

PERBANDINGAN ALGORITMA SIMPLE LINEAR REGRESSION DAN SUPPORT VECTOR REGRESSION DALAM PREDIKSI JUMLAH PENDUDUK DI SULAWESI TENGGARA

Ahmad Fadli Ramadhan¹, Hashimatul Zaria², Rafi Iyad Madani Chaidir³, Rizal Adi Saputra^{4*}

^{1,2,3,4*} Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo

¹ahmadfadlyramadhan@gmail.com, ²hashimatulzaria@gmail.com,
³rafiyadmadani@gmail.com, ^{4*}rizaladisaputra@uho.ac.id

ABSTRACT

Central Statistics Agency (BPS) statistics shows that Southeast Sulawesi's population has been growing steadily, rising by about 8% from 2014 to 2023. It is important to pay attention to the possible effects of this phenomena on local life and regional growth. The long-term planning of infrastructure development, resource management, and allocation can all benefit by forecasting the population's future numbers. The goal of this research is to develop a reliable model that can accurately predict population growth. The data used consist of the population figures for Southeast Sulawesi from 2014 to 2023, categorized by regency/city. The Support Vector Regression (SVR) and Simple Linear Regression (SLR) algorithms were used to make the predictions. With an average Root Mean Square Error (RMSE) of 0.51% and Mean Absolute Percentage Error (MAPE) of 1.89%, the comparison results show that SLR performs better than SVR in the majority of models. These findings indicate the SLR is a more precise algorithm for forecasting Southeast Sulawesi population growth.

Keywords: *Population Growth, Prediction, Simple Linear Regression, Support Vector Regression*

I. PENDAHULUAN

Penduduk adalah kelompok manusia yang mendiami suatu wilayah tertentu dan tunduk pada peraturan yang berlaku, serta terlibat dalam interaksi yang berkesinambungan satu sama lain. Dalam konteks sosiologi, penduduk merujuk pada kumpulan individu yang menempati suatu wilayah geografi dan ruang spesifik [1].

Pertumbuhan penduduk adalah suatu keadaan di mana jumlah penduduk meningkat atau berkurang dari satu periode ke periode berikutnya, yang dinyatakan dalam bentuk persentase (%). Pertumbuhan penduduk terjadi akibat adanya pergerakan penduduk, melibatkan kelahiran, kematian, dan migrasi penduduk. Jika jumlah kelahiran lebih besar daripada jumlah kematian, maka populasi akan bertambah. Sebaliknya, jika jumlah kematian lebih besar daripada jumlah kelahiran, populasi akan berkurang. Hal yang sama berlaku untuk migrasi. Jika jumlah migrasi masuk lebih besar daripada migrasi keluar, maka jumlah penduduk akan bertambah. Sebaliknya, jika migrasi keluar lebih besar daripada migrasi masuk, jumlah penduduk akan berkurang [2].

Indonesia berada pada peringkat keempat setelah India, Tiongkok, dan Amerika Serikat sebagai negara dengan populasi terbesar di dunia. Saat ini, jumlah penduduk Indonesia mencapai 277,7 juta jiwa yang tersebar di seluruh wilayah [3].

Berdasarkan data populasi Indonesia tahun 2023 [4], provinsi Sulawesi Tenggara berada pada peringkat ke-22 dalam hal jumlah penduduk di Indonesia. Dari total 38 provinsi, Sulawesi Tenggara memiliki populasi sebanyak 2.726.590 jiwa.

Peramalan dapat diartikan sebagai prediksi masa depan seakurat mungkin dengan mempertimbangkan semua informasi yang tersedia, termasuk data historis dan pengetahuan tentang peristiwa masa depan yang mungkin memengaruhi hasil prediksi [5]. Penafsiran lain dari

peramalan, yaitu teknik analisis untuk memproyeksikan kejadian masa depan dengan menggunakan data masa lalu, baik secara kualitatif maupun kuantitatif, guna mengurangi ketidakpastian [6]. Peramalan memiliki penting yang besar dalam tata kelola, tidak hanya saat merencanakan, tetapi juga untuk manajemen strategis serta bahkan untuk manajemen risiko.

