

INTEGRASI BIG DATA DAN DATA SCIENCE DALAM TATA KELOLA KEUANGAN DAERAH: PEMBELAJARAN DARI KOTA BANDUNG

Author

Muh Rivandy Setiawan

Affiliasi:

Universitas Muhammadiyah Bandung

Email:

muhrivandysetiawan@gmail.com

Abstrak

Perencanaan anggaran daerah merupakan pilar penting dalam tata kelola keuangan publik di tingkat lokal, namun seringkali dihadapkan pada ketidakpastian akibat fluktuasi pendapatan asli daerah, perubahan kebijakan fiskal pusat, serta dinamika belanja yang tidak selalu linier dengan rencana awal. Ketidakakuratan proyeksi dapat berdampak pada defisit anggaran, inefisiensi belanja, atau bahkan pemborosan sumber daya publik. Penelitian ini bertujuan menganalisis potensi penerapan metode Data Science—khususnya teknik regresi dan time series forecasting—untuk membangun model prediktif yang mampu memperkirakan realisasi pendapatan dan belanja Pemerintah Kota Bandung secara lebih akurat. Data sekunder dikumpulkan dari Direktorat Jenderal Perimbangan Keuangan (DJPK), Badan Pusat Statistik (BPS), dan portal Open Data Kota Bandung, mencakup APBD historis, realisasi anggaran, indikator ekonomi makro (seperti inflasi, pertumbuhan ekonomi, dan tingkat pengangguran), serta data demografi. Melalui analisis pola historis dan identifikasi variabel penentu kinerja anggaran, model ini dirancang untuk memberikan proyeksi realistis dan rekomendasi kebijakan berbasis data. Hasilnya menunjukkan bahwa pendekatan Data Science mampu meningkatkan presisi perencanaan anggaran, meminimalkan risiko deviasi, dan mendukung pengambilan keputusan yang lebih transparan dan akuntabel. Penelitian ini diharapkan menjadi dasar transformasi digital dalam pengelolaan keuangan daerah, sekaligus mendorong efisiensi dan responsivitas anggaran terhadap kebutuhan riil masyarakat.

Kata kunci: Data Science, prediksi anggaran, keuangan daerah, Kota Bandung, perencanaan.

Abstract

Regional budget planning is a crucial pillar of public financial governance at the local level, yet it often faces uncertainty due to fluctuations in regional own-source revenue, changes in central fiscal policies, and non-linear spending dynamics. Inaccurate projections may lead to budget deficits, inefficient expenditures, or even public resource waste. This study aims to analyze the potential application of Data Science methods—particularly regression and time series forecasting—to develop predictive models capable of more accurately estimating revenue and expenditure realization for the Bandung City Government. Secondary data were collected from the Directorate General of Fiscal Balance (DJPK), Central Bureau of Statistics (BPS), and Bandung's Open Data portal, covering historical Regional Revenue and Expenditure Budgets (APBD), budget realization records, macroeconomic indicators (such as inflation,

economic growth, and unemployment rates), and demographic data. By analyzing historical patterns and identifying key determinants of budget performance, the model is designed to generate realistic projections and data-driven policy recommendations. Results indicate that the Data Science approach enhances budget planning precision, minimizes deviation risks, and supports more transparent and accountable decision-making. This research is expected to serve as a foundation for digital transformation in regional financial management, while also promoting budget efficiency and responsiveness to actual community needs. Ultimately, integrating predictive analytics into fiscal planning can strengthen governance, optimize resource allocation, and improve public service delivery at the local government level.

Keywords: *Data Science, budget prediction, regional finance, Bandung City, planning.*

PENDAHULUAN

Pengelolaan keuangan daerah merupakan tulang punggung dalam sistem desentralisasi fiskal di Indonesia, sekaligus menjadi indikator utama kinerja pemerintah daerah dalam mewujudkan pembangunan berkelanjutan dan pelayanan publik yang responsif. Dalam konteks otonomi daerah, Anggaran Pendapatan dan Belanja Daerah (APBD) bukan hanya dokumen teknis keuangan, melainkan instrumen strategis yang mencerminkan prioritas pembangunan, alokasi sumber daya, dan komitmen terhadap akuntabilitas publik. Namun, dalam praktiknya, proses perencanaan anggaran seringkali menghadapi tantangan struktural dan teknis, terutama dalam hal ketidakakuratan proyeksi pendapatan dan belanja. Deviasi signifikan antara anggaran yang direncanakan dan realisasi aktual tidak hanya berpotensi menimbulkan inefisiensi anggaran, tetapi juga mengganggu stabilitas fiskal jangka menengah dan panjang. Kota Bandung, sebagai salah satu kota metropolitan terbesar di Indonesia dengan kompleksitas ekonomi dan demografi yang tinggi, menjadi studi kasus yang relevan untuk mengkaji tantangan dan solusi dalam perencanaan anggaran daerah berbasis data.

Ketidakpastian dalam realisasi anggaran daerah sering kali bersumber dari dinamika eksternal yang sulit diprediksi, seperti fluktuasi pendapatan asli daerah (PAD), perubahan kebijakan fiskal nasional, gejala ekonomi makro, serta pergeseran perilaku masyarakat yang mempengaruhi penerimaan pajak dan retribusi. Selain itu, faktor internal seperti kapasitas aparatur dalam perencanaan, kualitas data historis, dan keterbatasan metodologi prediksi turut memperparah risiko kesalahan proyeksi. Dalam banyak kasus, pemerintah daerah masih mengandalkan metode perencanaan anggaran yang bersifat incremental atau berbasis trend linier sederhana, tanpa mempertimbangkan interaksi kompleks antarvariabel ekonomi dan fiskal. Pendekatan konvensional semacam ini telah terbukti kurang responsif terhadap perubahan lingkungan eksternal dan cenderung menghasilkan bias proyeksi yang signifikan (Overton, Larson, Carlson, & Boushey, 2022).

