ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

16th Digital Design In Architecture Symposium
16th DDAS (MSTAS) - Special Issue 2022

2022, Vol.23, pp. 60-67, DOI:10.18038/estubtda.1169816

DEMYSTIFYING MACHINE LEARNING FOR ARCHITECTURE STUDENTS

Arzu GÖNENÇ SORGUÇ [1,*], Müge KRUŞA YEMİŞCİOĞLU [2], Ozan YETKİN [3]

[1,3,] Department of Architecture, Faculty of Architecture, Middle East Technical University, Ankara, Turkey.
[1,2] METU Design Factory, Middle East Technical University, Ankara, Turkey.

ABSTRACT

With the developments in technology, mass data, new approaches/tools, and the increasing inclusion of machine learning applications, the necessity to teach these concepts and their applications have emerged in all research areas including architecture. In this context, a new course named "Machine Learning Applications in Architecture" containing lectures on data, data literacy, patterns, and various kinds of models along with a project conducted by the students was developed and started to be taught in spring'2020. Conducting a class on relatively new subjects for students was a great challenge. Yet, with a well-defined problem-based learning approach, the adaptation of students to the subject took place immediately. It is important to note that as students are equipped with information on machine learning concepts and applications with the given lectures, they were free to choose the project topics of their own which are believed to be one of the reasons for the success of the end results. As a result of this class, the project topics varied widely as coloring a given painting, predicting the era of a building, interpreting 2D drawings for 3D modeling, optimizing daylight gain, analyzing distinctive features of data in a city, and visualizing data to represent various aspects in data. The outcomes of the class are documented and analyzed to show how information in different fields such as computer science, engineering, statistics, and so on can broaden their thinking of how to attack problems in the architectural design domain. Finally, topics such as data, data literacy, pattern recognition, and intelligent models are projected to play a key role in the future of design education since it provides an interdisciplinary ground to think about problems at hand from a distinct perspective.

Keywords: Machine learning, Architecture education, Data literacy, Data-model matching

1. INTRODUCTION

Advances in technology, the abundance of data mass, tools easing the adopting new approaches, and increasing involvement of machine learning applications in architecture increase the interests of students in these areas and application studies. Ever since artificial intelligence and machine learning became a 'buzzword' for every domain, the necessity to teach these concepts and their applications have emerged for all research areas including architecture [1,2]. With the increasing involvement of computational design, data becomes the main input, and data visualizations have already turned into forms with algorithms/models developed or compiled by designers. Today, considering that form finding has started becoming an archaic term, data itself and data visualization are taking its place for design disciplines, especially for architecture.

In education, computational design and then machine learning become key topics with offered courses in various universities such as "Machine Learning for Creative Design" at Massachusetts Institute of Technology [3], "Artificial Intelligence in Architecture" at Institute for Advanced Architecture of Catalonia [4], "A.I., Machine Learning, and the Built Environment" at Harvard Graduate School of Design [5], "Data Science for Construction, Architecture and Engineering" at National University of Singapore [6], and so on. Therefore, it can be easily observed that the field of computational design started to focus more on trending subjects in computation that are machine learning, artificial intelligence, and data science.

*Corresponding Author: arzug@metu.edu.tr
Received: 01.09.2022 Published:23.12.2022

In this context, a new course named “Machine Learning Applications in Architecture” offering lectures on data, data literacy, patterns, and various kinds of models along with a project that is planned to be conducted by the students was developed and started to be taught in spring’2020. Then, in spring’2021 term, the course is offered to graduate students in the department of Architecture. With these experiences, we would like to share not only our teaching experience or outcomes but also the “to do” and “not to do” lists for future classes.

