SPEKTRUM AKADEMIKA

Volume. 1, Nomor. 2, 2026

KLASIFIKASI DESA PERKOTAAN–PERDESAAN DAN INTELIJEN BISNIS INTERAKTIF: ANALISIS EMPIRIS PODES 2025 DI PROVINSI KEPULAUAN BANGKA BELITUNG

Marsha Suciana[1], Muhamad Liswansyah Pratama[2]
[1, 2] Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur,
[2]Universitas Pembangunan Nasional "Veteran" Jawa Timur
24083010039@student.upnjatim.ac.id[1], muhamad_liswansyah.sada@upnjatim.ac.id[2]

Abstract

The official urban–rural status of villages in Indonesia is set by Statistics Indonesia, and understanding the village characteristics behind this label requires combining it with village census data while respecting statistical confidentiality. This study analyses 393 villages in the Bangka Belitung Islands Province using 51 candidate predictors from the 2025 Village Potential Statistics and the official 2025 status issued by Statistics Indonesia Regulation 268/2025, which classifies 151 villages as urban and 242 as rural. We computed 34 indicators across seven domains at province and regency level with primary and secondary cell suppression, evaluated logistic regression and random forest classifiers with repeated nested five-by-five cross-validation, and built an eight-page interactive business intelligence dashboard with a twelve-step data story. Rural villages were far more often agriculture-based (94.2% versus 29.8%), whereas pharmacies (74.8% versus 12.4%) and minimarkets (84.1% versus 23.6%) were concentrated in urban villages; electricity, primary schools and fourth-generation mobile signal were almost universal. Both candidate models recognised the official status well beyond a majority-class baseline, with a Matthews correlation coefficient of 0.741 for logistic regression and 0.774 for random forest, and 0.740 and 0.764 when whole subdistricts were held out; the difference between the models was not significant ($p = 0.236$). The main income source of the village was the most informative predictor for both models. Village characteristics are therefore strongly associated with the official status, chiefly through economic structure, and a confidentiality-aware dashboard can communicate these patterns. No final model is selected, no village-level predictions are produced, and user evaluation of the dashboard remains future work.

Keyword: business intelligence; data storytelling; machine learning; urban–rural classification; village potential statistics

Abstrak

Status perkotaan–perdesaan desa di Indonesia ditetapkan oleh Badan Pusat Statistik, dan pemahaman atas karakteristik desa di balik label tersebut memerlukan pemaduan dengan data pendataan desa tanpa melanggar kerahasiaan statistik. Penelitian ini menganalisis 393 desa/kelurahan di Provinsi Kepulauan Bangka Belitung dengan 51 kandidat prediktor dari Pendataan Potensi Desa 2025 dan status resmi 2025 menurut Peraturan Kepala Badan Pusat Statistik Nomor 268 Tahun 2025, yaitu 151 desa perkotaan dan 242 desa perdesaan. Sebanyak 34 indikator dalam tujuh domain dihitung pada tingkat provinsi dan

kabupaten/kota dengan supresi sel primer dan sekunder; regresi logistik dan random forest dievaluasi dengan validasi silang bersarang berulang lima kali lima; serta dibangun dashboard intelijen bisnis interaktif delapan halaman dengan alur cerita data dua belas langkah. Desa perdesaan jauh lebih sering berpenghasilan utama pertanian (94,2% berbanding 29,8%), sedangkan apotek (74,8% berbanding 12,4%) dan minimarket (84,1% berbanding 23,6%) terkonsentrasi di desa perkotaan; listrik, sekolah dasar, dan sinyal seluler generasi keempat hampir merata. Kedua model kandidat mengenali status resmi jauh di atas model pembanding kelas mayoritas, dengan koefisien korelasi Matthews 0,741 untuk regresi logistik dan 0,774 untuk random forest, serta 0,740 dan 0,764 ketika seluruh kecamatan uji tidak terlihat saat pelatihan; selisih kedua model tidak bermakna (p = 0,236). Sumber penghasilan utama desa merupakan prediktor paling informatif pada kedua model. Karakteristik desa berasosiasi kuat dengan status resmi, terutama melalui struktur ekonomi, dan dashboard yang memperhatikan kerahasiaan dapat mengomunikasikan pola tersebut. Model final belum dipilih, prediksi per desa tidak dihasilkan, dan evaluasi pengguna dashboard masih menjadi pekerjaan lanjutan.

Kata Kunci: data storytelling; intelijen bisnis; klasifikasi perkotaan–perdesaan; pembelajaran mesin; Potensi Desa

1. PENDAHULUAN

Klasifikasi desa perkotaan dan perdesaan merupakan salah satu kerangka wilayah yang dipakai Badan Pusat Statistik (BPS) untuk menyajikan statistik menurut tipe daerah. Status 2025 ditetapkan melalui Peraturan Kepala BPS Nomor 268 Tahun 2025 dengan memadukan tiga kelompok kriteria, yaitu kepadatan penduduk, persentase keluarga pertanian, serta keberadaan atau akses terhadap fasilitas perkotaan, yang diolah dari Podes 2024 dan data luas wilayah hasil pemutakhiran desa (Badan Pusat Statistik, 2025). Lampiran peraturan tersebut memuat status setiap desa sehingga dapat dipadukan dengan data karakteristik desa untuk memahami profil di balik label. Di sisi lain, Pendataan Potensi Desa (PODES) 2025 memuat ratusan isian tentang kondisi geografis, infrastruktur, pendidikan, kesehatan, transportasi, konektivitas digital, dan ekonomi desa, tetapi data mikronya hanya dapat diakses melalui perjanjian penggunaan data sehingga penyajiannya harus memperhatikan kerahasiaan.

Penelitian terdahulu menunjukkan bahwa karakteristik desa dapat dipakai untuk mengenali tipe wilayah dengan pembelajaran mesin. Br. Saragih dan Hartojo (2026) memadukan PODES dengan kepadatan penduduk WorldPop dan indeks lahan terbangun dari citra satelit untuk mengklasifikasikan 244 desa di Kabupaten Toba, dengan *random forest* (RF) sebagai algoritma terbaik. Ridwan dan Tanur (2026) memakai PODES 2024 dan Indeks Desa Membangun untuk mengklasifikasikan status pembangunan desa di Kabupaten Bekasi; RF dengan penyeimbangan kelas SMOTE memberi kinerja terbaik dan sektor ekonomi dominan termasuk prediktor paling berpengaruh. Pada konteks internasional, Hu dkk. (2022) mengidentifikasi kemiskinan tingkat desa dengan RF berbasis citra resolusi tinggi dan data geospasial, sedangkan Wang dkk. (2024) membatasi wilayah perkotaan–perdesaan melalui spasialisasi penduduk. Kedua studi berbasis PODES tersebut menilai model dengan akurasi, presisi, recall, atau F1, dan salah satunya memakai satu pembagian data latih–uji 70:30 (Ridwan & Tanur, 2026). Ketika target berupa status resmi yang disusun dari kriteria tertentu, prediktor yang beririsan konsep dengan kriteria tersebut juga berisiko membuat kinerja tampak lebih baik daripada kemampuan generalisasi sebenarnya.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Kebocoran data (*data leakage*) merupakan kegagalan yang meluas dalam sains berbasis pembelajaran mesin dan dapat menghasilkan kinerja yang terlalu optimistis (Kapoor & Narayanan, 2023). Untuk data berukuran kecil, validasi silang bersarang berulang dan uji permutasi dianjurkan agar estimasi kinerja tidak bergantung pada satu pembagian data (Steinert dkk., 2024). Ketidakseimbangan kelas membatasi kegunaan model dan telah memunculkan beragam strategi penanganan, mulai dari pembobotan kelas hingga pengambilan sampel ulang (Carvalho dkk., 2025; W. Chen dkk., 2024); koefisien korelasi Matthews (*Matthews correlation coefficient*, MCC) lebih informatif daripada akurasi dan F1 untuk klasifikasi biner yang tidak seimbang (Chicco & Jurman, 2020). Perbandingan dua model pada fold validasi silang yang sama memerlukan koreksi varians karena data latih antarfold saling tumpang tindih (Nadeau & Bengio, 2003).

