

Seleksi Fitur Berdasarkan Korelasi Pearson dalam Pemodelan Efisiensi Energi Bangunan

Feature Selection Based on Pearson Correlation in Building Energy Efficiency Modeling

Desmarita Leni¹, Aggrivina Dwiharzandis², Ruzita Sumiati³, Haris⁴, Sicilia Afriyani^{5*}

¹Program Studi Teknik Mesin, Universitas Muhammadiyah Sumatera Barat, Indonesia

²Program Studi Teknik Elektro, Universitas Muhammadiyah Sumatera Barat, Indonesia

^{3,4}Program Studi Teknik Mesin, Politeknik Negeri Padang, Indonesia

⁵Program Studi Teknik Sipil, Politeknik Negeri Padang, Indonesia

Email: ^{5*}sicilia@pnp.ac.id

Abstrak

Prediksi beban pemanasan dan pendinginan bangunan merupakan langkah penting untuk perencanaan dan pengelolaan sistem energi. Hal ini, tidak terlepas dari berkontribusi beban pemanasan dan pendinginan bangunan yang menyumbang 30% dari total konsumsi energi global. Penelitian ini bertujuan untuk menerapkan metode seleksi fitur berdasarkan korelasi Pearson dalam pemodelan prediksi beban pemanasan dan pendinginan bangunan menggunakan *Artificial Neural Network* (ANN). Korelasi Pearson digunakan untuk menganalisis hubungan antara variabel input dan variabel target. Fitur-fitur yang memiliki korelasi signifikan dengan variabel target digunakan sebagai dataset untuk pelatihan model, sedangkan yang tidak memiliki korelasi signifikan dihapus dari dataset pelatihan. Evaluasi dilakukan menggunakan metrik evaluasi seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared* untuk mengukur tingkat keakuratan dan kinerja model dalam memprediksi beban pemanasan dan pendinginan. Hasil pemodelan menunjukkan bahwa seleksi fitur berdasarkan korelasi Pearson menghasilkan prediksi yang sangat akurat untuk beban pemanasan dan pendinginan bangunan. Model ini menunjukkan kinerja yang baik selama pelatihan dan validasi dengan *Cross Validation* (CV) menggunakan $k = 10$. Hasil evaluasi model diperoleh nilai MAE 0.457, RMSE 0.628, dan *R-squared* 0.996 untuk beban pemanasan, sedangkan untuk beban pendinginan diperoleh nilai MAE sebesar 1.163, RMSE 1.74, dan *R-squared* 0.967. Hasil ini mengindikasikan bahwa seleksi fitur dengan korelasi Pearson dapat dijadikan pendekatan yang efektif untuk meningkatkan performa model prediksi menggunakan machine learning, terutama dalam konteks prediksi beban pemanasan dan pendinginan bangunan.

Kata kunci: *Artificial Neural Network*, Energi, Fitur Pemodelan, Korelasi Pearson, Seleksi

Abstract

Predicting the heating and cooling loads of buildings is a crucial step in energy system planning and management. This is not only due to the substantial contribution of building heating and cooling loads, which account for 30% of the total global energy consumption. This research aims to apply feature selection methods based on Pearson correlation in modeling the prediction of heating and cooling loads in buildings using Artificial Neural Networks (ANN). Pearson correlation is used to analyze the relationship between input variables and the target variable. Features that have a significant correlation with the target variable are used as the dataset for model training, while those with no significant correlation are removed from the

training dataset. Evaluation is performed using evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to measure the accuracy and performance of the model in predicting heating and cooling loads. The modeling results show that feature selection based on Pearson correlation produces highly accurate predictions for building heating and cooling loads. This model demonstrates good performance during training and validation with Cross Validation (CV) using $k = 10$. The model evaluation results yield an MAE of 0.457, RMSE of 0.628, and R-squared of 0.996 for heating loads, while for cooling loads, an MAE of 1.163, RMSE of 1.74, and R-squared of 0.967 are obtained. These results indicate that feature selection using Pearson correlation can be an effective approach to enhance the performance of prediction models using machine learning, especially in the context of predicting building heating and cooling loads.

Keywords: Artificial Neural Network, Energy, Modeling Features, Pearson Correlation, Selection

PENDAHULUAN

Bangunan berperan besar dalam konsumsi energi global, dengan sekitar 30% energi dunia digunakan untuk memenuhi kebutuhan pemanasan dan pendinginan bangunan [1]. Hal ini tidak terlepas dari bangunan yang memerlukan energi untuk mempertahankan suhu yang nyaman di dalamnya, terutama dalam iklim yang ekstrem. Beban pemanasan dan pendinginan bangunan dapat dipengaruhi oleh banyak faktor seperti desain bangunan, sistem HVAC (*Heating, Ventilation, and Air Conditioning*), iklim, serta bahan-bahan yang digunakan dalam konstruksi bangunan [1]. Menurut [2] orientasi bangunan terhadap matahari dan arah angin dapat mempengaruhi seberapa baik bangunan memanfaatkan panas matahari alami dan ventilasi alam.

Dengan merancang bangunan agar terkena matahari pada waktu yang tepat dan mengalirkan udara segar, konsumsi energi untuk pemanasan dan pendinginan dapat berkurang. Dalam studi [3], juga ditemukan bahwa memaksimalkan penggunaan pencahayaan alami di dalam bangunan dapat mengurangi kebutuhan pencahayaan buatan. Ini dapat dicapai melalui penggunaan jendela besar, atap kaca, dan rancangan interior yang mendukung pencahayaan alami. Secara keseluruhan, upaya yang dapat dilakukan untuk meminimalisir beban pemanasan dan

pendingin ruangan adalah dengan memiliki desain bangunan yang lebih hemat energi dengan sifat konservasi energi yang lebih baik.

