

Application of Support Vector Machine Algorithm for Energy Consumption Prediction

Ivani Valentine ^{1*}, Joko Triloka ^{2*}, Ridho Sholehurrohman ^{3**}

* Fakultas Ilmu Komputer, Institut Informatika Dan Bisnis Darmajaya

**Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung

ivani.2421211001p@mail.darmajaya.ac.id¹, joko.triloka@darmajaya.ac.id², ridho.sholehurrohman@fmipa.unila.ac.id³

Article Info

Article history:

Received 2025-11-07

Revised 2026-01-16

Accepted 2026-02-27

Keyword:

Energy Consumption,
Energy Efficiency,
SMOTE,
SVM Kernel.

ABSTRACT

Global energy consumption continues to increase due to population growth, urbanization, and rapid technological advancement, creating significant challenges for effective energy management. This study proposes an energy consumption classification model using Support Vector Machine (SVM) to identify consumption levels based on environmental, temporal, and operational features. Data preprocessing includes feature normalization using StandardScaler and class balancing with SMOTE to improve model stability. Experimental results show that the SVM with a linear kernel consistently achieves the best performance on the test data, outperforming more complex kernel configurations. The proposed model attains an average accuracy of 88.15%, precision of 91.08%, recall of 88.16%, and an F1-score of 87.54%. Feature analysis indicates that temperature and HVAC usage are the most influential factors in determining energy consumption levels. These findings demonstrate that selecting a kernel aligned with data characteristics is more critical than increasing model complexity. The proposed SVM-based approach has strong potential as a decision-support tool for building energy management, enabling better identification of consumption conditions and supporting energy efficiency strategies.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Energi merupakan kebutuhan mendasar yang mendukung hampir seluruh aktivitas manusia, baik dalam skala rumah tangga maupun sektor industri [1-2]. Namun, dengan pertumbuhan populasi, urbanisasi yang pesat, serta peningkatan penggunaan teknologi, konsumsi energi global terus meningkat secara signifikan [2-3]. Kondisi ini menciptakan tantangan besar bagi pengelolaan energi yang efisien, terutama dalam menjaga ketersediaan pasokan energi yang berkelanjutan di tengah tekanan lingkungan [2]. Selain itu, penggunaan energi yang tidak efisien dapat menyebabkan pemborosan sumber daya dan peningkatan emisi karbon, yang berkontribusi terhadap perubahan iklim global [1-3].

Salah satu tantangan utama dalam pengelolaan energi adalah prediksi konsumsi energi secara akurat. Kebutuhan energi dipengaruhi oleh berbagai faktor, seperti kondisi cuaca, pola aktivitas masyarakat, dan tren ekonomi, yang sering kali bersifat non-linear dan kompleks [4-5].

Pendekatan tradisional, seperti regresi linier, sering kali kurang mampu menangkap dinamika ini secara menyeluruh [5]. Akibatnya, pengelola energi menghadapi kesulitan dalam mengantisipasi kebutuhan energi, mengoptimalkan alokasi sumber daya, dan merancang kebijakan berbasis data yang tepat [3, 5-6].

Berbagai pendekatan berbasis pembelajaran mesin telah diterapkan untuk memprediksi pola konsumsi energi [4-8]. Misalnya, algoritma seperti Regresi Linier digunakan untuk memodelkan hubungan antar variabel, tetapi memiliki keterbatasan dalam menangani hubungan non-linear [7-8]. Sementara itu, Jaringan Saraf Tiruan (ANN) dan Random Forest (RF) menunjukkan performa yang baik dalam analisis data yang lebih kompleks [8]. Namun, ANN sering membutuhkan data dalam jumlah besar untuk menghasilkan model yang akurat, dan RF cenderung kurang efisien dalam menangani data berdimensi tinggi [6-8].

Sementara pendekatan ini cukup menjanjikan, penelitian terkait pemanfaatan Support Vector Machine (SVM) untuk

prediksi konsumsi energi masih relatif jarang dilakukan [9-10]. SVM memiliki keunggulan dalam menangani data berdimensi tinggi, mendeteksi pola non-linear, dan menghasilkan model yang efisien meskipun data yang tersedia tidak terlalu besar [9, 11]. Oleh karena itu, eksplorasi penggunaan SVM untuk prediksi energi berpotensi menjadi solusi yang lebih optimal [9-11]. Penelitian yang berjudul "Prediction of energy consumption in hotel buildings via support vector machines" membahas pengembangan model prediksi konsumsi energi pada bangunan hotel menggunakan metode Support Vector Machine (SVM). Penelitian ini menggunakan parameter cuaca dan parameter operasi sistem pendingin udara hotel sebagai input model. Proses analisis mencakup penentuan interval distribusi normal untuk mengeliminasi outlier, sehingga meningkatkan stabilitas prediksi. Dengan kernel RBF yang dioptimalkan, model menunjukkan akurasi tinggi dengan MSE 2,22% dan R^2 sebesar 0,94 [9].

Penelitian selanjutnya yang berjudul "Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm" mengembangkan metode prediksi konsumsi energi listrik jangka menengah menggunakan algoritma hibrida GA-SVM. Metode ini mengoptimalkan parameter SVM, seperti penalty factor (C) dan kernel parameter (γ), menggunakan algoritma genetika untuk meningkatkan akurasi prediksi. Dengan stratified sampling untuk membagi data pelatihan dan pengujian, model ini mencapai MAPE 3,05% dan R^2 0,95 pada Pipeline 1, menunjukkan akurasi tinggi dibandingkan metode lain seperti PSO-SVM, GA-BPNN, dan DT. Model ini unggul dalam menangani data kecil, mengurangi pengaruh bias, dan memastikan kestabilan prediksi, sehingga bermanfaat dalam perencanaan energi dan pemantauan operasional pipa minyak mentah [12].

