

DIMENSIONALITY REDUCTION WITH INDEPENDENT COMPONENT ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS IN SUPPORT VECTOR REGRESSION FOR CONCRETE COMPRESSIVE STRENGTH PREDICTION

REDUKSI DIMENSI DENGAN *INDEPENDENT COMPONENT ANALYSIS* DAN *PRINCIPAL COMPONENT ANALYSIS* PADA *SUPPORT VECTOR REGRESSION* UNTUK PREDIKSI KUAT TEKAN BETON

Chaca Alya Fahrani¹, Azzah Zhafira Shidiq^{1‡}, Atika Fawwaz¹, Nizar Hermawan¹, Mahendra Rakriandonny Setiadjie¹, and Septian Rahardiantoro¹

¹Statistics and Data Science Study Program, IPB University, Indonesia

[‡]corresponding author: azs_azzahzhafira@apps.ipb.ac.id

Copyright © 2026 Chaca Alya Fahrani, Azzah Zhafira Shidiq, Atika Fawwaz, Nizar Hermawan, Mahendra Rakriandonny Setiadjie, Septian Rahardiantoro. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Concrete compressive strength is an important parameter for assessing concrete quality, however, laboratory testing is relatively time-consuming and costly. This study applied dimensionality reduction using Principal Component Analysis (PCA) and Independent Component Analysis (ICA) prior to Support Vector Regression (SVR) modelling with a Radial Basis Function (RBF) kernel. Independent components in ICA were extracted using the FastICA algorithm. The objective of this study was to compare the performance of PCA-SVR and ICA-SVR in predicting concrete compressive strength. The results show that PCA produces six principal components based on a 95% cumulative explained variance threshold, explaining 97.31% of the total variance, while ICA yields seven optimal independent components based on negentropy and mutual information. The evaluation and statistical test results showed that the ICA-SVR model achieved an RMSE of 5.9435, an MAE of 4.1039, and an R^2 of 0.8629, indicating better predictive performance than PCA-SVR. However, the baseline SVR model without dimensionality reduction still achieved slightly better performance, with an RMSE of 5.8221 and an R^2 of 0.8685. Overall, ICA-SVR demonstrates better predictive performance than PCA-SVR on the dataset and configuration used in this study, although dimensionality reduction did not surpass the performance of the baseline SVR model.

Keywords: Compressive Strength, Dimensionality Reduction, ICA, PCA, SVR., Concrete

Abstrak

Kuat tekan beton merupakan parameter penting dalam penentuan mutu beton, namun pengujiannya di laboratorium membutuhkan waktu yang relatif lama dan biaya yang cukup besar. Penelitian ini menerapkan teknik reduksi dimensi menggunakan PCA dan ICA sebelum pemodelan SVR berbasis kernel RBF. Komponen independen pada ICA diekstraksi menggunakan algoritma FastICA. Penelitian ini bertujuan membandingkan kinerja PCA-SVR dan ICA-SVR dalam memprediksi kuat tekan beton. Hasil analisis menunjukkan bahwa PCA menghasilkan enam komponen utama berdasarkan *cumulative explained variance* sebesar 95% dengan total variansi yang dijelaskan sebesar 97.31%, sedangkan ICA menghasilkan tujuh komponen independen optimal berdasarkan evaluasi *negentropy* dan *mutual information*. Hasil evaluasi dan pengujian statistik menunjukkan bahwa model ICA-SVR menghasilkan nilai RMSE sebesar 5.9435, MAE sebesar 4.1039, dan R^2 sebesar 0.8629, yang lebih baik dibandingkan PCA-SVR. Namun, model SVR *baseline* tanpa reduksi dimensi masih memberikan performa yang sedikit lebih baik dengan nilai RMSE sebesar 5.8221 dan R^2 sebesar 0.8685. Secara keseluruhan, ICA-SVR menghasilkan performa prediksi yang lebih baik dibandingkan PCA-SVR pada dataset dan konfigurasi yang digunakan, meskipun reduksi dimensi belum mampu melampaui performa model SVR *baseline*.

Kata Kunci: ICA, Kuat Tekan Beton, PCA, Reduksi Dimensi, SVR.