Energy informatics adalah suatu disiplin ilmu yang berkaitan dengan pengumpulan, analisis, pemodelan, dan pemanfaatan data untuk mengelola, memantau, dan mengoptimalkan penggunaan energi [4–5]. Salah satu aspek penting dalam energy informatics adalah penggunaan metode machine learning untuk memodelkan dan memprediksi konsumsi energi pada sberbagai aplikasi bangunan termasuk beban pemanasan dan pendinginan [5], [6]. *Artificial Neural Network* (ANN) merupakan sebuah teknik atau pendekatan pengolahan informasi yang terinspirasi oleh cara kerja sistem saraf biologis, khususnya pada sel otak manusia dalam memproses informasi. *Neural Network* terdiri dari sejumlah besar elemen pemrosesan informasi (*neuron*) yang saling terhubung dan bekerja bersama-sama untuk menyelesaikan sebuah masalah tertentu, yang pada umumnya adalah masalah klasifikasi ataupun prediksi [7].

Banyak penelitian terdahulu yang telah membahas penggunaan metode *machine learning* dengan algoritma ANN untuk memprediksi efisiensi energi seperti [1], yang memprediksi kinerja energi bangunan menggunakan berbagai teknik data *mining*. Penelitian ini melibatkan penggunaan berbagai teknik termasuk

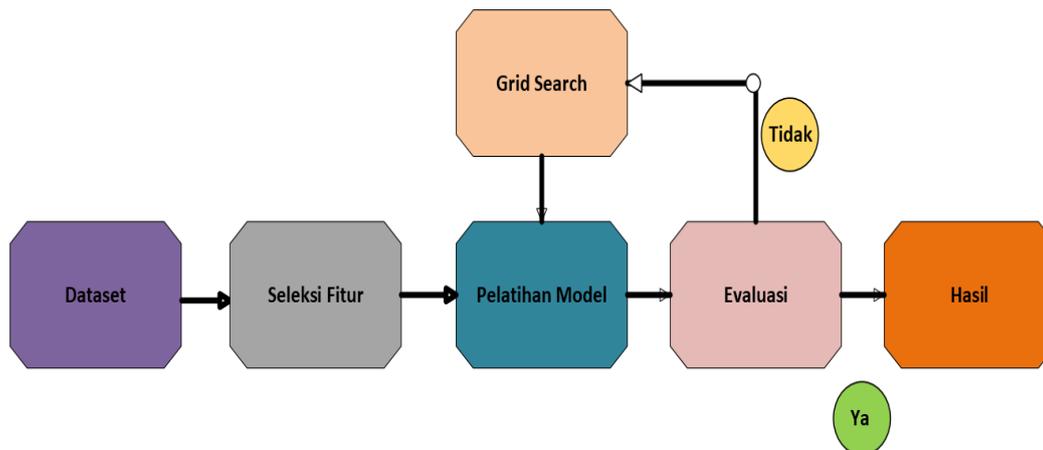
Support Vector Regression (SVR), *Artificial Neural Network (ANN)*, *Classification and Regression Tree*, *Chi-squared Automatic Interaction Detector*, *General Linear Regression*, dan model ensambel. Data yang digunakan terdiri dari 768 set data eksperimental dengan 8 parameter masukan dan 2 parameter keluaran (*cooling load* dan *heating load*). Hasil penelitian menunjukkan bahwa model ANN memiliki kinerja yang lebih baik dibandingkan dengan model lainnya, hal ini dapat dilihat dari kesalahan persentase rata-rata yang lebih rendah dari 4%. Pada studi yang dilakukan [8], juga ditemukan bahwa ANN lebih efektif dalam melakukan prediksi beban pemanasan dan pendinginan dibandingkan dengan pendekatan lain seperti analisis regresi dan *Random Forest*. Pada penelitian tersebut ketiga model machine learning diuji menggunakan data simulasi 12 bangunan yang berbeda dengan 8 fitur masukan dan 2 variabel *output*.

Berdasarkan permasalahan yang telah diuraikan di atas dan dampak positif dari penelitian terdahulu menggunakan ANN sebagai pemodelan untuk memprediksi beban pemanasan dan pendinginan bangunan. Penelitian ini, merancang sebuah pemodelan *machine learning* dengan ANN dan menerapkan seleksi fitur menggunakan korelasi Pearson dalam menentukan variabel input. Korelasi Pearson adalah metode statistik yang digunakan untuk mengukur hubungan linear antara dua

variabel numerik[9], sehingga dapat digunakan untuk mengidentifikasi fitur-fitur yang memiliki hubungan kuat dan relevan dengan variabel *output*. Dengan memilih fitur-fitur yang memiliki korelasi yang tinggi dengan variabel *output*, dapat mengurangi dimensi data dan meningkatkan efisiensi pemodelan.

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian eksperimental dengan menggunakan pendekatan kuantitatif. Penelitian eksperimental adalah jenis penelitian yang dilakukan dengan mengontrol variabel-variabel penelitian dan memvariasikan satu atau beberapa variabel independen untuk melihat pengaruhnya terhadap variabel dependen [10]. Pada penelitian ini dilakukan sebuah perancangan model *machine learning* untuk memprediksi konsumsi energi bangunan seperti beban pemanasan dan pendinginan berdasarkan fitur bangunan. Fitur bangunan yang akan digunakan sebagai input dalam dataset pelatihan ditentukan dengan korelasi Pearson, dimana fitur yang memiliki korelasi sangat lemah dihilangkan dari dataset. Pemodelan dibuat menggunakan bahasa pemrograman *python* dengan *framework Tensorflow* dan dijalankan di *google colab*. Tahapan penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Diagram Alir Penelitian

Dataset

Dataset yang digunakan dalam penelitian ini, berasal dari publikasi [6], tentang efisiensi energi 12 jenis bangunan yang disimulasikan dalam *Ecotect*. Semua bangunan memiliki volume yang sama (771,75 m³) dan menggunakan material yang sama untuk setiap blok, tetapi memiliki luas permukaan dan dimensi yang berbeda. Data simulasi mengasumsikan bahwa aktivitas di dalam bangunan bersifat statis.

Eksperimen ini menginvestigasi delapan parameter masukan dan dua parameter keluaran dari bangunan hunian, yaitu kepadatan relatif (RC), luas permukaan, luas dinding, luas atap, tinggi keseluruhan, orientasi, luas kaca, distribusi kaca, beban pendinginan dan beban pemanasan. Statistik dataset efisiensi energy yang digunakan pada penelitian ini dapat dilihat pada table 1.

