

Pengembangan Model *Government Decision Intelligence* (IC-Gov) Untuk Mendukung Pengambilan Kebijakan Pemerintah Berbasis Data

Mohamad Jamil^{1*}, Iis Hamsir Ayub Wahab², Seh Turuy³

^{1*,3} Progran Studi Teknik Informatika, Fakultas Teknik, Universitas Khairun, Ternate, Indonesia

² Progran Studi Teknik Elektro, Fakultas Teknik, Universitas Khairun, Ternate, Indonesia

^{1*}jamikhairun@gmail.com, ²hamsir@unkhair.ac.id, ³seh.turuy26@gmail.com

*) jamikhairun@gmail.com

Abstrak—Transformasi digital pemerintahan memerlukan dukungan analitik yang tidak hanya mampu mengolah data, tetapi juga menghasilkan informasi prediktif yang dapat digunakan dalam proses pengambilan kebijakan. Penelitian ini mengembangkan **IC-Gov** sebagai pendekatan *Government Decision Intelligence* yang mengintegrasikan *Machine Learning*, *evidence retrieval*, dan sintesis berbasis kecerdasan artifisial untuk mendukung analisis prioritas pembangunan daerah. Pemodelan menggunakan empat indikator, yaitu jumlah *blankspot* 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita, dengan target klasifikasi prioritas rendah, sedang, dan tinggi. Tiga algoritma, yaitu Random Forest, XGBoost, dan Gradient Boosting, dibandingkan menggunakan *accuracy*, *precision macro*, *recall macro*, *F1-score macro*, *confusion matrix*, dan ROC-AUC. Hasil pengujian menunjukkan Random Forest memperoleh *accuracy*, *precision macro*, *recall macro*, dan *F1-score macro* sebesar 1,00, diikuti Gradient Boosting dengan *accuracy* 0,95 dan *F1-score macro* 0,95, serta XGBoost dengan *accuracy* 0,87 dan *F1-score macro* 0,85. Nilai ROC-AUC masing-masing model mencapai 1,000, 1,000, dan 0,992. Analisis *feature importance* menunjukkan bahwa PDRB per kapita memiliki kontribusi relatif terbesar pada ketiga model. Hasil prediksi selanjutnya diintegrasikan dalam IC-Gov sebagai *predictive insight* yang dipadukan dengan *evidence* dan sintesis berbasis AI untuk menghasilkan informasi analitis yang lebih kontekstual. Penelitian ini menunjukkan bahwa integrasi analitik prediktif dan *evidence* dalam kerangka *Government Decision Intelligence* berpotensi memperkuat dukungan terhadap pengambilan kebijakan pembangunan daerah berbasis data tanpa menggantikan peran pengambil keputusan.

Kata Kunci: *Government Decision Intelligence*, IC-Gov, *Machine Learning*, *Predictive Insight*, Kebijakan Pembangunan Daerah.

Abstract—Digital government transformation requires analytical capabilities that not only process data but also generate predictive information to support policy decision-making. This study develops **IC-Gov** as a *Government Decision Intelligence* approach that integrates *Machine Learning*, *evidence retrieval*, and AI-based synthesis to support the analysis of regional development priorities. The modeling process employs four development indicators: the number of 4G *blank spots*, the percentage of roads in good condition, the education index, and GRDP per capita, with development priorities classified into low, medium, and high categories. Three algorithms—Random Forest, XGBoost, and Gradient Boosting—were evaluated using accuracy, macro precision, macro recall, macro F1-score, confusion matrix, and ROC-AUC. The results show that Random Forest achieved an accuracy, macro precision, macro recall, and macro F1-score of 1.00, followed by Gradient Boosting with an accuracy of 0.95 and a macro F1-score of 0.95, and XGBoost with an accuracy of 0.87 and a macro F1-score of 0.85. The ROC-AUC values were 1.000, 1.000, and 0.992, respectively. *Feature importance* analysis indicates that GRDP per capita has the highest relative contribution across all three models. The prediction results are subsequently integrated into IC-Gov as *predictive insights*, combined with supporting *evidence* and AI-based synthesis to generate more contextual analytical information. The findings demonstrate that integrating predictive analytics and *evidence* within a *Government Decision Intelligence* framework has the potential to strengthen data-driven regional development policy decision support while maintaining human decision-makers as the final authority.

Keywords: *Government Decision Intelligence*, IC-Gov, *Machine Learning*, *Predictive Insight*, Regional Development Policy.

1. PENDAHULUAN

Transformasi digital pemerintahan telah berkembang dari digitalisasi proses administratif menuju pemanfaatan data dan kecerdasan artifisial untuk mendukung tata kelola dan pengambilan kebijakan. OECD menempatkan *data-driven public sector* sebagai salah satu dimensi penting pemerintahan digital, dengan data sebagai aset strategis dalam perumusan kebijakan dan penyelenggaraan layanan [1], [2].

Di Indonesia, orientasi tersebut diperkuat melalui kebijakan Satu Data Indonesia yang mendorong tersedianya data pemerintah yang akurat, mutakhir, terpadu, dapat dipertanggungjawabkan, serta dapat dibagipakaikan antarinstansi [3]. Meskipun demikian, data pemerintahan saat ini masih tersebar pada berbagai aplikasi, basis data, *Application Programming Interface* (API), dokumen elektronik, dan sistem informasi sektoral. Heterogenitas tersebut menyebabkan informasi yang dibutuhkan pengambil kebijakan sering berada pada sumber yang berbeda. OECD menegaskan bahwa fragmentasi sistem dan proses pemerintahan dapat menghambat kemampuan pemerintah untuk bertindak secara cepat dan responsif [4], [5]. Kondisi ini menunjukkan adanya kesenjangan antara ketersediaan data (*data availability*) dan kemampuan memanfaatkannya untuk menghasilkan keputusan (*decision capability*). Tantangannya bukan hanya mengintegrasikan dan menampilkan data, tetapi mengubah data menjadi *evidence* dan *insight* yang relevan bagi pengambilan kebijakan.

Perkembangan *Large Language Models* (LLM) menawarkan pendekatan baru untuk mengatasi kesenjangan tersebut melalui kemampuan interaksi bahasa alami, sintesis informasi, dan analisis kontekstual [6], [7]. Namun, penggunaan LLM secara langsung memiliki keterbatasan karena pengetahuan model tidak selalu merepresentasikan data pemerintahan yang aktual dan keluaran generatif berpotensi menghasilkan informasi yang tidak didukung fakta. *Retrieval-Augmented Generation* (RAG) menjadi salah satu pendekatan untuk mengatasi keterbatasan tersebut dengan menghubungkan model bahasa dengan sumber pengetahuan eksternal [8], [9], [10]. RAG memungkinkan model memanfaatkan informasi eksternal untuk menghasilkan keluaran yang lebih spesifik dan berbasis sumber. Namun, kebutuhan pengambilan kebijakan tidak berhenti pada kemampuan menemukan informasi [11], [12]. Sistem perlu memahami persoalan, memilih sumber data yang relevan, menghubungkan *evidence*, mempertahankan konteks pertanyaan, serta menghasilkan informasi yang dapat digunakan sebagai dasar pertimbangan keputusan [13], [14]. Hal ini menunjukkan peluang pengembangan RAG sebagai bagian dari pendekatan *Decision Intelligence* yang mengintegrasikan data, analitik, kecerdasan artifisial, dan proses keputusan [15], [16].

