

SISTEM PREDIKSI JUMLAH SISWA BARU MENGGUNAKAN METODE *LONG SHORT TERM MEMORY* BERDASARKAN *DATA TIME SERIES*

Ja'far Maulana Ihsan¹, Freddy Wicaksono², Agust Isa Martinus³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Cirebon Jl. Fatahillah, Watubelah, Kec. Sumber, Cirebon, Jawa Barat, Indonesia, 45611

¹ixanxnax299@gmail.com, ²freddy.wicaksono@umc.ac.id, ³agust.isa@umc.ac.id

ABSTRACT

Education plays a crucial role in improving the quality of human resources. One of the indicators of a primary school's success is the stability of new student enrollment each year. However, fluctuations in student numbers often become obstacles in school planning, including facilities, teaching staff, and budgeting. This study aims to develop a prediction system for new student enrollment at SDN 1 Kedungjaya using the Long Short Term Memory (LSTM) method with a time series data approach. The data used consist of new student enrollment records from several previous years, followed by preprocessing, model training, and evaluation using Mean Squared Error (MSE). The results show that the LSTM model is capable of predicting new student enrollment with relatively low error rates. The system is implemented as a web-based application, enabling schools to utilize it as a decision support tool for future enrollment planning.

Keywords : Prediction, LSTM, Time Series, New Student Enrollment, Web based System

ABSTRAK

Pendidikan memiliki peran penting dalam meningkatkan kualitas sumber daya manusia. Salah satunya indikator keberhasilan sekolah dasar adalah stabilitas jumlah siswa baru yang mendaftar setiap tahunnya. Namun, fluktuasi jumlah siswa sering menjadi kendala dalam perencanaan sekolah, baik dari sisi sarana, tenaga pendidik, maupun anggaran. Penelitian ini bertujuan membangun sistem prediksi jumlah siswa baru di SDN 1 Kedungjaya menggunakan metode *Long Short Term Memory (LSTM)* serta pendekatan data *time series*. Data yang digunakan berupa jumlah siswa baru dari beberapa tahun terakhir, kemudian dilakukan tahap praproses data, pelatihan model, dan evaluasi menggunakan *Mean Squared Error (MSE)*. Hasil penelitian menunjukkan bahwa model *LSTM* mampu memprediksi jumlah siswa baru dengan tingkat kesalahan yang relatif rendah. Sistem ini diimplementasikan dalam bentuk aplikasi berbasis web sehingga dapat digunakan pihak sekolah untuk mendukung pengambilan keputusan dan perencanaan penerimaan siswa baru di masa mendatang.

Kata Kunci : Prediksi, LSTM, Time Series, Jumlah Siswa Baru, Sistem Berbasis Web

1. PENDAHULUAN

Pendidikan dasar memegang peran penting dalam membentuk kualitas sumber daya manusia di Indonesia. Sekolah Dasar (SD) menjadi fondasi utama dalam menanamkan pengetahuan, keterampilan, dan karakter generasi muda. Berdasarkan data Kemendikbudristek tahun ajaran 2024/2025, terdapat sekitar 23,93 juta siswa SD atau 45% dari total peserta didik di seluruh Indonesia. Hal ini menegaskan bahwa jenjang SD memiliki porsi terbesar dalam sistem pendidikan nasional sehingga perencanaannya perlu dilakukan secara matang.

Salah satu aspek penting dalam perencanaan pendidikan adalah prediksi jumlah siswa baru. Fluktuasi jumlah siswa dapat menimbulkan masalah, seperti keterbatasan ruang kelas, tenaga pendidik, maupun sarana prasarana. Sebaliknya, penurunan jumlah siswa berisiko pada efisiensi sumber daya dan keberlangsungan sekolah. Oleh karena itu, prediksi jumlah siswa baru menjadi faktor strategis dalam mendukung pengambilan keputusan pendidikan.

Metode prediksi tradisional, seperti *ARIMA* atau *Exponential Smoothing*, memiliki keterbatasan dalam menangkap pola data *non-linear* yang kompleks. Sebagai alternatif, metode kecerdasan buatan

berbasis jaringan saraf tiruan, khususnya *Long Short Term Memory (LSTM)*, mampu mengatasi kelemahan model konvensional dan memberikan hasil prediksi lebih akurat pada data *time series*.