Hasil analitis juga perlu dikomunikasikan. Dashboard intelijen bisnis (*business intelligence*, BI) memengaruhi kualitas keputusan secara tidak langsung melalui format, kemutakhiran, dan kelengkapan informasi (Hjelle dkk., 2024), tetapi kegunaannya perlu dievaluasi dengan instrumen usability yang sesuai (Almasi dkk., 2023) dan prinsip desain yang berpusat pada pengguna (Ansari & Martin, 2022, 2024). Penceritaan data (*data storytelling*) menambahkan konteks, alur, dan ajakan bertindak pada visualisasi (Lo Duca & McDowell, 2025). Bukti empiris menunjukkan cerita data dapat mempercepat pemahaman insight dibandingkan visualisasi konvensional (Shao dkk., 2024), dan berbagai genre serta alat narasi visual, termasuk *scrollytelling*, terus berkembang (Q. Chen dkk., 2024; Mörth dkk., 2023).

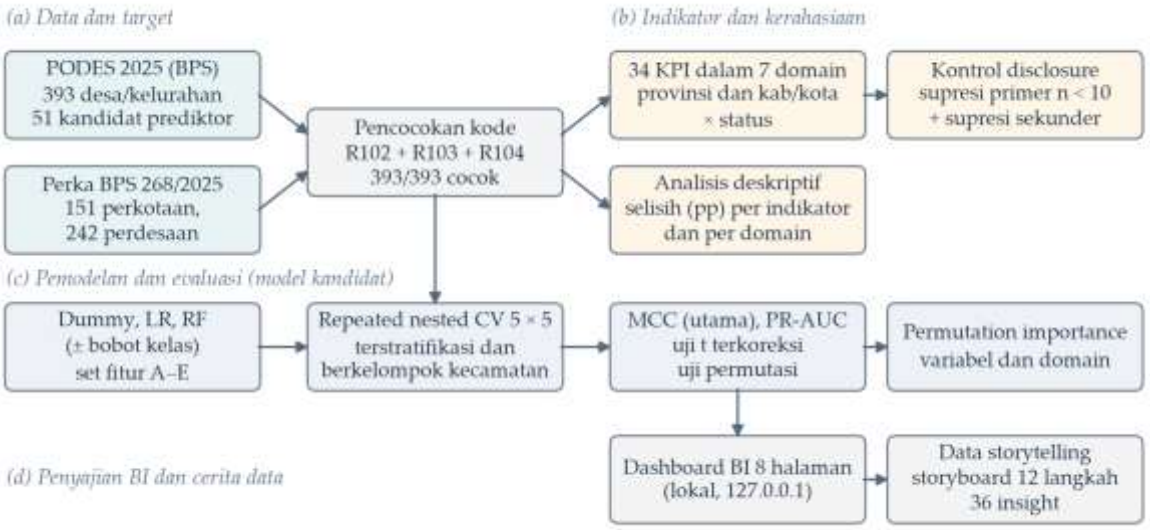
Berdasarkan celah tersebut, penelitian ini memadukan status resmi 2025 dengan karakteristik PODES 2025 untuk 393 desa/kelurahan di Provinsi Kepulauan Bangka Belitung. Tujuannya adalah (1) menggambarkan perbedaan karakteristik desa perkotaan dan perdesaan beserta variasinya antarkabupaten/kota, (2) menilai seberapa baik karakteristik PODES 2025 mengenali status resmi dengan desain validasi yang jujur terhadap ukuran sampel, ketidakseimbangan kelas, dan risiko irisan konsep, serta (3) merancang dashboard BI interaktif dan alur cerita data yang menyajikan temuan dengan kontrol pengungkapan (*disclosure control*). Penelitian ini tidak bertujuan mereplikasi metode klasifikasi BPS, tidak memilih model final, dan tidak menghasilkan prediksi status per desa.

2. METODE

Penelitian ini merupakan studi kuantitatif potong lintang dengan unit analisis desa/kelurahan. Alur kerja terdiri atas empat tahap: penyiapan data dan target, penyusunan indikator dengan kontrol pengungkapan, pemodelan dan evaluasi model kandidat, serta penyajian melalui dashboard BI dan cerita data (Gambar 1).

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 1.
Alur penelitian: data, indikator dan kontrol pengungkapan, pemodelan kandidat, serta penyajian BI dan cerita data



Catatan. Sumber: rancangan penelitian (digambar ulang dari dokumentasi teknis proyek).

2.1 Data dan unit analisis

Populasi penelitian adalah seluruh 393 desa/kelurahan di Provinsi Kepulauan Bangka Belitung yang tercakup dalam PODES 2025, sehingga tidak dilakukan penarikan sampel. Data PODES 2025 (Desa) untuk variabel dan wilayah terpilih bersumber dari BPS dan digunakan berdasarkan perjanjian penggunaan data dengan BPS Provinsi Kepulauan Bangka Belitung untuk keperluan penelitian dan analisis; rekaman data tersebut bersifat rahasia dan tetap menjadi milik BPS. Seluruh tabel dan gambar dalam artikel ini merupakan hasil olahan penulis dari data PODES 2025 dan lampiran peraturan status resmi, kecuali dinyatakan lain. Variabel target adalah status resmi 2025 (1 = perkotaan, 0 = perdesaan) dari lampiran Peraturan Kepala BPS Nomor 268 Tahun 2025 Buku 1 Sumatera (Badan Pusat Statistik, 2025). Status dicocokkan per desa dengan kode wilayah delapan digit (kode kabupaten, kecamatan, dan desa); seluruh 393 desa cocok tanpa duplikasi, dengan 151 desa perkotaan dan 242 desa perdesaan. Komposisi status per kabupaten/kota disajikan pada Tabel 1. Variabel R105 pada PODES (label perkotaan/perdesaan sebelumnya) hanya dipakai sebagai pembanding historis dan tidak pernah menjadi prediktor maupun target.

Tabel 1.
Komposisi status resmi 2025 per kabupaten/kota

Kode	Kabupaten/kota	Desa/kelurahan	Perkotaan	Perdesaan	% perkotaan
1971	Pangkalpinang	42	42	0	100,0
1901	Bangka	81	34	47	42,0
1906	Belitung Timur	39	16	23	41,0
1902	Belitung	49	19	30	38,8
1904	Bangka Tengah	63	19	44	30,2

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Kode	Kabupaten/kota	Desa/kelurahan	Perkotaan	Perdesaan	% perkotaan
1903	Bangka Barat	66	16	50	24,2
1905	Bangka Selatan	53	5	48	9,4
19	**Provinsi Kep. Bangka Belitung**	**393**	**151**	**242**	**38,4**

Catatan. Sumber: lampiran Peraturan Kepala BPS Nomor 268 Tahun 2025 yang dicocokkan dengan PODES 2025 (diolah). Urutan menurut proporsi perkotaan.

Prediktor berasal dari 51 kandidat variabel PODES 2025 yang dikelompokkan ke dalam tujuh domain: kondisi geografis, infrastruktur dasar, pendidikan, kesehatan, transportasi, konektivitas digital, dan aktivitas ekonomi. Perlu dicatat bahwa status 2025 disusun dari Podes 2024, sedangkan prediktor berasal dari PODES 2025, dan indikator jumlah penduduk serta keluarga pertanian yang menjadi kriteria resmi tidak tersedia pada ekstrak yang digunakan. Karena itu, analisis ini mengukur asosiasi antara karakteristik desa dan status resmi, bukan reproduksi perhitungan skor BPS.

2.2 Indikator dan KPI

Dari data PODES disusun 34 indikator kinerja utama (*key performance indicator*, KPI) yang dapat dihitung, berupa persentase desa yang memenuhi kondisi tertentu, rasio jumlah, jumlah total, dan rata-rata per desa. Setiap KPI memiliki rumus, penyebut, satuan, dan catatan agregasi agar tidak terjadi penghitungan ganda; misalnya keberadaan sekolah negeri dan swasta dijumlahkan per desa sebelum diuji lebih dari nol. Tiga indikator kriteria resmi (kepadatan penduduk, persentase keluarga pertanian, dan skor klasifikasi) ditandai tidak dapat dihitung. Indikator persentase yang nilainya hampir seragam di tingkat provinsi (≤ 3% atau ≥ 97%) tidak dipakai sebagai pembeda. Empat belas indikator persentase yang bervariasi dan menjadi dasar perbandingan didefinisikan pada Tabel 2. Selisih dinyatakan dalam poin persen (pp) sebagai nilai perkotaan dikurangi nilai perdesaan.

