Al-Khwarizmi Engineering Journal
ISSN (printed): 1818 – 1171, ISSN (online): 2312 – 0789
Vol. 20, No. 1, March, (2024), P P. 76- 88

Al-Khwarizmi Engineering Journal

Applying Scikit-Learn of Machine Learning To Predict Consumed Energy in Al-Khwarizmi College of Engineering, Baghdad, Iraq

Reem Omar Mahdi* **Wisam S. Hacham****

*,***Department of Mechatronics Engineering / Al-Khwarizmi College of Engineering / University of Baghdad/ Iraq*

*Email: omarreem07@gmail.com

Corresponding Author: **Email: wisam@kecbu.uobaghdad.edu.iq

(Received 20 July 2023; accepted 2 October 2023; Published 1 March 2024)

https://doi.org/10.22153/kej.2024.10.005

Abstract

Buildings account for 40% of energy usage worldwide. The physical characteristics of the building, the effectiveness of the heating and cooling systems, the inhabitants' activities, and the sustainability of the building are just a few of the many factors that influence a building's energy consumption. It is quite challenging to estimate the energy requirements of a structure. Estimating the building's energy demand is vital to increase sustainability and develop sustainable energy sources to lower carbon dioxide emissions from the burning of fossil fuels. The energy used in the lecture hall at the University of Baghdad (UOB), located in Baghdad, Iraq's Al-Khwarizmi College of Engineering, is explained in this study. The weather data and the building construction information were collected for a specific period and put into a specific data set. That data was used to find the value of energy consumption in the building using artificial intelligence and data analysis. A Python library called Scikit-learn is used to implement machine learning algorithms. In particular, the Multi-layer Perceptron regressor (MLPRegressor) algorithm was used to predict the consumption. The importance of this work lies in predicting the amount of energy consumed. The outcomes of this work can be used to predict the energy consumed by any building before it is built. The used methodology shows the ability to predict energy performance in educational buildings using previous results and train the model on them, and prediction accuracy depends on the amount of data available for the training in artificial intelligence (AI) steps to give the highest accuracy. The prediction was checked using root-mean-square error (RMSE) and coefficient of determination (R^2) and we arrived at 0.16 and 0.97 for RMSE and R^2, respectively.

Keywords: *Prediction, Energy Consumption, Regression algorithm, Scikit-Learn algorithm.*

1. Introduction

As the world's population grows, so does the amount of energy used in buildings. Numerous physical and sociological elements have an impact on it. The need for basic services like water, gas, and power has increased exponentially due to increasing consumer culture and urbanization activities; this demand is predicted to double by 2050 when compared to 2010[1]. Building-related activities account for more than 30% of world energy demand and a quarter of greenhouse gas emissions [2-4].

For instance, the building industry accounts for 36% of CO2 emissions and 40% of energy consumption in Europe [5]. In contrast, the construction sector is responsible for 38.9% of the total primary energy requirement (PER) in the United States [6]. Building stocks in China accounted for around 24.1% of the total public energy usage in 1996 and 27.5% in 2011. Building stocks are predicted to reach roughly 35% by 2020 [7]. Besides, in a cutting-edge city like Hong Kong, electricity generation accounts for 60% of the byproducts of fossil fuels, while buildings account for 89% of total power consumption[8].

Additionally, certain climate conditions around the world forced homes to use more energy for their heating, cooling, cooking, and refrigeration requirements. University buildings and commercial buildings, particularly offices, are thought to have high energy usage [9,10]. As a result, cutting back on energy use to achieve energy sustainability is recognized as a crucial strategy for lowering rising energy demand and emissions as well as tackling the problems associated with a warming globe because of the ongoing global rise in energy prices, so technological solutions are essential for reducing energy demand [11].

It is critical to examine the energy consumption patterns of buildings to establish energy efficiency measures that will assist us in achieving sustainability goals and net-zero greenhouse gas emission targets. The efficiency of energy-consuming equipment and appliances has improved with technological advancement. Building energy usage has been decreased by retrofitting outdated air conditioning and cooling systems, switching to new lighting technology, and changing operating habits through energy efficiency projects [12].

Here, an accurate prediction of energy used and its amount in the building will be performed to work on it to reach energy savings and different conservation measures. It will also help us select the best design for the building to reduce energy consumption.

There are two ways to predict how much energy a structure will require. The first one is based on a physical model, whereas the other is data-driven. The physical modeling technique is also known as the "forward modeling approach." The physical modeling technique involves precise information about the structure, mechanical systems, and inhabitants' activities that may not be readily available to produce a mathematical model to predict the building's energy usage. To simulate and predict the energy consumption of buildings, the forward modeling technique frequently makes use of commercial software, such as DOE-2, Design Builder, etc., as inputs. There are frequently only small differences in the outcomes across various software when the input values for the variables are the same or identical [13].

Meanwhile, the physical model could not account for the sociocultural elements that might have an impact on how much energy residents use. The data-driven technique circumvents the constraints of using physical models to anticipate consumed energy by using data analysis using well-known data sets. Typically, building sample simulations or collecting data is used to develop an energy-use database. Artificial neural networks (ANN), classification and regression trees (CART), multiple linear regression (MLR), and other data-driven methods are examples. Other data-driven strategies have also been looked into by researchers. For instance, Li et al. [14] built a hybrid teaching-learning artificial neural network model utilizing weather data, calendar data, occupancy patterns, and historical energy use data to anticipate the hourly electrical energy consumption for two educational facilities located in the USA and China.