1. Pendahuluan

Prediksi kuat tekan beton menjadi pendekatan yang relevan dalam pemodelan material konstruksi karena kuat tekan merupakan indikator utama mutu beton, sementara pengujiannya di laboratorium membutuhkan waktu, biaya, dan prosedur eksperimen tertentu. Prayogo (2018) menyatakan bahwa pengujian kuat tekan beton memiliki keterbatasan biaya dan waktu sehingga metode kecerdasan buatan diperlukan untuk memetakan hubungan *input-output* yang kompleks pada campuran beton. Selain itu, Christiono dan Prayogo (2020) menekankan pentingnya prediksi kuat tekan beton tanpa harus menunggu umur rencana 28 hari. Penelitian oleh Paudel *et al.* (2023) menunjukkan bahwa umur beton, semen, dan air merupakan peubah yang paling berpengaruh terhadap kuat tekan beton, sedangkan Wan *et al.* (2021) menunjukkan bahwa komposisi campuran beton dan *curing age* dapat digunakan untuk memprediksi kuat tekan beton menggunakan model *machine learning*.

Hubungan antara peubah penyusun beton dan kuat tekan tidak selalu bersifat linear sehingga diperlukan metode yang mampu menangkap pola nonlinear serta interaksi antarpeubah. *Support Vector Regression* (SVR) merupakan metode regresi berbasis kernel yang efektif untuk memodelkan hubungan nonlinear (Smola dan Schölkopf, 2004). Namun, kinerja SVR sangat dipengaruhi oleh representasi fitur yang digunakan. Wan *et al.* (2021) menunjukkan bahwa penggunaan fitur hasil reduksi dimensi dengan *Principal Component Analysis* (PCA) dapat meningkatkan kinerja SVR dalam prediksi kuat tekan beton. PCA membentuk komponen utama berdasarkan variasi terbesar data (Jolliffe dan Cadima, 2016), sedangkan

Independent Component Analysis (ICA) menghasilkan komponen yang saling independen secara statistik (Sela dan Kimmel, 2016). Oleh karena itu, reduksi dimensi diperlukan untuk menyederhanakan struktur data, mengurangi redundansi, serta menghasilkan representasi yang lebih informatif.

Sejumlah penelitian telah mengombinasikan teknik reduksi dimensi dengan model regresi. Sharief Basha *et al.* (2025) menggunakan PCA-SVR untuk peramalan radiasi surya, sedangkan Chowdhury *et al.* (2018) mengintegrasikan PCA dan ICA dengan SVR pada peramalan deret waktu finansial dan menunjukkan bahwa model PCA-ICA-SVR memiliki kinerja terbaik. Alshatri *et al.* (2026) juga membandingkan PCA dan ICA sebelum pemodelan SVR dan *Multi-Layer Perceptron* (MLP) pada prediksi harga kredit karbon, dengan hasil terbaik diperoleh oleh ICA-SVR. Namun, dalam konteks prediksi kuat tekan beton, Wan *et al.* (2021) masih berfokus pada pengaruh PCA terhadap beberapa model *machine learning*, sedangkan Zhang *et al.* (2025) menggunakan PCA dalam pengembangan model PCA-PSO-RVM. Oleh karena itu, penelitian ini bertujuan membandingkan kinerja PCA-SVR dan ICA-SVR dalam memprediksi kuat tekan beton.

2. Metodologi

2.1 Data

Penelitian ini menggunakan dataset *Concrete Compressive Strength* yang diperoleh dari UCI Machine Learning Repository. Dataset terdiri atas 1.030 observasi dan 9 peubah, yaitu 8 peubah prediktor dan 1 peubah target. Dataset ini merupakan data multivariat yang dirancang untuk memodelkan kuat tekan beton berdasarkan komposisi bahan penyusun dan faktor umur beton. Pemilihan dataset ini didasarkan pada karakteristik hubungan antara kuat tekan beton dan peubah prediktor yang bersifat multivariat dan nonlinear. Data diperoleh melalui hasil pengujian laboratorium terhadap campuran beton dengan komposisi bahan dan umur beton tertentu. Peubah yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1 : Peubah yang digunakan

Peubah	Satuan	Keterangan peubah
<i>Concrete Compressive Strength</i>	MPa	Y
<i>Cement</i>	kg/m ³	X1
<i>Blast Furnace Slag</i>	kg/m ³	X2
<i>Fly Ash</i>	kg/m ³	X3
<i>Water</i>	kg/m ³	X4
<i>Superplasticizer</i>	kg/m ³	X5
<i>Coarse Aggregate</i>	kg/m ³	X6
<i>Fine Aggregate</i>	kg/m ³	X7
<i>Age</i>	Hari	X8

2.2 Metode Penelitian

Penelitian ini menerapkan ICA sebagai metode reduksi dimensi berbasis *unsupervised learning* yang diintegrasikan dengan SVR untuk memprediksi kuat tekan beton. Model PCA-SVR digunakan sebagai pembanding berbasis variansi untuk mengevaluasi perbedaan performa reduksi dimensi terhadap hasil prediksi. Selain itu, model SVR tanpa reduksi dimensi digunakan sebagai *baseline*. Proses analisis dilakukan melalui beberapa tahapan sebagai berikut:

1. Persiapan dan eksplorasi data
Tahap ini dilakukan dengan memeriksa keberadaan data hilang (*missing value*) pada setiap peubah. Selain itu, dilakukan eksplorasi data untuk memahami karakteristik data melalui analisis distribusi menggunakan histogram, identifikasi pencilan (*outlier*), serta identifikasi multikolinearitas.
2. Pembagian dan standarisasi data
Dataset dibagi menjadi 80% data latih (*training*) dan 20% data uji (*testing*). Standarisasi dilakukan menggunakan metode Z-score pada data *training*, di mana parameter rata-rata (μ) dan standar deviasi (σ) yang dihasilkan kemudian diterapkan pada data *testing* untuk menghindari kebocoran data (*data leakage*).
3. Metode *Principal Component Analysis* (PCA)
 - a. Penentuan jumlah komponen (K)
Jumlah komponen utama pada PCA dipilih dari nilai *cumulative explained variance* dengan ambang batas (*threshold*) 95% (Alshatri *et al.*, 2026).
 - b. Transformasi data ke dalam ruang komponen utama
Fitting PCA pada data *training* dan hasil transformasinya diterapkan pada data *testing* untuk menghindari data *leakage*. Selanjutnya, data peubah prediktor yang telah distandarisasi ditransformasikan ke dalam ruang komponen utama berdasarkan jumlah K yang terpilih.
4. Metode *Independent Component Analysis* (ICA)
Proses ekstraksi fitur pada metode ICA dalam penelitian ini diimplementasikan menggunakan algoritma *Fast Independent Component Analysis* (FastICA).
 - a. Optimasi dan ekstraksi fitur menggunakan FastICA
Proses ekstraksi fitur dilakukan dengan mencari jumlah komponen independen (K) yang paling optimal melalui skema perulangan (*looping*) pada rentang $K = \{1, 2, \dots, 8\}$. Pada setiap iterasi, algoritma FastICA bekerja dengan tahapan sebagai berikut:
 - 1) Prapemrosesan FastICA
 - a) *Centering*: menggeser data agar berpusat di titik nol (*zero-mean*) dengan mengurangi setiap nilai peubah terhadap rata-ratanya sesuai Persamaan (1).

$$x_{centered} = x - E[x] \quad (1)$$
 - b) *Whitening*: transformasi linear untuk menghilangkan korelasi antarpeubah dan menyeragamkan variansi. Proses ini menggunakan *Singular Value Decomposition* (SVD) seperti

pada Persamaan (2).

$$z = \Sigma^{-\frac{1}{2}} U^T x \quad (2)$$

di mana U adalah matriks *unitary* eigenvektor dari x dan Σ adalah matriks diagonal nilai eigen dari data x .

2) Estimasi komponen independen (*fixed-point iteration*)

Melakukan estimasi vektor bobot w_i , menggunakan algoritma *fixed-point iteration* untuk memaksimalkan *non-Gaussianity* (*negentropy*) dengan skema parallel (*symmetric orthogonalization*) yang mengestimasi seluruh komponen independen secara simultan dengan tahapan berikut:

a) Inisialisasi vektor bobot awal w_i secara acak untuk $i = 1, 2, \dots, K$.

b) Melakukan iterasi pembaruan bobot dengan fungsi nonlinear $g(\cdot)$ (logcosh) yang dituliskan seperti pada Persamaan (3).

$$w_i^{new} = E\{zg(w_i^T z)\} - E\{g'(w_i^T z)\}w_i \quad (3)$$

c) Melakukan ortogonalisasi simetris agar komponen yang dihasilkan tetap saling bebas seperti pada Persamaan (4).

$$W = (WW^T)^{-\frac{1}{2}}W \quad (4)$$

d) Proses diulangi hingga mencapai konvergensi dengan maksimal 3000 iterasi dan toleransi 0.001.

3) Penentuan komponen optimal

Setelah algoritma mencapai konvergensi pada setiap nilai K , dilakukan evaluasi secara visual melalui grafik menggunakan dua metrik utama, yaitu:

a) *Negentropy*: ukuran *non-Gaussianity* untuk mengevaluasi sejauh mana komponen hasil transformasi menyimpang dari distribusi *Gaussian*. Nilai yang semakin tinggi menunjukkan tingkat *non-Gaussianity* yang semakin besar.

b) *Mutual information*: ukuran tingkat ketergantungan antar komponen, di mana nilai yang semakin kecil menunjukkan tingkat independensi antarkomponen yang semakin baik.