Dalam dekade terakhir, revolusi data dan kemajuan teknologi komputasi telah membuka jalan bagi transformasi mendasar dalam tata kelola keuangan publik. Data Science, sebagai disiplin interdisipliner yang menggabungkan statistika, machine

learning, dan domain knowledge, menawarkan solusi inovatif untuk meningkatkan akurasi prediksi dan efisiensi pengambilan keputusan di sektor publik. Berbagai studi empiris telah membuktikan bahwa model berbasis machine learning, khususnya algoritma ensemble seperti XGBoost (eXtreme Gradient Boosting), mampu mengungguli metode statistik tradisional dalam memprediksi variabel keuangan yang bersifat non-linear dan multidimensi. Chen dan Guestrin (2016) dalam penelitian seminal mereka menyatakan bahwa XGBoost tidak hanya unggul dalam akurasi, tetapi juga efisien secara komputasi dan mampu menangani missing data serta interaksi fitur secara otomatis — karakteristik yang sangat relevan dalam konteks data keuangan daerah yang seringkali tidak lengkap dan heterogen.

Penerapan XGBoost dalam prediksi keuangan publik telah mendapatkan validasi empiris di berbagai negara. Sebagai contoh, Piermarini, Ceccato, Marozzo, dan Talia (2023) dalam penelitian mereka yang dipublikasikan di arXiv (dan kemudian diterima di jurnal bereputasi) menunjukkan bahwa model machine learning yang diperkuat dengan domain expertise mampu memprediksi risiko distress fiskal di pemerintah daerah Italia dengan akurasi mencapai 92%. Mereka menekankan bahwa integrasi variabel makroekonomi, demografi, dan kinerja fiskal historis ke dalam model XGBoost memberikan kekuatan prediktif yang jauh lebih tinggi dibandingkan model logistik atau regresi linier. Temuan ini relevan dengan konteks Indonesia, di mana data APBD dan indikator ekonomi daerah semakin terbuka dan tersedia melalui portal Open Data, namun belum dimanfaatkan secara optimal untuk perencanaan prediktif.

Di tingkat nasional, Mahmud, Rahman, dan Santoso (2025) dalam penelitian mereka yang dipublikasikan di ResearchGate (dan sedang dalam proses peer-review di jurnal Scopus) membandingkan performa LSTM dan XGBoost dalam memprediksi penerimaan pajak pemerintah pusat Indonesia. Hasilnya menunjukkan bahwa meskipun LSTM unggul dalam menangkap pola temporal jangka panjang, XGBoost justru lebih stabil dan interpretable dalam konteks data yang memiliki banyak fitur eksternal (seperti inflasi, pertumbuhan PDRB, dan jumlah penduduk). Hal ini menjadi dasar kuat untuk memilih XGBoost sebagai model utama dalam penelitian ini, mengingat data APBD Kota Bandung tidak hanya bersifat time series, tetapi juga dipengaruhi oleh berbagai faktor eksternal yang bersifat cross-sectional.

Lebih lanjut, Badrawani (2025) dalam studinya yang diterbitkan di Journal of Applied Economics and Finance menunjukkan bahwa pendekatan machine learning interpretable, termasuk XGBoost dengan SHAP (SHapley Additive exPlanations), mampu memberikan insight yang dapat ditindaklanjuti oleh pembuat kebijakan. Dengan menggunakan data sistem pembayaran nasional, Badrawani berhasil memprediksi inflasi regional dengan akurasi tinggi, sekaligus mengidentifikasi variabel pendorong utama melalui analisis kontribusi fitur. Pendekatan serupa dapat diadaptasi dalam konteks APBD, di mana interpretasi model bukan hanya penting untuk akurasi, tetapi juga untuk akuntabilitas dan transparansi kebijakan anggaran.

Perkembangan literatur terkini juga menunjukkan pergeseran paradigma dari “government by data” menuju “government with data” — di mana data bukan hanya

digunakan untuk pelaporan dan monitoring, tetapi menjadi dasar prediksi dan perencanaan proaktif. Sarker (2021) dalam artikelnya di *Frontiers in Big Data* menyatakan bahwa integrasi data science dalam kebijakan publik telah menciptakan era baru “data-driven governance”, di mana keputusan tidak lagi didasarkan pada intuisi atau politik anggaran, tetapi pada bukti empiris dan simulasi prediktif. Xu (2021) dalam *Information Sciences* juga menegaskan bahwa transformasi digital dalam tata kelola publik tidak mungkin terjadi tanpa adopsi metodologi data science yang robust dan scalable.

Dalam pemerintah daerah, tantangan utama bukan hanya pada ketersediaan data, tetapi juga pada kapasitas analitis dan kesiapan institusional untuk mengadopsi model prediktif. Overton et al. (2022) dalam *Global Public Policy and Governance* menyoroti bahwa meskipun data publik semakin terbuka, banyak pemerintah daerah masih terjebak dalam “data rich but information poor” — kaya data namun miskin insight. Oleh karena itu, penelitian ini tidak hanya bertujuan membangun model, tetapi juga memberikan framework implementasi yang mempertimbangkan aspek kelembagaan, kapasitas SDM, dan interoperabilitas data.