Tabel 1. Statistik dataset

No.	Variabel	Kategori	Min	Max	Mean
1	Kepadatan relatif	Input	0.62	0.98	0.76
2	Luas permukaan (m ²)	Input	514.5	808.5	671.7
3	Luas dinding (m ²)	Input	245	416.5	318.5
4	Luas atap (m ²)	Input	110.25	220.5	176.6
5	Tinggi keseluruhan (m)	Input	3.5	7	5.25
6	Orientasi	Input	2	5	3.5
7	Luas kaca	Input	0	0.4	0.23
8	Distribusi kaca	Input	0	5	2.81
9	Beban pendinginan (kWh/m ²)	Output	6.01	43.1	22.3
10	Beban pemanasan (kWh/m ²)	Output	10.9	48.03	24.58

Fitur Selection

Seleksi fitur adalah proses memilih subset variabel input yang paling relevan dengan variabel output [11]. Tujuannya adalah mengurangi dimensi data, meningkatkan kinerja model, menghindari overfitting, dan menghemat waktu

komputasi [12]. Dalam penelitian ini, variabel input dipilih berdasarkan korelasi Pearson. Variabel input yang memiliki korelasi yang sangat rendah dengan variabel output akan dihilangkan dari dataset pelatihan. Korelasi Pearson dapat dihitung dengan menggunakan Persamaan 1

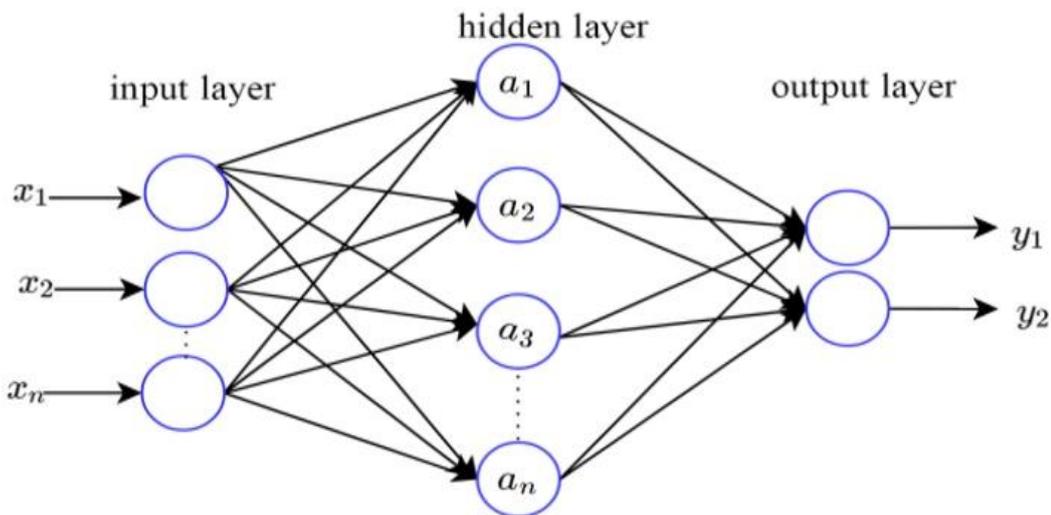
$$r_{xy} = \frac{\sum xy}{(n-1)S_x S_y} \dots\dots\dots(1)$$

Dimana r_{xy} adalah koefisien korelasi Pearson, $\sum xy$ adalah jumlah perkalian x dan y , n adalah ukuran sampel, x adalah variabel independen, y adalah variabel dependen, dan S adalah standar deviasi. Nilai koefisien korelasi berkisar dari -1 hingga 1. Nilai -1 menunjukkan korelasi

negatif yang kuat antara kedua variabel, nilai 0 menunjukkan tidak ada korelasi, dan nilai 1 menunjukkan korelasi positif yang kuat [13].

Pelatihan Model

Arsitektur Artificial Neural Network (ANN) yang digunakan dalam penelitian ini terdiri dari tiga lapisan utama yang saling terhubung, seperti yang terlihat pada Gambar 2.



Gambar 2. Arsitektur ANN

Layer pertama adalah layer input, yang mengakomodasi sejumlah neuron sesuai dengan jumlah variabel input dalam dataset. Layer kedua terdiri dari beberapa layer tersembunyi (*hidden layer*) yang bertugas untuk mengekstraksi informasi yang relevan dari input sebelum meneruskannya ke layer output. Layer-layer tersembunyi ini menggunakan fungsi aktivasi ReLU (*Rectified Linear Unit*) untuk mengatasi masalah vanishing gradient dan mempercepat konvergensi model. Layer terakhir adalah layer output yang memberikan hasil prediksi beban pemanasan dan pendinginan bangunan. Selama proses pelatihan model, juga dilakukan pencarian parameter terbaik dengan menggunakan teknik grid search

untuk mengoptimalkan kinerja model. Proses pencarian parameter ini mencakup pengujian *optimizer*, *learning rate*, dan *batch size* yang berbeda, yang bertujuan untuk mendapatkan konfigurasi terbaik untuk model ANN. Setelah pelatihan, model dievaluasi menggunakan tiga metrik evaluasi, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared*. Metrik evaluasi dapat dihitung menggunakan persamaan berikut :

Mean Absolute Error (MAE)

$$MAE = \frac{i}{N} \sum |y_i - z_i| \dots\dots\dots(2)$$

Dimana i adalah indeks dari data pada sample, N adalah jumlah total sample, y_i

adalah nilai aktual dari data ke-i, sedangkan z_i adalah nilai prediksi dari model untuk data ke-i.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(X_i) - Y_i)^2} \dots \dots \dots (3)$$

Dimana n adalah jumlah data yang digunakan untuk menguji model, $f(X_i)$ adalah nilai yang diprediksi oleh model untuk data ke-i, Y_i adalah nilai sebenarnya untuk data ke-i.

