



Implementasi Algoritma XGBoost Untuk Prediksi Capaian Bulanan Pendapatan Daerah Kota Bandung

Puteri Marchanda Izzati^{*1}, Fitriyani²

Email: ¹16213022@ars.ac.id, ²fitriyani@ars.ac.id

¹Sistem Informasi, Teknologi Informasi, Adhirajasa Reswara Sanjaya

²Sistem Informasi, Teknologi Informasi, Adhirajasa Reswara Sanjaya

Diterima: 01 Juli 2025 | Direvisi: - | Disetujui: 29 Juli 2025

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Pendapatan Asli Daerah (PAD) merupakan komponen utama pembiayaan pembangunan daerah. Di Kota Bandung, ketidaksesuaian antara target dan realisasi PAD masih menjadi tantangan dalam perencanaan fiskal. Penelitian ini bertujuan membangun model prediksi capaian bulanan PAD menggunakan algoritma XGBoost, yang dikenal efektif dalam menangani data non-linear. Data sekunder diperoleh dari Bapenda Kota Bandung, mencakup berbagai jenis pajak daerah tahun 2018–2024. Tahapan penelitian meliputi pembersihan data, rekayasa fitur, pembagian data, pelatihan model, dan evaluasi performa menggunakan metrik MAE, RMSE, dan R^2 . Hasil evaluasi menunjukkan MAE sebesar Rp5,6 miliar, RMSE Rp9,3 miliar, dan R^2 73% pada data uji. Validasi silang 5-fold menghasilkan MAE Rp3,49 miliar, RMSE Rp6,65 miliar, dan R^2 86%. Hasil ini membuktikan bahwa model memiliki akurasi dan kemampuan generalisasi yang baik. XGBoost dapat digunakan sebagai alat bantu strategis dalam perencanaan anggaran dan penyusunan APBD secara lebih presisi.

Kata kunci: *Machine Learning, Pendapatan Asli Daerah, Prediksi, Regresi XGBoost.*

Implementation of the XGBoost Algorithm for Predicting Monthly Regional Revenue Performance in Bandung City

Abstract

Local Own-Source Revenue (PAD) is a key pillar in financing regional development. In Bandung City, discrepancies between revenue targets and actual realization remain a challenge to effective fiscal planning. This study aims to develop a predictive model for monthly PAD achievement using the XGBoost algorithm, known for its strength in handling non-linear and complex data. The dataset, obtained from the Bandung Revenue Agency (Bapenda), includes various types of regional taxes from 2018 to 2024. The research process involved data cleaning, feature engineering, data splitting, model training, and performance evaluation using MAE, RMSE, and R^2 metrics. The evaluation on test data resulted in MAE of IDR 5.6 billion, RMSE of IDR 9.3 billion, and R^2 of 73%. Meanwhile, 5-fold cross-validation yielded MAE of IDR 3.49 billion, RMSE of IDR 6.65 billion, and R^2 of 86%. These results demonstrate high accuracy and generalization capability. XGBoost proves to be a reliable decision-support tool for data-driven fiscal planning.

Keywords: *Machine Learning, Prediction, Regional Original Revenue, Regression, XGBoost.*

1. PENDAHULUAN

Bandung, yang dikenal dengan julukan Paris van Java, memiliki potensi ekonomi yang signifikan, khususnya pada sektor pariwisata, perdagangan, dan jasa. Kondisi ini memperkuat urgensi optimalisasi Pendapatan Asli Daerah (PAD) sebagai sumber utama pembiayaan pembangunan daerah dan peningkatan kualitas pelayanan publik.[1]. Berdasarkan Undang-Undang Nomor 33 Tahun 2004, PAD terdiri dari pajak daerah, retribusi, hasil pengelolaan kekayaan daerah, serta pendapatan sah lainnya. Pengelolaan PAD di Kota Bandung menjadi tanggung jawab Badan Pengelolaan Pendapatan Daerah (BPPD), yang berperan dalam memaksimalkan penerimaan daerah melalui perencanaan dan pemantauan target serta realisasi pendapatan.[2]. Pendapatan Asli Daerah (PAD) memiliki peran krusial dalam memperkuat otonomi daerah. Semakin besar kontribusi PAD yang diperoleh, semakin tinggi pula tingkat kemandirian suatu daerah dalam menyelenggarakan fungsi pemerintahan dan pelaksanaan pembangunan.[3].

