ENERGY DEMAND PREDICTION FOR RESIDENTIAL BUILDINGS AT DIFFERENT CLIMATE CONDITIONS BASED ON DIFFERENT DATA-DRIVEN MODELS

O. K. İşeri[1], S. Akin[1] and I. G. Dino[1]
[1]Department of Architecture, Middle East Technical University, Turkey, Ankara

ABSTRACT

The prediction of energy demand provides insights to improve energy performance and to reduce the environmental impact of buildings. In this paper, a data-driven prediction model for the energy demand of small-scale building stock at two different cities in Turkey, by using various prediction methods, i.e., Multiple Linear Regression (MLR), Random Forest (RF), Multilayer Neural Networks (NNN), and Multiple-Output Multilayer Neural Networks is presented. These data-driven models are suitable to hasten the simulation process of Urban Building Energy Modeling (UBEM), where the model complexity is high and requires long-term planning, especially for residential buildings. The energy demand results are presented based on RMSE, MAE, r2, particularly for heating, cooling, and lighting demands of the city models.

INTRODUCTION

The existing building stock represents 40% of the total energy demand by including zone heating, cooling, lighting, domestic hot water, and equipment energy consumptions (EEA, 2008). On the other hand, as the effects of climate change are beginning to make an impact on our lives gradually, the urbanization process should consider organizing sustainable energy and resource use (Stocker et al., 2013). The preparation for the retrofit transformation starts with the estimation of the future conditions. Urban Building Energy Modeling (UBEM) is a useful method to track and examine the existing or new building stock at urban level. UBEM helps to monitor climate and energy demand patterns regarding the adaptation of different refurbishment scenarios in terms of energy efficiency (Hong et al., 2020; Kontokosta et al., 2018). Various studies focused on the energy demand, indoor occupant thermal comfort prediction in buildings for energy-efficieny design transformation (EIE, 2010; McKenna et al., 2013; Suganthi & Samuel, 2012).

The formation of the estimation process for UBEM can be realized with the adaptation of top-down (i.e., data-driven) or bottom-up models. The bottom-up models are generally applied to simulations methods with zone-based thermal models (Kavgic et al., 2010; Reinhart et al., 2013), and the top-down models are used for comprehending the overall energy demand and comfort trends of existing building stocks (Gassar et al., 2019; Howard et al., 2012). According to data availability, the top-down models are much faster than the bottom-up models and also propose a more straightforward solution to analyze larger building stock archetype datasets.

There are different top-down models used by researchers to predict the energy demands of building stock in the framework of UBEM (Kontokosta et al., 2018). Based on this purpose, several specialized algorithms are developed which can forecast the energy demand changes over time or estimate the retrofit scenarios for future conditions (Moghadam et al., 2019; Papadopoulos et al., 2018; Vásquez et al., 2016). For instance, the linear regression models are suitable for understanding the energy demand pattern of building stock. However, they are not efficient enough to provide a solution for the non-linear complex models (Sjögren et al., 2007; Tabasi et al., 2016). Lately, neural network models have become popular, and, these models are potent to estimate the energy demand of building stock by dealing with larger data sizes and complex models (Gajowniczek et al., 2017; Kankal et al., 2011).

The building stock for the developing countries is expected to have drastic changes in the future, e.g., in number, area, and energy demand. For climate change mitigation, understanding these countries' future energy trajectories is critical to identify future environmental challenges. As a developing country, Turkey, and its building stock have been started to be studied, since city planning and energy issues have gained importance in recent years. Several studies were conducted to forecast the electricity and gas demand of Turkey's building stock with the neural network algorithms (Hamzaçebi, 2007; Kaynar et al., 2011).

The main challenge for the top-down models is the lack of available the recorded energy usage data or the organized the building stock datasets based on construction properties, energy demand profiles, and the number of occupants (Chen et al., 2018; Duranton & Puga, 2014; Pauliuk et al., 2013). Related, various researchers have studied the classification of building

 DOI: 10.3217/978-3-85125-786-1-78

data (Davila et al., 2016; Filogamo et al., 2014; Tardioli et al., 2018), and the several city managament shared the building specifications dataset for the the urban analytics that could lead the formation of the UBEM (IIO, 2020; Nyc.gov, 2020; Paris, 2018).

