



Selected Application Tools for Creating Models in the Matlab Environment

Stella Hrehova^(✉)  and Jozef Husár 

Faculty of Manufacturing Technologies with a Seat in Prešov, Department of Industrial Engineering and Informatics, The Technical University of Košice, Bayerova 1, 080 01 Prešov, Slovak Republic

{stella.hrehova, jozef.husar}@tuke.sk

Abstract. The issue of analysis, data evaluation and prediction has resonated in every area for several years. With the introduction of the Industry 4.0 philosophy, data is becoming the centre of attention. The constant development of new measurement technologies and applications, as well as the possibilities of data storage and sharing, also contribute to the constant increase in the amount of data. Application developers who are already firmly established in the company are trying to develop new, user-friendly tools. These application are focused on advanced data evaluation and modelling capabilities. Matlab is one such application. Its advantage is the constant development of individual advanced tools - toolboxes specialized in various areas. The presented paper will describe selected data visualization options. We will also focus on the description of selected tools using machine learning techniques to find and design the best model for data prediction. The data is obtained through the SCADA user interface and relates to the issue of building heating. In the presented paper, we will point out the advanced data display and compare models created using the Regression Learner and Neural Net Fitting tools.

Keywords: Data · Visualisation · Neural network · Matlab · Heating process

1 Introduction

Data and methods of their acquisition, processing and evaluation have long been the subject of intense interest. By implementing current information and communication technologies in the spirit of Industry 4.0 principles in several areas, we gain much greater opportunities to obtain the required data [1, 2]. Thanks to increasingly advanced data acquisition technologies, we are able to capture large amounts of data. It is therefore understandable that new, advanced methodologies, procedures and applications are emerging in many areas that focus on data management. Undoubtedly, computer application programs have an important place in this area of interest, which makes it much easier for users to work with data management. During the current development in the field of data processing, the individual stages that are necessary for the interpretation of the data we obtained were correct. The following figure represents one of the many schemes of the individual stages of data processing [3] (Fig. 1).

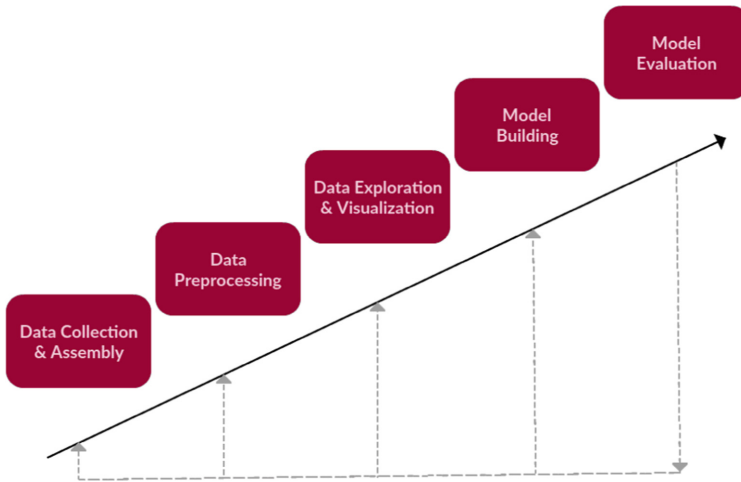


Fig. 1. Simple textual data task framework

- a) Data collection & Assambley - this part is already largely ensured by the application of the Industry 4.0 philosophy in practice. Various measuring units, sensors and other elements of advanced measuring technology are introduced, which are able to collect and store data.
- b) Data pre-processing - data pre-processing is generally required to correct for variance, baseline changes, peak shifts, noise, missing values, duplications, and several other features so that the “true” relevant basic structure or property of the data can be highlighted. The purpose of these adjustments is to prepare the source data so that they can be interpreted correctly [4].
- c) Data Exploration & Visualisation - data visualization is an efficient process for presenting data and information graphically or using pictures. It emerges as a powerful and widely applicable and acceptable tool for interpreting and analyzing large and complex data [5]. Visualisation is useful in cleaning data, examining data structure, detecting outliers and unusual groups, identifying trends and clusters, identifying local patterns, evaluating modeling outputs, and presenting results [6].
- d) Model building – the choice of a suitable model depends on the requirements. Currently, there are many applications that are able to create a model. However, we need to know the goal, what information we want to obtain from the source data, what method and tools we will use for their analysis and other information [7, 8]. Model creation is thus a complex activity where close cooperation with experts in several areas is required.
- e) Model Evaluation – many publications describe several options for evaluating the proposed model. However, we should remember that it is also important to examine the quality of the proposed model [9].