Realisasi Pendapatan Asli Daerah (PAD) Kota Bandung masih menghadapi sejumlah tantangan, di antaranya fluktuasi penerimaan serta keterbatasan pada metode proyeksi yang digunakan, yang berdampak pada kurang optimalnya perencanaan fiskal. Selain itu, tingkat kesadaran pajak masyarakat juga menjadi faktor yang memengaruhi capaian PAD. Dalam menghadapi permasalahan tersebut, diperlukan pendekatan berbasis data yang mampu menangkap pola kompleks dari data historis guna mendukung pengambilan keputusan yang lebih akurat dan adaptif.[4]. Tingkat kesadaran pajak merupakan salah satu determinan utama dalam pencapaian target dan realisasi Pendapatan Asli Daerah (PAD). Kesadaran pajak mencerminkan sejauh mana wajib pajak memiliki pemahaman, pengetahuan, dan kemampuan dalam menerapkan ketentuan perpajakan secara sukarela dan tepat. Peningkatan kesadaran ini berkontribusi pada peningkatan kepatuhan pajak, yang secara langsung mendorong kemauan membayar pajak dan mendukung optimalisasi penerimaan daerah.[5].

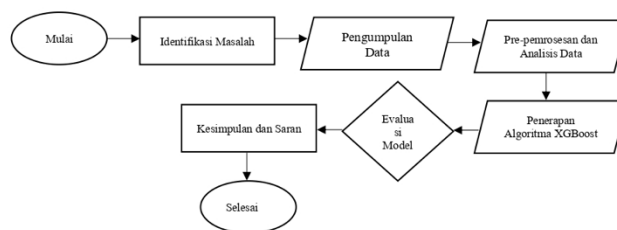
Algoritma Machine Learning (ML), khususnya Extreme Gradient Boosting (XGBoost), memiliki kapabilitas tinggi dalam melakukan prediksi akurat melalui pendekatan pemodelan non-linear dan efisiensi dalam pengolahan data. Sejumlah studi sebelumnya membuktikan keunggulan XGBoost dalam menyelesaikan permasalahan klasifikasi dan regresi di berbagai domain, seperti prediksi serangan siber, risiko gagal bayar, dan potensi bencana banjir. Namun, implementasi XGBoost dalam pengelolaan dan peramalan pendapatan daerah masih tergolong minim dan belum banyak dieksplorasi.[6].

Algoritma ini merupakan hasil pengembangan dari metode *gradient boosting decision tree* yang terkenal karena efisiensinya, kecepatan prosesnya, serta kemampuannya dalam memodelkan hubungan non-linear antar variabel dengan tingkat akurasi tinggi. Dalam konteks pengelolaan Pendapatan Asli Daerah (PAD), XGBoost memiliki potensi yang signifikan untuk membentuk model prediksi yang dapat memperkirakan target dan realisasi pendapatan secara tepat. Hal ini menjadikannya alat yang strategis dalam mendukung penyusunan Anggaran Pendapatan dan Belanja Daerah (APBD) yang lebih adaptif terhadap perubahan dan responsif terhadap kebutuhan daerah.[7].

Penelitian ini bertujuan merancang model prediksi capaian bulanan Pendapatan Asli Daerah (PAD) Kota Bandung dengan memanfaatkan algoritma XGBoost. Melalui pendekatan regresi, model ini dirancang untuk menangkap keterkaitan antara variabel-variabel historis pendapatan daerah guna menghasilkan proyeksi yang lebih presisi. Temuan dari penelitian ini diharapkan dapat menjadi instrumen pendukung dalam perencanaan fiskal dan penyusunan APBD yang lebih adaptif serta responsif terhadap dinamika perekonomian daerah.[8]. Selain itu, penelitian ini menawarkan pendekatan inovatif dalam penerapan algoritma XGBoost di sektor publik, khususnya untuk memprediksi capaian pendapatan daerah secara otomatis dan berkelanjutan tanpa ketergantungan pada evaluasi manual pasca realisasi anggaran.[9].