Penelitian ini mengusulkan sistem prediksi jumlah siswa baru menggunakan metode *LSTM* berdasarkan data *time series*. Sistem ini diharapkan dapat membantu sekolah maupun instansi terkait dalam merencanakan kebutuhan sumber daya pendidikan serta mendukung kebijakan pemerataan pendidikan.

2. METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen. Tujuannya adalah mengembangkan sistem prediksi jumlah peserta didik baru menggunakan metode *Long Short-Term Memory (LSTM)* berbasis data deret waktu. Data historis pendaftaran siswa dimanfaatkan sebagai bahan pelatihan dan pengujian model prediksi, sekaligus untuk menilai tingkat akurasi dibandingkan dengan data sebenarnya.

a. Data Selection

Dataset yang dipakai dalam penelitian ini terdiri atas data historis jumlah siswa baru di SDN 1 Kedungjaya dari tahun 2020 sampai 2025 dengan total 6 data observasi. Dataset tersebut memuat 5 variabel utama:

1. Tahun: Periode waktu pengamatan (2020-2025)
2. Jumlah_siswa: Target variable yang akan diprediksi (berkisar 62-85 siswa)
3. Jumlah_penduduk: Fitur demografis (169.871-190.627 jiwa)
4. Laki_laki: Jumlah penduduk laki-laki (87.997-97.932 jiwa)
5. Perempuan: Jumlah penduduk perempuan (81.874-92.695 jiwa)

Data ini dipilih karena menunjukkan hubungan korelatif antara jumlah siswa baru dengan faktor demografis penduduk, yang merupakan indikator relevan untuk prediksi jumlah siswa.

b. Data preprocessing

Dalam tahap ini dilakukan data *preprocessing* yang meliputi proses pembersihan data, penanganan terhadap kemungkinan adanya data yang hilang atau tidak konsisten, serta penghapusan nilai duplikat. Selanjutnya dilakukan proses normalisasi data agar setiap variabel, seperti jumlah siswa baru, jumlah penduduk, jumlah siswa laki-laki, dan perempuan berada pada rentang nilai yang sama. Tahapan ini bertujuan untuk meningkatkan kualitas data sehingga model *Long Short Term Memory (LSTM)* dapat lebih optimal dalam mengenali pola dan menghasilkan prediksi jumlah siswa baru secara akurat. Tahap *preprocessing* dilakukan melalui fungsi `prepare_data()` yang mengimplementasikan:

1. Normalisasi data: Menggunakan *MinMaxScaler* untuk menormalkan semua variabel ke rentang $[0,1]$, memastikan tidak ada bias terhadap variabel dengan magnitudo yang lebih besar
2. Validasi konsistensi: Pemeriksaan terhadap *missing values*, *outliers*, dan konsistensi format data

Normalisasi dilakukan secara terpisah untuk features (X) dan target (y) menggunakan *scaler* yang berbeda untuk memungkinkan *inverse transformation* yang akurat pada tahap prediksi.

c. Transformasi data

Dalam tahap ini dilakukan transformasi data dengan menyusun ulang dataset ke dalam format *Comma Separated Values (CSV)* agar lebih mudah diproses oleh sistem. Dataset disusun dengan struktur kolom berupa tahun, jumlah_siswa, jumlah_penduduk, laki_laki, dan perempuan. Proses ini memastikan data memiliki format yang konsisten serta dapat langsung dibaca oleh program *Python* untuk tahap pemodelan menggunakan *Long Short Term Memory (LSTM)*. Transformasi data mencakup beberapa tahap kritical:

1. Format *standardization*: Data dalam format *CSV* dikonversi menjadi struktur yang kompatibel dengan *TensorFlow/Keras*
2. *Sequence generation*: Implementasi *sliding window approach* dimana setiap *input sequence* (t) dipasangkan dengan target (t+1)
3. *Reshaping* untuk *LSTM*: Data direstruktur menjadi format 3D (*samples, timesteps, features*) sesuai *requirement LSTM*
4. *Train-test splitting*: Pemisahan data untuk *training* dan evaluasi model

Struktur sekuensial dibuat dengan pendekatan: `X.append(features_scaled[i])` dan `y.append(target_scaled[i+1])` untuk menciptakan pola temporal *learning*.

d. Data minning

Tahap data mining merupakan proses pemodelan terhadap data yang telah melalui tahap *preprocessing* dan transformasi. Pada penelitian ini, pemodelan dilakukan menggunakan bahasa pemrograman *Python* dengan menerapkan metode *Long Short-Term Memory (LSTM)*. Metode ini dipilih karena memiliki kemampuan dalam mengolah data deret waktu (*time series*) dan mengenali pola jangka panjang, sehingga dapat menghasilkan prediksi jumlah siswa baru dengan tingkat akurasi yang lebih baik.

