

Estimating Building Cooling and Heating Energy: A Comprehensive Comparison of Machine Learning Models; the Case of ASHRAE Standard Sample Parametric Design in Tehran

Fateme Akhlaghinezhad* 

Ph.D. Candidate, Faculty of Architecture and Urban Planning, Shahid Beheshti University, Tehran, Iran

Received: January 22, 2024

Revised: May 11, 2024

Accepted: May 20, 2024

(Pages: 53–82)

Akhlaghinezhad, Fateme. "Estimating Building Cooling and Heating Energy: A Comprehensive Comparison of Machine Learning Models; the Case of ASHRAE Standard Sample Parametric Design in Tehran". *Soffeh* 36, no. 3 (2026): 53–82.

DOI: <http://doi.org/10.48308/soffeh.2026.234552.1307>

Abstract:

Background and Objectives: It has become an imperative to Improve energy efficiency in the building sector, especially in the early stages of design, in order to reduce greenhouse gas emissions and fossil fuel consumption. This requires accurate energy forecasts to optimise decision makings. In recent years, artificial intelligence and machine learning techniques have been proposed to predict energy consumption and building performance. The simplicity of building modelling and the use of simulated building energy consumption data has improved machine learning models.

Materials and Methods: The present research uses a parametric design of the ASHRAE standard sample (Case 600) with the predicting features of building orientation, window-to-wall ratio (WWR), building dimensions

Keywords:

Boosting machine learning, Interpretable machine learning, Cross validation, Tuning of hyperparameters.



SOFFEH

Soffeh Journal, Shahid Beheshti University, Vol. 36, Issue 3, No. 114, 2026



P-ISSN: 1683-870X

*. Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

*. Corresponding Author Email Address: fakhlaghi1377@gmail.com
<http://doi.org/10.48308/soffeh.2026.234552.1307>

--including length, width, height and building use – and simulation using EnergyPlus and jEPlus software. Then, using several widely-used machine learning approaches in predicting and improving building energy performance – including regression approach with the help of Linear and Gaussian regression, classification approach with the help of Logistic regression, K-Nearest Neighbour, Decision Tree, Random Forests and Support Vector Machine, the Multilayer Neural Network approach, the clustering approach with the help of K-means, Hierarchical and Density-Based algorithms – the cooling and heating energy of the building is predicted, with the resulting data being classified.

The results are then combined with several boosting machine learning models – including AdaBoost, Gradient Boosting Machine and its derivatives – and a comparison is made about the accuracy of prediction models with different evaluation criteria, followed by an optimisation of the hyperparameter of prediction models, a 10-fold cross-validation, and finally an interpretation of the results with the help of interpretable machine learning in the best possible way (SHapley value).

Results and Conclusion: The results show that the use of the Boosting Machine model improves the prediction accuracy, and that compared to single learning models, and among the single machine learning models, the Artificial Neural Network and the Decision Tree have the highest prediction accuracy for the regression and classification problems of building cooling and heating energy prediction after cross-validation. The effect of adjusting the Artificial Neural Network in increasing the accuracy of the model prediction is more than other models, and the Label Encoding method is recommended for predicting the cooling energy and the One Hot Encoding method for predicting the heating energy for data preprocessing.

تخمین انرژی سرمایشی و گرمایشی ساختمان: مقایسه‌ای جامع از مدل‌های یادگیری ماشین؛ نمونه موردی: طراحی پارامتریک نمونه استاندارد اشری در اقلیم تهران



فاطمه اخلاقی نژاد

دریافت: ۲ بهمن ۱۴۰۲
بازبینی: ۲۲ اردیبهشت ۱۴۰۳
پذیرش: ۳۱ اردیبهشت ۱۴۰۳

(صفحه ۵۳-۸۲)

اخلاقی نژاد، فاطمه. «تخمین انرژی سرمایشی و گرمایشی ساختمان: مقایسه‌ای جامع از مدل‌های یادگیری ماشین؛ نمونه موردی: طراحی پارامتریک نمونه استاندارد اشری در اقلیم تهران». فصلنامه علمی معماری و شهرسازی صفه، سال ۳۶، شماره ۳ (۱۴۰۵): ۵۳-۸۲.

کلیدواژگان: یادگیری ماشین تقویتی، یادگیری ماشین تفسیرپذیر، اعتبارسنجی متقاطع، تنظیم هایپرپارامترها.

چکیده

اهداف و پیشینه: بهبود بهره‌وری انرژی در بخش ساختمان، مخصوصاً در مراحل اولیه طراحی، هدفی اساسی برای کاهش میزان انتشار گازهای گلخانه‌ای و نیز مصرف سوخت فسیلی است که این هدف مستلزم پیش‌بینی دقیق انرژی برای تصمیم‌گیری بهینه است. در سال‌های اخیر، هوش مصنوعی و روش‌های یادگیری ماشین برای پیش‌بینی مصرف انرژی و عملکرد ساختمان پیشنهاد شده‌اند. سادگی مدل‌سازی ساختمان‌ها و استفاده از داده‌های مصرف انرژی ساختمان شبیه‌سازی شده باعث بهبود مدل‌های یادگیری ماشین شده است.

مواد و روش‌ها: در این مقاله با طراحی پارامتریک نمونه استاندارد اشری (نمونه ۶۰۰) با متغیرهای پیش‌بینی جهت‌گیری ساختمان، نسبت پنجره به دیوار، ابعاد ساختمان شامل طول، عرض، ارتفاع و کاربری ساختمان و شبیه‌سازی به کمک نرم‌افزارهای EnergyPlus و JEPlus داده‌سازی می‌شود، سپس با استفاده از چندین رویکرد پراکارد یادگیری ماشین در پیش‌بینی و بهبود عملکرد انرژی ساختمان، شامل رویکرد رگرسیون با کمک رگرسیون خطی و مبتنی بر گاوس، رویکرد طبقه‌بندی با کمک الگوریتم رگرسیون لجستیک، K-نزدیک‌ترین

همسایه، درخت تصمیم، جنگل‌های تصادفی، ماشین بردار پشتیبان، رویکرد شبکه عصبی چندلایه، رویکرد خوشه‌بندی با کمک الگوریتم‌های K-میانگین، و سلسله‌مراتبی و مبتنی بر چگالی، به پیش‌بینی انرژی سرمایشی و گرمایشی ساختمان و گروه‌بندی داده‌های حاصل پرداخته می‌شود، در ادامه نتایج با چندین مدل یادگیری ماشین تقویتی شامل AdaBoost، ماشین تقویت گرادیان و مشتقات آن مقایسه می‌گردد، دقت مدل‌های پیش‌بینی با معیارهای ارزیابی مختلف سنجیده می‌شود، هایپرپارامتر مدل‌های پیش‌بینی بهینه‌سازی می‌گردد، به اعتبارسنجی متقاطع ۱۰ برابری پرداخته می‌شود، و در نهایت به کمک یادگیری ماشین تفسیرپذیر، نتایج به بهترین شکل تفسیر می‌گردد.

نتایج و جمع‌بندی: نتایج نشان می‌دهد که استفاده از مدل ماشین تقویت‌کننده، دقت پیش‌بینی را در مقایسه با مدل‌های یادگیری انفرادی بهبود می‌بخشد و در میان مدل‌های یادگیری ماشین انفرادی، شبکه عصبی مصنوعی و درخت تصمیم بالاترین دقت پیش‌بینی را در مسائل رگرسیونی و طبقه‌بندی دارند. تأثیر تنظیم شبکه عصبی مصنوعی در افزایش دقت پیش‌بینی مدل بیشتر از سایر مدل‌هاست و در پیش‌بینی انرژی سرمایشی روش Label Encoding و در پیش‌بینی انرژی گرمایشی روش One Hot Encoding برای پیش‌پردازش داده‌ها توصیه می‌شود.

۱. دانشجوی دوره دکتری، دانشکده معماری و شهرسازی، دانشگاه شهید بهشتی، تهران، ایران
fakhlaghi1377@gmail.com



۱۱۴ شماره ۳، پیاپی: ۱۴۰۵، سال سی‌وششم، پاییز ۱۴۰۵
*. Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ص ۵۳
*. Corresponding Author Email Address: fakhlaghi1377@gmail.com
<http://doi.org/10.48308/soffeh.2026.234552.1307>

پرسش‌های پژوهش

پرسش اصلی

– کدام الگوریتم یادگیری ماشین از نظر دقت در حوزه پیش‌بینی انرژی سرمایشی و گرمایشی ساختمان مناسب‌ترین است؟

پرسش‌های فرعی

۱. میزان و جهت تأثیر متغیرهای طراحی ساختمان شامل جهت گیری ساختمان، نسبت پنجره به دیوار، ابعاد ساختمان شامل طول، عرض، ارتفاع، و کاربری ساختمان بر انرژی سرمایشی و گرمایشی چیست؟
۲. تفاوت میان الگوریتم‌های یادگیری ماشین انفرادی و تقویتی چیست؟
۳. کدام روش پیش‌پردازش داده‌ها در مدل‌های مختلف یادگیری ماشین مناسب‌ترین است؟
۴. تأثیر بهینه‌سازی هایپر پارامترها در هریک از مدل‌های یادگیری ماشین چه میزان است؟

مقدمه

اژانس بین‌المللی انرژی افزایش ۲۰ درصدی مصرف انرژی در بخش ساختمان را تا سال ۲۰۴۰ پیش‌بینی کرده است،^۲ که این مستلزم اجرای استراتژی‌های کارآمد برای کاهش مصرف انرژی ساختمان است.^۳ حتی یک بهبود جزئی در بهره‌وری انرژی ساختمان‌ها می‌تواند تا حد زیادی مصرف انرژی را بکاهد. پیش‌بینی دقیق مصرف انرژی ساختمان مبنای توسعه استراتژی‌های مختلف برای بهره‌وری انرژی ساختمان است.^۴ فشار فزاینده‌ای برای تغییرات پایدار در بخش ساختمان اعمال می‌شود، هم از منظر ساخت‌وساز (کاهش مصرف انرژی از طریق تصمیمات عمدی طراحی ساختمان مانند استفاده از مواد انرژی کارا) و هم از دیدگاه رفتاری (مانند تغییر رفتار انسان در مصرف انرژی در داخل ساختمان‌ها). برای اثربخشی بیشتر چنین اقداماتی، توانایی پیش‌بینی مصرف انرژی ساختمان بر اساس ویژگی‌های آن حیاتی است.^۵

اولین گام در کاهش مصرف انرژی، محاسبه مقدار کاهش با استفاده از روش ارزیابی انرژی ساختمان است که یک شاخص عملکرد انرژی مقایسه‌ای را در اختیار تصمیم‌گیرندگان قرار می‌دهد. ارزیابی انرژی ساختمان به ۴ دسته اصلی تقسیم می‌شود: (۱) محاسبات مهندسی، (۲) شبیه‌سازی مبتنی بر مدل‌سازی (۳) مدل‌سازی‌های آماری، و (۴) یادگیری ماشین.^۶ در روش‌های مهندسی (جعبه سفید) از قوانین فیزیکی، دقیق‌ترین روش‌ها، و ریاضیات پیچیده یا دینامیک ساختمان با در نظر گرفتن جزئیات داخلی و خارجی ساختمان به‌منزله ورودی (مانند اطلاعات آب‌وهوا، پوسته ساختمان، سیستم تهویه مطبوع، موقعیت پنجره، و ...) برای استخراج مصرف انرژی ساختمان استفاده می‌شود. شبیه‌سازی بهره‌وری انرژی ساختمان (جعبه خاکستری) شامل نرم‌افزار و مدل‌های رایانه‌ای برای شبیه‌سازی عملکرد ساختمان با وضعیت از پیش تعریف شده است. به‌طور کلی، شبیه‌سازی رایانه‌ای را می‌توان برای کاربردهای مختلفی مانند طراحی سیستم روشنایی، بارهای داخلی، تهویه مطبوع، و رفتار ساکنان^۷ استفاده کرد. وجود داده‌های انرژی ساختمان امکان بهره‌گیری از روش‌های بالا را برای ارزیابی عملکرد انرژی فراهم کرده است. با این حال، فقدان ورودی دقیق یا در دسترس نبودن اطلاعات ساختمانی بسیار خاص، که اغلب ناشناخته هستند، منجر به پیش‌بینی ضعیف^۸ و نیز محدودیت تعمیم

2. N. Zhou, et al., "A Roadmap for China to Peak Carbon Dioxide Emissions and Achieve a 20% Share of Non-Fossil Fuels in Primary Energy by 2030", *Applied Energy*, 239 (2019): 793-819.

3. A. Karimi, et al., "Assessment of Outdoor Design Conditions on the Energy Performance of Cooling Systems in Future Climate Scenarios—A Case Study over Three Cities of Texas, Unites States", *Sustainability*, 14(22) (2022): 14848.

4. Z. Dong, et al., "Hourly Energy Consumption Prediction of an Office Building Based on Ensemble Learning and Energy Consumption Pattern Classification", *Energy and Buildings*, 241 (2021): 110929.

5. A. Bassi, et al., "Building Energy Consumption Forecasting", in *The 12th International Conference on Advances in Information Technology* (NewYork, NY, USA: ACM, 2021), 1-9.

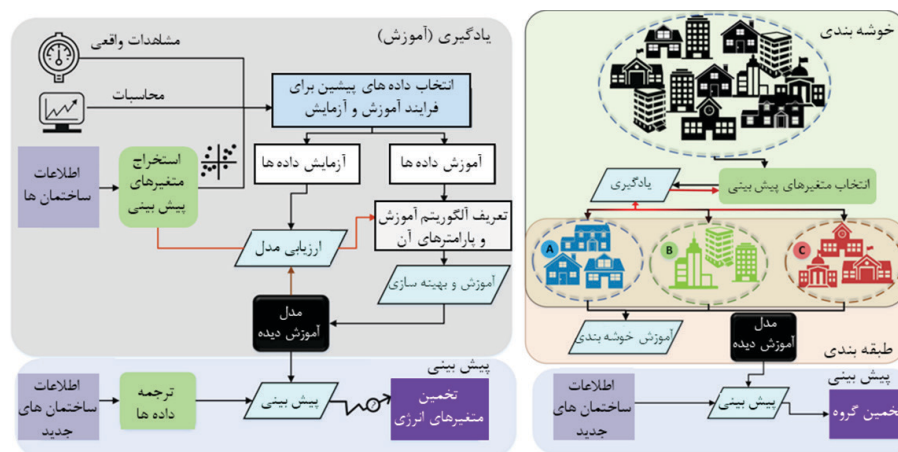
6. ML: Machine Learning

7. M.M. Ouf, et al., "Improving Occupant-Related Features in Building Performance Simulation Tools," *Building Simulation* 11, no. 4 (2018): 803-817.

8. E. Cuerda, et al., "Understanding the Performance Gap in Energy Retrofitting," *Energy and Buildings*, 209 (2020): 109688.

ت ۱. دیاگرام شماتیک یادگیری ماشین با نظارت (چپ) و بدون نظارت (راست)، مأخذ:

Saleh Seyedzadeh et al., "Machine Learning for Estimation of Building Energy Consumption and Performance: A Review", *Visualization in Engineering*, 6(1) (2018): 1-20.



نتایج در انواع مختلف ساختمان و اقلیم‌های دیگر می‌شود. در روش‌های آماری (جعبه سیاه) از داده‌های تاریخی ساختمان و اغلب از رگرسیون برای مدل‌سازی مصرف انرژی / عملکرد ساختمان‌ها استفاده می‌شود. این مدل‌ها، مدل‌های جایگزین مبتنی بر داده نیز نامیده می‌شوند؛ زیرا به جای تکیه بر جزئیات پیچیده سیستم، مبتنی بر داده‌های موجود هستند.

مفهوم یادگیری ماشین ارتباط نزدیکی با آمار محاسباتی دارد و می‌توان آن را زیرمجموعه مدل‌سازی آماری در نظر گرفت، بنابراین مدل‌های یادگیری ماشین، به مثابه یک جعبه سیاه، نیازی به اطلاعات در مورد سیستم‌های ساختمانی ندارند. یادگیری ماشینی می‌تواند بدون برنامه‌ریزی صریح «یاد بگیرد». این باعث می‌شود که این مدل با محیط غیرقطعی و سریع در حال تغییر در دنیای واقعی سازگارتر شود.^۹ یادگیری ماشین زیرمجموعه‌ای از هوش مصنوعی است که توانایی یادگیری از داده‌های ساختمان را با استفاده از الگوریتم‌های رایانه‌ای فراهم می‌کند.^{۱۰} از جمله روش‌های یادگیری ماشین، یادگیری‌های نظارت‌شده و بدون نظارت هستند که به ترتیب، روابط بین ویژگی‌های ورودی و اهداف خروجی را کشف و الگوهای پنهان را استخراج می‌کنند. این روش‌ها به تحلیل و پیش‌بینی مصرف انرژی ساختمان کمک می‌کنند. طرح کلی یادگیری نظارت‌شده و نظارت‌نشده برای مدل‌سازی انرژی ساختمان در «ت ۱» نشان داده شده است.

اگرچه مدل‌های یادگیری ماشین، اغلب در کارهای تحقیقاتی قبلی مطالعه شده است، بررسی ادبیات کامل یادگیری ماشین، یک طرح مقایسه‌ای بین الگوریتم‌های مختلف یادگیری انفرادی و تقویتی است و برای تعیین میزان متغیرهای پیش‌بینی‌کننده مصرف انرژی ساختمان و تأثیر این متغیرها ضروری به نظر می‌رسد. در این مقاله روش‌های مختلف یادگیری ماشین با تأکید بر مزایا و معایب بیان می‌گردد. سپس

9. L. Zhang, et al., "A Review of Machine Learning in Building Load Prediction", *Applied Energy*, 285 (2021): 116452.
10. S. Seyedzadeh, et al., "Machine Learning for Estimation of Building Energy Consumption and Performance: A Review", *Visualization in Engineering*, 6(1) (2018): 1-20.
11. WWR: Window to Wall Ratio
12. Shapley Value
13. Zhang, et al., "A Review of Machine Learning in Building Load Prediction", 116452.
14. Bassi, et al., "Building Energy Consumption Forecasting", 1-9.
15. Xuming He, *Handbook of Quantile Regression*, ed. Roger Koenker et al., *Handbook of Quantile Regression* (Chapman and Hall/CRC, 2017).
16. B. Yildiz, et al., "A Review and Analysis of Regression and Machine Learning Models on Commercial Building Electricity Load Forecasting", *Renewable and Sustainable Energy Reviews*, 73 (2017): 1104-1122.

گرمایشی و سرمایشی ساختمان بر اساس داده‌های یادگیری (۷۰٪ داده‌ها)، پیش‌بینی می‌شود، بهینه‌سازی پارامترها، ارزیابی و اعتبارسنجی مدل انجام می‌شود و در نهایت از طریق یادگیری ماشین تفسیرپذیر^{۱۲} پیش‌بینی‌ها تفسیر می‌شوند تا میزان و جهت تأثیر هریک از متغیرهای پیش‌بینی‌کننده بر مصرف انرژی روشن شود. نتایج نشان می‌دهد که چگونه استراتژی‌های طراحی مختلف به کاهش انرژی سرمایشی و گرمایشی ساختمان کمک می‌کنند.

