

ABSTRAK

PERBANDINGAN ALGORITMA *RANDOM FOREST REGRESSION* DAN *SUPPORT VECTOR REGRESSION* (SVR) UNTUK MESIN PERAMALAN

CURAH HUJAN

Oleh:

Rizka M. Imron
1217050126

Peramalan deret waktu (*time series*) menggunakan *machine learning* menghadapi tantangan dalam mengadaptasi algoritma regresi umum yang tidak dirancang untuk menangkap pola temporal. Penelitian ini mencoba menjawab tantangan tersebut dengan membuktikan bahwa algoritma *machine learning* yang tidak memiliki pemahaman bawaan tentang urutan waktu (temporal) dalam data seperti *Support Vector Regression* (SVR) dan *Random Forest Regression* (RFR) dapat diadaptasi secara efektif melalui rekayasa fitur (*feature engineering*). Dengan menciptakan fitur *lag historis*, siklus musiman, dan statistik bulanan, model diberi konteks temporal yang diperlukan untuk peramalan. Menggunakan data curah hujan bulanan univariat dari Stasiun pengamatan BMKG Kota Bandung periode (Januari 1993 – Maret 2025) dan metodologi CRISP-DM, penelitian ini membandingkan SVR dan RFR dalam empat skenario partisi data untuk mengevaluasi tiga aspek: akurasi, stabilitas, dan efisiensi. Hasil penelitian mengungkap adanya tradeoff yang jelas. *Random Forest Regression* menunjukkan akurasi prediktif tertinggi ($R^2 = 0,9283$; RMSE = 31,28 mm) namun dengan stabilitas lebih rendah ($\Delta R^2 = 0,051$). Sebaliknya, SVR memiliki akurasi yang solid ($R^2 = 0,8042$; RMSE = 51,70 mm) namun unggul secara signifikan dalam stabilitas ($\Delta R^2 = 0,016$) dan efisiensi komputasi (4-6x lebih cepat). Kesimpulannya, rekayasa fitur adalah kunci untuk mengadaptasi algoritma non deret waktu (*time series*), di mana pemilihan model terbaik bergantung pada prioritas antara akurasi (RFR) dan efisiensi (SVR). Penelitian ini tidak hanya memberikan model peramalan yang aplikatif, yang diwujudkan dalam purwarupa Gradio, tetapi juga menyajikan kontribusi metodologis untuk pengembangan sistem peramalan yang praktis dan efisien.

Kata Kunci: *Machine Learning*, *Random Forest Regression*, *Support Vector Regression*, Rekayasa Fitur, *Time Series*, Peramalan Curah Hujan, CRISP-DM.

ABSTRACT

COMPARISON OF RANDOM FOREST REGRESSION AND SUPPORT VECTOR REGRESSION (SVR) ALGORITHMS FOR RAINFALL FORECASTING MACHINES

Oleh:

Rizka M. Imron

1217050126

Time-series forecasting using machine learning faces the challenge of adapting general regression algorithms that are not inherently designed to capture temporal patterns. This study addresses this challenge by demonstrating that non-sequential machine learning algorithms, such as Support Vector Regression (SVR) and Random Forest Regression (RFR), can be effectively adapted through feature engineering. By creating historical lags, seasonal cycles, and monthly statistical features, the models are provided with the necessary temporal context for forecasting. Utilizing univariate monthly rainfall data from the BMKG Bandung observation station (January 1993 – March 2025) and the CRISP-DM methodology, this study compares SVR and RFR across four data partition scenarios to evaluate three aspects: accuracy, stability, and efficiency. The results reveal a clear trade-off: Random Forest Regression demonstrates the highest predictive accuracy ($R^2 = 0.9283$; RMSE = 31.28 mm) but with lower stability ($\Delta R^2 = 0.051$). Conversely, SVR shows solid accuracy ($R^2 = 0.8042$; RMSE = 51.70 mm) while significantly excelling in stability ($\Delta R^2 = 0.016$) and computational efficiency (4-6x faster). In conclusion, feature engineering is the key to adapting non-time-series algorithms, where the optimal model selection depends on the priority between accuracy (RFR) and efficiency (SVR). This research not only delivers an applicable forecasting model, actualized in a Gradio prototype, but also presents a methodological contribution for developing practical and efficient forecasting systems.

Keywords: Machine Learning, Random Forest Regression, Support Vector Regression, Feature Engineering, Time Series, Rainfall Forecasting, CRISP-DM.