Penelitian lainnya berjudul "Analyzing the Relationship Between Meteorological Parameters and Electric Energy Consumption Using Support Vector Machine and Cooling Degree Days Algorithm" menjelaskan tentang hubungan antara parameter meteorologi dan konsumsi listrik, dengan fokus pada suhu, kelembapan, dan kecepatan angin. Dengan menggunakan kumpulan data dari BMKG dan PLN, para peneliti menerapkan metodologi CRISP-DM, algoritma Cooling Degree Days (CDD), dan model Support Vector Machine (SVM). CDD mengukur korelasi antara parameter cuaca dan penggunaan listrik selama musim dingin, sementara SVM memperkirakan konsumsi listrik dari Januari hingga April 2024, mencapai metrik akurasi tinggi: presisi (0,796), recall (0,793), skor F1 (0,793), MAPE (17,2%), RMSE (0,41), MAE (0,167), dan akurasi (0,98). Hasilnya menyoroti tren listrik yang dipengaruhi oleh variasi cuaca, dengan konsumsi rata-rata diperkirakan sebesar 20,9 Watt per hari [13].

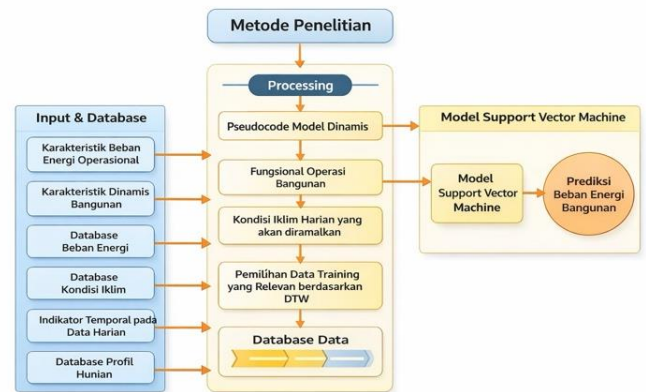
Penelitian ini bertujuan untuk mengembangkan model prediktif berbasis Support Vector Machine (SVM) yang dapat digunakan untuk memproyeksikan konsumsi energi secara akurat. Model ini memanfaatkan kemampuan SVM dalam

mendeteksi pola-pola kompleks dan non-linear untuk menganalisis hubungan antara variabel seperti suhu, waktu, dan aktivitas manusia dengan kebutuhan energi. Selain itu, hasil prediksi ini akan digunakan untuk memberikan rekomendasi yang mendukung optimalisasi pengelolaan energi dan mengurangi pemborosan.

II. METODE PENELITIAN

A. Tahapan Penelitian

Studi ini dilaksanakan secara terstruktur melalui serangkaian tahapan untuk memperoleh hasil prediksi konsumsi energi. Tahapan penelitian ditunjukkan pada Gambar 1.

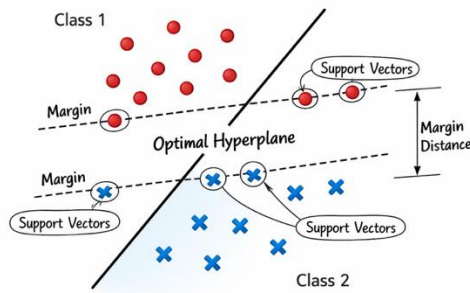


Gambar 1. Tahapan penelitian prediksi konsumsi energi

B. Support Vector Machine

Support Vector Machine (SVM), dikembangkan oleh Vapnik et al, di laboratorium AT&T Bell [14], [15]. SVM merupakan metode supervised learning yang sangat andal untuk menangani tugas klasifikasi dan regresi [12-15]. SVM menggunakan kombinasi linier dari fungsi kernel untuk menghasilkan prediksi. Pendekatan SVM didasarkan pada prinsip minimisasi risiko struktural Structural Risk Minimization (SRM), yang memprioritaskan pengurangan batas atas kesalahan generalisasi daripada hanya fokus pada kesalahan empiris [9], [15]. [16]. Dalam prosesnya, SVM memanfaatkan fungsi kernel untuk memetakan data ke ruang berdimensi tinggi, memungkinkan pemisahan linear dalam klasifikasi atau pencarian hiperbidang optimal yang memaksimalkan margin antar kelompok data pada masalah regresi [16].

Gambar 2 memperlihatkan bahwa Optimal hyperplane dalam SVM yang digunakan untuk memisahkan dua dataset, dengan vektor yang berada di dekat hyperplane disebut sebagai Support Vectors (SV). Keakuratan model SVM sangat dipengaruhi oleh pemilihan parameter kernel [16], karena parameter tersebut memiliki pengaruh besar terhadap kinerja metode kernel. Jumlah parameter ini ditentukan oleh margin yang memisahkan dataset [15], [16].