This paper aims to apply the different data-driven models for two different neighborhood regions and to estimate different energy demands for Kültür Mahallesi, Izmir, and Bahçelievler, Ankara located in Turkey. The simulation-based generated energy demand data are used for the training of the top-down models. The main aim of this paper to compare four different top-down model performance based-on the error prediction rate.

METHODOLOGY

In this chapter, the methodology of the four different data-driven models were presented in terms of the how the energy demand data was generated, the specifics of the data, the different types of the prediction algorithms.

The Generated Energy Demand Data (kWh/m^2)

The selected zones are one of the most densely populated neighborhoods in Turkey, namely Kültür in Izmir, and Bahçelievler in Ankara (Figure). Izmir is in ASHRAE climate zone *3A* ($2500 < CDD_{10^\circ C}$ (*Cooling Degree Days*) < 3500), therefore, the cooling energy demand is an essential objective for the region. Ankara is in *4B* ($CDD10^\circ C \leq 250$, $HDD_{18^\circ C}$ (*Heating Degree Days*) ≤ 3000) (Ashrae, 2009), and the heating energy demand has the most significant share for the total energy demand. The building typology of the two regions consists of mostly residential buildings with retail units on the ground floors. In the Izmir case, there are five-six storey 525 buildings in total 459.567 m^2, Ankara case, there are three-four storey 560 buildings in total 574.353 m^2.

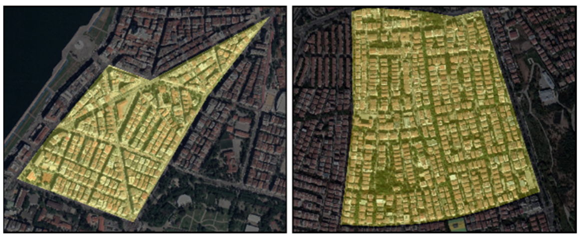

Figure 1: Kültür, Izmir (left); Bahçelievler, Ankara (right)

The publicly sharing option of the city datasets of Turkey is under the protection by the law (KVKK, 2018). Therefore, the training data was generated by employing the bottom-up building energy simulation method. There is no available energy demand data for the selected regions in Turkey (TÜİK, 2010). Therefore, the dataset was generated. There are different data sources were used about 3d and thermal modeling, e.g., geographical information data (*.gis*) for building layouts and number of flats; google images for building height, window to wall ratio; and statistical information for construction date, occupant density, and profiles. The *.gis* data was taken from the local municipality, and buildings' height information was retrieved from the google images. Other essential information was derived from the Turkish Statistical Government Bureau (TÜİK, 2010) based on city researches, e.g., occupant type and density, conditioning type.

The building stock for each neighborhood is modeled based on TS-825 and Ashrae standards (ASHRAE, 2013; TSE, 2008). All of the building energy simulations were executed with EnergyPlus simulation software which works with python (Crawley et al., 2000). All simulations were executed separately for each flat unit, and peripherical buildings were introduced as context geometries to the simulations, thus, the computing cost of the simulations decreased.

The building energy simulations require two essential inputs as geometrical and climate data. The climate data for each region were selected from the years between 2003-2017, which are generated based on typical meteorological year (TMY) methodology (Jiang, 2010). TMY is useful for representing the general climate trends of the region, preventing the extreme conditions.

The Characteristics of the Data

For Izmir, the data covers the heating, cooling, and lighting energy demands, but for Ankara, only heating and lighting energy demands were included as objective (*Y*). For Izmir, the features of *the number of people per sqm* for each zone was defined in accordance with the function of the zone, e.g., living room or bedroom, and only residential building typology exists. On the other hand, *the number of people per sqm* is constant for the whole zone, and there are two different building typologies as residential and retail for Ankara. Therefore, the two datasets separate from each other.

The heating and cooling demand ratios were defined according to (TÜİK, 2010), and because of the zone cooling system ratio was highly low for the case Ankara, the zone with cooling systems were ignored for the Ankara. The zone partitioning is made in the unit plan as living, service, and bedroom zones, and the cooling system was applied as an option for living and bedroom. On the other hand, units with only heating space conditioning systems were determined as a single zone. The objectives and variables for training data of the city models can be seen, both for Izmir and Ankara, in Table 1.