According to previous publications, relatively few studies have attempted to predict the annual consumption of energy for residential buildings using data from actual energy usage. Based on the outcomes of simulations using commercial software, several studies have concentrated on monthly [15], daily [16-18], or hourly [16-20] energy use their benefit appears to be finding the best solution to reduce energy consumption. Additionally, in the prediction model, the impacts of occupant behaviors on energy use are typically disregarded, and the bulk of the variables focus on meteorological data [16-19, and 21] or design variables of the building's façade [15 and 21-24]. As a result of this, estimates of homes' energy use vary, and socioeconomic and demographic information is frequently left out.

Due to the vast potential of artificial intelligence (AI), it is highly likely to play a significant role in advancing the study of energy conservation in buildings. Consequently, numerous individuals passionate about the energy sector are compelled to explore the application of this technology. Artificial intelligence (AI) is the study of developing intelligent machines capable of mimicking human cognition and doing tasks that would typically need human intelligence. Machine learning, natural language processing, computer vision, robotics, and expert systems are other subfields of artificial intelligence. A subset of (AI) artificial intelligence comprises algorithms aimed to learn from data and progressively improve performance without the need for programming skills is called "Machine learning". Deep learning is a subfield of machine learning and artificial intelligence that entails training artificial neural networks with numerous layers to detect complicated patterns and representations in data. These networks are modeled after the form and functions of neural networks in the human brain. Deep learning techniques are particularly well-suited for big data processing because they can manage massive amounts of data while accurately detecting subtle patterns [25].

A subset of artificial intelligence (AI) involves teaching computers to analyze data make predictions and form conclusions without the need, for programming which it called Machine learning (ML). In simpler terms, machine learning algorithms learn on their own by using data or experience, without needing any explicit instructions. The process of machine learning involves creating mathematical models and algorithms that can identify patterns, connections, and insights in datasets. These models are trained on labeled or unlabeled data to identify patterns and predict or respond to new, unexpected data. There are various types of machine learning methodologies, including supervised, unsupervised, and reinforcement learning[26].

The input data and appropriate target labels are coupled to train the supervised learning algorithm. As opposed to supervised learning, which uses labels to guide learning, unsupervised learning uses data that has not yet been labeled. Where the trial-and-error method used in reinforcement learning is used to train an algorithm. The approach obtains new knowledge by interacting with the environment and receiving positive or negative feedback. It aims to maximize advantages by behaving correctly in different situations .

ML libraries are software packages that provide numerous tools, features, and techniques for developing and executing machine learning models. These libraries simplify and accelerate the process of developing machine learning applications. They generally include pre-implemented algorithms, feature engineering, data pretreatment methods, metrics for model evaluation, and visualization tools. ML libraries that are widely used include TensorFlow [28], PyTorch [29], Keras [30], Scikit-learn [27], and many more. Scikit-learn, a popular Python package, provides a variety of tools for machine learning applications, including regression, clustering, classification, dimensionality reduction, and model selection.

The TensorFlow library, developed by Google, is open-source and focused on numerical computation and deep learning. It offers a flexible framework for building and implementing machine learning models on many platforms. PyTorch is another popular open-source library that's mostly used for deep learning. Because PyTorch has an intuitive interface and dynamic computational graphs, it is suitable for both research and production applications. Theano or TensorFlow is the foundation upon which the Keras high-level neural network library is constructed. It provides an accessible API for creating and honing deep learning models .

The goal of this work is to develop a forecast model for energy consumption based on data from the Al-Khwarizmi Engineering College building at the University of Baghdad (UOB), in Iraq. Information was obtained on the building's architectural attributes, the social and demographic makeup of its tenants, and the quantity of energy consumed. The energy prediction model was enhanced using a multiple linear regression (MLR) technique to estimate power use. To find the model that generated the best predictions with the fewest inputs, we experimented with different training/validation ratios and counts of input variables. Predicting the energy requirements of academic buildings can be aided by the study's conclusions.

2. Multi-campus University Setting

The facilities of the UOB are dispersed over its several campuses, which are located throughout the Baghdad Governorate of Iraq in a variety of geographically dispersed places. The main campus, which occupies 325 hectares, is situated near Al-Jadriya. Al-Waziriyah, Abu Ghraib, Bab al-Mu'azzam, al-Nahda, and al-Adhamiya are the locations of other universities. All colleges are located in the province of Baghdad, which enjoys a climate that is hot and dry in summer and cold and rainy in winter, and has all seasons: summer from June to August, fall from September to November, winter from December to February, and spring from March to May. Al-Jadriya Campus is located in the heart of the Iraqi capital, Baghdad, on the Rusafa side of the Al-Jadriya complex, and is surrounded by the Tigris River from three directions, meaning that it is located on a peninsula, the capital of Baghdad. The minimum, maximum, and mean temperatures in Baghdad across all months of the year are shown in Table (1) [31].

Electricity also operates all facilities. Some college campuses have a central heating and cooling system that relies on electricity. This campus contains a mixture of old and new buildings, where many retrofit projects have been undertaken. It also contains 13x absorption coolers. Absorption coolers are thermally driven refrigeration systems that use a heat source, such as natural gas, propane, or solar energy, to generate cooling. The "13x" in the term refers to the coefficient of performance (COP) of the absorption cooler, indicating the ratio of cooling output to heat input. These coolers are commonly used in applications where electricity availability is limited or costly, or where alternative fuel sources are preferred.

