Automation in Building Occupant Profile Development: A Machine Learning- and Persona-Enabled Approach

Sheik Murad Hassan Anik[1]; Xinghua Gao, Ph.D., M.ASCE[2]; and Na Meng, Ph.D.[3]

[1]Dept. of Computer Science, Virginia Polytechnic Institute and State Univ., Blacksburg, VA. Email: murad@vt.edu
[2]Myers-Lawson School of Construction, Virginia Polytechnic Institute and State Univ., Blacksburg, VA (corresponding author). Email: xinghua@vt.edu
[3]Dept. of Computer Science, Virginia Polytechnic Institute and State Univ., Blacksburg, VA. Email: nm8247@vt.edu

ABSTRACT

The user persona is a communication tool for designers to generate a mental model that describes the archetype of users. Developing building occupant personas has proven to be an effective method for human-centered smart building design, considering occupant comfort, behavior, and energy consumption. The current approaches to developing building occupant personas face a major obstacle of manual data processing and analysis. This study proposes a machine learning-based approach for occupant characteristics' classification and prediction with a view toward partially automating the building occupant persona generation process. We investigate the 2015 Residential Energy Consumption Dataset using six machine learning techniques for predicting 16 occupant characteristics, such as age, education, and thermal comfort. The models achieved moderate accuracy in predicting most of the occupant characteristics and significantly higher accuracy (over 90%) for attributes including the number of occupants in the household, their age group, and preferred usage of primary cooling equipment. The results of the study show the feasibility of using machine learning techniques for occupant characteristic prediction and automating the development of building occupant persona to minimize human effort.

INTRODUCTION

People spend most of their time indoors, and the impact of a building on its occupants is significant (Bell et al. 2003). Smart buildings integrate intelligence and adaptability to meet the drivers for building progression: energy, efficiency, longevity, comfort, and satisfaction (Buckman et al. 2014). Human-centeredness has become crucial for smart building design and operation, and developing building occupant personas effectively creates profiles for occupant-centric design (Agee et al. 2021; Brangier & Bornet 2011). The increasing availability of data allows smart buildings to adapt and prepare for change (Buckman et al. 2014). Maximizing occupant comfort while minimizing energy consumption is a key objective in smart buildings (Mo et al. 2002). Human-Building Interaction (HBI) affects human well-being and the surrounding environment (Alavi et al. 2019). To optimize performance and well-being, an iterative, human-centered approach to building design must be employed, placing human needs at the center of design and decision-making processes.

Agee et al. (2021) introduced a human-centered approach to smart housing, leading to the development of data-driven smart housing personas that focus on occupants' needs. Personas

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

© ASCE

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

help designers to avoid designing for themselves or technology and instead focus on users' needs (Takai & Ishii 2010). Designing buildings that meet the needs of the occupant persona can lead to improved building performance, such as energy efficiency and indoor air quality, as the building is designed to meet the specific needs of its occupants. It can also reduce the need for retrofits and renovations, resulting in cost savings over the building's lifecycle. Understanding the needs and preferences of potential occupants can help building-owners and developers to better market and lease the building, leading to higher occupancy rates and increased revenue. Developing building occupant personas has many benefits, but faces challenges like limited data availability and manual, time-consuming processes (Agee et al. 2021). Automating even a single step can accelerate persona development. Recent work has addressed data scarcity, with Anik et al. (2022) presenting a cost-effective framework for indoor data collection and Song et al. (2019) discussing data collection and analysis methods for human comfort. As more data becomes available and machine learning technologies advance, new opportunities for automating and accelerating persona development arise (Mitchell 1997; Kelleher et al. 2020).

In this study, we utilize machine learning techniques on the Residential Energy Consumption Survey Dataset (RECS) to classify and predict building occupant characteristics. This has the potential to streamline certain aspects of building occupant persona creation through automation. According to Higginbotham et al. (2000), research inquiries revolve around the examination of the formulated hypothesis. To test the hypothesis of the feasibility of using machine learning techniques in occupant characteristics classification and prediction, this work aims to answer the following research questions; RQ1: How effectively can machine learning tools predict individual building occupant characteristics? RQ2: How do machine learning algorithms compare with each other when predicting building occupant characteristics?