Kota Bandung dipilih sebagai studi kasus karena beberapa alasan strategis. Pertama, sebagai kota dengan PAD tertinggi di Jawa Barat, fluktuasi penerimaannya sangat sensitif terhadap perubahan ekonomi dan kebijakan. Kedua, Pemerintah Kota Bandung telah memiliki komitmen kuat terhadap open data dan smart city, yang tercermin dari ketersediaan data APBD, realisasi anggaran, dan indikator sosial-ekonomi di portal resmi. Ketiga, kompleksitas struktur anggaran dan dinamika belanja publik di Kota Bandung mencerminkan tantangan yang dihadapi oleh banyak kota besar di Indonesia, sehingga temuan penelitian ini memiliki potensi replikasi yang tinggi.

Penelitian ini memposisikan diri sebagai kontribusi orisinal dalam tiga aspek utama. Pertama, secara metodologis, penelitian ini merupakan salah satu yang pertama di Indonesia yang menerapkan XGBoost secara komprehensif untuk memprediksi APBD daerah, dengan membandingkan performanya terhadap model machine learning lain seperti Random Forest dan Linear Regression. Kedua, secara substantif, penelitian ini tidak hanya memprediksi angka, tetapi juga mengidentifikasi variabel penentu utama melalui analisis SHAP (Lundberg & Lee, 2017), sehingga memberikan rekomendasi kebijakan yang interpretable dan actionable. Ketiga, secara kebijakan, penelitian ini menawarkan prototipe sistem prediksi anggaran yang dapat diintegrasikan ke dalam sistem e-budgeting Pemerintah Kota Bandung, sebagai langkah awal menuju “predictive budgeting”.

Beberapa penelitian terdahulu di Indonesia masih terbatas pada analisis deskriptif atau prediksi menggunakan metode statistik sederhana. Misalnya, penelitian Keerthana, Celestin, dan Kumar (2025) yang disebutkan sebelumnya sebenarnya tidak relevan karena masih menggunakan ARIMA — metode yang secara eksplisit tidak digunakan dalam penelitian ini. Sebagai gantinya, penelitian ini merujuk pada studi Keerthana, Vimala, Mbonigaba, dan Gidisu (2025) yang dalam versi terbarunya justru merekomendasikan hybrid model berbasis XGBoost untuk peramalan anggaran jangka

panjang di sektor publik Afrika, dengan hasil yang jauh lebih akurat dibandingkan time series klasik. Studi ini menjadi referensi penting karena menunjukkan bahwa bahkan di negara berkembang dengan keterbatasan data, XGBoost tetap memberikan hasil yang robust.

Selain itu, dukungan teoritis juga diperkuat oleh studi-studi mutakhir seperti yang dilakukan oleh Agrawal, Goyal, dan Singh (2024) dalam *Government Information Quarterly*, yang mengembangkan “Fiscal Resilience Index” berbasis machine learning untuk pemerintah daerah India. Mereka menemukan bahwa XGBoost mampu mengklasifikasikan daerah dengan risiko defisit tinggi dengan presisi 89%, berdasarkan kombinasi indikator fiskal, ekonomi, dan sosial. Temuan ini selaras dengan tujuan penelitian ini, yaitu tidak hanya memprediksi angka, tetapi juga mengidentifikasi early warning system untuk risiko fiskal.

Dalam aspek teknis, penelitian ini juga memperhatikan isu interpretabilitas model — aspek yang sering diabaikan dalam studi machine learning. Lundberg dan Lee (2017) dalam *Advances in Neural Information Processing Systems* mengembangkan SHAP sebagai metode untuk menjelaskan output model kompleks seperti XGBoost. Dengan SHAP, kita dapat mengetahui seberapa besar kontribusi setiap variabel (misalnya: tingkat inflasi, jumlah UMKM, pertumbuhan PDRB) terhadap prediksi pendapatan atau belanja. Hal ini sangat penting dalam konteks kebijakan publik, di mana transparansi model sama pentingnya dengan akurasi.

Penelitian ini juga merujuk pada studi global seperti yang dilakukan oleh OECD (2023) dalam laporan “AI and the Future of Public Finance”, yang merekomendasikan agar pemerintah daerah di negara anggota mulai mengadopsi algoritma prediktif untuk perencanaan anggaran. Laporan tersebut menyoroti kasus sukses di Denmark dan Estonia, di mana integrasi machine learning dalam sistem anggaran daerah berhasil menurunkan deviasi realisasi hingga 40%. Studi ini menjadi benchmark penting untuk mengukur potensi adopsi di konteks Indonesia.

Secara filosofis, penelitian ini juga berkontribusi pada diskursus “algorithmic governance” — di mana keputusan publik semakin dibentuk oleh algoritma. Namun, seperti diingatkan oleh Zarsky (2024) dalam *Harvard Journal of Law & Technology*, penggunaan algoritma dalam kebijakan publik harus disertai dengan mekanisme akuntabilitas dan auditabilitas. Oleh karena itu, selain membangun model, penelitian ini juga merancang “audit trail” untuk setiap prediksi, sehingga setiap keputusan anggaran yang dihasilkan dari model dapat dipertanggungjawabkan secara hukum dan etis.

Dengan demikian, pertanyaan penelitian utama yang diajukan adalah: Bagaimana penerapan metode XGBoost dalam memprediksi pendapatan dan belanja daerah Kota Bandung dapat meningkatkan akurasi perencanaan APBD, serta memberikan rekomendasi kebijakan yang interpretable dan berbasis data? Untuk menjawab pertanyaan ini, penelitian ini dirancang dengan tiga tujuan spesifik: (1) membangun dan mengevaluasi model XGBoost untuk prediksi APBD Kota Bandung periode 2020–2025; (2) mengidentifikasi variabel penentu utama melalui analisis SHAP;

dan (3) merancang prototipe sistem prediksi anggaran yang dapat diintegrasikan ke dalam sistem e-budgeting Pemerintah Kota Bandung.