2. DATA LITERACY

At the beginning, conducting a class on relatively new subjects for students was a great challenge. Yet, with a well-defined problem-based learning approach, the adaptation of students on the subject took place immediately. In this period, terms, data literacy, and the concepts of data mining, labeling, and matching are introduced to the students. In this vein, along with the concepts of data science, the importance of domain knowledge is worth mentioning considering the specific knowledge of students of the department of architecture related to visualization, representation, and classification. Hence, it can be said that coming from an architectural background helped students to understand and implement data visualization and having knowledge on computational design aided them to comprehend mathematical concepts behind machine learning models. Based on that, it should be noted that data literacy and domain knowledge have filled a critical gap when it comes to implementing machine learning algorithms to specific problems observed and defined by the students.

Analyzing the projects completed in the scope of the course, the process of each project can be summarized with four main steps that are: formulating the problem, acquiring and processing the data, determining the model, and visualization of results. Yet, it can be observed that each project focused on different steps of the overall process. For example, in one project aiming to predict the corresponding three-dimensional (3D) model of a tree based on a given two-dimensional (2D) drawing, it can be observed that labeling the acquired data and determining the model architecture was the focus. On the other hand, in another project aiming to enhance the visual connection between interior and exterior of a building provided by visibility prediction based on 2D plan layout data, it can be said that formulating the problem and visualization of results was the focus. Therefore, exemplified by these two projects, outputs of each project completed in the scope of this course have shown diversity. For comparison, the data preparation phase of projects; 3D tree model prediction from 2D drawings (Figure 1), and the study analyzing and aiming to enhance the visual connection between interior and exterior (Figure 2) are presented below.

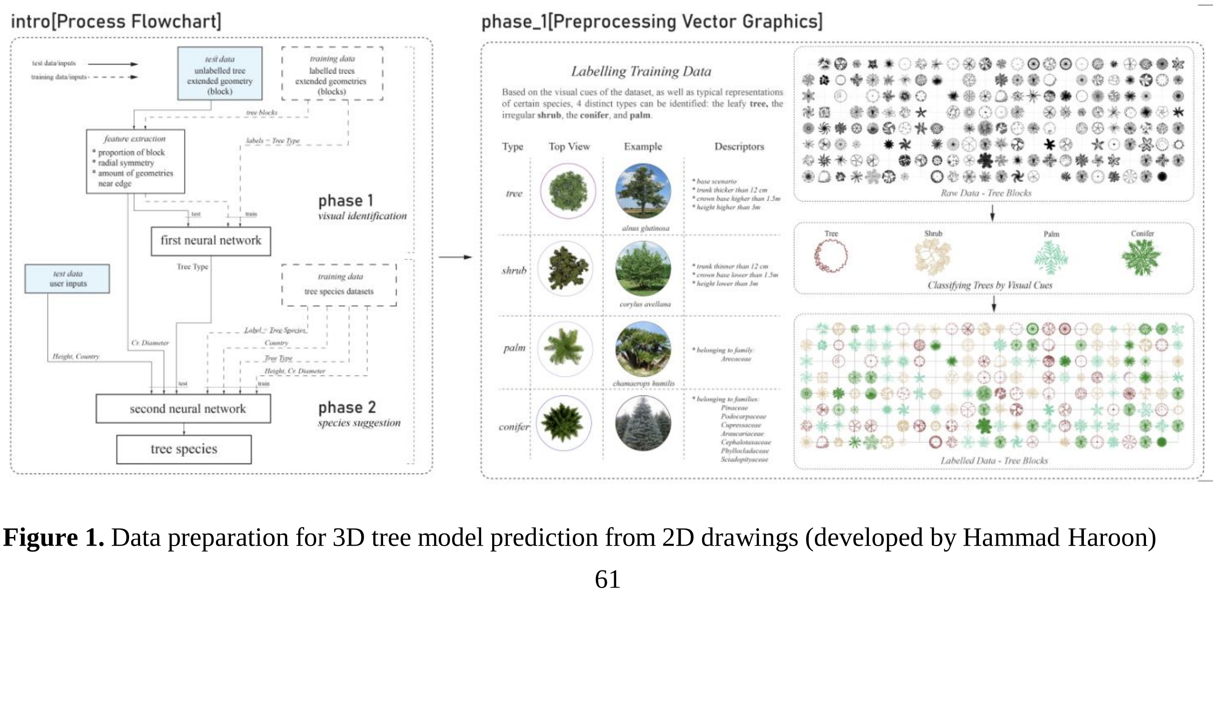

Figure 1. Data preparation for 3D tree model prediction from 2D drawings (developed by Hammad Haroon)