Table 1,
The average minimum, maximum, and mean temperature in Baghdad

Average	Month												
Temp.(°C)	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	**Year**
Minimum	5	7	11.1	15.8	21.5	25.4	28	17	23.4	17.7	10.7	6.4	16.6
Maximum	16.8	19.6	24.6	30.7	37	42.2	45.2	47	41.3	33.7	24.1	18.2	31.6
Mean	10.9	13.3	17.8	23.3	29.3	33.8	36.6	36	32.3	25.7	17.4	12.3	24.1

The UOB has 24 colleges, 3 high study institutes, and 9 centers spread across four main campuses. In this paper, the energy consumption in Al-Khwarizmi College of Engineering (Figure 1), which is located at the Al-Jadriya campus, will be discussed, and artificial intelligence (AI) will be applied to predict the amount of energy consumption. To use it to impose mechanisms to retrofit the building to reduce the energy consumed in these buildings and arrive at buildings with less energy consumption and zero carbon emissions in the future.

Fig . 1. Lecturing building of Al-Khwarizmi College of Engineering.

3. The Dataset Characteristics

The following subsections will illustrate building types, weather information, and education calendars to demonstrate whole the characteristics of the dataset incorporated in the used methodology. Table (2) illustrates all these characteristics briefly.

3.1 Types of Buildings

All campuses have varied applications for the buildings, which also result in different energy consumption patterns and operational schedules. The buildings are divided into the following seven groups based on their main uses. To be used for different modeling and analytical purposes, these categories are given as metadata.

- Instructional Facilities: These include all structures utilized for basic and applied research as well as
- lectures, workshops, tutorials, labs, and other instructional activities;
- Library Buildings: structures that house libraries as well as expansive computer laboratories and quiet study areas for students;
- Administrative buildings: structures that are only utilized for university administration, including the HR, IT, and finance departments. These are only utilized from 8 a.m. to 3 p.m. during working hours;
- Residential buildings: structures that house students;
- Mixed-Use Buildings: Some structures combine administrative, research, and instructional activities; they include lecture halls, offices, and study areas;
- Sports Facilities: These structures are utilized for sporting and leisure activities; and
- Other structures: These include facilities that do not fit into any of the categories mentioned above, such as greenhouses and animal kennels that are open round-the-clock.

  The College of Engineering Khwarizmi is classified as a mixed-use building, as the building contains classrooms, laboratories, and some administrative rooms. It also contains a library for the college.

3.2 Weather information

Weather information is crucial for correctly forecasting consumption patterns since it has a substantial correlation to utility usage [32] [33]. We can see from the dataset that the UOB network uses more energy during the summer than it does during the winter.

We must utilize the HAVC system and air conditioning in some sections of the building since the summer weather is too hot and dry. Therefore, gathering weather data is important information for any dataset connected to energy.

The Iraqi Meteorological Network offers weather information to the general population. Several important factors are included in the weather recordings from the weather network. In this dataset, we have retrieved dew point, apparent temperature, relative humidity, air temperature, wind speed, and wind direction.

3.3 Education Calendar

Working days and weekends are only depicted on the university's academic calendar during the summer. We programmatically encode all of this data into a code. The largest energy use time is said to be at the beginning of the semester and at the beginning of summer. The institution uses power to provide cooling for the buildings and student housing since a high number of students are required to attend during this time. Applications involving energy forecasting would benefit from this knowledge.

Table 2,
The Dataset Characteristics.

	Title	Discretion
Buildings Types	1) Instructional Facilities: 2) Library Buildings: 3) Administrative buildings: 4) Residential buildings: 5) Mixed-Use Buildings: 6) Sports Facilities: 7) Other structures	1) Include lectures, workshops, tutorials, labs, and other instructional activities; 2) Include libraries as well as expansive computer laboratories and quiet study areas for students; 3) Including the HR, IT, and finance departments. 4) Including the house of the students; 5) combine administrative, research, and instructional activities; they include lecture halls, offices, and study areas; 6) Including sporting and leisure activities; 7) Include facilities that do not fit into any of the categories mentioned above.
Weather Information	Baghdad weather information	The Iraqi Meteorological Network offers weather information to the general population. Several important factors are included in the weather recordings from the weather network such as dew point, apparent temperature, relative humidity, air temperature, wind speed, and wind direction
Education Calendar	Working time	The targeted building's working hours are from 8 A.M. to 3 P.M., with peak occupancy typically observed between 9 A.M. and 1:30 P.M. Our study focuses on the summer months of 2021 (March, April, May, and June), characterized by high occupancy and increased energy consumption.

4. Data Discourse

An existing data set located in Kaggle was used, which contains a huge amount of information that I need it is UNICON Dataset, Harsha Moraliyage, et al. [39]. This reference describes all data and information contained in the data set, it contains a detailed description of the characteristics of the data collected. After that, I used this dataset to train the model to get high-efficiency results. Then, my dataset was created according to my needs, which contains all the information to use in the test to achieve the specific goal, which is to measure the amount of energy consumed in the building.