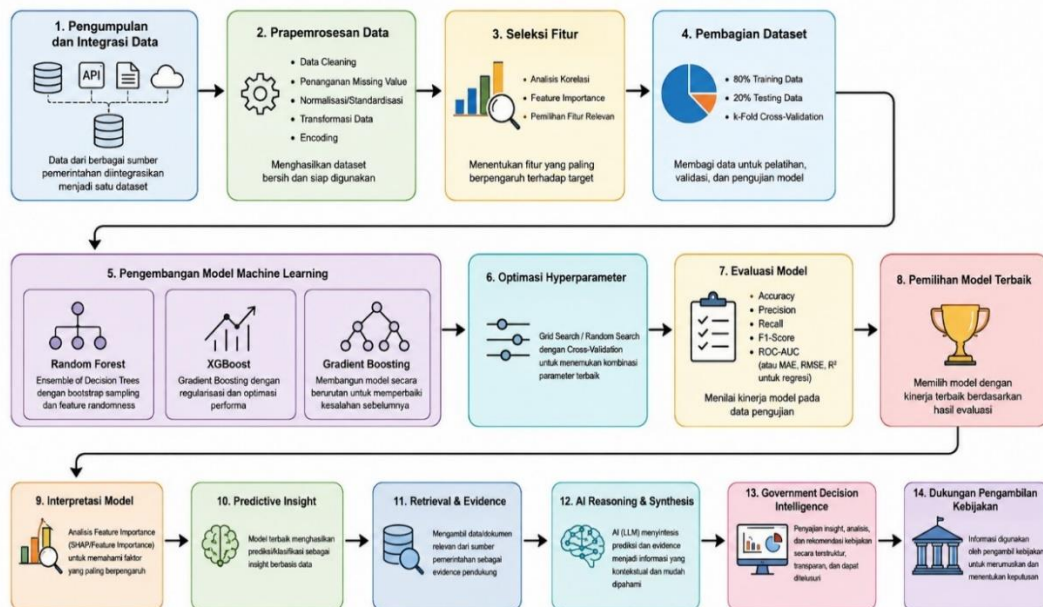
Penerapan *Decision Intelligence* dalam pemerintahan memiliki tantangan yang lebih kompleks. Selain sumber data yang heterogen, pertanyaan kebijakan sering bersifat kontekstual dan *multi-turn*. Pengguna dapat mengajukan pertanyaan mengenai suatu indikator, kemudian melanjutkan dengan pertanyaan tentang perbandingan wilayah, tren, kesenjangan, atau prioritas intervensi tanpa mengulangi konteks sebelumnya. Sistem karena itu perlu mempertahankan konteks sekaligus memastikan bahwa informasi yang dihasilkan dapat ditelusuri ke sumber datanya. Kebutuhan ini sejalan dengan RAGAS yang menekankan pentingnya relevansi konteks, *faithfulness*, dan kualitas jawaban dalam evaluasi sistem berbasis RAG [6], [17].

Penelitian sebelumnya telah membahas RAG untuk meningkatkan akses terhadap pengetahuan [18], [19] evaluasi kualitas *retrieval* dan *generation* [20], [21], serta penggunaan RAG dalam persoalan pengambilan keputusan. Namun, masih terdapat ruang untuk mengintegrasikan data pemerintahan multi-sumber, contextual reasoning, evidence traceability, dan Decision Intelligence dalam satu model yang berorientasi pada kebutuhan kebijakan pemerintah. Oleh karena itu, penelitian ini bertujuan mengembangkan dan mengevaluasi model Government Decision Intelligence (IC-Gov) untuk mendukung pengambilan kebijakan pemerintah berbasis data. Kontribusi penelitian terletak pada pengembangan arsitektur yang menempatkan AI sebagai lapisan analitis untuk mentransformasikan data pemerintahan menjadi *evidence* dan *insight* yang relevan, kontekstual, serta dapat ditelusuri dalam mendukung proses pengambilan kebijakan.

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan model Government Decision Intelligence (IC-Gov) yang mengintegrasikan sumber data pemerintahan, *intelligent retrieval*, pengelolaan konteks, kecerdasan artifisial, dan *evidence traceability*. IC-Gov tidak diposisikan sebagai chatbot atau dashboard, tetapi sebagai lapisan kecerdasan (*intelligence layer*) di atas ekosistem data dan sistem informasi pemerintahan. Model ini dirancang untuk memahami kebutuhan informasi, menentukan sumber data yang relevan, memperoleh *evidence*, melakukan analisis berbantuan AI, dan menghasilkan informasi yang dapat ditelusuri kembali ke sumbernya.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan pengembangan model berbasis Machine Learning (ML) untuk membangun *Government Decision Intelligence* yang mampu menghasilkan *predictive insight* dari data pemerintahan. Model dikembangkan dengan mengintegrasikan sumber data pemerintahan, proses prapemrosesan data, pemodelan prediktif, pemilihan model terbaik, serta pemanfaatan hasil prediksi sebagai bagian dari dukungan pengambilan kebijakan. Secara umum, tahapan penelitian meliputi pengumpulan dan integrasi data, prapemrosesan data, pemilihan fitur, pengembangan model *Machine Learning*, pelatihan dan validasi model, evaluasi kinerja, serta integrasi hasil prediksi ke dalam *Government Decision Intelligence*. Gambar 1 diperlihatkan alur tahapan penelitian.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan dan Integrasi Data

Data penelitian bersumber dari Portal Satu Data Maluku Utara yang menyediakan berbagai dataset sektoral Pemerintah Provinsi Maluku Utara melalui katalog *open data*. Dataset yang digunakan dipilih untuk merepresentasikan beberapa dimensi pembangunan daerah, yaitu infrastruktur digital, infrastruktur dasar, pendidikan, dan ekonomi. Variabel yang digunakan meliputi jumlah *blankspot* 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita. Data diperoleh dalam format terstruktur, termasuk CSV, Excel, dan JSON. Khusus indikator ekonomi, data PDRB per kapita dilengkapi menggunakan data resmi Badan Pusat Statistik (BPS) Provinsi Maluku Utara. Selanjutnya, data dari berbagai sumber diintegrasikan melalui proses penyelarasan nama wilayah, periode pengamatan, atribut, dan tipe data. Kabupaten/kota dan tahun digunakan sebagai atribut penghubung (*key*) dalam proses integrasi sehingga setiap indikator dapat disusun dalam satu struktur dataset. Tabel 1 diperlihatkan sampel dataset indikator pembangunan daerah yang digunakan dalam penelitian.

Tabel 1. Sampel Dataset Sampel Dataset Indikator Pembangunan Daerah

Kabupaten/Kota	Blankspot 4G (2025)	Jalan Baik (2025, %)	Indeks Pendidikan (2025)	PDRB/Kapita (2024, Juta Rp)
Halmahera Barat	2	69,567	10,93	21,39
Kepulauan Sula	24	10,077	11,18	30,92
Halmahera Selatan	29	32,168	10,74	70,66
Halmahera Utara	15	70,663	11,325	34,57
Kota Ternate	6	68,208	14,01	70,95

2.2 Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan dataset hasil integrasi memiliki kualitas dan struktur yang sesuai sebelum digunakan dalam pemodelan *Machine Learning*. Proses ini meliputi pemeriksaan data duplikat, penanganan *missing value*, penyalarsan tipe dan format data, serta identifikasi nilai yang tidak konsisten pada setiap indikator. Selanjutnya dilakukan transformasi data untuk menyamakan skala dan representasi variabel apabila diperlukan. Data juga diperiksa berdasarkan wilayah dan tahun untuk memastikan setiap observasi memiliki keterkaitan yang konsisten antarindikator. Hasil tahap ini berupa dataset yang telah bersih dan terstruktur sehingga dapat digunakan pada proses pemilihan fitur dan pengembangan model prediktif. Tabel 2 disajikan contoh hasil Prapemrosesan

Tabel 2. Hasil Prapemrosesan variabel Penelitian

Variabel	Format Awal	Hasil Prapemrosesan	Tipe Data
Blankspot 4G	2; 24; 29	2; 24; 29	Integer
Jalan Kondisi Baik	69,567%	69.567	Float
Indeks Pendidikan	10,93	10.930	Float
PDRB per Kapita	Rp21,39 juta	21.390	Float

2.3 Penentuan Variabel Target Prediksi

Tahap ini dilakukan untuk menentukan variabel masukan (*features*) dan variabel target yang digunakan dalam pemodelan *Machine Learning*. Variabel masukan merupakan indikator pembangunan daerah yang telah melalui proses integrasi dan prapemrosesan, yaitu jumlah *blankspot* 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita. Keempat variabel tersebut merepresentasikan dimensi infrastruktur digital, infrastruktur dasar, pendidikan, dan ekonomi. Variabel target (Y) ditetapkan sebagai prioritas pembangunan daerah yang diklasifikasikan ke dalam tiga kelas, yaitu tinggi, sedang, dan rendah. Penetapan kelas target tidak ditentukan secara langsung dari nilai variabel masukan, tetapi mengacu pada prioritas pembangunan yang tercantum dalam dokumen perencanaan daerah atau hasil validasi ahli. Pendekatan ini digunakan untuk memperoleh *ground truth* yang menjadi acuan dalam proses pelatihan dan evaluasi model.