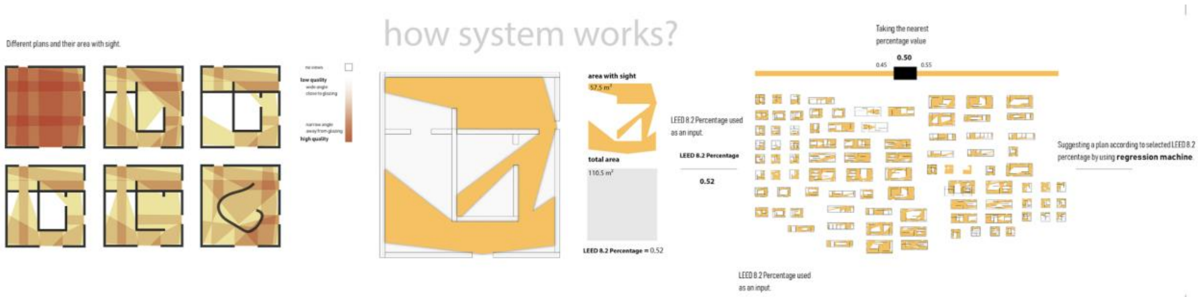

Figure 2. A case on sight of a space (developed by Nilra Ayşegül Zoraloğlu)

3. DATA AND MODEL MATCHING

Following the data preparation processes, it is seen that matching data with the design process in terms of form, function, generation process is found valuable to be explored in architecture with many examples and in-class assignments. In the two semesters that the course is offered, different approaches are embraced. For the first term, students are asked to locate, collect, or create their own datasets depending on their problem definition. Then, for the second term, New York City Database (https://opendata.cityofnewyork.us/data/) is assigned considering the wide variety of data sets included that can serve for a wide variety of problems.

As a result of this difference, students were able to adapt themselves in terms of the tools used for the project, data processing strategies, model selection, and data visualization techniques. For example, when the students are asked to find or generate their own datasets, most of the students tend to use the tools that they already know such as Grasshopper; but when the case is switched in the next term where students were assigned the New York City Database [7], students tend to learn new tools such as Python. Moreover, data processing techniques were switched from feature extraction from geometry to statistical methods in general depending on the difference between terms. Furthermore, model selection is also affected since available methods and community behind tools differ greatly. Yet, the switch from expecting datasets from students to assigning a predefined dataset created a difficulty in terms of producing visual outcomes of the models. Finally, problem definitions, datasets and used models for two terms are presented in the Table 1 below in order to provide a general picture.

Table 1. Matrix of problem definitions, data sets and selected models (developed by the authors based on student projects)

Problem	Term	Data Set	Main Strategy	Model	Main Tool
Visibility Analysis	s'2020	2D Plan Layouts and Wall Locations (Generated)	Geometry Processing	Regression	Grasshopper
Tree Suggestion	s'2020	2D Tree Drawings and Names of Species (Collected)	Geometry Processing	Classification	Grasshopper
Date Prediction	s'2020	Building Photographs and Construction Dates (Collected)	Image Processing	Classification	C# (.NET)
Data Visualization	s'2020	Climate Data and Corresponding Locations (Obtained)	Data Clustering	Regression	Grasshopper
Image Coloring	s'2020	Painting Images of Famous Painters (Collected)	Image Processing	Regression	Python
Direction Control	s'2020	Wind Directions and Intensions (Generated)	Geometry Processing	Regression	Grasshopper
Air Quality Prediction	s'2021	Air Pollution and Particles in the Air (Obtained)	Statistical Analysis	Regression	Python
Vehicle Safety Prediction	s'2021	Vehicle Collisions and Crash Reports (Obtained)	Statistical Analysis	Classification	Python
Mapping of Constructions	s'2021	Construction Jobs and Locations (Obtained)	Data Clustering	Classification	Grasshopper
Energy Score Prediction	s'2021	Energy and Water Usage (Obtained)	Statistical Analysis	Regression	Python
Incident Prediction	s'2021	Incidents and Crime Reports (Obtained)	Statistical Analysis	Regression	Python