۱. ادبیات تحقیق

در سال‌های اخیر، روش‌های یادگیری ماشین متعددی در بخش ساختمان برای تخمین انرژی گرمایشی و سرمایشی، مصرف انرژی، آسایش حرارتی، و عملکرد ساختمان در شرایط مختلف پیشنهاد شده است. افزایش کاربرد یادگیری ماشین در سیستم‌های انرژی ساختمان در اواخر دهه ۱۹۹۰ آغاز شد^{۱۳} و از سال ۲۰۱۰ تحقیقات انرژی ساختمان در حوزه یادگیری ماشین رشد چشمگیری یافته است و تا کنون با توسعه سی‌ساله، روش‌های یادگیری ماشین به‌طور گسترده در مدل‌سازی سیستم انرژی ساختمان استفاده شده‌اند. چند مورد از الگوریتم‌های یادگیری ماشین معمولاً برای پیش‌بینی بار ساختمان استفاده می‌شوند، شامل: (۱) رگرسیون خطی، (۲) ماشین بردار پشتیبان، (۳) شبکه‌های عصبی، (۴) یادگیری عمیق، (۵) الگوریتم‌های مبتنی بر درخت، (۶) الگوریتم‌های تقویتی، و (۷) سایر موارد.^{۱۴} هریک از این الگوریتم‌ها در ادامه شرح داده می‌شود.

۱.۱. یادگیری بانظارت: رگرسیون و طبقه‌بندی

در یادگیری ماشین، مسائل با پاسخ کمی معمولاً در دسته رگرسیون و مسائل با پاسخ کیفی در دسته طبقه‌بندی قرار می‌گیرند؛ با این حال، مرز میان این دو همیشه کاملاً مشخص

نیست؛ به‌طور مثال، رگرسیون لجستیک با اینکه غالباً با عنوان روش طبقه‌بندی به کار می‌رود، به‌دلیل برآورد احتمال کلاس‌ها، می‌تواند در چارچوب رگرسیونی نیز تفسیر شود. همچنین برخی الگوریتم‌ها مانند K-نزدیک‌ترین همسایه، ماشین بردار پشتیبان و روش‌های تقویتی، قابلیت کاربرد در هر دو نوع مسئله را دارند. در انتخاب روش، ماهیت متغیر هدف اهمیت اصلی را دارد و نوع متغیرهای پیش‌بین معمولاً اهمیت کمتری دارد.^{۱۵} زیرا متغیرهای کیفی را می‌توان با روش‌های کدگذاری مناسب مانند Label Encoding یا One-Hot Encoding وارد مدل کرد. حتی متغیرهای کمی هدف نیز در صورت نیاز با باینری‌سازی قابل تبدیل به مسئله طبقه‌بندی هستند. از آنجاکه پیش‌بینی بار ساختمان ذاتاً یک مسئله رگرسیونی محسوب می‌شود، روش‌های یادگیری ماشین به‌طور گسترده در این حوزه به کار رفته‌اند،^{۱۶} هرچند در این پژوهش پیش‌بینی هم به‌صورت رگرسیونی و هم طبقه‌بندی انجام شده است.

۱.۱.۱. رگرسیون خطی (LR – Linear Regression)

روش‌های رگرسیونی از جمله رویکردهای آماری پرکاربرد برای پیش‌بینی و ارزیابی عملکرد انرژی ساختمان بر اساس داده‌های ساختمان و اقلیم هستند. رگرسیون ساده و چندمتغیره^{۱۷} مصرف انرژی را به یک یا چند متغیر مؤثر مرتبط می‌کند و رگرسیون غیرخطی نیز در مطالعات متعددی برای پیش‌بینی کاربرد انرژی بیان شده است؛^{۱۸} به‌طور نمونه، کاتالینا و همکاران^{۱۹} مدل‌های رگرسیونی‌ای را برای پیش‌بینی تقاضای گرمایش ماهانه ساختمان‌های مسکونی ارائه کردند که ورودی‌هایی مانند ضریب شکل ساختمان، ضریب انتقال حرارت پوسته^{۲۰}، نسبت پنجره به مساحت طبقه، ثابت زمانی ساختمان، و شرایط اقلیمی را شامل می‌شد. بنابر نتایج آن مطالعه، این مدل‌ها ابزاری ساده و کارآمد برای برآورد تقاضای گرمایش هستند.

۳.۱.۱. رگرسیون فرایند گاوسی (GPR – Gaussian process regression)

از اوایل سال ۲۰۰۰، محققان از رگرسیون فرایند گاوسی در کاربردهای مختلف استفاده کرده‌اند.^{۳۷} در زمینه انرژی ساختمان، اخیراً از فرایند گاوسی، به دلیل پتانسیل آن در تعیین عدم قطعیت پیش‌بینی‌ها، استفاده شده است. در مدل‌سازی انرژی ساختمان، معمولاً در بخش مقادیر مناسب برای برخی ویژگی‌ها (مانند عایق پوسته) عدم قطعیت وجود دارد. از این‌رو، ارزیابی عدم قطعیت ورودی در نتایج پیش‌بینی‌شده، رگرسیون فرایند گاوسی را، به یک رویکرد جایگزین برای مدل‌سازی انرژی ساختمان به جای مدل‌های رگرسیون خطی مرسوم تبدیل کرده است. راستوگی و همکاران^{۳۸} دقت رگرسیون گاوسی و خطی را در شبیه‌سازی عملکرد یک ساختمان مقایسه کردند و نشان دادند که دقت رگرسیون گاوسی ۴ برابر بهتر از رگرسیون خطی در مطالعات موردی شبیه‌سازی‌شده EnergyPlus واقع در ایالات متحده است. اشکال اصلی مدل‌سازی گاوسی هزینه محاسباتی گران، مخصوصاً در نمونه‌های آموزشی زیاد، و پیچیدگی تقریباً برابر با مکعب نمونه‌های مسئله است. مزیت این روش نیز همبستگی خودکار بین ویژگی‌ها در آن است. نگارنده پژوهش حاضر، با توجه به حجم داده کمتر (۵۷۶)، استفاده از رگرسیون گاوسی را منطقی می‌داند.

۴.۱.۱. شبکه عصبی مصنوعی (ANN) و پرسپترون چندلایه (MLP)

شبکه‌های عصبی مصنوعی یکی از روش‌های اصلی یادگیری ماشین است که به‌خاطر الهام‌گیری‌اش از ساختار عصبی زیستی، به‌طور گسترده در حوزه انرژی ساختمان^{۳۹} به کار آمده‌اند. این شبکه‌ها متشکل از لایه‌های ورودی، پنهان، و خروجی هستند که با بهره‌گیری از توابع فعال‌سازی (غالباً خطی و لجستیک) و

در پژوهشی دیگر^{۴۰}، همین محققان با کاهش ورودی‌ها به سه پارامتر اصلی شامل تلفات حرارتی کل، مساحت جبهه جنوبی، و اختلاف دمای تنظیم داخلی و دمای هوا، مدلی ساده‌تر و با دقت قابل‌قبولی را پیش‌بینی و عرضه کردند. رگرسیون خطی ساده برای پیش‌بینی بر اساس یک متغیر به کار می‌رود، اما در عمل، اغلب چندین متغیر پیش‌بین وجود دارد؛ از این‌رو از رگرسیون خطی چندگانه با اختصاص ضرایب مستقل به هر متغیر استفاده می‌شود. علامت ضرایب نشان‌دهنده نوع همبستگی با متغیر هدف است. از مهم‌ترین مزایای رگرسیون خطی می‌توان به سادگی، سرعت، قابلیت تفسیر بالا، و نبود نیاز به تنظیم هاپرپارامتر اشاره کرد. رگرسیون خطی چندگانه در این پژوهش مدل پایه کار است.

۲.۱.۱. رگرسیون لجستیک (Logistic Regression)

این رگرسیون یک روش آماری تثبیت‌شده برای تحلیل و مدل‌سازی متغیرهای وابسته باینری است. این روش با موفقیت برای توصیف احتمال متغیر هدف^{۴۱} است و طبقه‌بندی در یادگیری ماشینی نیز به کار می‌رود.^{۴۲} از نظر رگرسیون، در رگرسیون لجستیک^{۴۳} استدلال‌های یک مدل لجستیک که نوعی رگرسیون باینری است، تخمین زده می‌شود. بنابراین، نتایج آن در حالت باینری هستند: ۱ یا ۰. از رگرسیون لجستیک برای پیش‌بینی بار ساختمان، مصرف انرژی، آسایش حرارتی، و رفتار ساکنان به‌طور گسترده استفاده می‌شود.^{۴۴} رحمان و همکاران^{۴۵} الگوریتم‌های مختلف یادگیری ماشین را برای پیش‌بینی احساس حرارتی افراد اعمال کردند و به این نتیجه رسیدند که دقت پیش‌بینی رگرسیون لجستیک از مدل‌های K-نزدیک‌ترین همسایه و درخت تصمیم بیشتر و از مدل جنگل‌های تصادفی و شبکه عصبی عمیق کمتر است. در پژوهش حاضر از رگرسیون لجستیک برای پیش‌بینی کلاس باینری متغیر هدف استفاده شده است.

17. R.E. Edwards, et al., "Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study", Energy and Buildings, 49 (2012): 591-603; J.S. Hygh, et al., "Multivariate Regression as an Energy Assessment Tool in Early Building Design", Building and Environment, 57 (2012): 165-175.

18. Y.K. Penya, et al., "Short-Term Load Forecasting in Non-Residential Buildings," in IEEE Africon'11 (IEEE, 2011), 1-6; Min-Yuan Cheng and Minh-Tu Cao, "Accurately Predicting Building Energy Performance Using Evolutionary Multivariate Adaptive Regression Splines", Applied Soft Computing, 22 (2014): 178-188.

19. T. Catalina, et al., "Development and Validation of Regression Models to Predict Monthly Heating Demand for Residential Buildings", Energy and Buildings, 40(10) (2008): 1825-1832.

20. U-Value

است. انتخاب هسته مناسب برای ماشین‌های بردار پشتیبان می‌تواند کار بسیار چالش‌برانگیزی باشد. مزیت دیگر ماشین بردار پشتیبان نسبت به سایر مدل‌های یادگیری ماشین، منحصربه‌فرد بودن و بهینه‌سازی سراسری راه حل تولیدشده آن است؛ زیرا نیازی به بهینه‌سازی غیرخطی با خطر قرار گرفتن در حداقل محلی^{۴۰} ندارد. یکی از اشکالات اصلی ماشین بردار پشتیبان زمان محاسبه است که ترتیبی تقریباً برابر با مکعب نمونه‌های مسئله (N3) دارد و برای تعداد نمونه‌های بیشتر از ۱۰۰۰ کاربردی ندارد.

در بخش ساختمان، ماشین بردار پشتیبان مخصوص پیش‌بینی انرژی سرمایشی و گرمایشی^{۴۱}، مصرف انرژی^{۴۲}، و طبقه‌بندی مصرف انرژی ساختمان^{۴۳} است. در این پژوهش تنها به ماشین بردار پشتیبان طبقه‌بندی اکتفا می‌شود.

۱.۱.۶. الگوریتم‌های مبتنی بر درخت

روش مبتنی بر درخت دارای ساختاری مانند نمودار جریان (فلوچارت) است، که در آن هر گره داخلی نشان‌دهنده یک آزمایش روی یک ویژگی و هر شاخه نشان‌دهنده نتیجه یک آزمایش است. بالاترین گره در درخت، گره ریشه (روت) و پایین‌ترین گره، گره برگ (لیف) است. این روش شامل زیرمجموعه‌های مختلفی از جمله جنگل‌های تصادفی، درخت رگرسیون، و درخت طبقه‌بندی^{۴۴} می‌شود و برای مسائل رگرسیونی و طبقه‌بندی قابل استفاده است. در الگوریتم‌های مبتنی بر درخت، انتخاب ویژگی برای هر گره بخش مهمی از الگوریتم را تشکیل می‌دهد که در ادامه شرح داده می‌شود. انتخاب ویژگی به معنای تعیین میزان اهمیت هر متغیر پیش‌بینی‌کننده در پیش‌بینی متغیر هدف است که هرچه تأثیر یک متغیر بیشتر باشد، در درخت ترسیم‌شده، به گره ریشه نزدیک‌تر است. انتخاب ویژگی قابلیت ذاتی همه انواع

به‌روزرسانی وزن‌ها و بایاس‌ها، قادر به مدل‌سازی روابط پیچیده بین متغیرهای طراحی و مصرف انرژی هستند.^{۴۰} کاربردهای ANN در این حوزه شامل پیش‌بینی بار حرارتی و برودتی^{۴۱}، مدل‌سازی سیستم‌های تهویه مطبوع^{۴۲}، و پیش‌بینی پارامترهای اقلیمی داخلی (مانند دما و رطوبت)^{۴۳} است.

در میان ساختارهای متنوع ANN، «پرسپترون چندلایه» یکی از رایج‌ترین مدل‌های یادگیری عمیق برای حل مسائل رگرسیون و طبقه‌بندی است.^{۴۴} MLP با برخورداری از یک یا چند لایه پنهان، توانایی بالایی در برازش توابع غیرخطی پیچیده دارد؛ به‌طوری‌که ترکیب تابع فعال‌سازی لجستیک در لایه‌های پنهان و تابع خطی در لایه خروجی، غالباً بهینه‌ترین عملکرد را در مدل‌سازی‌های انرژی دارد.^{۴۵} تعیین تعداد نورون‌های لایه پنهان در این مدل‌ها معمولاً بر اساس رویکرد آزمون و خطا صورت می‌گیرد.^{۴۶} با توجه به گزارش‌های پیشین مبنی بر برتری کارایی MLP نسبت به سایر ساختارهای عصبی در پیش‌بینی دقیق پارامترهای ساختمان^{۴۷}، در پژوهش حاضر نیز از شبکه عصبی MLP با رویکرد رگرسیونی برای پیش‌بینی بارهای انرژی استفاده شده است.

۱.۱.۵. ماشین بردار پشتیبان (SVM – Support Vector Machine)

ماشین‌های بردار پشتیبان مدل‌های بسیار قوی در حل مسائل غیرخطی هستند و در تحقیقات و صنعت برای اهداف رگرسیونی^{۴۸} و طبقه‌بندی^{۴۹} هستند. از آنجاکه این مدل‌ها را می‌توان با تعداد کمی از نمونه‌های داده آموزش داد، می‌توانند راه حل مناسبی برای مدل‌سازی موارد مطالعه بدون داده‌های تاریخی ثبت‌شده باشند. ماشین‌های بردار پشتیبان با عملکرد هسته به صورت یک شبکه عصبی مصنوعی دولایه عمل می‌کنند، اما تعداد هابیرپارامترها کمتر از شبکه عصبی

21. T. Catalina, et al., "Multiple Regression Model for Fast Prediction of the Heating Energy Demand", Energy and Buildings, 57 (2013): 302-312.

22. D. Cali, et al., "Analysis of Occupants' Behavior Related to the Use of Windows in German Households", Building and Environment, 103 (July 2016): 54-69.

23. S. Verbruggen, "Window Use Habits as an Example of Habitual Occupant Behaviour in Residential Buildings", (Ghent University, 2021).

24. AR. Javed, et al., "Analyzing the Effectiveness and Contribution of Each Axis of Tri-Axial Accelerometer Sensor for Accurate Activity Recognition", Sensors, 20(8) (2020): 2216; J. Tolles and W.J. Meurer, "Logistic Regression: Relating Patient Characteristics to Outcomes", Jama, 316(5) (2016): 533-534.

25. Weihao Liu, et al., "Modeling Window and Thermostat Use Behavior to Inform Sequences of Operation in Mixed-Mode Ventilation Buildings", Science and Technology for the Built Environment, 27(9) (2021): 1204-1220; M. Vellei, et al., "The Influence of Relative Humidity on Adaptive Thermal Comfort", Building and Environment, 124 (2017): 171-185; S. Lu, et al., "Data-Driven Simulation of a Thermal Comfort-Based Temperature Set-Point Control with ASHRAE RP884", Building and Environment, 156 (2019): 137-146; Sun Ho Kim and Hyeun Jun Moon, "Case Study of an Advanced Integrated Comfort Control Algorithm with Cooling, Ventilation, and Humidification Systems Based on Occupancy Status", Building and Environment, 133 (2018): 246-264.
26. Saif Ur. Rehman, et al., "Personalised Comfort: A Personalised Thermal Comfort Model to Predict Thermal Sensation Votes for Smart Building Residents", Enterprise Information Systems, 16(7) (2022): 1-23.

نمونه‌های مختلف به‌دست آید. مزیت اصلی آن خوانایی مدل و سرعت طبقه‌بندی است. لیو و همکاران^{۴۸} از درخت تصمیم برای تقسیم مصرف انرژی سیستم به ۸ الگو بر اساس عوامل کلیدی، که بر عملکرد انرژی ساختمان‌های اداری تأثیر می‌گذارند، استفاده کردند. این مدل دارای معایب خاص خود است، از جمله: (۱) به‌راحتی بیش از حد برازنده می‌شوند و (۲) آنها به تولید ویژگی‌های تقسیم‌کننده بیشتری متمایل هستند. در پژوهش پیش رو از درخت تصمیم طبقه‌بندی با استفاده از هر دو شاخص معرفی شده پیش گفته استفاده شده است.

۱.۱.۷. یادگیری ماشین گروهی (Ensemble Machine Learning):

الگوریتم‌های یادگیری گروهی (مانند جنگل تصادفی و روش‌های تقویت گرادیان)، به‌دلیل توانایی در جبران محدودیت داده‌های آموزشی، اصلاح خطاهای مدل‌های پایه، و بهبود قدرت تعمیم‌دهی، در ادبیات پیش‌بینی انرژی ساختمان جایگاه ویژه‌ای دارند.^{۴۹} از آنجاکه مبانی نظری و تاریخیچه توسعه این الگوریتم‌ها در مطالعات متعددی (مانند مطالعات میلر و همکاران^{۵۰} به تفصیل بررسی شده است، در پژوهش حاضر تمرکز بر کاربرد عملی و مقایسه سیستماتیک برای پیش‌بینی بارهای حرارتی است.

– جنگل‌های تصادفی (RF – Random Forests): جنگل‌های تصادفی یکی از قدرتمندترین الگوریتم‌های یادگیری ماشینی تحت نظارت و نمونه‌ای از یادگیری گروهی (۲، ۱، ۷) است که قادر به انجام رگرسیون و طبقه‌بندی است. آرایا و همکاران^{۵۱} از جنگل‌های تصادفی برای تشخیص ناهنجاری در مصرف انرژی ساختمان بهره بردند و احمد و همکاران^{۵۲} از این الگوریتم برای پیش‌بینی مصرف انرژی سیستم تهویه مطبوع ساعتی استفاده کردند. جنگل‌های تصادفی چندین جنگل با چندین درخت

الگوریتم‌های مبتنی بر درخت است.

– انتخاب ویژگی (Feature Selection): انتخاب ویژگی فرایندی برای انتخاب زیرمجموعه‌ای از مرتبط‌ترین ویژگی‌هاست^{۴۵} که می‌تواند نمایانگر یک ماتریس ویژگی کامل برای استفاده در مدل طبقه‌بندی یادگیری ماشین باشد. این مفهوم اصلی یادگیری ماشینی است که بر عملکرد مدل تأثیر می‌گذارد. بنابراین، در پژوهش حاضر از روش‌های چنگانه انتخاب ویژگی مانند بهره اطلاعات و ضریب Gini برای انتخاب بهترین ویژگی کمک‌کننده استفاده می‌شود. در زیر جزئیات روش‌های انتخاب ویژگی آورده شده است:

بهره اطلاعات (IG – Information Gain): IG با محاسبه انتروپی (ناخالصی) و اطلاعات عرضه‌شده در یک ویژگی کار می‌کند.^{۴۶} یک ویژگی اگر اطلاعات بیشتری را شامل شود و آنتروپی کمتری داشته باشد، اهمیت بیشتری دارد. آنتروپی هر متغیر تصادفی چیزی جز یک محدودیت ناخالصی نیست.