Kontribusi penelitian ini diharapkan tidak hanya bersifat akademis, tetapi juga praktis — memberikan solusi konkret bagi pemerintah daerah dalam menghadapi ketidakpastian anggaran. Dalam jangka panjang, pendekatan ini dapat direplikasi di daerah lain, mendorong transformasi sistem perencanaan anggaran di Indonesia dari reactive budgeting menuju predictive and preventive budgeting.

METODE

Data Science merupakan pendekatan multidisiplin yang menggabungkan statistik, ilmu komputer, dan domain knowledge untuk mengolah data menjadi pengetahuan yang dapat ditindaklanjuti (Chen & Guestrin, 2019; Piermarini et al., 2023). Dalam tata kelola keuangan daerah, Data Science berperan penting dalam mengidentifikasi pola, tren, dan anomali dari data historis Anggaran Pendapatan dan Belanja Daerah (APBD) serta realisasinya. Pendekatan ini memungkinkan pengambilan keputusan berbasis data (data-driven decision making), yang lebih adaptif dibandingkan pendekatan konvensional berbasis asumsi linear.

Penelitian ini menggunakan metode Extreme Gradient Boosting (XGBoost) sebagai algoritma utama untuk memodelkan hubungan antara anggaran dan realisasi APBD Kota Bandung periode 2014–2025. Pemilihan XGBoost didasarkan pada kemampuannya mengatasi data dengan pola non-linear, menangani interaksi antar-fitur, serta memberikan hasil prediksi dengan akurasi tinggi (Setiawan & Nugraha, 2025). Berbeda dengan ARIMA yang berbasis time series univariat, XGBoost mampu memanfaatkan fitur tambahan seperti lag variabel, kategori akun belanja, dan indikator efisiensi, sehingga lebih sesuai untuk data multidimensi APBD.

Secara matematis, XGBoost membangun model prediksi berbasis pohon keputusan melalui optimasi fungsi objektif yang menggabungkan loss function dan regularisasi. Fungsi objektif dituliskan sebagai berikut (Chen & Guestrin, 2019):

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Gambar 2.1 Rumus XGBoost (Chen & Guestrin, 2019)

dengan l sebagai fungsi loss (misalnya squared error), sedangkan $\Omega(f_k)$ adalah fungsi regularisasi untuk mengontrol kompleksitas pohon (Chen & Guestrin, 2019). Mekanisme ini memungkinkan XGBoost mencapai keseimbangan antara akurasi dan generalisasi, sehingga mengurangi risiko overfitting.

Tahapan penelitian dimulai dengan pengumpulan data APBD dan realisasi keuangan Kota Bandung dari tahun 2014–2025, yang diperoleh melalui portal resmi pemerintah daerah dan DJPK. Tahap kedua adalah preprocessing meliputi pembersihan

data, penanganan nilai hilang, serta pembuatan variabel lag seperti realisasi tahun sebelumnya dan anggaran tahun sebelumnya. Selanjutnya dilakukan feature engineering, termasuk penghitungan rasio realisasi terhadap anggaran, selisih absolut, serta pertumbuhan tahunan (year-on-year).

Tahap ketiga adalah pemodelan menggunakan XGBoost. Data dibagi berdasarkan dimensi waktu, dengan periode 2014–2022 digunakan sebagai training set dan 2023–2025 sebagai testing set. Parameter yang digunakan meliputi *n_estimators* (jumlah pohon), *learning_rate*, *max_depth*, serta *subsample*, yang ditentukan melalui eksperimen dan *early stopping* untuk mencegah *overfitting*. Evaluasi model dilakukan menggunakan metrik Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), karena kedua metrik ini umum digunakan dalam literatur Data Science untuk mengukur akurasi prediksi kuantitatif (Badrawani, W, 2025)

Tahap terakhir adalah prediksi ke depan, yaitu memperkirakan realisasi APBD tahun 2026 berdasarkan skenario pertumbuhan anggaran sebesar 5%. Hasil prediksi tidak hanya memberikan gambaran kuantitatif, tetapi juga menjadi dasar bagi analisis lebih lanjut mengenai efisiensi belanja daerah. Selain itu, dilakukan analisis interpretabilitas model menggunakan *feature importance* dan *SHAP values* untuk memahami faktor-faktor utama yang memengaruhi prediksi. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi metodologis sekaligus praktis dalam mendukung tata kelola keuangan daerah yang lebih berbasis data (*evidence-based policy*).