4. RESULTS

As students are equipped with information on machine learning concepts and applications with the given lectures, and hands-on workshops, they were free to choose the project topics of their own which are believed to be one of the reasons for the success of the end results. The project topics varied widely as coloring a given painting, predicting the era of a building, interpreting 2D drawings for 3D modeling, optimizing daylight gain, analyzing unique features of data in a city, and visualizing data to represent different aspects in data. The outcomes of the class are documented and analyzed to show how information in such a different field such as computer science, engineering, statistics can broaden their thinking of how to attack problems in the architectural design domain. Finally, topics such as data, data literacy, pattern recognition, and intelligent models are projected to play a key role in the future of design education since it provides an interdisciplinary ground to think about problems at hand from a different perspective.

This result can be observed in project results in different aspects. For example, in one of the examples completed in spring'2020, the problem is defined as predicting the era of a building from its photograph (Figure 3). Since the dataset was limited compared to other well-known image processing examples like classifying objects or animals, data augmentation was implemented by rotating, scaling, and cropping the images and adding them as new samples to enlarge the limited dataset which inevitably increased the accuracy of the model. Hence, it can be said that available approaches in other domains (such as computer engineering) were successfully discovered and implemented to overcome challenges of the architectural domain (such as lack of available data).

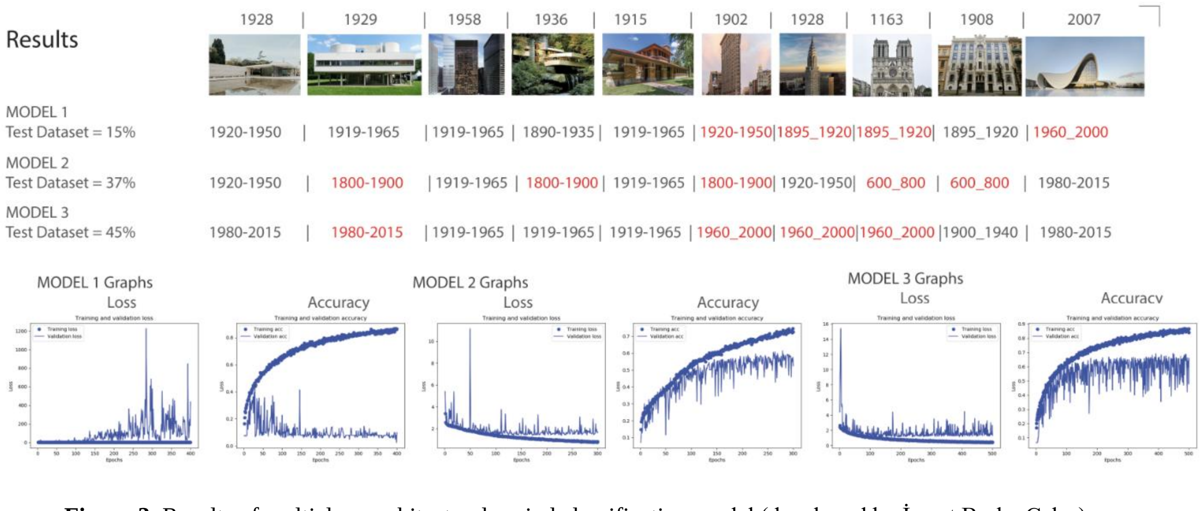

Figure 3. Results of multiclass architectural period classification model (developed by İsmet Berke Çakır)

Similarly, in another example from the same term, image processing was used in a different way to the defined problem of coloring a given black and white image (Figure 4). In this project, the dataset is curated in a way to provide different coloring strategies, unlike available approaches in other domains that also aim to color the given image. Hence, the project was able to color the given image with the color palettes of famous artists such as Vincent van Gogh, Raphael, Frida Kahlo, Edgar Degas, Pablo Picasso, and Amadeo Modigliani. Therefore, it can be said that the project successfully transformed the well-known problem with a touch of domain knowledge and proposed a novel approach.