Tabel 2.
Definisi operasional 14 indikator pembeda

Kode KPI	Indikator	Domain	Rumus (pembilang / penyebut)	Variabel PODES 2025
C07	Penghasilan utama pertanian	Aktivitas ekonomi	COUNT(R906A = 1) / COUNT(desa) × 100	R906A
H03	Memiliki apotek	Kesehatan	COUNT(R702LK2 > 0) / COUNT(desa) × 100	R702LK2
C05	Memiliki minimarket/swalayan	Aktivitas ekonomi	COUNT(R905EK2 > 0) / COUNT(desa) × 100	R905EK2
G02	Di dalam/tepi kawasan hutan	Kondisi geografis	COUNT(R303A IN {1,2}) / COUNT(desa) × 100	R303A
C01	Memiliki kantor bank	Aktivitas ekonomi	COUNT((R902A1 + R902A2) > 0) / COUNT(desa) × 100	R902A1, R902A2
I03	Penerangan jalan utama (sebagian besar)	Infrastruktur dasar	COUNT(R502A = 1) / COUNT(desa) × 100	R502A
E03	Memiliki SMA	Pendidikan	COUNT((R701GK2 + R701GK3) > 0) / COUNT(desa) × 100	R701GK2, R701GK3
E02	Memiliki SMP	Pendidikan	COUNT((R701EK2 + R701EK3) > 0) / COUNT(desa) × 100	R701EK2, R701EK3
C04	Memiliki pasar permanen	Aktivitas ekonomi	COUNT(R905BK2 > 0) / COUNT(desa) × 100	R905BK2
H02	Memiliki puskesmas	Kesehatan	COUNT((R702DK2 + R702EK2) > 0) / COUNT(desa) × 100	R702DK2, R702EK2
G01	Berbatasan dengan laut	Kondisi geografis	COUNT(R302A = 1) / COUNT(desa) × 100	R302A

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Kode KPI	Indikator	Domain	Rumus (pembilang / penyebut)	Variabel PODES 2025
E04	Memiliki SMK	Pendidikan	COUNT((R701IK2 + R701IK3) > 0) / COUNT(desa) × 100	R701IK2, R701IK3
H01	Memiliki rumah sakit	Kesehatan	COUNT(R702AK2 > 0) / COUNT(desa) × 100	R702AK2
T03	Ada angkutan umum trayek tetap	Transportasi	COUNT('A' IN R801C1) / COUNT(desa) × 100	R801C1

Catatan. Semua indikator berupa persentase desa (%). COUNT(desa) = jumlah desa/kelurahan dalam kelompok. Kode variabel mengikuti kuesioner PODES 2025; definisi lengkap 34 KPI tersedia dalam kamus data paket penelitian.

2.3 Kontrol pengungkapan

Seluruh indikator hanya disajikan pada tingkat provinsi dan kabupaten/kota menurut status (semua, perkotaan, perdesaan); tidak ada data, nama, atau prediksi tingkat desa yang ditampilkan. Sel dengan kurang dari 10 desa disembunyikan (supresi primer). Supresi sekunder kemudian diterapkan secara bertahap agar tidak ada relasi aditif, yaitu semua = perkotaan + perdesaan dan provinsi = jumlah kabupaten/kota, yang hanya memiliki satu sel tersembunyi sehingga nilainya dapat dihitung balik. Setiap sel memiliki tiga keadaan: bernilai, disembunyikan, atau tidak berlaku (kelompok tanpa desa), dan sel yang disembunyikan tidak pernah ditampilkan sebagai nol. Aturan ini merupakan langkah kehati-hatian dan tidak dimaksudkan sebagai bukti kepatuhan terhadap ketentuan izin penggunaan data.

2.4 Desain pemodelan dan validasi

Tiga jenis model dibandingkan: model pembanding kelas mayoritas (*dummy*), regresi logistik (*logistic regression*, LR), dan RF, masing-masing tanpa dan dengan bobot kelas seimbang pada set fitur utama. Seluruh praproses, yaitu pengodean aturan lompatan kuesioner, imputasi, pengodean kategori, standardisasi, dan penyaringan variansi, berada di dalam *pipeline* sehingga hanya disesuaikan pada data latih setiap fold. Penyetelan hiperparameter dilakukan di loop dalam dengan metrik MCC; ruang pencarian LR adalah $C \in \{0{,}01; 0{,}1; 1; 10\}$, sedangkan RF memakai 300 pohon dengan kedalaman maksimum {tanpa batas, 8} dan jumlah sampel minimum daun {1, 3}.

Estimasi kinerja memakai validasi silang bersarang berulang 5 × 5 (25 fold uji; *seed* 42) dengan dua skema. Skema terstratifikasi menjadi estimasi utama dalam provinsi, sedangkan skema berkelompok kecamatan (47 kecamatan sebagai grup) menguji generalisasi ke kecamatan yang tidak terlihat saat pelatihan. Fold luar dibentuk sekali per skema dan dipakai bersama oleh semua model. Validasi berkelompok kabupaten/kota tidak dijalankan karena hanya terdapat tujuh kabupaten/kota dan Kota Pangkalpinang seluruhnya perkotaan.

Untuk menilai risiko irisan konsep dengan kriteria resmi, disusun lima set fitur. Set A (51 fitur) memuat seluruh kandidat dan diperlakukan sebagai batas atas. Set B (40 fitur) menghapus 11 fitur yang beririsan langsung dengan fasilitas kriteria resmi, set C (37 fitur) juga menghapus tiga fitur proksi, sedangkan set D (36 fitur) dan E (50 fitur) menghapus variabel sumber penghasilan utama (R906A) dari set C dan A. Set B–E merupakan analisis sensitivitas dan tidak diklaim bebas kebocoran.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

2.5 Evaluasi dan analisis tambahan

MCC dipakai sebagai metrik utama karena memperhitungkan keempat sel matriks kebingungan (Chicco & Jurman, 2020):

$$MCC = (TP \times TN - FP \times FN) / \sqrt{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]}$$

dengan TP, TN, FP, dan FN masing-masing menyatakan positif benar, negatif benar, positif palsu, dan negatif palsu, serta kelas positif adalah perkotaan. Metrik pendukung meliputi akurasi seimbang, recall, presisi, dan F1 kelas perkotaan, area di bawah kurva presisi–recall (PR-AUC), dan area di bawah kurva ROC (ROC-AUC). Nilai dilaporkan sebagai rata-rata ± simpangan baku (SD) 25 fold uji. Perbedaan antarmodel atau antarset fitur diuji dengan uji t sampel ulang terkoreksi Nadeau–Bengio dua sisi pada 25 fold yang sama (Nadeau & Bengio, 2003). Uji permutasi label (200 permutasi untuk LR dan 100 untuk RF) menilai apakah kinerja melampaui kebetulan (Steinert dkk., 2024); karena hiperparameter ditetapkan dari label asli, uji ini cenderung menguntungkan model asli. Kontribusi prediktor diukur dengan *permutation importance* pada data uji lima fold pertama dengan sepuluh pengulangan dan skor MCC, lalu diagregasi per variabel dan per domain. Metode ini mengukur kontribusi setiap fitur terhadap prediksi model, tetapi tidak selalu menangkap interaksi fitur yang kompleks (O'Brien Quinn dkk., 2024); karena itu, importance dibaca sebagai kontribusi prediktif, bukan hubungan sebab-akibat.

2.6 Perancangan dashboard dan cerita data

Dashboard dibangun dengan Streamlit dan Plotly serta dijalankan secara lokal (127.0.0.1) dengan pemeriksaan awal yang memastikan berkas, versi paket, kontrak target 393/151/242, dan aturan supresi terpenuhi sebelum aplikasi dibuka. KPI yang tampil dihitung dengan fungsi yang sama dengan tabel artikel dan melewati fungsi supresi yang sama. Cerita data disusun dalam tahapan konteks, pertanyaan, bukti, insight, interpretasi, implikasi, dan keterbatasan, dengan menambahkan konteks, alur, dan ajakan bertindak pada visualisasi sebagaimana gagasan kerangka S-DIKW (Lo Duca & McDowell, 2025). Setiap insight dicatat dalam katalog beserta jenisnya (fakta atau interpretasi), angka pendukung, sumber, dan batas interpretasi. Jaminan mutu dilakukan dengan pengujian otomatis, pencocokan setiap sel KPI dengan perhitungan independen, dan pencocokan teks narasi dengan angka sumber. Evaluasi usability dengan pengguna belum dilakukan dalam penelitian ini.

