

Perbandingan Akurasi Linear Regression dan Support Vector Regression dalam Prediksi Suhu Rata-Rata

Gideon Namlea Lesnusa¹, Dwi Shinta Angreni², Rizka Ardiansyah³

dionnamlea@gmail.com¹, ds.angreni@untad.ac.id², rizka@untad.ac.id³

^{1,2,3} Universitas Tadulako, Sulawesi Tengah

Informasi Artikel

Diterima : 24 Apr 2024

Direvisi : 13 Mei 2024

Disetujui : 25 Jul 2024

Kata Kunci

Prediksi cuaca, python, linear regression, support vector regression, time series

Abstrak

Cuaca di Indonesia sangat bervariasi dan dipengaruhi oleh letak geografis, topografi, dan iklim regional. Pola cuaca berbeda antara bagian barat dan timur Indonesia. Penelitian ini mengeksplorasi model time series untuk memprediksi data cuaca di Kota Palu, wilayah yang kompleks karena dipengaruhi oleh berbagai faktor cuaca. Fokusnya adalah pada pola cuaca unik yang tercermin dari geografi dan topografi Kota Palu. Evaluasi dilakukan terhadap model time series, termasuk Linear Regression dan *Support Vector Regression (SVR)*, untuk memperkirakan kondisi cuaca di Kota Palu. Hasil evaluasi menunjukkan bahwa model SVR memiliki *RMSE* sebesar 0.6302, sedangkan linear regression memiliki *RMSE* sebesar 0.6328. Penelitian ini berpotensi meningkatkan peringatan dini dan pengambilan keputusan terkait cuaca ekstrem.

Keywords

Forecast weather, python, linear regression, support vector regression, time series

Abstract

The weather in Indonesia varies significantly and is influenced by geographical location, topography, and regional climate. Weather patterns differ between the western and eastern parts of Indonesia. This study explores time series models to predict weather data in Palu City, a region that is complex due to various weather factors. The focus is on the unique weather patterns reflected by the geography and topography of Palu City. Evaluation was conducted on time series models, including Linear Regression and Support Vector Regression (SVR), to estimate weather conditions in Palu City. The evaluation results show that the SVR model has an RMSE of 0.6302, while linear regression has an RMSE of 0.6328. This research has the potential to improve early warning and decision-making regarding extreme weather.

A. Pendahuluan

Cuaca merupakan faktor penting yang memengaruhi berbagai aktivitas manusia dan kehidupan sehari-hari. Memprediksi cuaca merupakan tugas yang menantang karena melibatkan banyak variabel yang memengaruhi, seperti tekanan atmosfer, kecepatan angin, curah hujan, suhu, dan fenomena atmosfer lainnya. Prediksi sering kali dilakukan berdasarkan data yang tersedia dalam periode waktu yang singkat[1].

Secara tidak disadari, suhu udara semakin terasa panas seiring berjalannya waktu. Pertumbuhan populasi manusia menyebabkan peningkatan konsumsi sumber daya alam dibandingkan sebelumnya. Fenomena ini mendorong perkembangan teknologi secara cepat, namun juga mengakibatkan penurunan kualitas tanah, air, dan udara[2]. Suhu tertinggi di permukaan bumi terjadi di daerah tropis sekitar khatulistiwa, dan semakin mendekati kutub, suhu menjadi lebih dingin. Saat mendaki gunung, suhu udara terasa semakin dingin seiring dengan kenaikan ketinggian[3].

Dalam melakukan prediksi cuaca, berbagai faktor perlu dipertimbangkan, seperti suhu atmosfer, kadar kelembaban udara, cahaya matahari yang terpancar, arah serta kecepatan angin, curah hujan, dan variabel atmosfer lainnya. Dalam proses prediksi cuaca, terdapat berbagai kondisi yang diamati untuk menyusun ramalan yang akurat[4].

Penelitian ini menganalisis data dari BMKG Stasiun Meteorologi Mutiara SIS Aljufri dengan menggunakan teknik data mining. Hal ini memungkinkan untuk memanfaatkan potensi besar dari data cuaca yang tersedia supaya memperbaiki akurasi dan ketepatan ramalan cuaca di Kota Palu. Hal ini dapat membantu masyarakat dan pemerintah setempat dalam mengambil tindakan yang lebih proaktif terhadap cuaca ekstrem, seperti banjir dan tanah longsor.