Implementasi arsitektur *LSTM* dengan spesifikasi:

1. Model arsitektur: *Sequential model* dengan 1 *LSTM layer* (50 units) + 1 *Dense layer* (1 unit)
2. *Activation function*: *ReLU* untuk *LSTM layer*
3. *Optimizer*: *Adam* dengan *learning rate default*
4. *Loss function*: *Mean Squared Error (MSE)*
5. *Training parameters*: 200 *epochs* dengan *verbose=0* untuk efisiensi

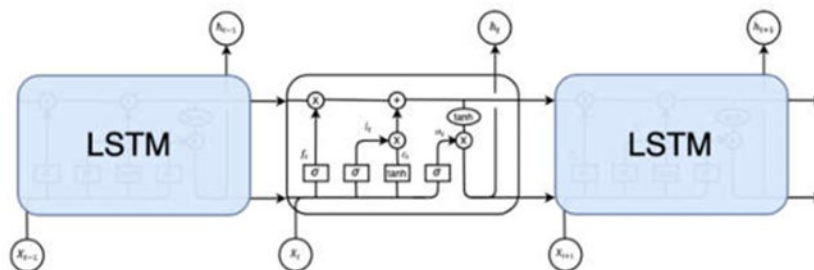
Model dirancang untuk menangkap temporal dependencies dalam data *enrollment* dengan mempertimbangkan faktor demografis sebagai *predictor variables*.

2.1 Long Short Term Memory

Struktur *LSTM* memiliki tiga jenis gerbang, yaitu gerbang *input* (*input gate*), gerbang lupa (*forget gate*), dan gerbang *output* (*output gate*). Gerbang *input* berfungsi untuk mengatur seberapa banyak informasi yang akan disimpan ke dalam memori jangka panjang. Gerbang lupa bertugas menentukan informasi mana yang perlu dihapus dari memori jangka panjang. Sedangkan gerbang *output* mengatur jumlah informasi yang dikeluarkan dari memori jangka panjang. Lapisan dalam *LSTM* sendiri terdiri atas lapisan *input*, lapisan tersembunyi, dan lapisan *output*, sebagaimana ditampilkan pada gambar berikut.

Gambar 1 Arsitektur Long Short Term Memory

2.2 Evaluasi Model Prediksi



Evaluasi model merupakan tahap penting dalam menguji kinerja suatu model machine learning. Proses ini bertujuan untuk menilai sejauh mana model yang dibangun mampu memberikan prediksi yang akurat dan dapat diandalkan berdasarkan data yang digunakan. Pada penelitian ini, evaluasi difokuskan pada kinerja model *Long Short-Term Memory (LSTM)* dalam memprediksi jumlah siswa baru menggunakan data deret waktu.

a. Evaluasi Metriks

Model yang telah dilatih kemudian diuji menggunakan beberapa metrik evaluasi, yaitu *Mean Squared Error (MSE)*, *Mean Absolute Percentage Error (MAPE)*, dan koefisien determinasi (R^2). Dalam tahap pelatihan, *MSE* digunakan sebagai fungsi *loss* untuk meminimalkan rata-rata kuadrat selisih antara nilai aktual dan hasil prediksi. Selanjutnya, *MAPE* dan R^2 dimanfaatkan pada tahap evaluasi untuk mengukur besarnya kesalahan relatif dalam bentuk persentase, serta menilai sejauh mana model mampu menjelaskan variasi data yang ada.

2.3 Arsitektur Sistem

Arsitektur sistem menggambarkan hubungan antara komponen-komponen utama yang membentuk sistem, baik dari sisi perangkat lunak (*software*) maupun proses komunikasi antar bagian. Arsitektur ini dirancang untuk memisahkan proses pengolahan data (*backend*) dari tampilan antarmuka pengguna (*frontend*), sehingga sistem menjadi lebih terstruktur, mudah dikelola, dan dapat dikembangkan di masa mendatang.