2. METODE PENELITIAN

Jenis penelitian ini merupakan penelitian kuantitatif dengan pendekatan prediktif, yang menggunakan data historis capaian Pendapatan Asli Daerah (PAD) secara time-series untuk membangun model regresi menggunakan algoritma XGBoost.[10]. Objek dalam penelitian ini adalah data capaian bulanan Pendapatan Asli Daerah (PAD) Kota Bandung yang dikelola oleh Badan Pendapatan Daerah (Bapenda) Kota Bandung dalam kurun waktu tahun 2018 hingga 2024. Pendapatan Asli Daerah merupakan indikator penting yang mencerminkan kemampuan daerah dalam membiayai pembangunan secara mandiri tanpa ketergantungan terhadap dana transfer dari pemerintah pusat.



Gambar 1. Flowchart Alur Penelitian

2.1. Identifikasi Masalah

Permasalahan utama yang diidentifikasi dalam penelitian ini adalah ketidaktepatan dalam memprediksi capaian bulanan Pendapatan Asli Daerah (PAD) Kota Bandung, yang ditunjukkan oleh deviasi signifikan antara target dan realisasi pendapatan. Ketidaksesuaian ini berdampak negatif terhadap efisiensi perencanaan anggaran, pelaksanaan program pembangunan, serta kualitas pelayanan publik. Analisis terhadap dokumen resmi pemerintah seperti APBD dan Laporan Realisasi Anggaran menunjukkan bahwa metode peramalan konvensional belum mampu menangkap dinamika dan kompleksitas data PAD secara akurat. Oleh karena itu, dibutuhkan pendekatan baru berbasis teknologi prediktif, seperti algoritma XGBoost, untuk meningkatkan akurasi estimasi pendapatan daerah dan mendukung pengambilan keputusan fiskal yang lebih tepat.

2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini bersifat sekunder dan diperoleh dari Badan Pendapatan Daerah (Bapenda) Kota Bandung serta dokumen resmi lainnya seperti APBD, Laporan Realisasi Anggaran (LRA), dan data statistik dari instansi terkait. Dataset mencakup capaian bulanan Pendapatan Asli Daerah (PAD) berdasarkan jenis pendapatan, tahun anggaran, dan kategori pajak, seperti pajak restoran, reklame, BPHTB, hotel, dan penerangan jalan, selama periode 2018-2024. Data ini digunakan untuk menangkap tren pendapatan daerah secara historis sebagai dasar analisis prediktif.

2.3. Pra-pemrosesan dan Analisis Data

Sebelum data digunakan dalam pelatihan model machine learning, dilakukan proses pra-pemrosesan untuk memastikan data dalam kondisi bersih, terstruktur, dan relevan. Tahapan ini meliputi:

1. Pembersihan data, termasuk penghapusan nilai kosong, duplikasi, dan penyesuaian tipe data menggunakan Python dan library *pandas*.
2. Transformasi data menggunakan metode Z-score normalization melalui *StandardScaler* dari Scikit-learn, guna menyeragamkan skala fitur numerik.
3. Rekayasa fitur, dengan mempertahankan fitur numerik yang relevan dan menghapus kolom non-informatif agar struktur data optimal bagi model XGBoost.
4. Pembagian dataset, dilakukan dengan rasio 80:20 untuk data latih dan uji, guna memastikan model dapat melakukan generalisasi dengan baik.

2.4. Penerapan Algoritma XGBoost

Model prediksi dikembangkan menggunakan algoritma XGBoost karena kemampuannya dalam menangani data numerik, pola non-linear, dan outlier. Pelatihan dilakukan dengan teknik *hyperparameter tuning* untuk mengoptimalkan kinerja model. Validasi dilakukan pada data uji guna memastikan akurasi prediksi capaian PAD Kota Bandung secara andal dan efisien.

2.5. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik Root Mean Squared Error (RMSE) untuk mengukur rata-rata selisih antara nilai aktual dan prediksi. Semakin kecil nilai RMSE, semakin akurat model dalam melakukan prediksi. Nilai RMSE dihitung secara otomatis selama proses pelatihan dan pengujian model.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \tag{1}$$

Dimana:

y_i = Nilai aktual,

\hat{y}_i = Nilai prediksi,

n = Jumlah data yang diteliti.