Comma-separated values (CSV) files are used to hold the energy data that has been gathered from every electrical board and building Meta. A full explanation of the properties of the data obtained is provided in Table 3. Below is an explanation of the CSV file used:

• Test.csv - This file includes metadata of all we need, such as floor space, energy consumption information for National Meter Identifiers (NMIs), and weather data which contains weather information gathered from each of the weather stations.

Table 3
A summary of the used dataset.

Class	Data field	Description	Unit	Data output
Test.csv	ID	Unique identifier	-	**Energy consumption.csv**
	Campus_ID	Identification for the foreign key referencing the campus ID	-	
	Room_area	Building's room space	m^2	
	Apparent_temperature	Temperature of the specified location	°C	
	Relative_humidity	Relative humidity	-	
	Air_temperature	The outside weather temperature	°C	
	Wind_speed	Wind speed	m/s	
	Wind_direction	Wind direction angle	°	
	Dew_point_temperature	Dew point temperature	°C	
	Hour	Hour of day for determining consumption	hr	

5. Methodology

The UNICON dataset will be used as applied by [39] to train the function because it contains a huge number of data points that give a good training result when our dataset is used to predict. After that, we start preprocessing. Preprocessing refers to the set of techniques and steps used to prepare and transform raw data into a format suitable for further analysis and modeling. The main objectives of preprocessing are to clean, transform, and standardize the data to improve the quality and reliability of the subsequent analysis. Preprocessing techniques may vary depending on the specific characteristics of the data and the requirements of the analysis or modeling task. In this work, the data was split into two parts, with 80 % training and 20 % testing the data. The area of the building and the weather changes in Iraq affect the energy consumption of the targeted building. After that, the feature scaling technique will be used in data preprocessing by standardizing the numerical features of a dataset. It ensures that all features have a similar scale or range, which can improve the performance and stability of machine learning algorithms. The standardization scaling method transforms the features to have a zero mean and unit variance. It subtracts the mean value of each feature and divides it by the standard deviation.

Standardization preserves the shape of the distribution and is less affected by outliers. The selection of the feature scaling technique depends on the specific characteristics of the dataset and the requirements of the machine learning algorithm. The standardization method was used to scale the input values and get a specific scalar which is saved to use later when we enter our data set. Figure 2 illustrates the workflow briefly.

The energy consumption in the lecture building of the AL-Khwarizmi Engineering College will be predicted using the Scikit-learn library, which contains supervised machine learning [34]. Regression models will be chosen because this study deals with prediction. The neural network model in Sklearn will be used. In Scikit-learn, the neural network models for supervised learning tasks are implemented through the \`MLPRegressor\` class, which stands for Multi-Layer Perceptron. These models are based on artificial neural networks and will be used for solving regression problems. In this work, the \`MLPRegressor\` class will be used for regression tasks, where the goal is to predict a continuous numerical value based on input features. Multiple linear regression (MLR) will be employed in the prediction of building energy consumption as used by [35] and may be applied early in the design process to enhance building performance. In Scikit-learn, neural network models are composed of several layers of sequentially arranged, networked nodes known as neurons. After receiving input signals, each neuron produces an output by applying an activation function and then sends the result to the layer below. Hidden layers, which lie between the input and output layers, aid in the discovery of intricate linkages and patterns in the data. After you've established the model, you can train it using the 'fit' method by feeding it the goal values for the input data. During the training phase, the model updates its weights and biases to decrease the gap between predicted and actual output values. Gradient descent and backpropagation are two typical techniques for this optimization process. After training the model, use the 'predict' technique to make predictions on fresh, previously unknown data.

After propagating the input data through its layers, the neural network model produces the predicted results. This might be beneficial for a variety of purposes, including estimating power use, saving electricity and energy, identifying appliances, projecting costs, and other energy-related operations [36-38]. All of the model's parameters will be displayed in Table 3.

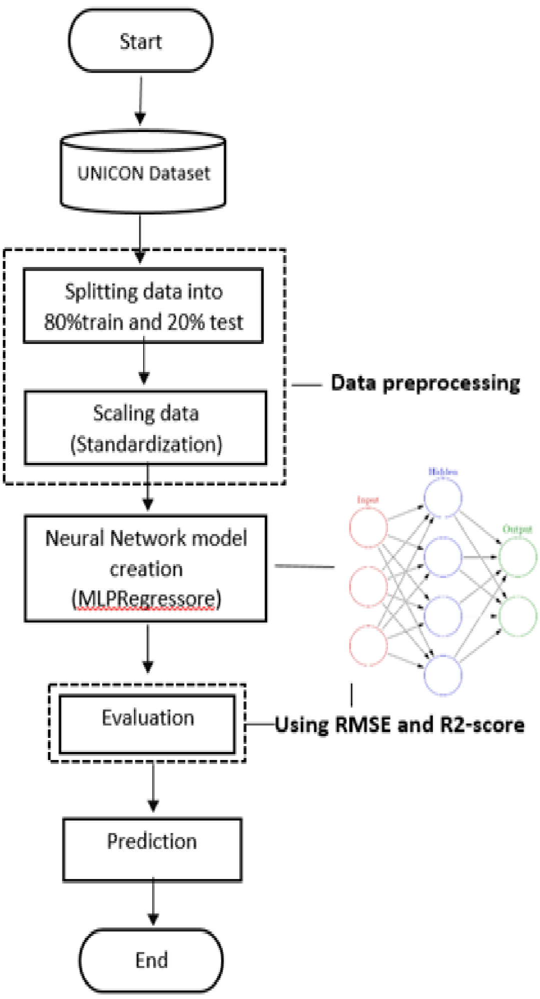

Fig. 2. Flowchart of Key Workflow.