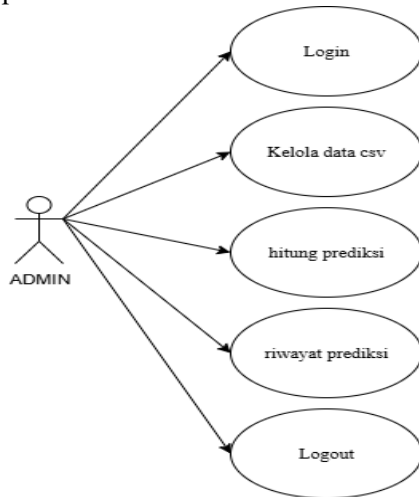
Secara umum, sistem prediksi jumlah siswa baru ini menggunakan arsitektur berbasis *client-server* yang diimplementasikan dalam bentuk *RESTful API*. *Frontend* berperan sebagai *client* yang

menerima masukan dari pengguna, menampilkan hasil prediksi, dan menyediakan fitur interaktif seperti grafik dan tabel riwayat. *Backend* berperan sebagai server yang mengelola data, menjalankan logika bisnis, memproses prediksi menggunakan *machine learning*, dan mengirimkan hasil kembali ke *frontend*.

2.4 Perancangan Sistem

a. Use Case Diagram

Use case diagram menunjukkan hubungan fungsionalitas sistem dengan aktor yang menggunakannya. Pada sistem ini aktor utamanya adalah admin. Beberapa *use case* (kasus penggunaan) utama meliputi kelola data *csv*, menghitung prediksi, melihat riwayat hasil prediksi.



Gambar 2 Use Case Diagram

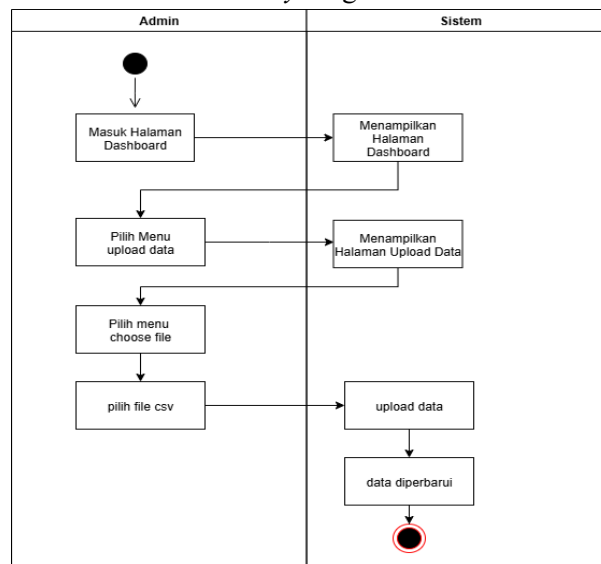
b. Activity Diagram

Activity diagram memodelkan rangkaian aktivitas dalam sistem secara berurutan.

1. Activity diagram kelola data *csv*

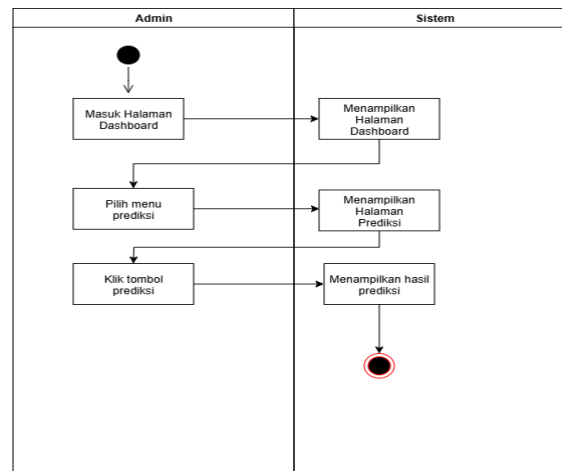
Proses pengelolaan data CSV pada sistem ini diawali ketika admin login ke aplikasi dan mengakses halaman dashboard. Setelah itu, admin memilih fitur unggah data yang telah disediakan dalam sistem. Setelah data *csv* di pilih, admin kemudian menekan tombol *upload* data untuk memperbarui data *csv*, kemudian sistem merespons dengan menampilkan notifikasi bahwa data *csv* diperbarui dan siap dipakai untuk proses prediksi.