Dataset penelitian selanjutnya direpresentasikan sebagai:

$X = \{\text{Blankspot 4G, Jalan Kondisi Baik, Indeks Pendidikan, PDRB per Kapita}\}$

$Y = \{\text{Prioritas Tinggi, Prioritas Sedang, Prioritas Rendah}\}$

Variabel X digunakan sebagai fitur prediktor, sedangkan variabel Y digunakan sebagai label kelas. Dataset tersebut selanjutnya digunakan dalam proses pelatihan dan pengujian Random Forest, XGBoost, dan Gradient Boosting untuk memperoleh model klasifikasi prioritas pembangunan daerah dengan kinerja terbaik. Tabel 3 menyajikan variabel, peran, satuan, dan sumber data yang digunakan dalam pemodelan

2.4 Penentuan Variabel Target Prediksi

Dataset yang telah melalui tahap prapemrosesan dan penentuan variabel selanjutnya dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*). Data pelatihan digunakan untuk membangun pola klasifikasi, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam memprediksi tingkat prioritas pembangunan pada data yang tidak digunakan selama proses pelatihan. Pembagian dataset dilakukan dengan proporsi 80% data pelatihan dan 20% data pengujian menggunakan *stratified sampling* untuk mempertahankan proporsi kelas prioritas tinggi, sedang, dan rendah pada kedua kelompok data. Pada data pelatihan diterapkan Stratified K-Fold Cross-Validation untuk memperoleh estimasi performa model yang lebih stabil. Dataset pelatihan dibagi menjadi beberapa *fold*, di mana setiap *fold* secara bergantian digunakan sebagai data validasi, sedangkan *fold* lainnya digunakan untuk pelatihan. Proses tersebut diterapkan pada Random Forest, XGBoost, dan Gradient Boosting menggunakan skema validasi yang sama sehingga perbandingan performa antaralgoritma dilakukan secara konsisten. Data pengujian sebesar 20% tetap dipisahkan selama proses pelatihan dan optimasi model dan hanya digunakan pada tahap evaluasi akhir untuk mengukur kemampuan generalisasi model terbaik terhadap data yang belum pernah digunakan sebelumnya.

2.5 Pengembangan Model *Machine Learning*

Tahap pengembangan model dilakukan menggunakan tiga algoritma *ensemble learning*, yaitu Random Forest, XGBoost, dan Gradient Boosting. Ketiga algoritma menggunakan dataset pelatihan dan variabel prediktor yang sama untuk menghasilkan klasifikasi tingkat prioritas pembangunan daerah. Random Forest membangun sejumlah *decision tree* dari kombinasi sampel dan fitur yang berbeda, kemudian menentukan kelas berdasarkan agregasi

hasil prediksi setiap pohon. Model ini digunakan karena mampu menangani hubungan nonlinier antarvariabel dan menyediakan informasi mengenai tingkat kepentingan fitur (*feature importance*). XGBoost menggunakan pendekatan *gradient boosting* dengan membangun model secara bertahap, di mana setiap model berikutnya diarahkan untuk memperbaiki kesalahan prediksi model sebelumnya [22], [23]. Algoritma ini dilengkapi mekanisme regularisasi untuk mengendalikan kompleksitas model dan mengurangi risiko *overfitting*. Gradient Boosting juga membangun sejumlah *weak learner* secara berurutan dengan meminimalkan kesalahan prediksi pada setiap iterasi. Algoritma ini digunakan sebagai model pembandingan untuk mengevaluasi efektivitas pendekatan *boosting* terhadap karakteristik dataset pembangunan daerah. Optimasi parameter dilakukan pada data pelatihan menggunakan Grid Search yang dikombinasikan dengan *Stratified K-Fold Cross-Validation*. Parameter yang dioptimasi meliputi jumlah estimator (*n_estimators*), kedalaman pohon (*max_depth*), dan *learning rate* sesuai karakteristik masing-masing algoritma. Kombinasi parameter terbaik kemudian digunakan untuk membangun model akhir sebelum dilakukan pengujian menggunakan *testing data*. Untuk memperoleh konfigurasi model yang optimal, setiap algoritma menggunakan sejumlah *hyperparameter* yang disesuaikan dengan karakteristik proses pembelajaran masing-masing. Parameter yang dioptimasi pada Random Forest, XGBoost, dan Gradient Boosting disajikan pada Tabel 3.

Tabel 3. Konfigurasi Hyperparameter Model Machine Learning

Model	Hyperparameter	Nilai yang Diuji	Metode Optimasi
Random Forest	<i>n_estimators</i>	100, 200, 300	Grid Search
	<i>max_depth</i>	5, 10, 20, None	Grid Search
	<i>min_samples_split</i>	2, 5, 10	Grid Search
	<i>min_samples_leaf</i>	1, 2, 4	Grid Search
XGBoost	<i>n_estimators</i>	100, 200, 300	Grid Search
	<i>max_depth</i>	3, 5, 7	Grid Search
	<i>learning_rate</i>	0.01, 0.05, 0.10	Grid Search
	<i>subsample</i>	0.8, 1.0	Grid Search
Gradient Boosting	<i>n_estimators</i>	100, 200, 300	Grid Search
	<i>max_depth</i>	3, 5, 7	Grid Search
	<i>learning_rate</i>	0.01, 0.05, 0.10	Grid Search
	<i>subsample</i>	0.8, 1.0	Grid Search

2.6. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan Random Forest, XGBoost, dan Gradient Boosting dalam mengklasifikasikan tingkat prioritas pembangunan daerah. Pengujian dilakukan menggunakan *testing data* yang telah dipisahkan dari data pelatihan sehingga evaluasi dapat menggambarkan kemampuan generalisasi model terhadap data yang belum digunakan dalam proses pelatihan. Kinerja masing-masing model diukur menggunakan Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. *Accuracy* digunakan untuk mengukur proporsi prediksi yang diklasifikasikan dengan benar terhadap keseluruhan data pengujian [24]. *Precision* mengukur ketepatan prediksi pada masing-masing kelas, sedangkan *Recall* mengukur kemampuan model dalam mengidentifikasi data yang termasuk dalam kelas prioritas tertentu. F1-Score digunakan untuk mengukur keseimbangan antara *Precision* dan *Recall* [25]. Karena target prediksi terdiri atas tiga kelas, yaitu prioritas tinggi, sedang, dan rendah, evaluasi *Precision*, *Recall*, dan F1-Score menggunakan pendekatan macro-average agar setiap kelas memperoleh bobot yang sama. ROC-AUC digunakan sebagai metrik tambahan dengan pendekatan *one-vs-rest* untuk mengukur kemampuan model dalam membedakan masing-masing kelas. Hasil evaluasi ketiga algoritma selanjutnya dibandingkan berdasarkan metrik yang sama. Macro F1-Score digunakan sebagai metrik utama dalam pemilihan model, sedangkan Accuracy, Precision, Recall, dan ROC-AUC digunakan sebagai metrik pendukung. Model dengan performa evaluasi paling baik selanjutnya dipilih untuk diintegrasikan sebagai komponen *predictive intelligence* dalam model Government Decision Intelligence (IC-Gov)

2.7 Integrasi Model Machine Learning dengan IC-Gov

Tahap ini bertujuan untuk mengintegrasikan model *Machine Learning* yang terpilih berdasarkan hasil evaluasi ke dalam arsitektur Government Decision Intelligence (IC-Gov). Model *Machine Learning* berfungsi sebagai *predictive intelligence layer* yang mengolah indikator pembangunan daerah untuk menghasilkan prediksi tingkat prioritas pembangunan pada masing-masing wilayah. Hasil prediksi tidak secara langsung digunakan sebagai keputusan kebijakan, tetapi diperlakukan sebagai *predictive insight* yang menjadi salah satu masukan dalam proses analisis IC-Gov. Setiap hasil prediksi dihubungkan kembali dengan indikator yang digunakan, sumber data,

wilayah, periode, serta informasi pendukung lainnya sehingga hasil analisis tetap memiliki keterlacakan (*traceability*). IC-Gov selanjutnya menggabungkan *predictive insight* dengan mekanisme *evidence retrieval* untuk memperoleh informasi yang relevan dari sumber data pemerintahan. Informasi tersebut diproses melalui lapisan *AI reasoning* untuk menghasilkan penjelasan kontekstual mengenai kondisi wilayah, indikator yang berkontribusi terhadap prediksi, serta informasi yang relevan sebagai bahan pertimbangan kebijakan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen Model Machine Learning