Ada banyak cara untuk meramalkan dalam melakukan prediksi. Salah satu metode peramalan yang dapat digunakan adalah Simple Linear Regression (SLR) dan Support Vector Regression (SVR). SLR menggali hubungan antara variabel respons dengan satu atau lebih variabel independen, sementara SVR menggunakan metode yang menghasilkan model regresi untuk meramalkan kuantitas di masa yang akan datang berdasarkan data sebelumnya [7][8]. Dengan membandingkan keduanya, dapat menentukan pendekatan yang paling cocok dan melakukan prediksi jumlah penduduk yang lebih akurat.

Pada prinsipnya, jumlah penduduk di suatu wilayah akan terus meningkat setiap tahunnya, dan pertumbuhan ini dapat memiliki dampak positif atau negatif terhadap suatu wilayah. Penanganan yang tepat terhadap pertumbuhan penduduk dapat meningkatkan kualitas suatu wilayah, baik dari segi ekonomi dengan peningkatan pendapatan per kapita seiring dengan pertumbuhan jumlah penduduk, maupun dalam hal inovasi dengan munculnya ide-ide baru yang mendorong dari pertumbuhan penduduk di suatu wilayah. Namun, jika tidak dipersiapkan dengan baik sebelumnya, sebaliknya, dapat menurunkan kualitas wilayah tersebut [9].

Perencanaan terstruktur yang tepat oleh pemerintah dapat mencegah konsekuensi negatif dari pertumbuhan penduduk yang tidak diantisipasi dan tidak direncanakan sebelumnya. Untuk membuat rencana yang sesuai untuk masa depan, diperlukan proyeksi jumlah penduduk, yang merupakan

perkiraan jumlah penduduk yang akan datang yang didasarkan pada data sebelumnya. Proyeksi ini digunakan oleh pemerintah untuk mengestimasi permintaan sumber daya seperti makanan, air, energi, dan pelayanan, serta untuk membuat rancangan perencanaan kota.

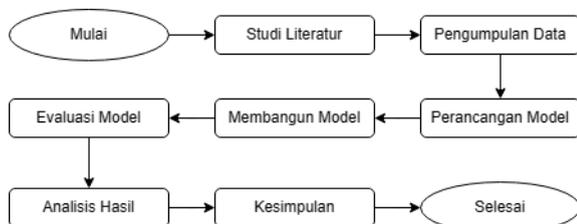
Penelitian untuk melakukan prediksi menggunakan metode Simple Linear Regression dan Support Vector Regression sudah banyak dilakukan sebelumnya, salah satunya adalah oleh Veri Arinal dan Muhammad Azhari dengan judul “Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia” [10]. Penelitian tersebut melakukan penerapan metode regresi linier sederhana (atau Simple Linear Regression) untuk prediksi harga beras, menghasilkan nilai. Data dibagi menjadi 70% untuk data training dan 30% untuk data testing. Hasil analisis yang dilakukan dalam memprediksi harga beras menggunakan regresi linear menghasilkan nilai RMSE 337.996 +/-0.000.

Penelitian lain juga dilakukan oleh Muhaimin Hasanudin, Ifan Prihandi, Sifania Nazua dengan judul “a Comparative Study of Iconnet Jabodetabek and Banten Using Linear Regression and Support Vector Regression” [11]. Penelitian tersebut bertujuan untuk memprediksi pertumbuhan pelanggan di masa depan berdasarkan data kuantitatif dari tahun sebelumnya. Penelitian tersebut menggunakan algoritma regresi linear dan Support Vector Regression (SVR). Hasil penelitian menunjukkan bahwa algoritma SVR unggul dalam hal Mean Absolute Percentage Error (MAPE) dengan nilai MAPE sebesar 0.02%, MAE 0.10, dan RMSE 0.99. Sementara untuk regresi linear, nilai MAPE sebesar 36.28%, MAE 201, dan RMSE 0.80.

Berdasarkan penelitian yang telah diuraikan, peneliti memilih untuk membandingkan antara Simple Linear Regression dan Support Vector Regression dalam melakukan prediksi pertumbuhan jumlah penduduk Sulawesi Tenggara. Tujuan utama penelitian ini adalah memberikan kontribusi kepada pemerintah dalam perencanaan masa depan kota dengan menghindari konsekuensi negatif dan membuat estimasi dalam pemanfaatan sumber daya dan pelayanan.

II. METODE PENELITIAN

Metode penelitian yang diterapkan dalam studi ini mengikuti sebuah alur kerja yang terencana dengan cermat, sebagaimana diilustrasikan dalam skema pada Gambar 1.