 DOI: 10.3217/978-3-85125-786-1-78

Table 1:
Example of a table features of the generated data, train/test/validation, Izmir/Ankara

Feature	Data	Type
Heating, *kWh/m²*	Continuous	Y_1
Lighting, *kWh/m²*	Continuous	Y_2
Cooling*, *kWh/m²*	Continuous	Y_3
Building Function**	Discrete,2	X_{27}
Floor Area, *sqm*	Continuous	X_1
Floor Number	Discrete,6	X_2
Number of people per sqm** (constant)	Continuous	X_{23}
Num. of pe. per sqm service*	Continuous	X_{24}
Num. of pe. per sqm living*	Continuous	X_{25}
Num.. of pe. per sqm bedroom*	Continuous	X_{26}
Construction Type	Discrete,2	X_3
Occupancy Schedule	Discrete,3	X_4
Window Count	Discrete ,4	X_5
Window Area, *sqm*	Continuous	X_6
Window Orientation	Discrete,2	X_7-X_{14}
Context Opening	Discrete,2	X_{15}-X_{22}

**only for Izmir, ** only for Ankara*

The cooling conditioning system ratio is just only 20% for the region of Izmir (TÜİK, 2010), and therefore, the total number of data for cooling represents 20% and heating is 80% of total data. For Ankara, all the building units have a heating system, and the total number of data is higher than Izmir. On the other hand, the objective and zone types are more than one for the case of Izmir. The data shape of the Ankara dataset is 5928,20; and it is 2023,25 for the Izmir dataset.

Regression for Building Energy Demand

The data-driven models are practical tools for energy simulation models in terms of computing cost and alternative interpretation for different cases. Regression models can form the functions for inter-relation between independent variables to estimate the dependent variables, which is building energy demand in this study (Fumo & Rafe Biswas, 2015). According to the number of model variables, the complexity could increase, which means that the inclusion of different algorithms to train the data and to validate the outputs becomes hard to manage.

Multiple Linear Regression

The multiple linear regression (MLR) aim to form the relationship of the variables by fitting a linear mathematical equation to the data. If there is more than one predictor variable ($X_1,\ldots X_n$), and these variables are related to predicting the response (Y) value, the MLR (Freedman, 2009) is suitable for the task.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots + \beta_n X_n + \mathcal{E} \quad (1)$$

In the given equation (1), Y is the dependent variable, $X_1, X_2, \ldots X_n$ are the predictor variables, $\beta_0, \beta_1, \ldots \beta_n$ are the regression coefficients, and $\mathcal{E}$ is the error term of the model.

Various studies were adapted MLR to predict the energy demand of residential stock by using physical and operational parameters of buildings (Fumo & Rafe Biswas, 2015; Gassar et al., 2019).

Random Forest

Random Forest (RF) is a supervised machine learning algorithm for the classification or regression purposes that contains a large number of decision trees. Statistically, the algorithm aims to decrease the variance and bias of the model. A decision tree is composed of a recursive loop to split the non-overlapping predictions randomly. The predictions are the result of the sum of all trees' mean value.

There are different examples of RF in the field of building energy modeling to predict the energy demand of buildings or to classify the prominent features for energy demand (Ahmad et al., 2017; Kontokosta & Tull, 2017).

Neural Network for Regression

Neural Networks (NN) are the imitated representation of the human learning process by using the input data to regress or classify output variables. NN is composed of multiple layers, i.e., input, hidden, output layer. Each layer consists of different nodes that transmit the information with weighted connections and transfer functions. According to the layer and node number, the complexity or learning capacity of the model could increase or decrease. Based on the output activation function, the model can be used for regression or classification. In this study, multi-layer perceptron neural networks with single-output (NN_S) and multiple-output (NN_M) examples are used.

NN can predict the energy demand of building stocks because of their capacity to handle complex datasets (Howard et al., 2012; Kaynar et al., 2011). For building energy models, the predictors can vary from the physical to operational characteristics of the data, e.g., window area, window direction, occupant density. In this study, two different cases contain a different number of data sizes and features to compare the performance for the urban building energy demand's estimation process, e.g., non-binary and all parameters.

NN models could come across the over-fitting problem if the precision of the algorithm reaches 100%, the state of learning could transform to unintended memorization, which decreases the predictive performance of the algorithm for new data. The problem could be overcome with hyperparameter tuning (Østergård et al., 2018). MLR and RF algorithms have lower learning capacity rather than the NN models. The data size is also directly related to the tuning process and learning performance. Therefore, the performance of the algorithm could be increased by controlling the batch size, training-test split ratio, and the number of predictors. Hyper-

 DOI: 10.3217/978-3-85125-786-1-78

parameter tuning could lead to balance the loss function value to generalize the algorithm performance in NN models, e.g., regularization or dropout, regulating learning rate or neuron number, starting weight with randomization. The selection of relevant hyperparameter values could lead to acceptable learning ratios for the algorithm—this study exemplifies the process of tuning in the aspect of the NN_M and NN_S model training process.