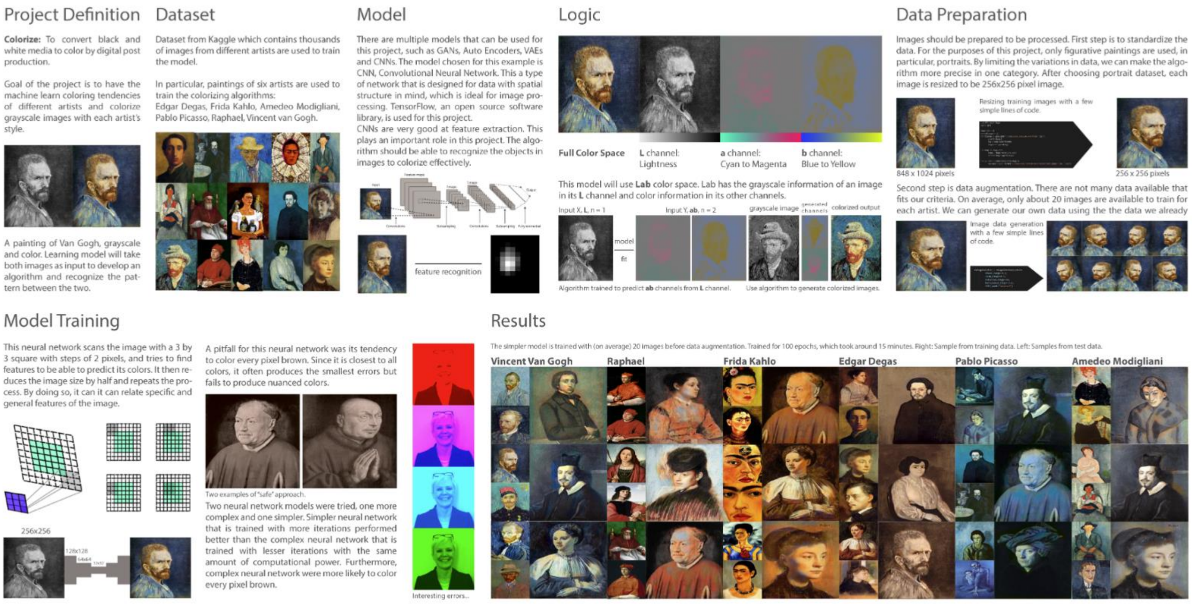

Figure 4. Colorizing with painting (developed by Ege Doğan)

Moreover, when the works developed in spring'2021 are analyzed, it is possible to see that the focus was more on the engineering side since the source of the data is already provided. For example, in the example of predicting energy star score from energy and water usage data, pairwise feature visualizations and a correlation matrix were used to demonstrate and select which features of the given data affect the energy score more. In another example from the same term, data were analyzed by means of a tree structure so that features of data were put in a hierarchical way from general to specific and construction works in NYC are visualized accordingly.