R-squared

$$R = \frac{\sum_{i=1}^n (f(X_i) - \bar{f})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (f(X_i) - \bar{f})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \dots \dots \dots (4)$$

Dimana $f(X_i)$ adalah nilai prediksi dari variabel dependen (Y) berdasarkan variabel independen (X) pada observasi ke-i, \bar{f} adalah rata-rata dari semua nilai prediksi $f(X_i)$ pada seluruh observasi, Y_i adalah nilai observasi aktual dari variabel dependen pada observasi ke-i, \bar{Y} adalah rata-rata dari semua nilai observasi Y_i pada seluruh observasi, dan n adalah jumlah total observasi.

Validasi model

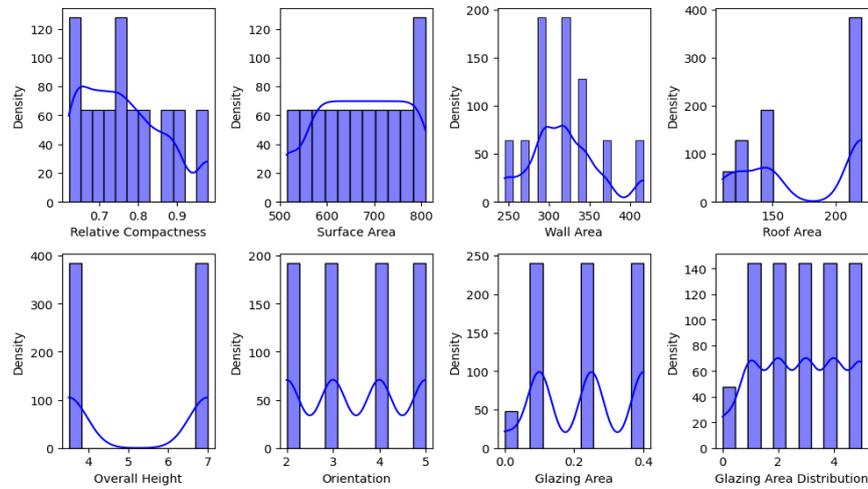
Cross-Validation (CV) adalah salah satu teknik validasi model yang umum digunakan dalam *machine learning*. Pada CV, dataset dibagi menjadi beberapa subset yang disebut *fold*, di mana salah satu *fold* digunakan sebagai data pengujian dan *fold-fold* lainnya digunakan sebagai data pelatihan [14]. Pada penelitian ini model di validasi menggunakan CV dengan $k=10$, yang berarti data dibagi menjadi 10 *fold*. Dengan menggunakan CV $k=10$, penelitian ini dapat mengevaluasi model ANN dengan

lebih kuat karena model diuji pada berbagai kombinasi data pengujian dan pelatihan, yang meminimalkan risiko *overfitting* dan memastikan bahwa model dapat menggeneralisasi dengan baik pada data baru. Setiap *fold* dievaluasi menggunakan metrik evaluasi seperti MAE, RMSE, dan R-squared. Sehingga hasil evaluasi dari keseluruhan iterasi digunakan untuk menghitung rata-rata performa model secara keseluruhan.

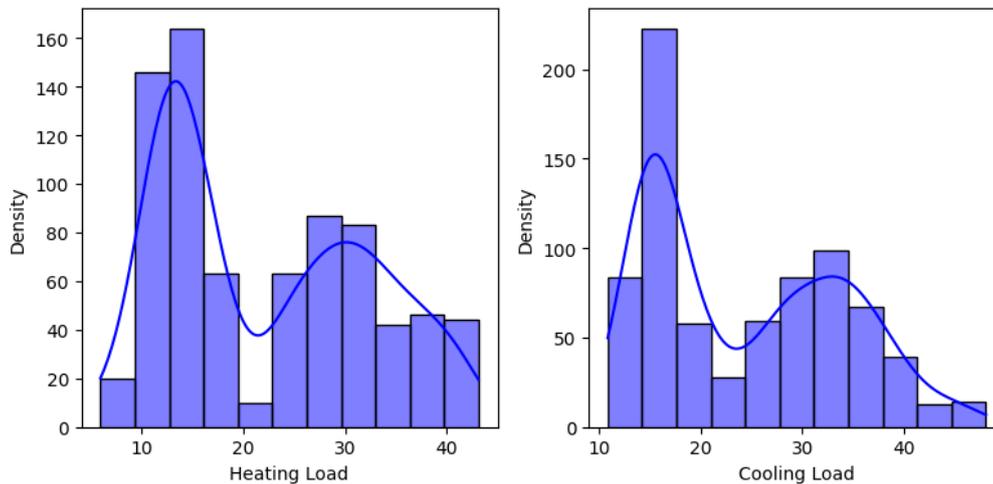
HASIL DAN PEMBAHASAN

Dataset

Dataset ini merupakan data dari simulasi, untuk menghasilkan berbagai bentuk bangunan yang berbeda dengan mengambil kubus dasar ($3,5 \times 3,5 \times 3,5$). Ada 12 bentuk bangunan yang dihasilkan, dan setiap bentuk bangunan terdiri dari 18 elemen dasar (kubus-kubus kecil). Meskipun volume semua bangunan sama, yaitu 771,75 m³, namun setiap kubus memiliki luas permukaan dan dimensi yang berbeda [6]. Simulasi dilakukan di daerah Athens Yunani, dengan menggabungkan 12 bentuk bangunan, tiga variasi area kaca, lima distribusi area kaca, dan empat orientasi, terdapat total $12 \times 3 \times 5 \times 4 = 720$ sampel bangunan yang dihasilkan. Selanjutnya, ada 12 bentuk bangunan untuk keempat orientasi tanpa area kaca. Dengan demikian, secara keseluruhan, ada $12 \times 3 \times 5 \times 4 + 12 \times 4 = 768$ bangunan yang telah diuji. Setiap bangunan dari 768 bangunan yang dihasilkan memiliki delapan parameter bangunan yang menjadi variabel masukan dan dua variabel output. Distribusi data efisiensi energi yang digunakan pada penelitian ini dapat dilihat pada gambar 3.