2.6 Kesimpulan dan Saran

Model XGBoost berhasil memprediksi capaian bulanan PAD Kota Bandung dengan tingkat akurasi yang baik berdasarkan evaluasi RMSE. Penelitian ini merekomendasikan penerapan model prediktif berbasis machine learning sebagai alat bantu dalam perencanaan fiskal Bapenda, guna menghasilkan target pendapatan yang lebih realistis dan mendukung kebijakan berbasis data.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pada tahap ini disajikan hasil dari proses analisis dan pengujian data yang telah dilakukan menggunakan algoritma XGBoost. Berikut merupakan implementasi model prediksi terhadap capaian Pendapatan Asli Daerah (PAD) Kota Bandung berdasarkan data historis. Data yang digunakan untuk pelatihan dan pengujian model mencakup periode tahun 2018 hingga 2024, sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Data Pendapatan Pajak Daerah Kota Bandung

No	Jenis Pajak	Target	Realisasi
1	HOTEL	375.000.000.000,00	424.004.747.247,00
2	RESTORAN	360.000.000.000,00	398.056.124.888,00
3	HIBURAN	70.000.000.000,00	67.823.508.868,00
4	REKLAME	50.000.000.000,00	45.718.375.580,00
5	PPJ	250.000.000.000,00	289.137.071.754,00
6	PARKIR	40.000.000.000,00	23.819.475.820,00
7	PAJAK AIR TANAH	35.000.000.000,00	27.690.547.532,00
8	PBB P2	600.000.000.000,00	542.814.240.037,00
9	BPHTB	878.126.837.178,00	579.899.638.472,00

3.2. Preprocessing Data

Tahap pra-pemrosesan data berperan krusial dalam menyiapkan data mentah agar dapat diolah secara optimal oleh algoritma machine learning, khususnya XGBoost. Proses ini bertujuan untuk mengonversi data mentah menjadi bentuk yang bersih, terstruktur, dan relevan dengan permasalahan yang akan diprediksi.

1. Ekstraksi Fitur Waktu

Kolom tanggal diubah menjadi dua fitur numerik, yaitu tahun dan bulan, guna membantu model mengenali pola musiman dan tren tahunan dalam data pendapatan daerah.

2. Pembuatan Fitur Baru

Untuk memperkaya informasi dari data yang ada, dua fitur turunan dibuat berdasarkan perbandingan antara capaian bulanan, yaitu:

a. Persentase Capaian

$$\text{Persentase Capaian} = \frac{\text{Realisasi}}{\text{Target}} \times 100\% \tag{2}$$

b. Selisih Capaian

$$\text{Selisih Capaian} = \text{Realisasi} - \text{Target} \tag{3}$$

Tabel 2. Pembuatan Fitur Baru

Tanggal	Jenis Pajak	Target/bulan	Realisasi/bulan	Selisih	%
2024-1	HOTEL	31.250.000.000,00	39.086.662.979,00	7836662979	125%
2024-2	RESTORAN	30.000.000.000,00	34.772.388.217,00	4772388217	116%
2024-3	HIBURAN	5.833.333.333,00	5.766.464.398,00	-66868935	99%
2024-4	REKLAME	4.166.666.666,00	3.722.748.305,00	-443918361	89%
2024-5	PPJ	20.833.333.333,00	26.399.681.112,00	5566347779	127%
2024-6	PARKIR	3.333.333.333,00	1.845.124.957,00	-1488208376	55%
2024-7	PAJAK AIR TANAH	2.196.666.666,00	2.211.788.330,00	15121664	101%
2024-8	PBB P2	50.000.000.000,00	23.451.143.680,00	-26548856320	47%
2024-9	BPHTB	73.177.236.431,00	97.128.877.831,00	23951641400	133%

3. Seleksi Fitur

Pada tahap ini dilakukan seleksi fitur dengan hanya menggunakan kolom-kolom bertipe numerik sebagai input, sesuai kebutuhan XGBoost Regressor. Fitur target yang diprediksi adalah kolom Realisasi Pendapatan Daerah, sementara fitur numerik lainnya digunakan sebagai variabel independen. Langkah ini memastikan model menerima input yang tepat untuk proses pelatihan.

4. Transformasi Data

Transformasi data dilakukan menggunakan metode Z-score normalization untuk menyeimbangkan skala antar fitur. Proses ini mencegah dominasi fitur tertentu dalam pelatihan model dan dilakukan dengan fungsi *StandardScaler* dari pustaka Scikit-Learn. Hasil transformasi membantu meningkatkan stabilitas dan konsistensi performa model XGBoost.