Gambar 2. Optimal hyperplane pada support vector machine (SVM) [16]

Dalam definisi regresi, SVM memerlukan beberapa parameter input untuk eksplorasi [16]. Misalkan parameter input adalah serangkaian titik data $(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots (x_n, y_n)$, di mana x_i adalah parameter input normal yang dimana i mewakili satu sampel input dan y_i adalah parameter konsumsi energi didalam sampel input i . Jumlah total sampel adalah 1. Oleh karena itu, SVM memperkirakan fungsi tersebut menggunakan bentuk berikut.

$$f(x) = \omega \cdot \phi(x) + b \tag{1}$$

dimana $\phi(x)$ merupakan representasi ruang fitur berdimensi tinggi yang dipetakan secara non-linier dari ruang berdimensi rendah x . Selanjutnya, koefisien ω dan b digunakan untuk meminimalkan fungsi risiko [9, 15-16].

$$\frac{1}{2} \|\omega\|^2 + C \frac{1}{n} \sum_{i=1}^n L_\epsilon(y_i, f(x_i)) \tag{2}$$

Dimana $\|\omega\|^2$ merupakan koefisien regularisasi [16]. Dengan meminimalkan $\|\omega\|^2$ kita dapat membuat fungsi sedatar mungkin, sehingga dapat mengendalikn fungsi kapasitas fungsional. Selanjutnya, fungsi $\frac{1}{n} \sum_{i=1}^n L_\epsilon(y_i, f(x_i))$ merupakan adalah Error empiris yang diukur dengan loss-function ϵ , yang didefinisikan sebagai berikut.

$$L_\epsilon(y_i, f(x_i)) = \begin{cases} 0, & y_i - f(x_i) \leq \epsilon \\ |y_i - f(x_i)| - \epsilon, & \text{otherwise} \end{cases} \tag{3}$$

Definisi tersebut merupakan rentang nilai ϵ dimana jika nilai prediksi berada dalam rentang tersebut, kerugiannya adalah nol, dan jika titik prediksi berada di luar rentang tersebut, maka kerugiannya adalah besarnya selisih antara nilai prediksi dan jarak ϵ dari rentang tersebut. C disebut konstanta regularisasi, yang juga dikenal sebagai faktor penalti. Semuanya ditentukan oleh pengguna kecuali default dari sistem [16].

Selanjutnya dikenalkan variabel slack ζ_i dan ζ_i^* yang digunakan untuk membentuk batasan pada permasalahan sebagai berikut.

Meminimalkan:

$$R(\omega, \xi_i^*) = \frac{1}{2} \|\omega\|^2 + C \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{4}$$

Subjek ke:

$$\begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{5}$$

C adalah konstanta terregulasi yang lebih besar dari 0 untuk melakukan penyeimbangan antara kesalahan training dan flat model. C merupakan penalti untuk kesalahan prediksi yang lebih besar dari ϵ . ξ_i dan ξ_i^* adalah slack variabel yang membentuk jarak dari nilai aktual ke nilai batas ϵ yang sesuai. Tujuan SVM adalah meminimalkan i, ξ_i^* dan ω^2 [9,15-16].

Optimasi dengan kendala di atas dapat diubah dengan menggunakan perkalian Lagrangian dibuah menjadi permasalahan pemrograman kuadrat. Oleh karena itu, bentuk penyelesaiannya dapat diberikan oleh persamaan berikut [9, 15-16]:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{6}$$

Dimana α_i dan α_i^* adalah pengali Lagrange yang menjadi subjek pada batasan berikut.

$$\begin{aligned} & \sum_{i=1}^n (\alpha_i - \alpha_i^*) \\ & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n \\ & 0 \leq \alpha_i^* \leq C \quad i = 1, \dots, n \end{aligned}$$

$K(\cdot)$ adalah fungsi kernel dan nilainya adalah produk dalam dari dua vektor x_i dan x_j dalam ruang fitur $\phi(x_i)$ dan $\phi(x_j)$ dan memenuhi kondisi sesuai dengan parameter.

Maka,

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

C. Kernel SVM

Algoritma support vector machine (SVM) adalah teknik pembelajaran mesin terbimbing untuk klasifikasi dengan menggunakan fungsi kernel dalam proses mendapatkan hasil dengan cepat, dan akurat [13-16]. Secara umum, terdapat empat fungsi kernel yang sering digunakan pada algoritma SVM adalah sebagai berikut [15-16].

Pertama, kernel linear yang merupakan kernel dasar atau paling sederhana. Kernel ini biasanya digunakan untuk klasifikasi linear. Selanjutnya, kernel polynomial adalah kernel yang menggunakan derajat polinomial sebagai parameter dalam menentukan kompleksitas batas keputusan. Kemudian kernel radial basis function (RBF) merupakan salah satu fungsi kernel yang paling populer dan banyak

digunakan untuk SVM [15-16]. Kernel RBF merepresentasikan hubungan yang kompleks dan non-linier antar dataset, sehingga memiliki batas keputusan yang fleksibel dengan pola melingkar. Terakhir, Kernel Sigmoid adalah kernel yang dikembangkan berdasarkan prinsip jaringan saraf tiruan.

D. Confusion Matrix

Confusion Matrix adalah sebuah matrix yang dapat digunakan untuk melakukan evaluasi pada model klasifikasi dengan cara melakukan prediksi pada objek apakah benar atau salah [17]. Confusion Matrix berisikan informasi mengenai klasifikasi aktual dan presisi yang dilakukan model [17-18].

