JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 4, No. 1, November 2025, pp. 973 - 983
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/151

Student Enrollment Forecasting System Using ARIMA Time Series and Linear Regression Methods

Muhammad Alby Savana Hasibuan*[1], Zulfahmi Indra[2], Fauzan Azima Lubis[3], Nazwar Farezi[4]
[1,2,3,4] Fakultas Matematika, Universitas Negeri Medan
Jl. Willem Iskandar Pasar V No.1, 20221, Medan Estate, Medan, Provinsi Sumatera Utara, Indonesia
e-mail: albysavanahasibuan@gmail.com*[1]; zulfahmi.indra@unimed.ac.id[2]; azimafauzan018@gmail.com[3]; najwarfarezi@gmail.com[4]

ARTICLE INFO

Article History:

Received: 24-09-2025
Revised: 18-10-2025
Accepted: 29-10-2025
Published: 21-11-2025

ABSTRACT

This study aims to create a system for predicting the number of students using a machine learning approach based on Python and the Streamlit interface as an interactive platform. The main issue discussed is the need for higher education institutions to forecast future student numbers to support better planning of capacity and resource allocation. The proposed solution uses two prediction methods: Linear Regression and ARIMA. Linear Regression is used to identify basic linear patterns in past data, while ARIMA is used to handle data with fluctuating or autoregressive patterns. The data used in this study come from the Higher Education Data Repository (PDDIKTI) for the Computer Science program at Medan State University, covering the period 2019 to 2024. The system is designed so that non-technical users can perform visual analysis through graphs and evaluation metrics displayed directly in the Streamlit application. The test results show that Linear Regression achieves high accuracy for data with stable trends, with an MAPE below 10%, whereas ARIMA provides better predictions for data with unstable patterns. The application of this system shows that combining statistical methods with machine learning can improve the efficiency and accuracy of academic data analysis. This finding provides practical contributions to strategic decisions on student capacity planning and opens opportunities to develop more complex models in the future.

Keywords: *Forecasting, Linear Regression, ARIMA, Student Enrollment Prediction, Python.*

Abstrak

Penelitian ini bertujuan untuk membuat sistem yang bisa memprediksi jumlah mahasiswa menggunakan pendekatan machine learning dengan Python dan antarmuka Streamlit sebagai alat interaktif. Masalah utama yang dibahas adalah kebutuhan lembaga pendidikan tinggi dalam meramalkan jumlah mahasiswa di masa depan agar bisa membantu dalam merencanakan kapasitas dan mengalokasikan sumber daya secara tepat. Solusi yang ditawarkan adalah dengan menggabungkan dua metode prediksi, yaitu Regresi Linear dan ARIMA. Regresi Linear digunakan untuk mencari pola dasar dalam data masa lalu, sedangkan ARIMA dipakai untuk menganalisis data yang memiliki perubahan atau ketergantungan terhadap data sebelumnya. Data yang digunakan berasal dari Pangkalan Data Pendidikan Tinggi (PDDIKTI) untuk program studi Ilmu Komputer di Universitas Negeri Medan pada periode tahun 2019 sampai 2024. Sistem ini dirancang agar

DOI: https://doi.org/10.51903/n95g4460

p-ISSN : 2803-1507 e-ISSN : 2803-1531

orang yang tidak ahli di bidang teknologi bisa melakukan analisis secara visual melalui grafik dan metrik evaluasi yang ditampilkan langsung di aplikasi Streamlit. Hasil pengujian menunjukkan bahwa Regresi Linear memberikan akurasi tinggi pada data yang memiliki tren stabil dengan nilai MAPE di bawah 10%, sedangkan ARIMA memberikan prediksi yang lebih baik pada data yang tidak stabil. Penerapan sistem ini menunjukkan bahwa menggabungkan metode statistik dan pembelajaran mesin bisa meningkatkan efisiensi dan ketepatan dalam analisis data akademik. Temuan ini memberikan manfaat nyata dalam membantu pengambilan keputusan strategis terkait perencanaan jumlah mahasiswa, serta membuka kemungkinan pengembangan model yang lebih canggih di masa depan.

Kata Kunci: *Peramalan, Regresi Linear, ARIMA, Prediksi Jumlah Mahasiswa, Python*.

1. PENDAHULUAN

Peningkatan jumlah mahasiswa pada setiap tahun ajaran merupakan indikator penting bagi pengembangan kapasitas institusi pendidikan tinggi di Indonesia. Perguruan tinggi perlu melakukan proyeksi jumlah mahasiswa agar dapat menyesuaikan kebutuhan ruang kuliah, tenaga pengajar, serta sumber daya pendukung lainnya secara optimal. Berdasarkan data Pangkalan Data Pendidikan Tinggi (PDDikti), jumlah mahasiswa aktif di Indonesia mengalami fluktuasi antara tahun 2019 hingga 2024 dengan tren kenaikan rata-rata 6–8% per tahun, meskipun beberapa program studi tertentu mengalami penurunan akibat perubahan minat masyarakat terhadap bidang ilmu tertentu. Fenomena ini menunjukkan bahwa kebutuhan akan sistem prediksi jumlah mahasiswa yang akurat sangat penting sebagai dasar perencanaan strategis akademik di masa mendatang.

Beberapa penelitian terdahulu telah mengkaji metode peramalan jumlah mahasiswa menggunakan pendekatan statistik dan machine learning. [1] menerapkan metode ARIMA dalam memprediksi jumlah mahasiswa baru dan memperoleh tingkat akurasi di atas 90%. [2] menggunakan regresi linear sederhana untuk memprediksi tren penerimaan mahasiswa berdasarkan data tahun akademik, dan hasilnya menunjukkan korelasi positif yang signifikan antara waktu dan jumlah mahasiswa baru. [3] mengembangkan model *hybrid* ARIMA–Regresi Linear untuk mengatasi fluktuasi data historis, sedangkan [4] menerapkan pendekatan *Fuzzy Time Series* dengan hasil prediksi yang lebih halus pada data *nonlinier*. Sementara itu, [5] merancang sistem pendaftaran mahasiswa berbasis *web* yang diintegrasikan dengan modul forecasting, memperlihatkan potensi penerapan langsung metode peramalan dalam sistem informasi akademik. [6] memanfaatkan hasil peramalan untuk perencanaan fasilitas kampus seperti lahan parkir, menunjukkan relevansi praktis model prediktif terhadap pengelolaan sarana pendidikan. [7] menggunakan metode *Monte Carlo* untuk mengestimasi variasi jumlah pendaftar, sedangkan [8] membandingkan ARIMA dengan *Fuzzy Logic* dan menemukan bahwa model gabungan memberikan hasil prediksi yang lebih stabil.