Gambar 3 Activity diagram kelola data *csv*



2. Activity diagram hitung prediksi

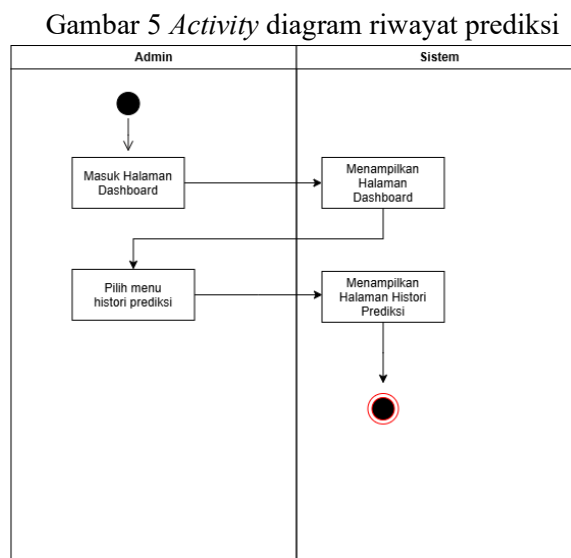
Proses *activity* diagram hitung prediksi jumlah siswa baru dalam sistem ini dimulai ketika admin mengakses aplikasi dan masuk ke halaman *dashboard*. Setelah berhasil masuk, admin memilih menu prediksi yang tersedia di dalam sistem. Setelah data dimasukkan dengan lengkap, admin kemudian menekan tombol prediksi untuk memulai proses perhitungan, kemudian menampilkan hasil prediksi jumlah siswa baru berdasarkan data historis sebelumnya yang dimasukkan oleh admin.



Gambar 4 Activity diagram hitung prediksi

3. Activity diagram riwayat prediksi

Proses diawali ketika admin masuk ke dalam sistem dan mengakses halaman *dashboard*. Setelah berhasil masuk, admin memilih riwayat hasil prediksi yang tersedia pada tampilan sistem. Sistem kemudian merespons dengan menampilkan halaman riwayat yang memuat daftar hasil prediksi yang telah dilakukan sebelumnya.



Gambar 5 Activity diagram riwayat prediksi

2.5 Kode Program

Program berikut membangun sebuah model prediksi berbasis jaringan saraf *Long Short-Term Memory (LSTM)* yang dirancang untuk mengolah data *time series* dengan arsitektur sederhana menggunakan satu lapisan *LSTM* dan satu lapisan *Dense* sebagai *output*, kemudian mengevaluasi kinerjanya melalui perhitungan metrik *Mean Squared Error (MSE)*, *Mean Absolute Percentage*

Error (*MAPE*), dan koefisien determinasi (R^2) berdasarkan hasil prediksi dibandingkan dengan data aktual.

Gambar 6 Kode Program model *LSTM*
Gambar 7 Kode Program evaluasi model

```
def build_model(input_shape):
    model = Sequential([
        LSTM(50, activation='relu', input_shape=input_shape), # LSTM
        Dense(1)
    ])
    model.compile(optimizer='adam', loss='mse')
# Evaluasi model (MSE, MAPE, R2)
y_pred_scaled = model.predict(X, verbose=0)
y_pred = scaler_y.inverse_transform(y_pred_scaled)
y_true = scaler_y.inverse_transform(y)

mse = mean_squared_error(y_true, y_pred)
mape = mean_absolute_percentage_error(y_true, y_pred)
r2 = r2_score(y_true, y_pred)
```

3. HASIL DAN PEMBAHASAN

3.1. Input Data

Data jumlah siswa baru yang digunakan dalam pembahasan ini adalah sebagai berikut:

Tabel 1 Data Siswa

Tahun	Siswa	Jumlah Penduduk	Penduduk Laki-laki	Penduduk Perempuan
2020	62	169.871	87.997	81.874
2021	68	172.918	89.045	83.873
2022	75	171.833	88.642	83.371
2023	70	171.835	88.491	83.344
2024	81	184.482	94.825	89.657
2025	85	190.627	97.932	92.695

3.2. Data normalisasi

Untuk menyesuaikan data dengan skala *input LSTM*, dilakukan normalisasi menggunakan metode *Min-Max Scaling*, hasilnya sebagai berikut:

Tabel 2 Normalisasi data

Tahun	Siswa	Jumlah Penduduk	Penduduk Laki-laki	Penduduk Perempuan
2020	0.0000	0.0000	0.0000	0.0000
2021	0.2609	0.1469	0.1055	0.1847
2022	0.5652	0.0946	0.0468	0.1384
2023	0.3478	0.0947	0.0497	0.1358
2024	0.8261	0.7042	0.6871	0.7189
2025	1.0000	1.0000	1.0000	1.0000