(a)



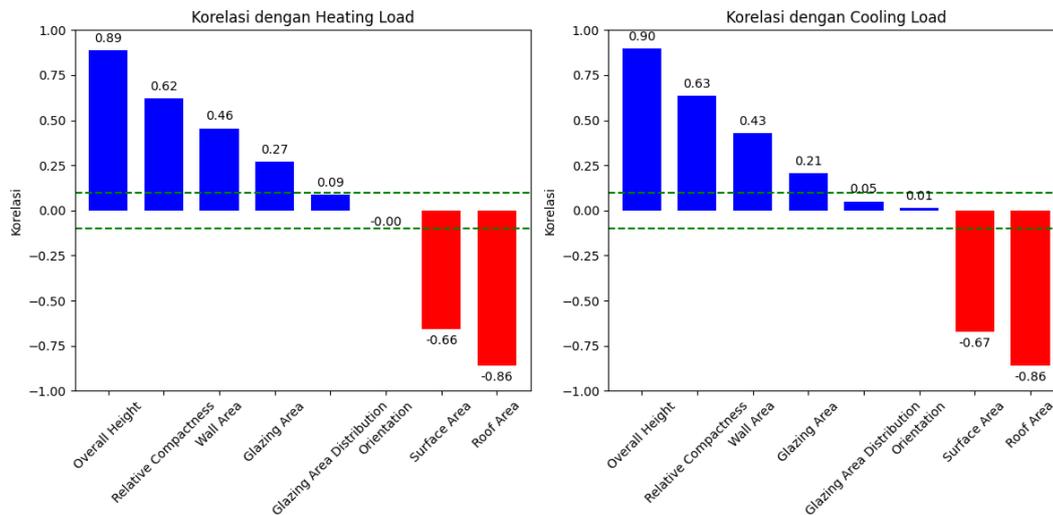
(b)

Gambar 3. Distribusi variable input (a) dan distribusi variable output (b)

Fitur seleksi

Seleksi fitur adalah proses memilih subset fitur (variabel) dari sekumpulan fitur yang tersedia dalam dataset. Hal ini dilakukan untuk mengidentifikasi dan mempertahankan hanya fitur-fitur yang paling relevan atau penting untuk membangun model prediksi yang baik [15].

Seleksi fitur pada penelitian ini menggunakan korelasi pearson, dimana variabel input yang memiliki korelasi sangat kecil (-0.1 hingga 0.1) dengan variabel output seperti beban pemanasan dan pendinginan dihilangkan dari dataset untuk pelatihan model ANN. Hasil seleksi fitur menggunakan korelasi pearson dapat dilihat pada gambar 4.



Gambar 4. Hasil Korelasi Fitur Bangunan dengan Beban Pemanasan dan Pendinginan

Berdasarkan hasil korelasi pearson dapat dilihat bahwa ada dua variabel input yang memiliki nilai korelasi dibawah nilai yang sudah ditentukan, yaitu distribusi kaca dan orientasi. Variabel ini memiliki korelasi yang sangat lemah dan bahkan tidak memiliki korelasi dengan kedua variabel target. Namun, terdapat hasil yang berbeda dari variabel tinggi keseluruhan bangunan (*Overall height*), yang memiliki korelasi positif yang sangat kuat dengan beban pemanasan dan pendinginan dengan masing-masing nilai sebesar 0.89 dan 0.90. Hasil ini mengindikasikan bahwa semakin tinggi suatu bangunan maka akan mengakibatkan peningkatan beban pemanasan dan pendinginan untuk bangunan tersebut. Hasil ini sejalan dengan penelitian terdahulu di bidang manajemen energi bangunan [16], yang menunjukkan bahwa tinggi bangunan merupakan salah satu faktor yang mempengaruhi konsumsi energi, terutama dalam hal pemanasan dan pendinginan. Bangunan yang lebih tinggi cenderung memiliki volume interior yang lebih besar, yang pada gilirannya memerlukan lebih banyak energi untuk mempertahankan suhu di dalamnya. Selain itu, juga dapat dilihat bahwa kepadatan relatif, luas dinding dan luas kaca memiliki

korelasi positif dengan beban pemanasan dan pendinginan. Ini menunjukkan bahwa semakin tinggi kepadatan relatif bangunan (seberapa padatnya bangunan), semakin besar luas dinding, dan semakin besar luas kaca, maka semakin besar pula kebutuhan energi untuk pemanasan dan pendinginan bangunan. Hasil yang berbeda diperoleh dari luas atap (*Roof Area*) dan luas permukaan (*Surface Area*) yang memiliki korelasi negative yang cukup kuat untuk kedua variabel target yaitu -0.66 dan -0.86 untuk beban pemanasan, sedangkan untuk beban pendinginan sebesar -0.67 dan -0.86. Luas atap yang lebih besar dapat memungkinkan penggunaan energi panas matahari secara maksimal, yang dapat membantu mengurangi kebutuhan pemanasan dalam ruangan. Selain itu, luas atap yang besar juga dapat berperan dalam pengurangan hilangnya panas dari bangunan.

Pelatihan Model

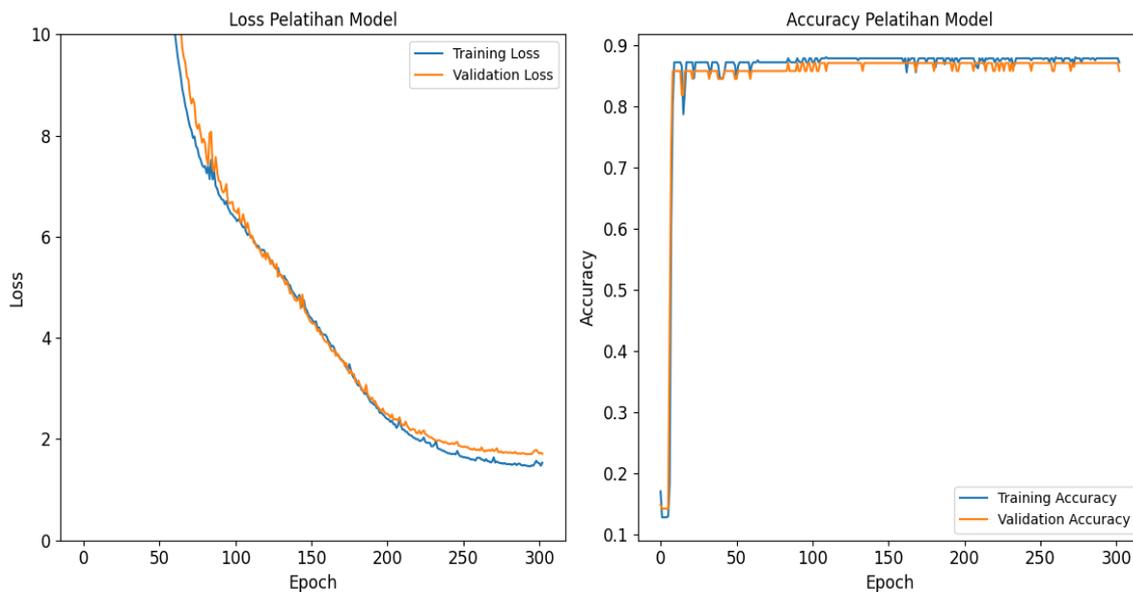
Setelah seleksi fitur dengan korelasi pearson, diperoleh enam variabel input yang akan digunakan dalam pelatihan model ANN untuk memprediksi beban pemanasan dan pendinginan. Model ANN dilatih untuk memprediksi kedua output