TABEL I
CONFUSION MATRIX

Aktual Kelas	Prediksi Kelas	
	Positif	Negatif
Positif	True Positive	False Negative
Negatif	False Positive	True Negative

Tabel 1 menunjukkan hubungan antara hasil prediksi model dan nilai sebenarnya (ground truth) dari data uji [18]. Confusion Matrix terdiri dari empat elemen utama, yaitu True Positive (TP), yang menunjukkan jumlah kasus di mana model memprediksi kelas positif dengan benar; False Positive (FP), yaitu jumlah kasus di mana model salah memprediksi kelas positif padahal seharusnya negatif; True Negative (TN), yang merepresentasikan jumlah kasus di mana model dengan benar memprediksi kelas negatif; dan False Negative (FN), yaitu jumlah kasus di mana model salah memprediksi kelas negatif padahal seharusnya positif [17-18].

Selanjutnya berbagai metrik evaluasi seperti *accuracy*, presisi, *recall*, dan F1-score dapat dihitung untuk memberikan penilaian menyeluruh terhadap performa model sebagai berikut.

1) Accuracy

Accuracy mengukur seberapa baik model SVM memprediksi suatu data dengan tepat.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2) Precision

Precision adalah pengukuran yang digunakan untuk memprediksi contoh-contoh kelas yang sama di antara contoh-contoh data. Presisi dihitung sebagai proporsi antara prediksi yang benar-benar positif dan kumpulan semua nilai positif yang nyata.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

3) Recall

Recall digunakan untuk mengukur kemampuan model dengan persentase jumlah data aktual yang berhasil diprediksi dengan benar.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

4) F1-Score

F1-Score adalah metrik yang digunakan untuk mengevaluasi keseimbangan antara precision dan recall dalam sebuah model klasifikasi

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

III. HASIL DAN PEMBAHASAN

A. Dataset

Dataset pada penelitian ini terdapat pada <https://www.kaggle.com/datasets/ajinilpatel/energy-consumption-prediction/data>. Dataset yang digunakan adalah catatan konsumsi energi sebuah bangunan yang terdiri dari 12 kolom fitur utama dengan 5000 baris data sampel. Detail dataset berdasarkan variabel temporal (bulan, jam, hari, dan status libur), lingkungan (suhu dan kelembapan), karakteristik fisik (luas area dan jumlah penghuni), serta operasional (penggunaan HVAC, lampu, dan energi terbarukan). Deskripsi dataset sebagaimana ditunjukkan pada Gambar 3.

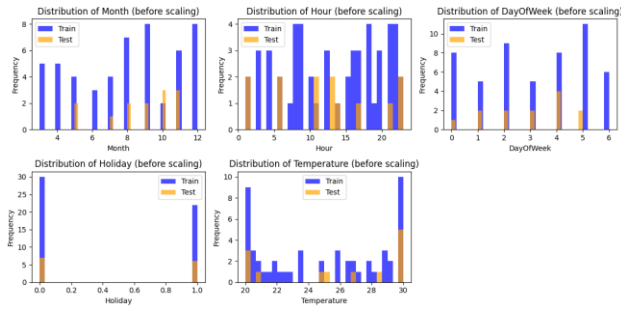
Gambar 3. Deskripsi Dataset

Pada penelitian ini, prediksi konsumsi energi diformulasikan sebagai permasalahan klasifikasi, di mana nilai konsumsi energi yang bersifat kontinu ditransformasikan ke dalam beberapa kelas tingkat konsumsi, seperti konsumsi rendah, sedang, dan tinggi, berdasarkan ambang batas tertentu. Pendekatan klasifikasi dipilih untuk mendukung kebutuhan pengambilan keputusan operasional yang bersifat kategorikal, misalnya dalam menentukan prioritas pengelolaan energi, pengendalian beban, dan perencanaan operasional bangunan. Dengan formulasi ini, penggunaan metrik evaluasi klasifikasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* menjadi relevan dan sesuai dengan tujuan penelitian.

B. Preprocessing Data

Sebelum data digunakan untuk training dan testing, data dilakukan proses preproesing yang meliputi normalisasi menggunakan StandardScaler untuk memastikan kontribusi fitur yang seimbang terhadap model. Selain itu, apabila distribusi kelas target tidak merata, metode SMOTE

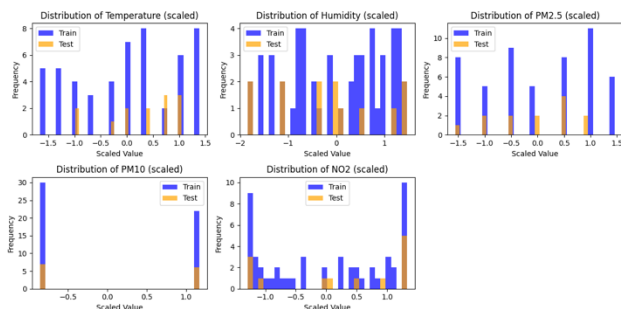
(Synthetic Minority Oversampling Technique) diterapkan untuk meningkatkan proporsi kelas minoritas. Visualisasi dataset pertama sebelum proses preprocessing data ditunjukkan pada Gambar 4 yang memperlihatkan adanya perbedaan distribusi antara data pelatihan dan data pengujian pada beberapa fitur, khususnya fitur temporal dan fitur Holiday yang menunjukkan ketidakseimbangan kelas.