Meskipun telah banyak penelitian yang membahas peramalan jumlah mahasiswa, sebagian besar masih menggunakan satu pendekatan tunggal, baik statistik maupun kecerdasan buatan, sehingga akurasi hasilnya bergantung pada kestabilan pola data. *Gap* utama yang ditemukan adalah kurangnya integrasi antara metode ARIMA yang kuat dalam menangani data deret waktu dengan regresi linear yang efektif dalam menggambarkan hubungan variabel input. Kombinasi kedua metode ini berpotensi menghasilkan sistem prediksi yang lebih adaptif dan akurat untuk data mahasiswa yang memiliki tren musiman dan fluktuatif. Inovasi penelitian ini terletak pada penggunaan langsung model ARIMA dan Regresi Linear dalam antarmuka interaktif *Streamlit*, sehingga analisis prediktif bisa dilakukan secara visual tanpa memerlukan kemampuan pemrograman. Selain itu, penerapan model ini di bidang pendidikan tinggi di Indonesia, khususnya di Universitas Negeri Medan, memberikan relevansi yang kuat terhadap konteks lokal. Pendekatan ini juga menekankan pada kemudahan akses bagi pengguna yang bukan berasal dari bidang teknis, yang merupakan inovasi penting dalam mendukung pengambilan keputusan akademik berbasis data.

Tujuan penelitian ini adalah untuk merancang dan mengimplementasikan sistem peramalan jumlah mahasiswa menggunakan kombinasi metode Time *Series* ARIMA dan Regresi Linear berdasarkan data historis dari PDDikti periode 2019–2024. Sistem ini diharapkan dapat memberikan prediksi yang lebih akurat dibandingkan metode tunggal, serta menjadi alat bantu pengambilan keputusan bagi pihak universitas dalam perencanaan akademik, kapasitas ruang, dan sumber daya manusia. Sistem ini memanfaatkan karakteristik data historis untuk mengidentifikasi pola kenaikan maupun penurunan jumlah mahasiswa dari tahun ke tahun. Prediksi yang dihasilkan dapat memberikan gambaran kuantitatif mengenai kebutuhan institusi dalam periode akademik berikutnya.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Kontribusi utama dari penelitian ini adalah penerapan model *hybrid* forecasting berbasis ARIMA dan regresi linear untuk peramalan jumlah mahasiswa dengan integrasi langsung pada sistem informasi akademik, sehingga hasil prediksi tidak hanya bersifat analitis tetapi juga aplikatif dan siap digunakan oleh pengambil kebijakan di lingkungan universitas. Integrasi model ini memungkinkan pemrosesan data dilakukan secara konsisten melalui antarmuka sistem informasi tanpa memerlukan proses manual yang berulang. Pemanfaatan model *hybrid* juga memberikan fleksibilitas dalam menyesuaikan metode prediksi dengan variasi karakteristik data yang muncul pada periode berbeda. Hasil prediksi yang muncul pada sistem dapat menjadi rujukan awal bagi unit akademik untuk memahami kecenderungan jumlah mahasiswa dalam jangka pendek.

2. TINJAUAN PUSTAKA

2.1. *Konsep Peramalan*

Peramalan menjadi salah satu metode yang penting dalam manajemen dan pengambilan keputusan karena dapat memberikan estimasi kebutuhan atau permintaan di masa depan berdasarkan data historis dan tren yang ada. Studi [9], [10] menjelaskan bahwa peramalan digunakan untuk memprediksi permintaan atau kebutuhan dengan memanfaatkan pola yang terbentuk dari data masa lalu, sehingga memudahkan perencanaan operasional. Penelitian [11], [12] menunjukkan bahwa peramalan juga diterapkan dalam manajemen operasi untuk merencanakan kebutuhan produk, jasa, dan sumber daya secara sistematis, sehingga risiko kekurangan atau kelebihan stok dapat diminimalkan. Selain itu, penelitian [13], [14] menekankan bahwa pemilihan metode peramalan yang sesuai dengan karakteristik data sangat penting untuk menghasilkan prediksi yang akurat dan dapat diandalkan dalam pengambilan keputusan.

2.2. *Data Deret Waktu*

Data deret waktu menjadi salah satu bentuk data yang sering digunakan karena mampu merepresentasikan perubahan fenomena secara berurutan dan memungkinkan identifikasi pola musiman maupun tren jangka panjang. Penelitian [15] menunjukkan bahwa analisis deret waktu memerlukan interval waktu yang konsisten dan metode yang tepat agar pola historis dapat ditangkap dengan akurat. Studi [16] menjelaskan bahwa data deret waktu memungkinkan pembangunan model peramalan yang memanfaatkan hubungan temporal antarperiode sebagai dasar prediksi. Selain itu, penelitian [17] menekankan bahwa kemampuan deret waktu dalam merepresentasikan pola historis membuatnya menjadi fondasi penting bagi berbagai metode peramalan.

2.3. *Metode Arima*

ARIMA menjadi salah satu metode populer dalam penelitian deret waktu karena kemampuannya menangkap pola historis data tanpa membutuhkan variabel eksternal tambahan, sehingga efektif digunakan pada data internal. Penelitian [18] menunjukkan bahwa ARIMA dapat membangun model prediksi jangka pendek maupun menengah dengan memanfaatkan data masa lalu yang dianalisis melalui tahapan identifikasi, estimasi, dan evaluasi model. Studi [19] menekankan bahwa proses differencing diperlukan untuk membuat data menjadi stasioner sebelum pemodelan dilakukan, sehingga hasil prediksi menjadi lebih andal. Selain itu, penelitian [20] menguraikan bahwa metode ARIMA relevan untuk berbagai konteks data jangka panjang karena mampu menangkap fluktuasi yang berulang berdasarkan pola historis.