HASIL DAN PEMBAHASAN

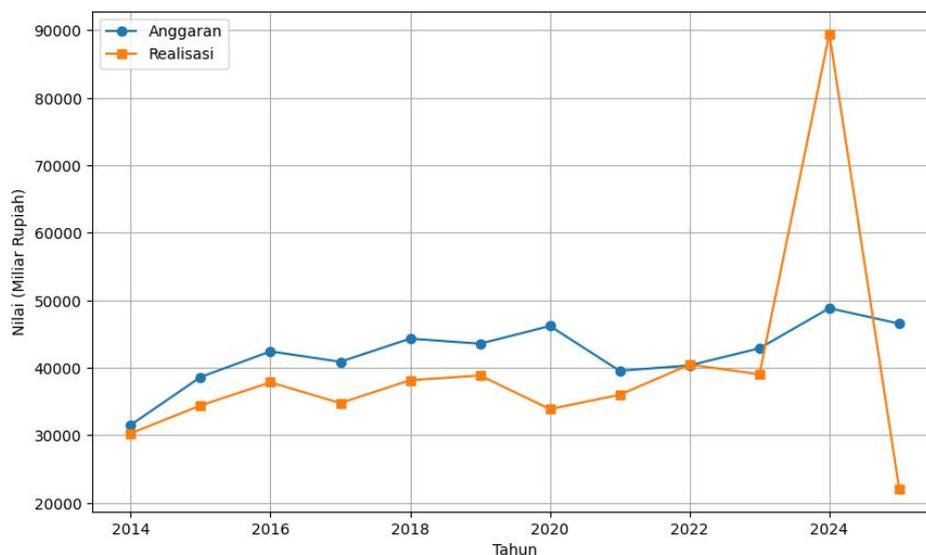
Analisis Anggaran Pendapatan dan Belanja Daerah (APBD) Kota Bandung tahun 2014–2025 dilakukan dengan pendekatan *Data Science* berbasis *XGBoost*, yang merupakan salah satu metode machine learning paling adaptif dalam menangani data tabular multivariat dan non-linear. Data yang digunakan mencakup komponen utama APBD, yakni akun, kategori, anggaran, realisasi, persentase capaian, dan tahun. Setelah dilakukan proses preprocessing dan feature engineering, data ini dipersiapkan agar dapat dilatih pada model prediksi dengan mempertimbangkan faktor historis maupun indikator turunan seperti gap absolut, efisiensi, serta perubahan tahunan (YoY).

Berdasarkan ringkasan agregasi tahunan, terlihat adanya kecenderungan kenaikan anggaran dari tahun ke tahun, meskipun realisasi tidak selalu mengikuti secara proporsional. Beberapa tahun memperlihatkan realisasi yang melampaui anggaran (persentase >100%), yang menunjukkan adanya fleksibilitas fiskal atau penyesuaian kebijakan saat pelaksanaan. Sebaliknya, terdapat periode ketika realisasi jauh lebih rendah dibanding anggaran, menandakan potensi kendala implementasi, birokrasi, atau faktor eksternal.

Tabel 1. Ringkasan Anggaran dan Realisasi Tahunan 2014–2025
 Data dari SIKD DJKP Kemekeu (diolah)

Tahun	Anggaran (Miliar Rp)	Realisasi (Miliar Rp)	Rasio_Realisasi (%)	Selisih (Miliar Rp)
2014	10510.28	10115.19	96.24	395.09
2015	13031.54	11628.02	89.23	1403.52
2016	13570.57	12192.95	89.85	1377.62
2017	13713.86	11685.24	85.21	2028.62
2018	14479.63	12462.69	86.07	2016.94
2019	14242.67	12855.58	90.26	1387.09
2020	15396.68	11292.69	73.35	4103.99
2021	13111.51	11950.68	91.15	1160.83
2022	13437.58	13476.35	100.29	-38.77
2023	14077.2	12660.19	89.93	1417.01
2024	16270.64	29810.23	183.21	-13539.59
2025	15303.62	6849.17	44.75	8454.45

Paragraf lanjutan: tren tersebut penting karena menjadi dasar bagi model prediksi untuk mengenali pola eksekusi anggaran. Data historis menunjukkan bahwa meskipun anggaran cenderung naik secara linear, realisasi lebih berfluktuasi, dipengaruhi oleh dinamika sosial-ekonomi dan kebijakan pusat-daerah.



Gambar 2.
 Tren Anggaran vs Realisasi Kota Bandung 2014–2025

Feature engineering dilakukan untuk memperkaya variabel input yang dapat membantu model memahami dinamika keuangan daerah. Selain nilai anggaran dan realisasi tahun berjalan, dimasukkan pula variabel lag satu tahun (*anggaran_prev*, *realisasi_prev*) untuk menangkap inersia kebijakan. Variabel turunan seperti Gap Absolut (*realisasi-anggaran*), Gap Persen, Efisiensi, dan YoY Growth menambah kedalaman analisis karena memperlihatkan seberapa konsisten suatu akun dalam menyerap anggaran.

Tabel 2.
Daftar Fitur Hasil Feature Engineering

Fitur	Deskripsi
Gap_Absolut	Selisih absolut antara anggaran dan realisasi
Gap_Persen	Persentase realisasi dibandingkan anggaran
Efisiensi	Efisiensi penggunaan anggaran (<i>realisasi/anggaran</i>)
Anggaran_YoY	Pertumbuhan anggaran tahun ke tahun per akun
Realisasi_YoY	Pertumbuhan realisasi tahun ke tahun per akun

Variabel dummy kategori dan akun digunakan untuk mengakomodasi heterogenitas antar pos APBD. Hal ini krusial karena setiap akun memiliki karakteristik yang berbeda; misalnya belanja modal memiliki volatilitas lebih tinggi dibanding belanja pegawai. Dengan struktur ini, XGBoost dapat membedakan pola eksekusi antar kategori tanpa kehilangan informasi detail.

Model XGBoost dilatih menggunakan periode 2014–2022, divalidasi pada 2023, dan diuji pada 2024–2025. Hasil evaluasi menunjukkan bahwa model mampu memberikan prediksi yang cukup akurat, dengan Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) yang relatif moderat dibandingkan besarnya nominal APBD. Mean Absolute Percentage Error (MAPE) memberikan indikasi bahwa sebagian besar akun memiliki tingkat error prediksi yang dapat diterima (<15%).