The difference between NN_S and NN_M models relies on the number of output variables that they estimate. In some cases, multiple outputs could be advantageous for the learning process because multiple output variables work as an extra feature in the learning data. For this study, three types of energy demand were predicted for the selected region of Izmir in terms of cooling, heating, and lighting. The same procedure for Ankara was followed for the heating and lighting output variables. NN_M and NN_S results are presented with spatial distribution on the observation regions and comparative plots (Fonseca & Schlueter, 2015).

The Evaluation of Performance

The prediction of the output variables that are based on the self and inter-correlation of independent variables makes possible the process of regression on pre-defined metrics. For this study, Mean Squared Error (MSE) is used as evaluation metrics for training and validation loss calculation. It is the average of the squared difference between the target value and the value predicted, and it punishes the small errors that could cause over-estimation (2).

$$\text{MSE} = \frac{1}{N}\sum (y - y\_pred)^2 \qquad (2)$$

Three different metrics are used in the study to compare the performance of algorithms between training and test data, e.g., Root-Mean-Squared-Error(RMSE) (3), and Coefficient of Determination (r^2) (4) and Mean Absolute Error (MAE) (5).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y-y\_pred)^2}{N}} \qquad (3)$$

$$r^2 = 1 - \frac{y\_pred}{y} \qquad (4)$$

$$\text{MAE} = \frac{1}{N}\sum |y - y_{pred}| \qquad (5)$$

RESULTS

The Statistics of the Data

The training data were generated with EnergyPlus building energy simulation software using statistical indicators for thermal characteristics of the model and *.gis* information for the building stock geometries. Table 2 points out the statistical evaluation of the generated data.

Table 2:
Statistical representations of objectives, Izmir/Ankara

Izmir (kWh/sqm) - Count:2023						
Feature	**Mean**	**Min**	**Max**	**Std.**	**Skew**	**Kurt**
Heating	34.65	11.08	117.38	12.04	1.55	4.81
Lighting	8.01	5.72	13.18	1.73	1.00	0.35
Cooling	32.00	7.14	94.70	13.88	0.80	0.34
Ankara (kWh/sqm) - Count:5928						
Feature	**Mean**	**Min**	**Max**	**Std.**	**Skew**	**Kurt**
Heating	69.99	20.09	152.33	21.72	0.42	-0.43
Lighting	16.74	15.26	19.67	0.96	0.50	-0.77

The heating demand of Izmir values was found lower than Ankara because of the fewer heating degree days; however, the data distribution was found more homogenous for the case of Ankara according to the kurtosis value. In addition, lower heating demand values are dense because there is a positively skewed formation for both curves. For the lighting demand, the values are higher for Ankara due to less amount of solar radiation (*W/m²*). The cooling values of Izmir are similar to the heating values except for the kurtosis value, which is 0.34. It states that the data distribution curve depth is lower than the normal distribution curve.

The Results for Multi-Linear Regression

Table 3 points out the MLR algorithm performance results. During the training process, different test-ratio values were tested; therefore, the most successful results were added in the table based on comparison metrics. In the first case, which is the binary feature extracted data, the window orientation ($X_7 – X_{14}$) and context opening parameters ($X_{15} – X_{22}$) were taken out from the set to observe the performance of the algorithm only with continuous and non-binary variables. The other set of training was executed with all parameters. As a result, the performance of the values was found higher than the non-binary continuous and categorical variables as it is expected.

Table 3:
The results for Multi-Linear Regression

Y	Parameter	r^2	MAE	RMSE	Test Set	Loc.
Cooling	*Non-binary	0.564	0.106	0.131	10%	Izm
Heating		0.407	0.078	0.104		
Lighting		0.376	0.169	0.220	20%	
Cooling	All, 25	0.715	0.088	0.111	10%	
Heating		0.657	0.066	0.087		
Lighting		0.471	0.163	0.209		
Heating	*Non-binary	0.826	0.073	0.093	10%	Ank
Lighting		0.779	0.103	0.137		
Heating	All, 20	0.969	0.031	0.040	20%	
Lighting		0.811	0.099	0.128		

* X_1, X_2, X_{24}, X_{25}, X_{26}, X_6

r^2 values state that the learning capacity of the MLR model increases when the size of the dataset increases. On the other hand, the case of Izmir consists of more data features, but it seems that these extra features did

 DOI: 10.3217/978-3-85125-786-1-78

not develop any rise for the performance of the model. As a result, the performance metrics of Ankara was found more accurate than the case of Izmir for heating and lighting output variables.