Mutual information digunakan sebagai ukuran pendukung untuk memperkuat interpretasi independensi antarkomponen yang diperoleh melalui *negentropy*.

b. Transformasi data ke dalam ruang komponen independen

FastICA difit pada data *training* untuk menghindari data *leakage*, kemudian matriks *unmixing* (W) yang diperoleh digunakan untuk mentransformasikan data *testing* ke dalam ruang komponen independen sebagaimana ditunjukkan pada Persamaan (5).

$$s = Wz \quad (5)$$

di mana s adalah matriks fitur independen atau peubah prediktor baru yang akan digunakan dalam model SVR.

5. Pemodelan *Support Vector Regression* (SVR)

Tahap ini dilakukan secara seragam baik untuk fitur hasil PCA maupun FastICA.

a. Pelatihan model (*Hyperparameter Tuning*)

Fitur hasil transformasi digunakan sebagai *input* model SVR dengan menggunakan kernel *Radial Basis Function* (RBF). Optimasi dilakukan pada parameter *regularization* (C), *gamma* (γ), dan *epsilon* (ϵ) menggunakan metode *Grid Search* dengan *Repeated K-Fold Cross-Validation*. Parameter terbaik dipilih berdasarkan nilai *Root Mean Square Error* (RMSE) terkecil.

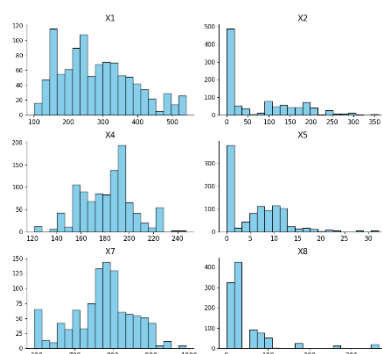
b. Pengujian dan evaluasi model

Model terbaik hasil optimasi kemudian diuji menggunakan data *testing* yang belum pernah digunakan selama proses pelatihan. Performa model dibandingkan menggunakan koefisien determinasi (R^2), RMSE, dan *Mean Absolute Error* (MAE).

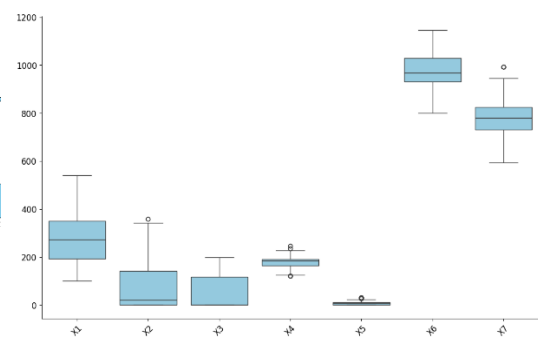
3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Sebelum memasuki tahap pemodelan, dilakukan analisis eksplorasi data untuk memahami karakteristik awal dari data yang digunakan. Analisis ini bertujuan mengidentifikasi pola distribusi data, mendeteksi keberadaan *outlier*, serta evaluasi multikolinearitas antarpeubah prediktor menggunakan *Variance Inflation Factor* (VIF).



Gambar 1: Histogram peubah X



Gambar 2: Boxplot untuk mendeteksi pencilan

Tahap eksplorasi data diawali dengan analisis distribusi menggunakan histogram serta identifikasi pencilan menggunakan boxplot. Berdasarkan Gambar 1, sebagian besar peubah prediktor menunjukkan penyimpangan dari distribusi normal. Hasil analisis *skewness* menunjukkan bahwa peubah X8 memiliki tingkat kemencengan tertinggi dengan nilai *skewness* sebesar 3.2692, diikuti oleh X5 dan X2 dengan nilai *skewness* masing-masing sebesar 0.9081 dan 0.8007. Nilai *skewness* tersebut mengindikasikan adanya karakteristik *non-Gaussian* pada beberapa peubah prediktor sehingga mendukung penggunaan metode ICA yang memanfaatkan sifat *non-Gaussianity* dalam proses ekstraksi komponen independen (Jain *et al.*, 2023). Sementara itu, Gambar 2 menunjukkan adanya beberapa pencilan pada data. Pencilan tersebut tetap dipertahankan dalam analisis karena masih dianggap sebagai bagian dari karakteristik alami data.