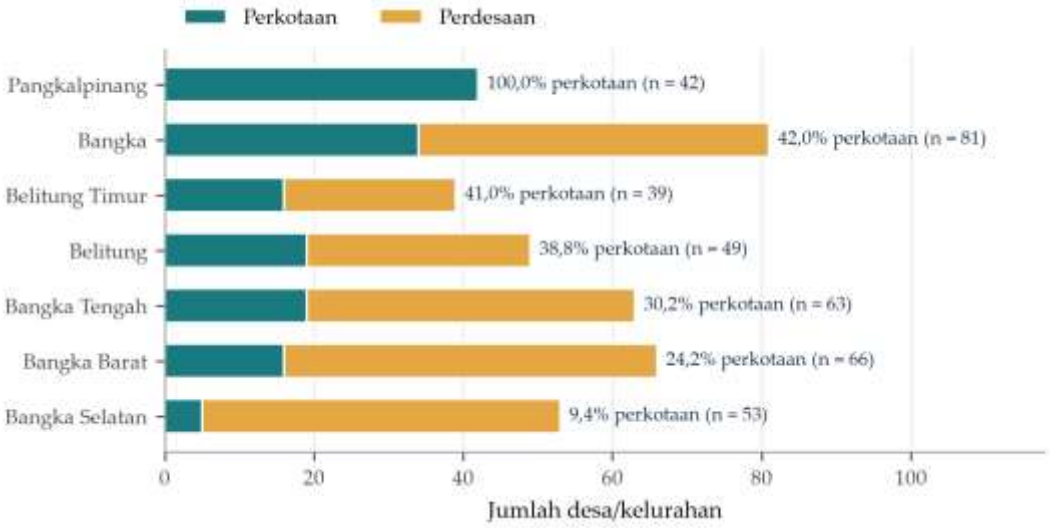
3. HASIL DAN PEMBAHASAN

3.1 Komposisi status perkotaan–perdesaan

Sebanyak 242 dari 393 desa/kelurahan (61,6%) berstatus perdesaan dan 151 (38,4%) berstatus perkotaan. Komposisi tersebut sangat timpang antarwilayah (Tabel 1 dan Gambar 2): seluruh desa/kelurahan di Kota Pangkalpinang berstatus perkotaan, sedangkan di Kabupaten Bangka Selatan hanya 9,4%. Proporsi perkotaan di lima kabupaten lain berada pada rentang 24,2–42,0%. Pada tingkat kecamatan, 7 dari 47 kecamatan seluruhnya perkotaan, 14 seluruhnya perdesaan, dan 26 campuran. Label R105 pada PODES berbeda dari status resmi 2025 pada 23 desa (15 desa berlabel perkotaan pada R105 kini perdesaan dan 8 sebaliknya); perbedaan ini menunjukkan bahwa R105 tidak dapat menggantikan status resmi 2025, tetapi tidak dapat dibaca sebagai perubahan status per desa yang terverifikasi.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 2.
Komposisi status perkotaan–perdesaan per kabupaten/kota



Catatan. Sumber: Perka BPS 268/2025 dan PODES 2025 (diolah); data sama dengan Tabel 1.

3.2 Perbedaan karakteristik desa perkotaan dan perdesaan

Tabel 3 dan Gambar 3 menyajikan 14 indikator persentase yang bervariasi. Perbedaan terbesar terdapat pada sumber penghasilan utama pertanian, yaitu 29,8% desa perkotaan berbanding 94,2% desa perdesaan (selisih 64,4 pp). Keberadaan apotek (74,8% berbanding 12,4%) dan minimarket/swalayan (84,1% berbanding 23,6%) menunjukkan selisih lebih dari 60 pp, diikuti letak di dalam atau di tepi kawasan hutan (41,7% berbanding 88,8%), kantor bank (40,4% berbanding 3,3%), dan penerangan jalan utama (74,2% berbanding 40,9%). Desa perdesaan juga lebih sering berbatasan dengan laut (48,8% berbanding 33,8%), sedangkan keberadaan angkutan umum trayek tetap praktis setara (76,8% berbanding 77,7%).

Tabel 3.
Persentase desa menurut status pada 14 indikator yang bervariasi, tingkat provinsi

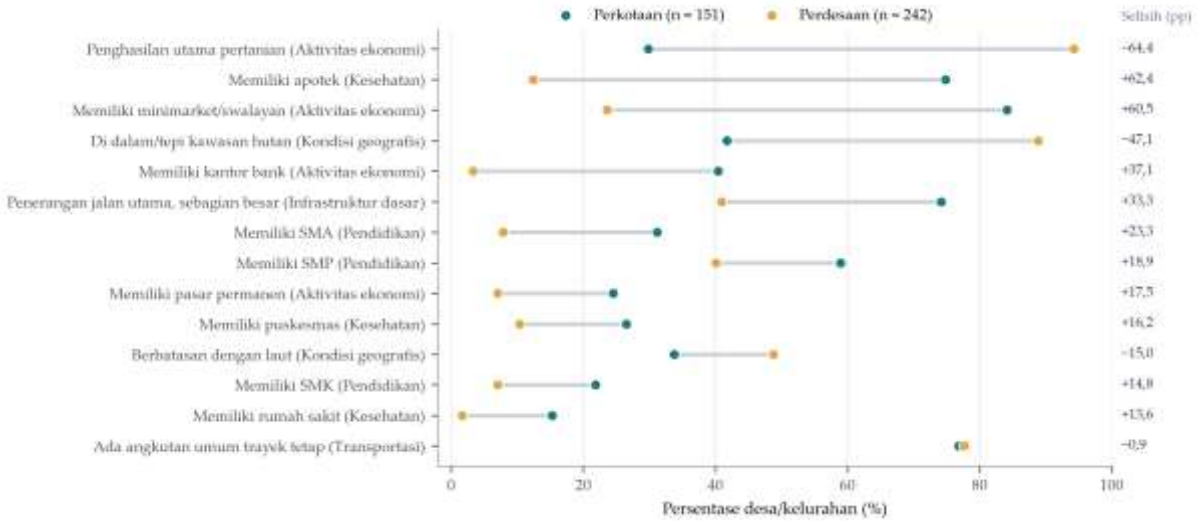
Kode KPI	Indikator	Domain	Perkotaan (%)	Perdesaan (%)	Provinsi (%)	Selisih (pp)
C07	Penghasilan utama pertanian	Aktivitas ekonomi	29,8	94,2	69,5	−64,4
H03	Memiliki apotek	Kesehatan	74,8	12,4	36,4	+62,4
C05	Memiliki minimarket/swalayan	Aktivitas ekonomi	84,1	23,6	46,8	+60,5
G02	Di dalam/tepi kawasan hutan	Kondisi geografis	41,7	88,8	70,7	−47,1
C01	Memiliki kantor bank	Aktivitas ekonomi	40,4	3,3	17,6	+37,1
I03	Penerangan jalan utama (sebagian besar)	Infrastruktur dasar	74,2	40,9	53,7	+33,3
E03	Memiliki SMA	Pendidikan	31,1	7,8	16,8	+23,3
E02	Memiliki SMP	Pendidikan	58,9	40,1	47,3	+18,9
C04	Memiliki pasar permanen	Aktivitas ekonomi	24,5	7,0	13,7	+17,5
H02	Memiliki puskesmas	Kesehatan	26,5	10,3	16,5	+16,2
G01	Berbatasan dengan laut	Kondisi geografis	33,8	48,8	43,0	−15,0

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Kode KPI	Indikator	Domain	Perkotaan (%)	Perdesaan (%)	Provinsi (%)	Selisih (pp)
E04	Memiliki SMK	Pendidikan	21,9	7,0	12,7	+14,8
H01	Memiliki rumah sakit	Kesehatan	15,2	1,6	6,9	+13,6
T03	Ada angkutan umum trayek tetap	Transportasi	76,8	77,7	77,3	−0,9

Catatan. Selisih = perkotaan − perdesaan (poin persen); tanda negatif berarti lebih tinggi di perdesaan. n perkotaan = 151, n perdesaan = 242. Sumber: PODES 2025 dan Perka BPS 268/2025 (diolah).

Gambar 3.
Perbedaan persentase desa perkotaan dan perdesaan pada 14 indikator, tingkat provinsi



Catatan. Selisih = perkotaan − perdesaan (poin persen). Data sama dengan Tabel 3.