5. Pembagian Data Set

Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%) secara acak namun konsisten menggunakan *random_state*. Tujuannya adalah untuk melatih model dengan data yang cukup dan mengevaluasi kemampuannya dalam memprediksi data baru, sehingga hasil evaluasi lebih objektif dan tidak bias.

3.3. Implementasi Algoritma XGBoost

Setelah proses pra-pemrosesan selesai, model prediksi dibangun menggunakan algoritma XGBoost Regressor dengan pustaka Python yang terintegrasi dalam ekosistem Scikit-learn. Parameter dasar yang digunakan mencakup *objective='reg:squarederror'*, *eval_metric='rmse'*, dan *random_state=42* untuk memastikan hasil pelatihan yang konsisten.

Model dioptimalkan menggunakan Grid Search dengan 5-Fold Cross Validation untuk mencari kombinasi hyperparameter terbaik, seperti jumlah estimator, kedalaman pohon, learning rate, dan rasio subsample. Evaluasi performa dilakukan

menggunakan metrik MAE, RMSE, dan R² untuk menilai akurasi dan generalisasi model terhadap prediksi realisasi pendapatan daerah.

Tabel 3. Parameter XGBoost

No.	Nama Parameter	Nilai yang Diuji	Keterangan
1	n_estimators	100, 200, 300	Jumlah total pohon keputusan yang dibangun oleh model
2	max_depth	4, 6, 8	Kedalaman maksimum setiap pohon
3	learning_rate	0.01, 0.1, 0.2	Laju pembelajaran untuk setiap iterasi boosting
4	subsample	0.6, 0.8, 1.0	Proporsi data latih yang digunakan dalam setiap iterasi
5	colsample_bytree	0.6, 0.8, 1.0	Proporsi fitur yang digunakan untuk membangun setiap pohon

3.4. Evaluasi Model

Evaluasi model dilakukan terhadap data uji guna mengukur akurasi prediksi Realisasi Pendapatan Daerah, dengan menggunakan metrik MAE, RMSE, dan koefisien determinasi (R²) sebagai indikator kinerja.

Tabel 4. Nilai Evaluasi Model

Metrik Evaluasi	Nilai Hasil Model
MAE	5,6M
RMSE	9,3M
R ²	0,73

1. Perhitungan Evaluasi MAE

$$MAE = \frac{1}{n} \sum_{i=0}^n |y_i - y^{\wedge}i| \tag{4}$$

$$MAE = \frac{\sum |y_i - y^{\wedge}i|}{n} = \frac{112.013.435.510.800}{20} = 5.600.671.775,54$$

2. Perhitungan Evaluasi RMSE

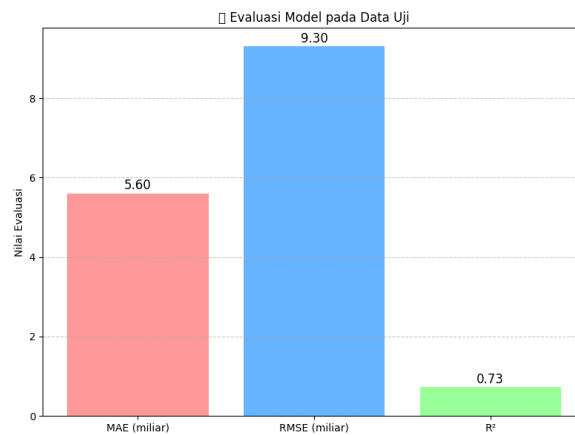
$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - y^{\wedge}i)^2}{n}} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} = \sqrt{\frac{1.1731 \times 10^{20}}{20}} = 9.300.930.296,08$$