RELATED WORKS

This work delves into three key areas of prior research to provide a comprehensive background. These areas encompass occupant behavior and its impact on building energy performance, human factors' role in indoor environmental quality, and, the utilization of machine learning for modeling occupant behavior.

(i) Occupant behavior significantly impacts energy efficiency. Studies like Pan et al. (2017) and Sun et al. (2017), identified behavior patterns and recommended energy-saving strategies. Hu et al. (2020) stressed integrating behavior into energy policies, urging research on standards and incentives. Ortiz et al. (2019) probed comfort and energy use, categorizing occupants by behavior. Agee et al. (2021) and Malik et al. (2022) analyzed residential behaviors in the U.S. and India, advocating occupant-centric designs. These studies underscore behavior's role in energy savings and policy development.

(ii) Indoor Environmental Quality (IEQ) is multifaceted, impacting health and well-being (Shan et al. 2018; Khan et al. 2022). IEQ involves factors like air quality, lighting, temperature, humidity, noise, and aesthetics. Recognizing the intricate link between human factors and IEQ is crucial for architects, designers, and building managers (Pereira et al. 2020). Occupancy density affects IEQ, with overcrowding elevating CO2, heat, and discomfort (Bortolini et al. 2021). Bortolini and Forcada's study on forty-two rooms in Portugal showed higher CO2 levels and PM2.5 values, highlighting inadequate ventilation. Human behavior can introduce pollutants or improve IEQ (Fabi et al. 2013; Kim et al. 2017), as activities like smoking or adjusting thermostats impact the indoor environment.

© ASCE

(iii) Machine learning is harnessed to model occupant behavior and its energy impact. Amasyali et al. (2021) proposed data-driven models predicting energy consumption and comfort, aided by genetic algorithms. Carlucci et al. (2020) employed a machine learning model to predict energy use, revealing behavior's role and energy-saving opportunities. Li et al. (2020) used machine learning for load prediction, enhancing building design decisions. ML shows potential in boosting occupant satisfaction and energy efficiency. Deng et al. (2018) forecasted comfort with neural networks considering occupants' behavior. Peng et al. (2018) automated Heating, Ventilation, and Air Conditioning (HVAC) systems using machine learning for energy reduction. Kim et al. (2018) developed personalized comfort models, improving thermal preference prediction.

These studies enhance insights into how occupant behavior affects building performance, energy use, and comfort, utilizing machine learning for modeling and optimization. Persona development needs complete energy and behavior data; their absence creates hurdles. Despite behavior's complexity, identifying patterns aids energy-efficient building design. This research leverages machine learning and real-world data for faster occupant characteristic classification.

RESEARCH METHOD

Figure 1 illustrates the framework of this research, which is inspired by a study conducted by Zhongguo et al. (2017). The process begins with investing the chosen dataset. This study focuses on modeling occupant characteristics with housing and energy consumption attributes. A real-world dataset containing all these attributes together are not widely available. U.S. Energy Information Administration (EIA) conducts Residential Energy Consumption Survey (RECS) through a five-year period and releasing a dataset that contains data about occupant, housing, and, energy consumption data. It usually takes 2 to 3 years processing time before they release the dataset. This study started on 2021 and utilized the latest available RECS 2015 dataset from EIA at that moment. The metadata is extracted, and the data is cleaned in the pre-processing step. Then, 16 target variables are chosen to represent occupant characteristics, such as age, education, income, thermal preference, etc. The rest of the attributes remain as descriptive variables and are filtered through the feature selection step, in which the irrelevant and redundant variables are dropped. The dataset is then trained and evaluated through the 10-fold cross validation process.