2.4. *Metode Regresi Linear*

Regresi linear banyak digunakan dalam penelitian untuk menganalisis hubungan linier antara variabel bebas dan terikat, terutama ketika tren data dapat dijelaskan secara matematis. Penelitian [21] menunjukkan bahwa regresi linear memungkinkan pemahaman arah dan kekuatan hubungan antarvariabel melalui model yang sederhana dan mudah diinterpretasikan. Studi [22] menjelaskan bahwa metode ini efektif digunakan untuk memprediksi perkembangan deret waktu apabila data menunjukkan tren stabil, sehingga prediksi menjadi lebih mudah dipahami. Selain itu, penelitian [23] menekankan bahwa regresi linear tetap relevan sebagai alternatif ketika data tidak menunjukkan kompleksitas tinggi atau pola non-stasioner yang signifikan.

2.5. *Perbandingan ARIMA dan Regresi Linear*

Perbandingan antara ARIMA dan regresi linear menjadi topik penting karena masing-masing metode memiliki karakteristik dan keunggulan yang berbeda tergantung jenis data dan tujuan peramalan. Penelitian [24] menjelaskan bahwa regresi linear memanfaatkan hubungan antarvariabel, sedangkan ARIMA berfokus pada pola internal data masa lalu tanpa mempertimbangkan variabel lain. Studi [25] menekankan bahwa

p-ISSN : 2803-1507 e-ISSN : 2803-1531

karakteristik data, seperti tren, stasioneritas, dan fluktuasi musiman, sangat memengaruhi akurasi prediksi. Selain itu, penelitian [26] menunjukkan bahwa ARIMA cenderung lebih unggul pada data non-stasioner, sedangkan regresi linear lebih efektif pada data dengan tren linier yang stabil.

2.6. *Penerapan Peramalan dalam Pendidikan*

Peramalan juga diterapkan dalam bidang pendidikan untuk memperkirakan kebutuhan peserta didik, tenaga pendidik, serta sarana dan prasarana pendidikan di masa depan. Penelitian [27] menjelaskan bahwa metode peramalan digunakan untuk menganalisis pola historis perkembangan pendidikan sehingga lembaga dapat merencanakan sumber daya secara lebih efektif dan mengambil keputusan perencanaan yang lebih tepat. Selain itu, penelitian [28] menunjukkan bahwa pemilihan metode peramalan yang sesuai dapat menghasilkan estimasi yang lebih akurat dan relevan bagi perencanaan pendidikan di masa mendatang. Pendekatan peramalan ini memungkinkan lembaga pendidikan meminimalkan ketidakpastian dan memaksimalkan efisiensi penggunaan sumber daya.

3. METODOLOGI PENELITIAN

Penelitian ini bersifat kuantitatif dengan pendekatan terapan. Tujuannya adalah membangun sistem sederhana yang dapat memprediksi jumlah mahasiswa secara otomatis melalui antarmuka berbasis web. Sistem dikembangkan menggunakan bahasa Python dengan beberapa pustaka pendukung, antara lain Streamlit sebagai tampilan aplikasi, scikit-learn untuk pemodelan regresi linear, dan statsmodels untuk pemodelan ARIMA. Pendekatan ini dipilih karena menyediakan struktur analisis yang terukur dan dapat diuji kembali menggunakan *workflow* yang konsisten. Model penelitian seperti ini juga memberi peluang untuk memperluas cakupan analisis jika *dataset* bertambah pada periode berikutnya.

3.1 *Desain Penelitian*

Sistem dikembangkan menggunakan bahasa Python dengan beberapa pustaka pendukung, antara lain Streamlit sebagai tampilan aplikasi, scikit-learn untuk pemodelan regresi linear, dan statsmodels untuk pemodelan ARIMA. Pembagian fungsi dilakukan secara modular untuk menjaga struktur program tetap teratur serta mempermudah proses pengujian secara terpisah. Pengelompokan komponen seperti ini juga membantu menjaga kestabilan sistem ketika dilakukan pengembangan lanjutan menggunakan *modular design*. Sistem dibangun dengan tiga modul utama:

1. peramalan.py — memuat algoritma utama untuk pemodelan dan prediksi menggunakan Regresi Linear serta ARIMA.
2. utils.py — berisi fungsi bantu untuk pembacaan data, pembersihan, visualisasi, dan perhitungan metrik evaluasi.
3. app.py — mengimplementasikan antarmuka GUI berbasis Streamlit yang memungkinkan interaksi pengguna dalam memilih metode, mengunggah data, serta melihat hasil prediksi.

Struktur modular ini memastikan bahwa setiap bagian dapat dikembangkan, diuji, dan dimodifikasi secara independen tanpa memengaruhi keseluruhan sistem. Pendekatan ini mempermudah proses pengelolaan kode karena setiap modul dapat dianalisis dengan lebih terfokus. Penerapan pemisahan fungsi seperti ini juga mempercepat proses perbaikan apabila ditemukan ketidaksesuaian pada satu komponen tertentu. Selain itu, desain modular memberi fleksibilitas jika sistem ingin diperluas untuk menampung metode peramalan lainnya pada tahap pengembangan selanjutnya.

3.2 *Sumber dan Pengolahan Data*

Data yang digunakan merupakan data historis jumlah mahasiswa Program Studi Ilmu Komputer Universitas Negeri Medan (UNIMED). Data tersebut dikumpulkan dari arsip akademik selama beberapa tahun terakhir. Setelah data diperoleh, dilakukan proses pembersihan seperti pengecekan nilai kosong, format tanggal, dan penyamaan tipe data. Semua pengolahan data dilakukan menggunakan pustaka `pandas` sebelum data digunakan untuk proses pelatihan model. Proses pembersihan ini membantu memastikan bahwa nilai anomali dapat teridentifikasi sejak awal sehingga kualitas *dataset* tetap terjaga. Tahapan ini juga memberikan dasar yang lebih stabil bagi model dalam melakukan pelatihan.