Proses eksperimen dilakukan untuk mengevaluasi kemampuan model *Machine Learning* dalam mengklasifikasikan tingkat prioritas pembangunan daerah berdasarkan indikator yang digunakan dalam penelitian. Pengujian melibatkan tiga algoritma, yaitu Random Forest, XGBoost, dan Gradient Boosting, dengan target klasifikasi berupa prioritas pembangunan rendah, sedang, dan tinggi. Model menerima empat variabel masukan, yaitu jumlah *blankspot* 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita. Dataset eksperimen dibentuk berdasarkan data pembangunan daerah yang telah melalui tahapan integrasi dan prapemrosesan. Untuk kebutuhan pengujian model, data dasar tersebut dikembangkan menjadi data simulasi terkontrol dengan tetap mempertahankan karakteristik indikator pada masing-masing wilayah. Dataset eksperimen terdiri atas **120 data simulasi** yang dikelompokkan ke dalam tiga kelas prioritas pembangunan. Distribusi data pada masing-masing kelas disajikan pada Tabel 4.

Tabel 4. Distribusi Data Pada Tiap Kelas

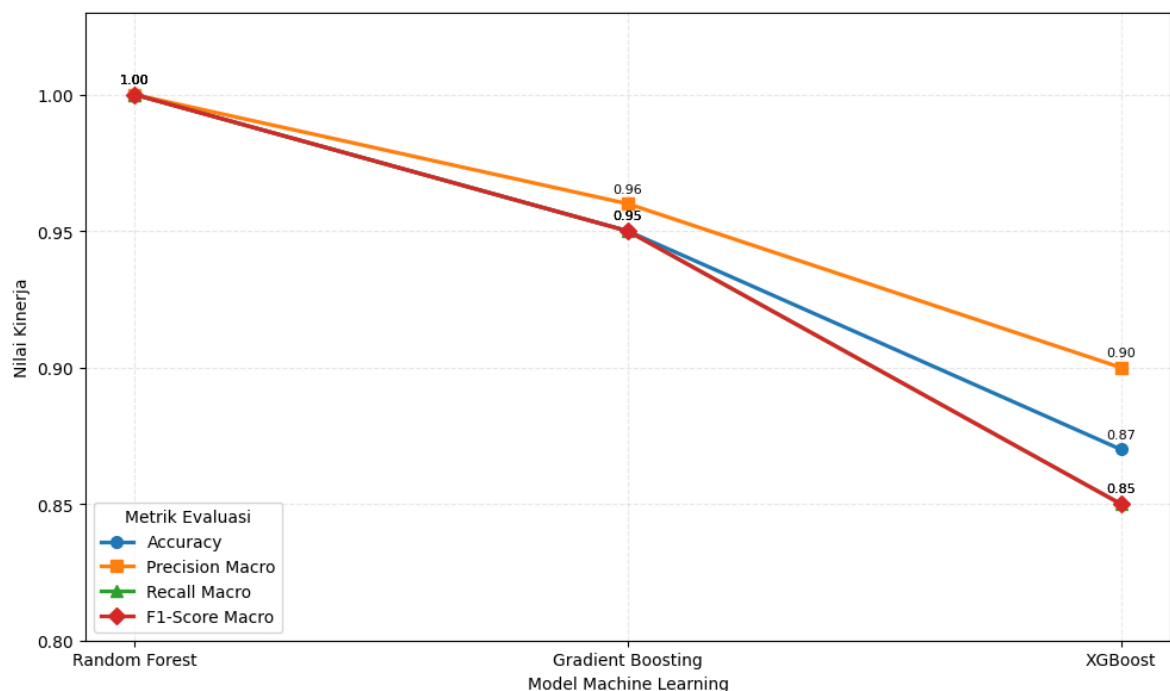
Kelas Prioritas	Jumlah Data Simulasi	Persentase (%)
Rendah	36	30,00
Sedang	36	30,00
Tinggi	48	40,00
Total	120	100,00

Berdasarkan Tabel 4, kelas prioritas rendah dan sedang masing-masing terdiri atas 36 data atau 30,00% dari keseluruhan dataset, sedangkan kelas prioritas tinggi terdiri atas 48 data atau 40,00%. Komposisi tersebut menunjukkan bahwa seluruh kelas memiliki representasi dalam dataset eksperimen, meskipun proporsi kelas tinggi sedikit lebih besar dibandingkan dua kelas lainnya. Kondisi ini menjadi salah satu pertimbangan penggunaan Macro F1-Score dalam evaluasi karena metrik tersebut menghitung kinerja setiap kelas secara setara tanpa memberikan bobot lebih besar kepada kelas dengan jumlah data lebih banyak. Dataset selanjutnya dibagi menjadi data pelatihan dan data pengujian menggunakan pendekatan *stratified split* sehingga proporsi kelas rendah, sedang, dan tinggi tetap dipertahankan pada kedua kelompok data. Pada data pelatihan, optimasi parameter dilakukan menggunakan *Grid Search* yang dikombinasikan dengan *Stratified K-Fold Cross-Validation*. Pendekatan tersebut digunakan agar konfigurasi model tidak hanya bergantung pada satu pembagian data dan setiap kelas tetap terwakili selama proses validasi. Random Forest, XGBoost, dan Gradient Boosting kemudian dilatih menggunakan data dan skema evaluasi yang sama. Kinerja ketiga algoritma dievaluasi menggunakan Accuracy, Precision, Recall, Macro F1-Score, dan ROC-AUC. *Accuracy* digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan, sedangkan *Precision* dan *Recall* digunakan untuk mengevaluasi ketepatan dan kemampuan model mengenali kelas prioritas. *Macro F1-Score* digunakan untuk memberikan penilaian yang seimbang terhadap ketiga kelas, sementara ROC-AUC digunakan untuk mengevaluasi kemampuan diskriminasi model dalam membedakan kelas prioritas rendah, sedang, dan tinggi. Selanjutnya Tabel 5 diperlihatkan hasil perbandingan kinerja dari masing-masing model

Tabel 5. Hasil Perbandingan Kinerja Model

Model	Accuracy	Precision Macro	Recall Macro	F1-Score Macro
Random Forest	1,00	1,00	1,00	1,00
Gradient Boosting	0,95	0,96	0,95	0,95
XGBoost	0,87	0,90	0,85	0,85