Gambar 1. Skema Metode Penelitian

2.1 Studi Literatur

Studi literatur adalah langkah awal dalam penelitian ini. Proses ini melibatkan pengumpulan dan evaluasi literatur dari berbagai sumber untuk membentuk landasan teoritis. Identifikasi topik dan analisis

metodologi penelitian sebelumnya dilakukan di tahap ini.

2.2 Pengumpulan Data

Langkah berikutnya melibatkan proses pengambilan data jumlah penduduk Sulawesi Tenggara dari tahun 2014 hingga 2023, diperoleh dari Badan Pusat Statistik (BPS). Data ini kemudian diuraikan berdasarkan kabupaten/kota untuk analisis yang lebih terperinci.

2.3 Preprocessing Data

Preprocessing data menjadi fase penting yang melibatkan tindakan merapikan index data, mengecek apakah ada nilai kosong, dan normalisasi data menggunakan MinMaxScaler dan Standard Scaler.

Kedua metode tersebut adalah dua teknik populer yang digunakan untuk normalisasi data dalam machine learning. MinMaxScaler menskalakan data dalam rentang tertentu, biasanya 0 hingga 1, sedangkan Standard Scaler menskalakan data sehingga memiliki rata-rata 0 dan standar deviasi 1 [12].

MinMaxScaler menskalakan data ke dalam rentang tertentu. Rumusnya adalah:

$$X_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

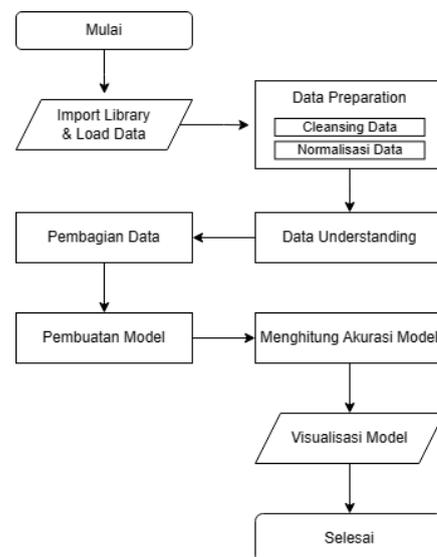
Standard Scaler menskalakan data sehingga memiliki rata-rata 0 dan standar deviasi 1. Rumusnya adalah:

$$X_{scaled} = \frac{x - \mu}{\sigma} \quad (2)$$

Di mana x adalah nilai asli, x_{min} dan x_{max} adalah nilai minimum dan maksimum dalam data, μ adalah rata-rata data, dan σ adalah deviasi standar. Dengan menggunakan rumus-rumus ini, proses preprocessing data dapat dijalankan dengan efektif untuk memastikan data yang siap digunakan dalam analisis kuantitatif.

2.4 Perancangan Model

Adapun rancangan dari model prediksi yang akan dibangun adalah pada Gambar 2.



Gambar 2. Skema Perancangan Model

2.5 Membangun Model

Setelah model terancang, proses pembangunan model dilakukan sesuai dengan skema yang telah dibuat sebelumnya. Langkah ini mencakup implementasi konsep dan struktur yang telah dirancang.

2.6 Evaluasi Model

Tahap evaluasi model menjadi kritis dalam menilai kinerja dari model-model tersebut. Metrik kuantitatif seperti Mean Absolute Percentage Error (MAPE) dan Root Mean Squared Error (RMSE).

Mean Absolute Percentage Error (MAPE) dan Root Mean Squared Error (RMSE) adalah metrik kuantitatif yang digunakan untuk mengukur akurasi prediksi pada model machine learning, terutama dalam masalah peramalan dan regresi.

MAPE mengukur kinerja model prediksi dengan menghitung rata-rata selisih persentase absolut antara nilai aktual dan nilai prediksi. Rumus untuk MAPE adalah:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (3)$$

Di mana n adalah jumlah observasi, y_i adalah nilai aktual, dan \hat{y}_i adalah nilai prediksi untuk observasi ke- i . Nilai MAPE yang lebih rendah menunjukkan prediksi yang lebih akurat, dengan nilai 0% mewakili prediksi yang sempurna [13].

RMSE mengukur kinerja model prediksi dengan menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. Rumus untuk RMSE adalah:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Di mana n adalah jumlah observasi, y_i adalah nilai aktual, dan \hat{y}_i adalah nilai prediksi untuk observasi ke- i . Nilai RMSE yang lebih rendah menunjukkan prediksi yang lebih akurat, dengan nilai 0 mewakili prediksi yang sempurna [14].