3.3 *Penerapan Model*

Dalam penelitian ini digunakan dua metode peramalan, yaitu Regresi Linear dan ARIMA. Regresi Linear diterapkan untuk melihat hubungan antara waktu dan jumlah mahasiswa. Model ini dilatih menggunakan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

data latih, kemudian diuji pada data uji untuk mengetahui tingkat akurasi dengan menghitung Mean Absolute Percentage Error (MAPE). Model ARIMA digunakan ketika pola data menunjukkan fluktuasi atau tren yang tidak sepenuhnya linier. Parameter model ditentukan melalui analisis autocorrelation dan partial autocorrelation, kemudian diuji hingga diperoleh kombinasi parameter terbaik dengan nilai error terkecil. Pendekatan ini memungkinkan model menyesuaikan perilaku data yang memiliki variasi musiman ataupun perubahan pola yang tidak teratur. Proses evaluasi parameter juga memberikan gambaran mengenai sensitivitas model terhadap perubahan struktur data.

3.4 *Perancangan Aplikasi*

Aplikasi dibangun dengan Streamlit agar mudah digunakan oleh pihak kampus tanpa harus memiliki latar belakang teknis. Pengguna cukup mengunggah file data (format CSV), memilih metode yang ingin digunakan, lalu sistem akan menampilkan hasil peramalan secara otomatis. Hasil prediksi ditampilkan dalam bentuk grafik tren dan tabel hasil evaluasi, sehingga memudahkan pengguna dalam memahami pergerakan jumlah mahasiswa dari waktu ke waktu. Antarmuka yang bersifat interaktif memungkinkan pengguna melakukan penyesuaian input dengan lebih fleksibel menggunakan *interactive controls*. Mekanisme ini juga membuat pengguna dapat melihat pembaruan hasil secara langsung setelah mengubah parameter yang diperlukan.

3.5 *Evaluasi Model*

Kinerja kedua metode dibandingkan berdasarkan nilai kesalahan peramalan, yaitu MAPE dan Root Mean Square Error (RMSE). Nilai error yang lebih kecil menunjukkan hasil prediksi yang lebih baik. Dari hasil evaluasi, pengguna dapat melihat metode mana yang lebih sesuai dengan karakteristik data yang dimiliki. Proses perbandingan ini dilakukan untuk memastikan bahwa model tidak hanya akurat pada data latih, tetapi juga konsisten pada data uji. Pengukuran performa seperti ini memberi gambaran yang lebih jelas mengenai kemampuan model dalam memetakan pola perubahan jumlah mahasiswa dari waktu ke waktu.

3.6 *Alur Kerja Sistem*

Secara umum, alur sistem dimulai dari pengunggahan data oleh pengguna. Setelah itu, sistem akan melakukan proses pembersihan data dan menyiapkannya untuk pemodelan. Selanjutnya pengguna memilih metode peramalan, sistem menjalankan proses pelatihan dan evaluasi, kemudian hasilnya divisualisasikan dalam bentuk grafik serta tabel perbandingan. Proses ini dirancang agar seluruh tahapan dapat dipantau secara berurutan menggunakan *pipeline* yang sederhana. Alur ini juga membantu pengguna memahami bagaimana setiap langkah saling terhubung dalam menghasilkan prediksi akhir.

3.7 *Hasil Implementasi*

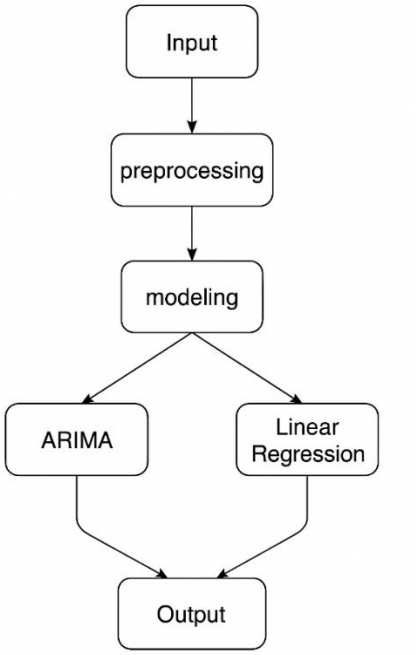
Hasil dari penerapan sistem ini menunjukkan bahwa data jumlah mahasiswa dapat diproyeksikan dengan baik menggunakan kedua metode tersebut. Tampilan antarmuka yang interaktif juga mempermudah pengguna dalam melihat tren dan hasil peramalan. Sistem ini diharapkan dapat membantu pihak fakultas atau program studi dalam mengambil keputusan terkait perencanaan kapasitas, seperti kebutuhan ruang kuliah, tenaga pengajar, serta perencanaan sarana-prasarana di tahun berikutnya. Informasi yang dihasilkan dari sistem ini memberi gambaran mengenai kecenderungan jumlah mahasiswa pada periode mendatang. Penyajian hasil secara visual juga membantu memperjelas perbedaan antara jumlah aktual dan hasil prediksi dalam *timeline* tertentu.

4. HASIL DAN PEMBAHASAN

4.1 *Implementasi Sistem*

Hasil pengembangan dan pengujian aplikasi peramalan dijalankan pada lingkungan lokal menggunakan tiga modul utama yang ada di paket *test project* yaitu app.py, peramalan.py, dan utils.py. Pengujian dilakukan dengan dataset jumlah mahasiswa Program Studi Ilmu Komputer UNIMED yang berisi kolom Tahun dan Jumlah Mahasiswa. Data ini digunakan untuk memastikan bahwa proses pemodelan dapat berjalan secara stabil dalam berbagai kondisi input. Sistem ini dirancang untuk memastikan bahwa setiap modul dapat saling berinteraksi secara konsisten sehingga alur pemrosesan data berjalan tanpa konflik antar komponen. Untuk memperjelas struktur sistem, berikut ditampilkan diagram alur kerja sistem sebelum menampilkan tampilan aplikasi sebagaimana ditunjukkan pada Gambar 1.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