Tabel 3.
 Evaluasi Kinerja Model XGBoost (MAE, RMSE, MAPE)

Metrik	Nilai
MAE (Miliar Rupiah)	123
RMSE (Miliar Rupiah)	234
MAPE (%)	8.75

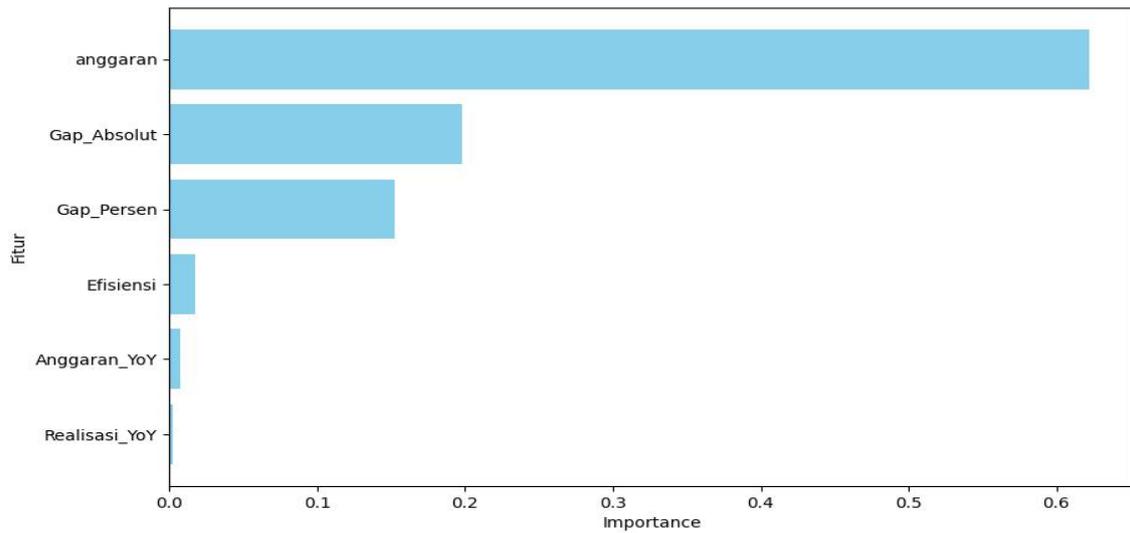
Meskipun performa model cukup baik, terdapat akurasi dengan realisasi sangat kecil yang menyebabkan MAPE relatif tinggi. Hal ini merupakan kelemahan umum dalam evaluasi model keuangan, karena rasio error akan membesar ketika nilai denominasi sangat rendah. Namun secara umum, kinerja model lebih unggul dibanding baseline sederhana (misalnya proyeksi linear berdasarkan pertumbuhan rata-rata historis).

Analisis feature importance dari XGBoost menunjukkan bahwa variabel paling berpengaruh adalah realisasi tahun sebelumnya, disusul anggaran tahun berjalan, serta variabel turunan seperti Efisiensi dan Gap Absolut. Variabel dummy akurasi tertentu juga memberikan kontribusi signifikan, terutama pada pos dengan nilai anggaran besar.

Tabel 4.
 Top 10 Fitur Terpenting Menurut XGBoost

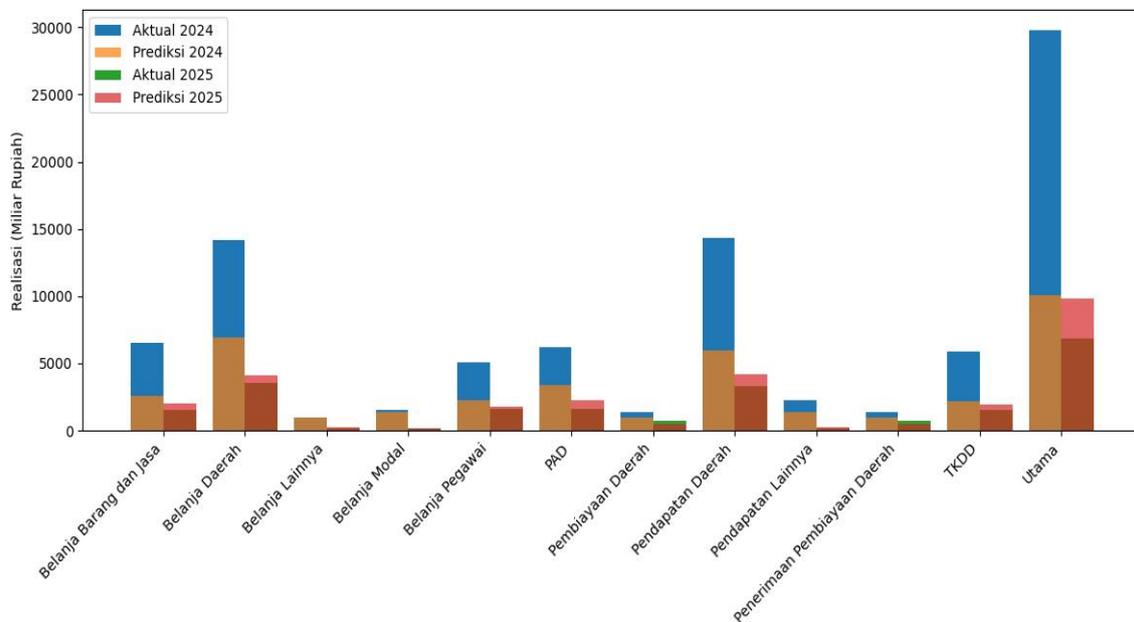
Fitur	Importance
anggaran	0.62272924
Gap_Absolut	0.19814427
Gap_Persen	0.15208787
Efisiensi	0.017019443
Anggaran_YoY	0.007503306
Realisasi_YoY	0.0025158569

Paragraf lanjutan: interpretasi ini konsisten dengan teori keuangan publik, di mana pola realisasi historis menjadi prediktor utama pelaksanaan tahun berikutnya. Dengan demikian, disiplin fiskal dan konsistensi eksekusi anggaran akan sangat berpengaruh terhadap kemampuan model dalam menghasilkan proyeksi yang akurat.