Dataset description. The RECS is a periodic study conducted by the EIA that provides detailed information about energy usage in U.S. homes (EIA. 2015). RECS is a multi-year effort consisting of a household survey, data collection from household energy suppliers, and end-use consumption and expenditure estimation. RECS focuses on primary residences, excluding secondary homes, vacant units, and non-residential areas. Consequently, RECS estimates are ideal for comparing home characteristics within the residential sector rather than providing sector-wide totals. The total number of responding households in the 2015 RECS is 5,686. Each record includes a total of 759 attributes. These attributes have been categorized into 12 sections by the EIA: Structural Characteristics, Kitchen Appliances, Home Appliances, Space Heating, Air Conditioning, Water Heating, Miscellaneous, Fuels Used, Housing Unit Measurement, Fuel Bills, Housing Unit Characteristics, and Energy Insecurity and Assistance.

Pre-processing. The dataset included some records contain infinite, missing, blank, or null values. For machine learning models to understand all records, these values need to be transformed. Infinite values are replaced with a large enough number, i.e., the largest integer. Blank or null values are replaced with -1 to indicate that the value is missing. One target

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

© ASCE

variable, HHAGE, which refers to the age of the occupant, is continuous, ranging from 18 to 110. It is difficult to fit this variable into classification models. Past studies (Yarlagadda et al. 2015; Lin et al. 2020) used age groups to tackle this issue. This study follows the age range used in (Lin et al. 2020). The records are categorized into the following age groups (in years): Children (0 to 12), Young Adult (13 to 30), Middle Adult (31 to 50), Senior Adult (51 to 70), and Senior (71 to 110).

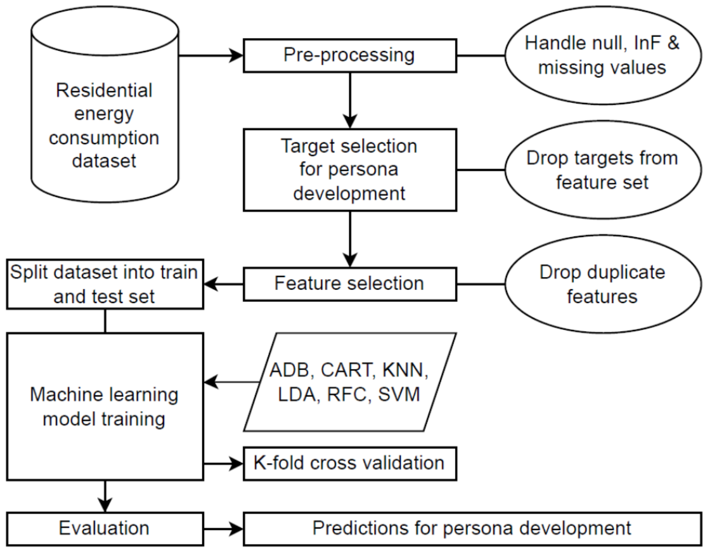

Figure 1. Occupant characteristics prediction and building occupant persona development.

Feature selection. Handling all 759 features is challenging and slows down the machine learning process, leading to a feature selection process. After manually examining each of the 759 features, the authors drop 370 features and use 389 features in the model training. The dataset included imputation flags for most of the attributes which refer to whether a record was imputed or not. A record can be imputed if it is not directly obtained from the occupants, but it is observed from other sources such as the structure, surrounding, or measuring device. Either way, the records hold valid data irrespective of the source, making the imputation flag an unnecessary component in this study as it will only increase the computation cost without adding any additional information. Hence, a total of 217 imputation flags are dropped. The energy consumption attributes recorded in the dataset contained both energy units and corresponding costs in US dollars. The dollar amount is not within the scope of this study and is thus dropped from the dataset. The dropped 370 features include imputation flags, utility bills in US dollar amounts (actual units are retained), and replicated weights (used for variance estimation). These features can be safely dropped because they do not represent any data that might affect the output of the machine learning models for developing building occupant personas. The attribute related to the number of phones is also removed because it directly correlates with the number of occupants or adults living in the house.