The Results for Random Forest

Table 4 shows the performance of the Random Forest (RF) algorithm with a similar comparison strategy to MLR. For each output variable, the performance of the model increases when the number of the feature increases, except the training process of lighting energy demand in Ankara. RF algorithm is capable of predicting energy demand values if the size of the data is at an acceptable level compared to the region of Ankara. The difference between the target and the predictor variables are highly close to zero for the heating and lighting demands of Ankara. On the other hand, for the smaller size datasets, the ratio for the test data could be increased to obtain similarly distributed test and training set.

Table 4:
The results for Random Forest

Y	Parameter	r^2	MAE	RMSE	Test Set	Loc.
Cooling	Non-binary	0.557	0.105	0.131	20%	Izm
Heating		0.560	0.064	0.092	20%	
Lighting		0.314	0.180	0.226	20%	
Cooling	All, 25	0.699	0.095	0.118	10%	
Heating		0.763	0.052	0.073	20%	
Lighting		0.451	0.167	0.210	20%	
Heating	Non-binary	0.975	0.029	0.037	10%	Ank
Lighting		0.977	0.028	0.035		
Heating	All, 20	0.977	0.027	0.036		
Lighting		0.809	0.096	0.128		

The Results for Neural Network

The architecture of NN was composed differently for İzmir and Ankara cases. Ankara case is trained with four hidden and one dropout layer, and three layers and one dropout for Izmir. As activation function rectified linear unit (relu) was selected. The hyperparameter tuning was executed for better training performance as managing the number of hidden layers and neurons, training batch size, training time, and learning rate.

Table 5:
The results for single-output Neural Network

Y	Parameter	r^2	MAE	RMSE	Test Set	Loc.
Cooling	Non-binary	0.424	0.095	0.124	10%	Izm
Heating		0.343	0.064	0.097		
Lighting		0.203	0.181	0.209		
Cooling	All, 25	0.700	0.071	0.092		
Heating		0.761	0.042	0.058		
Lighting		0.639	0.107	0.145		
Heating	Non-binary	0.770	0.062	0.080	10%	Ank
Lighting		0.788	0.100	0.134		
Heating	All, 20	0.978	0.017	0.024		
Lighting		0.852	0.091	0.117		

Table 5 points out that NN_S is performed successfully for Ankara with all parameters, even it could not pass the performance of RF for other comparison cases. Similarly, the results of Izmir with non-binary parameters were found lower than the Ankara case. Consequently, the number of instances was found more effective than the number of attributes as the results of the Izmir case were less accurate.

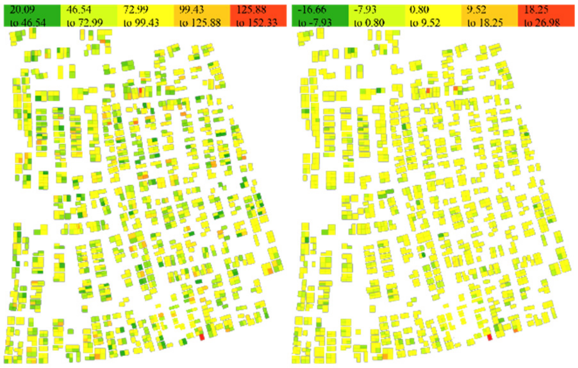

Figure 2: Ankara heating demand (left); predicted demand error spatial distribution (right), kWh/sqm

Figure 2 presents the spatial distribution of the simulation-based generated heating demand and data-driven based generated heating demand prediction error for the region of Ankara. For the map at the left, the heating energy demand changes averagely between 72.99 to 99.43 kWh/sqm. The error values are distributed close to between 0.80 and 9.52 kWh/sqm for Ankara.