Tabel 2: Hasil pengujian VIF

Peubah	X1	X2	X3	X4	X5	X6	X7	X8
VIF	15.4686	3.3304	4.1489	82.1508	5.4724	84.9595	72.7954	1.6994

Selanjutnya, multikolinearitas antarpeubah prediktor dievaluasi menggunakan nilai VIF yang disajikan pada Tabel 2. Hasil pengujian menunjukkan bahwa peubah X1, X4, X6, dan X7 memiliki nilai VIF lebih dari 10, yang mengindikasikan adanya multikolinearitas. Kondisi tersebut menunjukkan adanya redundansi informasi antarpeubah prediktor sehingga penggunaan metode reduksi dimensi seperti PCA dan ICA dipertimbangkan sebelum pemodelan menggunakan SVR.

3.2 Pembagian dan Standarisasi Data

Pada tahap ini, data dibagi menjadi 824 data *training* dan 206 data *testing* menggunakan *random state* 42 agar hasil pembagian tetap konsisten (*reproducible*) ketika dijalankan kembali. Selanjutnya, standarisasi dilakukan menggunakan metode Z-score (*Standard Scaler*) sehingga setiap peubah memiliki rata-rata nol dan standar deviasi satu. Proses *fitting* parameter rata-rata (μ) dan standar deviasi (σ) hanya dilakukan pada data *training*, kemudian parameter tersebut digunakan untuk mentransformasikan data *testing* guna menghindari data *leakage*. Standarisasi diperlukan karena metode reduksi dimensi maupun SVR sensitif terhadap perbedaan skala antarpeubah.

3.3 Pemodelan *Principal Component Analysis* (PCA)

Reduksi dimensi menggunakan PCA dilakukan untuk mereduksi delapan peubah prediktor menjadi beberapa komponen utama dengan mempertahankan sebagian besar variasi data. Jumlah komponen ditentukan berdasarkan *cumulative explained variance* dengan ambang batas 95%. Hasil analisis menunjukkan bahwa lima komponen pertama hanya menjelaskan 87.37% variasi data sehingga belum memenuhi ambang batas yang ditetapkan. Setelah ditambahkan komponen keenam, nilai *cumulative explained variance* meningkat menjadi 97.31%. Rincian hasil PCA disajikan pada Tabel 3.

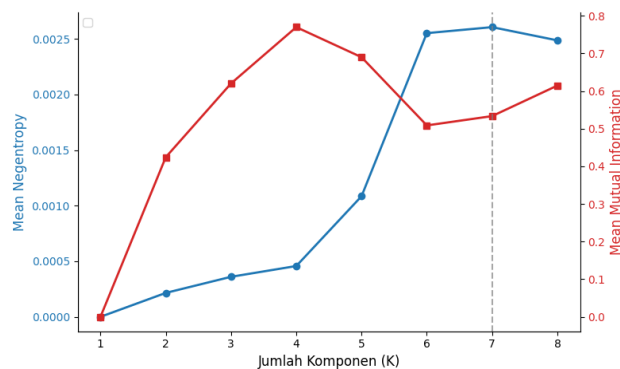
Berdasarkan Tabel 3, sebagian besar variasi data telah dijelaskan oleh komponen-komponen awal, dengan kontribusi yang semakin menurun pada komponen berikutnya. Komponen keenam menjadi titik penting karena nilai *cumulative explained variance* telah melampaui ambang batas 95%. Sementara itu, penambahan komponen ketujuh dan kedelapan hanya meningkatkan variasi yang dijelaskan masing-masing sebesar 2.30% dan 0.39%. Oleh karena itu, enam komponen utama dipilih untuk selanjutnya digunakan sebagai fitur *input* pada pemodelan SVR.

Tabel 3: Nilai eigen, *explained variance*, dan *cumulative variance* pada PCA

Jumlah komponen (K)	Nilai eigen	<i>Explained variance</i> (%)	<i>Cumulative explained variance</i> (%)
1	2.2440	28.02	28.02
2	1.4345	17.91	45.93
3	1.3481	16.83	62.76
4	1.0255	12.80	75.56
5	0.9461	11.81	87.37
6	0.7961	9.94	97.31
7	0.1844	2.30	99.61
8	0.0311	0.39	100.00

3.4 Metode *Independent Component Analysis* (ICA)

Pada metode ICA, algoritma FastICA digunakan untuk mengekstraksi komponen yang saling independen dari data. Dalam penelitian ini, jumlah komponen independen (K) pada FastICA tidak ditentukan secara langsung, melainkan diuji secara bertahap dari $K = 1$ hingga $K = 8$ sesuai dengan jumlah fitur pada data. Untuk setiap nilai K , algoritma FastICA diterapkan, kemudian hasil ekstraksi komponennya dievaluasi menggunakan rata-rata *negentropy* dan estimasi *mutual information* (Saeedi *et al.*, 2025). Hasil evaluasi tersebut kemudian divisualisasikan melalui grafik pada Gambar 3 untuk memudahkan analisis dalam menentukan jumlah komponen (K) yang optimal.