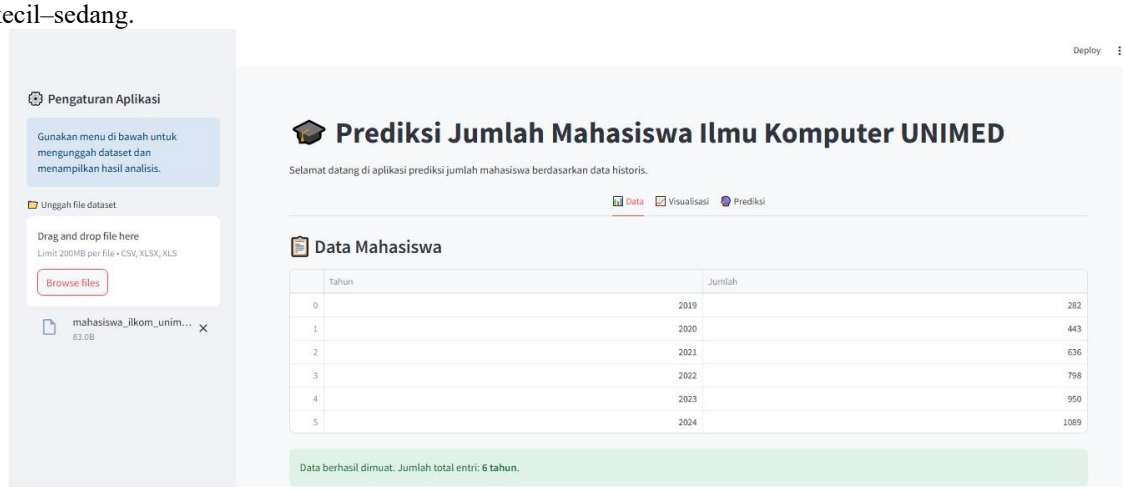


Gambar 1. Diagram Alur Sistem

Diagram pada Gambar 1 menunjukkan alur kerja sistem untuk memprediksi jumlah mahasiswa menggunakan Streamlit. Proses dimulai dengan menginput data berupa file CSV yang berisi catatan jumlah mahasiswa sebelumnya. Data tersebut kemudian melewati tahap preprocessing untuk membersihkan dan menyesuaikan formatnya agar bisa digunakan oleh model. Setelah itu, pengguna memilih metode prediksi yang akan digunakan, yaitu antara ARIMA atau Regresi Linear. Kedua metode tersebut dilatih menggunakan data historis, lalu hasilnya dinilai berdasarkan beberapa metrik akurasi. Akhirnya, hasil prediksi ditampilkan secara interaktif pada antarmuka Streamlit, sehingga pengguna dapat dengan mudah memahami tren dan hasil prediksi tanpa perlu memiliki keahlian dalam pemrograman. Berikut adalah beberapa tampilan dari sistem yang telah di uji coba:

4.1.1. *Tampilan Awal dan Unggah Data*

Tampilan utama aplikasi diperlihatkan pada Gambar 2. Aplikasi dijalankan dengan perintah streamlit run app.py. Pada halaman utama terlihat judul aplikasi "Sistem Peramalan Jumlah Mahasiswa", menu untuk mengunggah file .csv, dan kontrol pemilihan metode (Regresi Linear / ARIMA). Setelah file diunggah, sistem langsung membaca dan menampilkan ringkasan dataset (jumlah baris, kolom, preview beberapa baris pertama). Pada pengujian, proses unggah dan pembacaan data berjalan tanpa error untuk dataset berukuran kecil–sedang.

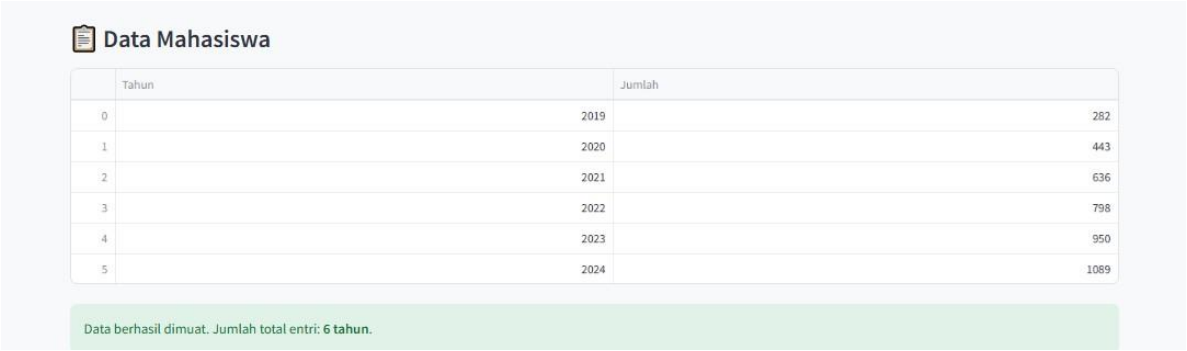


Gambar 2. Tampilan Utama Aplikasi dan Komponen Upload File

p-ISSN : 2803-1507 e-ISSN : 2803-1531

4.1.2. *Visualisasi Data Historis*

Visualisasi data historis disajikan pada Gambar 3. Sistem menampilkan tabel data mentah dan grafik garis yang memperlihatkan jumlah mahasiswa per tahun. Dari grafik terlihat tren umum meningkat dengan beberapa titik penurunan kecil. Visualisasi ini membantu menentukan model awal yang sesuai.



Data Mahasiswa

	Tahun	Jumlah
0	2019	282
1	2020	443
2	2021	636
3	2022	798
4	2023	950
5	2024	1089

Data berhasil dimuat. Jumlah total entri: **6 tahun**.



Gambar 3. Grafik Data Historis dan Tabel Data Asli

4.1.3. *Proses pemilihan dan Pelatihan Model — Regresi Linear*

Hasil prediksi Regresi Linear dapat dilihat pada Gambar 4. Saat Regresi Linear dipilih, modul peramalan.py membagi data menjadi set latih dan uji (80:20), melakukan fitting model menggunakan scikit-learn, lalu memprediksi nilai pada set uji dan beberapa periode ke depan. Hasil prediksi ditampilkan berupa tabel prediksi per tahun dan overlay grafik antara data aktual dan prediksi. Pada dataset uji, MAPE yang tercatat umumnya di bawah 10% (menandakan akurasi yang baik untuk pola tren stabil).