Target variables. This research revolves around the specified dataset on residential energy usage (EIA. 2015). The authors individually examined the dataset to pinpoint these target attributes, followed by a collaborative assessment by all authors drawing on their respective expertise. Through a comprehensive manual analysis of all dataset attributes, features

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

contributing to understanding individual occupants' behavior and traits were chosen. A total of 16 attributes were discovered, closely linked to occupant characteristics, presenting valuable insights for constructing building occupant personas: EQUIPMUSE (Main heating equipment household behavior), TEMPHOME (Winter temperature when someone is home), TEMPGONE (Winter temperature when no one is home), TEMPNITE (Winter temperature at night), USEWWAC (Most-used individual air conditioning unit household behavior, including values), TEMPHOMEAC (Summer temperature when someone is home), TEMPGONEAC (Summer temperature when no one is home), TEMPNITEAC (Summer temperature at night), HHAGE (Respondent age), EMPLOYHH (Respondent employment status), EDUCATION (Highest education completed by respondent), NHSLDMEM (Number of household members), NUMADULT (Number of household members age 18 or older), NUMCHILD (Number of household members age 17 or younger), ATHOME (Number of weekdays someone is at home), and MONEYPY (Annual gross household income for the last year).

Machine learning models. Assessing classifiers holds significance in academic and industrial domains (Zhang et al. 2017). In the StatLog initiative, King et al. (1995) conducted a comparison of diverse classification algorithms, including KNN, NB, LR, and NN, using substantial real-world datasets. Their findings emphasized the pivotal role of the specific dataset in determining performance, debunking the idea of a universally superior algorithm. Building upon previous research (Fernandez et al. 2014; Macia et al. 2014), this study examines prominent classification algorithms using the guidelines outlined by Pedregosa et al. (2011). This work simultaneously investigates six machine learning models on the classification task of 16 target variables. The models have been employed on the library recommended settings (Pedregosa et al. 2011) to conduct a comparative analysis, achieve reasonable performance throughout all target attributes, and avoid biasness towards any specific target variable. The following classification models are used in this work: Linear Discriminant Analysis (LDA), K-Nearest Neighbors Classifier (KNN), Decision Tree Classifier (CART), Support Vector Machine (SVM), AdaBoost Classifier (ADB), and Random Forest Classifier (RFC).

RESULTS

Assessing the performance of classification models is pivotal in machine learning evaluations. In this research, six machine learning models underwent evaluation through a 10-fold cross-validation setup. The results, showcased in Table 1, present the average accuracy attained by each model during the cross-validation process.

Notably, the best average scores for each target are highlighted in bold. Among the 16 target variables, the CART, LDA, and RFC classifiers achieved the highest average accuracy (63%), while KNN and SVM achieved the lowest (42%). Notably, CART and LDA achieved over 95% average accuracy for three specific targets: household members (NHSLDMEM), adults (NUMADULT), and children (NUMCHILD). Most models also surpassed 75% accuracy for classifying the main cooling equipment control behavior (USEWWAC), with an average accuracy of 82%, indicating moderate performance. Except for two variables, level of education (EDUCATION) and total household income (MONEYPY); where accuracy was 35% and 28% respectively, models achieved over 50% accuracy for all targets. It is worth noting that MONEYPY data is distributed across 16 classes, potentially influencing the lower accuracy. However, the limited five-class distribution of EDUCATION data contradicts this explanation. A plausible reason could be that education and gross household income might be independent of the energy usage and housing detail input variables employed in this study.