Gambar 4. Visualisasi data pertama sebelum proses preprocessing

Visualisasi data pada Gambar 4 menunjukkan pola distribusi data berdasarkan fitur temporal dan lingkungan sebelum dilakukan normalisasi. Pada fitur Month, data pelatihan memiliki distribusi yang lebih merata dibandingkan data pengujian, yang cenderung terkonsentrasi pada beberapa bulan tertentu. Distribusi jam (Hour) memperlihatkan puncak pada jam-jam tertentu, menunjukkan potensi waktu puncak aktivitas, dengan data pengujian yang tetap representatif terhadap pola umum data pelatihan. Pola distribusi hari dalam seminggu (DayofWeek) terlihat cukup seragam antara data pelatihan dan pengujian. Fitur Holiday menunjukkan distribusi yang sangat tidak seimbang, dengan mayoritas data berasal dari hari kerja, baik untuk pelatihan maupun pengujian. Distribusi suhu (Temperature) menunjukkan rentang yang relatif sempit dengan beberapa puncak tertentu, yang mencerminkan kondisi lingkungan.

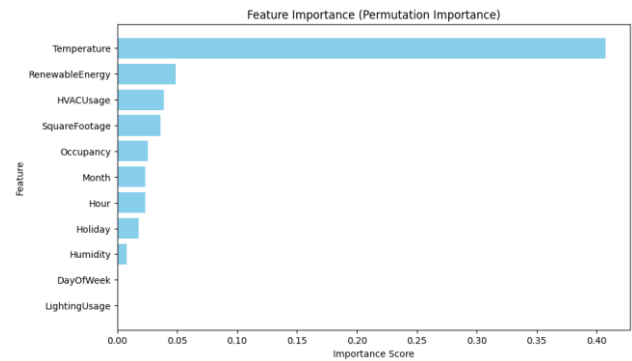
Selanjutnya, diberikan visualisasi dataset pertama hasil proses preprocessing data sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Visualisasi hasil preprocessing

Visualisasi data hasil proses preprocessing menunjukkan pola yang lebih seragam antara data pelatihan dan pengujian untuk semua fitur. Suhu (Temperature) dan

kelembapan (Humidity) memiliki distribusi yang lebih tersebar di berbagai nilai skala, sementara PM2.5 dan NO2 menunjukkan puncak pada nilai tertentu. Sebaliknya, PM10 cenderung terkonsentrasi pada dua titik ekstrem, menunjukkan distribusi yang lebih terfokus. Keseragaman distribusi ini memastikan bahwa data pengujian tetap representatif terhadap pola data pelatihan setelah normalisasi. Penggunaan SMOTE dilakukan setelah proses diskretisasi variabel target, sehingga oversampling diterapkan pada label kelas konsumsi energi. Pendekatan ini bertujuan untuk mengurangi bias model terhadap kelas mayoritas tanpa mengubah struktur numerik data asli.



Gambar 6. Visualisasi data hasil feature importance

Visualisasi feature importance sebagaimana ditunjukkan pada Gambar 6 menunjukkan bahwa temperature adalah fitur yang paling dominan dalam memengaruhi prediksi model, dengan skor jauh lebih tinggi dibanding fitur lainnya. Fitur ini sangat relevan dalam kasus seperti pengelolaan energi, di mana suhu memengaruhi penggunaan sistem HVAC. Fitur lainnya, seperti RenewableEnergy dan HVACUsage, juga memiliki pengaruh signifikan, diikuti oleh SquareFootage dan Occupancy, yang mencerminkan dampak luas bangunan dan jumlah penghuni terhadap kebutuhan energi.

Fitur berbasis waktu, seperti Month dan Hour, memiliki kontribusi moderat, sedangkan DayOfWeek dan Holiday menunjukkan dampak yang lebih kecil. Humidity dan LightingUsage memiliki pengaruh paling rendah, menunjukkan bahwa variabilitas atau relevansinya terhadap output model tidak signifikan. Temuan ini memberikan implikasi praktis bagi pengelola energi, bahwa strategi efisiensi energi dapat difokuskan pada pengendalian suhu dan optimasi penggunaan HVAC sebagai prioritas utama. Namun demikian, perlu dicatat bahwa model SVM dengan kernel non-linear bersifat kurang transparan, sehingga feature importance digunakan sebagai indikator relatif untuk mendukung interpretasi hasil, bukan sebagai penjelasan kausal langsung.

Selanjutnya, data hasil preprocessing dibagi menjadi 80% data training dan 20% data testing. Tujuannya adalah untuk memastikan model dapat dilatih dengan data yang cukup untuk mengenali pola atau hubungan dalam dataset sekaligus diuji dengan data yang belum pernah dilihat model

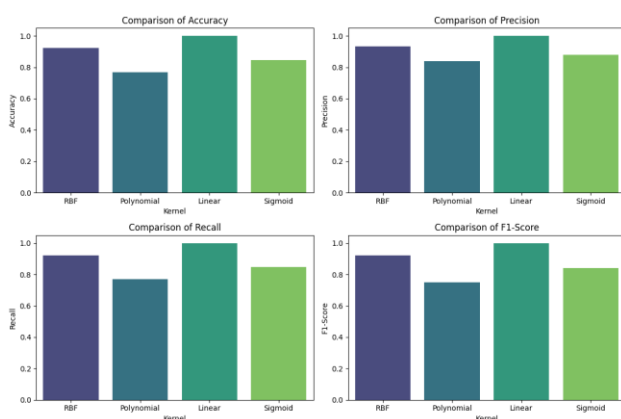
sebelumnya. Pembagian ini bertujuan untuk mengukur kemampuan generalisasi model, yaitu seberapa baik model memberikan prediksi pada data yang tidak digunakan selama proses pelatihan, sehingga berfungsi sebagai skema validasi hold-out dalam evaluasi performa model. Dalam training data untuk membangun model SVM, penulis menggunakan beberapa kernel, diantaranya kernel linear, kernel polynomial, kernel RBF (*Radial Basis Function*), dan kernel sigmoid, untuk mengeksplorasi performa masing-masing kernel dalam memprediksi target.