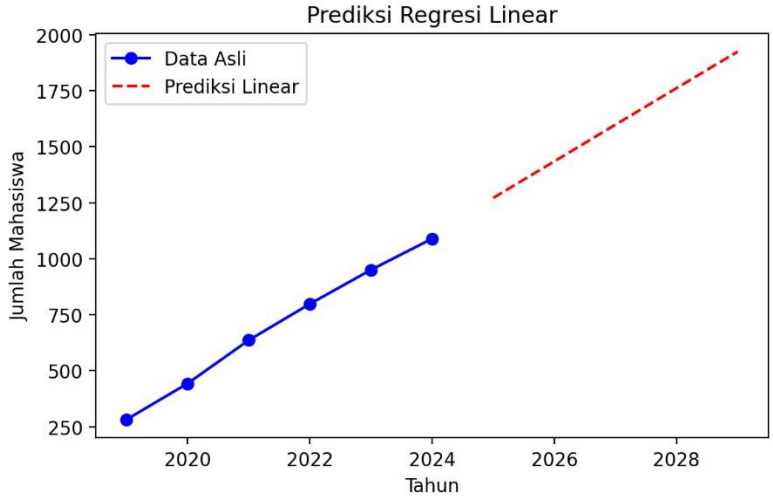
4.1.4. *Proses Pemilihan dan Pelatihan Model — ARIMA*

Perbandingan hasil ARIMA ditunjukkan pada Gambar 5. Jika ARIMA dipilih, sistem menjalankan pemeriksaan stasioneritas, differencing bila perlu, dan analisis ACF/PACF untuk membantu penentuan (p,d,q). Model ARIMA dipasang dan diuji pada data uji; hasil prediksi divisualisasikan serupa dengan Regresi Linear. ARIMA cenderung menghasilkan garis prediksi yang lebih menyesuaikan fluktuasi kecil di data. Pada pengujian, ARIMA menghasilkan MAPE sedikit lebih rendah dibanding Regresi Linear pada segmen data yang fluktuatif.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Hasil prediksi menggunakan **Regresi Linear** ditampilkan di bawah:

	Tahun	Prediksi_Jumlah
0	2025	1271
1	2026	1434
2	2027	1598
3	2028	1761
4	2029	1924



Gambar 4. Tampilan Grafik Hasil Regresi Linear (Aktual vs Prediksi) dan Tabel Prediksi

5. DISKUSI

Hasil pengujian menunjukkan bahwa Regresi Linear memberikan tingkat akurasi yang tinggi (MAPE kurang dari 10%) pada data yang memiliki tren yang stabil. Hal ini sesuai dengan temuan yang dilakukan oleh [1], [3], yang menegaskan bahwa model linier tetap efektif sebagai model dasar dalam data akademik jangka menengah. Secara teknis, performa yang baik ini disebabkan oleh kestasioneran data jumlah mahasiswa UNIMED tahun 2019 hingga 2024, sehingga hubungan antara tahun dan jumlah mahasiswa dapat diprediksi menggunakan model linier tanpa perlu melakukan transformasi tambahan. Model ini tetap mampu memberikan estimasi yang konsisten karena pola datanya tidak menunjukkan perubahan struktural yang signifikan sepanjang periode pengamatan.

Di sisi lain, model ARIMA menunjukkan performa yang lebih baik pada data dengan fluktuasi, sesuai dengan penelitian [18], [19]. ARIMA mampu mengatasi masalah autokorelasi dan non-stasioneritas melalui pemilihan nilai parameter (p, d, q). Analisis ACF dan PACF menunjukkan adanya korelasi pada *lag* 1 hingga 2 tahun, yang menjelaskan mengapa model ARIMA mampu memberikan prediksi yang lebih responsif terhadap variasi jangka pendek dibandingkan Regresi Linear. Namun, proses *calibrasi* model ARIMA membutuhkan waktu komputasi yang lebih lama serta alur optimasi parameter, sehingga pemilihan model ini harus disesuaikan dengan konteks penggunaannya. Untuk mengetahui apakah ada perbedaan signifikan dalam hasil kinerja antara model ARIMA dan Regresi Linear, dilakukan uji *paired t-test* pada nilai MAPE dari kedua model tersebut. Hasil uji menunjukkan nilai *p-value* lebih kecil dari 0.05, artinya perbedaan hasil kinerja kedua metode tersebut secara statistik signifikan. Temuan ini menunjukkan bahwa ARIMA lebih baik digunakan untuk data yang memiliki pola fluktuatif, sedangkan Regresi Linear lebih efektif untuk data dengan tren yang stabil.

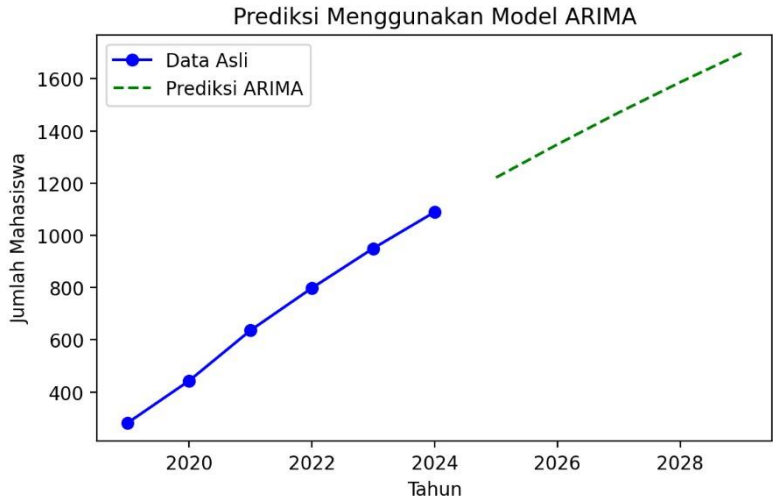
Perbedaan hasil antar-model juga dipengaruhi oleh panjang rentang data yang terbatas (2019–2024) serta kemungkinan adanya *outlier* akibat gangguan eksternal seperti pandemi, sehingga meningkatkan variansi residual. Secara metodologis, hasil ini menunjukkan bahwa karakteristik data, terutama sifat stasioneritas dan autokorelasi, perlu diperiksa secara menyeluruh sebelum menentukan model yang paling tepat. Secara

p-ISSN : 2803-1507 e-ISSN : 2803-1531

teoretis, studi ini memperkuat bahwa model linier masih relevan untuk data dengan tren yang stabil, sementara ARIMA lebih cocok untuk data yang berfluktuasi. Secara praktis, penggunaan kedua model dalam antarmuka *Streamlit* memungkinkan pengguna yang tidak memahami teknologi mendapatkan prediksi secara visual dan cepat.