In the presented paper we will focus on the areas of visualization and model creation in the Matlab software application. The presented data were obtained via a user interface created through the system Supervisory Control and Data Acquisition (SCADA). It is

basically industrial control system, witch processing data in real time. It helps to control and monitor the operations in the many industrial areas [10]. The source data were obtained through user interface, which is displayed on the next figure (Fig. 2).



Fig. 2. SCADA user interface

The user is able to export data in different time periods through this interface. For the submitted paper, we used data from the year 2019, in the scheme once an hour for each day of the month (see Fig. 3).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Datum a čas	K1TSPALIN	K1TVRAT	K1TVYSTU	K1VYKON	K2TSPALIN	K2TVRAT	K2TVYSTU	K2VYKON	K3TSPALIN	K3TVRAT	K3TVYSTU
2	1.1.2019 0:00	41	-3276,8	49,5	49	21	-3276,8	39	0	21	-3276,8	42
3	1.1.2019 1:00	46	-3276,8	55,5	64	21	-3276,8	38	0	21	-3276,8	40,5
4	1.1.2019 2:00	44	-3276,8	55	64	36	-3276,8	51	99	22	-3276,8	39,5
5	1.1.2019 3:00	43	-3276,8	52,5	48	24	-3276,8	44	0	22	-3276,8	38,5
6	1.1.2019 4:00	35	-3276,8	66	0	47	-3276,8	45,5	18	39	-3276,8	64,5
7	1.1.2019 5:00	49	-3276,8	57	64	48	-3276,8	57	64	48	-3276,8	47
8	1.1.2019 6:00	55	-3276,8	66	64	45	-3276,8	64	64	32	-3276,8	58,5
9	1.1.2019 7:00	50	-3276,8	64,5	0	50	-3276,8	59,5	39	49	-3276,8	56
10	1.1.2019 8:00	53	-3276,8	62,5	64	50	-3276,8	59,5	64	29	-3276,8	51

Fig. 3. Part of source data

As part of the presentation of the Matlab application, only certain variables were selected from the given data, which have a high informative value - gas consumption per day SPZPLYNDEN, electricity consumption per day SPELDEN, outdoor temperature TVONKU and required heating temperatures for the given circuits UKTVODZ. The table is extracted in.csv format, which is suitable for Matlab application. However, be aware that the decimal separator must be a dot so that Matlab does not have problems identifying data. To import data to Matlab environment we used the Import option.

2 Chosen Methods and Matlab Tools

2.1 Machine and Deep Learning

Advanced machine learning analysis is becoming increasingly popular as it can be applied in almost any field. Machine learning is a sub-area of artificial intelligence (AI) in which computers derive new information or decisions by learning data using algorithms and programs [11]. The goal of machine learning is to model machine learning algorithms based on input data in a defined solution space. The following figure shows an overview of the learning types for machine learning [12] (Fig. 4).

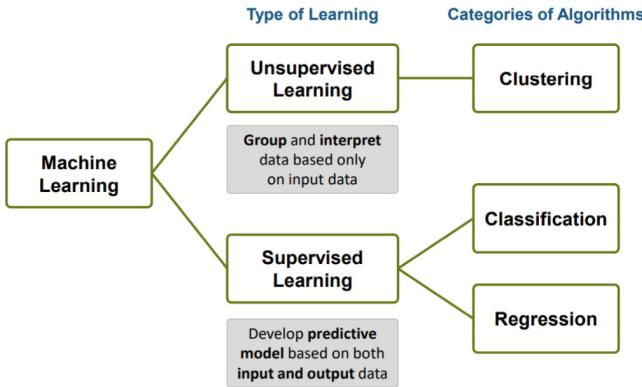


Fig. 4. Overview – Machine learning

In this paper, we have focused on the Supervised Learning section, in which input and output data are marked for classification so as to provide a basis for future data processing learning [13].

Deep learning - deep learning is a subset of machine learning in artificial intelligence that is able to learn from data that is not marked or unstructured [14]. It is machine learning technique that learns features and tasks directly from data. The term “deep” is related to the depth of neural network architecture, which is used when-ever the network contains at least two hidden layers [15].