Figure 3: Izmir cooling demand (left); predicted demand error spatial distribution (right), kWh/sqm

Figure 3 is the spatial distribution of the simulation-based generated cooling demand and data-driven based generated cooling demand prediction error for the region of Izmir. The cooling energy demand averagely varies between 24.66 and 42.16 kWh/sqm, which is lower than the median value. The error gap is close to the yellow bar that is bounded between -3.90 to 8.42 kWh/sqm for Kültür, Izmir

 DOI: 10.3217/978-3-85125-786-1-78

Table 6:
The results for multiple-output Neural Network

Y	Parameter	r^2	MAE	RMSE	Test Set	Loc.
Cooling	Non-binary	0.531	0.093	0.117	10%	Izm
Heating		0.376	0.067	0.092		
Lighting		0.111	0.162	0.211		
Cooling	All, 25	0.756	0.065	0.084		
Heating		0.692	0.059	0.075		
Lighting		0.454	0.096	0.139		
Heating	Non-binary	0.936	0.030	0.040	10%	Ank
Lighting		0.635	0.098	0.133		
Heating	All, 20	0.976	0.020	0.027		
Lighting		0.690	0.085	0.117		

Table 6 shows the results of Izmir and Ankara for non-binary and all parameters' cases with NN_M. The training performances were found higher for Ankara. The non-binary case in Ankara performed better than the all parameters case in Izmir. Therefore, the results differed more when the data size of the number of instances increased for NN_M. The difficulty of the prediction was more for Izmir as the output variable composes of three output variables. Therefore, the architecture of NN was found different for two regions (Figure 4, top-right). The simpler architecture was proposed for the cases of Izmir. However, the hyperparameter tuning was found longer for Izmir, and the model demanded more time to converge. For both of the regions, the estimation of lighting value error was higher than the other energy demand types, according to Figure 4.

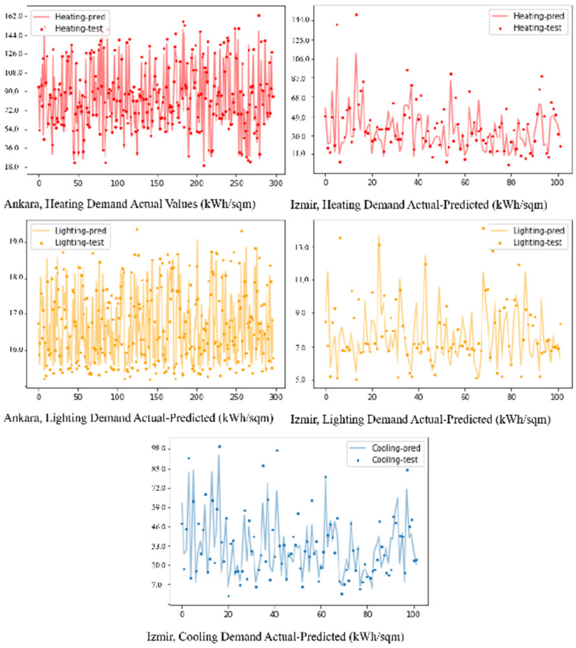

Figure 4: Visualization of target and predicted difference for multiple-output Neural Network

Based on the dataset and the problem, different data-driven problems could be faced, e.g., under-fitting and over-fitting problems. As a suggestion to the problem, hyperparameter tuning could be an effective solution. Lastly, the prediction performance was found more most descriptive for the cooling demand in the Izmir case and the heating demand in Ankara for all comparative cases.

DISCUSSION

The model complexity proceeds in parallel with the learning capacity based on the accuracy of the comparison metrics of four data-driven algorithms. The RF was successful for each case based on comparison metrics, i.e., r^2, MSE, RMSE. However, if the size of the dataset increases in terms of the number of instances and attributes, the NN_M and NN_S algorithms could perform better. Due to the shape of the output variable for NN_M, the algorithm demands more complex architecture for the regression. However, it is still faster compared to the bottom-up simulation-based approaches in the model training process. NN_S and NN_M are capable of managing the datasets of this study. Nevertheless, RF and NN_M reached the most successful results. On the other hand, the tuning process was difficult for NN_M. For instance, the over-fitting problem for the Izmir cases and the under-fitting problem for the Ankara cases have occurred. In addition, applying the random weight initialization as tuning action was helpful for the problem; otherwise, the algorithm performance can be stuck to lower the performances.