ضریب جینی (Gini): زمانی که ضریب جینی صفر باشد، یعنی همه داده‌ها در این مجموعه متعلق به یک دسته هستند و حداکثر اطلاعات مفید را می‌توان به‌دست آورد. در این پژوهش در مدل درخت تصمیم،

با استفاده از هر دو شاخص بهره اطلاعات و ضریب جینی، نمودار درختی ترسیم شده است تا میزان اهمیت هریک از متغیرهای پیش‌بینی روشن شود.

– درخت تصمیم (DT – Decision Tree): از میان الگوریتم‌های مبتنی بر درخت در این پژوهش فقط به شرح درخت تصمیم اکتفا می‌شود که یک روش طبقه‌بندی پایه است و می‌تواند مجموعه‌ای از قوانین if-then یا توزیع احتمال شرطی تحت یک شرایط مشخصی در نظر گرفته شود.^{۴۷} از ریشه تا گره برگ، درخت تصمیم همه فضای ویژگی را با توجه به طبقه‌بندی ویژگی و انتخاب آن، گام‌به‌گام تقسیم می‌کند تا طبقه‌بندی

(مانند شبکه‌های عصبی) مؤثر گزارش شده است، اما به دلیل حساسیت نسبی به نویز داده‌ها^{۶۱} در مقایسه با روش‌های گرادیانی مدرن، در پژوهش حاضر به‌منزله یک روش مکمل برای تحلیل تفاوت‌های عملکردی در شرایط داده‌ای خاص مورد استفاده است.

۱.۱.۸. K-نزدیک‌ترین همسایه (KNN – K Nearest Neighbor)

الگوریتم K-نزدیک‌ترین همسایه، یک روش بدون محدودیت است که برای طبقه‌بندی و رگرسیون استفاده می‌شود و کاربرد گسترده‌ای در تشخیص الگو دارد.^{۶۲} در این الگوریتم ورودی شامل K-نزدیک‌ترین نمونه‌های آموزشی در فضای ویژگی است. از این الگوریتم نیز برای پیش‌بینی مصرف انرژی ساختمان استفاده می‌شود^{۶۳} و در این پژوهش برای پیش‌بینی نوع طبقه‌بندی متغیر هدف استفاده شده است، همان‌طور که وحید و همکاران^{۶۴} از طبقه‌بندی KNN برای پیش‌بینی مصرف انرژی روزانه بر اساس طبقه‌بندی استفاده کرده‌اند.

۱.۲. یادگیری بدون نظارت - خوشه‌بندی

خوشه‌بندی یکی از روش‌های شناخته‌شده یادگیری ماشین است که مربوط به شناسایی روابط، الگوها، و توزیع‌های ضمنی در مجموعه داده‌هاست. خوشه‌بندی یک روش یادگیری بدون نظارت است و می‌تواند ساختار پنهان در مجموعه‌ای از داده‌های بدون برچسب را توصیف کند. در انرژی ساختمان، کاربرد اصلی این روش طبقه‌بندی ساختمان‌ها با استفاده از ویژگی‌های مختلف است. خوشه‌بندی نمونه‌های داده را در زیرمجموعه‌ها به گونه‌ای گروه‌بندی می‌کند که نمونه‌های مشابه با هم گروه‌بندی می‌شوند، درحالی‌که نمونه‌های مختلف به گروه‌های مختلف تعلق دارند. در این پژوهش نیز

تصمیم ایجاد می‌کند و برای ایجاد هر درخت از قاعده بوت استرپ^{۵۳} استفاده می‌کند که به معنای نمونه‌برداری تصادفی با جایگزینی از نمونه اصلی به همان اندازه است.^{۵۴} هرچه تصمیم قوی‌تر باشد، دقت بیشتری دارد. با تجمیع نتایج پیش‌بینی، تطبیق بیش از حد مدل را می‌توان در الگوریتم‌های درخت تصمیم‌گیری پیشرفته مانند درختان تقویت‌شده گرادیان یا جنگل‌های تصادفی کاهش داد. در این پژوهش از جنگل‌های تصادفی به روش طبقه‌بندی استفاده شده است.

ماشین تقویت گرادیان (GBM-Gradient Boost Machine): الگوریتم‌های تقویت گرادیان^{۵۵} با ترکیب مکرر «یادگیرندگان ضعیف»، مدل‌هایی با دقت پیش‌بینی بالا ایجاد می‌کنند. در پژوهش حاضر، سه الگوریتم پیشرفته این خانواده شامل XGBoost، LightGBM و CatBoost ارزیابی می‌شوند:

XGBoost: با استفاده از ساختار درختی بهینه‌سازی شده و در یادگیری سریع و دقت بالا شناخته شده است.^{۵۶}

LightGBM: با بهره‌گیری از ویژگی‌های هیستوگرام، کارایی حافظه و سرعت آموزش را در داده‌های حجیم بهبود می‌بخشد.^{۵۷} **CatBoost:** این الگوریتم با رویکرد ویژه در مدیریت داده‌های غیرعددی و جلوگیری از بیش‌برازش^{۵۸}، کارایی بالایی در مدل‌سازی‌های انرژی نشان داده است.^{۵۹}

از آنجا که به مقایسه سیستماتیک این سه مدل برای پیش‌بینی انرژی ساختمان در ادبیات پژوهش کمتر توجه شده است، در این مطالعه، هر سه الگوریتم به همراه مدل پایه GBM بر روی مجموعه داده شبیه‌سازی شده (۵۷۶ نمونه) اعمال و مقایسه می‌شوند.

AdaBoost: الگوریتم AdaBoost^{۶۰} با استفاده از درخت‌های تصمیم ساده (Stumps) و تخصیص وزن به نمونه‌های دارای خطای بالاتر، عمل می‌کند. اگرچه این روش در بهینه‌سازی ساختار مصرف انرژی و رفع برخی نقایص مدل‌های پایه

27. S.T.S. Bukkapatnam and Ch. Cheng, "Forecasting the Evolution of Nonlinear and Nonstationary Systems Using Recurrence-Based Local Gaussian Process Models", Physical Review E, 82(5) (2010): 56206; Xi. Jiang, et al., "Adaptive Gaussian Process for Short-Term Wind Speed Forecasting", in ECAI 2010 (IOS Press, 2010), 661-666; E. Grosicki, et al., "A Weighted Linear Prediction Method for Near-Field Source Localization", IEEE Transactions on Signal Processing, 53(10) (2005): 3651-3660.

28. P. Rastogi, et al., "Gaussian-Process-Based Emulators for Building Performance Simulation", in International Building Performance Simulation Association, 2017.

29. J. Massana, et al., "Short-Term Load Forecasting in a Non-Residential Building Contrasting Models and Attributes", Energy and Buildings, 92 (2015): 322-330.

30. R. Platon, et al., "Hourly Prediction of a Building's Electricity Consumption Using Case-Based Reasoning, Artificial Neural Networks and Principal Component Analysis", Energy and Buildings, 92 (2015): 10-18; S. Ferlito, et al., "Predictive Models for Building's Energy Consumption", in 2015 Xviii Aisem Annual Conference (IEEE, 2015), 1-4; Sung-Min Hong, et al., "Improved Benchmarking Comparability for Energy Consumption in Schools", Building Research & Information, 42(1) (2014): 47-61.

31. Seyedzadeh, et al., "Machine Learning for Estimation of Building Energy Consumption and Performance"; A.G.A. Sholahudin, et al., "Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments", Applied Mechanics and Materials, 819 (2016): 541-545.

32. S. Faizollahzadeh Ardabili, et al., "Modeling and Simulation Controlling System of HVAC Using Fuzzy and Predictive (Radial Basis Function, RBF) Controllers", Building Engineering, 6 (2016): 301-308.

الف. خوشه‌بندی سلسله‌مراتبی

خوشه‌بندی سلسله‌مراتبی با استفاده از پارتیشن‌بندی بازگشتی، به دوروش کلی قابل اجرا است: ۱) تجمعی (Agglomerative) / پایین به بالا؛ ادغام متوالی خوشه‌های اولیه (که هر نمونه یک خوشه است) تا رسیدن به ساختار مطلوب؛ ۲) تقسیمی (Divisive) / بالا به پایین؛ تقسیم متوالی خوشه واحد اولیه به زیرخوشه‌ها.

معیار شباهت در این روش‌ها بر پایه فواصل بین خوشه‌ای^{۷۷} تعیین می‌شود که شامل تک‌پیوندی (کوتاه‌ترین فاصله)^{۷۸}، پیوند کامل (دورترین فاصله)^{۷۹}، و پیوند میانگین (روش حداقل واریانس)^{۸۰} است. خروجی این فرایند یک دندروگرام است که نمایش بصری از ساختار تودرتوی داده‌ها در سطوح مختلف شباهت است؛ به‌طوری‌که با برش دندروگرام در سطح مشخص (در این پژوهش سطح ۶)، تعداد خوشه‌ها تعیین می‌شود.

مزایای این رویکرد شامل نبود نیاز به پیش‌فرض برای تعداد خوشه‌ها و تفسیرپذیری بالای نتایج است. با این حال، محدودیت‌هایی نظیر نبود امکان بازنگری در خوشه‌های تشکیل شده و هزینه محاسباتی بالا به دلیل محاسبه ماتریس فواصل، از نقاط ضعف آن است.

ب. خوشه‌بندی مبتنی بر چگالی

الگوریتم خوشه‌بندی فضایی مبتنی بر چگالی برنامه‌های کاربردی با نويز خوشه‌هایی از اشکال دلخواه را کشف می‌کند و برای پایگاه داده‌های فضایی بزرگ کارآمد است.^{۸۱} این الگوریتم با جستجوی همسایگی هر شیء در پایگاه داده، خوشه‌ها را جستجو و بررسی می‌کند که آیا شامل بیش از حداقل تعداد داده است. در این مدل خوشه‌بندی، تا زمانی که چگالی (تعداد نقاط داده) در همسایگی از آستانه‌ای فراتر رود، به رشد خوشه ادامه داده می‌شود؛ یعنی همسایگی یک شعاع معین

به‌منظور آشنایی با داده‌ها حاصل از شبیه‌سازی، ابتدا با استفاده از روش‌های مختلف خوشه‌بندی معرفی شده، همه داده‌ها طبقه‌بندی می‌شود. خوشه‌بندی داده‌های ساختمان قبل از اعمال الگوریتم‌های یادگیری بانظارت می‌تواند تنوع داده‌های مرتبط با عملیات ساختمان پیچیده را کاهش دهد، بنابراین عملکرد هر مدل فرعی و درنهایت عملکرد کلی مدل را بهبود می‌بخشد.^{۶۵}

روش‌های خوشه‌بندی برای شناخت الگوی مصرف انرژی ساختمان^{۶۶}، تشخیص داده‌های پرت^{۶۷}، تشخیص رفتار ساکنان^{۶۸}، و طبقه‌بندی سیستم روشنایی^{۶۹} استفاده می‌شوند و به‌طور گسترده در مطالعات موجود یک روش مهم پشتیبانی برای کمک به یادگیری ماشین در حل مشکلات دنیای واقعی قلمداد می‌شود. در مطالعات پیشین اغلب از الگوریتم‌های K-means^{۷۰}، شامل خوشه‌بندی سلسله‌مراتبی^{۷۱} و مبتنی بر چگالی^{۷۲} برای اهداف یادگیری بدون نظارت نظیر مصرف انرژی ساختمان استفاده شده است.^{۷۳}

۱.۲.۱-K-میانگین (K-means)

متداول‌ترین الگوریتم خوشه‌بندی K-means است که به‌طور مکرر تا حداکثر محلی^{۷۴} دنبال می‌شود.^{۷۵} این الگوریتم با انتخاب تصادفی K مرکز (مرکز خوشه) شروع می‌شود و هر داده به نزدیک‌ترین نقطه مرکزی مربوط می‌شود. سپس همه مراکز با استفاده از میانگین همه نقاط داده در یک گروه دوباره محاسبه می‌شوند. این فرایند تا زمانی ادامه می‌یابد که یک معیار توقف (مثلاً رسیدن به حداقل تجمع فاصله‌ها) را برآورده کند. لارا و همکاران^{۷۶} برای ارزیابی عملکرد انرژی مدارس در ایتالیا و مشخص کردن ساختمان مرجع برای هر گروه، از این خوشه‌بندی استفاده کردند.



۱.۳.۱. منبع داده‌ها

تولید و گردآوری داده‌ها از ارکان اصلی مدل‌های یادگیری ماشین است؛ زیرا داده‌ها پایه و ستون فقرات فرایند پیش‌بینی را تشکیل می‌دهند. در مسائل طراحی ساختمان‌های جدید، توسعه یک مدل جانشین داده‌محور بر مبنای داده‌های تاریخی، راهکاری عملی برای پیش‌بینی عملکرد انرژی به‌شمار می‌رود. در این رویکرد، از ویژگی‌های ساختمانی و داده‌های اقلیمی برای پیش‌بینی پارامترهای انرژی نمونه‌های جدید استفاده می‌شود. از آنجا که گردآوری میدانی بسیاری از متغیرهای مؤثر — مانند ویژگی‌های ساختاری، نرخ و ساعات اشغال، تجهیزات و شرایط بهره‌برداری — هزینه‌بر و زمان‌بر است، در اغلب مطالعات از داده‌های شبیه‌سازی شده استفاده می‌شود.^{۸۲}

در این پژوهش، داده‌های ورودی و خروجی از طریق شبیه‌سازی پارامتریک با نرم‌افزارهای EnergyPlus و JEPlus تولید شده‌اند. بدین ترتیب، داده‌های آب‌وهوایی و مشخصات هندسی مدل مرجع (نمونه استاندارد ۶۰۰ اشری) در EnergyPlus تعریف شده و در JEPlus، پارامترهای طراحی با عنوان متغیرهای پیش‌بینی‌کننده و خروجی‌های انرژی با عنوان متغیرهای هدف در نظر گرفته شده‌اند. سپس از این داده‌ها برای آموزش و ارزیابی مدل‌های یادگیری ماشین استفاده شده‌اند.

۱.۳.۲. تقسیم داده‌ها

عملکرد مدل‌های داده‌محور مبتنی بر هوش مصنوعی شامل مراحل پاک‌سازی داده‌ها (حذف داده‌های پرت و تکمیل داده‌های مفقود) و پیش‌پردازش است. سپس داده‌ها به‌صورت تصادفی به دو مجموعه آموزشی (۷۰٪) و آزمایشی (۳۰٪) تقسیم می‌شوند؛ به‌گونه‌ای که مدل با داده‌های آموزشی مواجه است و عملکرد آن بر اساس داده‌های آزمایشی ارزیابی می‌شود. دقت پیش‌بینی با مقایسه خروجی مدل و مقادیر واقعی سنجیده می‌شود.

باید دارای حداقل تعداد داده باشد. در این مدل، برخلاف مدل K-means، شکل خوشه‌ها دلخواه است و به‌راحتی داده‌های پرت را با عنوان نویز خارج می‌کند و نیازی به تعیین تعداد خوشه نیست. در پژوهش پیش رو خوشه‌بندی داده‌ها به روش K-means و DBSCAN در «ت ۳» مقایسه شده است.

۱.۳.۳. پیش‌پردازش داده‌ها

دقت پیش‌بینی مدل‌های یادگیری ماشین به دو عامل بستگی دارد:

(۱) عملکرد الگوریتم طراحی شده که هریک از الگوریتم‌های پرکاربرد در پیش‌بینی مصرف انرژی در بخش قبل مفصل توضیح داده شد؛

(۲) کیفیت داده‌های مورد استفاده.

بنابراین آشنایی با روش‌های پیش‌پردازش داده در این بخش ضروری به‌نظر می‌رسد. رویکردهای یادگیری ماشین عمدتاً مبتنی بر مجموعه داده‌های جمع‌آوری شده از مطالعات میدانی مختلف یا نتایج حاصل از شبیه‌سازی است. پیش‌بینی‌ها بر اساس یک مجموعه داده عمومی در مقیاس بزرگ و طبقه‌بندی‌کننده است. در یادگیری ماشین، مجموعه‌ای از ویژگی‌ها ورودی هستند و کلاس مصرف انرژی (طبقه‌بندی) یا مقدار مصرف انرژی (رگرسیون) خروجی پیش‌بینی می‌شوند. روش‌های پشتیبانی که برای افزایش دقت مدل پیش‌بینی به کار می‌روند، شامل انتخاب ویژگی، خوشه‌بندی، نرمال‌سازی، کدگذاری داده‌های متنی، اعتبارسنجی متقاطع، و تنظیم دقیق هایپرپارامترهای هر الگوریتم هستند که در ادامه به‌تفصیل بررسی خواهند شد و چندین معیار ارزیابی مدل یادگیری ماشین نیز مرور می‌شود.

33. Xin Shi, et al., "Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing", IEEE Access, 6 (2018): 30559-66; O.M. Tzuc, et al., "Artificial Intelligence Techniques for Modeling Indoor Building Temperature under Tropical Climate Using Outdoor Environmental Monitoring", Energy Engineering, 146(2) (2020).

34. B. Laarabi, et al., "Artificial Neural Network Modeling and Sensitivity Analysis for Soiling Effects on Photovoltaic Panels in Morocco", Superlattices and Microstructures, 127 (2019): 139-150.

35. Seyedzadeh, et al., "Machine Learning for Estimation of Building Energy Consumption and Performance: A Review"; M. Tzuc, et al., "Artificial Intelligence Techniques for Modeling Indoor Building Temperature under Tropical Climate Using Outdoor Environmental Monitoring"; M.H. Beale, et al., "Neural Network Toolbox", User's Guide, MathWorks, 2 (2010): 77-81.

در یک مقیاس جدید با ارزش واقعی خود ظاهر می‌شوند. در مقیاس‌بندی MinMax از «معادله ۱» برای نرمال‌سازی داده‌ها

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

استفاده می‌شود: روش دیگر مقیاس‌بندی، مقیاس استاندارد است که به توزیع استاندارد با میانگین صفر و انحراف معیار یک (واریانس واحد) کمک می‌کند. ویژگی‌ها را با کم کردن مقدار میانگین از ویژگی و سپس تقسیم نتیجه بر انحراف معیار ویژگی، استاندارد می‌کند که در «معادله ۲» آمده است. در این پژوهش از مقیاس

$$X_{norm} = \frac{X_i - X_{mean}}{std} \quad (2)$$

۴.۳.۱. کدگذاری داده‌ها به روش Label Encoding (LE) در پروژه‌های یادگیری ماشینی معمولاً مجموعه داده‌ها دارای ستون‌های دسته‌بندی متفاوتی هستند، به‌طوری‌که برخی از ستون‌ها دارای عناصر خود در دسته متغیرهای ترتیبی هستند؛ به‌طور مثال ستون سطح درآمد دارای عناصر کم، متوسط، یا زیاد است، در این مورد می‌توان این عناصر را با ۱ و ۲ و ۳ جانشین کرد. از طریق این نوع رمزگذاری (Label Encoding)^{۸۷}، معنای عنصر حفظ و وزن‌های بالاتری به عناصر دارای اولویت بالاتر اختصاص داده می‌شود.