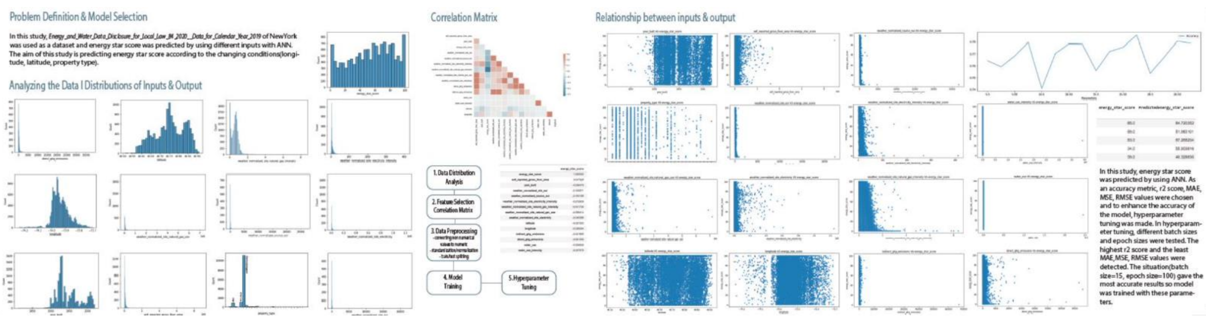

Figure 5. Results of energy star score prediction model (developed by İlkim Canlı)

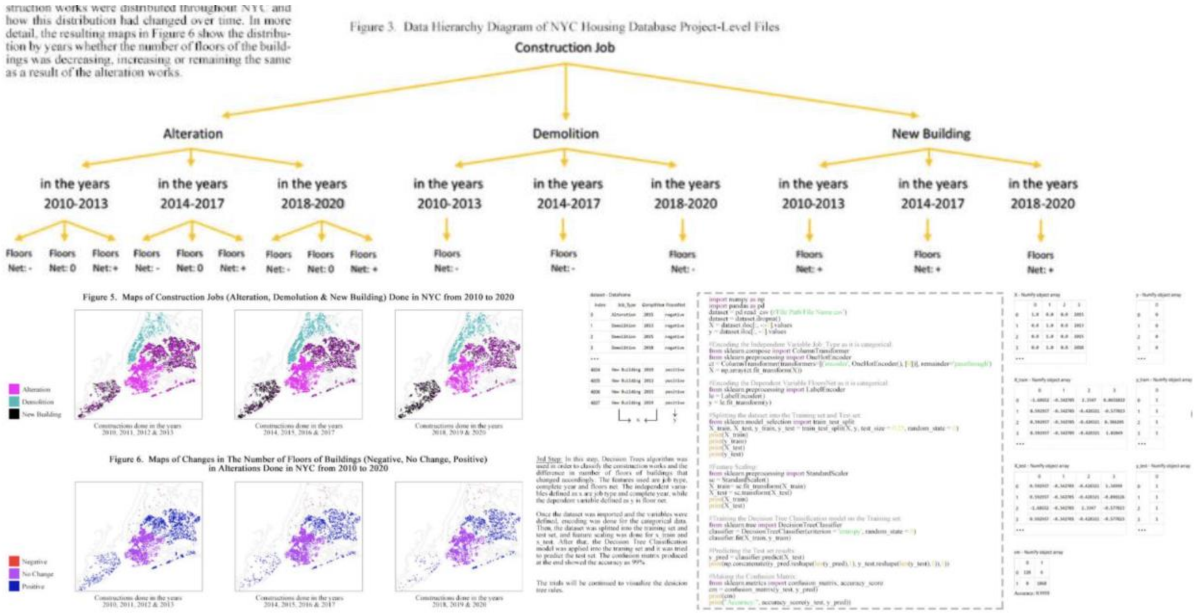

Figure 6. Model for understanding and mapping construction works in NYC (developed by Hatice Hilal Topuz)

5. DISCUSSIONS

In the first semester, students are asked to find or provide their own datasets, then in the second term, New York datasets are provided to be worked on. In each semester, the synthetic nature of the datasets and the role of augmentation are discussed. In both cases, they were free to choose their problem according to their curiosity. It is believed that defining the question and finding their curiosity which is what they want to learn is crucial instead of what they can do with the obtained data and the models. Yet, while constructing their questions, considering these exercises are their first encounter with ML models, students are directed toward quantifiable issues instead of topics like perception, creativity, and so on.