Dalam training dataset, algoritma SVM dirancang untuk membangun hyperplane yang secara efektif memisahkan titik data dari berbagai kelas. Proses tersebut melibatkan pemetaan data masukan ke ruang fitur berdimensi lebih tinggi untuk meningkatkan kemampuan pemisahan data. Penyesuaian parameter (hyperparameter tuning) juga dilakukan untuk setiap kernel guna meningkatkan akurasi dan efisiensi model.

C. Prediksi Konsumsi Energi

Dalam tahapan prediksi konsumsi energi sebagaimana ditunjukkan pada Gambar 2, model yang telah dibangun pada tahap sebelumnya kemudian diuji menggunakan data uji untuk menghasilkan prediksi kelas konsumsi energi. Evaluasi performa model dilakukan menggunakan confusion matrix sehingga diperoleh metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan F1-score.

Perbandingan performa model Support Vector Machine (SVM) dengan berbagai jenis kernel ditunjukkan pada Gambar 7. Berdasarkan gambar tersebut, kernel Linear menunjukkan performa terbaik pada seluruh metrik evaluasi. Kernel ini mencapai nilai *accuracy* sebesar 98,77%, *precision* 99,17%, *recall* 98,80%, dan F1-score 98,79%, yang menunjukkan kemampuan model dalam meminimalkan kesalahan prediksi positif maupun negatif secara seimbang. Hasil ini mengindikasikan bahwa kernel Linear sangat andal dalam memprediksi kelas konsumsi energi pada dataset yang digunakan.



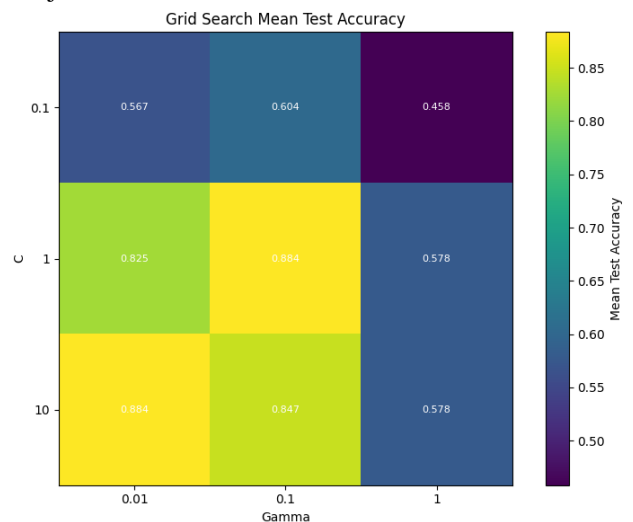
Gambar 7. Performa Metrik Kernel RBF, Polynomial, Linear, dan Sigmoid

Selanjutnya, Kernel RBF juga menunjukkan performa yang baik dengan *accuracy* sebesar 92,31%, *precision*

93,27%, dan *recall* 92,31%, meskipun masih berada di bawah kernel Linear. Sementara itu, kernel Sigmoid dan Polynomial memiliki performa yang lebih rendah. Kernel Sigmoid mencatat *accuracy* sebesar 84,62% dengan F1-score 84,04%, sedangkan kernel Polynomial menunjukkan performa terendah dengan *accuracy* sebesar 76,92% dan F1-score 75,11%. Rendahnya performa kedua kernel tersebut menunjukkan bahwa batas keputusan yang terlalu kompleks kurang sesuai dengan karakteristik data konsumsi energi dalam penelitian ini. Secara keseluruhan, hasil pada Gambar 7 menegaskan bahwa kernel Linear merupakan konfigurasi terbaik berdasarkan evaluasi langsung pada data uji.

Secara agregat, algoritma SVM menunjukkan performa yang stabil dengan rata-rata *accuracy* sebesar 88,15%, *precision* 91,08%, *recall* 88,16%, dan F1-score 87,54%. Nilai rata-rata ini mencerminkan konsistensi kinerja SVM pada berbagai konfigurasi kernel, dengan kernel Linear sebagai konfigurasi yang memberikan performa paling optimal.

Selanjutnya, dilakukan penyesuaian hyperparameter menggunakan metode grid search untuk menganalisis pengaruh parameter C dan Gamma terhadap performa model, khususnya pada kernel RBF. Hasil grid search mean test *accuracy* ditunjukkan pada Gambar 8, yang memperlihatkan bahwa nilai Gamma yang lebih kecil (0.01–0.1) menghasilkan akurasi yang lebih tinggi dibandingkan nilai Gamma yang besar. Selain itu, peningkatan nilai C hingga 1 dan 10 cenderung meningkatkan performa model. Berdasarkan Gambar 8, heatmap menggambarkan rata-rata akurasi pengujian model berdasarkan variasi dua parameter, yaitu C dan Gamma, yang dioptimalkan menggunakan metode grid search. Warna dalam grafik merepresentasikan tingkat akurasi, di mana warna terang (kuning) menunjukkan akurasi tinggi, sementara warna gelap (biru hingga ungu) menunjukkan akurasi rendah.