The next table compares these techniques in some areas [16] (Table 1).

Table 1. Comparison

	Machine learning	Deep learning
Training dataset	Small	Large
Choose your own features	Yes	No
# of classifiers available	Many	Few
Training time	Short	Long

2.2 Matlab and Description of Chosen Tools

The Matlab system is a tool both for comfortable interactive work and for the development of a wide range of application programs. Its possibilities are constantly expanding into several areas by developing new extensions - toolboxes. These extensions allow users who do not need to be IT experts to use a wide range of application, software and analytics tools. In the form of guides, they make it possible to intuitively create process models and thus obtain the necessary information. The data flow in the Matlab environment is shown in the following figure [17] (Fig. 5).

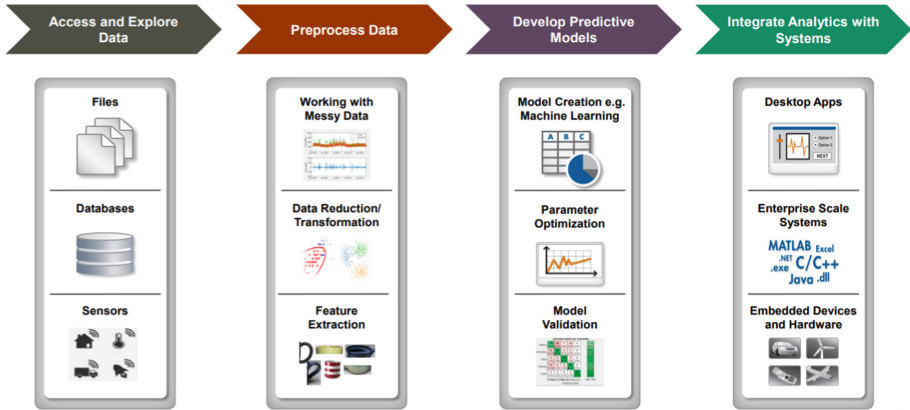


Fig. 5. Data analytics workflow in Matlab

The descriptions of used tools will present the real data that were presented in the introductory chapter.

1. Visualisation – in the field of data visualization, Matlab offers a really wide range of possibilities. It offers suitable graphs when marking the required variables. The representation of the graphical form can be changed both programmatically (using commands) and through a graphical interface (Property Editor), where the user selects the individual properties of the graphical display. When plotting a simple dependence, it also offers the possibility of displaying basic statistical characteristics in a clear table.
2. Regression Learner basically allows us to create regression models interactively without writing code and measuring the accuracy and performance of the models. It allows fast experimentation with multiple regression algorithms and quickly compares the performance of different regression models and functions [18].
3. Neural Net Fitting – it is application which allows create, visualize, and train a two-layer feed-forward network to solve data fitting problems. It is possible to choose from three different algorithms for updating weights. They are Levenberg–Marquardt, Bayesian regularization, and scaled conjugate gradient. Levenberg–Marquardt backpropagation algorithm is one of the most popular algorithms in the MATLAB toolbox [19]. It is one of the fast algorithms which has reasonable performance

and it is used more than the other ones. This algorithm has better regression (curve fitting).

2.3 Used Methods to Evaluate of Created Models

Toolbox Machine and deep learning offers several methods for creating models. The MSE characteristics and the correlation coefficient R are mainly used to assess the suitability of the resulting models [20].

- MSE is used to determine how close the regression line is to the measured data. An MSE value close to 0 means that the model complies with the data.

$$MSE = (1/\text{Number of sample}) \times \sum \text{Square errors} \quad (1)$$

- R is a statistical measurement of the relationship between variables and their interconnection.

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (2)$$

where R is the correlation coefficient, x_i values of the x -variable in a sample, \bar{x} mean the values of the x -variables, y_i values of the y -variable in a sample, \bar{y} mean of the values of the y -variables.

3 Model Design

When presenting the described tools, we assume that the data is pre-processed and cleaned up. For using chosen applications within Machine a Deep Learning Toolbox we work with average values of gained data.

3.1 Visualisation

During visualization, we are interested in the course of the monitored parameters, especially if they are dependent on the same variable. From the point of view of our data, these are, for example, the required values of the water temperature in the individual circuits of the heating system. Here are effective tools *parallelplot*, respectively *stackedplot*, when you just need to mark the data imported in the table or matrix (Fig. 6).