Sebaliknya, sembilan indikator persentase hampir seragam di tingkat provinsi sehingga tidak membedakan status, antara lain keluarga pengguna listrik PLN (99,8%), desa yang memiliki sekolah dasar (98,7%), dan desa dengan sinyal internet 4G/5G (99,7%). Pada tingkat domain, rata-rata selisih absolut terbesar terdapat pada aktivitas ekonomi (44,9 pp dari empat indikator), disusul kondisi geografis (31,1 pp) dan kesehatan (30,7 pp); infrastruktur dasar hanya memiliki satu indikator yang bervariasi (33,3 pp), pendidikan 19,0 pp, transportasi 0,9 pp, dan seluruh indikator konektivitas digital hampir seragam. Rata-rata ini bersifat deskriptif dan bergantung pada indikator yang tersedia di setiap domain.

3.3 Variasi antarkabupaten/kota dalam status yang sama

Perbandingan antarkabupaten/kota yang tidak memperhitungkan status cenderung mencerminkan komposisi status. Karena itu, Tabel 4 dan Gambar 4 membandingkan empat indikator di dalam kelompok status yang sama setelah supresi. Di antara sel yang dapat ditampilkan, proporsi desa perdesaan yang memiliki apotek berkisar 6,4–20,5% dan yang memiliki minimarket 16,0–36,7%, sedangkan penghasilan utama pertanian di desa perdesaan tetap tinggi (89,4–100,0%). Di kelompok perkotaan, proporsi desa yang memiliki apotek berkisar 61,8–78,9%, tetapi penghasilan utama pertanian sangat beragam (2,4–57,9%). Sel Kabupaten Belitung Timur dan kelompok perdesaan Kabupaten Bangka Selatan disembunyikan melalui supresi sekunder, sedangkan kelompok perkotaan Bangka Selatan (5 desa) disembunyikan

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

melalui supresi primer. Dari 816 sel KPI provinsi dan kabupaten/kota × status, 50 sel terkena supresi primer, 101 sel supresi sekunder, dan 48 sel tidak berlaku.

Tabel 4.
Persentase desa pada empat indikator menurut kabupaten/kota di dalam status yang sama (setelah supresi)

	Desa perkotaan (%)				Desa perdesaan (%)			
Kabupaten/kota	**Pertanian**	**Apotek**	**Minimarket**	**Bank**	**Pertanian**	**Apotek**	**Minimarket**	**Bank**
Pangkalpinang	2,4	78,6	90,5	52,4	–	–	–	–
Bangka	41,2	61,8	67,6	26,5	89,4	6,4	23,4	6,4
Belitung Timur	d	d	d	d	d	d	d	d
Belitung	5,3	73,7	89,5	31,6	90,0	13,3	36,7	0,0
Bangka Tengah	57,9	78,9	78,9	36,8	100,0	20,5	20,5	4,5
Bangka Barat	50,0	68,8	93,8	50,0	98,0	8,0	16,0	4,0
Bangka Selatan	s	s	s	s	d	d	d	d

Catatan. s = disembunyikan karena kelompok < 10 desa (supresi primer); d = disembunyikan agar tidak dapat dihitung balik (supresi sekunder); – = tidak berlaku (tidak ada desa perdesaan). Pertanian = penghasilan utama pertanian; apotek, minimarket/swalayan, dan kantor bank = keberadaan di desa. Jumlah desa per kelompok tersedia pada Tabel 1.

Gambar 4.
Variasi antarkabupaten/kota di dalam status yang sama setelah supresi



Catatan. Sel berarsir disembunyikan; nilai yang disembunyikan tidak ditampilkan sebagai nol. Data sama dengan Tabel 4.

3.4 Kinerja model kandidat

Tabel 5 dan Gambar 5 merangkum kinerja lima model pada set A dengan skema terstratifikasi. Model pembanding kelas mayoritas menghasilkan MCC 0,000 dan PR-AUC 0,384, setara prevalensi kelas perkotaan. LR mencapai MCC 0,741 ± 0,079 dan RF 0,774 ± 0,066; bobot kelas seimbang menaikkan MCC menjadi 0,745 (LR) dan 0,788 (RF) serta meningkatkan recall perkotaan. Namun, tidak ada perbedaan yang bermakna secara statistik: LR dibandingkan RF menghasilkan p terkoreksi 0,236, bobot kelas pada RF p = 0,546, dan pada LR p = 0,871. Uji permutasi menunjukkan kinerja kedua model jauh di atas kebetulan: MCC LR 0,747

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

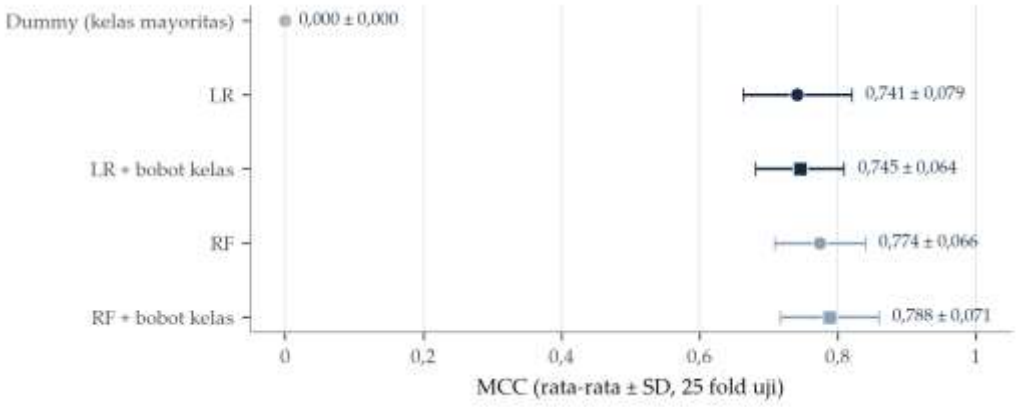
dibandingkan rata-rata 0,005 pada label acak (p = 0,0050; 200 permutasi) dan MCC RF 0,780 dibandingkan 0,002 (p = 0,0099; 100 permutasi). Nilai p tersebut merupakan nilai terkecil yang mungkin untuk jumlah permutasi yang dipakai.

Tabel 5.
Kinerja model kandidat pada set fitur A, validasi terstratifikasi (rata-rata 25 fold uji)

Model	MCC ± SD	Akurasi seimbang	Recall perkotaan	Presisi perkotaan	F1 perkotaan	PR-AUC	ROC-AUC
Dummy (kelas mayoritas)	0,000 ± 0,000	0,500	0,000	0,000	0,000	0,384	0,500
LR	0,741 ± 0,079	0,859	0,779	0,892	0,828	0,936	0,949
LR + bobot kelas	0,745 ± 0,064	0,865	0,802	0,877	0,835	0,934	0,949
RF	0,774 ± 0,066	0,878	0,815	0,900	0,853	0,948	0,962
RF + bobot kelas	0,788 ± 0,071	0,892	0,861	0,877	0,867	0,950	0,963

Catatan. Recall, presisi, dan F1 untuk kelas perkotaan. Uji t terkoreksi Nadeau–Bengio: LR vs RF p = 0,236; bobot kelas RF p = 0,546; bobot kelas LR p = 0,871. Model final belum dipilih.

Gambar 5.
MCC lima model kandidat pada set A, skema terstratifikasi (rata-rata ± SD, 25 fold uji)



Catatan. Data sama dengan Tabel 5.

3.5 Generalisasi wilayah dan sensitivitas set fitur

Ketika seluruh kecamatan uji tidak terlihat saat pelatihan, MCC rata-rata nyaris tidak berubah (LR 0,740; RF 0,764), tetapi simpangan baku antarfold membesar dan MCC fold terendah turun (Tabel 6 dan Gambar 6). Selisih LR dan RF pada skema berkelompok kecamatan juga tidak bermakna (p = 0,607). Hasil ini menunjukkan bahwa kinerja tidak semata-mata bertumpu pada kemiripan desa dalam kecamatan yang sama, tetapi ketidakpastiannya lebih besar.