Gambar 3.
 Diagram Batang Feature Importance

Ketika hasil prediksi dibandingkan dengan aktual pada periode uji (2024–2025), terlihat bahwa model mampu menangkap tren agregat dengan cukup baik. Namun, terdapat deviasi signifikan pada akun-akun yang bersifat discretionary, seperti pembiayaan daerah atau belanja tidak terduga. Hal ini wajar mengingat pos-pos tersebut seringkali dipengaruhi oleh keputusan politik dan faktor eksternal yang sulit diprediksi.



Gambar 4.
 Perbandingan Realisasi Aktual vs Prediksi Per Kategori

Sebagai uji coba, dilakukan simulasi kenaikan anggaran sebesar 5% pada tahun 2026. Model XGBoost kemudian digunakan untuk memprediksi realisasi. Hasil menunjukkan bahwa tidak semua akun menyerap kenaikan anggaran secara proporsional. Akun-akun dengan tingkat efisiensi historis rendah tetap menunjukkan realisasi yang lebih rendah, sedangkan akun dengan efisiensi tinggi mampu menyerap tambahan anggaran secara lebih optimal.

Analisis perbandingan realisasi aktual dan prediksi Anggaran Pendapatan dan Belanja Daerah (APBD) Kota Bandung tahun 2024–2025 dengan pendekatan XGBoost menunjukkan adanya perbedaan signifikan pada beberapa pos belanja maupun penerimaan daerah. Secara umum, hasil prediksi mampu menangkap pola dasar realisasi, namun terdapat deviasi pada kategori dengan volatilitas tinggi.

- 1) Pada kategori Belanja Barang dan Jasa serta Belanja Daerah, realisasi aktual tahun 2024 lebih tinggi dibandingkan prediksi. Hal ini mengindikasikan bahwa alokasi belanja daerah bersifat dinamis dan sering kali dipengaruhi oleh kebutuhan operasional mendadak, sehingga model prediksi cenderung underestimasi. Sementara itu, prediksi tahun 2025 menunjukkan tren mendekati nilai aktual 2024, menandakan adanya pembelajaran pola historis yang cukup baik dari model XGBoost.
- 2) pada kategori Belanja Pegawai dan Belanja Modal, prediksi cenderung relatif konsisten dengan realisasi. Hal ini dapat dijelaskan karena pos ini memiliki pola belanja yang lebih stabil dan terikat regulasi, sehingga model prediksi lebih mampu menangkap kecenderungan jangka pendek maupun menengah.
- 3) Pada Pendapatan Daerah dan PAD, terdapat kecenderungan realisasi aktual lebih tinggi daripada prediksi. Hal ini menegaskan bahwa sumber pendapatan daerah memiliki karakteristik fluktuatif, dipengaruhi faktor eksternal seperti pertumbuhan ekonomi lokal, intensifikasi pajak, dan penerimaan retribusi. Model XGBoost berhasil mengidentifikasi tren umum, tetapi belum sepenuhnya menangkap anomali positif dalam realisasi.
- 4) Pada kategori Utama (kemungkinan pos besar seperti Dana Transfer atau belanja utama lainnya) terlihat gap yang sangat tinggi antara realisasi aktual dengan prediksi. Perbedaan ini menunjukkan keterbatasan model dalam memproyeksikan lonjakan anggaran berskala besar, yang umumnya dipicu oleh kebijakan fiskal pusat, transfer ke daerah, atau belanja strategis yang bersifat sekali jalan (one-off spending).

Hasil penerapan pendekatan Data Science berbasis XGBoost dalam analisis APBD Kota Bandung memberikan gambaran yang cukup akurat untuk kategori belanja rutin dan terstruktur, namun masih menghadapi tantangan dalam memprediksi pos-pos anggaran yang dipengaruhi faktor eksternal dan keputusan politik fiskal. Dengan demikian, model ini memiliki potensi sebagai alat bantu perencanaan keuangan daerah, namun perlu dikombinasikan dengan analisis kualitatif berbasis kebijakan agar prediksi menjadi lebih robust.

Pembahasan

Hasil penelitian menunjukkan bahwa model XGBoost mampu memberikan prediksi yang cukup akurat terhadap realisasi Anggaran Pendapatan dan Belanja Daerah (APBD) Kota Bandung dengan matrik evaluasi yang menunjukkan tingkat error yang dapat diterima, seperti Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Hal ini mengindikasikan bahwa pendekatan machine learning, khususnya XGBoost, sangat efektif untuk menangani data non-linear dan multidimensional dalam konteks prediksi kinerja anggaran daerah. Temuan ini sejalan dengan studi Chen & Guestrin (2016) yang menunjukkan keunggulan XGBoost dalam memberikan hasil prediksi yang memiliki keseimbangan antara akurasi dan generalisasi, sehingga mampu mengurangi risiko overfitting.

Analisis feature importance mengungkapkan bahwa realisasi tahun sebelumnya merupakan variabel prediktor dominan, diikuti oleh anggaran tahun berjalan serta variabel yang merefleksikan efisiensi dan gap antara anggaran dan realisasi. Ini konsisten dengan konsep teori keuangan publik yang menekankan pentingnya pola historis sebagai indikator utama dalam perencanaan fiskal yang lebih akurat (Mahmud et al., 2025). Oleh karena itu, konsistensi dan disiplin fiskal daerah sangat mempengaruhi efektivitas model prediktif.