2.7 Analisis Hasil

Hasil dari evaluasi model dianalisis secara mendalam untuk membandingkan performa antara model Support Vector Regression dan Simple Linear Regression. Analisis ini memberikan wawasan yang lebih baik tentang keefektifan masing-masing model dalam meramalkan pertumbuhan penduduk di Sulawesi Tenggara.

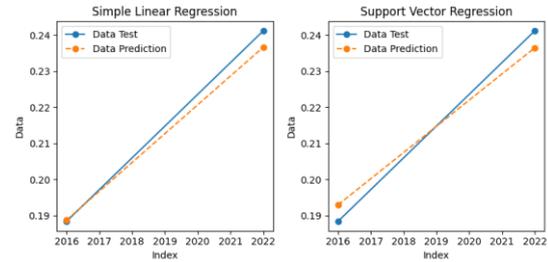
2.8 Kesimpulan

Dari hasil analisis, kesimpulan ditarik dan rekomendasi diberikan untuk penerapan model prediktif ini dalam konteks perencanaan pembangunan dan pengelolaan sumber daya di wilayah tersebut.

III. HASIL DAN PEMBAHASAN

3.1 Perbandingan Data Uji dan Data Prediksi

Setelah melakukan prediksi menggunakan kedua metode, model kini memiliki data prediksi. Berikut adalah grafik perbandingan antara data uji dan data hasil prediksi dari salah satu model, yaitu Kabupaten Buton.



Gambar 3. Grafik Perbandingan Data Uji dan Data Hasil Prediksi Model Kabupaten Buton

3.2 Algoritma Simple Linear Regression

HASIL EVALUASI KERJA SIMPLE LINEAR REGRESSION		
Kabupaten	MAPE	RMSE
Buton	0,01031	0,00320
Muna	0,00640	0,00339
Konawe	0,00400	0,00247
Kolaka	0,01533	0,01004
Konawe Selatan	0,00341	0,00268
Bombana	0,03050	0,01239
Wakatobi	0,02020	0,00542
Kolaka Utara	0,01175	0,00391
Buton Utara	0,00261	0,00024
Konawe Utara	0,00771	0,00089
Kolaka Timur	0,01767	0,00496
Konawe Kepulauan	0,10642	0,00068
Muna Barat	0,00582	0,00097
Buton Tengah	0,02440	0,00597
Buton Selatan	0,02206	0,00421
Kota Kendari	0,01934	0,02022
Kota Baubau	0,01272	0,00583

Gambar 4. Hasil Evaluasi Kerja Simple Linear Regression

Evaluasi kinerja Simple Linear Regression untuk pertumbuhan penduduk di berbagai kabupaten di Sulawesi Tenggara memberikan gambaran akurasi prediksi dengan menggunakan metrik MAPE (Mean Absolute Percentage Error) dan RMSE (Root Mean Squared Error). Sebagai contoh, kabupaten Buton menunjukkan MAPE sebesar 0,01031 dan RMSE sebesar 0,00320. Ini mengindikasikan bahwa model prediktif memiliki tingkat akurasi yang relatif tinggi dalam meramalkan pertumbuhan penduduk di kabupaten tersebut.

3.3 Algoritma Support Vector Regression (SVR)

Dalam mengevaluasi kinerja Support Vector Regression (SVR) untuk meramalkan pertumbuhan penduduk di Sulawesi Tenggara, digunakan metrik evaluasi MAPE (Mean Absolute Percentage Error) dan RMSE (Root Mean Squared Error). Kabupaten Buton Utara dan Konawe Utara menunjukkan tingkat akurasi yang tinggi, dengan MAPE masing-masing sebesar 0,00146 dan 0,01265, serta RMSE yang relatif rendah masing-masing sebesar 0,00014 dan 0,00132.