Similarly, you can easily use the *stackedplot* option, which allows display individual columns of imported data as separate rows in the graphics window, which are also interconnected. In this case, outdoor temperature, energy and gas consumption for the selected date (Fig. 7).

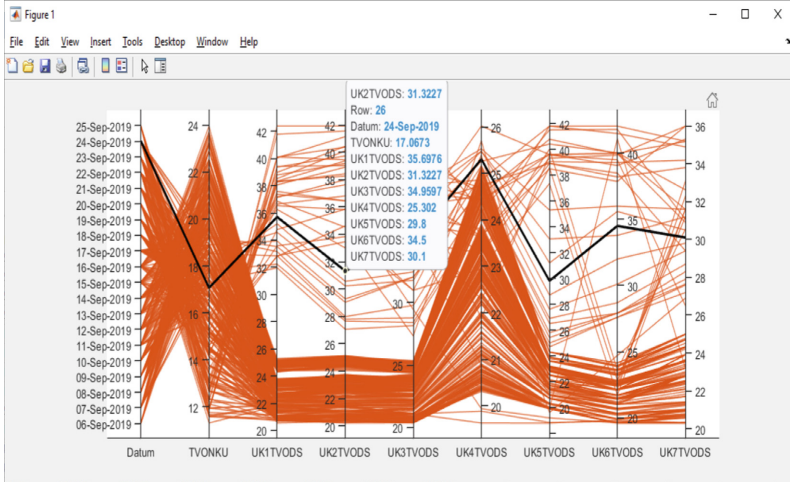


Fig. 6. Representation of chosen data using *parallelplot*

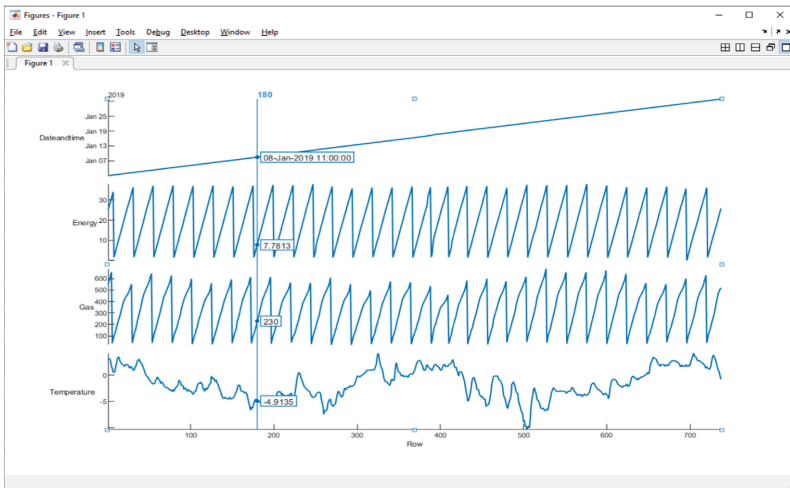


Fig. 7. Representation of chosen data using *stackedplot*

3.2 Model Design Using Machine and Deep Learning

Regression Learner

The average values of the outdoor temperature (predictor) and the corresponding average value of electricity consumption (response) were used to present the creation of the model using Regression Learner. After setting the required parameters, training was started using all options. We used this option so that the system itself found the most optimal variant (see Fig. 8).