Metrics are numerical measurements used in machine learning and data analysis to evaluate the quality or effectiveness of a model or algorithm. They provide unbiased benchmarks, such as mean_squared_error and r2_score, for assessing a model's performance on a given task or dataset [40].

R-squared determines the proportion of the dependent variable's variance that can be explained by the independent variables in the model. The scale ranges from 0 to 1, with higher values indicating a better match. In other words, R-squared reflects the adequacy of a regression model in explaining the variance and predicting the building's energy use.

To predict energy consumption, we use our dataset to generate anticipated energy values using a trained algorithm to achieve high efficiency. By using weight values we can predict energy usage accurately and make well-informed choices regarding our energy consumption.

Table 3.
List of parameters used in the algorithm.

Name of Parameter	Used Parameter
hidden_layer_sizes	60,170,150,80,80,160
solver	sgd
learning_rate	**adaptive**
Activation function	**relu**
Alpha value	0.0001

6. Results

The system was taught using the UNICON dataset. It displayed predictive performance. Through the utilization of a network framework, the experimental data-driven technique demonstrated that the neural network (NN) model exhibited precision, in forecasting energy usage. With a root mean square error (RMSE) of 0.16 and an impressive R2 score of 0.97, the educated model delivers results. The R2 score shows how well the model captures patterns and trends accurately. It suggests that the model can explain 97% of the variations, in energy consumption data. Furthermore, Figure 3 illustrates the model's precision with a small root mean square error (RMSE) of 0.16 indicating that the typical gap, between predicted and actual energy consumption figures is minimal.

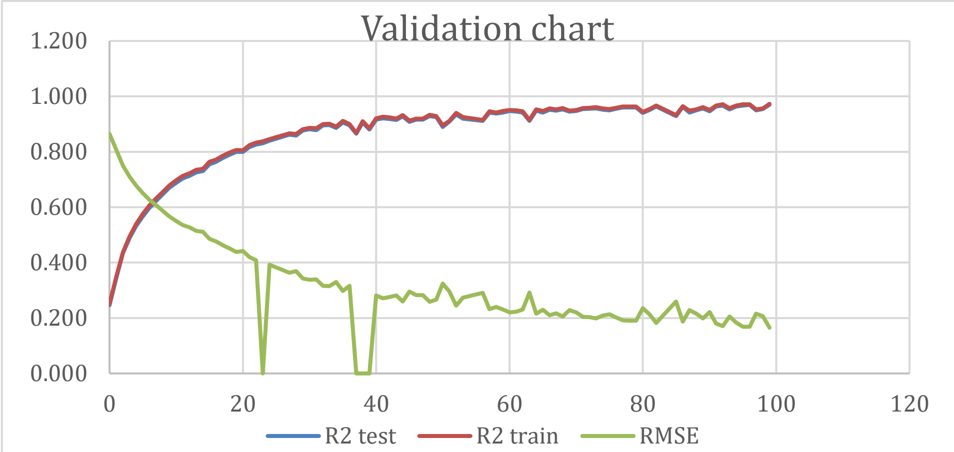

Fig .3. Validation chart which shows RMSE value (root mean square error) and R2-score value (proportion of variance)

The system is highly reliable and ideal, for predicting the energy usage of the building as, per the evaluation criteria. Its outstanding performance enables stakeholders and decision-makers to accurately foresee and gauge energy consumption levels. With these foresight skills, they can make choices regarding how resources are allocated, manage energy, and optimize efficiency leading to improved energy efficiency and cost reductions, at the facility, in question.

Assessing the precision of a model a chart illustrating prediction errors could come in handy. The model generates projected values that are then contrasted with the outcomes in the dataset. In Figure 4 you can see a representation of the differences, between the predicted and observed values.

The prediction error plot serves as a tool to identify heteroscedasticity or noise, in data. Noise refers to fluctuations or inconsistencies in data while heteroscedasticity indicates the varying levels of variability in the target variable, across ranges. By comparing expected and actual data discrepancies one can easily spot patterns and trends.

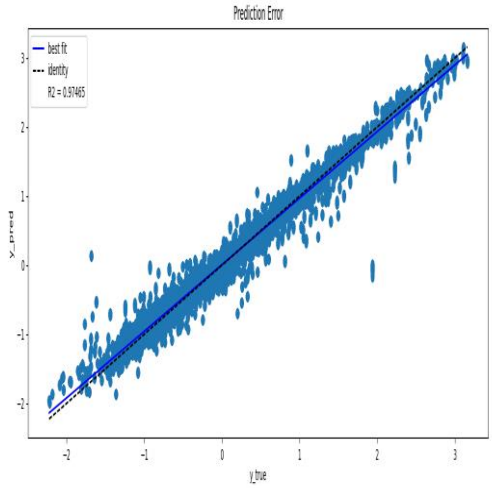

Fig.4. Prediction error plot between actual and predicted values.

A plot showing prediction errors is a tool, for assessing how much variability exists in a model. When the points on the plot are scattered randomly around zero it suggests that the model accurately captures the underlying patterns with deviation. On the other hand, if there are patterns or significant deviations, from zero it suggests that either the model did not consider all aspects of the data or that the predictions were highly unpredictable.