© ASCE

Table 1. Average accuracy of the machine learning models over 10-fold validation process

Models	ADB	CART	KNN	LDA	RFC	SVM	Mean
ATHOME	**0.58**	0.43	0.43	0.54	0.57	0.55	0.52
EDUCATION	0.38	0.3	0.28	**0.40**	**0.40**	0.33	0.35
EMPLOYHH	**0.71**	0.61	0.48	0.70	**0.71**	0.48	0.62
EQUIPMUSE	0.55	0.61	0.37	0.60	**0.66**	0.38	0.53
HHAGE	0.59	0.52	0.40	0.59	**0.61**	0.38	0.52
MONEYPY	0.34	0.26	0.21	0.31	**0.35**	0.22	0.28
NHSLDMEM	0.37	**0.99**	0.53	0.96	0.76	0.37	0.66
NUMADULT	0.77	**0.99**	0.64	0.96	0.77	0.54	0.78
NUMCHILD	0.69	**0.99**	0.67	0.97	0.75	0.68	0.79
TEMPGONE	**0.56**	0.51	0.32	0.51	0.55	0.37	0.47
TEMPGONEAC	0.47	0.63	0.31	0.58	**0.64**	0.31	0.49
TEMPHOME	0.55	0.60	0.37	0.55	**0.63**	0.41	0.52
TEMPHOMEAC	0.56	0.66	0.32	0.62	**0.67**	0.30	0.52
TEMPNITE	**0.58**	0.54	0.33	0.50	**0.58**	0.35	0.48
TEMPNITEAC	0.57	0.61	0.30	0.53	**0.63**	0.31	0.49
USEWWAC	0.86	0.85	0.74	0.83	**0.88**	0.76	0.82

DISCUSSION

Several findings can be derived by examining the results presented in the previous section. Firstly, the machine learning models achieve nearly 99% accuracy for certain target variables, such as the number of household members (NHSLDMEM), the number of adults (NUMADULT), and the number of children (NUMCHILD) living in the household. In contrast, an average accuracy of 37% is observed for other variables, including winter temperature when no one is at home (TEMPGONE), winter temperature at night (TEMPNITE), highest level of education obtained by the respondent (EDUCATION), and annual gross household income (MONEYPY). These results clearly demonstrate that the models can predict NHSLDMEM, NUMADULT, NUMCHILD, and USEWWAC with a high degree of confidence. The overall performance of the machine learning models on the target variables EDUCATION and MONEYPY suggests that the behavior of occupants with different educational backgrounds and income levels may not be significantly distinct.

These results demonstrate that machine learning models can be employed with a certain degree of confidence to automate some steps in the building occupant persona generation process. However, they also highlight that there is room for improvement in this area. The evaluation outcomes of the models emphasize the following findings: 1) Machine learning tools can be utilized to semi-automate the building occupant persona development process; 2) Given sufficient training data, machine learning models can achieve 90% or higher accuracy in predicting certain occupant characteristics; and 3) Some occupant characteristics may not depend on occupant behavior or energy consumption properties.

Answer to the research questions. RQ1: The machine learning models' performance varied across the 16 target variables, as shown in Table 1. Temperature preference attributes achieved 50% average accuracy, reasonable given their 7-category distribution. Variations in accuracy among these variables were within 3%, reflecting consistent input features. The variables related to the number of occupants yielded notably high accuracy, with the decision tree classifier achieving 99% for all three. This suggests strong correlations with input features like energy consumption and housing structure, revealing information about occupants. Respondent age achieved 51% average accuracy, respondent employment status reached 61%, while education level hit only 35%. These lower scores suggest insufficient correlated input features for classification. These variables focus on individual respondents, unlike most input features which

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

represent entire housing units. This mismatch makes it challenging for models to classify individual occupant traits from household data.

RQ2: Table 1 displays the average Accuracy achieved by each model. The random forest classifier (RFC), decision tree classifier (CART), and linear discriminant analysis classifier (LDA) all share the top spot with an average accuracy of 63%. RFC stands out by achieving the highest accuracy for 9 out of 16 target variables. This success of RFC can be attributed to its ability to make precise predictions, crucial for strategic decision-making. The Adaboost classifier (ADB) ranks second, securing the best accuracy for 4 out of 16 target variables. Notably, the decision tree classifier attained a remarkable 99% accuracy for 3 target variables. On the other hand, the K nearest neighbor classifier (KNN) and support vector machine classifier (SVM) achieved an overall average accuracy of 42%.