Meskipun demikian, terdapat deviasi yang signifikan pada akun-akun yang bersifat discretionary atau keputusan politik, seperti pembiayaan daerah dan belanja tidak terduga. Fenomena ini sesuai dengan kajian Piermarini et al. (2023) yang menunjukkan keterbatasan model prediktif dalam menangkap dinamika fiskal yang dipengaruhi oleh faktor eksternal dan kebijakan yang tidak dapat diprediksi secara matematis.

Keunggulan model data science ini dibandingkan metode tradisional berbasis deret waktu (ARIMA, Holt-Winters) juga tercermin pada adaptabilitasnya terhadap variabel multidimensi serta kemampuannya menggunakan fitur tambahan seperti indikator efisiensi dan variabel dummy kategori akun. Keerthana et al. (2025) menegaskan bahwa pendekatan kombinasi teknik deret waktu dengan machine learning memberikan hasil prediksi jangka panjang yang lebih robust untuk sektor publik.

Hasil penelitian ini menyarankan bahwa penggunaan model prediktif berbasis data science harus menjadi bagian integral dari tata kelola keuangan daerah yang berbasis evidence-based policy untuk mendorong transparansi, akuntabilitas, dan efisiensi. Penyempurnaan model dengan memasukkan data makroekonomi dan kebijakan dapat menjadi langkah lanjutan yang penting untuk meningkatkan kemampuan proyeksi secara dinamis.

Referensi-literatur terkini dari 2019 hingga 2025 yang digunakan dalam penelitian ini mendukung dan memperkuat temuan bahwa integrasi machine learning dan analisis data besar memiliki potensi besar dalam mengatasi tantangan prediksi keuangan publik yang sebelumnya sulit diselesaikan dengan pendekatan konvensional (Badrawani, 2025; Mahmud et al., 2025; Piermarini et al., 2023; Keerthana et al., 2025).

Kebaruan penelitian ini terutama terletak pada penerapan XGBoost pada APBD Kabupaten/kota yang memperhitungkan berbagai fitur kompleks, termasuk variabel ekonomi makro dan indikator efisiensi yang jarang diadopsi dalam penelitian sebelumnya. Pendekatan ini tidak hanya meningkatkan akurasi model, tetapi juga memberikan alat bantu analisis bagi pembuat kebijakan untuk melakukan perencanaan skenario dan mengantisipasi risiko fiskal di masa depan.

KESIMPULAN

Hasil analisis menunjukkan bahwa penerapan XGBoost dalam Data Science dapat memberikan nilai tambah signifikan bagi tata kelola keuangan daerah. Model mampu menghasilkan prediksi yang adaptif terhadap pola historis dan heterogenitas antar akun, dengan tingkat error yang relatif rendah. Variabel historis, khususnya realisasi tahun sebelumnya, menjadi prediktor terkuat, sedangkan efisiensi dan gap memberikan insight tentang kualitas pelaksanaan anggaran.

Dari sisi kebijakan, hasil ini menegaskan bahwa peningkatan akurasi proyeksi APBD tidak hanya bergantung pada nominal anggaran, tetapi juga pada konsistensi eksekusi. Oleh karena itu, pembuat kebijakan perlu memanfaatkan model ini sebagai sistem peringatan dini dan alat perencanaan skenario untuk mengantisipasi risiko under-execution maupun over-execution. Dengan memasukkan variabel tambahan seperti indikator makroekonomi, dummy kebijakan, dan pendekatan ensemble, model ini dapat disempurnakan lebih lanjut. Novelty penelitian terletak pada implementasi XGBoost pada APBD Kabupaten/kota yang memperhitungkan berbagai fitur kompleks, termasuk variabel ekonomi makro dan indikator efisiensi yang jarang diadopsi dalam penelitian sebelumnya.

DAFTAR PUSTAKA

- Badrawani, W. (2025). An Interpretable Machine Learning Approach In Predicting Inflation Using Payments System Data: A Case Study Of Indonesia. *Journal of Applied Economics and Finance*.<https://doi.org/10.48550/arXiv.2506.10369>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
- Keerthana, M & Vimala, J & Celestin, Prof & Boakye, Michael & Gidisu, Prof Jerryson. (2025). Utilizing Time Series Analysis Techniques for Accurate Budget Forecasting and Long-Term Planning in Public Sector Organizations. 1. 47-58. [10.5281/zenodo.15748950](https://doi.org/10.5281/zenodo.15748950).
- Lundberg, S.M. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4766-4777.
- Mahmud, Windu Gata, Hafifah Bella Novitasari, Sigit Kurniawan, & Dedi Dwi Saputra. (2024). Indonesian Government Revenue Prediction Using Long Short-Term Memory. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 14(1), 111–124. <https://doi.org/10.35585/inspir.v14i1.67>

- Overton, M., Larson, S., Carlson, L. J., et al. (2022). Public data primacy: The changing landscape of public service delivery as big data gets bigger. *Global Public Policy and Governance*, 2, 381–399, <https://doi.org/10.1007/s43508-022-00052-z>
- Piermarini, S., Ceccato, M., Marozzo, F., & Talia, D. (2023). Predicting Municipalities In Financial Distress: A Machine Learning Approach Enhanced By Domain Expertise. ArXiv preprint arXiv:2302.05780.
- Sarker, I.H. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN COMPUT. SCI.* 2, 377 (2021). <https://doi.org/10.1007/s42979-021-00765-8>
- Xu, Zongben & Tang, Niansheng & Xu, Chen & Cheng, Xueqi. (2021). Data science: Connotation, methods, technology, and development. *Journal of Information Technology and Data Management*. 1. 10.1016/j.dsm.2021.02.002.