Studying the error map predictions gives us insights, into how our model is performing and helps identify areas where we can make enhancements. By conducting this analysis we can refine the model explore strategies and deepen our comprehension of the precision and reliability of our forecasts.

The graph displaying prediction errors illustrates the variance, between estimated values enabling us to assess the models' accuracy, variability, and inconsistency. This tool plays a role, in assessing, improving, and studying prediction models.

By utilizing our data and a forecasting model we can obtain estimates of consumption for March, May, June, and July generating updated values, for, over 120 days. This process involved applying a customized weight and factor derived from our training dataset leading to capabilities.

These forecasts offer perspectives that can be used to enhance energy efficiency in the building. This information allows us to implement various strategies and ideas to enhance energy efficiency and reduce overall consumption. By leveraging the power of data-driven predictions, we can make informed decisions and implement effective measures to achieve energy-saving goals and contribute to sustainable practices in the building.

In the figure above, the amount of predicted energy consumption during working days was illustrated, such as on day 60, which falls on Thursday, April 29, 2021, the amount of predicted energy is nearly 540 Kw/h. The amount of energy consumption increased with the increase in days due to the dry and hot weather in Iraq and the requirement to start turning on the central cooling system used in the targeted building.

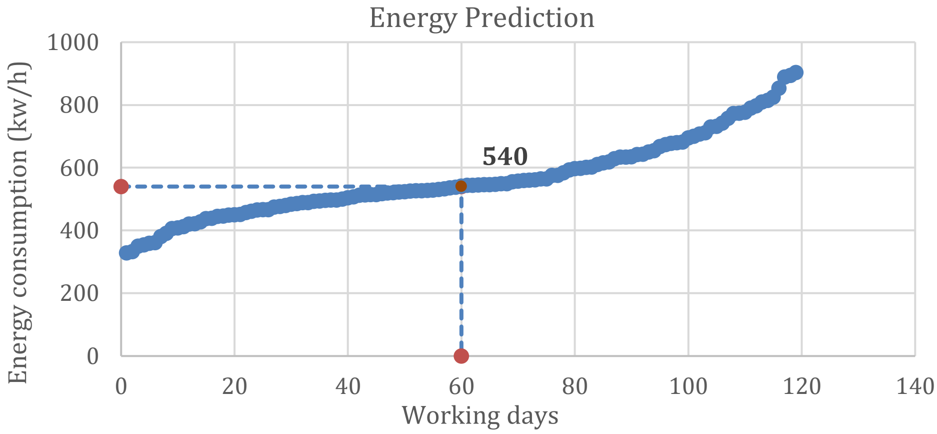

Fig. 5. Energy prediction of targeted building for 120 days of working.

7. Discussion

The Scikit-learn library was chosen for this study due to its ease of use, especially when compared to other libraries such as Keras that may require more powerful computers. Given that the objective of our paper is to predict energy consumption, regression techniques were deemed suitable and thus selected for implementation. Extensive research conducted on previous studies revealed that artificial neural networks consistently yielded the best results.

We decided to use this model because it showed accuracy during training and was efficient. Before training the data we removed duplicates and rows with values using a Data Cleaning method. We then standardized the data to ensure it was processed optimally.

This approach aids in making the data consistent and addressing its fluctuations and diverse value range. The dataset, for training was divided into two parts; testing and training and validation metrics, like root mean square error and R2 score, were employed to avoid overfitting and ensure model effectiveness. The model demonstrated a high accuracy level of 0.97 suggesting its ability to predict results that mirror real-life situations closely.

The current research highlights how artificial intelligence (AI) can improve energy efficiency and control in schools and universities. By using abilities, administrators and energy managers can effectively organize resources regulate demand, and distribute energy. Apart, from cost savings this proactive energy management strategy promotes sustainability. Reduces harm. This method contributes to a future by cutting down on energy wastage and greenhouse gas emissions.

8. Conclusion

In this study we present a dataset focusing on energy consumption in an university campus, in Iraq. The main aim of the research was to leverage AI techniques to develop models for predicting energy usage patterns and supporting energy management practices.

The use of AI algorithms and machine learning methods has significantly improved the estimation of energy consumption in buildings. By training AI models on data, weather conditions, occupancy rates, building sizes, and other relevant factors it becomes possible to forecast energy usage trends over time.

The results from training the model were highly promising with an MSE (Mean Squared Error) of 0.16 and an R2 score of 0.97 achieved. A neural network model was employed to estimate the energy consumption in buildings at the University of Baghdad by considering input variables and adjusting training, to validation ratios accordingly.

The study focuses on utilizing intelligence to predict energy usage in a university building, to reduce energy consumption. An illustration of how AI could forecast energy demand was shared by the Al Khwarizmi College of Engineering in Iraq. The models created in the research successfully captured the connections between factors affecting energy usage leading to precise forecasts and proactive decision making.

To enhance AI-powered energy prediction models, additional research and advancement are essential. It is crucial to fine-tune algorithms to incorporate real-time data streams and integrate state-of-the-art technology for improved accuracy and reliability. Collaboration among industry experts, academia, and policymakers is also vital to drive the progression and implementation of AI-driven energy management solutions, within institutions.