Kompleksitas prediksi cuaca di Stasiun Meteorologi SIS Al Jufri di Kota Palu yang semakin meningkat akibat volume data cuaca yang melimpah. Tujuan dari penelitian ini adalah menerapkan model time series seperti regresi linear telah terbukti efektif dalam melakukan prediksi angka dengan tingkat akurasi yang cukup baik[5]. Dan juga Support Vector Regression untuk memperbaiki ramalan cuaca di Kota Palu. Dengan membandingkan kinerja kedua model ini, diharapkan menghasilkan pemahaman yang lebih dalam tentang fenomena iklim dan perubahan cuaca secara menyeluruh. Diharapkan dengan memperoleh wawasan yang lebih komprehensif tentang pola cuaca, dapat meningkatkan ketepatan dalam meramalkan kondisi atmosfer di daerah tersebut.

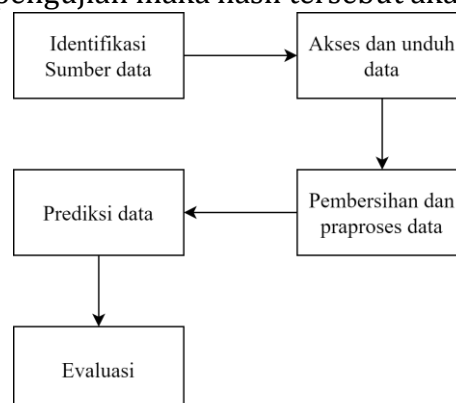
Penelitian ini merujuk dari penelitian sebelumnya yang membahas support vector machine dan juga linear regression dalam memprediksi cuaca dengan parameter suhu rata-rata. Tujuan penelitian ini adalah untuk menciptakan suatu pendekatan baru yang nantinya digunakan untuk menentukan prediksi cuaca dengan salah satu parameternya adalah suhu rata-rata dengan hasil yang baik dan tentunya akurat[6]. Langkah penelitian berikutnya akan melibatkan evaluasi prediksi kondisi atmosfer dengan mempertimbangkan sejumlah parameter, termasuk suhu, tekanan udara, dan kecepatan angin, menggunakan metode analisis regresi linear. Penelitian ini memiliki tujuan dengan memprakirakan cuaca yang akan terjadi maka akan meminimalisir kemungkinan buruk yang bisa saja terjadi kedepannya[7]. Penelitian berikut adalah perbandingan support vector

untuk prediksi cuaca. Penelitian ini bertujuan untuk membandingkan metode imputasi dan support vector regression, dari kedua metode tersebut nantinya akan ditentukan mana yang lebih untuk meramal curah hujan[8].

B. Metode Penelitian

1. Metode Penelitian

Dalam konteks penelitian ini, teknik yang diterapkan adalah linear regression dan support vector regression. Model regresi linear umumnya membuat dua variabel yaitu variabel dependen dan juga independen, variabel dependen sendiri merupakan variabel yang akan diprediksi atau dikenal juga sebagai variabel target. Sementara itu, variabel independen merupakan variabel yang berfungsi untuk memprediksikan variabel dependen, variabel independen juga dikenal sebagai variabel prediktor[9]. Sedangkan support vector regression adalah metode yang mencoba membuat model untuk memprediksi nilai target dari data. Cara kerjanya adalah dengan mengubah data ke dalam bentuk ruang vektor yang lebih kompleks, kemudian mencari pola hubungan antara input dan output menggunakan garis lurus[10]. Langkah awal penelitian ini yaitu identifikasi sumber data dimana langkah ini dilakukan untuk menentukan sumber data yang akan digunakan dan juga menentukan berapa banyak data yang akan digunakan berdasarkan periode waktu dari data tersebut. Tahap berikut adalah akses dan unduh data dimana pada tahap ini akan dilakukan akses dan unduh pada data cuaca dari website BMKG, pada tahap akses data ini perlu untuk mendaftar di website BMKG untuk bisa mengunduh data yang dibutuhkan. Tahap selanjutnya adalah pembersihan dan praproses data, karena data yang telah di unduh merupakan data perbulan, oleh karena itu data-data tersebut harus disatukan dalam sebuah file csv, pada tahap ini juga dilakukan pembersihan data yang tidak lengkap. Selanjutnya adalah tahap pemisahan data dimana pada tahap ini data dipisahkan menjadi dua yaitu data untuk menjadi set pelatihan dan set pengujian. Informasi dari data latih akan dimanfaatkan untuk melatih model, sementara data uji akan digunakan untuk mengevaluasi kinerja model tersebut. Tahap selanjutnya adalah prediksi data, pada proses ini data yang telah dikumpulkan akan di uji menggunakan linear regression dan juga support vector regression, selanjutnya hasil dari kedua metode tersebut akan dibandingkan untuk mencari hasil yang lebih baik antara kedua metode tersebut. Tahap terakhir yaitu evaluasi model, dimana setelah mendapatkan hasil dari pengujian maka hasil tersebut akan di uji.