3.3. Dataset Komprehensif

Dataset ini menunjukkan hubungan multi-dimensional antara pertumbuhan demografis dan partisipasi pendidikan dari 2020-2025:

- Jumlah Siswa: Target prediksi dengan tren growth 37% (62→85 siswa)

- b. Jumlah Penduduk: Basis demografis dengan pertumbuhan 12% overall, spike signifikan 2024
- c. Distribusi Gender: Komposisi laki-laki (~51%) dan perempuan (~49%) relatif stabil
- d. Rasio Partisipasi: Persentase siswa terhadap populasi menunjukkan tren positif dari 0.037% ke 0.045%

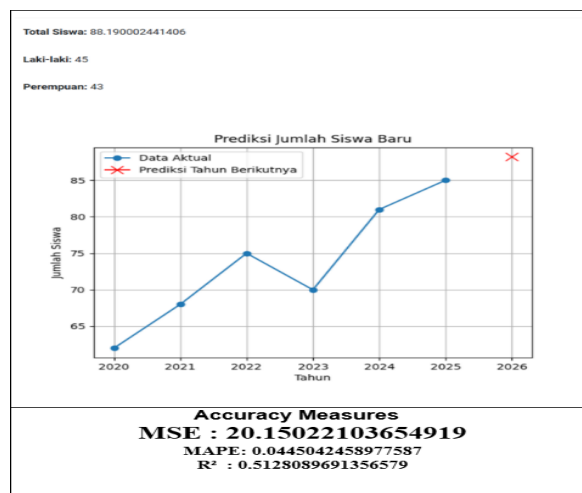
Tabel 3 Dataset Komprehensif

Tahun	Siswa	Jumlah Penduduk	Penduduk Laki-laki	Penduduk Perempuan	Rasio
2020	62	169.871	87.997	81.874	0.037%
2021	68	172.918	89.045	83.873	0.039%
2022	75	171.833	88.642	83.371	0.044%
2023	70	171.835	88.491	83.344	0.041%
2024	81	184.482	94.825	89.657	0.044%
2025	85	190.627	97.932	92.695	0.045%

3.4. Hasil Evaluasi Model

Hasil evaluasi model dengan algoritma *Long Short Term Memory (LSTM)* menunjukkan bahwa nilai *MSE* yang diperoleh sebesar 20.15, yang berarti rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual relatif kecil. Nilai *MAPE* sebesar 0.0445 atau sekitar 4,45%, mengindikasikan bahwa tingkat kesalahan relatif model dalam memprediksi data cukup rendah. Sementara itu, nilai R^2 yang diperoleh sebesar 0.51, menunjukkan bahwa model mampu menjelaskan sekitar 51% variasi data. Dengan demikian, model ini dapat dikatakan bekerja cukup baik karena menghasilkan error yang rendah serta kemampuan penjelasan data yang moderat.