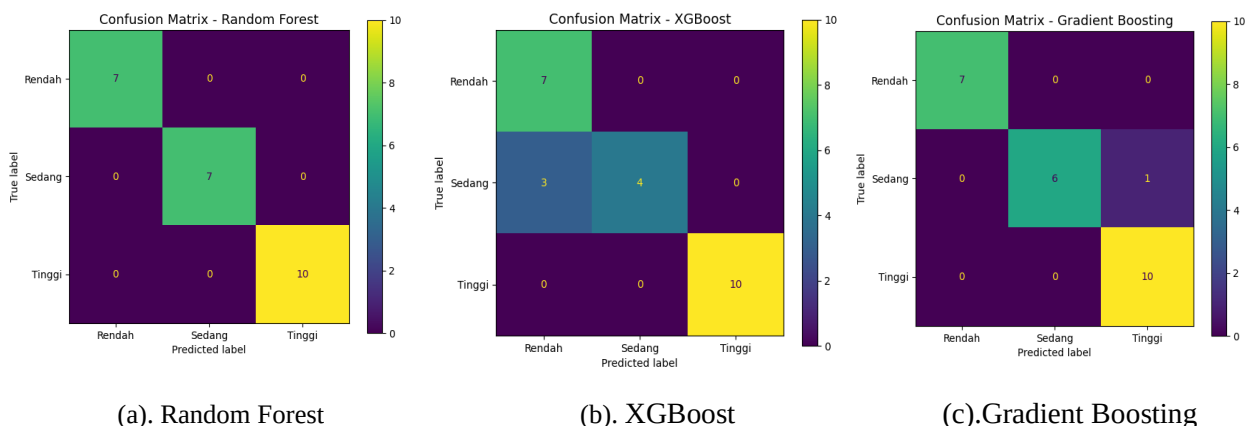
Berdasarkan hasil pengujian pada Tabel 2, ketiga algoritma menunjukkan kemampuan yang berbeda dalam mengklasifikasikan tingkat prioritas pembangunan daerah ke dalam kelas rendah, sedang, dan tinggi. Random Forest memperoleh nilai *accuracy*, *precision macro*, *recall macro*, dan *F1-score macro* masing-masing sebesar 1,00. Hasil tersebut menunjukkan bahwa pada data pengujian yang digunakan, seluruh observasi berhasil diklasifikasikan sesuai dengan kelas aktualnya. Nilai *precision* dan *recall* yang sama-sama mencapai 1,00 juga menunjukkan bahwa tidak terdapat kesalahan klasifikasi antar kelas pada data pengujian. Gradient Boosting menunjukkan kinerja yang mendekati Random Forest dengan *accuracy* sebesar 0,95, *precision macro* 0,96, *recall macro* 0,95, dan *F1-score macro* 0,95. Hasil ini menunjukkan bahwa Gradient Boosting mampu mengklasifikasikan sebagian besar data dengan benar, meskipun masih terdapat sejumlah kecil observasi yang ditempatkan pada kelas prioritas yang berbeda dari kelas aktualnya. Nilai *precision*, *recall*, dan *F1-score* yang relatif berimbang menunjukkan bahwa kemampuan model dalam mengenali ketiga kelas prioritas juga relatif konsisten. Sementara itu, XGBoost menghasilkan *accuracy* sebesar 0,87, *precision macro* 0,90, *recall macro* 0,85, dan *F1-score macro* 0,85. Dibandingkan dua algoritma lainnya, hasil tersebut menunjukkan adanya lebih banyak kesalahan klasifikasi. Perbedaan antara *precision* sebesar 0,90 dan *recall* sebesar 0,85 juga mengindikasikan bahwa ketika XGBoost menetapkan suatu kelas prioritas, prediksinya relatif tepat, tetapi masih terdapat sebagian data dari kelas aktual yang belum berhasil dikenali dengan benar. Untuk memberikan gambaran yang lebih jelas mengenai perbedaan kinerja ketiga algoritma pada seluruh metrik evaluasi, hasil pengujian selanjutnya divisualisasikan dalam bentuk kurva perbandingan sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Perbandingan Kinerja Model *Machine Learning* Berdasarkan Metrik Evaluasi

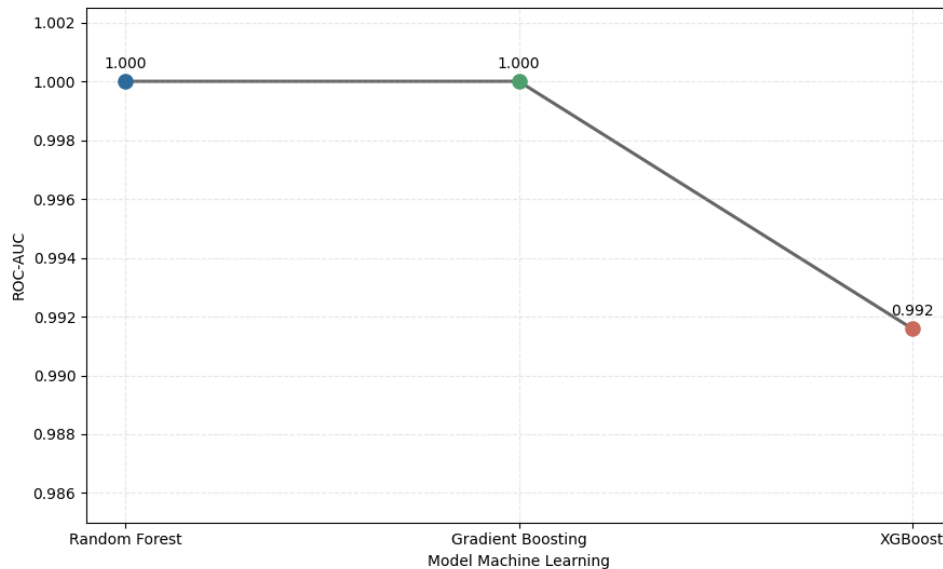
3.2 Evaluasi Kinerja Klasifikasi Model

Evaluasi lebih lanjut dilakukan untuk mengetahui kemampuan masing-masing model dalam membedakan tingkat prioritas pembangunan rendah, sedang, dan tinggi. Evaluasi tidak hanya didasarkan pada nilai metrik agregat, tetapi juga mempertimbangkan pola kesalahan klasifikasi dan kemampuan diskriminasi model. Pola kesalahan klasifikasi dianalisis menggunakan *confusion matrix*, sedangkan kemampuan model dalam membedakan kelas dievaluasi menggunakan kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under the Curve* (AUC). Sebagai tahap awal evaluasi, *confusion matrix* digunakan untuk memperlihatkan kesesuaian antara kelas aktual dan kelas hasil prediksi sekaligus mengidentifikasi pola kesalahan klasifikasi yang terjadi pada masing-masing algoritma. Perbandingan *confusion matrix* Random Forest, Gradient Boosting, dan XGBoost disajikan pada Gambar 3.



Gambar 3. *Confusion matrix* Random Forest, XGBoost dan Gradient Boosting

Berdasarkan Gambar 3, Random Forest menunjukkan hasil klasifikasi yang konsisten pada seluruh kelas prioritas. Sebanyak 7 data pada kelas rendah, 7 data pada kelas sedang, dan 10 data pada kelas tinggi berhasil diklasifikasikan sesuai dengan kelas aktualnya. Tidak terdapat data yang berpindah ke kelas lain, sehingga seluruh data pengujian berada pada diagonal utama *confusion matrix*. Gradient Boosting juga menunjukkan kemampuan klasifikasi yang tinggi. Seluruh 7 data kelas rendah berhasil diklasifikasikan dengan benar, sedangkan pada kelas sedang terdapat 6 data yang diklasifikasikan dengan benar dan 1 data diprediksi sebagai kelas tinggi. Sementara itu, seluruh 10 data kelas tinggi berhasil dikenali dengan benar. Temuan ini menunjukkan bahwa kesalahan Gradient Boosting pada data pengujian terkonsentrasi pada batas antara kelas sedang dan tinggi. Pola yang berbeda terlihat pada XGBoost. Model ini mampu mengklasifikasikan seluruh 7 data kelas rendah dan 10 data kelas tinggi dengan benar. Namun, dari 7 data aktual pada kelas sedang, hanya 4 data yang diprediksi secara tepat, sedangkan 3 data lainnya diklasifikasikan sebagai prioritas rendah. Selanjutnya penurunan kinerja XGBoost terutama disebabkan oleh keterbatasan model dalam membedakan kelas sedang dari kelas rendah pada data pengujian. Temuan ini sejalan dengan nilai *recall macro* dan *F1-score macro* XGBoost yang lebih rendah dibandingkan Random Forest dan Gradient Boosting. Meskipun *confusion matrix* telah menunjukkan pola ketepatan dan kesalahan klasifikasi pada masing-masing kelas, evaluasi selanjutnya diperlukan untuk melihat kemampuan model dalam membedakan kelas prioritas berdasarkan probabilitas prediksi. Oleh karena itu, kinerja Random Forest, Gradient Boosting, dan XGBoost dianalisis lebih lanjut menggunakan kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under the Curve* (AUC). Karena target klasifikasi terdiri atas tiga kelas, yaitu rendah, sedang, dan tinggi, evaluasi ROC menggunakan pendekatan *One-vs-Rest* (OvR) dengan *macro averaging*, sehingga kemampuan diskriminasi terhadap setiap kelas diperhitungkan secara setara. Nilai ROC-AUC yang diperoleh selanjutnya digunakan untuk membandingkan kemampuan diskriminasi Random Forest, Gradient Boosting, dan XGBoost. Perbandingan nilai ROC-AUC ketiga model disajikan pada Gambar 4.