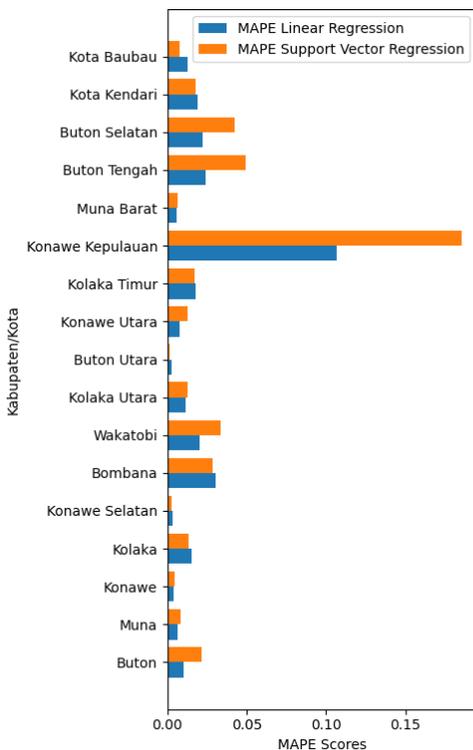
HASIL EVALUASI KERJA SUPPORT VECTOR REGRESSION		
Kabupaten	MAPE	RMSE
Buton	0,02179	0,00462
Muna	0,00823	0,00431
Konawe	0,00455	0,00279
Kolaka	0,01357	0,00875
Konawe Selatan	0,00248	0,00226
Bombana	0,02873	0,01122
Wakatobi	0,03346	0,00679
Kolaka Utara	0,01267	0,00413
Buton Utara	0,00146	0,00014
Konawe Utara	0,01265	0,00132
Kolaka Timur	0,01698	0,00462
Konawe Kepulauan	0,18549	0,00110
Muna Barat	0,00646	0,00137
Buton Tengah	0,04925	0,00939
Buton Selatan	0,04244	0,00638
Kota Kendari	0,01754	0,01750
Kota Baubau	0,00803	0,00315

Gambar 5. Hasil Evaluasi Kerja Support Vector Regression

3.4 Perbandingan Hasil MAPE

Perbandingan nilai Mean Absolute Percentage Error (MAPE) antara model Simple Linear Regression (SLR) dan Support Vector Regression (SVR) menunjukkan perbedaan performa dalam meramalkan pertumbuhan penduduk di Sulawesi Tenggara.

Dalam konteks hasil evaluasi, SLR secara konsisten menunjukkan nilai MAPE yang lebih rendah dibandingkan dengan SVR, menandakan tingkat akurasi prediksi yang lebih baik. Perbandingannya dapat dilihat pada grafik batang di Gambar 5.

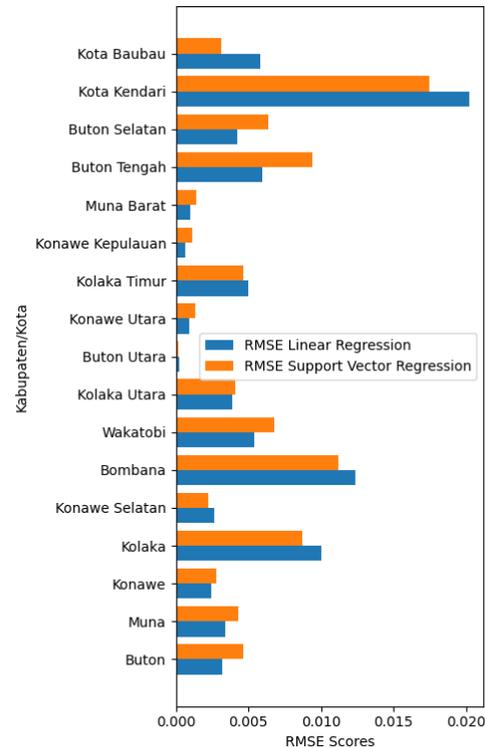


Gambar 6. Perbandingan Skor MAPE antara SLR dan SVR berdasarkan Model

3.5 Perbandingan Hasil RMSE

Perbandingan nilai Root Mean Squared Error (RMSE) antara model Simple Linear Regression (SLR) dan Support Vector Regression (SVR) dapat dilihat dalam Gambar 6. Dari hasil evaluasi tersebut, terlihat bahwa

SLR cenderung menunjukkan nilai RMSE yang lebih rendah dibandingkan dengan SVR pada sebagian besar kabupaten. Hal ini memberikan indikasi visual tentang keunggulan SLR dalam memberikan prediksi yang lebih akurat.



Gambar 7. Perbandingan Skor RMSE antara SLR dan SVR berdasarkan Model

3.6 Perbandingan Rata-rata Evaluasi Metode

Dengan membandingkan rata-rata MAPE dan RMSE dari setiap metode, dapat diamati seperti yang tergambar pada Gambar 7. Semakin kecil persentase dari nilai rata-rata, semakin rendah tingkat kesalahan pada setiap model dari metode tersebut.