sekaligus, dengan variasi dataset yang dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Selama tahap pelatihan ini, dilakukan pencarian parameter terbaik untuk model ANN, seperti *optimizer*, *learning rate*, dan *batch size*, melalui penggunaan *grid search* parameter. Parameter terbaik adalah parameter yang mampu memperoleh nilai MAE dan RMSE terkecil dan nilai *R-squared* terbesar. Selain itu, pada tahap

pelatihan juga diterapkan metode *early stopping* untuk mengoptimalkan proses pelatihan model dengan cara yang lebih efisien. Penggunaan *early stopping* bertujuan untuk menghentikan pelatihan model jika terjadi *overfitting* atau tidak terdapat peningkatan pada nilai validation loss setelah beberapa iterasi. Hasil pencarian parameter terbaik menggunakan *grid search*, dapat dilihat pada table 2 dan kinerja model dapat dilihat pada Gambar 4.

Tabel 2. Hasil Pencarian Parameter Terbaik dengan *Grid Search*

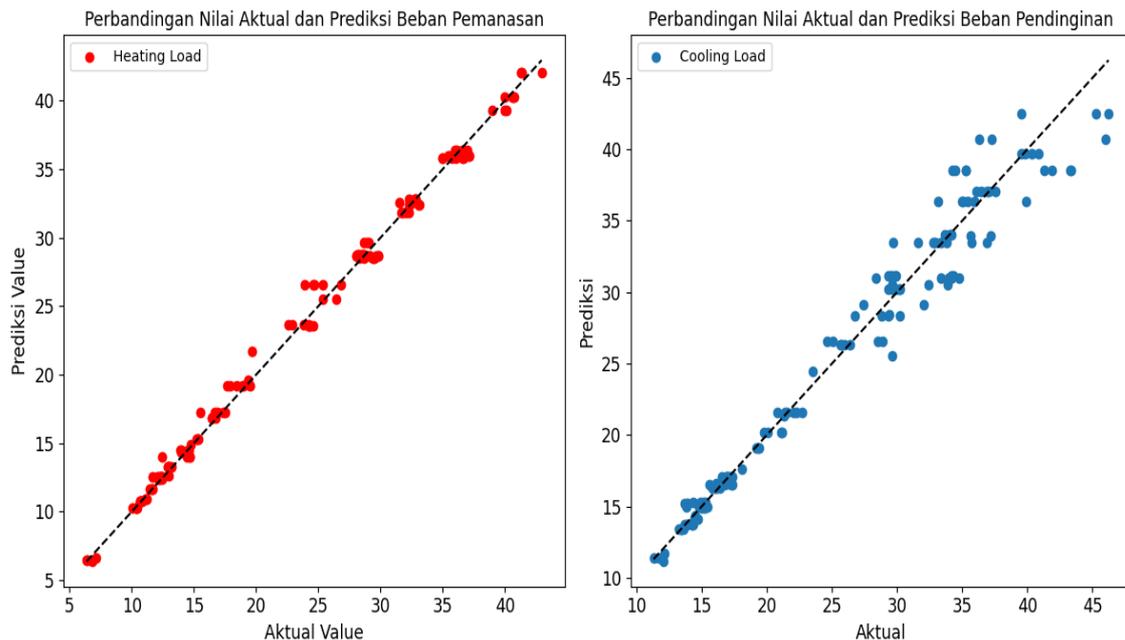
No	Parameter	Value
1	<i>Optimizer</i>	Adam
2	<i>Learning rate</i>	0.001
3	<i>Batch size</i>	32
4	<i>Epoch</i>	309



Gambar 5. *Loss* dan *Accuracy* Pelatihan Model

Berdasarkan hasil pencarian parameter terbaik dengan *grid search*, diperoleh optimizer terbaik adalah Adam, *learning rate* 0.001, *batch size* 32 dan *epoch* 309. Kinerja dari *setting-an* parameter model ini dapat diamati dengan melihat grafik *loss* saat pelatihan model, hal ini dapat membantu untuk memahami

terjadinya *overfitting* dan *underfitting* pada model prediksi. *Settingan* parameter terbaik ini, memperoleh nilai *loss* pelatihan sebesar 1.461 dan *loss* pengujian 1.733. Hasil prediksi dengan aktual dapat dilihat pada gambar 6 dan hasil metric evaluasi saat pelatihan dapat dilihat pada table 3.