Label Encoding روشی است برای تبدیل ستون‌های طبقه‌بندی به ستون‌های عددی تا بتوان آنها را با مدل‌های یادگیری ماشینی که فقط داده‌های عددی را دریافت می‌کنند، تطبیق داد.^{۸۸} این یک مرحله پیش‌پردازش مهم در یک پروژه یادگیری ماشینی است؛ اما برای داده‌هایی که نسبت به هم اولیوی ندارند، قابل‌استفاده نیست و یا ممکن است برچسبی با ارزش بالا نسبت به برچسبی که ارزش کمتری دارد، دارای اولویت کمتری در نظر گرفته شود.

از آنجاکه تقسیم تصادفی داده‌ها می‌تواند منجر به نوسان در نتایج شود، برای افزایش قابلیت اعتماد، از اعتبارسنجی متقاطع K برابری استفاده شده است. در این پژوهش، نرخ تصادفی برابر با ۱ برای همه الگوریتم‌ها در حالت پیش‌فرض در نظر گرفته و در نهایت، نتایج بر اساس اعتبارسنجی متقاطع ۱۰ برابری گزارش شده‌اند.^{۸۹} همچنین، در فرایند بهینه‌سازی هایپرپارامترها، نرخ تصادفی یکی از پارامترهای قابل‌تنظیم لحاظ شده است.

۳.۳.۱. نرمال‌سازی داده‌ها

فرایند پیش‌پردازش داده‌ها گامی ضروری در آماده‌سازی داده‌های خام برای استفاده در مدل‌های یادگیری ماشین است.^{۹۰} این مرحله معمولاً شامل پاک‌سازی داده‌ها، پر کردن داده‌های مفقود، تشخیص داده‌های پرت، کاهش نویز، و نرمال‌سازی است. در پژوهش حاضر، از آنجاکه داده‌ها از شبیه‌سازی‌های عددی حاصل شده و اندازه‌گیری میدانی با سنسور انجام نشده است، تنها نرمال‌سازی داده‌ها گام اصلی پیش‌پردازش قلمداد و به کار گرفته شده است.

نرمال‌سازی با هدف افزایش سرعت یادگیری و بهبود کارایی مدل‌های پیش‌بینی انجام می‌شود.^{۹۱} در این روش، داده‌های ورودی به بازه استاندارد — معمولاً بین ۰ تا ۱ یا بین ۱- تا ۱+ — تبدیل می‌شوند تا مانع از غلبه متغیرهای دارای مقیاس بزرگ‌تر بر مدل شود. نرمال‌سازی می‌تواند بر روی هر متغیر به‌صورت جداگانه یا گروهی اعمال شود و حذف آن ممکن است موجب کاهش دقت و طبقه‌بندی نادرست مدل گردد.^{۹۲}

مقیاس‌بندی

یکی از روش‌های نرمال‌سازی داده‌های ورودی برای ساختاربخشی به مجموعه داده‌ها و حفظ یکپارچگی آنها، مقیاس‌بندی MinMax است که بدین ترتیب داده‌های ورودی

36. F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Psychological Review, 65(6) (1958): 386-408.

37. Tzuc, et al., "Artificial Intelligence Techniques for Modeling Indoor Building Temperature under Tropical Climate Using Outdoor Environmental Monitoring".

38. SVR: Support Vector Regression

39. SVC: Support Vector Classification

40. Local Minimal

41. Qiong Li, et al., "Applying Support Vector Machine to Predict Hourly Cooling Load in the Building", Applied Energy, 86(10) (2009): 2249-2256; Qiong Li, et al., "Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks", Energy Conversion and Management, 50(1) (2009): 90-96.

۵.۳.۱. کدگذاری داده‌ها به روش One Hot Encoding (OHE)

برای تبدیل متغیرهای طبقه‌بندی شده به داده‌های عددی قابل استفاده در مدل‌های یادگیری ماشین، از روش One Hot Encoding استفاده می‌شود.^{۹۸} در این روش، به جای اختصاص یک مقدار عددی ترتیبی به دسته‌ها، برای هر دسته یک ستون مجزا ایجاد می‌شود تا از اعمال سوگیری ترتیبی در داده‌ها جلوگیری شود؛ به‌طور مثال، در یک متغیر دو حالتی مانند جنسیت، کدگذاری مستقیم با ۰ و ۱ می‌تواند به‌طور ضمنی نوعی برتری برای یکی از دسته‌ها ایجاد کند، درحالی‌که One Hot Encoding هر دو دسته را به‌صورت مستقل و هم‌ارز نمایش می‌دهد.^{۹۹}

با وجود مزیت افزایش شفافیت برای مدل، این روش می‌تواند موجب افزایش ابعاد داده‌ها، ایجاد ماتریس‌های پراکنده، افزایش پیچیدگی محاسباتی، و در برخی موارد بیش‌برازش شود؛ به‌ویژه زمانی که تعداد دسته‌ها زیاد و حجم نمونه محدود باشد. از این‌رو، استفاده از آن باید با توجه به ساختار داده و نوع مدل انجام شود. در این پژوهش، به‌منظور مقایسه دقت پیش‌بینی، هر دو روش Label Encoding و One Hot Encoding برای تبدیل متغیر طبقه‌بندی «کاربر» به مقادیر عددی به کار گرفته شده‌اند.

۴.۱. ارزیابی مدل (Evaluation Model)

عملکرد مدل‌های رگرسیونی و طبقه‌بندی بر اساس معیارهای استاندارد و مقایسه مقادیر واقعی (T) و پیش‌بینی شده (P) ارزیابی شد.

در مدل‌های رگرسیونی، از شاخص‌های MAE ، MSE ، $RMSE$ و $MAPE$ استفاده شد. R^2 و $RMSE$ نشان‌دهنده کیفیت تخمین و مجذور خطاست، درحالی‌که

42. Zhitong Ma, et al., "Applying Support Vector Machines to Predict Building Energy Consumption in China", Energy Procedia, 152 (2018): 780-786; R.K. Jain, et al., "Forecasting Energy Consumption of Multi-Family Residential Buildings Using Support Vector Regression: Investigating the Impact of Temporal and Spatial Monitoring Granularity on Performance Accuracy", Applied Energy, 123 (2014): 168-178; Hyun Chul Jung, et al., "Prediction of Building Energy Consumption Using an Improved Real Coded Genetic Algorithm Based Least Squares Support Vector Machine Approach", Energy and Buildings, 90 (2015): 76-84; M. Shao, et al., "Prediction of Energy Consumption in Hotel Buildings via Support Vector Machines", Sustainable Cities and Society, 57 (2020): 102128.

43. Xiaoli Li, et al., "Classification of Energy Consumption in Buildings with Outlier Detection", IEEE Transactions on Industrial Electronics, 57(11) (2009): 3639-3644.

MAE و $MAPE$ به‌ترتیب میانگین خطای مطلق و درصد خطا را تعیین می‌کنند. ضریب تعیین R^2 نیز کیفیت برازش مدل را (از ۰ تا ۱) نشان می‌دهد.

در مدل‌های طبقه‌بندی، با استفاده از ماتریس درهم‌ریختگی (Confusion Matrix) و تعیین مقادیر TN ، FP ، TP و FN ، چهار معیار اصلی محاسبه شد:

Precision: نسبت پیش‌بینی‌های صحیح مثبت به کل پیش‌بینی‌های مثبت،

Recall: نسبت نمونه‌های مثبت به‌درستی شناسایی شده به کل نمونه‌های مثبت واقعی،

Accuracy: نرخ دقت کلی در دسته‌بندی صحیح کلاس‌ها، F-Score: میانگین هارمونیک بین Precision و Recall برای سنجش پایداری مدل.

در این پژوهش این معیارها برای مقایسه و اعتبارسنجی دقیق عملکرد مدل‌های یادگیری ماشین به کار گرفته شدند.

۵.۱. تطبیق بیش از حد (Overfit)

برای هر روش پیش‌بینی یادگیری ماشینی، تطبیق بیش از حد یک نگرانی مهم است.^{۹۱} این مقوله به معنای دقت بالای مدل در نمونه‌های آموزشی است، درحالی‌که الگوی تخمینی احتمالاً به نقاط داده جدید تعمیم داده نمی‌شود و به این دلیل اتفاق می‌افتد که یک مدل دارای تطبیق بیش از حد، به جای شناسایی الگوی کلی رابطه ورودی / خروجی، با ویژگی‌های نویز موجود در داده‌های آموزشی مطابقت دارد. تطبیق بیش از حد معمولاً نقطه ضعف یک مدل بیش از حد پیچیده است؛ به‌طور مثال در مورد یک مدل ماشین تقویت‌گرایان، اگر کاربر تعداد بسیار زیادی تکرار K و عمق بسیار زیادی از درخت‌های تصمیم d را انتخاب کند، ممکن است تطبیق بیش از حد اتفاق بیفتد.

44. T. Ahmad and H. Chen, "Short and Medium-Term Forecasting of Cooling and Heating Load Demand in Building Environment with Data-Mining Based Approaches", Energy and Buildings, 166 (2018): 460-476; M.W. Ahmad, et al., "Trees vs Neurons: Comparison between Random Forest and ANN for High-Resolution Prediction of Building Energy Consumption", Energy and Buildings, 147 (2017): 77-89; Jui-Sheng Chou and Dac-Khuong Bui, "Modeling Heating and Cooling Loads by Artificial Intelligence for Energy-Efficient Building Design", Energy and Buildings, 82 (2014): 437-446.

45. Richard De Dear and Gail Schiller Brager, "The Adaptive Model of Thermal Comfort and Energy Conservation in the Built Environment", International Journal of Biometeorology, 45 (2001): 100-108.

46. H. Peng, et al., "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8) (2005): 1226-1238.

داده شد. اگرچه مدل‌های یادگیری ماشین، اغلب در کارهای تحقیقاتی قبلی مورد بحث بوده‌اند و الگوریتم‌ها به‌صورت مجزا یا با هدف مقایسه تعداد محدودی از الگوریتم‌ها بررسی شدند، اما در هیچ تحقیقی به مقایسه جامع الگوریتم‌ها پرداخته نشده است، بنابراین یک طرح مقایسه‌ای جامع بین الگوریتم‌های مختلف یادگیری انفرادی و تقویتی ضروری به‌نظر می‌رسد. در «جدول ۱» خلاصه‌ای از یافته‌های نمونه‌ای از مطالعات پژوهشی با رویکردها، ویژگی‌ها، و پارامترهای مختلف یادگیری ماشین آمده است.

۲. روش تحقیق

مجموعه داده‌های پیش‌بینی‌کننده در این پژوهش یک پایگاه داده مصنوعی است که برای آموزش مدل‌ها استفاده می‌شود تا مبنایی ثابت برای تحقیق، توسعه مدل انرژی، و تعیین معیارها فراهم گردد و می‌توان آن را بسیار سریع‌تر، در چند ثانیه، به جای سال‌ها تولید کرد. همان‌طور که در پیش‌تر ذکر گردید، این مجموعه داده از شبیه‌سازی مدل با استفاده از EnergyPlus به‌دست آمده است. EnergyPlus مدل‌سازی انرژی ساختمان را انجام می‌دهد و از شبیه‌سازی مبتنی بر رایانه برای تجزیه و تحلیل دقیق مصرف انرژی استفاده می‌شود. ویژگی‌های موجود در مجموعه داده (متغیرهای پیش‌بینی‌کننده) عبارتند از: جهت‌گیری ساختمان، نسبت پنجره به دیوار، ابعاد ساختمان شامل طول، عرض، ارتفاع، و کاربری ساختمان که با استفاده از نرم‌افزار EnergyPlus به‌صورت پارامتریک تعریف شده است (جدول ۲). متغیرهای هدف موجود در مجموعه داده نیز شامل انرژی سرمایشی و گرمایشی ساختمان می‌شود.

۶.۱. اعتبارسنجی متقاطع (Cross Validation-CV)

کمبود تعداد نقاط آموزشی ممکن است تخمین ضعیفی از رابطه ورودی / خروجی ایجاد کند. اعتبارسنجی متقاطع، و به‌ویژه اعتبارسنجی متقاطع K برابری^{۹۲}، سنتی‌ترین روش برای غلبه بر کمبود داده است.^{۹۳} روش اعتبارسنجی متقاطع K برابری شامل تقسیم تصادفی مجموعه داده‌های آموزشی به K زیرنمونه، با اندازه تقریباً مساوی است. مدل یادگیری ابتدا با استفاده از ریزنمونه اول و سپس با ریزنمونه دوم و همچنان تا ریزنمونه Kام با عنوان مجموعه داده آموزشی تخمین زده می‌شود و از باقی ریزنمونه‌ها در هر مجموعه آزمایشی برای تخمین متریک دقت پیش‌بینی استفاده خواهد شد و در نهایت میانگین متریک مورد نظر در K مجموعه انجام‌شده مدل یادگیری، با عنوان متریک نهایی معرفی می‌شود. اگرچه هیچ قانون رسمی برای انتخاب مقدار K وجود ندارد، اما در عمل از $K = 5$ یا $K = 10$ استفاده می‌شود.^{۹۴} هرچه K بیشتر باشد، رویکرد از نظر محاسباتی قوی‌تر است، در این پژوهش از اعتبارسنجی متقاطع ۱۰ برابری استفاده شده است.

۷.۱. تنظیم هاپیرپارامتر (Tuning)

فراپارامترها متغیرهای مشخص‌کننده ساختار الگوریتم و نحوه آموزش مدل هستند (مانند نرخ یادگیری، عمق درخت، تعداد همسایگان، و ...). هاپیرپارامترها قبل از آموزش، به‌صورت پیش‌فرض مقداردهی اولیه می‌شوند و سپس با بهینه کردن فراپارامترهای هر الگوریتم می‌توان بر دقت مدل پیش‌بینی افزود.^{۹۵}

تا به اینجا انگیزه و ضرورت استفاده از یادگیری ماشین در تخمین انرژی ساختمان، روش‌های مختلف یادگیری ماشین، نحوه ارزیابی مدل پیش‌بینی، اهمیت متغیرهای پیش‌بینی، پیش‌پردازش آنها، و اعتبارسنجی مدل یادگیری مفصل توضیح

منبع	موضوع	الگوریتم یادگیری ماشین									
		الگوریتم انفرادی								الگوریتم تقویتی	
		رگرسیون خطی	رگرسیون لجستیک	رگرسیون گاوسی	نزدیکترین همسایه K	درخت تصمیم	شبکه عصبی	پردازش تفریق	جنگل تصادفی	AdaBoost	GradientBoost
Catalina, et al., "Multiple Regression Model for Fast Prediction of the Heating Energy Demand".	پیش‌بینی تقاضای گرمایشی ساختمان‌های مسکونی	✓									
Antanasijević, et al., "Modeling of Energy Consumption and Related GHG (Greenhouse Gas) Intensity and Emissions in Europe Using General Regression Neural Networks".	پیش‌بینی مصرف انرژی و انتشار گازهای گلخانه‌ای مرتبط با انرژی	✓					✓				
Rastogi, et al., "Gaussian-Process-Based Emulators for Building Performance Simulation".	شبیه‌سازی انرژی ساختمان‌های با کارایی بالا	✓		✓							
Seyedzadeh, et al., "Machine Learning for Estimation of Building Energy Consumption and Performance: A Review".	مروری بر الگوریتم‌های یادگیری ماشین در پیش‌بینی مصرف انرژی ساختمان			✓	✓		✓	✓			
Touzani, et al., "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings".	مدل‌سازی پایه مصرف انرژی بر اساس یک ماشین تقویت گرادیان	✓							✓		✓
Chowdhury, et al., "Energy Consumption Prediction Using Light Gradient Boosting Machine Model".	پیش‌بینی مصرف انرژی در یک طرح مقایسه‌ای					✓					✓
Dong, et al., "Hourly Energy Consumption Prediction of an Office Building Based on Ensemble Learning and Energy Consumption Pattern Classification".	پیش‌بینی مصرف انرژی ساختمان بر اساس یادگیری گروهی از الگوریتم‌های انفرادی	✓				✓	✓	✓			
Fayyaz, et al., "Thermal Comfort Model for HVAC Buildings Using Machine Learning".	پیش‌بینی احساس حرارتی افراد		✓				✓	✓	✓	✓	
پژوهش حاضر	پیش‌بینی بار مصرفی ساختمان در یک طرح مقایسه‌ای جامع	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

47. R. Mikučionienė, et al., "Evaluation of Energy Efficiency Measures Sustainability by Decision Tree Method", Energy and Buildings, 76 (2014): 64-71.
48. J. Liu, et al., "An Energy Performance Evaluation Methodology for Individual Office Building with Dynamic Energy Benchmarks Using Limited Information", Applied Energy, 206 (2017): 193-205.
49. Dong, et al., "Hourly Energy Consumption Prediction of an Office Building Based on Ensemble Learning and Energy Consumption Pattern Classification", 110929; J. Silva, et al., "Energy Consumption Forecasting Using Ensemble Learning Algorithms", Advances in Intelligent Systems and Computing, 1004 (2020): 5-13; Z. Wang, et al., "A Novel Ensemble Learning Approach to Support Building Energy Use Prediction", Energy and Buildings, 159 (2018): 109-122; T.G. Dietterich, "Ensemble Methods in Machine Learning", in International Workshop on Multiple Classifier Systems (Springer, 2000), 1-15.

جدول ۱ (بالا). خلاصه‌ای از تحقیقات در زمینه استفاده از الگوریتم‌های مختلف یادگیری ماشین، تدوین: نگارنده.

جدول ۲ (پایین). مقادیر پارامتریک متغیرهای پیش‌بینی‌کننده در نرم‌افزار Eplus، تدوین: نگارنده.

آستانه تغییر	متغیر پیش‌بینی‌کننده
۰ و ۴۵ و ۹۰ و ۱۳۵ و ۱۸۰ و ۲۲۵ و ۲۷۰ و ۳۱۵	جهت‌گیری ساختمان
۲/۷ و ۵/۴	ارتفاع ساختمان
۳×۱۶ و ۶×۸ و ۱۲×۱۶	ابعاد ساختمان
مسکونی، اداری، آموزشی، تجاری	کاربری ساختمان
۳۰٪ و ۵۰٪ و ۷۰٪	نسبت پنجره به دیوار

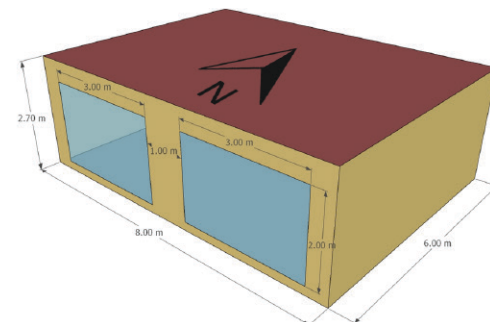
۱.۲. مورد پژوهش

نمونه ۶۰۰ برای اعتبارسنجی مدل ANSI / ASHRAE استاندارد ۱۴۰^۶ به‌طور گسترده در شبیه‌سازی ساختمان استفاده می‌شود. در این مطالعه، نمونه ۶۰۰ برای انجام شبیه‌سازی‌ها عرضه شده است که یک زون مکعبی (۶ × ۸ × ۲/۷ متر) با دو پنجره (هرکدام ۳ × ۲ متر) در دیوار جنوبی است (جهت‌گیری مبنا = ۰) که بنابه نسبت پنجره به دیوار، ابعاد پنجره‌ها تغییر کرده است (ت ۲).