CONCLUSION

This paper presents different data-driven models for the urban building energy demand prediction by comparing different cities' neighborhoods. The selected four data-driven models were used to predict heating, cooling, and lighting energy demand of a neighborhood in a hot-humid climate of Izmir and heating, lighting energy demand of a neighborhood in a cold-dry climate of Ankara. The building energy simulations were conducted with a simulation-based brute-force methodology to produce a training dataset. The generated building energy demand data consist of the physical and operational properties of the flats in the buildings for each district. Then, the two data are used for the data-driven models. Two different cases are tested for each region with extracting the binary parameters from the cluster of features and using all available parameters to compare the performance of algorithms in terms of different features. The RF and NN_M have performed the most accurate results in each case. The cases of Ankara were found more accurate based on the comparison metrics, the size of the dataset consist of more instances. In conclusion, the data-driven models provide opportunities for the UBEM simulations, and the authors developed a broad framework for the dataset arrangement for urban energy modeling.

 DOI: 10.3217/978-3-85125-786-1-78

REFERENCES

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, *147*, 77–89. https://doi.org/https://doi.org/10.1016/j.enbuild.2017.04.038

ASHRAE. (2009). ASHRAE climatic design conditions 2009/2013/2017. Retrieved February 4, 2020, from http://ashrae-meteo.info/

ASHRAE. (2013). *ASHRAE Standard 90.1-2013 -- Energy Standard For Buildings Except Low-rise Residential Buildings*.

Chen, Y., Hong, T., Luo, X., & Hooper, B. (2018). Development of City Buildings Dataset for Urban Building Energy Modeling. *Energy & Buildings*. https://doi.org/10.1016/j.enbuild.2018.11.008

Crawley, B. D. B., Pedersen, C. O., Lawrie, L. K., & Winkelmann, F. C. (2000). EnergyPlus : Energy Simulation Program. *ASHRAE Journal*, (April), 49–56. https://doi.org/10.1.1.122.6852

Davila, C. C., Reinhart, C., & Bemis, J. (2016). Modeling Boston: A workflow for the generation of complete urban building energy demand models from existing urban geospatial datasets. *Energy*, *117*, 237–250. https://doi.org/10.1016/j.energy.2016.10.057

Duranton, G., & Puga, D. (2014). Chapter 5 - The Growth of Cities. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 781–853). https://doi.org/https://doi.org/10.1016/B978-0-444-53540-5.00005-7

EEA, E. E. A. (2008). Energy and environment report 2008. In *Energy*. https://doi.org/10.2800/10548

EIE, I. E. E. (2010). *Monitoring of EU and national energy efficiency targets : ODYSSEE-MURE 2010*.

Filogamo, L., Peri, G., Rizzo, G., & Giaccone, A. (2014). On the classification of large residential buildings stocks by sample typologies for energy planning purposes. *Applied Energy*, *135*, 825–835. https://doi.org/10.1016/j.apenergy.2014.04.002

Fonseca, J. A., & Schlueter, A. (2015). Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*, *142*, 247–265. https://doi.org/https://doi.org/10.1016/j.apenergy.2014.12.068

Freedman, D. A. (2009). *Statistical Models: Theory and Practice* (2nd ed.). https://doi.org/10.1017/CBO9780511815867

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, *47*, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Gajowniczek, K., Nafkha, R., & Ząbkowski, T. (2017). *Electricity peak demand classification with artificial neural networks*. 307–315. https://doi.org/10.15439/2017F168

Gassar, A. A. A., Yun, G. Y., & Kim, S. (2019). Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy*, *187*, 115973. https://doi.org/10.1016/j.energy.2019.115973

Hamzaçebi, C. (2007). Forecasting of Turkey's net electricity energy consumption on sectoral bases. *Energy Policy*, *35*(3), 2009–2016. https://doi.org/https://doi.org/10.1016/j.enpol.2006.03.014

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, *168*, 106508. https://doi.org/https://doi.org/10.1016/j.buildenv.2019.106508

Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, *45*, 141–151. https://doi.org/10.1016/j.enbuild.2011.10.061

IIO. (2020). İstanbul İstatistik Ofisi. Retrieved May 7, 2020, from https://istatistik.istanbul/

Jiang, Y. (2010). Generation of typical meteorological year for different climates of China. *Energy*, *35*(5), 1946–1953. https://doi.org/10.1016/j.energy.2010.01.009

Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Applied Energy*, *88*(5), 1927–1939. https://doi.org/https://doi.org/10.1016/j.apenergy.2010.12.005

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, *45*(7), 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021

Kaynar, O., Yilmaz, I., & Demirkoparan, F. (2011). Forecasting of natural gas consumption with neural network and neuro fuzzy system. *EGU 2010*, *26*(2), 221–238.