References

[1] D. Larcher, and J. M. Tarascon, "Science for Environment Policy Science for Environment Policy Towards the battery of the future.," science for environment policy, vol. 20, no. 7, pp. 19-29, 2018

[2] Y. Himeur, A. Alsalemi, F. Bensaali and A. Amira, , "Building power consumption datasets: Survey, taxonomy and future directions.," Energy and Buildings, vol. 227, p. 110404, 2020.

[3] X. Liang, T. Hong and . G. Shen, "Improving the accuracy of energy baseline models for commercial buildings with occupancy data.," Applied energy, vol. 179, pp. 247-260, 2016.

[4] J. Chou and D. Bui, "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design.," Energy and Buildings, vol. 82, pp. 437-446, 2014.

[5] E. P. a. Council, "Directive 2010/31/EU of the European Parlia-ment and of the Council of 19 May on the energy performance of buildings(recast).," Official Journal of the European Union L153 (0) , 2010.

[6] A. Kwok and N. Rajkovich, "Addressing climate change in comfort standards.," Building and environment, vol. 45, no. 1, pp. 18-22, 2010.

[7] R. Yao, B. Li and . K. Steemers, "Energy policy and standard for built environment in China.," Renewable Energy, vol. 30, no. 13, pp. 1973-1988, 2005.

[8] M. Leung, C. Norman, . L. Lai and T. Chow, "The use of occupancy space electrical power demand in building cooling load prediction.," Energy and Buildings, vol. 55, pp. 151-163, 2012.

[9] M. Gul and S. Patidar, "Understanding the energy consumption and occupancy of a multi-purpose academic building.," Energy and Buildings, vol. 87, pp. 155-165, 2015 .

[10] X. Zhang, K. Grolinger, M. Capretz and L. Seewald , " Forecasting residential energy consumption: Single household perspective.," In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 110-117, 17 December 2018.

[11] S. Makonin, B. Ellert, I. Bajic and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014.," Sci Data, vol. 3, no. 1, p. 160037, 2016 .

[12] T. Walter, P. Price and . M. Sohn, "Uncertainty estimation improves energy measurement and verification procedures.," Applied Energy, vol. 130, pp. 230-236, 2014 .

[13] D. Zhu, T. Hong, D. Yan and . C. Wang, "September. A detailed loads comparison of three building energy modeling programs: EnergyPlus, DeST and DOE-2.1 E. In Building Simulation," Springer Berlin Heidelberg, vol. 6, pp. 323-335, 2013 .

[14] K. Li, X. Xie, . W. Xue, X. Dai, . X. Chen and X. Yang , "A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction.," Energy and Buildings, vol. 174, pp. 323-334, 2018.

[15] M. Shen, H. Sun and . Y. Lu, "Household electricity consumption prediction under multiple behavioural intervention strategies

using support vector regression.," Energy Procedia, vol. 142, pp. 2734-2739, 2017 .

[16] R. Jain, . K. Smith, . P. Culligan and J. Taylor, "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, vol. 123, pp. 168-178, 2014 .

[17] A. Rahman, V. Srikumar and . A. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks.," Applied energy, vol. 212, pp. 372-385, 2018 .

[18] Z. Wang, X. Liu, H. Shen, Y. Wang and . H. Li, "Energy performance prediction of vapor-injection air source heat pumps in residential buildings using a neural network model.," Energy and Buildings, vol. 228, p. 110499, 2020 .

[19] S. E. M. C. S. N. P. K. R. L. B. a. L. C. O. Paudel, " A relevant data selection method for energy consumption prediction of low energy building based on support vector machine.," Energy and Buildings, vol. 138, pp. 240-256, 2017 .

[20] T. Kim and S. Cho, " Predicting residential energy consumption using CNN-LSTM neural networks.," Energy, vol. 182, pp. 72-81, 2019 .

[21] D. Bui, T. Nguyen, . T. Ngo and H. Nguyen-Xuan, "[24] Bui, D.K., Nguyen, T.N., Ngo, T.D. and Nguyen-Xuan, H., 2020. An artificial neural network (ANN) expert system enhanced with the electromagnetism-based firefly algorithm (EFA) for predicting the energy consumption in buildings.," Energy, vol. 190, p. 116370, 2020 .

[22] Z. Guo, H. Moayedi, . L. Foong and . M. Bahiraei, "Optimal modification of heating, ventilation, and air conditioning system performances in residential buildings using the integration of metaheuristic optimization and neural computing.," Energy and Building, vol. 214, p. 109866, 2020 .

[23] J. Valbé, M. Martí, . P. Casanovas, A. Jakulin, D. Mladenic and B. fortuna, "Stemming and lemmatisation: improving knowledge management through language processing techniques.," Stemming and Lemmatisation, pp. 1000-1016, 2007 .

[24] B. Chegari, . M. Tabaa, E. Simeu, F. Moutaouakkil and H. Medromi, "Multi-objective optimization of building energy performance and indoor thermal comfort by combining artificial neural networks and metaheuristic algorithms.," Energy and Building, vol. 239, p. 110839, 2021 .

[25] K. Das and . R. Behera, "A survey on machine learning: concept, algorithms and applications.," International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, no. 2, pp. 1301-1309, 2017.

[26] V. V.K., The Hundred-Page Machine Learning Book., Quebec City, Canada: Andriy Burkov, 2020 .

[27] "user guide of scikit learn," [Online]. Available: https://scikit-learn.org/0.21/_downloads/scikit-learn-docs.pdf.