Gambar 8. Grid Search Mean Test Accuracy dan Gamma

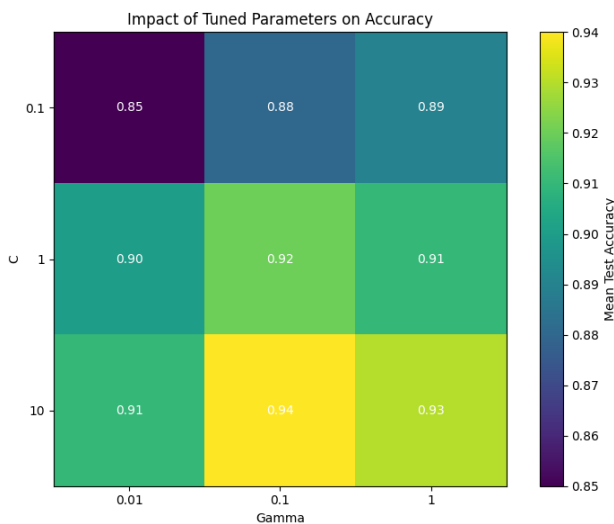
Selain itu, nilai C yang lebih besar, seperti 1 dan 10, cenderung menghasilkan akurasi lebih baik dibandingkan nilai C yang lebih kecil, seperti 0.1. Kombinasi parameter

terbaik tercapai pada $C = 1$ atau $C = 10$ dengan $\text{Gamma} = 0.1$, keduanya memberikan akurasi tertinggi sebesar 0.884. Dengan demikian, parameter optimal yang direkomendasikan untuk model adalah $C = 1$ dan $\text{Gamma} = 0.1$, karena kombinasi ini memberikan performa terbaik secara konsisten.

Hasil ini diperkuat oleh Grafik heatmap Gambar 9 menunjukkan pengaruh parameter yang dituning, yaitu C dan Gamma , terhadap rata-rata akurasi pengujian model. Warna pada grafik merepresentasikan tingkat akurasi, dengan warna terang (kuning) menunjukkan akurasi tertinggi, sedangkan warna gelap (ungu) menunjukkan akurasi terendah. Berdasarkan grafik, kombinasi parameter C yang lebih tinggi (10) dan Gamma yang sedang (0.1) menghasilkan akurasi tertinggi sebesar 0.94.

Secara umum, peningkatan nilai C cenderung meningkatkan akurasi, terutama pada nilai Gamma kecil hingga sedang. Sebaliknya, nilai Gamma besar (1) menunjukkan akurasi yang lebih rendah dibandingkan nilai Gamma kecil (0.01 atau 0.1) untuk sebagian besar kombinasi C . Dengan demikian, parameter optimal untuk model adalah $C = 10$ dan $\text{Gamma} = 0.1$, yang memberikan performa terbaik dengan akurasi maksimal. Kombinasi ini disarankan untuk digunakan dalam implementasi model untuk hasil prediksi yang lebih akurat.

Dalam konteks pengelolaan energi bangunan, performa model dengan tingkat akurasi yang tinggi serta keseimbangan antara precision dan recall menunjukkan bahwa algoritma SVM, khususnya dengan kernel Linear, memiliki potensi yang baik untuk digunakan sebagai alat bantu pengambilan keputusan operasional, seperti identifikasi kondisi konsumsi energi tinggi dan perencanaan strategi efisiensi energi.



Gambar 9. Impact of Tuned Parameter on Accuracy dan Gamma

D. Pembahasan

Berdasarkan hasil eksperimen menggunakan Support Vector Machine (SVM) dengan berbagai kernel, yaitu linear, polynomial, RBF dan sigmoid, hasil evaluasi menunjukkan bahwa kernel Linear memberikan performa terbaik dengan

akurasi 98,77%, precision 99,17%, recall 98,80%, dan F1-score 98,79%. Kernel ini unggul dalam memberikan prediksi akurat dengan keseimbangan yang baik antara kesalahan positif dan negatif. Selanjutnya, analisis tuning parameter C dan Gamma menggunakan metode grid search menunjukkan bahwa kombinasi parameter C yang lebih tinggi dan Gamma yang lebih kecil menghasilkan akurasi terbaik. Pada Gambar 8, heatmap menunjukkan bahwa parameter optimal adalah $C = 1$ atau $C = 10$ dan $\text{Gamma} = 0.1$, dengan akurasi tertinggi sebesar 88,4%. Sementara itu, Gambar 9 mengonfirmasi bahwa kombinasi $C = 10$ dan $\text{Gamma} = 0.1$ memberikan akurasi maksimal sebesar 94%. Parameter C yang lebih tinggi meningkatkan kemampuan model dalam menangani margin kecil, sedangkan Gamma kecil meningkatkan kemampuan generalisasi model.

Secara keseluruhan, algoritma SVM menunjukkan kinerja yang sangat baik dengan rata-rata akurasi sebesar 88,15%, precision 91,08%, recall 88,16%, dan F1-score 87,54%. Hasil ini menggarisbawahi pentingnya memilih kernel dan parameter optimal untuk memastikan performa model yang maksimal. Kombinasi $C = 10$ dan $\text{Gamma} = 0.1$ direkomendasikan sebagai konfigurasi terbaik untuk implementasi lebih lanjut.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) mampu memberikan performa yang sangat baik dalam memprediksi kelas konsumsi energi bangunan. Penerapan tahapan preprocessing data, yang mencakup normalisasi fitur dan penyeimbangan kelas target, berkontribusi dalam meningkatkan kestabilan model serta memastikan setiap fitur memiliki peran yang seimbang dalam proses pembelajaran.