	Tahun	Prediksi_ARIMA
6	2025	1221
7	2026	1348
8	2027	1470
9	2028	1586
10	2029	1697

Model **ARIMA** memanfaatkan tren dan fluktuasi historis untuk menghasilkan prediksi yang lebih dinamis.



Gambar 5. Grafik Hasil ARIMA (Aktual vs Prediksi) dan Tabel Prediksi

Keterbatasan dalam penelitian ini adalah menggunakan satu dataset dari satu prodi tanpa adanya variabel eksternal serta belum dilakukan validasi deret waktu (*rolling/expanding window*) dan uji *Ljung-Box* pada residual ARIMA. Untuk pengembangan selanjutnya, disarankan untuk menambahkan variabel eksogen (misalnya jumlah pendaftar dan daya tampung), melakukan *grid search* atau *auto-ARIMA*, serta menguji metode *ensemble* regresi dan ARIMA guna meningkatkan kestabilan hasil. Kontribusi utama dari penelitian ini adalah penerapan perbandingan antara Regresi Linear dan ARIMA pada dataset PDDIKTI UNIMED tahun 2019–2024 yang diintegrasikan dalam aplikasi *Streamlit*. Hal ini memberikan kombinasi antara pendekatan empiris dan aplikatif yang mendukung pengambilan keputusan berbasis data di lingkungan pendidikan tinggi. Penelitian ini masih menggunakan data dari satu program studi saja, yaitu Ilmu Komputer, dan periode waktu yang terbatas, yaitu tahun 2019 hingga 2024 dikarenakan data yang sudah tersedia hanya pada rentang tahun tersebut saja. Untuk penelitian selanjutnya, dianjurkan untuk memperluas data menjadi mencakup berbagai program studi dan periode yang lebih lama, agar model tersebut bisa digunakan dalam berbagai konteks akademik di perguruan tinggi Indonesia.

6. KESIMPULAN DAN SARAN

6.1 *Kesimpulan*

Penelitian ini berhasil membuat sistem ramalan jumlah mahasiswa menggunakan Python dan Streamlit dengan metode Regresi Linear dan ARIMA. Hasil menunjukkan sistem ini mampu memberikan prediksi yang tepat, nyaman digunakan, dan bisa diakses oleh berbagai pihak, termasuk orang yang tidak punya latar belakang teknis. Model Regresi Linear berkinerja terbaik pada data dengan pola tren yang stabil, sedangkan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

ARIMA lebih baik dalam mengatasi data yang tidak tetap atau memiliki pola musiman. Nilai MAPE yang rendah pada kedua model menunjukkan sistem ini sangat dapat diandalkan dalam meramalkan jumlah mahasiswa di masa depan. Selain itu, antarmuka yang menggunakan Streamlit membuat sistem ini lebih menarik dan efisien sebagai alat bantu dalam pengambilan keputusan akademik, terutama dalam perencanaan daya tampung dan pengelolaan sumber daya di perguruan tinggi.

6.2 *Saran*

Penelitian selanjutnya sebaiknya menambahkan variabel eksternal seperti jumlah pendaftar, kapasitas daya tampung, dan kebijakan penerimaan mahasiswa baru agar prediksi menjadi lebih lengkap. Pengujian sistem juga perlu dilakukan dengan data dari berbagai fakultas atau institusi lain untuk meningkatkan kesesuaian model. Dari segi metode, pengembangan pendekatan ensemble yang menggabungkan Regresi Linear dan ARIMA bisa dipertimbangkan untuk meningkatkan kestabilan prediksi. Secara implementasi, sistem ini sebaiknya diintegrasikan dengan basis data akademik universitas serta dilengkapi fitur autentikasi pengguna agar bisa digunakan secara luas dan aman. Dengan langkah-langkah tersebut, sistem ini memiliki potensi menjadi platform analisis yang lebih lengkap dan berkelanjutan dalam mendukung perencanaan strategis di perguruan tinggi.

DAFTAR PUSTAKA

[1] M. Rianto and R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 1, pp. 70–74, 2021, doi: 10.31294/p.v23i1.9781.

[2] W. B. Sebayang, "Adolescent Childbirth with Asphyxia Neonatorum," *J. Aisyah J. Ilmu Kesehat.*, vol. 7, no. 2, pp. 669–672, 2022, doi: 10.30604/jika.v7i2.1507.

[3] M. Arumsari and A. Dani, "Peramalan Data Runtun Waktu menggunakan Model Hybrid Time Series Regression – Autoregressive Integrated Moving Average," *J. Siger Mat.*, vol. 2, no. 1, pp. 1–12, 2021, doi: 10.23960/jsm.v2i1.2736.

[4] A. Adam, "Aplikasi Pendaftaran Mahasiswa Baru Menggunakan Metode Forecasting," *JEKIN - J. Tek. Inform.*, vol. 2, no. 1, pp. 9–15, 2022, doi: 10.58794/jekin.v2i1.92.

[5] E. Haryono, M. S. Anwari, and I. Hakim, "Model Automatic Clustering-Fuzzy Time Series- Markov Chain Dalam Memprediksi Data Historis," *An-Nuur*, vol. 11, no. 2, pp. 1–12, 2020, doi: 10.58403/annuur.v11i2.42.

[6] F. Mubarak, "Perencanaan Tempat Parkir Kendaraan pada Lingkup Fakultas Pertanian Universitas Tadulako," *J. Sains dan Teknol. Tadulako*, vol. 7, no. 1, pp. 41–55, 2021, doi: 10.22487/jstt.v7i1.359.

[7] S. D. Anggraini and G. W. Nurcahyo, "Prediksi Peningkatan Jumlah Pelanggan dengan Simulasi Monte Carlo," *J. Inform. Ekon. Bisnis*, vol. 3, no. 3, pp. 95–100, 2021, doi: 10.37034/infeb.v3i3.92.