Tabel 6.
Kinerja LR dan RF pada skema terstratifikasi dan berkelompok kecamatan (set A)

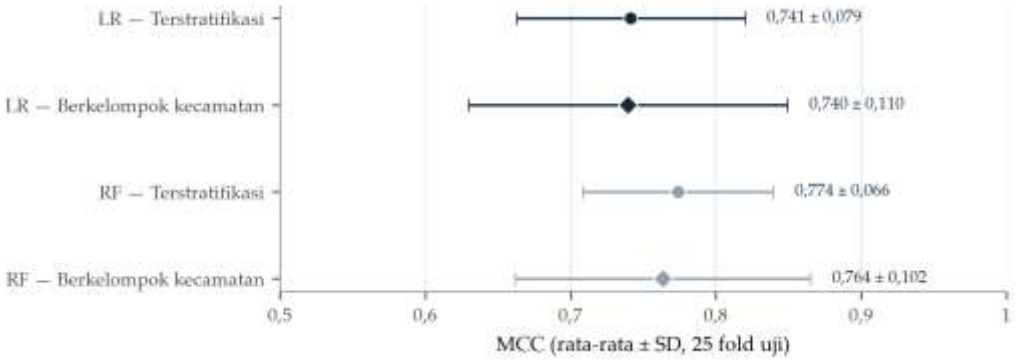
Model	Skema validasi	MCC ± SD	MCC minimum fold	Akurasi seimbang	PR-AUC	ROC-AUC
LR	Terstratifikasi	0,741 ± 0,079	0,484	0,859	0,936	0,949

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Model	Skema validasi	MCC ± SD	MCC minimum fold	Akurasi seimbang	PR-AUC	ROC-AUC
LR	Berkelompok kecamatan	0,740 ± 0,110	0,403	0,857	0,919	0,939
RF	Terstratifikasi	0,774 ± 0,066	0,597	0,878	0,948	0,962
RF	Berkelompok kecamatan	0,764 ± 0,102	0,516	0,871	0,941	0,957

Catatan. Berkelompok kecamatan: seluruh desa dalam satu kecamatan berada pada fold uji yang sama (47 kecamatan). LR vs RF: p = 0,236 (terstratifikasi) dan p = 0,607 (berkelompok kecamatan). Validasi lintas kabupaten/kota belum dijalankan.

Gambar 6.
MCC LR dan RF pada skema terstratifikasi dan berkelompok kecamatan



Catatan. Data sama dengan Tabel 6.

Tabel 7 menunjukkan bahwa penghapusan fitur yang beririsan konsep dengan kriteria resmi (set B dan C) hanya sedikit mengubah MCC, yaitu 0,732–0,762 untuk LR dan 0,740–0,773 untuk RF. Penghapusan R906A tidak menurunkan MCC secara bermakna (A dibandingkan E dan C dibandingkan D; p = 0,436–0,879), tetapi PR-AUC turun bermakna pada keempat perbandingan (p = 0,001–0,034). Dengan demikian, R906A terutama memperbaiki pemeringkatan probabilitas, sementara kemampuan klasifikasi pada ambang tetap sebagian besar dapat dikompensasi oleh variabel lain.

Tabel 7.
Sensitivitas kinerja terhadap set fitur (skema terstratifikasi)

Set	Jumlah fitur	Isi	MCC LR	MCC RF	PR-AUC LR	PR-AUC RF
A	51	Seluruh kandidat (utama; memuat fitur beririsan konsep dengan kriteria resmi)	0,741	0,774	0,936	0,948
B	40	A tanpa 11 fitur irisan konseptual langsung (sementara)	0,762	0,773	0,940	0,947
C	37	B tanpa R501A1, R905EK2, R905JK2 (sementara)	0,732	0,740	0,929	0,941
D	36	C tanpa R906A (eksploratif)	0,716	0,714	0,906	0,914
E	50	A tanpa R906A (eksploratif)	0,747	0,757	0,920	0,929

Catatan. Uji t terkoreksi untuk penghapusan R906A: MCC tidak berubah bermakna (A vs E dan C vs D; p = 0,436–0,879), PR-AUC turun bermakna (p = 0,001–0,034). Set B–E bukan set yang terbukti bebas kebocoran.

3.6 Faktor pembeda menurut model

Pada kedua model, R906A (sumber penghasilan utama sebagian besar penduduk) merupakan variabel dengan penurunan MCC terbesar saat diacak, yaitu 0,053 pada LR dan 0,154 pada RF (Tabel 8 dan Gambar 7). Variabel berikutnya adalah jumlah minimarket/swalayan

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

(R905EK2; 0,044 dan 0,030), jumlah keluarga pengguna listrik PLN pada LR (R501A1; 0,031), dan jarak ke SMP terdekat pada RF (R701EK4; 0,019). Secara agregat, domain aktivitas ekonomi paling informatif (LR 0,112; RF 0,208), sedangkan domain lain berada di bawah 0,03. Simpangan baku yang besar untuk sebagian besar variabel selain R906A menunjukkan bahwa peringkat di bawah posisi teratas kurang stabil, antara lain karena variabel yang berkorelasi saling berbagi kontribusi.

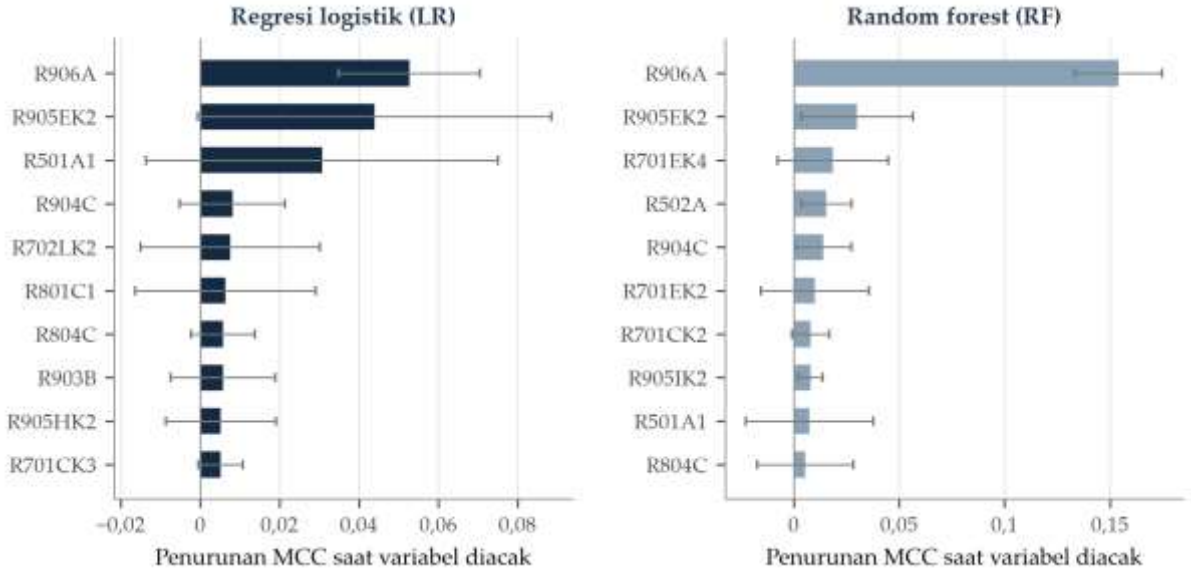
Tabel 8.
Sepuluh variabel dengan permutation importance tertinggi (set A, skema terstratifikasi)

Model	No.	Variabel	Label (ringkas)	Domain	Penurunan MCC (rata-rata)	SD
LR	1	R906A	Sumber penghasilan utama sebagian besar penduduk (lapangan usaha)	Aktivitas ekonomi	0,053	0,018
LR	2	R905EK2	Jumlah minimarket/swalayan/supermarket	Aktivitas ekonomi	0,044	0,045
LR	3	R501A1	Jumlah keluarga pengguna listrik PLN	Infrastruktur dasar	0,031	0,044
LR	4	R904C	Fasilitas kredit yang diterima warga: Kredit Usaha Kecil	Aktivitas ekonomi	0,008	0,013
LR	5	R702LK2	Jumlah apotek	Kesehatan	0,008	0,023
LR	6	R801C1	Keberadaan angkutan umum	Transportasi	0,006	0,023
LR	7	R804C	Sinyal telepon seluler di sebagian besar wilayah	Konektivitas digital	0,006	0,008
LR	8	R903B	Jumlah koperasi aktif: Kopinkra/usaha mikro	Aktivitas ekonomi	0,006	0,013
LR	9	R905HK2	Jumlah hotel	Aktivitas ekonomi	0,005	0,014
LR	10	R701CK3	Jumlah SD swasta	Pendidikan	0,005	0,006
RF	1	R906A	Sumber penghasilan utama sebagian besar penduduk (lapangan usaha)	Aktivitas ekonomi	0,154	0,021
RF	2	R905EK2	Jumlah minimarket/swalayan/supermarket	Aktivitas ekonomi	0,030	0,027
RF	3	R701EK4	Jarak ke SMP terdekat (desa tanpa SMP)	Pendidikan	0,019	0,026
RF	4	R502A	Penerangan di jalan utama desa/kelurahan	Infrastruktur dasar	0,015	0,012
RF	5	R904C	Fasilitas kredit yang diterima warga: Kredit Usaha Kecil	Aktivitas ekonomi	0,014	0,013
RF	6	R701EK2	Jumlah SMP negeri	Pendidikan	0,010	0,026
RF	7	R701CK2	Jumlah SD negeri	Pendidikan	0,008	0,009
RF	8	R905IK2	Jumlah penginapan (hostel/motel/losmen/wisma)	Aktivitas ekonomi	0,008	0,006
RF	9	R501A1	Jumlah keluarga pengguna listrik PLN	Infrastruktur dasar	0,008	0,030
RF	10	R804C	Sinyal telepon seluler di sebagian besar wilayah	Konektivitas digital	0,005	0,023