While structuring datasets provides students more control over their data and choosing a model, provided datasets lead them to find cross relations and merging different datasets as well as exploring different models. For both cases, it is seen that, when the obtained/collected/generated data are too complex, it becomes hard for students, who are new to the subject, to find the correlations and thus, understand and assess the models and results. Although there are many sources providing ML models, it is believed that fundamental statistical knowledge including manual calculations of probability, ANN etc. is crucial.

In each semester, students are offered tutorials on Python and Lunchbox add-on of Rhino-Grasshopper [8]. Based on the background and existing skills of students, they either developed their own model, or use Python models provided in databases like GitHub [9]. Although the models that are developed by students were less complicated than others, it is seen that at the end, they had more control and knowledge about how it works, and their results.

6. CONCLUSION

In this paper, the outcomes of the course "Machine Learning in Architecture" class are presented with the conduction of the terms and the lessons learned throughout the two semesters. The first point to mention is the importance of data literacy and adapting to the changing role of data visualization in computational design. Then, the introduction of machine learning's fundamental topics like probability, statistics, feature extraction, labeling is found vital for the conduction. As the most critical part, the importance of the domain knowledge, which is architectural knowledge in this case, has played a crucial role throughout the term. The specific knowledge of any domain is important for any ML application in that domain considering the fine tasks like data preparation, labeling and model selection.

Above all course objectives, at the end of the semester, students are expected to accept that ML is merely a tool for understanding, modeling, and predicting, and therefore no more mystical than any computational tool. Considering the literacy between data and information is being lost, it is aimed to re-establish this relation again for students. Hence, it is very important for students to understand the following: ML is neither a magical tool that only takes data and handles the rest by itself nor a very complicated one that needs to be analyzed and structured to its every bit.

ACKNOWLEDGEMENTS

We would like to thank and acknowledge our students who experienced this journey with us in Machine Learning Applications in Architecture course: Ege Doğan, Hammad Haroon, Handan Akyürek, İsmet Berke Çakır, Nilra Ayşegül Zoraloğlu, Seren Ertan, Burcu Atakul, Şevval Simruy Baygül, İlkim Canlı, Aslı Zeynep Doğan, Mehmet Burak Kaya, Dilara Keçeci, Atike Yağmur Köseoğlu, Mehmet Oğuz Nas, Bekir Enes Özel, Zeynep Şan, and Hatice Hilal Topuz.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] Khean N, Fabbri A & Haeusler MH. Learning Machine Learning as an Architect, How to? Proceedings of the 36th eCAADe Conference - Volume 1, Lodz University of Technology, Lodz, Poland, 19-21 September 2018; 95-102

[2] Carpo, M. The Second Digital Turn - Design Beyond Intelligence, MIT Press, Cambridge, 2017

[3] Mueller C & Danhaive R. Special Subject: Building Technology - Machine Learning for Creative Design | MIT Architecture. Retrieved April 22, 2022, from https://architecture.mit.edu/subject/spring-2019-4s42, 2019

[4] León, D. A. Artificial Intelligence in Architecture: MaCAD Students Learning How to Radically Innovate the AEC Sector. IAAC. Retrieved April 22, 2022, from https://iaac.net/artificial-intelligence-macad; 2021

[5] del Castillo y Lopez, JLG. Machine Learning, and the Built Environment. Retrieved April 22, 2022, from https://pll.harvard.edu/course/ai-machine-learning-and-built-environment, 2022

[6] Miller C. Data Science for Construction, Architecture and Engineering. edX. Retrieved April 22, 2022, from https://www.edx.org/course/Data-Science-for-Construction-Architecture-and-Engineering, 2022

[7] City of New York. NYC Open Data. Retrieved April 22, 2022, from https://opendata.cityofnewyork.us/data

[8] Lunchbox. Proving Ground Apps. https://apps.provingground.io/lunchbox. Published June 2, 2021. Accessed May 05, 2022,

[9] GitHub. https://github.com. Accessed May 05, 2022.