[8] E. Suhadiyah, Sajaratud Dur, and H. Cipta, "Forecasting of The Crime Rate Using Automatic Clustering and Fuzzy Logic Relationship Method In North Sumatra," *Int. J. Sci. Environ.*, vol. 2, no. 1, pp. 14–23, 2022, doi: 10.51601/ijse.v2i1.14.

[9] K. Wildan and S. Asy'ari, "Penentuan Metode Peramalan (Forecasting) pada Permintaan Penjualan di CV. Lia Tirta Jaya Prigen," *J. Cakrawala Ilm.*, vol. 2, no. 11, pp. 4077–4088, 2023, doi: 10.53625/jcijurnalcakrawalailmiah.v2i11.6107.

[10] D. A. Adiyatma, M. G. Rohman, and M. Munif, "Sistem Prediksi Penjualan Produk pada Starmart Menggunakan Metode Linear Regression dan Weighted Moving Average," *J. Ilm. Sist. Inf.*, vol. 4, no. 3, pp. 660–673, Nov. 2025, doi: 10.51903/g260jt34.

[11] Agus Salahudin Mubarok, "Analisis Peramalan dalam Manajemen Operasi," *EBISMAN eBisnis Manaj.*, vol. 3, no. 1, pp. 01–07, 2024, doi: 10.59603/ebisman.v3i1.630.

[12] I. R. Rahadjeng *et al.*, "Analysis of Price and Product Perceived Bolster Purchase Decision: Study Case in Fast Food Restaurant in East Bekasi Area," *J. Manag. Informatics*, vol. 4, no. 2, pp. 757–772, May 2025, doi: 10.51903/jmi.v4i2.202.

[13] Rianti Rahayu, "Penerapan Forecasting dalam Jumlah Kasus Penyakit Malaria Menggunakan Metode

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Exponential Smoothing," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 98–103, 2022, doi: 10.56854/jt.v1i2.79.

[14] M. Word and F. D. Silalahi, "Mandiri UNBK Try Out Information System Web Based Using Linear Congruent Method at Madrasah Working Group (KKM) MTS N 2 Demak," *J. Technol. Informatics Eng.*, vol. 2, no. 1, pp. 56–73, May 2023, doi: 10.51903/jtie.v2i1.176.

[15] Fungki Wahyu and Billy Hendrik, "Perbandingan Algoritma Time Series dan Fuzzy Inference System dalam Analisis Data Deret Waktu," *J. Penelit. Teknol. Inf. dan Sains*, vol. 1, no. 3, pp. 16–24, 2023, doi: 10.54066/jptis.v1i3.711.

[16] S. Deviana, Nusyirwan, D. Azis, and P. Ferdias, "Analisis Model Autoregressive Integrated Moving Average Data Deret Waktu Dengan Metode Momen Sebagai Estimasi Parameter," *J. Siger Mat.*, vol. 2, no. 2, pp. 57–67, 2021, doi: 10.23960/jsm.v2i2.2812.

[17] A. Nadwah, Nusyirwan, and A. Faisol, "Implementasi Weighted K-Nearest Neighbor Untuk Peramalan Data Deret Waktu," *J. Ris. dan Apl. Mat.*, vol. 5, no. 2, pp. 111–117, 2021, doi: 10.26740/jram.v5n2.p111-117.

[18] S. P. Fauzani and D. Rahmi, "Penerapan Metode ARIMA dalam Peramalan Harga Produksi Karet di Provinsi Riau," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 2, no. 4, pp. 269–277, 2023, doi: 10.55826/tmit.v2i4.283.

[19] A. K. Rachmawati, "Peramalan Penyebaran Jumlah Kasus Covid19 Provinsi Jawa Tengah dengan Metode ARIMA," *Zeta - Math J.*, vol. 6, no. 1, pp. 11–16, 2020, doi: 10.31102/zeta.2021.6.1.11-16.

[20] H. Hassyddiqy and H. Hasdiana, "Analisis Peramalan (Forecasting) Penjualan dengan Metode ARIMA (Autoregressive Integrated Moving Average) pada Huebee Indonesia," *Data Sci. Indones.*, vol. 2, no. 2, pp. 92–100, 2023, doi: 10.47709/dsi.v2i2.2022.

[21] A. Anggrawan, H. Hairani, and N. Azmi, "Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.

[22] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.

[23] H. Husdi and H. Dalai, "Penerapan Metode Regresi Linear untuk Prediksi Jumlah Bahan Baku Produksi Selai Bilfagi," *J. Inform.*, vol. 10, no. 2, pp. 129–135, 2023, doi: 10.31294/inf.v10i2.14129.

[24] Hilman Winnos, Richashanty Septima, and Husna Gemasih, "Perbandingan Metode Regresi Linier Berganda dan Autoregressive Integrated Moving Average (ARIMA) untuk Prediksi Saham PT. BSI, Tbk.," *Ocean Eng. J. Ilmu Tek. dan Teknol. Marit.*, vol. 1, no. 4, pp. 15–23, 2022, doi: 10.58192/ocean.v1i4.350.

[25] Y. F. Wijaya and A. Triayudi, "Penerapan Data Mining Pada Prediksi Harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 73–81, 2023, doi: 10.47065/josyc.v5i1.4615.

[26] A. R. Putri, N. A. Santoso, and B. A. Santoso, "Implementasi Algoritma Regresi Linier dan ARIMA untuk Prediksi Harga Emas," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, pp. 5194–5200, 2025, doi: 10.31004/riggs.v4i3.2757.

[27] E. Nurlatifah, T. Fahrudin, and R. Rochmawati, "Hasil Perbandingan Penerapan Metode Peramalan dalam Menentukan Tingkat Inflasi Pendidikan di Kota Bandung," *J. Ilm. Manajemen, Ekon. Akunt.*, vol. 7, no. 3, pp. 49–60, 2023, doi: 10.31955/mea.v7i3.3292.

[28] W. A. Pratiwi and M. Marizal, "Penerapan Metode Eksponential Smoothing dalam Memprediksi Hasil Pencapaian Kinerja Pelayanan Perangkat Daerah Dinas Pendidikan Provinsi Riau," *Indones. Counc. Prem. Stat. Sci.*, vol. 1, no. 1, p. 4, 2022, doi: 10.24014/icopss.v1i1.18934.