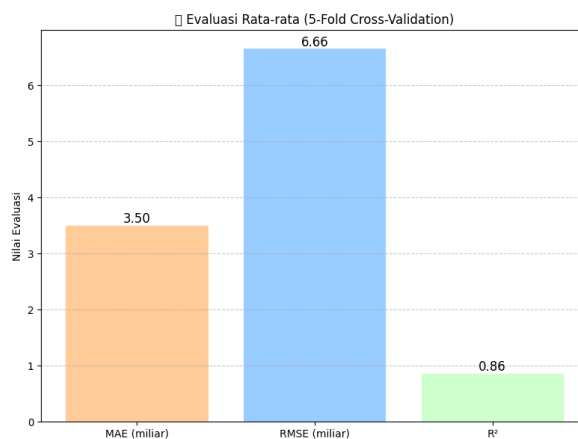
3. Perhitungan Evaluasi R Score

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{6}$$

$$R^2 = 1 - \frac{1731 \times 10^{20}}{6411 \times 10^{20}} = 1 - 0,27 = 0,73$$



Gambar 2. Hasil Evaluasi Model



Gambar 3. Hasil 5-Fold Cross-Validation

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma XGBoost Regressor efektif dalam membangun model prediksi capaian Pendapatan Asli Daerah (PAD) Kota Bandung berdasarkan data historis. Model memberikan hasil akurat dengan nilai MAE sebesar 5,6M, RMSE 9,3M, dan R² sebesar 0,73 pada data uji. Validasi menggunakan 5-Fold Cross-Validation menghasilkan performa yang lebih stabil dengan rata-rata MAE 3,4M, RMSE 6,6M, dan R² 0,86. Hasil ini membuktikan kemampuan generalisasi model yang baik tanpa overfitting. Selain akurasi, XGBoost juga menunjukkan efisiensi komputasi dan fleksibilitas parameter, sehingga layak diterapkan dalam pengambilan keputusan fiskal berbasis data di sektor pemerintahan daerah.

DAFTAR PUSTAKA

[1] S. V. Sofwan, M. Iqbaal, and S. Ramadhan, "PENGARUH PAJAK HOTEL DAN PAJAK RESTORAN TERHADAP PENDAPATAN ASLI DAERAH (PAD) PADA PEMERINTAH KOTA BANDUNG PERIODE 2013-2020,"

PENGARUH PAJAK HOTEL DAN PAJAK RESTORAN TERHADAP PENDAPATAN ASLI DAERAH (PAD) PADA PEMERINTAH KOTA BANDUNG PERIODE 2013-2020, 2021.

- [2] T. T. Wulansari and D. Novandi, "Implementasi Linear Regression Untuk Estimasi Realisasi Pendapatan Asli Daerah Pemerintah Kota," *Implementasi Linear Regression Untuk Estimasi Realisasi Pendapatan Asli Daerah Pemerintah Kota*, 2023.
- [3] M. Cahyaning, P. Putri, and ; Candradewini Candradewini, "PENDAPATAN DAERAH KOTA BANDUNG," 2025.
- [4] M. Ramdhani *et al.*, "Prediksi Capaian Bulanan Pajak Daerah Kabupaten Bandung Barat Menggunakan Metode Logistic Regression," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 881–890, 2024, doi: 10.47065/josh.v5i4.5330.
- [5] F. Fitriyani, "Analisis Kesadaran Pajak dan Dampak Kenaikan PPN terhadap Pola Konsumsi Masyarakat," *Jurnal Ekonomi Bisnis Antartika*, vol. 3, no. 1, pp. 69–74, Mar. 2025, doi: 10.70052/jeba.v3i1.698.
- [6] A. S. I. Jinan, "Studi Literatur Penerapan Machine Learning untuk Analisis Data Konsumen pada Minat Beli Konsumen," 2024, doi: 10.38035/jmpd.v2i4.
- [7] G. Dwilestari, "PREDIKSI ADOPSI HEWAN PELIHARAAN MENGGUNAKAN METODE XGBOOST," 2024.
- [8] S. Al Khairi, "PEMANFAATAN MACHINE LEARNING UNTUK PREDIKSI TARIF TOL MENGGUNAKAN ALGORITMA XGB REGRESSOR," 2024.
- [9] A. Santriawan, Gunadi Widi Nurcahyo, and Billy Hendrik, "Prediksi Penjualan Sepeda Motor Yamaha dengan Jaringan Syaraf Tiruan dan Backpropagation (Studi Kasus: CV Sinar Mas)," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 185–194, May 2024, doi: 10.37859/coscitech.v5i1.6709.
- [10] Muhammad Ramdani and Adhitia Erfina, "Pengaruh kualitas siak terhadap kepuasan mahasiswa di stikes sukabumi menggunakan regresi linier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 50–55, Apr. 2023, doi: 10.37859/coscitech.v4i1.4477.