MEAN HASIL KINERJA MODEL		
	Mean MAPE	Mean RMSE
SVR	2.74%	0.53%
SLR	1.89%	0.51%

Gambar 8. Rata-rata Evaluasi Model

IV. KESIMPULAN

Berdasarkan hasil pengujian, dapat disimpulkan bahwa metode Simple Linear Regression (SLR) lebih sesuai untuk digunakan dalam memprediksi jumlah penduduk Sulawesi Tenggara. SLR menghasilkan nilai rata-rata MAPE sebesar 1.89% dan RMSE sebesar 0.51%. Lebih lanjut, pada sebagian besar model Kabupaten/Kota, SLR menunjukkan performa yang lebih baik dengan MAPE dan RMSE yang lebih rendah dibandingkan dengan Support Vector Regression (SVR).

Meskipun demikian, perlu diperhatikan bahwa terdapat beberapa Kabupaten/Kota di mana SVR dapat bersaing atau bahkan melampaui SLR dalam kinerjanya. Oleh karena itu, disarankan untuk melakukan studi lebih lanjut terkait metode prediksi lainnya.

V. REFERENSI

- [1] L. Hurriah, "Analisis Fenomena Sosial Mudik Juga Menjadi Faktor Kepadatan Penduduk," in *Prosiding Seminar Sastra Budaya dan Bahasa (SEBAYA)*, 2023, pp. 139–147.
- [2] Y. G. Winarti, "Analisis Hubungan Antara Pertumbuhan Penduduk, Konsumsi Makanan, dan Tingkat Partisipasi Angkatan Kerja Perempuan Terhadap Pertumbuhan Ekonomi di Kota Magelang," *Jurnal Jendela Inovasi Daerah*, vol. 3, no. 2, pp. 1–16, 2020.
- [3] C. M. Annur, "10 Negara dengan Jumlah Penduduk Terbanyak di Dunia Pertengahan 2023," Databoks. Accessed: Jan. 09, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/07/28/10-negara-dengan-jumlah-penduduk-terbanyak-di-dunia-pertengahan-2023>
- [4] Kementerian Dalam Negeri Republik Indonesia, "Data Kependudukan Berdasarkan Provinsi 2023," Satu Data. Accessed: Jan. 09, 2024. [Online]. Available: <https://e-database.kemendagri.go.id/kemendagri/dataset/1102/tab-el-data>
- [5] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.
- [6] H. D. E. Sinaga and N. Irawati, "Perbandingan double moving average dengan double exponential smoothing pada peramalan bahan medis habis pakai," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 4, no. 2, pp. 197–204, 2018, doi: <https://doi.org/10.33330/jurteksi.v4i2.60>.
- [7] G. A. Shewa and F. I. Ugwuowo, "A new hybrid estimator for linear regression model analysis: computations and simulations," *Sci Afr*, vol. 19, p. e01441, 2023.
- [8] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos Solitons Fractals*, vol. 138, p. 109942, 2020.
- [9] D. S. Seruni, M. T. Furqon, and R. C. Wihandika, "Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 4, pp. 1075–1082, 2020.
- [10] V. Arinal and M. Azhari, "Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 341–346, 2023.
- [11] M. Hasanudin, I. Prihandi, and S. Nazua, "a Comparative Study of Iconnet Jabodetabek and Banten Using Linear Regression and Support Vector Regression," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 1, pp. 119–127, 2024.
- [12] V. R. Prasetyo, M. Mercifia, A. Averina, L. Sunyoto, and B. Budiarto, "Prediksi Rating Film Pada Website IMDB Menggunakan Metode Neural Network," *NERO (Networking Engineering Research Operation)*, vol. 7, no. 1, pp. 1–8, 2022.
- [13] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int J Forecast*, vol. 32, no. 3, pp. 669–679, 2016, doi: <https://doi.org/10.1016/j.ijforecast.2015.12.003>.
- [14] P. Singla, M. Duhan, and S. Saroha, "10 - Different normalization techniques as data preprocessing for one step ahead forecasting of solar global horizontal irradiance," in *Artificial Intelligence for Renewable Energy Systems*, A. K. Dubey, S. K. Narang, A. L. Srivastav, A. Kumar, and V. García-Díaz, Eds., in Woodhead Publishing Series in Energy. , Woodhead Publishing, 2022, pp. 209–230. doi: <https://doi.org/10.1016/B978-0-323-90396-7.00004-3>.