Gambar 6. Perbandingan Hasil Prediksi Pada Pelatihan Model

Tabel 3. Metrik Evaluasi Pelatihan Model

Metrik evaluasi	Heating Load	Cooling Load
Mean Absolute Error (MAE)	0.457	1.163
Root Mean Squared Error (RMSE)	0.628	1.74
R-squared (R2)	0.996	0.967

Hasil prediksi menggunakan ANN menunjukkan hasil prediksi yang sangat baik, hal ini dapat dilihat dari hasil prediksi beban pemanasan yang memperoleh sebesar MAE 0.457, RMSE 0.68, dan R-squared 0.996, sedangkan hasil prediksi beban pendinginan diperoleh nilai MAE 1.163, RMSE 1.74 dan R-squared 0.967. Nilai R-squared yang mendekati 1 untuk prediksi beban pemanasan, menunjukkan bahwa model mampu dengan sangat baik menjelaskan variasi dalam data beban pemanasan. Sementara itu, nilai R-squared sekitar 0.967 untuk prediksi beban pendinginan juga menunjukkan performa yang baik, meskipun sedikit lebih rendah dibandingkan dengan prediksi beban pemanasan. Berdasarkan hasil keseluruhan dapat dilihat bahwa ANN mampu memprediksi beban pemanasan bangunan dengan lebih baik dari beban pendinginan.

Perbedaan dalam performa antara prediksi beban pemanasan dan beban pendinginan dapat dipengaruhi oleh beberapa faktor, seperti karakteristik fisik bangunan, musim iklim, atau sumber data yang digunakan [1], [6], [16], [17].

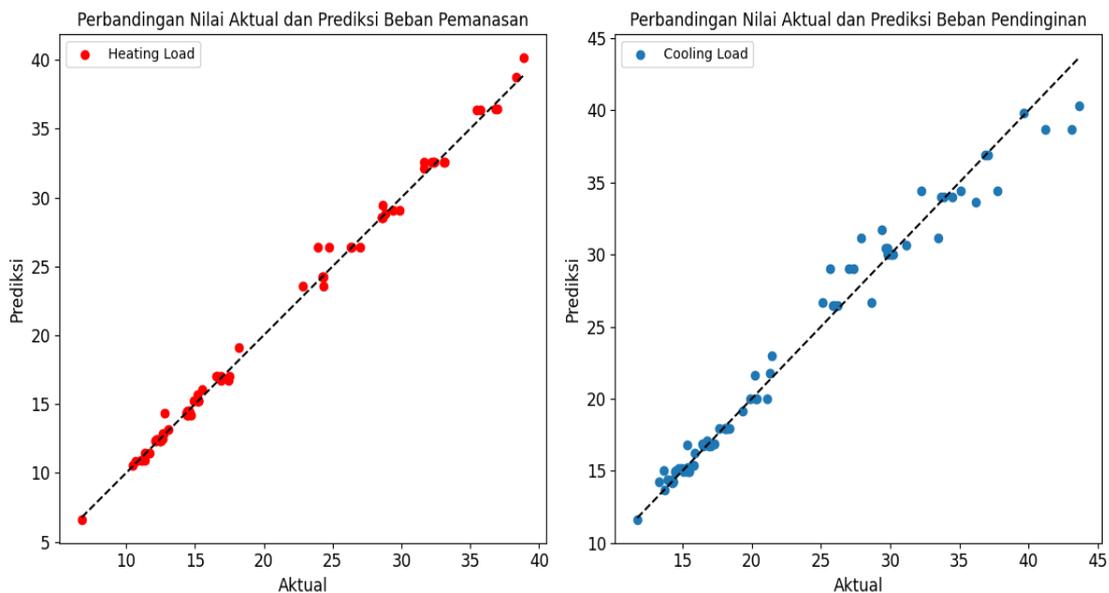
Evaluasi Model

Cross-validation adalah suatu metode yang sering digunakan untuk mengukur kinerja model machine learning dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Dalam penelitian ini, validasi model yang digunakan adalah k-fold cross-validation dengan k=10 untuk mengevaluasi keakuratan prediksi model ANN. Dataset akan dibagi menjadi 10 subset yang sama besar, dan model dilatih dan dievaluasi sebanyak 10 kali. Pada setiap iterasi, salah satu subset akan menjadi subset validasi, dan model akan

dilatih menggunakan 9 subset lainnya sebagai data latih. Pemilihan CV dengan jumlah $k=10$, tidak terlepas dari keseimbangan yang baik antara varians dan bias dalam estimasi performa model. Jika jumlah K terlalu kecil maka estimasi performa model cenderung memiliki varian yang tinggi, jumlah K 10 cukup besar untuk memberikan perkiraan yang stabil namun tidak terlalu besar sehingga masih efisien dalam penggunaan data [18]. Selain itu, jumlah K 10 juga membantu mengurangi risiko *overfitting* karena model dievaluasi pada berbagai subset data yang berbeda, sehingga lebih mungkin untuk mendeteksi pola yang lebih umum [19].

Hasil prediksi model ANN menggunakan CV dengan k 10 dapat dilihat

pada gambar 7 dan hasil metrik evaluasi menggunakan CV dengan k 10 dapat dilihat pada tabel 4. Hasil metrik evaluasi model menggunakan cross validation dengan k 10 menunjukkan bahwa model ANN memperoleh MAE sebesar 0.452, RMSE sebesar 0.65, dan R-squared sebesar 0.976 untuk prediksi beban pemanasan, sedangkan hasil evluasi beban pendinginan diperoleh nilai MAE 0.832, RMSE 1.32, dan R-squared 0.976. Hasil prediksi model menggunakan CV dengan k 10, tidak jauh berbeda dengan peforma model saat pelatihan. Hal ini menunjukkan bahwa model ANN yang digunakan memiliki performa yang stabil dan dapat diandalkan dalam melakukan prediksi pada dataset yang lebih besar dan beragam.