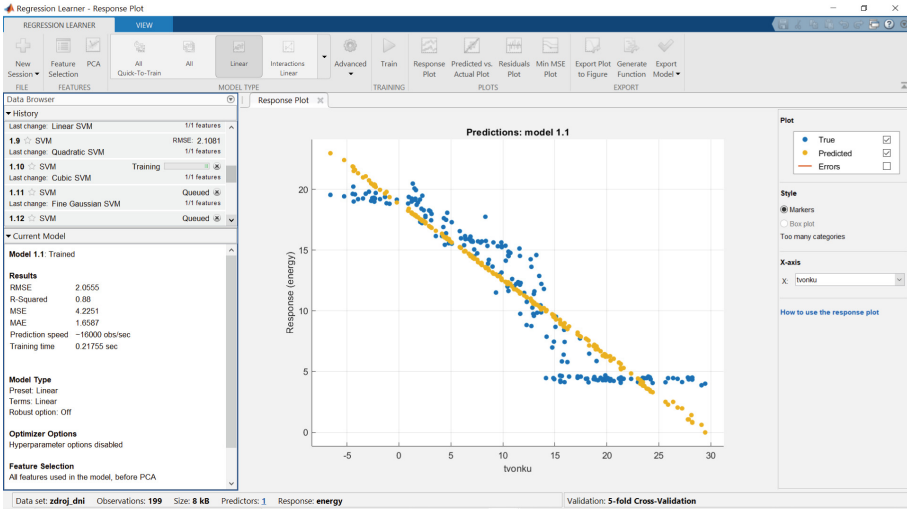


Fig. 8. Environment of Regression Learner for selected variables

From the picture it is possible to observe the course when the data are gradually trained and the optimal variant is sought. At the end of the training, the best option is automatically marked. The MSE and R values show how the model is “good”. The resulting dependence is shown in the following figure (Fig. 9).

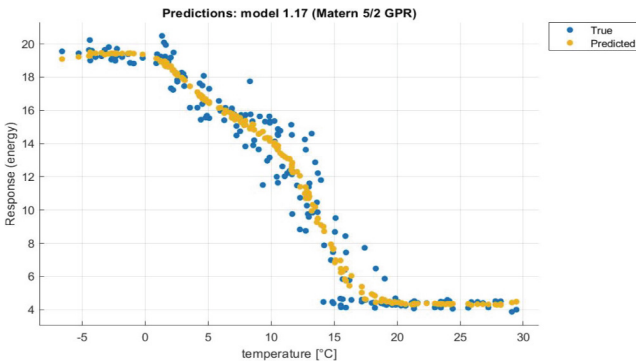


Fig. 9. The best model from regression

Neural Net Fitting

To use this tool, we will generate the variables needed to train the model x_{train} - there will be an average outdoor temperature and y_{train} - will be an average energy consumption. This tool is created in the form of a guide, where we proceed step by step and use built-in algorithms to design our model. We have a choice of three training algorithms, we used the possibility of the Levenberg-Marguard algorithm, the principle of which was described, for example, in the literature [21]. If the resulting values of

MSE and correlation coefficient do not correspond to our ideas, respectively they are unacceptable, we can retrain the training data set. We can repeat this process several times. Another option is to increase the number of inner layers or increase the number of training data. The *plotregression* output is displayed on the Fig. 10.

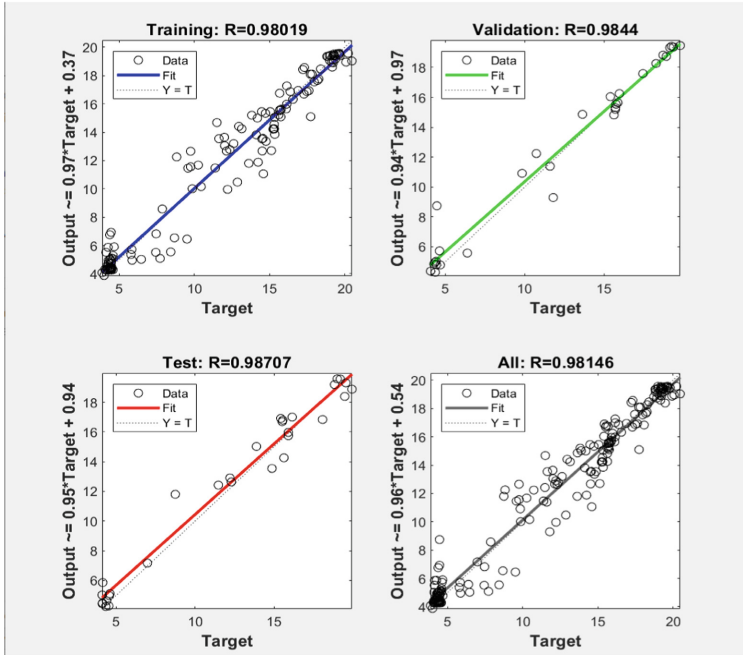


Fig. 10. The output from interface - Regression

Both described tools have the ability to generate functions that are then used for new data. The Neural Net Fitting tool also has the option of creating a schema in the Simulink environment (Fig. 11).

4 Result

To select the best model, we compared the individual qualitative indicators (Table 2).

The model created by Neural Net Fitting shows slightly better values. In order to compare the created individual models, the real values from 2020 were exported via the user interface. The following figure presents a graphical comparison of the predicted data obtained from both models with the real measured values (Fig. 12).