Gambar 1. Alur metode penelitian

2. Dataset

Data mengenai kondisi cuaca diperoleh dari sumber yang terdokumentasi di situs web Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data cuaca ini adalah data timeseries cuaca Kota Palu yang dipantau langsung oleh Stasiun Meteorologi Kelas II Mutiara Sis Al-Jufri Palu. Data yang menjadi subjek penelitian ini mencakup rentang waktu dari tahun 2020 hingga tahun 2023. Analisis deret waktu melibatkan studi pola perubahan nilai-nilai variabel dalam interval waktu yang ditentukan, seperti perhari, perminggu, perbulan bahkan pertahun. Melalui analisis deret waktu, data mengenai ukuran nilai-nilai ini dapat diambil untuk membantu dalam melakukan prediksi masa depan[11].

Tabel 1. Dataset cuaca

No	Tanggal	<i>temp_min</i>	<i>temp_max</i>	<i>temp_rata_rata</i>
1	01-01-2020	23.4	33	27.3
2	02-01-2020	23.2	33.2	27.2
3	03-01-2020	22.4	32.6	26.5
4	04-01-2020	22.8	33.2	27.3
--	-----	-----	-----	-----
1312	30-12-2023	25.4	35.2	29.6
1313	31-12-2023	25.8	34.8	28.5

3. Pembersihan dan praproses data

Data yang telah dikumpulkan sebelumnya merupakan data cuaca perbulan kota Palu. Agar data-data tersebut bisa digunakan untuk memprediksi maka langkah selanjutnya adalah menggabungkan semua data tersebut menjadi satu kedalam sebuah file csv, lalu data-data akan diperiksa terlebih dahulu. Tujuan dari proses ini adalah mengeliminasi data-data yang kosong agar tidak mengganggu output dari proses prediksi.

Tabel 2. Data yang telah di bersihkan

No	Tanggal	<i>temp_min</i>	<i>temp_max</i>	<i>temp_rata_rata</i>
1	01-01-2020	23.4	33	27.3
2	02-01-2020	23.2	33.2	27.2
3	03-01-2020	22.4	32.6	26.5
4	04-01-2020	22.8	33.2	27.3
5	05-01-2020	23	33.2	26.7
6	06-01-2020	23.4	32.4	27
7	07-01-2020	23.6	33.2	27.9
8	08-01-2020	23.6	34.3	28
9	09-01-2020	23.3	33.4	27.7
10	10-01-2020	23.2	33.8	28

4. Prediksi data

Penelitian ini memanfaatkan metode regresi linear dan regresi vektor pendukung sebagai algoritma yang digunakan., dimana diakhir kedua algoritma ini akan dibandingkan. Prediksi ini akan dibuat pada platform Google Collab menggunakan bahasa pemrograman python. Variabel yang akan digunakan untuk memprediksi cuaca adalah *temp_max* dan *temp_rata_rata*, variabel ditentukan dengan terlebih dahulu membuat tabel korelasi dengan tujuan mengilustrasikan seberapa besar dampak variabel independen terhadap hasil prediksi[12]. Setelah data diimpor kedalam google collab langkah selanjutnya adalah melakukan prediksi untuk mendapat prediksi data cuaca dengan cara memisahkan data

pengujian dan data training. Setelah data di proses data tersebut akan ditampilkan nilai actual dan predicted menggunakan algoritma linear regression dan support vector regression.

5. Evaluasi

Pada evaluasi kedua model yaitu linear regression dan support vector regression akan dibandingkan, model mana di antara kedua model yang menunjukkan akurasi pengujian yang lebih baik. Ketepatan suatu metode dalam mengestimasi kesalahan pengukuran dapat diperhatikan dari nilai RMSE yang rendah[13]. Kinerja yang baik dari sebuah algoritma tercermin dari rendahnya nilai RMSE. Sebaliknya, kinerja metode estimasi dianggap buruk jika RMSE-nya tinggi. Perhitungan RMSE dilakukan dengan mengurangi nilai aktual dari nilai prediksi yang bersifat kontinu, hasilnya kemudian dipangkatkan dua, dijumlahkan, dan diambil rata-ratanya[14].