 DOI: 10.3217/978-3-85125-786-1-78

Kontokosta, C. E., Prof, A., Informatics, U., & Johnson, N. (2018). *A Dynamic Spatial-Temporal Model of Urban Carbon Emissions for Data-Driven Climate Action by Cities*.

Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy*, *197*, 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005

KVKK, K. V. K. K. (2018). *KİŞİSEL VERİLERİN KORUNMASI VE İŞLENMESİ POLİTİKASI*. Ankara.

McKenna, R., Merkel, E., Fehrenbach, D., Mehne, S., & Fichtner, W. (2013). Energy efficiency in the German residential sector: A bottom-up building-stock-model-based analysis in the context of energy-political targets. *Building and Environment*, *62*, 77–88. https://doi.org/https://doi.org/10.1016/j.buildenv.2013.01.002

Moghadam, S. T., Coccolo, S., Mutani, G., Lombardi, P., Scartezzini, J.-L., & Mauree, D. (2019). A new clustering and visualization method to evaluate urban heat energy planning scenarios. *Cities*, *88*, 19–36. https://doi.org/https://doi.org/10.1016/j.cities.2018.12.007

Nyc.gov. (2020). NYC Analytics. Retrieved from https://www1.nyc.gov/site/analytics/index.page

Østergård, T., Jensen, R. L., & Maagaard, S. E. (2018). A comparison of six metamodeling techniques applied to building performance simulations. *Applied Energy*, *211*, 89–103. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.10.102

Papadopoulos, S., Bonczak, B., & Kontokosta, C. E. (2018). Pattern recognition in building energy performance over time using energy benchmarking data. *Applied Energy*, *221*, 576–586. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.03.079

Paris, C. of. (2018). Paris Open City Data. Retrieved February 19, 2020, from https://opendata.paris.fr/pages/home/

Pauliuk, S., Sjöstrand, K., & Müller, D. B. (2013). Transforming the Norwegian Dwelling Stock to Reach the 2 Degrees Celsius Climate Target. *Journal of Industrial Ecology*, *17*(4), 542–554. https://doi.org/10.1111/j.1530-9290.2012.00571.x

Reinhart, C., Dogan, T., Jakubiec, A., Rakha, T., & Sang, A. (2013). UMI - An Urban Simulation Environment For Building Energy Use, Daylighting And Walkability. *13th Conference of International Building Performance Simulation Association, Chambéry, France, August 26-28 - 476*, 476–483. Retrieved from http://web.mit.edu/SustainableDesignLab/publications/umi_introduction.pdf

Sjögren, J. U., Andersson, S., & Olofsson, T. (2007). An approach to evaluate the energy performance of buildings based on incomplete monthly data. *Energy and Buildings*, *39*(8), 945–953. https://doi.org/10.1016/j.enbuild.2006.10.010

Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M. M. B., Allen, S. K., Boschung, J., … Midgley, P. M. (2013). Climate change 2013 the physical science basis: Working Group I contribution to the fifth assessment report of the intergovernmental panel on climate change. In *IPCC* (Vol. 9781107057). https://doi.org/10.1017/CBO9781107415324

Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting - A review. *Renewable and Sustainable Energy Reviews*, *16*(2), 1223–1240. https://doi.org/10.1016/j.rser.2011.08.014

Tabasi, S., Aslani, A., & Forotan, H. (2016). Prediction of Energy Consumption by Using Regression Model. *Computational Research Progress in Applied Science & Engineering ©PEARL Publication CRPASE*, *02*(03), 110–115.

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. P. (2018). Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. *Building and Environment*, *140*, 90–106. https://doi.org/https://doi.org/10.1016/j.buildenv.2018.05.035

TSE. (2008). *Ts 825: Binalarda Isı Yalıtım Kuralları* (Vol. 825).

TÜİK. (2010). *TÜRKİYE İSTATİSTİK KURUMU Turkish Statistical Institute*.

Vásquez, F., Løvik, A. N., Sandberg, N. H., & Müller, D. B. (2016). Dynamic type-cohort-time approach for the analysis of energy reductions strategies in the building stock. *Energy and Buildings*, *111*, 37–55. https://doi.org/https://doi.org/10.1016/j.enbuild.2015.11.018

 DOI: 10.3217/978-3-85125-786-1-78