It can be seen from the figure that both models copy the real data quite well except for the area between months 4 and 5. This situation could have arisen due to various influences that will need to be defined and incorporated when implementing the models.

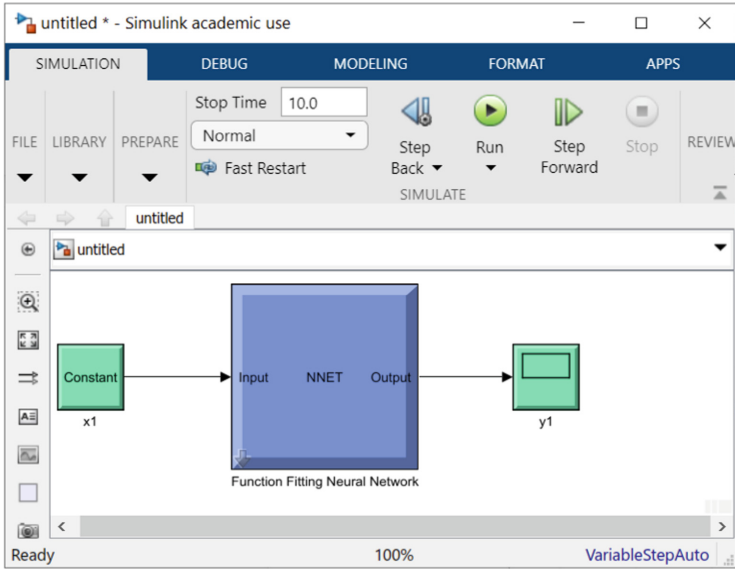


Fig. 11. Scheme in Simulink environment

Table 2. Model comparison

Coefficient	Regression learner	Neural net fitting
MSE	1,4196	1,3394
R	0,96	0,980,192

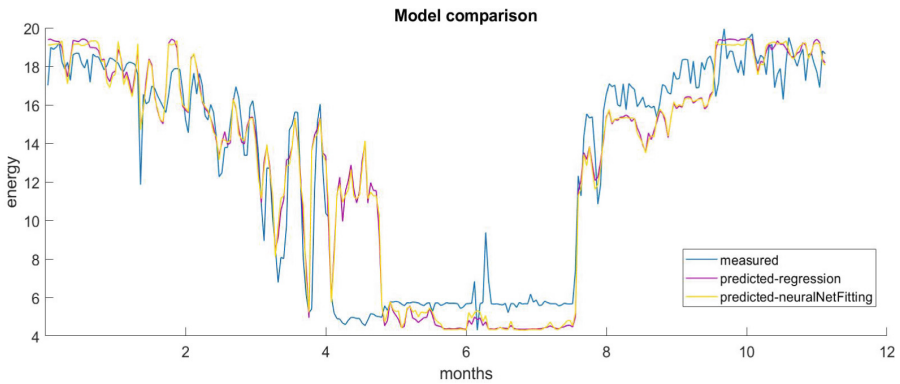


Fig. 12. Model comparison

5 Conclusion

Data can be a source of a lot of useful information if we can process it properly. They do not only tell about the state of the past, but provide opportunities for forecasting in the future. This potential only allows them to rapidly develop application environments that focus on data management. However, compared to others, MATLAB is easier and faster for many users to apply modern tools for learning. In this paper, we pointed out the use of the Machine and Deep Learning toolbox in creating two prediction models. These models are based on the use of historical data to predict the consumption of electricity when heating a selected building. The quality of the models was compared using the MSE and R indicators, where there are only small deviations. The next goal will be to create a user interface through Matlab tools so that the results are more accessible to a wider range of users.

Acknowledgements. This work was supported by the Slovak Research and Development Agency under the contracts No. APVV-19-0590 and also by the projects VEGA 1/0700/20, KEGA 055TUKE-4/2020 granted by the Ministry of Education of the Slovak Republic.