Tabel 3. Tingkat akurasi

Metode	MAE	MSE	RMSE	Coefficient of determination
Linear regression	0.4626	0.4005	0.6328	0.66
Support vector regression	0.4599	0.3972	0.6302	0.66

C. Hasil dan Pembahasan

Hasil dari penelitian yang sudah dilakukan menunjukkan bahwa model linear regression dan juga support vector regression bisa memprediksi cuaca dengan baik setelah mendapatkan nilai MAE, MSE, RME dibawah 0. Namun dari antara kedua model tersebut yang paling akurat adalah model SVR karena mendapatkan nilai yang lebih rendah dimana itu menunjukan bahwa tingkat akurasi dari prediksinya terbilang baik. Hasil menunjukkan bahwa model SVR memiliki kemampuan yang lebih baik dalam menangani relasi non-linear antara variabel prediktor dan variabel target, yang penting dalam memprediksi fenomena cuaca yang kompleks. Distribusi kesalahan prediksi juga menunjukkan pola yang berbeda antara kedua model, dengan SVR cenderung menghasilkan kesalahan prediksi yang lebih kecil. Faktor-faktor seperti suhu udara, kelembaban, dan tekanan udara telah dipertimbangkan dalam pengembangan model. Model SVR lebih unggul dalam menangani faktor-faktor ini secara efektif, yang mengarah pada prediksi cuaca yang lebih akurat. Validasi model dilakukan menggunakan metode cross-validation untuk memastikan keandalan prediksi model terhadap data baru. Hasil validasi menunjukkan bahwa kedua model tetap konsisten dalam kinerjanya di berbagai kondisi cuaca dan mampu menghasilkan prediksi yang stabil. Perbandingan dengan penelitian terdahulu menunjukkan bahwa temuan penulis mendukung dan memperluas pengetahuan sebelumnya tentang prediksi cuaca. Model SVR telah terbukti lebih akurat daripada model-model sebelumnya dalam konteks ini. Hasil prediksi dari kedua model tersebut dicantumkan pada tabel dibawah.

Tabel 4. Hasil akhir dari perbandingan linear regresion dan SVR

No	Actual	Predicted(linear regression)	Predicted(SVR)
1	27.5	27.346	27.292
2	24.8	25.142	25.472
3	27.8	27.747	27.722

4	27.2	27.446	27.397
5	26.1	25.844	25.944
6	26.2	25.042	25.413
7	27.8	27.847	27.834
8	25.8	26.645	26.608
9	26.4	26.344	26.343
10	27.4	27.547	27.503
11	27.6	27.547	27.503
12	26.2	27.046	26.987
13	27.6	27.046	26.987
14	27.4	26.845	26.794
15	27.0	27.046	26.987

D. Simpulan

Dari evaluasi yang dilakukan terhadap prediksi cuaca menggunakan metode Linear Regression dan Support Vector Regression (SVR) untuk data tahun 2020 hingga 2023 didapati bahwa model SVR menunjukkan RMSE bernilai 0.6302, mencerminkan tingkat kesalahan prediksi model yang relatif rendah, sedangkan model Linear Regression memiliki RMSE sebesar 0.6328. Meskipun perbedaannya tipis, hasil ini menunjukkan bahwa model SVR sedikit lebih baik dalam memprediksi kondisi cuaca di Kota Palu selama periode tersebut. Oleh karena itu, dalam konteks ini, model SVR dapat dianggap sebagai pilihan yang lebih optimal untuk memprediksi cuaca di wilayah tersebut, yang memiliki kompleksitas yang dipengaruhi oleh faktor-faktor geografis dan topografi yang beragam. Dalam penelitian ini, penulis menyajikan tinjauan komprehensif mengenai perbandingan antara regresi linear dan Support Vector Regression (SVR) dengan pendekatan yang inovatif, memperkenalkan metode evaluasi baru yang menggabungkan analisis kinerja tradisional dengan prediksi rata-rata yang lebih relevan dengan kebutuhan praktis. Melalui analisis mendalam terhadap faktor-faktor pengaruh dan kasus penggunaan yang spesifik, penelitian ini tidak hanya memberikan wawasan baru tentang kekuatan dan kelemahan masing-masing metode, tetapi juga menawarkan panduan yang lebih cerdas dalam memilih model regresi yang paling sesuai dengan konteks aplikasi yang dihadapi.

