekomendasi DSS, serta dekomposisi lebih lanjut terhadap pola skor prioritas yang lebih tinggi pada aset berklasifikasi Tanah dibandingkan Tanah dan Bangunan yang ditemukan pada analisis robustness.

CATATAN.

Kode sumber GeoFeature Extractor dan ML Pipeline yang digunakan dalam penelitian ini tersedia secara terbuka melalui Zenodo (DOI: 10.5281/zenodo.20792444) dan repository GitHub rozlan0/skripsi-dss-pln-aset-idle di bawah lisensi CC BY 4.0.

DAFTAR PUSTAKA

- [1] Alonso, W. (1964). *Location and Land Use*. Harvard University Press.
- [2] Asih Aryani Soemitro, R., & Suprayitno, H. (2025). Pemikiran Awal tentang Konsep Dasar Manajemen Aset Fasilitas. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, 2, 1–14.
- [3] Breiman, L. (2001). “Random Forests”. *Machine Learning*, 45(1), 5–32.
- [4] Brier, G. W. (1950). “Verification of Forecasts Expressed in Terms of Probability”. *Monthly Weather Review*, 78(1).
- [5] Cellmer, R. (2023). “Points of Interest and Housing Prices”. *Real Estate Management and Valuation*, 31(1), 69–77.
- [6] Clauretie, T. M., & Li, H. (2019). “Land Values: Size Matters”. *The Journal of Real Estate Finance and Economics*, 58(1), 80–110.
- [7] Colwell, P. F., & Scheu, T. (1989). “Optimal lot size and configuration”. *Journal of Urban Economics*, 26(1), 90–109.
- [8] Direktorat Jenderal Kekayaan Negara. (2019). *Laporan kinerja Direktorat Jenderal Kekayaan Negara tahun 2019*. Kementerian Keuangan Republik Indonesia.
- [9] Direktorat Jenderal Kekayaan Negara. (2021). *Laporan kinerja Direktorat Jenderal Kekayaan Negara tahun 2021*. Kementerian Keuangan Republik Indonesia.
- [10] Haklay, M. (2010). “How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets”. *Environment and Planning B: Planning and Design*, 37(4), 682–703.
- [11] Hanley, J. A., & McNeil, B. J. (1982). “The meaning and use of the area under a receiver operating characteristic (ROC) curve”. *Radiology*, 143(1), 29–36.
- [12] Hartikainen, A. (2023). SHapley Additive exPlanations (SHAP) untuk interpretasi model penilaian properti.

- [13] He, H., & Garcia, E. A. (2009). "Learning from Imbalanced Data". *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- [14] Ho, W. K. O., Tang, B.-S., & Wong, S. W. (2020). "Predicting property prices with machine learning algorithms". *Journal of Property Research*.
- [15] Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1137–1143.
- [16] Lee, B. K., Lessler, J., & Stuart, E. A. (2010). "Improving propensity score weighting using machine learning". *Statistics in Medicine*, 29(3), 337–346.
- [17] Levantesi, S., & Piscopo, G. (2020). "The Importance of Economic Variables on London Real Estate Market: A Random Forest Approach". *Risks*, 8(4).
- [18] Meyes, R., Lu, M., Puiseau, C. W. de, & Meisen, T. (2019). Ablation Studies in Artificial Neural Networks (arXiv:1901.08644).
- [19] Narvaez, L., & Griffiths, S. (2013). *Spatial Configuration and Bid Rent Theory: How urban space shapes the urban economy*. Proceedings of the 9th International Space Syntax Symposium. Seoul, Korea.
- [20] Niculescu-Mizil, A., & Caruana, R. (2005). "Predicting good probabilities with supervised learning". *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, 625–632.
- [21] Nithin, A. H., Hobbs, M., Sriramula, S., & Sripada, Y. (2022). "An approach to improve asset maintenance and management priorities using machine learning techniques". *Safety and Reliability*, 41(3), 151–169.
- [22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research*, 12, 2825–2830.
- [23] PT PLN (Persero). (2025a). *Peraturan Direksi PT PLN (Persero) Nomor 0001.KS/DIR/2025 tentang Kebijakan Strategis Manajemen Aset Properti PT PLN (Persero)*. PT PLN (Persero).
- [24] PT PLN (Persero). (2025b). *Peraturan Pelaksana PT PLN (Persero) Nomor 0004.SPO/DIR/2025 tentang Standar Prosedur Pendayagunaan Aset Properti PT PLN (Persero)*. PT PLN (Persero).
- [25] PT PLN (Persero). (2025c). *Laporan Tahunan PT PLN (Persero) 2024*. PT PLN (Persero).
- [26] PT PLN (Persero). (2026). *Laporan Keuangan Konsolidasian PT PLN (Persero) per 31 Desember 2025*. PT PLN (Persero).
- [27] Rosenbaum, P. R., & Rubin, D. B. (1983). "The central role of the propensity score in observational studies for causal effects". *Biometrika*, 70(1), 41–55.
- [28] Rossini, P. (2020). Land Value vs Accessibility: Teaching Bid Rent Theory Using Empirical Evidence from Australia.
- [29] Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligence systems (9th ed.)*. Pearson Education.
- [30] Wan Zahari, W. Y., & Ismail, M. (2015). "Determinants of Land Use and Property Value". *Advanced Science Letters*, 21, 1150–1153.
- [31] Yilmazer, S., & Kocaman, S. (2020). "A mass appraisal assessment study using machine learning based on multiple regression and random forest". *Land Use Policy*, 99, 104889.
- [32] Yue, Y., Zhuang, Y., Yeh, A. G. O., Xie, J.-Y., Ma, C.-L., & Li, Q.-Q. (2017). "Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy". *International Journal of Geographical Information Science*, 31(4), 658–675.

- [33] Zhao, P., Su, X., Ge, T., & Fan, J. (2016). “Propensity Score and Proximity Matching Using Random Forest”. *Contemporary Clinical Trials*, 47, 85–92.