Gambar 4. Perbandingan Nilai ROC-AUC Model *Machine Learning*

3.3 Kontribusi Indikator terhadap Prediksi Prioritas Pembangunan

Setelah kinerja klasifikasi masing-masing model dievaluasi, analisis selanjutnya diarahkan pada interpretasi kontribusi indikator yang digunakan dalam pembentukan prediksi prioritas pembangunan. Analisis ini bertujuan untuk mengetahui seberapa besar peran masing-masing indikator, yaitu jumlah *blankspot* 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita, terhadap keputusan yang dihasilkan oleh model. Interpretasi dilakukan menggunakan nilai *feature importance* yang dihasilkan oleh Random Forest, Gradient Boosting, dan XGBoost. Perbandingan kontribusi fitur pada ketiga model digunakan untuk melihat apakah terdapat pola yang konsisten mengenai indikator yang memiliki pengaruh relatif lebih besar dalam proses klasifikasi. Analisis ini penting karena penelitian tidak hanya diarahkan untuk memperoleh model dengan kinerja prediksi yang tinggi, tetapi juga untuk memberikan informasi yang dapat diinterpretasikan dalam konteks penentuan prioritas pembangunan daerah. Untuk melihat variasi kontribusi masing-masing indikator pada ketiga algoritma, nilai *feature importance* yang diperoleh dari hasil pemodelan dibandingkan secara langsung. Hasil perbandingan tersebut disajikan pada Tabel 6.

Tabel 6. Nilai *Feature Importance* Model *Machine Learning*

Fitur	Random Forest	XGBoost	Gradient Boosting
Blankspot 4G	0,233017	0,276641	0,207405
Jalan Baik (%)	0,159273	0,200334	0,151325
Indeks Pendidikan	0,261842	0,239278	0,262022
PDRB per Kapita	0,345868	0,283747	0,379248

Berdasarkan Tabel 6, terdapat pola yang relatif konsisten mengenai kontribusi keempat indikator terhadap prediksi prioritas pembangunan. PDRB per kapita memiliki nilai *feature importance* tertinggi pada Random Forest sebesar 0,345868 dan Gradient Boosting sebesar 0,379248, serta menjadi fitur dengan kontribusi tertinggi pada XGBoost sebesar 0,283747. Hasil ini menunjukkan bahwa dalam pola data yang digunakan pada eksperimen, variasi PDRB per kapita memberikan kontribusi relatif paling besar terhadap pembentukan keputusan klasifikasi pada ketiga model. Indeks pendidikan juga memberikan kontribusi yang relatif besar, dengan nilai *feature importance* sebesar 0,261842 pada Random Forest, 0,239278 pada XGBoost, dan 0,262022 pada Gradient Boosting. Sementara itu, *blankspot* 4G memiliki kontribusi sebesar 0,233017, 0,276641, dan 0,207405 pada masing-masing model. Menariknya, pada XGBoost, kontribusi *blankspot* 4G berada cukup dekat dengan PDRB per kapita, yang menunjukkan bahwa pola pemisahan kelas yang dibentuk XGBoost relatif lebih banyak memanfaatkan indikator konektivitas digital dibandingkan dua model lainnya. Persentase jalan dalam kondisi baik memiliki nilai *feature importance* paling rendah pada ketiga model, yaitu 0,159273 pada Random Forest, 0,200334 pada XGBoost, dan 0,151325 pada Gradient Boosting. Namun, nilai *feature importance* yang lebih rendah tidak berarti bahwa kondisi

jalan tidak penting dalam pembangunan daerah. Nilai tersebut hanya menunjukkan bahwa, dalam konteks dataset dan model yang digunakan pada penelitian ini, kontribusi relatif fitur jalan terhadap proses klasifikasi lebih kecil dibandingkan indikator lainnya.

3.4 Hasil Prediksi dan Uji Skenario Prioritas Pembangunan

Setelah evaluasi kinerja dan interpretasi kontribusi indikator dilakukan, pengujian selanjutnya diarahkan pada kemampuan model dalam menghasilkan prediksi tingkat prioritas pembangunan berdasarkan kombinasi kondisi indikator yang berbeda. Pengujian skenario dilakukan untuk memberikan gambaran mengenai bagaimana pola yang telah dipelajari model diterapkan ketika menerima masukan berupa kondisi konektivitas digital, infrastruktur jalan, pendidikan, dan kapasitas ekonomi suatu wilayah. Skenario pengujian dibentuk berdasarkan variasi empat indikator yang digunakan dalam pemodelan, yaitu jumlah blankspot 4G, persentase jalan dalam kondisi baik, indeks pendidikan, dan PDRB per kapita. Variasi tersebut merepresentasikan kondisi pembangunan dengan tingkat kesenjangan yang berbeda, mulai dari wilayah dengan kondisi relatif baik hingga wilayah yang menghadapi keterbatasan pada beberapa indikator pembangunan. Pengujian skenario tidak digunakan untuk menetapkan kelas prioritas secara manual, melainkan sebagai data masukan baru untuk mengamati respons model yang telah dilatih. Setiap kombinasi indikator diproses menggunakan tahapan prapemrosesan yang sama dengan data pelatihan sebelum diberikan kepada model untuk menghasilkan probabilitas dan kelas prediksi rendah, sedang, atau tinggi. Kelas yang dihasilkan merupakan keluaran prediktif model berdasarkan pola yang dipelajari dari dataset eksperimen. Untuk menjaga konsistensi interpretasi, skenario pengujian dapat dikelompokkan menjadi tiga karakteristik kondisi pembangunan. Skenario 1 menggambarkan kondisi indikator yang relatif lebih baik, Skenario 2 merepresentasikan kondisi pembangunan yang berada pada tingkat menengah atau moderat, sedangkan Skenario 3 menggambarkan kondisi dengan keterbatasan yang relatif lebih besar pada beberapa indikator pembangunan. Rancangan skenario pengujian disajikan pada Tabel 7.

Tabel 7. Hasil Prediksi Skenario Prioritas Pembangunan

Model	Prediksi Prioritas	Probabilitas Rendah	Probabilitas Sedang	Probabilitas Tinggi
Random Forest	Sedang	0,255000	0,400000	0,345000
XGBoost	Rendah	0,419205	0,328163	0,252631
Gradient Boosting	Sedang	0,000000134	0,999961	0,000039

Berdasarkan Tabel 7, Random Forest dan Gradient Boosting menghasilkan prediksi prioritas sedang, sedangkan XGBoost memprediksi prioritas rendah. Random Forest memberikan probabilitas tertinggi pada kelas sedang sebesar 0,400, tetapi nilainya relatif dekat dengan kelas tinggi sebesar 0,345. Sebaliknya, Gradient Boosting menunjukkan prediksi kelas sedang yang lebih tegas dengan probabilitas sebesar 0,999961. XGBoost memprediksi kelas rendah dengan probabilitas sebesar 0,419205, diikuti kelas sedang sebesar 0,328163 dan kelas tinggi sebesar 0,252631. Distribusi probabilitas yang relatif berdekatan menunjukkan adanya ketidakpastian yang lebih besar dibandingkan Gradient Boosting. Hasil ini juga sejalan dengan evaluasi sebelumnya yang menunjukkan bahwa XGBoost memiliki lebih banyak kesalahan dalam membedakan kelas rendah dan sedang. Perbedaan tersebut menunjukkan bahwa hasil prediksi *Machine Learning* perlu dipandang sebagai informasi prediktif (*predictive insight*), bukan sebagai keputusan kebijakan secara langsung. Dalam IC-Gov, hasil prediksi selanjutnya dikombinasikan dengan data dan *evidence* pendukung untuk memberikan konteks sebelum digunakan dalam proses sintesis dan dukungan pengambilan keputusan.