CONCLUSION

This study investigates the application of machine learning techniques to simplify the process of creating building occupant personas. The research utilizes data from the 2015 Residential Energy Consumption Survey and constructs six machine learning models. By introducing a machine learning-based approach to crafting occupant personas, this study expands the existing knowledge in this field. The outcomes show that, with sufficient data, it's possible to automatically predict occupant characteristics with a reasonable level of confidence, thus improving our comprehension of occupants and customizing living conditions to better suit their needs. However, it's important to note that complete automation wasn't achieved in this research, suggesting potential for further exploration in this domain. The study suggests more investigation focused on achieving full automation, which presents a promising avenue for future studies. Despite extensive experimentation, no single model out the six emerged as superior in classifying all 16 occupant characteristics with greater confidence than others. Therefore, adopting a model selection strategy, wherein the most effective models are chosen for each specific characteristic, is recommended. This approach aims to enhance overall predictive accuracy.

REFERENCES

Agee, P., Gao, X., Paige, F., McCoy, A., and Kleiner, B. (2021). A human-centred approach to smart housing. *Building research & information*, 49(1), 84-99.

Alavi, H. S., Churchill, E. F., Wiberg, M., Lalanne, D., Dalsgaard, P., Fatah gen Schieck, A., and Rogers, Y. (2019). Introduction to human-building interaction (hbi) interfacing hci with architecture and urban design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(2), 1-10.

Amasyali, K., and El-Gohary, N. M. (2021). Real data-driven occupant-behavior optimization for reduced energy consumption and improved comfort. *Applied Energy*, 302, 117276.

Anik, S. M. H., Gao, X., Meng, N., Agee, P. R., and McCoy, A. P. (2022). A cost-effective, scalable, and portable IoT data infrastructure for indoor environment sensing. *Journal of Building Engineering*, 49, 104027.

Bell, J., Mabb, J., Garcia Hansen, V., Bergman, R., and Morawska, L. (2003). Occupant health and productivity: an Australian perspective. In *Summary Book of The CIB 2003 International Conference on Smart and Sustainable Built Environment (SASBE 2003)* (pp. 1-8). Queensland University of Technology.

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

© ASCE

Bortolini, R., and Forcada, N. (2021). Association between building characteristics and indoor environmental quality through post-occupancy evaluation. *Energies*, 14(6), 1659.
Brangier, E., and Bornet, C. (2011). Persona: A method to produce representations focused on consumers' needs. *Human Factors and ergonomics in Consumer Product Design: methods and techniques*, 37-61.
Buckman, A. H., Mayfield, M., and BM Beck, S. (2014). What is a smart building?. *Smart and Sustainable Built Environment*, 3(2), 92-109.
Carlucci, S., De Simone, M., Firth, S. K., Kjærgaard, M. B., Markovic, R., Rahaman, M. S., and Van Treeck, C. (2020). Modeling occupant behavior in buildings. *Building and Environment*, 174, 106768.
Deng, Z., and Chen, Q. (2018). Artificial neural network models using thermal sensations and occupants' behavior for predicting thermal comfort. *Energy and Buildings*, 174, 587-602.
Fabi, V., Andersen, R. V., and Corgnati, S. P. (2013). Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings. *HVAC&R Research*, 19(5), 635-645.
Higginbotham, D. J. (2000). Formulating research questions. *The efficacy of augmentative and alternative communication: Toward evidence-based practice*, 43-57.
Hu, S., Yan, D., Azar, E., and Guo, F. (2020). A systematic review of occupant behavior in building energy policy. *Building and Environment*, 175, 106807.
Kelleher, J. D., Mac Namee, B., and D'arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.
Khan, J., Hussain, T., Javed, M. T., and Meraj, S. (2022). Effect of Indoor Environmental Quality on Human Comfort and Performance: A Review. *Ergonomics for Improved Productivity: Proceedings of HWWE 2017* Volume 2, 335-345.
Kim, J., Hong, T., Jeong, J., Lee, M., Lee, M., Jeong, K., and Jeong, J. (2017). Establishment of an optimal occupant behavior considering the energy consumption and indoor environmental quality by region. *Applied Energy*, 204, 1431-1443.
Kim, J., Zhou, Y., Schiavon, S., Raftery, P., and Brager, G. (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and environment*, 129, 96-106.
King, R. D., Feng, C., and Sutherland, A. (1995). Statlog: comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence an International Journal*, 9(3), 289-333.
Li, X., and Yao, R. (2020). A machine-learning-based approach to predict residential annual space heating and cooling loads considering occupant behaviour. *Energy*, 212, 118676.
Lin, Z., Yang, R., Li, K., Yi, G., Li, Z., Guo, J., and Huang, G. (2020). Establishment of age group classification for risk stratification in glioma patients. *BMC neurology*, 20, 1-11.
Macia, N., and Bernadó-Mansilla, E. (2014). Towards UCI+: a mindful repository design. *Information Sciences*, 261, 237-262.
Malik, J., Bardhan, R., Hong, T., and Piette, M. A. (2022, September). Developing occupant archetypes within urban low-income housing: A case study in Mumbai, India. In *Building Simulation* (Vol. 15, No. 9, pp. 1661-1683). Beijing: Tsinghua University Press.
Mitchell, T. M. (1997). Artificial neural networks. *Machine learning*, 45(81), 127.
Mo, Z. (2002, August). Intelligent buildings and intelligent agents-a human-centered framework for building controls. In *Proceedings of the 41st SICE Annual Conference. SICE 2002.* (Vol. 5, pp. 3151-3156). IEEE.