در مرحله بعد، داده‌های حاصل از شبیه‌سازی که شامل ۵۷۶ حالت مختلف است، وارد مرجع کد پایتون ۳٫۹ شد، عملیات تقسیم داده‌ها به دو دسته آموزشی و آزمایشی با نسبت ۳۰٪ : ۷۰٪ و پیش‌پردازش آنها صورت گرفت (نرمال‌سازی داده‌ها، تبدیل ستون «کاربری ساختمان» به داده‌های عددی به هر دو روش ذکرشده در بخش‌های ۴٫۳، ۱ و ۵٫۳). سپس مدل یادگیری ماشین با استفاده از الگوریتم‌های مختلف پیش‌بینی اعمال شد. درنهایت معیارهای ارزیابی متناسب با اهداف رگرسیونی یا طبقه‌بندی، به‌منظور دریافت دقت پیش‌بینی محاسبه شد. همچنین برای اعتبارسنجی نتایج پیش‌بینی، از اعتبارسنجی متقاطع ۱۰ (برابری در مرحله آخر استفاده شد. قابل‌ذکر است که در ابتدا از الگوریتم‌های یادگیری با هاپیر پارامترهای پیش‌فرض استفاده شد و سپس برای رسیدن به بهترین دقت پیش‌بینی در هر الگوریتم ماشین، هاپیر پارامترهای هر کدام بهینه‌سازی شد.

برای استفاده از الگوریتم‌های طبقه‌بندی، متغیر هدف (انرژی گرمایشی و سرمایشی ساختمان) به داده‌های باینری تبدیل شد تا قابلیت پیش‌بینی کلاس هر متغیر فراهم شود. به این صورت که میانگین هریک از متغیرهای هدف در کل داده‌ها محاسبه شد (میانگین انرژی گرمایشی $98/41 \text{ kWh/m}^2$ و میانگین انرژی سرمایشی $143/07 \text{ kWh/m}^2$). مقادیر کمتر از میانگین صفر، و مقادیر بیشتر از میانگین ۱ قلمداد شدند. این



ت ۲. هندسه مدل مورد نظر در حالت پایه (Ashrae).

مقادیر U عناصر پوسته ساختمان مطابق با اقلیم شهر تهران، برنامه زمان بندی حضور افراد، تجهیزات، روشنایی، و دمای تنظیم گرمایش و سرمایش مطابق مبحث ۱۹ مقررات ملی ساختمان در ایران و متناسب با کاربری مورد نظر است (جدول ۳)^{۹۷}، در سطح گسترده‌تر از «جدول ۳» و مرتبط با مبحث ۱۹، که در این مقاله نیامده، مطالب دیگری نیز مورد نظر بوده است؛ به‌طور مثال سطح لباس افراد بر اساس مدل لباس پویا ASHRAE 55 محاسبه شده است.^{۹۸} نرخ هوای تازه واردشده به اتاق، مطابق مبحث ۱۹ مقررات ملی ساختمان، به ازای هر ساعت، ۱ در نظر گرفته شده است. روشنایی مصنوعی در زون مورد نظر زمانی فعال می‌شود که سطح روشنایی طبیعی در ارتفاع میز کار (۸۰ سانتی‌متر) کمتر از ۵۰ لوکس باشد. به‌منظور جلوگیری از اتلاف محاسباتی، از سیستم مکانیکی ایدئال استفاده شده است.

بارهای داخلی				ضریب انتقال حرارت در جداره غیرنورگذر (U_{W/m^2K})		
تجهیزات/مساحت (w/m^2)	ست پوینت گرمایشی ($^{\circ}C$)	ست پوینت سرمایشی ($^{\circ}C$)	مساحت (m^2) / تعداد افراد (person)	ضریب بهره گرمایی خورشیدی پنجره (SHGC)	کف	سقف دیوار
۴	۲۰	۲۸	۴۲٫۵	۰٫۴	۰٫۵	۱٫۸
۱۴	۳۲-۳۰-۲۸	۱۵-۱۷-۲۰	۱۰	۰٫۵	۰٫۵	۱٫۲
۴	۳۲-۳۰-۲۸	۱۵-۱۷-۲۰	۵٫۴	۰٫۵	۰٫۵	۱٫۲
۵	۳۲-۳۰-۲۸	۱۵-۱۷-۲۰	۱۷	۰٫۵	۰٫۵	۱٫۲

جدول ۳. مشخصات پوسته ساختمان مطابق با مبحث ۱۹ مقررات ملی ساختمان.

50. C. Miller, et al., "The ASHRAE Great Energy Predictor III Competition", Science and Technology for the Built Environment, 26(10) (2020): 1427-1447.

51. D.B. Araya, et al., "An Ensemble Learning Framework for Anomaly Detection in Building Energy Consumption", Energy and Buildings, 144 (2017): 191-206.

52. Ahmad and Rezgui, "Trees vs Neurons: Comparison between Random Forest and ANN for High-Resolution Prediction of Building Energy Consumption".

53. Bootstrap

54. He, Handb. Quantile Regres.

55. R.E. Schapire, "The Strength of Weak Learnability", Machine Learning, 5 (1990): 197-227; Y. Freund, "Boosting a Weak Learning Algorithm by Majority", Information and Computation, 121(2) (1995): 256-285; Y. Freund and R.E. Schapire, "Experiments with a New Boosting Algorithm", in Icml, vol. 96 (Citeseer, 1996), 148-156; J. Friedman, et al., "Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors)", The Annals of Statistics, 28(2) (2000): 337-407; S. Touzani, et al., "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings", Energy and Buildings, 158 (2018): 1533-1543.

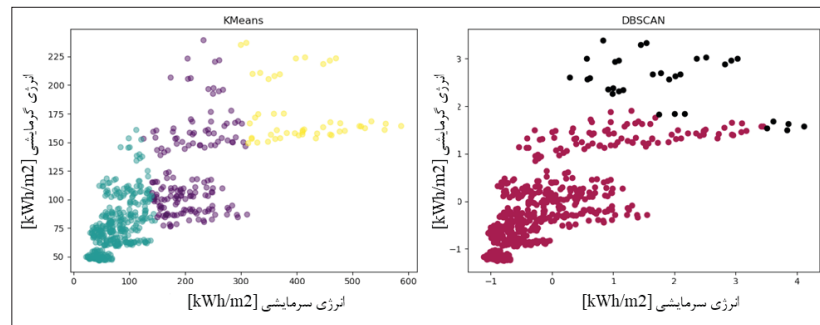
الگوی طبقه‌بندی پیشنهادی در شهر تهران برای پیش‌بینی ساختمان‌های پرمصرف (۱) و کم‌مصرف (۰) می‌شوند. مشابه این الگو در پژوهش فرلگ و همکاران^{۹۹} برای طبقه‌بندی بارمصرفی ساختمان (سرمایشی / گرمایشی) بر اساس نقطه تنظیم ترموستات استفاده شده است.

۳. نتایج و تحلیل‌ها

۳.۱. خوشه‌بندی داده‌ها

به‌منظور درک ساختار داده‌های بدون برچسب حاصل از شبیه‌سازی و کاهش پیچیدگی آنها، پیش از اجرای مدل‌های یادگیری بانظارت، از روش‌های خوشه‌بندی استفاده شد. در این مرحله، همه ۵۷۶ سناریوی شبیه‌سازی شده بدون در نظر گرفتن روابط علی، در سه گروه با سطوح مصرف انرژی «کم، متوسط، و زیاد» دسته‌بندی شدند. این تحلیل با دو الگوریتم انجام شد: ۱) K-means: با استفاده از سه خوشه و ۱۲ تکرار (به روش آزمون و خطا)، داده‌ها گروه‌بندی شدند، نتایج نشان داد که هیچ داده پرتی^{۱۰۰} در این روش شناسایی نمی‌شود؛ ۲) DBSCAN: با تنظیم شعاع Epsilon روی ۰/۳ و حداقل ۷ همسایه (به‌روش آزمون و خطا)، خوشه‌بندی تکرار شد که در این مدل، برخی داده‌ها همان «نویز» هستند (ت ۳).

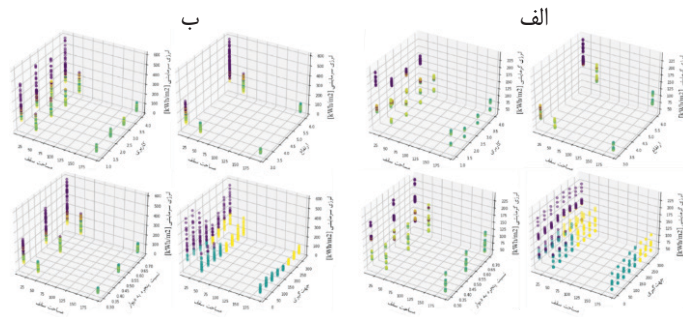
ت ۳. خوشه‌بندی داده‌ها به روش K-means و DBSCAN، تدوین: نگارنده.



تفاوت در شناسایی داده‌های پرت میان این دو روش به درک دقیق‌تری از توزیع داده‌ها می‌انجامد. در مراحل آتی پژوهش، اثر این دسته‌بندی‌ها بر سایر متغیرهای کالبدی و عملکردی ساختمان بررسی خواهد شد.

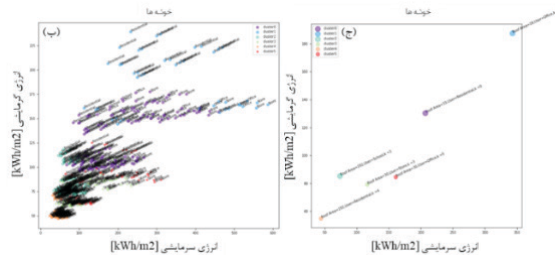
اگرچه متغیرهای پیش‌بینی‌کننده از پیش و بر اساس طراحی پارامتریک تعیین شده‌اند، رابطه میان متغیر هدف یعنی انرژی‌های گرمایشی و سرمایشی ساختمان با سایر ویژگی‌ها از طریق خوشه‌بندی K-means قابل تحلیل است. برطبق نتایج در «ت ۴-الف»، ساختمان‌های تجاری با مساحت کمتر، ارتفاع بیشتر، نسبت پنجره به دیوار بالاتر، و جهت‌گیری جنوبی و جنوب شرقی و جنوب غربی عمدتاً در گروه ساختمان‌های با انرژی سرمایشی زیاد هستند؛ در مقابل، ساختمان‌های مسکونی با مساحت بیشتر، ارتفاع کمتر، نسبت پنجره به دیوار کمتر، و جهت‌گیری شمالی در گروه انرژی سرمایشی کم قرار دارند. همچنین، بر اساس «ت ۴-ب»، ساختمان‌های مسکونی با مساحت کمتر، ارتفاع بیشتر، نسبت پنجره به دیوار بالاتر، و جهت‌گیری شمال و شمال شرقی و شمال غربی در گروه ساختمان‌های با انرژی گرمایشی زیاد هستند، اما ساختمان‌های اداری و آموزشی با مساحت بیشتر، ارتفاع کمتر، نسبت پنجره به

ت ۴. خوشه‌بندی داده‌ها به روش K-means و ارتباط بین سایر متغیرها با (الف) انرژی سرمایشی، (ب) انرژی گرمایشی، تدوین: نگارنده.



ت ۵.

الف) نمودار دندروگرام در خوشه‌بندی سلسله‌مراتبی،
 ب) برش نمودار دندروگرام در سطح ۰.۶
 ج) میانگین مقادیر خوشه‌ها بعد از برش، تدوین: نگارنده.

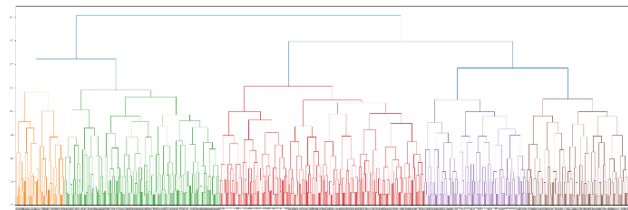


سرمایشی و گرمایشی و ماتریس پیچیدگی و متریک‌های مربوط به آن فقط برای حالت طبقه‌بندی انرژی‌های سرمایشی و گرمایشی محاسبه شد که به ترتیب در «جدول ۴» مشاهده می‌شود.

بر اساس نتایج موجود در «جدول ۵»، در پیش‌بینی بارهای انرژی ساختمان، مدل شبکه عصبی چندلایه (MLP) تنظیم‌شده بهترین عملکرد را در میان الگوریتم‌های رگرسیونی داشت، درحالی‌که در دسته‌بندی، الگوریتم جنگل تصادفی تنظیم‌شده بیشترین دقت را نشان داد. به‌طور کلی، الگوریتم‌های طبقه‌بندی عملکرد بهتری نسبت به مدل‌های رگرسیونی نشان دادند. با توجه به تصادفی بودن تقسیم داده‌ها، از اعتبارسنجی متقاطع ۱۰برابری ($K = 10$) برای ارزیابی استفاده شد.

تنظیم فرایامترها در پیش‌بینی انرژی سرمایشی به‌ترتیب برای الگوریتم‌های شبکه عصبی چندلایه (MLP)، k-نزدیک‌ترین همسایه (KNN)، درخت تصمیم (DT)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF)، منجر به افزایش میانگین $0.4/0.56$ ، $0.4/0.4$ ، $0.9/1.4$ و $2/2$ درصدی معیارهای R^2 و Accuracy و در هر دو روش پیش‌پردازش گردید. در پیش‌بینی انرژی گرمایشی، این بهبود به‌ترتیب $1/2.9$ ، $1/6$ ، $8/4$ و $5/0.8$ درصد بود. بیشترین تأثیر تنظیم در هر دو حالت مربوط به MLP و کمترین تأثیر مربوط به DT (در سرمایش) و RF (در گرمایش) بوده است. همچنین، اثر تنظیم بر SVM، MLP، RF در انرژی سرمایشی بیشتر از گرمایشی بود.

الف



دیوار کمتر، و جهت‌گیری جنوبی و جنوب شرقی و جنوب غربی در گروه انرژی گرمایشی کم هستند.

از خوشه‌بندی سلسله‌مراتبی نیز، نمودار دندروگرام «ت ۵-الف» و بعد از انتخاب سطح شباهت ۰.۶ برای برش نمودار دندروگرام، گروه‌بندی داده‌ها با جزئیات بیشتر به‌صورت «ت ۵-ب» و بعد از آن میانگین‌گیری در هر خوشه و گروه‌بندی کلی داده‌ها در «ت ۵-ج» مشاهده می‌شود.

۲.۳. دقت مدل پیش‌بینی در الگوریتم‌های انفرادی یادگیری ماشین

در این پژوهش ابتدا الگوریتم‌های یادگیری ماشین به دو دسته رگرسیونی و طبقه‌بندی تقسیم شدند. پس از تقسیم مدل به داده‌های آموزشی و آزمایشی به‌صورت ۳۰٪ و ۷۰٪، از دو روش Label Encoding و One Hot Encoding برای تبدیل ستون «کاربر» به داده‌های عددی استفاده شده است (مسکونی = ۰، اداری = ۱، آموزشی = ۲ و تجاری = ۳ در روش Label Encoding) و سپس همه داده‌ها با استفاده از مقیاس استاندارد، مقیاس‌بندی شد. ابتدا با هاپرپارامترهای پیش‌فرض هر الگوریتم و سپس در حالت تنظیم‌شده با هاپرپارامترهای بهینه، ارزیابی مدل پیش‌بینی انجام شد و درنهایت، پس از اعتبارسنجی متقاطع ۱۰برابری، دقت پیش‌بینی مدل‌ها از طریق معیارهای ارزیابی، مجدداً مقایسه شد. شاخص‌های خطا شامل MAPE، MAE، RMSE و R^2 برای حالت رگرسیونی انرژی

56. Bassi et al., "Building Energy Consumption Forecasting", 1-9; Seyedzadeh, et al., "Tuning Machine Learning Models for Prediction of Building Energy Loads", Sustainable Cities and Society, 47 (2019): 101484; Z. Wang, et al., "Building Thermal Load Prediction through Shallow Machine Learning and Deep Learning", Applied Energy, 263 (2020): 114683; Cheng Fan, et al., "A Short-Term Building Cooling Load Prediction Method Using Deep Learning Algorithms", Applied Energy, 195 (2017): 222-233; M. Al-Rakhami, et al., "An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings", IEEE Access, 7 (2019): 48328-48338.

57. S.R. Chowdhury, et al., “Energy Consumption Prediction Using Light Gradient Boosting Machine Model”, in International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy (Springer, 2020), 413-422; S. Liang et al., “Energy Consumption Prediction Using the GRU-MMAAttention-LightGBM Model with Features of Prophet Decomposition”, PloS One, 18(1) (2023): e0277085; RuoChen Zhao, et al., “Building Cooling Load Prediction Based on Lightgbm”, IFAC-PapersOnLine, 55(11) (2022): 114-119; Z. Yin, et al., “Pump Feature Construction and Electrical Energy Consumption Prediction Based on Feature Engineering and LightGBM Algorithm”, Sustainability, 15(1) (2023): ←

جدول ۴. ارزیابی مدل‌های یادگیری ماشین با دو روش One Hot و Label Encoding در پیش‌بینی انرژی سرمایشی و گرمایشی، تدوین: نگارنده.