3.5 Integrasi Predictive Insight dan Evidence dalam IC-Gov

Hasil prediksi yang diperoleh dari model *Machine Learning* selanjutnya diintegrasikan ke dalam IC-Gov sebagai *predictive insight*. Informasi tersebut tidak digunakan secara langsung sebagai keputusan kebijakan, tetapi menjadi salah satu masukan analitis yang memberikan indikasi mengenai tingkat prioritas pembangunan suatu wilayah berdasarkan kondisi indikator yang dianalisis. Untuk memperkuat konteks hasil prediksi, IC-Gov melakukan proses *retrieval* terhadap data dan dokumen yang tersedia pada basis pengetahuan. Proses ini digunakan untuk memperoleh *evidence* yang relevan dengan indikator dan wilayah yang dianalisis. Dengan mekanisme tersebut, keluaran model tidak hanya berupa kelas prioritas rendah, sedang, atau tinggi, tetapi dapat dilengkapi dengan informasi pendukung yang menjelaskan kondisi pembangunan yang mendasari hasil analisis. Selanjutnya, *predictive insight* dan *evidence* digunakan sebagai konteks dalam proses *AI reasoning and synthesis*. Pada tahap ini, IC-Gov menggabungkan hasil prediksi, probabilitas model, dan informasi pendukung untuk menghasilkan penjelasan analitis yang lebih mudah dipahami dalam konteks kebijakan pembangunan daerah. Integrasi tersebut

kemudian diimplementasikan dalam antarmuka IC-Gov yang memungkinkan pengguna memperoleh informasi dan analisis melalui interaksi berbasis bahasa alami. Sebagaimana ditunjukkan pada Gambar 5, antarmuka IC-Gov menyajikan hasil analisis dalam bentuk respons interaktif yang menghubungkan informasi berbasis data dengan konteks kebijakan pembangunan daerah. Aplikasi IC-Gov dapat diakses secara daring melalui laman <https://icgov.mohjamil.my.id>.



Gambar 5. Tampilan Hasil Implementasi

Berdasarkan Gambar 5, IC-Gov menyediakan mekanisme interaksi berbasis bahasa alami yang memungkinkan pengguna memperoleh informasi pembangunan secara lebih kontekstual. Informasi yang disajikan tidak hanya berupa keluaran prediksi, tetapi diperkaya dengan *evidence* yang relevan sehingga pengguna dapat menelusuri dasar informasi yang digunakan dalam penyusunan respons. Dengan pendekatan tersebut, hasil *Machine Learning* ditempatkan sebagai salah satu komponen analitis dalam proses pembentukan *insight*, bukan sebagai keputusan kebijakan secara otomatis. Integrasi antara *predictive insight*, *retrieval evidence*, dan sintesis berbasis AI menunjukkan bahwa IC-Gov dapat menjembatani keluaran analitik dengan kebutuhan informasi dalam proses pengambilan keputusan. Hasil integrasi tersebut selanjutnya menjadi dasar untuk melihat bagaimana IC-Gov berfungsi sebagai mekanisme *Government Decision Intelligence* dalam mendukung analisis kebijakan pembangunan daerah.

3.6 Implementasi Government Decision Intelligence dalam Dukungan Pengambilan Kebijakan

Implementasi Government Decision Intelligence pada IC-Gov diarahkan untuk mengintegrasikan data pembangunan, hasil prediksi Machine Learning, dan evidence pendukung menjadi informasi analitis yang dapat digunakan dalam proses pengambilan kebijakan. Pendekatan ini menempatkan hasil prediksi bukan sebagai keputusan akhir, tetapi sebagai predictive insight yang membantu mengidentifikasi kondisi dan indikasi tingkat prioritas pembangunan berdasarkan pola data yang dianalisis. Dalam proses tersebut, hasil prediksi dilengkapi dengan probabilitas kelas untuk memberikan informasi mengenai tingkat keyakinan model. IC-Gov kemudian menghubungkan hasil tersebut dengan evidence yang tersedia pada basis pengetahuan sehingga informasi prediktif dapat dibaca bersama dengan konteks yang mendasarinya. AI selanjutnya berperan dalam menyintesis kedua komponen tersebut menjadi penjelasan analitis yang lebih mudah dipahami dan relevan dengan kebutuhan pengguna. Mekanisme tersebut membentuk alur Government Decision Intelligence yang menghubungkan data, prediksi, evidence, dan sintesis AI dalam satu proses analisis. IC-Gov tidak hanya memberikan informasi mengenai kemungkinan tingkat prioritas pembangunan, tetapi juga menyediakan konteks yang dapat membantu pengguna memahami dasar analitis dari informasi yang dihasilkan. Dalam konteks pengambilan kebijakan, keluaran IC-Gov dapat dimanfaatkan sebagai bahan awal untuk mengidentifikasi wilayah atau indikator yang memerlukan perhatian lebih lanjut, membandingkan kondisi pembangunan, serta mendukung penyusunan alternatif intervensi. Keputusan akhir tetap berada pada pengambil kebijakan dengan mempertimbangkan aspek perencanaan, regulasi, anggaran, kondisi lapangan, dan pertimbangan lain yang tidak seluruhnya direpresentasikan dalam model. Untuk menunjukkan bagaimana keluaran analitik tersebut diterjemahkan menjadi informasi pendukung kebijakan, contoh hubungan antara hasil prediksi, evidence, dan insight yang dihasilkan IC-Gov disajikan pada Tabel 8.

Tabel 8. Implementasi Government Decision Intelligence dalam Dukungan Pengambilan Kebijakan

Komponen	Keluaran IC-Gov	Fungsi dalam Pengambilan Kebijakan
Indikator pembangunan	Blankspot 4G, jalan baik, indeks pendidikan, dan PDRB per kapita	Menggambarkan kondisi pembangunan wilayah
Predictive insight	Prioritas rendah, sedang, atau tinggi	Memberikan indikasi tingkat prioritas pembangunan
Probabilitas model	Probabilitas pada masing-masing kelas prioritas	Menunjukkan tingkat keyakinan model terhadap hasil prediksi
Evidence	Data dan dokumen pendukung yang relevan	Memberikan konteks dan dasar informasi terhadap hasil analisis
AI reasoning & synthesis	Penjelasan analitis berbasis hasil prediksi dan evidence	Membantu interpretasi hasil analisis dalam konteks kebijakan
Government Decision Intelligence	Insight dan alternatif pertimbangan kebijakan	Mendukung proses pengambilan keputusan

4. KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi *Machine Learning* dalam IC-Gov dapat digunakan untuk menghasilkan *predictive insight* mengenai prioritas pembangunan daerah berdasarkan indikator *blankspot* 4G, kondisi jalan, indeks pendidikan, dan PDRB per kapita. Dari tiga algoritma yang diuji, Random Forest menunjukkan kinerja tertinggi pada data pengujian dengan *accuracy*, *precision macro*, *recall macro*, dan *F1-score macro* sebesar 1,00, diikuti Gradient Boosting dan XGBoost. Hasil *confusion matrix* menunjukkan bahwa perbedaan kinerja terutama muncul pada klasifikasi kelas sedang, sedangkan nilai ROC-AUC ketiga model menunjukkan kemampuan diskriminasi yang tinggi. Interpretasi model menunjukkan bahwa PDRB per kapita memiliki kontribusi relatif terbesar terhadap pembentukan prediksi pada ketiga algoritma, diikuti oleh indeks pendidikan, *blankspot* 4G, dan kondisi jalan dengan proporsi yang berbeda pada masing-masing model. Hasil pengujian skenario juga memperlihatkan bahwa model dapat menghasilkan kelas dan tingkat probabilitas yang berbeda pada kombinasi indikator yang sama. Temuan ini menegaskan bahwa keluaran model perlu diperlakukan sebagai informasi prediktif dan tidak digunakan secara langsung sebagai keputusan kebijakan. Kontribusi utama penelitian terletak pada pengembangan IC-Gov yang mengintegrasikan prediksi *Machine Learning*, probabilitas, interpretasi indikator, *evidence*, dan sintesis berbasis AI dalam kerangka *Government Decision Intelligence*.