Gambar 3: Grafik penentuan jumlah komponen ICA berdasarkan *negentropy* dan *mutual information*

Berdasarkan Gambar 3, algoritma FastICA pada seluruh pengujian jumlah komponen telah mencapai kondisi konvergen. Penentuan jumlah komponen (K) optimal dilakukan dengan mempertimbangkan *trade-off* antara nilai *negentropy* dan *mutual information*. Hasil evaluasi menunjukkan bahwa nilai *mean mutual information* mencapai titik minimum pada $K = 6$ sebesar 0.5087, sedangkan nilai *negentropy* tertinggi diperoleh pada $K = 7$ sebesar 0.0026. Perubahan nilai *mutual information* dari $K = 6$ ke $K = 7$ relatif kecil, yaitu sebesar 0.0248, sehingga peningkatan

ketergantungan antarkomponen masih dapat diterima. Di sisi lain, $K = 7$ menghasilkan nilai *negentropy* tertinggi yang menunjukkan tingkat *non-Gaussianity* yang lebih tinggi dibandingkan jumlah komponen lainnya. Mengingat prinsip utama algoritma FastICA adalah memaksimalkan *non-Gaussianity* untuk memperoleh komponen yang saling independen, maka $K = 7$ dipilih sebagai jumlah komponen independen optimal untuk digunakan sebagai peubah *input* pada tahap pemodelan SVR.

3.5 Pemodelan Support Vector Regression (SVR)

Pemodelan dilakukan menggunakan SVR sebagai *baseline* serta pada fitur hasil reduksi dimensi dengan jumlah komponen optimal pada pendekatan PCA dan ICA. Kernel RBF digunakan karena mampu menangkap hubungan nonlinear pada data (Li et al., 2024). Optimasi *hyperparameter* menggunakan *grid Search* dengan *repeated 5-fold cross-validation* meliputi parameter $C = \{0.1, 1, 10, 100, 1000\}$, $\gamma = \{\text{scale}, \text{auto}, 0.0001, 0.001, 0.01, 0.1\}$, dan $\varepsilon = \{0.01, 0.05, 0.1, 0.5\}$. Parameter optimal yang diperoleh pada model SVR adalah $C = 100$, $\gamma = 0.1$, dan $\varepsilon = 0.5$, pada PCA-SVR diperoleh $C = 100$, $\gamma = 0.1$, dan $\varepsilon = 0.5$, sedangkan pada ICA-SVR diperoleh $C = 100$, $\gamma = \text{scale}$, dan $\varepsilon = 0.5$. Kinerja model dievaluasi pada data *testing* menggunakan metrik RMSE, MAE, dan R^2 yang disajikan pada Tabel 4.

Tabel 4: Perbandingan performa model SVR, PCA-SVR, dan ICA-SVR

Model	RMSE	MAE	R^2
SVR	5.8221	4.1401	0.8685
PCA-SVR	6.8617	4.9632	0.8172
ICA-SVR	5.9435	4.1039	0.8629

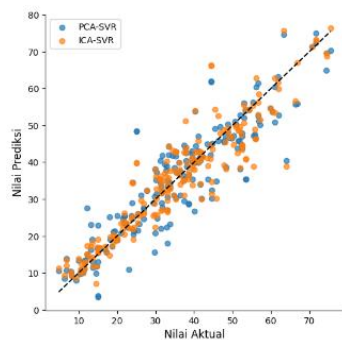
Analisis sensitivitas terhadap variasi nilai *hyperparameter* C , γ , dan ε menunjukkan bahwa setiap parameter memberikan pengaruh yang berbeda terhadap performa model. Parameter regularisasi (C) memengaruhi kompleksitas model, di mana nilai $C = 100$ terpilih sebagai nilai optimal pada seluruh model. Nilai C yang lebih besar cenderung meningkatkan kompleksitas model sehingga berpotensi menyebabkan *overfitting*. Risiko tersebut dapat diminimalkan melalui penggunaan *repeated 5-fold cross-validation* pada proses *tuning* sehingga estimasi performa model menjadi lebih stabil. Parameter γ pada kernel RBF mengatur fleksibilitas model dalam menangkap pola nonlinear pada data. Pada ICA-SVR, penggunaan $\gamma = \text{scale}$ memberikan kinerja yang baik karena menyesuaikan skala fitur hasil transformasi FastICA. Sementara itu, nilai $\varepsilon = 0.5$ pada seluruh model menunjukkan bahwa model lebih stabil ketika diberikan toleransi kesalahan yang lebih besar sehingga dapat mengurangi pengaruh *noise* kecil pada data.