انرژی سرمایشی									
رگرسیون	One Hot Encoding					Label Encoding			
	الگوریتم	R ²	RMSE	MAE	MAPE	R ²	RMSE	MAE	MAPE
	رگرسیون خطی	۰/۶۰	۶۸/۰۵	۵۲/۰۰	۰/۴۹	۰/۵۷	۷۰/۱۳	۵۳/۱۱	۰/۴۹
	رگرسیون گاوسی	۰/۹۳	۲۶/۹۷	۱۴/۹۹	۰/۱۱	۰/۹۸	۱۲/۶۳	۷/۷۷	۰/۰۵
	شبکه عصبی چندلایه	۰/۶۴	۶۳/۸۴	۴۸/۱۰	۰/۴۰	۰/۶۲	۶۶/۲۶	۵۰/۶۹	۰/۴۳
	شبکه عصبی چند لایه تنظیم شده	۰/۹۹	۸/۳۵	۶/۰۱	۰/۰۵	۰/۹۸	۱۲/۰۱	۸/۹۰	۰/۰۷
		Accuracy	Precision	Recall	F-Score	Accuracy	Precision	Recall	F-Score
	رگرسیون لجستیک	۰/۸۹	۰/۸۹	۰/۸۸	۰/۸۹	۰/۹۰	۰/۹۰	۰/۸۹	۰/۸۹
	K-نزدیک‌ترین همسایه	۰/۷۵	۰/۷۵	۰/۷۳	۰/۷۴	۰/۷۹	۰/۸۰	۰/۷۷	۰/۷۸
	K-نزدیک‌ترین همسایه تنظیم شده	۰/۷۹	۰/۸۱	۰/۷۷	۰/۷۷	۰/۸۲	۰/۸۴	۰/۸۰	۰/۸۱
طبقه‌بندی	درخت تصمیم	۰/۹۲	۰/۹۳	۰/۹۱	۰/۹۱	۰/۹۲	۰/۹۲	۰/۹۱	۰/۹۱
	درخت تصمیم تنظیم شده	۰/۹۲	۰/۹۳	۰/۹۲	۰/۹۲	۰/۹۲	۰/۹۳	۰/۹۲	۰/۹۲
	ماشین بردار پشتیبان	۰/۷۴	۰/۷۵	۰/۷۲	۰/۷۲	۰/۸۸	۰/۸۸	۰/۸۸	۰/۸۸
	ماشین بردار پشتیبان تنظیم شده	۰/۹۱	۰/۹۲	۰/۹۱	۰/۹۱	۰/۹۴	۰/۹۵	۰/۹۳	۰/۹۴
	جنگل‌های تصادفی	۰/۹۳	۰/۹۴	۰/۹۲	۰/۹۲	۰/۹۱	۰/۹۲	۰/۹۰	۰/۹۰
	جنگل‌های تصادفی تنظیم شده	۰/۹۴	۰/۹۴	۰/۹۳	۰/۹۳	۰/۹۴	۰/۹۴	۰/۹۳	۰/۹۳
	انرژی گرمایشی								
	رگرسیون	One Hot Encoding					Label Encoding		
الگوریتم		R ²	RMSE	MAE	MAPE	R ²	RMSE	MAE	MAPE
رگرسیون خطی		۰/۸۰	۱۶/۲۸	۱۴/۱۱	۰/۱۶	۰/۷۳	۱۹/۰۶	۱۵/۱۹	۰/۱۷
رگرسیون گاوسی		۰/۹۵	۷/۵۹	۳/۷۱	۰/۰۳	۰/۹۸	۴/۶۶	۲/۵۰	۰/۰۲
شبکه عصبی چندلایه		۰/۸۲	۱۵/۴۸	۱۲/۷۳	۰/۱۴	۰/۷۲	۱۹/۵۰	۱۴/۸۵	۰/۱۷
شبکه عصبی چند لایه تنظیم شده		۰/۹۹	۱/۵۲	۱/۱۵	۰/۰۱	۰/۹۹	۱/۲۸	۰/۸۷	۰/۰۰
		Accuracy	Precision	Recall	F-Score	Accuracy	Precision	Recall	F-Score
رگرسیون لجستیک		۰/۹۸	۰/۹۸	۰/۹۹	۰/۹۸	۰/۹۴	۰/۹۵	۰/۹۲	۰/۹۳
K-نزدیک‌ترین همسایه		۰/۹۱	۰/۹۱	۰/۹۲	۰/۹۱	۰/۸۹	۰/۸۸	۰/۸۹	۰/۸۸
K-نزدیک‌ترین همسایه تنظیم شده		۰/۹۷	۰/۹۶	۰/۹۷	۰/۹۶	۰/۹۴	۰/۹۴	۰/۹۲	۰/۹۳
طبقه‌بندی	درخت تصمیم	۰/۹۷	۰/۹۶	۰/۹۷	۰/۹۶	۰/۹۶	۰/۹۵	۰/۹۶	۰/۹۶
	درخت تصمیم تنظیم شده	۰/۹۹	۰/۹۹	۱/۰۰	۰/۹۹	۰/۹۹	۰/۹۹	۱/۰۰	۰/۹۹
	ماشین بردار پشتیبان	۰/۹۵	۰/۹۴	۰/۹۵	۰/۹۴	۰/۹۴	۰/۹۵	۰/۹۲	۰/۹۳
	ماشین بردار پشتیبان تنظیم شده	۰/۹۹	۰/۹۹	۱/۰۰	۰/۹۹	۰/۹۹	۰/۹۹	۱/۰۰	۰/۹۹
	جنگل‌های تصادفی	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۹۹	۰/۹۹	۱/۰۰	۰/۹۹
	جنگل‌های تصادفی تنظیم شده	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰

ریشه (Gini = ۰/۴۵، Entropy = ۰/۹۲۷) کاهش چشمگیری داشت که نشان‌دهنده طبقه‌بندی مؤثر داده‌هاست. متغیرهای «مساحت سقف» و «ارتفاع» به‌ترتیب برای پیش‌بینی انرژی سرمایشی و گرمایشی بیشترین بهره اطلاعاتی را داشتند و در گره ریشه قرار گرفتند؛ به‌طور مثال، در زونی با مساحت سقف کمتر از ۳۰ متر مربع، ارتفاع کمتر از ۴/۵ متر، WWR کمتر از ۴۰٪ و کاربری مسکونی / اداری / آموزشی، انرژی سرمایشی در طبقه پایین‌تر از میانگین (۱۴۳/۰۷ کیلووات‌ساعت بر متر مربع) قرار می‌گیرد.

۳.۳. اعتبارسنجی متقاطع (Cross validation)

به‌منظور اعتبارسنجی نتایج «جدول ۴»، اعتبارسنجی متقاطع ۱۰ برابری انجام شد تا میزان خطای پیش‌بینی قابل‌اعتماد باشد که نتایج بهترین الگوریتم‌ها در «جدول ۵» قابل‌مشاهده است. با توجه به اینکه در اعتبارسنجی متقاطع، میانگین ۱۰ بار تکرار هر کدام از مدل‌های پیش‌بینی که داده‌های آموزشی و آزمایشی متفاوتی دارند، محاسبه می‌شود، بنابراین نتایج «جدول ۵»

در پیش‌بینی انرژی سرمایشی، روش One Hot Encoding (OHE) دقت رگرسیون خطی، شبکه عصبی چندلایه (پایه و تنظیم‌شده)، و جنگل تصادفی را افزایش داد، اما در سایر مدل‌ها روش Label Encoding (LE) عملکرد بهتری داشت. در الگوریتم‌های درخت تصمیم (تنظیم‌شده) و جنگل تصادفی (تنظیم‌شده) تغییر روش پیش‌پردازش تأثیری بر دقت نداشت. در پیش‌بینی انرژی گرمایشی، تنها رگرسیون گاوسی با OHE کاهش دقت نسبت به LE نشان داد. به‌جز شبکه عصبی چندلایه تنظیم‌شده، درخت تصمیم تنظیم‌شده، ماشین بردار پشتیبان تنظیم‌شده، و جنگل تصادفی تنظیم‌شده که تغییری نداشتند، سایر الگوریتم‌ها با OHE عملکرد بهتری داشتند. بنابراین، برای پیش‌بینی انرژی گرمایشی روش One Hot Encoding و برای انرژی سرمایشی روش Label Encoding توصیه می‌شود.

نمودار درختی مدل تنظیم‌شده با معیارهای بهره اطلاعات (IG) و بهره اطلاعات ترسیم شد. حداکثر ناخالصی در گره‌های برگ (Gini = ۰/۱۴۲، Entropy = ۰/۳۹۱) در مقایسه با گره

→

art 789; C. Bentéjac, et al.,
"A Comparative Analysis
of Gradient Boosting
Algorithms", Artificial
Intelligence Review, 54 (2021):
1937-1967.

جدول ۵. ارزیابی مدل‌های یادگیری ماشین انفرادی با دو روش One Label Encoding و Hot Encoding در پیش‌بینی انرژی سرمایشی و گرمایشی پس از اعتبارسنجی متقاطع ۱۰ برابری، تدوین: نگارنده.

انرژی سرمایشی									
رگرسیون	One Hot Encoding					Label Encoding			
	الگوریتم	R ²	RMSE	MAE	MAPE	R ²	RMSE	MAE	MAPE
	شبکه عصبی چند لایه تنظیم شده	۰/۹۴	۲۰/۶۵	۱۴/۶۵	۰/۱۱	۰/۹۷	۱۵/۳۴	۱۱/۰۶	۰/۰۸
طبقه‌بندی	Accuracy Precision Recall F-Score					Accuracy Precision Recall F-Score			
	رگرسیون لجستیک	۰/۹۱	۰/۹۱	۰/۸۶	۰/۸۷	۰/۹۱	۰/۹۱	۰/۹۰	۰/۸۹
	درخت تصمیم تنظیم شده	۰/۹۴	۰/۹۲	۰/۹۲	۰/۹۲	۰/۹۱	۰/۸۸	۰/۹۳	۰/۸۹
انرژی گرمایشی									
رگرسیون	الگوریتم	R ²	RMSE	MAE	MAPE	R ²	RMSE	MAE	MAPE
	شبکه عصبی چند لایه تنظیم شده	۰/۹۹	۲/۶۲	۲/۰۳	۰/۰۲	۰/۹۷	۲/۷۸	۲/۰۲	۰/۰۲
	Accuracy Precision Recall F-Score					Accuracy Precision Recall F-Score			
طبقه‌بندی	درخت تصمیم تنظیم شده	۰/۹۸	۰/۹۶	۰/۹۹	۰/۹۸	۰/۹۸	۰/۹۶	۰/۹۹	۰/۹۸
	ماشین بردار پشتیبان تنظیم شده	۰/۹۸	۰/۹۶	۰/۹۹	۰/۹۸	۰/۹۸	۰/۹۵	۰/۹۹	۰/۹۷

۴.۳. الگوریتم‌های تقویت کننده

در مرحله بعد، تنها الگوریتم‌های تقویتی در حالت تنظیم شده، برای هر دو مسئله رگرسیونی و طبقه‌بندی، با استفاده از اعتبارسنجی متقاطع و پیش‌پردازش Label Encoding مقایسه شدند. نتایج در «جدول ۶» نشان می‌دهد که مدل‌های تقویتی در مقایسه با الگوریتم‌های انفرادی یادگیری ماشین، دقت بالاتری نشان می‌دهند. در میان مدل‌های رگرسیونی، LightGBM بهترین عملکرد را در پیش‌بینی انرژی سرمایشی و گرمایشی داشت. در بخش طبقه‌بندی نیز، AdaBoost و CatBoost برای پیش‌بینی انرژی سرمایشی، GradientBoost و CatBoost برای پیش‌بینی انرژی گرمایشی، بیشترین دقت را به‌دست آوردند. به‌طورکلی، اثر تنظیم فرآیندها در AdaBoost و سبس LightGBM، به‌ویژه در حالت رگرسیونی، از سایر مدل‌ها بیشتر بود. همچنین در هر دو مسئله سرمایش و گرمایش، مشتقات Gradient Boosting عملکرد بهتری نسبت به AdaBoost در حالت رگرسیونی نشان دادند. بالاین‌حال، در حالت طبقه‌بندی، این روند

معتبرتر از «جدول ۴» است و شبکه عصبی چندلایه تنظیم شده در پیش‌بینی انرژی سرمایشی ساختمان با ۰/۹۷ و ۰/۹۴، ضریب R^2 در هریک از روش‌های Label Encoding و One Hot Encoding به‌ترتیب، بهترین الگوریتم رگرسیونی است و در الگوریتم‌های طبقه‌بندی، رگرسیون لجستیک و درخت تصمیم تنظیم شده با ۰/۹۱ و ۰/۹۴ Accuracy به‌ترتیب در روش‌های ذکر شده، نیز از بهترین الگوریتم‌های پیش‌بینی انرژی سرمایشی هستند که نشان می‌دهد روش One Hot Encoding در پیش‌بینی انرژی سرمایشی به روش رگرسیونی اثر منفی و به روش طبقه‌بندی اثر مثبت می‌گذارد. در ارزیابی متقاطع مدل پیش‌بینی انرژی گرمایشی، شبکه عصبی چندلایه، درخت تصمیم و ماشین بردار پشتیبان (همگی از نوع تنظیم شده) بالاترین دقت پیش‌بینی را دارند. در غالب الگوریتم‌ها، اعمال روش One Hot Encoding تأثیر مثبت قابل‌توجهی بر افزایش دقت مدل پیش‌بینی دارد.

58. Overfitting
59. Yue Pan and Limao Zhang, "Data-Driven Estimation of Building Energy Consumption with Multi-Source Heterogeneous Data", Applied Energy, 268 (2020): 114965; Y.D. Qin, et al., "Exploring New Building Energy Saving Control Strategy Application under the Energy Internet of Things", in IOP Conference Series: Earth and Environmental Science, vol. 701 (IOP Publishing, 2021), 12060; Bentéjac, et al., "A Comparative Analysis of Gradient Boosting Algorithms".

جدول ۶. ارزیابی مدل‌های یادگیری ماشین تقویتی به روش Label Encoding در پیش‌بینی انرژی سرمایشی و گرمایشی با اعتبارسنجی متقاطع ۱۰ برابری، تدوین: نگارندگان.

	انرژی گرمایشی				انرژی سرمایشی			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
رگرسیون	۰/۰۹	۷/۵۹	۹/۴۷	۰/۹۳	۰/۲۷	۳۲/۶۴	۴۴/۲۳	۰/۷۴
	۰/۰۱	۱/۰۶	۱/۳۵	۰/۹۹	۰/۱۲	۱۵/۹۳	۲۱/۲۳	۰/۹۴
	۰/۰۰	۱/۰۹	۱/۵۰	۰/۹۹	۰/۰۸	۱۲/۹۰	۲۰/۱۱	۰/۹۵
	۰/۰۱	۰/۹۷	۱/۳	۰/۹۹	۰/۰۷	۸/۸۷	۱۲/۹۷	۰/۹۸
	۰/۰۰	۰/۵۹	۰/۷۷	۱/۰۰	۰/۰۷	۸/۳۱	۱۱/۲۷	۰/۹۸
	انرژی گرمایشی				انرژی سرمایشی			
	F-Score	Recall	Precision	Accuracy	F-Score	Recall	Precision	Accuracy
طبقه‌بندی	۰/۹۷	۰/۹۷	۰/۹۶	۰/۹۷	۰/۹۵	۰/۹۶	۰/۹۵	۰/۹۶
	۰/۹۸	۰/۹۹	۰/۹۶	۰/۹۸	۰/۹۰	۰/۹۳	۰/۸۸	۰/۹۲
	۰/۹۷	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۰	۰/۹۰	۰/۹۱	۰/۹۳
	۰/۹۸	۰/۹۹	۰/۹۶	۰/۹۸	۰/۹۱	۰/۹۲	۰/۹۲	۰/۹۴
	۰/۹۷	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۱	۰/۹۳	۰/۸۹	۰/۹۳

این پژوهش، بهینه‌ساز شبه‌نیوتنی (lbfgs) نسبت به «adam» کارایی بیشتری داشت. همچنین، استفاده از دو لایه پنهان با بیش از ۷ نرون، نتایج بهتری نسبت به تک‌لایه ارائه داد که در پژوهش سیدزاده^{۱۰۳} نیز یافت می‌شود.

مدل‌های تقویتی (Ensemble Boosting): در همه مدل‌های تقویتی (XGBoost, CatBoost, LightGBM, ...) پارامتر feature_fraction برای جلوگیری از بیش‌برازش روی ۰/۵ تنظیم شد. برخلاف برخی مستندات^{۱۰۴}، در این پژوهش نرخ یادگیری بالاتر منجر به نتایج بهتری شد. همچنین مشاهده گردید که افزایش مراحل تقویت (Boosting stages) به‌طور کلی عملکرد مدل را بهبود می‌بخشد.

۳.۶. یادگیری ماشین تفسیرپذیر (Interpretable Machine Learning)

عدم شفافیت در نحوه تصمیم‌گیری مدل‌های «جعبه سیاه» یکی از چالش‌های اصلی در پذیرش یادگیری ماشین در صنعت ساختمان است که برای رفع آن، در این پژوهش از الگوریتم SHAP^{۱۰۵} (برگرفته از تئوری بازی‌ها و ارزش Shapley) استفاده شد تا سهم هر ویژگی^{۱۰۶} در پیش‌بینی نهایی به‌صورت منصفانه مشخص گردد.

SHAP با ارزیابی تغییرات خروجی مدل در حضور یا غیاب هر ویژگی، اهمیت و جهت اثرگذاری آن را تعیین می‌کند. در این مطالعه، مدل جنگل تصادفی با عنوان الگوریتم پایه برای تفسیرپذیری انتخاب شد تا میزان و جهت تأثیر متغیرهای مستقل بر تقاضای انرژی ساختمان تبیین گردد. بدین‌منظور، متغیر «کاربری» با مقادیر (۰: مسکونی، ۱: اداری، ۲: آموزشی، ۳: تجاری) کدگذاری و در مدل لحاظ شد.

– **نمودار خلاصه SHAP:** نمودار خلاصه SHAP (ت ۶) اهمیت هر ویژگی را با تأثیر آن ترکیب می‌کند. هر نقطه در

برای پیش‌بینی گرمایشی حفظ شد، اما در پیش‌بینی سرمایشی، AdaBoost از مدل‌های مبتنی بر Gradient Boosting بهتر عمل کرد. در مجموع، در میان مدل‌های مبتنی بر تقویت گرادیان، LightGBM و CatBoost تنظیم‌شده بهترین نتایج را دادند؛ هرچند برای پیش‌بینی انرژی گرمایشی در حالت طبقه‌بندی، GradientBoost نیز عملکرد بسیار مناسبی داشت.

۳.۵. تنظیم هاپرپارامترها (Tuning)

مطابق با یافته‌های پژوهشی^{۱۰۱}، تنظیم فرایپارامترها نقش کلیدی در بهبود عملکرد مدل‌های یادگیری ماشین داشت. نتایج بهینه‌سازی برای الگوریتم‌های مختلف به شرح زیر است: الگوریتم KNN و مدل‌های درختی: در مدل KNN استفاده از روش OHE منجر به افزایش مقدار بهینه K (تا ۳۳) و در نتیجه افزایش هزینه محاسباتی شد. در مقابل، در الگوریتم درخت تصمیم، پیش‌پردازش به روش OHE برای انرژی سرمایشی و LE برای گرمایشی، با کاهش عمق و تعداد برگ‌ها، منجر به ایجاد مدل‌های ساده‌تر و سریع‌تر گردید؛ درحالی‌که افزایش عمق به بیش از ۹ تأثیری بر دقت نداشت.

ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF): در مدل SVM، تابع هسته «poly» در اکثر سناریوها برتری داشت. در الگوریتم RF افزایش تعداد یادگیرنده‌های پایه از ۱۰۰ به ۲۰۰ دقت را بهبود بخشید؛ با این حال، در پیش‌بینی بار گرمایشی با روش OHE، دستیابی به حداکثر دقت با تنها ۵۰ درخت (نصف مقدار پیش‌فرض) میسر شد که با نتایج تحقیق سیدزاده^{۱۰۲} همخوانی دارد.

شبکه عصبی چندلایه (MLP): تابع فعال‌سازی «relu» بهترین عملکرد را نشان داد (به‌جز در حالت Label Encoding گرمایشی که «logistic» برتر بود). با توجه به حجم داده‌های

60. L. Sun, et al., "The Prediction of Building Heating and Ventilation Energy Consumption Based on Adaboost-Bp Algorithm", in IOP Conference Series: Materials Science and Engineering, vol. 782 (IOP Publishing, 2020), 32008; Yoav Freund and Robert E Schapire, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", Computer and System Sciences, 55(1) (1997): 119-139.