Integrasi tersebut memungkinkan hasil analitik disajikan bersama konteks pendukung sehingga dapat digunakan sebagai bahan pertimbangan dalam memahami kondisi dan prioritas pembangunan daerah. Dengan demikian, IC-Gov ditempatkan sebagai sistem pendukung keputusan yang memperkuat proses analisis berbasis data, sementara keputusan akhir tetap berada pada pengambil kebijakan. Untuk pengembangan selanjutnya, penelitian dapat memperluas cakupan data, periode observasi, dan indikator pembangunan serta melakukan pengujian pada data wilayah dan periode yang berbeda. Langkah tersebut diperlukan untuk menguji konsistensi dan generalisasi model sebelum pemanfaatannya pada konteks pengambilan kebijakan yang lebih luas.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Universitas Khairun yang telah mendanai hibah penelitian ini, serta Pemerintah Provinsi Maluku Utara atas dukungan sumber data yang diberikan dalam pelaksanaan penelitian ini. Apresiasi juga disampaikan kepada seluruh pihak yang telah berkontribusi melalui penyediaan data, informasi, masukan, dan dukungan selama proses pengembangan serta pengujian IC-Gov. Dukungan tersebut menjadi bagian penting dalam penyelesaian penelitian dan pengembangan pendekatan *Government Decision Intelligence* untuk mendukung pengambilan kebijakan pembangunan daerah berbasis data dan *evidence*

REFERENSI

- [1] D. Saputra, P. R. Saputra, and S. Mersa, "Transformasi Digital Pemerintahan dalam Mewujudkan Public Value," *Al-Zayn J. Ilmu Sos. Huk.*, pp. 2782–2795, 2026, doi: <https://doi.org/10.61104/alz.v4i4.9706>.
- [2] E. Yanto *et al.*, "Pemerintahan Digital Dan Pembuatan Kebijakan Berbasis Bukti: Membangun Kerangka Kerja Tata Kelola Data Cerdas Untuk Pengambilan Keputusan Di Sektor Publik," *Indones. J. Intellect. Publ.*, vol. 6, no. 2, pp. 194–204, 2026, doi: 10.51577/ijpublication.
- [3] B. Adinegoro, M. Ruputhy, I. Pambudi, and T. Arrahman, "Kebijakan Satu Data Indonesia: Sebuah Antitesis Semangat Keterbukaan Dan Informasi Publik," *J. Ilmu Adm.*, vol. 16, pp. 1–11, 2025, doi: 10.23969/kebijakan.v16i01.21869.
- [4] Dewi, Khaidir, and Adnan, "Manajemen Pelayanan Publik Digital: Analisis Komparatif Implementasi Digital Governance Di Negara Maju Dan Negara Berkembang," *Integr. Perspect. Soc. Publick*, vol. 3, no. 6, pp. 4278–4289, 2026, [Online]. Available: <https://ipssj.com/index.php/ojs/article/view/1592%0Ahttps://ipssj.com/index.php/ojs/article/download/1592/1486>
- [5] S. W. Mursalim and R. Fajar, "DiSH (Digital Service Hub): Model Transformasi Pelayanan Publik di Lingkungan Pemerintahan Kecamatan Lembang," *J. Identitas Adm. Publik Univ. Muhammadiyah Bandung*, vol. 05, no. 02, pp. 104–125, 2026, doi: 10.52496/identitas.v5i2.903.
- [6] R. Sylvia, "Penggunaan AI Berbasis Large Language Models (LLM) sebagai Media Interaktif dalam Pendidikan Bahasa dan Hukum di Perguruan Tinggi," *DISIPLIN Maj. Civ. Akad. Sekol. Tinggi Ilmu Huk. Sumpah Pemuda*, vol. 31, no. 4, pp. 233–242, 2025, [Online]. Available: <https://ojs.stihpada.ac.id/index.php/disiplin>
- [7] Dyah Adila Perdana, "Integrasi Large Language Models Dalam Pembelajaran Analisis Teks Sastra Arab Di Perguruan Tinggi," *J. Ilmu Ekon. Pendidik. dan Tek.*, vol. 2, no. 5, pp. 32–40, 2025, [Online]. Available: <https://sihojurnal.com/index.php/identik/article/view/907>
- [8] M. Murtiyoso, I. Tahyudin, and B. Berlilana, "A Systematic Review of Retrieval-Augmented Generation for Enhancing Domain-Specific Knowledge in Large Language Models," *Sinkron*, vol. 9, no. 2, pp. 969–977, 2025, doi: 10.33395/sinkron.v9i2.14824.
- [9] E. Karakurt and A. Akbulut, "Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) for Enterprise Knowledge Management and Document Automation: A Systematic Literature Review," *Appl. Sci.*, vol. 16, no. 1, 2026, doi: 10.3390/app16010368.
- [10] P. Zhao *et al.*, "Retrieval-Augmented Generation for AI-Generated Content: A Survey," *Data Sci. Eng.*, vol. 11, no. 1, pp. 1–29, 2026, doi: 10.1007/s41019-025-00335-5.
- [11] F. Neha, D. Bhati, and D. K. Shukla, "Retrieval-Augmented Generation (RAG) in Healthcare: A Comprehensive Review," 2025. doi: 10.3390/ai6090226.
- [12] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, and F. L. Wang, "Retrieval-augmented generation for educational application: A systematic survey," *Comput. Educ. Artif. Intell.*, vol. 8, no. March, p. 100417, 2025, doi: 10.1016/j.caeai.2025.100417.





- [13] M. Lee, S. An, and M. S. Kim, "PlanRAG: A Plan-then-Retrieval Augmented Generation for Generative Large Language Models as Decision Makers," *Proc. 2024 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL 2024*, vol. 1, pp. 6537–6555, 2024, doi: 10.18653/v1/2024.naacl-long.364.
- [14] R. F. McCloud, M. A. Bekalu, T. Vaughan, L. Maranta, E. Peck, and K. Viswanath, "Evidence for Decision-Making: The Importance of Systematic Data Collection as an Essential Component of Responsive Feedback," *Glob. Heal. Sci. Pract.*, vol. 11, pp. 1–7, 2023, doi: 10.9745/GHSP-D-22-00246.
- [15] A. Y. Aytar, K. Kaya, and K. Kılıç, "A synergistic multi-stage RAG architecture for boosting context relevance in data science literature," *Nat. Lang. Process. J.*, vol. 13, p. 100179, 2025, doi: <https://doi.org/10.1016/j.nlp.2025.100179>.
- [16] M. Klesel and H. F. Wittmann, "Retrieval-Augmented Generation (RAG)," *Bus. Inf. Syst. Eng.*, vol. 67, no. 4, pp. 551–561, 2025, doi: 10.1007/s12599-025-00945-3.
- [17] Z. H. Cho *et al.*, "Large language models in materials science: assessing RAG evaluation frameworks through graphene synthesis," *RSC Adv.*, vol. 16, no. 13, pp. 11306–11313, 2026, doi: <https://doi.org/10.1039/d5ra09726f>.
- [18] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, and F. L. Wang, "Retrieval-augmented generation for educational application: A systematic survey," *Comput. Educ. Artif. Intell.*, vol. 8, p. 100417, 2025, doi: <https://doi.org/10.1016/j.caeai.2025.100417>.
- [19] A. James, M. Trovati, and S. Bolton, "Retrieval-Augmented Generation to Generate Knowledge Assets and Creation of Action Drivers," *Appl. Sci.*, vol. 15, no. 11, pp. 1–17, 2025, doi: 10.3390/app15116247.
- [20] A. Salemi and H. Zamani, "Evaluating Retrieval Quality in Retrieval-Augmented Generation," 2024, *Association for Computing Machinery, Washington DC, USA*. doi: 10.1145/3626772.3657957.
- [21] S. Knollmeyer *et al.*, "Evaluating Retrieval Augmented Generation: A Comprehensive Review of Evaluation Dimensions, Question Types, and Application," *SN Comput. Sci.*, vol. 7, no. 5, 2026, doi: 10.1007/s42979-026-05134-x.
- [22] L. R. Sitompul, A. A. Nababan, M. L. Manihuruk, W. A. Ponsen, and S. Supriyandi, "Comparison of Xgboost, Random Forest and Logistic Regression Algorithms in Stroke Disease Classification," *Sinkron*, vol. 9, no. 2, pp. 957–968, 2025, doi: 10.33395/sinkron.v9i2.14794.
- [23] J. Montomoli *et al.*, "Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients," *J. Intensive Med.*, vol. 1, no. 2, pp. 110–116, 2021, doi: 10.1016/j.jointm.2021.09.002.
- [24] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10 October, pp. 1–27, 2023, doi: 10.1371/journal.pone.0291908.
- [25] S. Riyanto, I. S. Sitanggang, T. Djatna, and T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 1082–1090, 2023, doi: 10.14569/IJACSA.2023.01406116.