Berdasarkan Tabel 4, model ICA-SVR menghasilkan nilai RMSE dan MAE yang lebih rendah serta nilai R^2 yang lebih tinggi dibandingkan PCA-SVR. Hasil tersebut menunjukkan bahwa pendekatan ICA-SVR memberikan performa prediksi yang lebih baik dibandingkan PCA-SVR

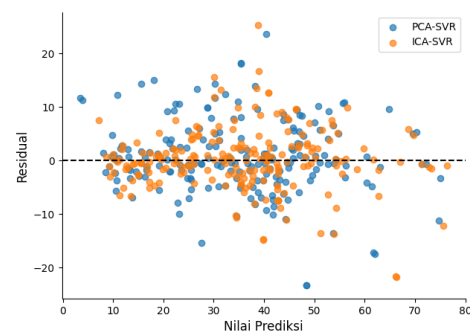
pada dataset dan konfigurasi penelitian ini. Namun demikian, model SVR tanpa reduksi dimensi (*baseline*) masih menghasilkan performa yang sedikit lebih baik dengan nilai RMSE sebesar 5.8221 dan nilai R^2 sebesar 0.8685. Hasil ini menunjukkan bahwa penerapan reduksi dimensi belum mampu meningkatkan performa prediksi dibandingkan penggunaan seluruh peubah asli. Meskipun demikian, di antara kedua metode reduksi dimensi yang dibandingkan, ICA-SVR menghasilkan performa yang lebih mendekati model *baseline* dibandingkan PCA-SVR.

Perbedaan performa antara PCA-SVR dan ICA-SVR diuji menggunakan *Wilcoxon signed-rank test* berdasarkan hasil evaluasi *repeated cross-validation*. Hasil analisis menunjukkan bahwa perbedaan performa kedua model signifikan pada taraf signifikansi 5%, sehingga performa ICA-SVR yang lebih baik dibandingkan PCA-SVR pada penelitian ini didukung secara statistik. Selain itu, evaluasi menggunakan *repeated 5-fold cross-validation* menunjukkan bahwa ICA-SVR memiliki rata-rata RMSE sebesar 6.4149 ± 0.6815 , MAE sebesar 4.5585 ± 0.3971 dan R^2 sebesar 0.8522 ± 0.0337 . Nilai standar deviasi yang relatif kecil mengindikasikan bahwa performa model cukup stabil terhadap variasi pembagian data selama proses *repeated cross-validation*.

Meskipun demikian, penerapan PCA-SVR dan ICA-SVR menyebabkan interpretabilitas hubungan langsung antara peubah asli dan *output* model menjadi berkurang karena fitur hasil transformasi tidak lagi merepresentasikan peubah awal secara individual, melainkan berupa kombinasi linear (PCA) atau komponen independen (ICA) dari seluruh peubah prediktor. Kondisi tersebut menunjukkan adanya *trade-off* antara penyederhanaan representasi data dan kemudahan interpretasi model, sehingga penelitian ini lebih menekankan pada aspek akurasi prediksi dibandingkan interpretasi kontribusi masing-masing peubah.



Gambar 4: *Plot* nilai aktual dan nilai prediksi model PCA-SVR dan ICA-SVR



Gambar 5: *Plot* residual model PCA-SVR dan ICA-SVR

Evaluasi model juga diperkuat melalui visualisasi hubungan nilai aktual dan prediksi pada Gambar 4 serta analisis residual pada Gambar 5. Berdasarkan Gambar 4, sebagian besar titik observasi pada kedua model berada di sekitar garis diagonal, dengan sebaran titik pada ICA-SVR terlihat lebih rapat dibandingkan PCA-SVR. Hal ini menunjukkan bahwa prediksi ICA-SVR lebih mendekati nilai aktual. Sementara itu, *plot* residual

pada Gambar 5 menunjukkan bahwa residual kedua model menyebar secara acak di sekitar garis nol tanpa pola tertentu. Ringkasan statistik residual menunjukkan bahwa ICA-SVR memiliki rata-rata residual sebesar 0.1823 dengan standar deviasi 5.9552, sedangkan PCA-SVR memiliki rata-rata residual sebesar 0.5989 dengan standar deviasi 6.852. Selain itu, rentang residual ICA-SVR juga relatif lebih sempit. Hasil tersebut mengindikasikan bahwa kesalahan prediksi ICA-SVR lebih stabil dan memiliki variasi yang lebih rendah dibandingkan PCA-SVR.