Catatan. Importance = penurunan MCC rata-rata pada data uji saat nilai variabel diacak (5 fold, 10 pengulangan). Nilai menggambarkan kontribusi prediktif, bukan hubungan sebab-akibat. Label diringkas dari kamus PODES 2025.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 7.
Sepuluh variabel dengan permutation importance tertinggi pada LR dan RF (rata-rata ± SD)



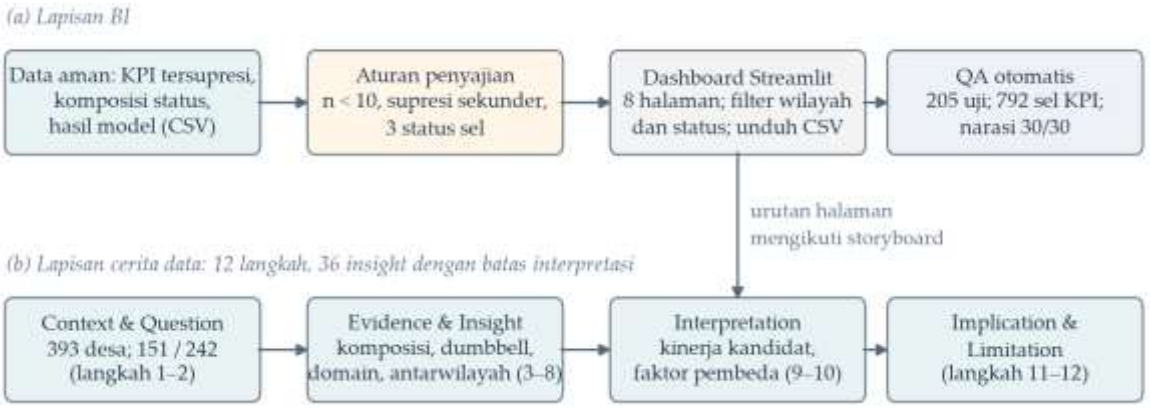
Catatan. Label variabel pada Tabel 8.

3.7 Dashboard intelijen bisnis dan cerita data

Dashboard yang dihasilkan terdiri atas delapan halaman yang urutannya mengikuti alur cerita data (Gambar 8 dan Tabel 9). Halaman Ringkasan membuka konteks dengan jumlah desa dan komposisi status; halaman Perkotaan vs Perdesaan menyajikan grafik dumbbell 14 indikator dan ringkasan tujuh domain; halaman Antarwilayah menampilkan heatmap indikator dalam status yang sama dengan sel tersupresi berwarna abu-abu; halaman Peta Spasial memberi orientasi geografis tingkat kabupaten/kota; halaman Kinerja Model dan Faktor Pembeda menyajikan hasil model dengan label kandidat; sedangkan halaman Kamus Indikator serta Metode & Keterbatasan menyediakan definisi, sumber, dan batas interpretasi. Storyboard dua belas langkah memetakan setiap halaman ke tahap cerita, dan katalog berisi 36 insight (28 fakta dan 8 interpretasi) yang masing-masing disertai batas interpretasi.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 8.
Lapisan BI dan alur cerita data pada dashboard



Catatan. Angka jaminan mutu berasal dari laporan QA proyek (Tabel 9).

Tabel 9.
Ringkasan dashboard intelijen bisnis dan hasil jaminan mutu

Aspek	Hasil
Halaman dashboard	8 (Ringkasan; Perkotaan vs Perdesaan; Antarwilayah; Peta Spasial; Kinerja Model; Faktor Pembeda; Kamus Indikator; Metode & Keterbatasan)
KPI terhitung	34 KPI dalam 7 domain (ditambah 2 indikator konteks: jumlah desa dan % perkotaan); 3 indikator kriteria resmi tidak dapat dihitung
Tingkat agregasi tampil	Provinsi dan kabupaten/kota × status (Semua/Perkotaan/Perdesaan)
Sel KPI kabupaten/kota × status	816 sel; 50 n < 10; 101 disembunyikan (supresi sekunder); 48 tidak berlaku
Katalog insight	36 insight (28 fakta, 8 interpretasi), masing-masing dengan sumber dan batas interpretasi
Storyboard	12 langkah: konteks → pertanyaan → bukti → insight → interpretasi → implikasi → keterbatasan
Pengujian otomatis	205 uji lulus, 0 gagal
Konsistensi KPI	792 sel tampil cocok dengan perhitungan independen
Konsistensi narasi	30/30 teks narasi cocok dengan angka sumber
Dokumentasi visual	15 tangkapan layar (desktop, tablet, ponsel)
Evaluasi pengguna	Belum dilakukan (tidak ada klaim usability atau dampak keputusan)

Catatan. Dashboard dijalankan lokal (127.0.0.1). Peta Spasial memakai batas wilayah BIG (referensi 2023) dan tidak dimuat dalam naskah karena ketentuan penggunaan belum dikonfirmasi.

Jaminan mutu menunjukkan 205 uji otomatis lulus, 792 sel KPI yang tampil cocok dengan perhitungan independen, dan 30 dari 30 teks narasi cocok dengan angka sumber. Hasil ini menunjukkan konsistensi teknis antara data, tampilan, dan narasi, tetapi belum menunjukkan bahwa dashboard mudah digunakan atau memperbaiki keputusan pengguna.

3.8 Pembahasan

Temuan deskriptif dan prediktif mengarah pada pola yang sama: perbedaan desa perkotaan dan perdesaan di Kepulauan Bangka Belitung paling tajam pada struktur ekonomi dan layanan berbasis pasar, bukan pada layanan dasar yang sudah hampir merata. Pola ini sejalan dengan temuan Ridwan dan Tanur (2026) bahwa sektor ekonomi dominan termasuk prediktor terkuat status pembangunan desa, serta dengan Br. Saragih dan Hartojo (2026) yang

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

menempatkan RF sebagai algoritma terbaik untuk klasifikasi perkotaan–perdesaan tingkat desa. Perbedaannya, penelitian ini memakai status resmi terbaru sebagai target, melaporkan MCC dengan validasi silang bersarang berulang, dan menguji generalisasi ke kecamatan yang tidak terlihat; dengan desain tersebut, keunggulan RF atas LR tidak terbukti bermakna.

Kedekatan R906A dengan kriteria resmi persentase keluarga pertanian perlu dibaca hati-hati. R906A adalah kategori penghasilan mayoritas desa, bukan persentase keluarga pertanian, sehingga kedekatannya merupakan irisan konsep dan bukan bukti kebocoran langsung. Meski demikian, sejalan dengan peringatan Kapoor dan Narayanan (2023), hasil set A diperlakukan sebagai batas atas; analisis sensitivitas menunjukkan bahwa MCC relatif stabil tanpa fitur beririsan, tetapi tidak membuktikan bahwa set mana pun bebas kebocoran.

Dari sisi BI, temuan antarwilayah menunjukkan bahwa perbandingan sebaiknya dilakukan di dalam kelompok status yang sama, karena perbedaan antarkabupaten/kota banyak dipengaruhi komposisi status. Penyajian dengan supresi primer dan sekunder membuat sebagian sel tidak dapat ditampilkan, sehingga dashboard secara eksplisit membedakan nilai yang disembunyikan dari nilai nol. Desain cerita data yang menambahkan konteks, alur, dan implikasi pada visualisasi sejalan dengan kerangka S-DIKW (Lo Duca & McDowell, 2025), dan bukti bahwa cerita data dapat mempercepat pemahaman (Shao dkk., 2024) menjadi alasan penyusunan alur tersebut. Namun, klaim tentang manfaat bagi pengguna memerlukan evaluasi usability dan pengujian keputusan seperti yang dilakukan Hjelle dkk. (2024) serta Ansari dan Martin (2022, 2024).