References

1. Peraković, D., Periša, M., Sente, R.E.: Information and communication technologies within Industry 4.0 concept. In: *Advances in Design, Simulation and Manufacturing*. DSMIE 2018. LNME. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-93587-4_14
2. Trojanowska, J., Kolinski, A., Galusik, D., Varela, M.L.R., Machado, J.: A methodology of improvement of manufacturing productivity through increasing operational efficiency of the production process. In: Hamrol, A., Ciszak, O., Legutko, S., Jurczyk, M. (eds) *Advances in Manufacturing*. LNME. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-68619-6_3
3. Mayo, M.: Data a general approach to preprocessing text data. <https://www.kdnuggets.com/2017/12/general-approach-preprocessing-text-data.html>. Accessed 21 Oct 2021
4. Mishra, P., Biancolillo, A., Roger, J.M., Marini, F., Rutledge, D.N.: New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC - Trends Anal. Chem.* **132**, 116045 (2020). <https://doi.org/10.1016/j.trac.2020.116045>
5. Gandhi, P., Pruthi, J.: Data visualization techniques: traditional data to big data. In: Anuncia, S., Gohel, H., Vairamuthu, S. (eds) *Data Visualization*. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2282-6_4
6. Unwin, A.: Why is data visualization important? What is important in data visualization? *Harvard Data Sci. Rev.* **2**(1) (2020). <https://doi.org/10.1162/99608f92.8ae4d525>
7. Husar, J., Knapcikova, L., Balog, M.: Implementation of material flow simulation as a learning tool. In: *Advances in Design, Simulation and Manufacturing*. DSMIE 2018. LNME. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-93587-4_4
8. Kliment, M., et al.: Production efficiency evaluation and products' quality improvement using simulation. *Int. J. Simul. Model.* **19**(3), 470–481 (2020)
9. Hicks, S.C., Peng, R.D.: Evaluating the success of a data analysis. [arXiv:1904.11907](https://arxiv.org/abs/1904.11907) (2019)
10. Manda, V.K., Poosapati, V., Katneni, V.: Super SCADA systems: a prototype for next gen SCADA system. *IAETSD J. Adv. Res. Appl. Sci.* (2018)

11. Chen, Q., Lee, S.: A machine learning approach to predict customer usage of a home workout platform. *Appl. Sci.* **11**, 9927 (2021). <https://doi.org/10.3390/app11219927>
12. Diether, A.: Machine learning for financial applications, mathworks. <https://www2.humusoft.cz/www/papers/finkonf-2017/diethert.pdf>. Accessed 21 Nov 2021
13. Bitna, K., Young, H.P.: Beginer's guide to neural networks for the MNIST dataset using MATLAB. *Korean J. Math.* **26**(2), 337–348 (2018). <https://doi.org/10.11568/kjm.2018.26.2.337>
14. Al Zamily, J.Y., Abu Naser, S.S.: Lemon classification using deep learning. *Int. J. Acad. Pedagog. Res. (IJAPR)* **3**(12) (2019)
15. Hošovský, A., Piteř, J., Trojanová, M., Židek, K.: Computational intelligence in the context of Industry 4.0. Implementing Industry 4.0 in SMEs. Palgrave Macmillan, Cham (2021). https://doi.org/10.1007/978-3-030-70516-9_2
16. Phitchayanon, P.: Machine learning & deep learning with Matlab. https://muit.mahidol.ac.th/muit_training/matlab2019/Machine-Learning-and-classification-App.pdf. Accessed 22 Nov 2021
17. Mathworks: Matlab for Data analytics. <http://lib.itpcas.ac.cn/documents/18/0/Data+Analytics+MATLAB.pdf>. Accessed 21 Nov 2021
18. Siirtola, P., Röning, J.: Comparison of regression and classification models for user-independent and personal stress detection. *Sensors* **20**, 4402 (2020). <https://doi.org/10.3390/s20164402>
19. BKA, M.A.R., Ngamkhanong, C., Wu, Y., Kaewunruen, S.: Recycled aggregates concrete compressive strength prediction using artificial neural networks (ANNs). *Infrastructures* **6**(2), 17 (2021). <https://doi.org/10.3390/infrastructures6020017>
20. Ocampo, I., López, R.R., Camacho-León, S., Nerguizian, V., Stiharu, I.: Comparative evaluation of artificial neural networks and data analysis in predicting liposome size in a periodic disturbance micromixer. *Micromachines* **12**, 1164 (2021). <https://doi.org/10.3390/mi12101164>
21. Vagaská, A., Gombár, M.: Comparison of usage of different neural structures to predict AAO layer thickness. *Tehnicky Vjesnik* **24**(2) (2017). <https://doi.org/10.17559/TV-20140423164817>