61. Geeksforgeeks, "https://www.Geeksforgeeks.Org/Boosting-in-Machine-Learning-Boosting-and-Adaboost", 01/01/2024, 2023.

62. N.S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", American Statistician, 1992, 175-85.

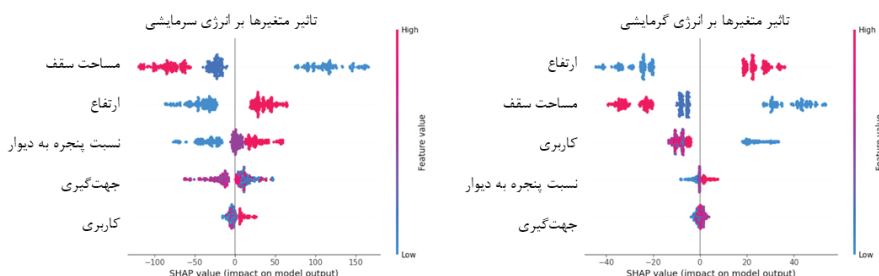
63. G. Sudheer and A. Suseelatha, "Short Term Load Forecasting Using Wavelet Transform Combined with Holt-Winters and Weighted Nearest Neighbor Models", International Journal of Electrical Power & Energy Systems, 64 (2015): 340-346.
64. F. Wahid and D. Kim, "A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings", International Journal of Smart Home, 10(2) (2016): 97-108.

ت ۶ جهت و میزان تأثیر متغیرهای پیش‌بینی کننده (فیچر) بر متغیر هدف (انرژی سرمایشی و گرمایشی) در نمودار خلاصه SHAP.

نمودار خلاصه یک مقدار Shapley برای یک ویژگی و یک حالت است. موقعیت روی محور γ توسط ویژگی و در محور x با مقدار Shapley تعیین می‌شود. رنگ نشان‌دهنده مقدار هر ویژگی از کم به زیاد است.

با توجه به «ت ۶»، فشرده‌تر شدن ساختمان (کاهش ابعاد) باعث افزایش انرژی سرمایشی و گرمایشی می‌شود و تأثیر این متغیر بر انرژی سرمایشی بیشتر از انرژی گرمایشی است. بهتر است سطح خارجی ساختمان در این اقلیم افزایش یابد؛ شاید با این نتیجه، پیشنهاد ساختمان‌های دارای شکستگی در پلان و نما بهترین گزینه طراحی در این اقلیم باشند. تأثیر ارتفاع در انرژی گرمایشی ساختمان بیشتر از انرژی سرمایشی آن است و با افزایش ارتفاع هم انرژی سرمایشی و هم انرژی گرمایشی افزایش می‌یابد. با افزایش نسبت پنجره به دیوار نیز، هم انرژی سرمایشی و هم انرژی گرمایشی افزایش می‌یابد که این تأثیر بر انرژی سرمایشی به مراتب بیشتر از انرژی گرمایشی است. جهت‌گیری ساختمان الگوی منظمی ندارد، بنابراین قابل تفسیر نیست. همچنین تأثیر کاربری در انرژی گرمایشی به مراتب بیشتر از انرژی سرمایشی است و کاربری‌های غیرمسکونی انرژی سرمایشی بیشتر و انرژی گرمایشی کمتری نسبت به کاربری مسکونی دارند.

– **نمودار وابستگی SHAP:** نمودار وابستگی SHAP برای تحلیل اثرات متقابل متغیرها بر نتایج مدل استفاده شد (ت ۷).



نتایج نشان‌دهنده الگوهای رفتاری متفاوتی در تقاضای انرژی گرمایشی و سرمایشی ساختمان است:

۱. پارامترهای کالبدی مشترک (ابعاد، ارتفاع، و WWR):

تعامل ارتفاع و مساحت: در هر دو بخش سرمایش و گرمایش، افزایش ارتفاع به‌طور کلی باعث افزایش تقاضای انرژی می‌شود. با این حال، تعامل ارتفاع با مساحت زون تعیین‌کننده است؛ در ارتفاع ۳ متری، افزایش مساحت منجر به افزایش مصرف می‌شود، در حالی که در ارتفاع ۶ متری، افزایش مساحت باعث کاهش تقاضای انرژی می‌گردد.

– **فشرده‌گی ساختمان:** فشرده شدن ابعاد ساختمان تقاضای هر دو نوع انرژی را افزایش می‌دهد. برای بهینه‌سازی، در ساختمان‌های فشرده، ارتفاع کمتر و در ساختمان‌های بزرگ‌تر، ارتفاع بیشتر پیشنهاد می‌شود.

نسبت پنجره به دیوار (WWR): با افزایش WWR، مصرف انرژی در هر دو بخش صعودی است. این اثر منفی را می‌توان با افزایش مساحت زون تعدیل کرد.

۲. تحلیل مقایسه‌ای کاربری و جهت‌گیری:

کاربری: کاربری مسکونی بیشترین بار گرمایشی را دارد، در حالی که کاربری تجاری در صدر تقاضای سرمایشی قرار دارد. جالب توجه است که افزایش مساحت در کاربری تجاری باعث کاهش بار سرمایشی می‌شود، اما در سایر کاربری‌ها اثر معکوس دارد.

جهت‌گیری: جهت‌گیری جنوبی کمترین بار گرمایشی و جهت‌گیری شمالی کمترین بار سرمایشی را ایجاد می‌کند. در بخش گرمایش، چرخش به سمت جنوب غربی و در بخش سرمایش، چرخش به سمت شمال شرقی عملکرد بهتری نسبت به جهات شرقی و غربی متناظر دارند.

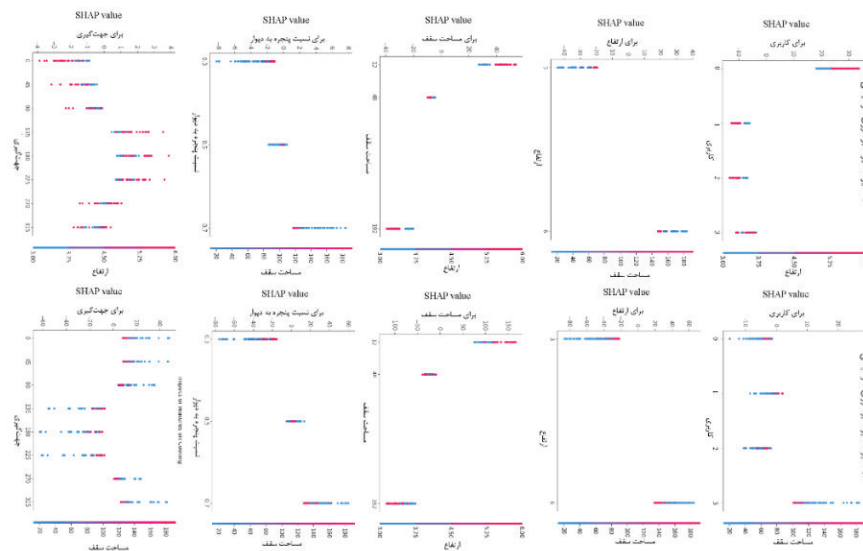
تعامل جهت‌گیری با ابعاد: در ساختمان‌های با جهت‌گیری



65. Zhang, et al., "A Review of Machine Learning in Building Load Prediction", 116452.

66. Junjing Yang, et al., "K-Shape Clustering Algorithm for Building Energy Usage Patterns Analysis and Forecasting Model Accuracy Improvement", Energy and Buildings, 146 (2017): 27-37.

ت ۷. جهت و میزان تأثیر متغیرهای پیش‌بینی‌کننده (فیچر) بر متغیرهدف (۱) انرژی گرمایشی، (۲) انرژی سرمایشی در نمودار وابستگی SHAP.



آنها، LightGBM (در رگرسیون) و شبکه عصبی چندلایه (در مدل‌های انفرادی) بهترین دقت را داشتند. برخی از مهم‌ترین نتایج این پژوهش را می‌توان به صورت زیر خلاصه کرد:

۱. تأثیر پارامترهای کالبدی: فشردگی ساختار، افزایش ارتفاع، و افزایش نسبت پنجره به دیوار همبستگی مستقیمی با افزایش مصرف انرژی (سرمایشی و گرمایشی) دارند. همچنین، مساحت سقف و ارتفاع، بیشترین اهمیت را در مدل‌های پیش‌بینی دارند.

۲. عملکرد الگوریتم‌ها: الگوریتم‌های طبقه‌بندی نسبت به روش‌های رگرسیونی عملکرد دقیق‌تری داشتند. تنظیم هایپرپارامترها به‌ویژه در شبکه عصبی و الگوریتم‌های تقویتی (AdaBoost و LightGBM) تأثیر بسزایی در بهبود دقت مدل داشت.

۳. پیش‌پردازش: روش One Hot Encoding برای پیش‌بینی انرژی گرمایشی و Label Encoding برای پیش‌بینی انرژی سرمایشی توصیه می‌شود.

جنوبی و غربی، برای کاهش بار گرمایشی افزایش ارتفاع، و برای کاهش بار سرمایشی افزایش مساحت ساختمان توصیه می‌شود. در مقابل، برای جهت‌گیری‌های شمالی، ارتفاع و مساحت کمتر منجر به عملکرد انرژی بهتر می‌گردد.

به‌منظور تحقق اهداف توسعه پایدار در صنعت ساختمان، تحلیل جامع عوامل طراحی شامل جهت‌گیری، ابعاد، کاربری، و نسبت پنجره به دیوار ضروری است. نتایج این پژوهش نشان می‌دهد که استفاده از الگوریتم‌های تقویتی یادگیری ماشین، به دلیل بهره‌گیری از تعداد زیاد یادگیرندگان پایه، عملکرد بسیار بهتری نسبت به الگوریتم‌های انفرادی دارند؛ موضوعی که با یافته‌های برخی پژوهش‌ها^{۱۰۷} نیز همسوست. همچنین، نتایج مطالعه حاضر بیانگر برتری شبکه‌های عصبی مصنوعی در دقت پیش‌بینی نسبت به سایر مدل‌های یادگیری ماشین است و نتایج پژوهش رحمان و همکاران^{۱۰۸} را تأیید می‌کند.

در میان مدل‌های رگرسیونی، رگرسیون فرایند گاوسی دقت بالاتری نسبت به رگرسیون خطی داشت که این نتیجه با گزارش رستوگی و همکاران^{۱۰۹} سازگار است. در مدل‌های مبتنی بر درخت نیز، جنگل‌های تصادفی با اتکا بر تجمیع چندین درخت تصمیم و تقویت طبیعی عملکرد، دقت بیشتری نسبت به الگوریتم درخت تصمیم نشان دادند؛ روندی که در پژوهش پکارل و همکارش^{۱۱۰} نیز تأیید شده است.

نتیجه‌گیری

در این پژوهش با هدف ارتقای دقت پیش‌بینی مصرف انرژی ساختمان، به بررسی جامع روش‌های یادگیری ماشین پرداخته شده است. نتایج نشان داد که الگوریتم‌های تقویتی (AdaBoost و زیرمجموعه‌های GradientBoost)، به دلیل ادغام یادگیرنده‌های پایه، عملکردی به‌مراتب دقیق‌تری از مدل‌های انفرادی (رگرسیونی و طبقه‌بندی) دارند. در میان

مطالعات آتی، مدل‌های توسعه‌یافته بر روی داده‌های واقعی ساختمان‌های موجود، با در نظر گرفتن سیستم‌های تهویه مطبوع واقعی و اقلیم‌های متنوع، ارزیابی شوند. همچنین، بررسی هزینه محاسباتی (زمان اجرا) و توسعه ابزارهای خودکار پیش‌پردازش داده‌ها از سرفصل‌های پیشنهادی برای پژوهش‌های آتی است.

References

- Ahmad, Mohammad Waseem, et al. "Trees vs Neurons: Comparison between Random Forest and ANN for High-Resolution Prediction of Building Energy Consumption". *Energy and Buildings*, 147 (2017): 77-89.
- Ahmad, Tanveer and Huanxin Chen. "Short and Medium-Term Forecasting of Cooling and Heating Load Demand in Building Environment with Data-Mining Based Approaches". *Energy and Buildings*, 166 (2018): 460-476.
- Al-Rakhami, Mabrook, et al. "An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings." *IEEE Access* 7 (2019): 48328-48338.
- Altman, Naomi S. "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression". in *American Statistician*, 1992, 175-185.
- Antanasijević, Davor, et al. "Modeling of Energy Consumption and Related GHG (Greenhouse Gas) Intensity and Emissions in Europe Using General Regression Neural Networks". *Energy*, 84 (2015): 816-24.
- Araya, Daniel B, et al. "An Ensemble Learning Framework for Anomaly Detection in Building Energy Consumption". *Energy and Buildings*, 144 (2017): 191-206.
- Ashrae. "ANSI/ASHRAE Standard 140-2017: Standard Method Of Test For The Evaluation Of Building Energy Analysis Computer Programs". 2017.
- ASHRAE. "Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE Standard 55-2020". 2020.
- Bassi, Abnash, et al. "Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models." In *The 12th International Conference on Advances in Information Technology*, 1-9. New York, NY, USA: ACM, 2021.
- Beale, Mark Hudson, et al. "Neural Network Toolbox". *User's Guide, MathWorks*, 2 (2010): 77-81.
- Bentéjac, Candice, et al. "A Comparative Analysis of Gradient Boosting Algorithms". *Artificial Intelligence Review*, 54 (2021): 1937-1967.
- Bukkapatnam, Satish T.S and Changqing Cheng. "Forecasting

۴. تفاوت کاربری: کاربری‌های غیرمسکونی در مقایسه با کاربری مسکونی، پتانسیل انرژی سرمایشی بالاتر و انرژی گرمایشی کمتری دارند.

محدودیت‌های این پژوهش شامل ماهیت شبیه‌سازی‌شده داده‌ها (عدم حضور داده‌های واقعی)، نادیده گرفتن رفتار کاربران، و تمرکز بر اقلیم تهران (تک‌زون) است. پیشنهاد می‌شود در

- the Evolution of Nonlinear and Nonstationary Systems Using Recurrence-Based Local Gaussian Process Models". *Physical Review E*, vol. 82, no. 5 (2010): 56206.
- Calì, Davide, et al. "Analysis of Occupants' Behavior Related to the Use of Windows in German Households". *Building and Environment*, 103 (July 2016): 54-69.
- Catalina, Tiberiu, et al. "Multiple Regression Model for Fast Prediction of the Heating Energy Demand". *Energy and Buildings*, 57 (2013): 302-312.
- Catalina, Tiberiu, et al. "Development and Validation of Regression Models to Predict Monthly Heating Demand for Residential Buildings". *Energy and Buildings*, 40(10) (2008): 1825-1832.
- Cheng, Min-Yuan and Minh-Tu Cao. "Accurately Predicting Building Energy Performance Using Evolutionary Multivariate Adaptive Regression Splines". *Applied Soft Computing*, 22 (2014): 178-188.
- Chou, Jui-Sheng and Dac-Khuong Bui. "Modeling Heating and Cooling Loads by Artificial Intelligence for Energy-Efficient Building Design". *Energy and Buildings*, 82 (2014): 437-446.
- Chowdhury, Shamit Roy, et al. "Energy Consumption Prediction Using Light Gradient Boosting Machine Model". In *International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy*, Springer, 2020, 413-422.
- Crone, Sven F., et al. "The Impact of Preprocessing on Support Vector Regression and Neural Networks in Time Series Prediction". In *DMIN*, 2006, 37-44.
- Cuerda, Elena, et al. "Understanding the Performance Gap in Energy Retrofitting: Measured Input Data for Adjusting Building Simulation Models". *Energy and Buildings*, 209 (2020): 109688.
- Dear, Richard De and Gail Schiller Brager. "The Adaptive Model of Thermal Comfort and Energy Conservation in the Built Environment". *International Journal of Biometeorology*, 45 (2001): 100-108.

67. C. Fan, et al., "Development of Prediction Models for Next-Day Building Energy Consumption and Peak Power Demand Using Data Mining Techniques", *Applied Energy*, 127 (2014): 1-10.
68. Fu Xiao and Cheng Fan, "Data Mining in Building Automation System for Improving Building Operational Performance", *Energy and Buildings*, 75 (2014): 109-118.
69. S. Petcharat, et al., "Assessment of Potential Energy Saving Using Cluster Analysis: A Case Study of Lighting Systems in Buildings", *Energy and Buildings*, 52 (2012): 145-152.
70. Fan Tang, et al., "Modeling and Short-Term Prediction of HVAC System with a Clustering Algorithm", *Energy and Buildings*, 82 (2014): 310-321; J.G. Jetcheva, et al., "Neural Network Model Ensembles for Building-Level Electricity Load Forecasts", *Energy and Buildings*, 84 (2014): 214-223; Zhun Yu, et al., "A Systematic Procedure to Study the Influence of Occupant Behavior on Building Energy Consumption", *Energy and Buildings*, 43(6) (2011): 1409-1417.

71. Hierarchical

72. DBSCAN: Density-Based Spatial Clustering of Applications with Noise

73. I. Prahastono, et al., "A Review of Electricity Load Profile Classification Methods", in 2007 42nd International Universities Power Engineering Conference (IEEE, 2007), 1187-1191; S. Jalori and T.A. Reddy, "A New Clustering Method to Identify Outliers and Diurnal Schedules from Building Energy Interval Data", ASHRAE Transactions 121, no. 2 (2015).

74. Local Maximum

75. S.Z Selim and M.A. Ismail, "K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality", IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 1 (1984): 81-87.

76. R.A. Lara, et al., "Energy Audit of Schools by Means of Cluster Analysis", Energy and Buildings, 95 (2015): 160-171.

77. A.K. Jain, M Narasimha Murty, and Patrick J Flynn, "Data Clustering: A Review", ACM Computing Surveys (CSUR), 31(3) (1999): 264-323.

Deng, Zhipeng and Qingyan Chen. "Reinforcement Learning of Occupant Behavior Model for Cross-Building Transfer Learning to Various HVAC Control Systems". *Energy and Buildings*, 238 (2021): 110860.

Dietterich, Thomas G. "Ensemble Methods in Machine Learning." In *International Workshop on Multiple Classifier Systems*, Springer 2000, 1-15.

Dong, Zhenxiang, et al. "Hourly Energy Consumption Prediction of an Office Building Based on Ensemble Learning and Energy Consumption Pattern Classification". *Energy and Buildings*, 241 (2021): 110929.

Edwards, Richard E., et al. "Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study". *Energy and Buildings*, 49 (2012): 591-603.

Ester, Martin, et al. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". In *Kdd*, 96 (1996): 226-231.

Faizollahzadeh Ardabili, Sina, et al. "Modeling and Simulation Controlling System of HVAC Using Fuzzy and Predictive (Radial Basis Function, RBF) Controllers". *Journal of Building Engineering* 6 (2016): 301-308.

Fan, Cheng, et al. "Development of Prediction Models for Next-Day Building Energy Consumption and Peak Power Demand Using Data Mining Techniques". *Applied Energy*, 127 (2014): 1-10.

Fan, Cheng, et al. "A Short-Term Building Cooling Load Prediction Method Using Deep Learning Algorithms". *Applied Energy*, 195 (2017): 222-233.

Fayyaz, Muhammad, et al. "Thermal Comfort Model for HVAC Buildings Using Machine Learning". *Arabian Journal for Science and Engineering*, 47(2) (2022): 2045-2060.

Ferlito, S., et al. "Predictive Models for Building's Energy Consumption: An Artificial Neural Network (ANN) Approach". In *2015 XVIII Aisem Annual Conference*, 1-4. IEEE, 2015.

Firląg, Szymon, et al. "Control Algorithms for Dynamic Windows for Residential Buildings". *Energy and Buildings*, 109 (December 2015): 157-173.

Freund, Yoav. "Boosting a Weak Learning Algorithm by Majority". *Information and Computation*, 121(2) (1995): 256-285.