Gambar 7. Perbandingan Hasil Prediksi Model Menggunakan CV dengan Jumlah k 10

Tabel 4. Metrik Evaluasi Model Menggunakan CV dengan jumlah k 10

Metrik Evaluasi	Heating Load	Cooling Load
Mean Absolute Error (MAE)	0.452	0.832
Root Mean Squared Error (RMSE)	0.65	1.32
R-squared (R2)	0.994	0.976

KESIMPULAN

Berdasarkan hasil pemodelan prediksi beban pemanasan dan pendinginan bangunan menggunakan Artificial Neural Network (ANN) dapat disimpulkan bahwa korelasi Pearson dapat digunakan sebagai landasan dalam menentukan variabel input yang paling relevan dengan variabel target. Berdasarkan hasil korelasi pearson, diperoleh 6 fitur bangunan yang memiliki korelasi diatas nilai yang sudah ditentukan, baik untuk prediksi beban pemanasan maupun beban pendinginan, dan terdapat 2 fitur bangunan yang memiliki nilai korelasi dibawah nilai yang sudah ditentukan. Variabel yang memiliki nilai korelasi dibawah nilai yang sudah ditentukan dihilangkan dari dataset pelatihan model prediksi beban pemanasan dan pendinginan bangunan. Hasil pemodelan dengan seleksi fitur menggunakan korelasi Pearson memperoleh hasil prediksi yang sangat baik untuk kedua variabel target (beban pemanasan dan pendinginan). Hal ini dapat dilihat dari nilai metrik evaluasi yang cukup baik saat pelatihan model dan saat validasi menggunakan *Cross Validation* (CV) dengan k 10. Hasil penelitian ini menunjukkan bahwa seleksi fitur dengan korelasi pearson dapat digunakan untuk membantu dalam meningkatkan peforma prediksi machine learning menggunakan algoritma ANN.

DAFTAR PUSTAKA

- [1] J.-S. Chou and D.-K. Bui, "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design," *Energy Build.*, vol. 82, pp. 437–446, Oct. 2014, doi: 10.1016/j.enbuild.2014.07.036.
- [2] B. Chenari, J. Dias Carrilho, and M. Gameiro da Silva, "Towards sustainable, energy-efficient and healthy ventilation strategies in buildings: A review," *Renew. Sustain. Energy Rev.*, vol. 59, pp. 1426–1447, Jun. 2016, doi: 10.1016/j.rser.2016.01.074.
- [3] A. Saleh, "Design Of Energy-Efficient Buildings And Benefits Of Enhancing Reliance On Natural Lighting Through Maximizing The External Reflected Component," *J. Al-Azhar Univ. Eng. Sect.*, vol. 11, no. 41, pp. 1321–1325, Oct. 2016, doi: 10.21608/aej.2016.19312.
- [4] Watson, Boudreau, and Chen, "Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community," *MIS Q.*, vol. 34, no. 1, p. 23, 2010, doi: 10.2307/20721413.
- [5] B. Huang, X. Bai, Z. Zhou, Q. Cui, D. Zhu, and R. Hu, "Energy informatics: Fundamentals and standardization," *ICT Express*, vol. 3, no. 2, pp. 76–80, Jun. 2017, doi: 10.1016/j.icte.2017.05.006.
- [6] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy Build.*, vol. 49, pp. 560–567, Jun. 2012, doi: 10.1016/j.enbuild.2012.03.003.
- [7] Z. Zhang, "A gentle introduction to artificial neural networks," *Ann. Transl. Med.*, vol. 4, no. 19, pp. 370–370, Oct. 2016, doi: 10.21037/atm.2016.06.20.
- [8] V. V. Mokeev, "Prediction of Heating Load and Cooling Load of Buildings Using Neural Network," in *2019 International Ural Conference on Electrical Power Engineering*

- (*UralCon*), IEEE, Oct. 2019, pp. 417–421. doi: 10.1109/URALCON.2019.8877655.
- [9] D. Leni, M. Muchlisinalahuddin, M. Maimuzar, H. Haris, and H. Hendra, “Analisis Heatmap Korelasi dan Scatterplot untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Pelabelan AC efisiensi Energi,” *J. Rekayasa Mater. Manufaktur dan Energi*, vol. 6, no. 1, Mar. 2023, doi: 10.30596/rmme.v6i1.13133.
- [10] J. Hegde and B. Rokseth, “Applications of machine learning methods for engineering risk assessment – A review,” *Saf. Sci.*, vol. 122, p. 104492, Feb. 2020, doi: 10.1016/j.ssci.2019.09.015.
- [11] P. Ghosh *et al.*, “Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques,” *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [12] M. Ali and T. Aittokallio, “Machine learning and feature selection for drug response prediction in precision oncology applications,” *Biophys. Rev.*, vol. 11, no. 1, pp. 31–39, Feb. 2019, doi: 10.1007/s12551-018-0446-z.
- [13] A. Adriansyah, D. Leni, and R. Sumiati, “Comparative analysis of energy-efficient air conditioner based on brand,” *J. Polimesin*, vol. 21, no. 04, Aug. 2023, doi: 10.30811/jpl.v21i4.3625.
- [14] D. Leni, “Prediksi Pelabelan Rating AC Efisiensi Energi Menggunakan Pemodelan Machine Learning,” *J. Ilm. Momentum*, vol. 19, no. 1, p. 12, May 2023, doi: 10.36499/jim.v19i1.7832.
- [15] Y. Sugianela and T. Ahmad, “Pearson Correlation Attribute Evaluation-based Feature Selection for Intrusion Detection System,” in *2020 International Conference on Smart Technology and Applications (ICoSTA)*, IEEE, Feb. 2020, pp. 1–5. doi: 10.1109/ICoSTA48221.2020.1570613717.
- [16] E. Resch, R. A. Bohne, T. Kvamsdal, and J. Lohne, “Impact of Urban Density and Building Height on Energy Use in Cities,” *Energy Procedia*, vol. 96, pp. 800–814, Sep. 2016, doi: 10.1016/j.egypro.2016.09.142.
- [17] S. Papadopoulos, E. Azar, W.-L. Woon, and C. E. Kontokosta, “Evaluation of tree-based ensemble learning algorithms for building energy performance estimation,” *J. Build. Perform. Simul.*, vol. 11, no. 3, pp. 322–332, May 2018, doi: 10.1080/19401493.2017.1354919.
- [18] D. Leni, “Pemilihan Algoritma Machine Learning Yang Optimal Untuk Prediksi Sifat Mekanik Aluminium,” *J. EngineEnergi, Manufaktur, dan Mater.*, vol. 7, no. 1, pp. 35–44, 2023.
- [19] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamsirband, and A. Varkonyi-Koczy, “State of the Art of Machine Learning Models in Energy Systems, a Systematic Review,” *Energies*, vol. 12, no. 7, p. 1301, Apr. 2019, doi: 10.3390/en12071301.