Gambar 8 Hasil Evaluasi Model

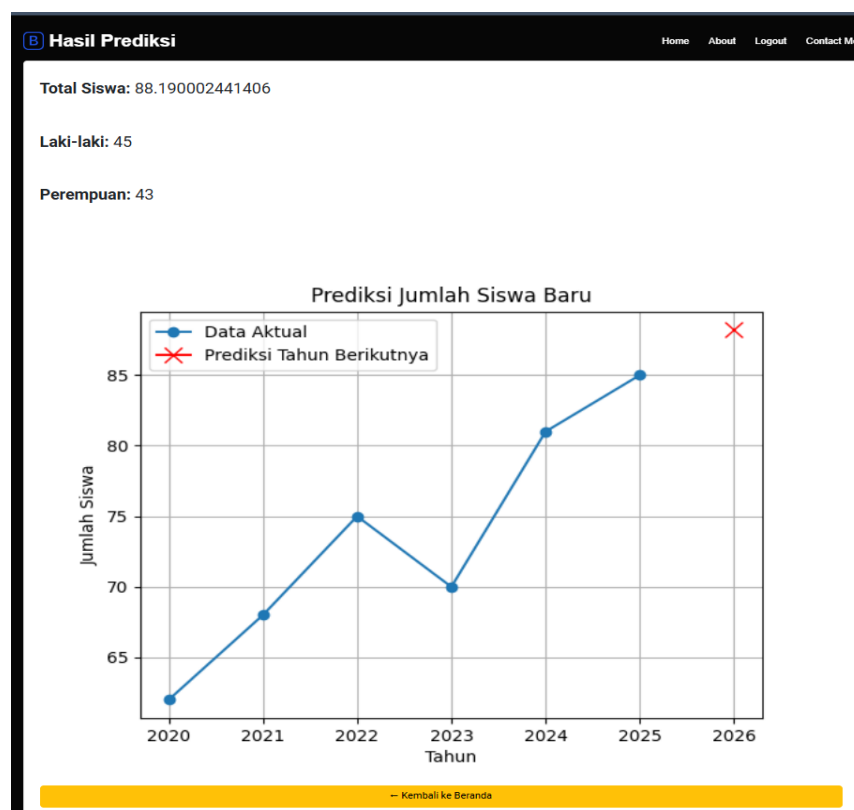


3.5. Hasil Pengujian

Tabel 4 validasi hasil pengujian

Nama Proyek	: Sistem Prediksi Jumlah Siswa Baru				
Id Kasus	: 09				
Prioritas Uji	: Tinggi				
Nama Modul	: Prediksi				
Nama Uji	: Validasi hasil prediksi				
Deskripsi	: Memastikan hasil prediksi ditampilkan sesuai perhitungan				
Pelaksana Uji	: Jafar				
Tanggal Uji	: 20 Agustus				
Kondisi Awal	: Proses prediksi telah selesai dilakukan				
No	Langkah Uji	Data Uji	Hasil yang diharapkan	Hasil yang terjadi	Status
1	Membuka halaman hasil prediksi	-	Hasil prediksi ditampilkan dalam tabel/grafik	Tabel hasil prediksi dan grafik muncul	Sukses
2	Membandingkan hasil	Data prediksi model	Hasil prediksi sesuai <i>output</i> dari model machine learning	Hasil prediksi sesuai data perhitungan model	Sukses
Kondisi Akhir: Hasil prediksi tervalidasi sesuai <i>output</i> model dan dapat ditampilkan dalam bentuk tabel maupun grafik					

Berdasarkan pengujian, hasil prediksi ditampilkan sesuai dengan perhitungan model. Sistem mampu menyajikan data dalam bentuk visualisasi yang jelas dan mudah dipahami pengguna.



Gambar 9 Hasil Pengujian

4. KESIMPULAN

Sistem prediksi jumlah siswa baru berhasil dibangun dengan memanfaatkan data historis *time series* melalui proses normalisasi, pelatihan model, dan menghasilkan prediksi periode berikutnya dalam antarmuka web yang mudah diakses. Hasil pengujian menunjukkan tingkat kesalahan yang relatif rendah, sehingga sistem ini dapat mendukung sekolah dalam perencanaan kapasitas kelas, pengadaan sarana prasarana, serta pengaturan sumber daya guru.

5. REFERENSI

- [1] Aji Santoso, Ade Irma Purnamasari, Irfan Ali.2024.”Sistem Prediksi Harga Beras Menggunakan Metode Recurrent Neural Network dan Long Short Term Memory”. e-ISSN : 2597-9922.
- [2] Sarwo, Hermawan.2016.”Prediksi Penerimaan Siswa Baru Pada Madrasah Aliyah AS-SAYAFI’IYAH 02 Menggunakan Metode Time Series.ISSN 1978-9262.
- [3] Yulistika Handini, Jusuf Wahyudi, Jhoanne Fredricka.2024.”Prediksi Jumlah Siswa Baru Menggunakan Metode Time Series”.e-ISSN 2615-3262.
- [4] Razanul Al Kiramy, Inggih Permana, Arif Marsal.2024.”Perbandingan Performa Algoritma RNN dan LSTM dalam Prediksi Jumlah Jamaah Umrah pada PT.Hajar Aswad”. ISSN(P): 2797-2313 | ISSN(E): 2775-8575.
- [5] M Abdul Dwiyanto Suyudi, Esmeralda C.Djamal, Asri Maspupah.2019.”Prediksi Harga Saham Menggunakan Metode Recurrent Neural Network”.ISSN: 1907-5022.
- [6] Abdillah Baradja, Sukoco.2023.”Pemanfaatan Recurrent Neural Network (RNN) Untuk Meningkatkan Akurasi Prediksi Mata Uang Pada Forex Trading”.e-ISSN: 2775-2588.