Freund, Yoav and Robert E Schapire. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting". *Computer and System Sciences*, 55(1) (1997): 119-139.

Freund, Yoav and Robert E Schapire. "Experiments with a New Boosting Algorithm". In *lcm1*, 96 (1996):148-156.

Friedman, Jerome, et al. "Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors)". *The Annals of Statistics*, 28(2) (2000): 337-407.

Fumo, Nelson. "A Review on the Basics of Building Energy Estimation". *Renewable and Sustainable Energy Reviews*, 31 (2014): 53-60.

Geeksforgeeks. "https://www. Geeksforgeeks. Org/Boosting-in-Machine-Learning-Boosting-and-Adaboost". 01/01/2024, 2023.

_____. "https://www.Geeksforgeeks.Org/MI-Label-Encoding-of-Datasets-in-Python/". 18 Apr, 2023, 2023.

_____. "https://www.Geeksforgeeks.Org/MI-One-Hot-Encoding-of-Datasets-in-Python/." 18 Apr, 2023, 2023.

Grosicki, Emmanuèle, et al. "A Weighted Linear Prediction Method for Near-Field Source Localization". *IEEE Transactions on Signal Processing*, 53(10) (2005): 3651-3660.

He, Xuming. *Handbook of Quantile Regression*. Edited by Roger Koenker, et al., *Handbook of Quantile Regression*. Chapman and Hall/CRC, 2017.

Hong, Sung-Min, et al. "Improved Benchmarking Comparability for Energy Consumption in Schools". *Building Research & Information*, vol. 42, no. 1 (2014): 47-61.

Hygh, Janelle S, et al. "Multivariate Regression as an Energy Assessment Tool in Early Building Design". *Building and Environment*, 57 (2012): 165-175.

"Iranian National Building Code, Part 19: Energy Conservation, Ministry of Roads and Urban Development". 2020.

Jain, Anil K., et al. "Data Clustering: A Review". *ACM Computing Surveys (CSUR)*, 31(3) (1999): 264-323.

Jain, Rishree K., et al. "Forecasting Energy Consumption of Multi-Family Residential Buildings Using Support Vector Regression: Investigating the Impact of Temporal and Spatial Monitoring Granularity on Performance Accuracy". *Applied Energy*, 123 (2014): 168-178.

Jalori, Saurabh and T Agami Reddy. "A New Clustering Method to Identify Outliers and Diurnal Schedules from Building Energy Interval Data". *ASHRAE Transactions*, vol. 121, no. 2 (2015).

Javed, AbdulRehman, et al. "Analyzing the Effectiveness and Contribution of Each Axis of Tri-Axial Accelerometer Sensor for Accurate Activity Recognition". *Sensors*, 20(8) (2020): 2216.

Jetcheva, Jorjeta G., et al. "Neural Network Model Ensembles for Building-Level Electricity Load Forecasts", *Energy and Buildings* 84 (2014): 214-223.

Jiang, Xiaoqian, et al. "Adaptive Gaussian Process for Short-Term Wind Speed Forecasting". In *ECAI 2010*, 661-66. IOS Press, 2010.

Jung, Hyun Chul, et al. "Prediction of Building Energy Consumption Using an Improved Real Coded Genetic Algorithm Based Least Squares Support Vector Machine Approach". *Energy and Buildings*, 90 (2015): 76-84.

Karimi, Alireza, et al. "Assessment of Outdoor Design

- Conditions on the Energy Performance of Cooling Systems in Future Climate Scenarios—A Case Study over Three Cities of Texas, United States". *Sustainability*, vol. 14, no. 22 (2022): 14848.
- Kim, Sun Ho and Hyeun Jun Moon. "Case Study of an Advanced Integrated Comfort Control Algorithm with Cooling, Ventilation, and Humidification Systems Based on Occupancy Status". *Building and Environment*, 133 (2018): 246-264.
- King, Benjamin. "Step-Wise Clustering Procedures". *the American Statistical Association*, 62(317) (1967): 86-101.
- Kuster, Corentin, et al. "Electrical Load Forecasting Models: A Critical Systematic Review". *Sustainable Cities and Society*, 35 (2017): 257-270.
- Laarabi, B, O. May Tzuc, D Dahlioui, A Bassam, M Flota-Bañuelos, and A Barhdadi. "Artificial Neural Network Modeling and Sensitivity Analysis for Soiling Effects on Photovoltaic Panels in Morocco". *Superlattices and Microstructures* 127 (2019): 139–50.
- Lara, Rigoberto Arambula, et al. "Energy Audit of Schools by Means of Cluster Analysis". *Energy and Buildings*, 95 (2015): 160-171.
- Lazos, Dimitris, et al. "Optimisation of Energy Management in Commercial Buildings with Weather Forecasting Inputs: A Review". *Renewable and Sustainable Energy Reviews*, 39 (2014): 587-603.
- Li, Qiong, et al. "Applying Support Vector Machine to Predict Hourly Cooling Load in the Building". *Applied Energy*, vol. 86, no. 10 (2009): 2249-2256.
- Li, Qiong, et al. "Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks". *Energy Conversion and Management*, 50(1) (2009): 90-96.
- Li, Xiaoli, et al. "Classification of Energy Consumption in Buildings with Outlier Detection". *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11 (2009): 3639-3644.
- Li, Xiwang and Jin Wen. "Review of Building Energy Modeling for Control and Operation". *Renewable and Sustainable Energy Reviews*, 37 (2014): 517-537.
- Liang, Shaokun, et al. "Energy Consumption Prediction Using the GRU-MMAAttention-LightGBM Model with Features of Prophet Decomposition". *PLoS One* 18, no. 1 (2023): e0277085.
- Liu, Jiangyan, et al. "An Energy Performance Evaluation Methodology for Individual Office Building with Dynamic Energy Benchmarks Using Limited Information". *Applied Energy*, 206 (2017): 193-205.
- Liu, Weihao, et al. "Modeling Window and Thermostat Use Behavior to Inform Sequences of Operation in Mixed-Mode Ventilation Buildings". *Science and Technology for the Built Environment*, 27(9) (2021): 1204-1220.
- Lu, Siliang, et al. "Data-Driven Simulation of a Thermal Comfort-Based Temperature Set-Point Control with ASHRAE RP884". *Building and Environment*, 156 (2019): 137-146.
- Ma, Zhitong, et al. "Applying Support Vector Machines to Predict Building Energy Consumption in China". *Energy Procedia*, 152 (2018): 780-786.
- Massana, Joaquim, et al. "Short-Term Load Forecasting in a Non-Residential Building Contrasting Models and Attributes". *Energy and Buildings*, 92 (2015): 322-330.
- May Tzuc, et al. "Artificial Intelligence Techniques for Modeling Indoor Building Temperature under Tropical Climate Using Outdoor Environmental Monitoring". *Energy Engineering*, 146(2) (2020).
- Mikučionienė, Rūta, et al. "Evaluation of Energy Efficiency Measures Sustainability by Decision Tree Method". *Energy and Buildings*, 76 (2014): 64-71.
- Miller, Clayton, et al. "The ASHRAE Great Energy Predictor III Competition: Overview and Results". *Science and Technology for the Built Environment*, 26(10) (2020): 1427-1447.
- Mottini, Alejandro and Rodrigo Acuna-Agost. "Relative Label Encoding for the Prediction of Airline Passenger Nationality". In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 671-676. IEEE, 2016.
- Murtagh, Fionn. "A Survey of Recent Advances in Hierarchical Clustering Algorithms". *The Computer Journal*, vol. 26, no. 4 (1983): 354-359.
- Ouf, Mohamed M., et al. "Improving Occupant-Related Features in Building Performance Simulation Tools". *Building Simulation*, 11(4) (2018): 803-817.
- Pan, Yue and Limao Zhang. "Data-Driven Estimation of Building Energy Consumption with Multi-Source Heterogeneous Data". *Applied Energy*, 268 (2020): 114965.
- Peng, Hanchuan, et al. "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8 (2005): 1226-1238.
- Penya, Yoseba K, et al. "Short-Term Load Forecasting in Non-Residential Buildings". In *IEEE Africon'11*, 1-6. IEEE, 2011.
- Petcharat, Siriwarin, et al. "Assessment of Potential Energy Saving Using Cluster Analysis: A Case Study of Lighting Systems in Buildings". *Energy and Buildings*, 52 (2012): 145-152.
- Platon, Radu, et al. "Hourly Prediction of a Building's Electricity Consumption Using Case-Based Reasoning, Artificial Neural Networks and Principal Component Analysis". *Energy and Buildings*, 92 (2015): 10-18.
- Pokharel, Sudan and Prashnna Ghimire. "Data-Driven MLmodels for Accurate Prediction of Energy Consumption in a Low-Energy House". *Innovations in Engineering Education*, 6(1) (2023): 12-20.
- Prahastono, Iswan, et al. "A Review of Electricity Load Profile
78. P.H.A. Sneath and R.R. Sokal, "Numerical Taxonomy", *Nature*, 193 (1962): 855-860.
79. B.King, "Step-Wise Clustering Procedures", *American Statistical Association*, 62(317) (1967): 86-101.
80. F. Murtagh, "A Survey of Recent Advances in Hierarchical Clustering Algorithms", *The Computer Journal*, 26(4) (1983): 354-359.
81. M. Ester, et al., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in *Kdd*, vol. 96, 1996, 226-231.
82. Nelson Fumo, "A Review on the Basics of Building Energy Estimation", *Renewable and Sustainable Energy Reviews*, 31 (2014): 53-60; Hai-xiang Zhao and F. Magoulès, "A Review on the Prediction of Building Energy Consumption", *Renewable and Sustainable Energy Reviews*, 16(6) (2012): 3586-3592; D. Lazos, et al., "Optimisation of Energy Management in Commercial Buildings with Weather Forecasting Inputs", *Renewable and Sustainable Energy Reviews*, 39 (2014): 587-603; Xiwang Li and Jin Wen, "Review of Building



Energy Modeling for Control and Operation", *Renewable and Sustainable Energy Reviews*, 37 (2014): 517-537.

83. M. Fayyaz, et al., "Thermal Comfort Model for HVAC Buildings Using Machine Learning", *Arabian Journal for Science and Engineering*, 47(2) (2022): 2045-2060.

84. C. Kuster, et al., "Electrical Load Forecasting Models: A Critical Systematic Review", *Sustainable Cities and Society*, 35 (2017): 257-270.

85. M.Q. Raza and A. Khosravi, "A Review on Artificial Intelligence Based Load Demand Forecasting Techniques for Smart Grid and Buildings", *Renewable and Sustainable Energy Reviews*, 50 (2015): 1352-1372.

86. S.F. Crone, et al., "The Impact of Preprocessing on Support Vector Regression and Neural Networks in Time Series Prediction", in *DMIN*, 2006, 37-44.

87. A. Mottini and R. Acuna-Agost, "Relative Label Encoding for the Prediction of Airline Passenger Nationality", in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (IEEE, 2016), 671-676.

Classification Methods". In *2007 42nd International Universities Power Engineering Conference*, 1187-1191. IEEE, 2007.

Qin, Y.D, et al. "Exploring New Building Energy Saving Control Strategy Application under the Energy Internet of Things". In *IOP Conference Series: Earth and Environmental Science*, 701:12060. IOP Publishing, 2021.

Rastogi, Parag, et al. "Gaussian-Process-Based Emulators for Building Performance Simulation". In *International Building Performance Simulation Association*, 2017.

Raza, Muhammad Qamar and Abbas Khosravi. "A Review on Artificial Intelligence Based Load Demand Forecasting Techniques for Smart Grid and Buildings". *Renewable and Sustainable Energy Reviews*, 50 (2015): 1352-1372.

Rehman, Saif Ur, et al. "Personalised Comfort: A Personalised Thermal Comfort Model to Predict Thermal Sensation Votes for Smart Building Residents". *Enterprise Information Systems*, 16(7) (2022): 1-23.

Rodríguez, Pau, et al. "Beyond One-Hot Encoding: Lower Dimensional Target Embedding". *Image and Vision Computing*, 75 (2018): 21-31.

Rosenblatt, Frank. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". *Psychological Review*, vol. 65, no. 6 (1958): 386-408.

Schapire, Robert E. "The Strength of Weak Learnability." *Machine Learning*, 5 (1990): 197-227.

Seger, Cedric. "An Investigation of Categorical Variable Encoding Techniques in Machine Learning: Binary versus One-Hot and Feature Hashing". *Kth Royal Institute of Technology School of Electrical Engineering and Computer Science*, 2018.

Selim, Shokri Z. and Mohamed A Ismail. "K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1 (1984): 81-87.

Seyedzadeh, Saleh, et al. "Machine Learning for Estimation of Building Energy Consumption and Performance: A Review". *Visualization in Engineering*, vol. 6, no. 1 (2018): 1-20.

Seyedzadeh, Saleh, et al. "Tuning Machine Learning Models for Prediction of Building Energy Loads". *Sustainable Cities and Society*, 47 (2019): 101484.

Shao, Minglei, et al. "Prediction of Energy Consumption in Hotel Buildings via Support Vector Machines". *Sustainable Cities and Society*, 57 (2020): 102128.

Shi, Xin, et al. "Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing". *Ieee Access*, 6 (2018): 30559-66.

Sholahudin, Azimil Gani Alam, et al. "Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments". *Applied Mechanics and Materials*, 819 (2016): 541-545.

Silva, Jose, et al. "Energy Consumption Forecasting Using Ensemble Learning Algorithms". *Advances in Intelligent Systems and Computing*, 1004 (2020): 5-13.

Sneath, Peter H.A. and Robert R. Sokal. "Numerical Taxonomy". *Nature*, 193 (1962): 855-860.

Sudheer, G. and A. Suseelatha. "Short Term Load Forecasting Using Wavelet Transform Combined with Holt-Winters and Weighted Nearest Neighbor Models". *International Journal of Electrical Power & Energy Systems*, 64 (2015): 340-346.

Sun, Lunan, et al. "The Prediction of Building Heating and Ventilation Energy Consumption Base on Adaboost-Bp Algorithm". In *IOP Conference Series: Materials Science and Engineering*, 782:32008. IOP Publishing, 2020.

Tang, Fan, et al. "Modeling and Short-Term Prediction of HVAC System with a Clustering Algorithm". *Energy and Buildings*, 82 (2014): 310-321.

Tolles, Juliana and William J Meurer. "Logistic Regression: Relating Patient Characteristics to Outcomes". *Jama*, vol. 316, no. 5 (2016): 533-534.

Touzani, Samir, et al. "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings". *Energy and Buildings*, 158 (2018): 1533-1543.

Vellei, Marika, et al. "The Influence of Relative Humidity on Adaptive Thermal Comfort". *Building and Environment*, 124 (2017): 171-185.

Verbruggen, Silke. "Window Use Habits as an Example of Habitual Occupant Behaviour in Residential Buildings". *Ghent University*, 2021.

Wahid, Fazli, and Dohyeun Kim. "A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings". *International Journal of Smart Home*, vol. 10, no. 2 (2016): 97-108.

Wang, Zeyu, et al. "A Novel Ensemble Learning Approach to Support Building Energy Use Prediction". *Energy and Buildings*, 159 (2018): 109-122.

Wang, Zhe, et al. "Building Thermal Load Prediction through Shallow Machine Learning and Deep Learning". *Applied Energy*, 263 (2020): 114683.

Xiao, Fu and Cheng Fan. "Data Mining in Building Automation System for Improving Building Operational Performance". *Energy and Buildings*, 75 (2014): 109-118.

Yang, Junjing, et al. "K-Shape Clustering Algorithm for Building Energy Usage Patterns Analysis and Forecasting Model Accuracy Improvement". *Energy and Buildings*, 146 (2017): 27-37.

Yildiz, Baran, et al. "A Review and Analysis of Regression and Machine Learning Models on Commercial Building Electricity Load Forecasting". *Renewable and Sustainable Energy Reviews*, 73 (2017): 1104-1122.

Yin, Zhiqiang, et al. "Pump Feature Construction and



Electrical Energy Consumption Prediction Based on Feature Engineering and LightGBM Algorithm". *Sustainability*, 15(1) (2023): art 789.

Yu, Zhun, et al. "A Systematic Procedure to Study the Influence of Occupant Behavior on Building Energy Consumption". *Energy and Buildings*, vol.43, no. 6 (2011): 1409-1417.

Zhang, Liang, et al. "A Review of Machine Learning in Building Load Prediction". *Applied Energy*, 285 (2021): 116452.

Zhao, Hai-xiang and Frédéric Magoulès. "A Review on the Prediction of Building Energy Consumption". *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6 (2012): 3586-3592.

Zhao, RuoChen, et al. "Building Cooling Load Prediction Based on Lightgbm". *IFAC-PapersOnLine*, 55(11) (2022): 114-119.

Zhou, Nan, et al. "A Roadmap for China to Peak Carbon Dioxide Emissions and Achieve a 20% Share of Non-Fossil Fuels in Primary Energy by 2030". *Applied Energy*, 239 (2019): 793-819.

88. Geeksforgeeks, "Https://Www.Geeksforgeeks.Org/ML-Label-Encoding-of-Datasets-in-Python/", 18, Apr, 2023, 2023.

89. Ibid.

90. C. Seger, "An Investigation of Categorical Variable Encoding Techniques in Machine Learning: Binary versus One-Hot and Feature Hashing", 2018; P. Rodríguez, et al., "Beyond One-Hot Encoding: Lower Dimensional Target Embedding". *Image and Vision Computing*, 75 (2018): 21-31.

91. He, Handb. Quantile Regres.

92. K-Fold-CV

93. He, Handb. Quantile Regres

94. Touzani, et al., "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings", 1533-1543.

95. Bassi, et al., "Building Energy Consumption Forecasting", 1-9.

96. Ashrae, "ANSI/ASHRAE Standard 140-2017: Standard Method Of Test For The Evaluation Of Building Energy Analysis Computer Programs", 2017.

104. Rehman, et al., "Personalised Comfort: A Personalised Thermal Comfort Model to Predict Thermal Sensation Votes for Smart Building Residents", 1-23; Z. Deng and Q. Chen, "Reinforcement Learning of Occupant Behavior Model for Cross-Building Transfer Learning to Various HVAC Control Systems", *Energy and Buildings*, 238 (2021): 110860.

105. SHapley Additive exPlanations

106. Feature

107. Touzani, et al., "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings", 1533-1543.

108. Rehman, et al., "Personalised Comfort: A Personalised Thermal Comfort Model to Predict Thermal Sensation Votes for Smart Building Residents", 1-23.

109. Rastogi, et al., "Gaussian-Process-Based Emulators for Building Performance Simulation".

97. "Iranian National Building Code, Part 19: Energy Conservation, Ministry of Roads and Urban Development", 2020.

98. ASHRAE, "Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE Standard 55-2020", 2020.

99. S. Firląg, et al., "Control Algorithms for Dynamic Windows for Residential Buildings", *Energy and Buildings*, 109 (December 2015): 157-173.

100. Outlier

101. Bassi, et al., "Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models", 1-9.

102. Seyedzadeh, et al., "Tuning Machine Learning Models for Prediction of Building Energy Loads", 101484.

103. Seyedzadeh, et al., "Machine Learning for Estimation of Building Energy Consumption and Performance", 1-20.

110. S. Pokharel and P. Ghimire, "Data-Driven ML models for Accurate Prediction of Energy Consumption in a Low-Energy House" *Innovations in Engineering Education*, 6(1) (2023): 12-20.