Penelitian ini memiliki sejumlah keterbatasan. Analisis bersifat potong lintang dan deskriptif, sehingga tidak ada klaim sebab-akibat. Status 2025 disusun dari Podes 2024, sedangkan prediktor berasal dari PODES 2025, dan dua dari tiga indikator kriteria resmi tidak tersedia sehingga metode BPS tidak direplikasi. LR dan RF masih merupakan model kandidat; model final belum dipilih, prediksi per desa di luar fold tidak tersedia, dan validasi lintas kabupaten/kota belum dijalankan. Sebagian nilai kabupaten/kota × status disembunyikan, peta batas wilayah tidak dimuat dalam naskah karena ketentuan penggunaan batas wilayah dan izin penyajian data belum dikonfirmasi, dan dashboard belum diuji dengan pengguna. Penelitian lanjutan dapat mengevaluasi usability dashboard, menguji generalisasi ke provinsi lain, dan memakai data kriteria resmi bila tersedia.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa status perkotaan–perdesaan resmi 2025 di Provinsi Kepulauan Bangka Belitung berasosiasi kuat dengan karakteristik desa dalam PODES 2025, terutama pertanian sebagai penghasilan utama dan kehadiran layanan berbasis pasar seperti apotek, minimarket, dan bank, sedangkan layanan dasar sudah hampir merata. Variasi antarkabupaten/kota tetap ada di dalam status yang sama, sehingga perbandingan wilayah perlu dikendalikan menurut status. Regresi logistik dan random forest mengenali status resmi jauh di atas model pembanding dan relatif stabil ketika kecamatan uji tidak terlihat, tetapi selisih keduanya tidak bermakna dan keduanya masih merupakan model kandidat. Dashboard intelijen bisnis delapan halaman dengan alur cerita data dua belas langkah dapat menyajikan temuan tersebut secara konsisten dengan kontrol pengungkapan. Penelitian selanjutnya disarankan mengevaluasi dashboard bersama pengguna, menguji generalisasi lintas wilayah, dan memastikan izin publikasi data sebelum hasil disebarluaskan.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Provinsi Kepulauan Bangka Belitung yang telah menyediakan data Pendataan Potensi Desa 2025 untuk keperluan penelitian dan analisis ini. Seluruh pengolahan, analisis, dan interpretasi dalam artikel ini merupakan tanggung jawab penulis.

Pendanaan: [PERNYATAAN PENDANAAN — diisi penulis. Bila tidak ada hibah, gunakan kalimat baku jurnal: "Studi ini tidak menerima hibah khusus dari lembaga pendanaan di sektor publik, komersial, atau nirlaba."]

Konflik kepentingan: [PERNYATAAN KONFLIK KEPENTINGAN — diisi penulis.]

Ketersediaan data: data PODES 2025 yang digunakan dalam penelitian ini berada di bawah ketentuan perjanjian penggunaan data dengan BPS; rekaman data bersifat rahasia, menjadi milik BPS, dan tidak dibagikan oleh penulis maupun melalui repositori terbuka. Akses terhadap data tersebut perlu mengikuti ketentuan layanan data BPS. Hasil yang dipublikasikan dalam artikel ini terbatas pada agregat tingkat provinsi dan kabupaten/kota yang telah melalui supresi sel.

Etika dan kerahasiaan data: artikel ini tidak memuat data, nama, atau prediksi tingkat desa. Indikator hanya disajikan pada tingkat provinsi dan kabupaten/kota menurut status, dengan supresi sel primer (kurang dari 10 desa) dan supresi sekunder sebagaimana diuraikan pada bagian Metode.

DAFTAR PUSTAKA

Almasi, S., Bahaadinbeigy, K., Ahmadi, H., Sohrabei, S., & Rabiei, R. (2023). Usability evaluation of dashboards: A systematic literature review of tools. *BioMed Research International, 2023*, Article 9990933. https://doi.org/10.1155/2023/9990933

Ansari, B., & Martin, E. G. (2022). Development of a usability checklist for public health dashboards to identify violations of usability principles. *Journal of the American Medical Informatics Association, 29*(11), 1847–1858. https://doi.org/10.1093/jamia/ocac140

Ansari, B., & Martin, E. G. (2024). Integrating human-centered design in public health data dashboards: Lessons from the development of a data dashboard of sexually transmitted infections in New York State. *Journal of the American Medical Informatics Association, 31*(2), 298–305. https://doi.org/10.1093/jamia/ocad102

Badan Pusat Statistik. (2025). *Peraturan Kepala Badan Pusat Statistik Nomor 268 Tahun 2025 tentang Klasifikasi Desa Perkotaan dan Perdesaan di Indonesia 2025, Buku 1 Sumatera* (Katalog 1204028). Badan Pusat Statistik.

Br. Saragih, M. T., & Hartojo, N. (2026). Integrating geospatial big data and machine learning for village level rural urban classification: Evidence from Toba Regency. *Engineering, MAthematics and Computer Science Journal (EMACS), 8*(1), 27–36. https://doi.org/10.21512/emacsjournal.v8i1.14159

Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: A review from a data perspective. *Journal of Big Data, 12*, Article 71. https://doi.org/10.1186/s40537-025-01119-4

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Chen, Q., Cao, S., Wang, J., & Cao, N. (2024). How does automation shape the process of narrative visualization: A survey of tools. *IEEE Transactions on Visualization and Computer Graphics, 30*(8), 4429–4448. https://doi.org/10.1109/TVCG.2023.3261320

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. *Artificial Intelligence Review, 57*(6), Article 137. https://doi.org/10.1007/s10462-024-10759-6

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*, Article 6. https://doi.org/10.1186/s12864-019-6413-7

Hjelle, S., Mikalef, P., Altwaijry, N., & Parida, V. (2024). Organizational decision making and analytics: An experimental study on dashboard visualizations. *Information & Management, 61*(6), Article 104011. https://doi.org/10.1016/j.im.2024.104011

Hu, S., Ge, Y., Liu, M., Ren, Z., & Zhang, X. (2022). Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *International Journal of Applied Earth Observation and Geoinformation, 107*, Article 102694. https://doi.org/10.1016/j.jag.2022.102694

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4*(9), Article 100804. https://doi.org/10.1016/j.patter.2023.100804

Lo Duca, A., & McDowell, K. (2025). Using the S-DIKW framework to transform data visualization into data storytelling. *Journal of the Association for Information Science and Technology, 76*(5), 803–818. https://doi.org/10.1002/asi.24973

Mörth, E., Bruckner, S., & Smit, N. N. (2023). ScrollyVis: Interactive visual authoring of guided dynamic narratives for scientific scrollytelling. *IEEE Transactions on Visualization and Computer Graphics, 29*(12), 5165–5177. https://doi.org/10.1109/TVCG.2022.3205769

Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning, 52*(3), 239–281. https://doi.org/10.1023/A:1024068626366

O'Brien Quinn, H., Sedky, M., Francis, J., & Streeton, M. (2024). Literature review of explainable tabular data analysis. *Electronics, 13*(19), Article 3806. https://doi.org/10.3390/electronics13193806

Ridwan, M., & Tanur, E. (2026). Classification of village development status in Bekasi Regency using ensemble learning and SMOTE-based class balancing. *Jurnal Aplikasi Statistika & Komputasi Statistik, 18*(1), 76–97. https://doi.org/10.34123/jurnalasks.v18i1.870

Shao, H., Martinez-Maldonado, R., Echeverria, V., Yan, L., & Gašević, D. (2024). Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension? In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (pp. 1–21). Association for Computing Machinery. https://doi.org/10.1145/3613904.3643022

Steinert, S., Ruf, V., Dzsotjan, D., Großmann, N., Schmidt, A., Kuhn, J., & Küchemann, S. (2024). A refined approach for evaluating small datasets via binary classification using machine learning. *PLOS ONE, 19*(5), Article e0301276. https://doi.org/10.1371/journal.pone.0301276

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Wang, H., Yu, X., Luo, L., & Li, R. (2024). Urban–rural boundary delineation based on population spatialization: A case study of Guizhou Province, China. *Sustainability, 16*(5), Article 1787. https://doi.org/10.3390/su16051787

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)