Hasil analisis menunjukkan bahwa suhu (temperature) merupakan faktor paling dominan dalam memengaruhi prediksi konsumsi energi, diikuti oleh penggunaan HVAC dan energi terbarukan, sementara fitur berbasis waktu memberikan kontribusi yang lebih moderat. Temuan ini menegaskan bahwa faktor lingkungan dan operasional memiliki peran penting dalam menentukan tingkat konsumsi energi bangunan.

Berdasarkan evaluasi terhadap berbagai kernel SVM, kernel Linear secara konsisten menunjukkan performa terbaik pada data uji, dengan tingkat akurasi yang tinggi serta keseimbangan yang baik antara precision dan recall. Meskipun kernel RBF dengan parameter optimal hasil tuning juga menunjukkan performa yang tinggi, kernel Linear terbukti lebih stabil dan efisien tanpa memerlukan penyesuaian parameter yang kompleks.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa pemilihan kernel yang sesuai dengan karakteristik data memiliki pengaruh yang lebih besar terhadap performa model dibandingkan peningkatan kompleksitas model itu sendiri. Model SVM dengan kernel Linear memiliki potensi yang kuat untuk diterapkan sebagai alat bantu pengambilan keputusan

dalam pengelolaan energi bangunan, khususnya dalam identifikasi kondisi konsumsi energi dan perencanaan strategi efisiensi energi.

DAFTAR PUSTAKA

- [1] Iea, Irena, UNSD, World Bank, and WHO, *Tracking SDG 7: The Energy Progress Report*, Washington, DC, USA: World Bank, 2023.
- [2] J. S. Puthumana, L. M. Ngaage, M. R. Borrelli, E. M. Rada, J. Caffrey, and Y. Rasko, "Risk factors for cooking-related burn injuries in children: WHO Global Burn Registry," *Bull. World Health Organ.*, vol. 99, no. 6, pp. 439–445, 2021, doi: 10.2471/BLT.20.279786L.
- [3] L. Olatomiwa, H. O. Idakwo, W. S. Olusola, and C. Ezech, "Global energy consumption trends," *Majlesi J. Electr. Eng.*, vol. 19, no. 1, art. no. 192509, 2025, doi: 10.57647/j.mjee.2025.1901.09.
- [4] H. N. Abraham, "Recent advancement in demand side energy management system for optimal energy utilization," *Energy Rep.*, vol. 11, pp. 5422–5435, 2024, doi: 10.1016/j.egy.2024.05.028.24.05.028
- [5] R. Qu, R. Kou, and T. Zhang, "The impact of weather variability on renewable energy consumption: Insights from explainable machine learning models," *Sustainability*, vol. 17, no. 1, art. no. 87, 2025, doi: 10.3390/su17010087.
- [6] M. Mahmood et al., "Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications," *Energy Convers. Manag.: X*, vol. 24, art. no. 100790, 2024, doi: 10.1016/j.ecmx.2024.100790.
- [7] T. Shojaei and A. Mokhtar, "Forecasting energy consumption with a novel ensemble deep learning framework," *J. Build. Eng.*, vol. 96, no. 1, art. no. 110452, 2024, doi: 10.1016/j.job.2024.110452.
- [8] I. Antonopoulos et al., "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," *Renew. Sustain. Energy Rev.*, vol. 130, art. no. 109899, 2020, doi: 10.1016/j.rser.2020.109899.
- [9] M. Shao et al., "Prediction of energy consumption in hotel buildings via support vector machines," *Sustain. Cities Soc.*, vol. 57, art. no. 102128, 2020, doi: 10.1016/j.scs.2020.102128.
- [10] Y. Ali et al., "Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review," *Accid. Anal. Prev.*, vol. 194, art. no. 107378, 2024, doi: 10.1016/j.aap.2023.107378.
- [11] W. Sholihah et al., "Revolutionizing healthcare: Comprehensive evaluation and optimization of SVM kernels for precise general health diagnosis," *Sci. J. Informatics*, vol. 10, no. 4, pp. 1–12, Nov. 2023, doi: 10.15294/sji.v10i4.46430
- [12] L. Xu, L. Hou, Z. Zhu, et al., "Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm," *Energy*, vol. 222, art. no. 119955, 2021, doi: 10.1016/j.energy.2021.119955.
- [13] N. W. Azizah et al., "Analyzing the relationship between meteorological parameters and electric energy consumption using support vector machine and cooling degree days algorithm," *J. Inf. Syst. Informatics*, vol. 6, no. 2, pp. 729–750, 2024, doi: 10.51519/journalisi.v6i2.719.
- [14] A. F. A. H. Alnuaimi and T. H. K. Albaldawi, "Concepts of statistical learning and classification in machine learning: An overview," *BIO Web Conf.*, vol. 97, art. no. 00129, 2024, doi: 10.1051/bioconf/20249700129.
- [15] A. Tsiamis, I. Ziemann, N. Matni, and G. J. Pappas, "Statistical learning theory for control: A finite-sample perspective," *IEEE Control Syst. Mag.*, vol. 43, no. 6, pp. 67–97, Dec. 2023, doi: 10.1109/MCS.2023.3310345.
- [16] K.-L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring kernel machines and support vector machines: Principles, techniques, and future directions," *Mathematics*, vol. 12, art. no. 3935, 2024, doi: 10.3390/math12243935.
- [17] R. H. Muhammadi et al., "Combination of support vector machine and lexicon-based algorithm in Twitter sentiment analysis," *Khazanah Informatika*, vol. 8, no. 1, pp. 59–71, Apr. 2022.
- [18] D. Valero-Carreras et al., "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, art. no. 106131, 2023, doi: 10.1016/j.cor.2022.106131.