[28] " user guide of tenser flow," [Online]. Available: https://www.tutorialspoint.com/tensorflow/tensorflow_tutorial.pdf.

[29] "user guide of pytorch," [Online]. Available: https://web.cs.ucdavis.edu/~yjlee/teaching/ecs289gwinter2018/Pytorch_Tutorial.pdf.

[30] "user guide of Keras," [Online]. Available: https://www.tutorialspoint.com/keras/keras_tutorial.pdf.

[31] Iraqiclimate," [Online]. Available: https://www.climatestotravel.com/climate/iraq#baghdad.

[32] Y. Geng, W. Ji, B. Lin, J. Hong and Y. Zhu, "Building energy performance diagnosis using energy bills and weather data.," Energy and Buildings, vol. 172, pp. 181-191, 2018 .

[33] J. Kočí, V. Kočí, J. Maděra and . R. Černý, "Effect of applied weather data sets in simulation of building energy demands: Comparison of design years with recent weather data.," Renewable and Sustainable Energy Reviews, vol. 100, pp. 22-32, 2019 .

[34] " scikit learn documentation," [Online]. Available: 1. Supervised learning — scikit-learn 1.2.2 documentation.

[35] S. Asadi, S. Amiri and M. Mottahedi, "On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design.," Energy and Buildings, vol. 85, pp. 246-255, 2014 .

[36] X. Liu, N. Iftikhar, H. Huo, . R. Li and P. Nielsen, "[35] Liu, X., Iftikhar, N., Huo, H., Li, R. and Nielsen, P.S., . Two approaches for synthesizing scalable residential energy consumption data.," Future Generation Computer Systems, vol. 95, pp. 586-600, 2019 .

[37] Y. Guo, Z. Tan, H. Chen, G. Li, J. Wang, R. Huang, J. Liu, and T. Ahmad, "Deep

learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving.," Applied Energy, vol. 225, pp. 732-745, 2018 .

[38] N. Ngo, " Early predicting cooling loads for energy-efficient design in office buildings by machine learning.," Energy and Buildings, vol. 182, pp. 264-273, 2019 .

[39] H. Moraliyage, . N. Mills, P. Rathnayake, D. De Silva and A. Jennings, "UNICON: An Open Dataset of Electricity, Gas and Water Consumption in a Large Multi-Campus University Setting.," in In 2022 15th International Conference on Human System Interaction(HSI), 2022, July..

[40] A. Cameron and F. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models.," Journal of econometrics, vol. 77, no. 2, pp. 329-342, 1997.

تطبيق Scikit-Learn للتعلم الآلي للتنبؤ بالطاقة المستهلكة في كلية الهندسة الخوارزمي، بغداد، العراق

ريم عمر مهدي*   وسام سعدي خضير**

*، **قسم الهندسة الميكاترونكس/ كلية الهندسة الخوارزمي/ جامعة بغداد

***البريد الالكتروني:**Omarreem07@gmail.com

****البريد الالكتروني:**wisam@kecbu.uobaghdad.edu.iq

الخلاصة

تمثل المباني 40% من استهلاك الطاقة في جميع أنحاء العالم. إن الخصائص الفيزيائية للمبنى، وفعالية أنظمة التدفئة والتبريد، وأنشطة السكان، واستدامة المبنى ليست سوى عدد قليل من العوامل العديدة التي تؤثر على استهلاك الطاقة في المبنى. من الصعب جدًا تقدير متطلبات الطاقة للهيكل. يعد تقدير الطلب على الطاقة في المبنى أمرًا حيويًا لزيادة الاستدامة وتطوير مصادر الطاقة المستدامة لخفض انبعاثات ثاني أكسيد الكربون الناتجة عن حرق الوقود الأحفوري. تم في هذه الدراسة شرح الطاقة المستخدمة في قاعة المحاضرات في جامعة بغداد (UOB) الواقعة في كلية الهندسة الخوارزمي في بغداد. تم جمع بيانات الطقس ومعلومات تشييد المباني لفترة محددة ووضعها في مجموعة بيانات محددة. وتم استخدام تلك البيانات لمعرفة قيمة استهلاك الطاقة في المبنى المذكور باستخدام الذكاء الاصطناعي وتحليل البيانات. يتم استخدام مكتبة Python تسمى Scikit-learn لتنفيذ خوارزميات التعلم الآلي. على وجه الخصوص، تم استخدام خوارزمية تراجع بيرسبترون متعدد الطبقات (MLPRegressor) للتنبؤ بالاستهلاك. وتكمن أهمية هذا العمل في التنبؤ بكمية الطاقة المستهلكة. يمكن استخدام نتائج هذا العمل للتنبؤ بالطاقة التي يستهلكها أي مبنى قبل بنائه. وتبين المنهجية المستخدمة إمكانية التنبؤ بأداء الطاقة في المباني التعليمية باستخدام النتائج السابقة وتدريب النموذج عليها، وتعتمد دقة التنبؤ على كمية البيانات المتوفرة للتدريب على خطوات الذكاء الاصطناعي (AI) لإعطاء أعلى دقة. تم التحقق من التنبؤ باستخدام جذر متوسط مربع الخطأ (RMSE) ومعامل التحديد (R^2) ووصلنا إلى 0.16 و0.97 لـ RMSE وR^2، على التوالي.