4. Simpulan dan Saran

Hasil reduksi dimensi menunjukkan bahwa PCA menghasilkan enam komponen utama, sedangkan FastICA menghasilkan tujuh komponen independen optimal. Berdasarkan hasil evaluasi dan pengujian statistik, model ICA-SVR menghasilkan performa yang lebih baik dibandingkan PCA-SVR. Namun, model SVR tanpa reduksi dimensi masih memberikan performa yang sedikit lebih baik. Dengan demikian, pada kasus data kuat tekan beton yang digunakan dalam penelitian ini, pendekatan ICA-SVR lebih baik dibandingkan PCA-SVR, meskipun reduksi dimensi belum mampu melampaui performa model *baseline*. Untuk pengembangan selanjutnya, penelitian dapat diperluas dengan membandingkan ICA-SVR terhadap algoritma lain seperti *Random Forest*, *XGBoost*, dan *Artificial Neural Network* (ANN), serta mengeksplorasi metode reduksi dimensi nonlinear seperti *Kernel PCA*, *Autoencoder*, atau *Nonlinear ICA*. Selain itu, penggunaan teknik optimasi *hyperparameter* yang lebih kompleks, seperti *Bayesian Optimization* atau *Genetic Algorithm*, serta dataset dengan karakteristik yang lebih beragam dapat dipertimbangkan untuk mengevaluasi kemampuan generalisasi model.

Daftar Pustaka

- Alshatri, N., Ghannam, S., & Khadeer Hussain, F. (2026). Intelligent data-driven models for the accurate multi-factor prediction of carbon credit prices. *Discover Artificial Intelligence*. <https://doi.org/10.1007/s44163-026-00894-0>
- Chowdhury, U. N., Chakravarty, S., & Hossain, M. T. (2018). Short-term financial time series forecasting integrating principal component analysis and independent component analysis with support vector regression. *Journal of Computer and Communications*, 6, 51–67. <https://doi.org/10.4236/jcc.2018.63004>
- Christiono, N., & Prayogo, D. (2020). Perbandingan kinerja artificial intelligence dalam memprediksi kuat tekan beton. *Dimensi Utama Teknik Sipil*, 7(2), 1–17. <https://doi.org/10.9744/duts.7.2.1-17>
- Jain, P., Knight, B. W. Jr., & Victor, J. D. (2023). Application of a Hermite-based measure of non-Gaussianity to normality tests and independent

- component analysis. *Frontiers in Neuroinformatics*, 17, 1113988. <https://doi.org/10.3389/fninf.2023.1113988>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Li, G., Fu, Y., Deng, Z., & Huang, Y. (2024). Predicting the capillary water absorption values of building stones by hybrid models. *Journal of Applied Science and Engineering*, 28(6): 1311-1320.
- Paudel, S., Pudasaini, A., Shrestha, R. K., & Kharel, E. (2023). Compressive strength of concrete material using machine learning techniques. *Cleaner Engineering and Technology*, 15, 100661. <https://doi.org/10.1016/j.clet.2023.100661>
- Prayogo, D. (2018). Prediksi kuat tekan beton dengan menggunakan metode artificial intelligence. *Seminar Nasional Ilmu Terapan (SNITER)*, 2(1).
- Saeedi, N., Baharvand, D., Shirini, K., & Gharehveran, S. S. (2025). Prediction of electrical energy consumption using principal component analysis and independent components analysis. *The Journal of Supercomputing*, 81(9).
- Sela, M., & Kimmel, R. (2016). *Randomized independent component analysis*. arXiv. <https://arxiv.org/abs/1609.06942>
- Sharief Basha, S. S., Nagaraja Rao, A., & Nida Fariz, T. K. (2025). Dimensionality reduction in solar radiation forecasting: A combined PCA-SVM framework for renewable energy applications. *Frontiers in Environmental Science*, 13, 1692076. <https://doi.org/10.3389/fenvs.2025.1692076>
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Wan, Z., Xu, Y., & Šavija, B. (2021). On the use of machine learning models for prediction of compressive strength of concrete: Influence of dimensionality reduction on the model performance. *Materials*, 14(4), 713. <https://doi.org/10.3390/ma14040713>
- Zhang, Y., Ye, Y., Wang, J., Tang, B., & Fu, F. (2025). Strength prediction of self-compacting concrete using improved RVM machine learning method. *International Journal of Concrete Structures and Materials*, 19, 101. <https://doi.org/10.1186/s40069-025-00835-8>