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

© ASCE

Ortiz, M. A., and Bluyssen, P. M. (2019). Developing home occupant archetypes: First results of mixed-methods study to understand occupant comfort behaviours and energy use in homes. *Building and Environment*, 163, 106331.

Pan, S., Wang, X., Wei, Y., Zhang, X., Gal, C., Ren, G., and Liu, J. (2017, December). Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in Shanghai residences. In *Building simulation* (Vol. 10, pp. 889-898). Tsinghua University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

Peng, Y., Rysanek, A., Nagy, Z., and Schlüter, A. (2018). Using machine learning techniques for occupancy-prediction-based cooling control in office buildings. *Applied energy*, 211, 1343-1358.

Pereira, P. F., Ramos, N. M., and Ferreira, A. (2020). Room-scale analysis of spatial and human factors affecting indoor environmental quality in Porto residential flats. *Building and environment*, 186, 107376.

Shan, X., Melina, A. N., and Yang, E. H. (2018). Impact of indoor environmental quality on students' wellbeing and performance in educational building through life cycle costing perspective. *Journal of Cleaner Production*, 204, 298-309.

Song, Y., Mao, F., and Liu, Q. (2019). Human comfort in indoor environment: a review on assessment criteria, data collection and data analysis methods. *IEEE Access*, 7, 119774-119786.

Sun, K., and Hong, T. (2017). A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. *Energy and Buildings*, 146, 383-396.

Takai, S., and Ishii, K. (2010). A use of subjective clustering to support affinity diagram results in customer needs analysis. *Concurrent Engineering*, 18(2), 101-109.

EIA (US Energy Information Administration). (2015). Residential Energy Consumption Survey Data, accessed April 24, 2023. https://www.eia.gov/consumption/residential/data/2015.

Yarlagadda, A., Murthy, J. V. R., and Prasad, M. K. (2015). A novel method for human age group classification based on Correlation Fractal Dimension of facial edges. *Journal of king saud university-computer and information sciences*, 27(4), 468-476.

Zhang, C., Liu, C., Zhang, X., and Almpanidis, G. (2017). An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications*, 82, 128-150.

Zhongguo, Y., Hongqi, L., Ali, S., and Yile, A. (2017). Choosing classification algorithms and its optimum parameters based on data set characteristics. *Journal of Computers*, 28(5), 26-38.

Downloaded from ascelibrary.org by VPI & SU on 06/19/24. Copyright ASCE. For personal use only; all rights reserved.

© ASCE