E. Ucapan Terima Kasih

Terima kasih kepada Tuhan Yang Maha Kuasa atas segala kebaikan dan kasih penyertaan-Nya dan juga kepada rekan-rekan Universitas Tadulako dan orang-orang yang telah memberikan dukungan dan kontribusi kepada penulis dalam menyelesaikan penelitian ini.

F. Referensi

- [1] A. M. Siregar, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning," *Petir*, vol. 13, no. 2, pp. 138–147, 2020, doi: 10.33322/petir.v13i2.998.
- [2] A. Tresnasena, F. Andika, and Respitawulan, "Prediksi Perubahan Suhu Rata-Rata Permukaan Global Tahun 2022-2032 Menggunakan Maple," *J. Mat.*, vol. 21, no. 1, pp. 35–41, 2022, [Online]. Available: <https://journals.unisba.ac.id/index.php/matematika>
- [3] M. Rusdi, "Komparasi Penggunaan Algoritma Support Vector Machine

- Dengan Particle Swarm Optimization Dalam Memprediksi Suhu Udara,” *Technol. J. Ilm.*, vol. 8, no. 4, p. 277, 2017, doi: 10.31602/tji.v8i4.1128.
- [4] H. Jayadianti, T. A. Cahyadi, N. A. Amri, and M. F. Pitayandanu, “Metode Komparasi Artificial Neural Network Pada Prediksi Curah Hujan - Literature Review,” *J. Tekno Insentif*, vol. 14, no. 2, pp. 48–53, 2020, doi: 10.36787/jti.v14i2.150.
- [5] F. Sodik, A. Sanusi Mashuri, and S. Huda, “Implementasi Algoritma Convolutional Neural Network Dan Linear Regresi Untuk Memprediksi Kebakaran Hutan Implementation of Convolutional Neural Network and Linear Regression Algorithm to Predict Forest Fires,” *Cogito Smart J. I*, vol. 9, no. 1, pp. 28–37, 2023.
- [6] M. N. Rifqi and R. T. Aldisa, “Penerapan Metode Support Vector Machine Dalam Memprediksi Prediksi Cuaca,” *J. Comput. Syst. Informatics*, vol. 5, no. 2, pp. 368–379, Feb. 2024, doi: 10.47065/josyc.v5i2.4961.
- [7] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, “Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 10–17, 2020, doi: 10.33633/joins.v5i1.2760.
- [8] A. M. Priyatno, A. Wiratmo, F. Syuhada, and P. Cholidhazia, “Perbandingan Imputasi Dan Parameter Support Vector Regression untuk Peramalan Cuaca,” *J. SIMETRIS*, vol. 10, no. 2, pp. 651–660, 2019, doi: 10.24176/simet.v10i2.3402.
- [9] M. Ryan and S. Alfiandy, “Prediksi Kecepatan Angin 12 Jam Kedepan Menggunakan Automatic Weather Observing System (AWOS) Berbasis Regresi Linear,” *Bul. GAW Bariri*, vol. 3, no. 2, pp. 1–7, 2022, doi: 10.31172/bgb.v3i2.63.
- [10] Putri R.A, Winahju W.S, and Mashuri Muhammad, “Penerapan Metode Ridge Regression dan Support Vector Regression (SVR) untuk Prediksi Indeks Batubara di PT XYZ,” *J. Sains Dan Seni Its*, vol. 9, no. 1, pp. 64–71, 2020.
- [11] C. A. Soekendro, “PREDIKSI CURAH HUJAN DI KAB.BANDUNG DENGAN ANALISIS TIME SERIES, MENGGUNAKAN MODEL SARIMA (Seasonal Autoregressive Integrated Moving Average),” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 2865–2875, 2021.
- [12] M. A. Musababa, “Implementasi Algoritma Linear Regression untuk Prediksi Produksi Tanaman Padi di Kabupaten Grobogan,” *Data Sci. Indones.*, vol. 3, no. 2, pp. 68–78, 2023.
- [13] E. Dewi, S. Mulyani, I. Septianingrum, N. Nurjanah, and R. Rahmawati, “Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi,” *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 8–1, no. 1, pp. 67–77, 2019, doi: 10.36774/jusiti.v8i1.602.
- [14] A. B. Raharjo, Z. Z. Dinanto, D. Sunaryono, and D. Purwitasari, “Prediksi Akumulasi Kasus Terkonfirmasi Covid-19 Di Indonesia Menggunakan Support Vector Regression,” *Techno.Com*, vol. 20, no. 3, pp. 372–381, 2021, doi: 10.33633/tc.v20i3.5062.