Early Cost Estimation in Customized Furniture Manufacturing Using Machine Learning

O. Kurasova, V. Marcinkevičius, V. Medvedev, and B. Mikulskienė

Abstract—**Accurate cost estimation at the early stage of a construction project is a key factor in the success of most projects. Many difficulties arise when estimating the cost during the early design stage in customized furniture manufacturing. It is important to estimate the product cost in the earlier manufacturing phase. The cost estimation is related to the prediction of the cost, which commonly includes calculation of the materials, labor, sales, overhead, and other costs. Historical data of the previously manufactured products can be used in the cost estimation process of the new products. In this paper, we propose an early cost estimation approach, which is based on machine learning techniques. The experimental investigation based on the real customized furniture manufacturing data is performed, results are presented, and insights are given.**

Index Terms—**Cost estimation, customized furniture manufacturing, analogous estimating, machine learning, prediction.**

I. INTRODUCTION

Customization of products is a challenging trend for the recent year. Competitive pressure, complex customer requirements, expectations of interested parties initiate additional requirements for manufacturers and products. This has particularly affected the manufacturing of customized furniture. Thereby, companies need decision support tools to ensure an immediate response to individual orders and a proper assessment of production procedures, costs, and timing. The task to estimate costs as precise and timely as possible has become critical in customized manufacturing [1].

More and more customers want to stand out and to order exclusive furniture. Customization changes pricing policy, as a specificity of customized furniture manufacturing, significantly influences the furniture cost. When considering the factors that influence the success of the results of manufacturing, it is evident that the cost is as essential as quality and functionality. The cost depends on furniture complexity (the furniture consists of many components, a part of them is complex), production volume (large volume decreases the cost of one item), design overhead (a new design needs additional designer time, or the existing design

can be simply customized), production peculiarities (much manual work, expensive machine tools are used). One more issue arises, when the cost has been estimated before manufacturing while the information about the actual costs is most limited, it is called early cost estimation. The accuracy of this estimation can strongly affect company profits: too low price will reduce profits, too high price can deter customers.

Different approaches have been developed for early cost estimation, however, they are missing in the furniture industry. Cost estimation is defined as the process of predicting the cost of a product before all stages of product development have been completed. As a rule, the accuracy of a cost estimation reflects the information currently available. Furniture manufacturing has a great variety of products and prices, from smallholders to large expensive furniture sets. Cost estimation of the product in the early design stage is useful for accelerating product entry to the market, reducing costs, and improving quality with a high level of competitiveness in the market [2]. It is very important for companies to properly and promptly react to changes in the growing complexity of customized manufacturing to be able to produce maximum precision in their pricing. Thus, machine learning-based approaches of cost estimation would be adapted for the furniture production domain. Machine learning is widely used in various fields of production [3]-[7] and can be an effective and accurate technique to estimate the cost of customized furniture. The objective of the research is to propose a machine learning-based approach to estimate the cost for customized furniture manufacturing in the early design phase.

The rest of the paper is organized as follows. A comprehensive review of related works is performed and presented in Section II. Section III is devoted to machine learning-based cost estimation. A use case analysis is presented in Section IV. Finally, Section V concludes the research.

II. COST ESTIMATION TECHNIQUES

Recently, the main factors for the successful development of manufacturing enterprises are related to product quality, competitive cost, fast delivery, and flexibility. Moreover, product customization puts extra pressure on manufacturing enterprises to survive and remain profitable, when allowed consumers to purchase products that are customized to their specifications [8]. In customized manufacturing, also called make-to-order, the production of an item begins only after a customer is confirmed an order. Here, early cost estimation is an extremely important factor.

Cost estimation is concerned with the prediction of costs

related to a set of activities and materials before they have been executed [9], [10]. Firstly, it is needed for manufacturing enterprises for financial plans and strategies. Secondly, this is especially important for customized manufacturing, when customers want to know the product price in advance. This preliminary price should be as close to the final price as possible. To calculate the price, the cost drivers should be identified considering customer requirements. The process usually takes a lot of time and requires many resources. To attract customers, the time for cost estimation should be decreased.

A simple way of the furniture price calculation is as follows: $P_w = C_p + P_p$, were P_w is the price of the product, C_p is the total cost of production, P_p is the profit. The cost C_p consists of direct costs and overhead costs. Direct cost includes direct material costs, direct labor cost, and other costs. Production costs, administrative costs, the cost of disposing belong to overhead costs.

Some important elements are related to cost estimation [11]. Firstly, the cost estimating technique requires extensive historical data including not only features that influence the cost, but also the costs of the previous products. Secondly, it is necessary to identify human and financial resources, how many people (cost estimators), and what their competence is required to estimate the cost, what financial support needs. It is also essential to know both company and customer expectations, outcome, and usage of the estimate. Finally, a schedule is also an important element, how much time is to collect the required data, how much time is needed for evaluation to complete the estimate, given the available resources and data.

In the literature, cost estimation techniques are divided into *qualitative* or *quantitative* [8] (Fig. 1). An idea of qualitative cost estimation techniques is based on a comparison of a new product with the previously manufactured products to identify their similarities. The similarities help to incorporate the past data into the new product, and there is no need for the cost estimate from scratch. Thus, using the cost estimator experience, the previous design and manufacturing data provide useful information to calculate a reliable cost estimate of a new product that is similar to the manufactured one according to past design. Qualitative cost estimation techniques can be categorized into *intuitive* and *analogical* techniques. The intuitive cost estimation techniques are grounded on the cost estimator experience [12]. Knowledge is used to generate cost estimates. The expert knowledge can be stored in the form of rules, decision trees, and judgments at a specific database to facilitate the decision-making process and prepare cost estimates for new products. When using *analogical techniques* (sometimes called top-down), the cost

of a new product is estimated according to the cost of similar known products. Here, the similarity is commonly evaluated without any quantitative measure.

Quantitative cost estimation techniques are based on a detailed analysis of product design, its features, called cost drivers, and corresponding manufacturing processes instead of simply relying on the past data or knowledge of a cost estimator. Although these techniques are known to provide more accurate results, the usage is normally restricted to the final design phases due to the requirement of detailed product design. Quantitative techniques can be further categorized into *parametric* and *analytical* techniques [13].

In *parametric* techniques (sometimes called statistical), historical data and empirical examinations are evaluated to gain information on the causal link between product features and costs [13]. In other words, parametric cost estimates are a result of a cost estimating using statistical relationships between historical costs and other product characteristics. Regression analysis is widely used for cost estimation as a well-defined mathematical approach [14], [15]. Moreover, this analysis enables to explain the significance of a variable and the relationships between other variables. Artificial neural networks are also suitable for cost estimations. However, they require a large amount of historical data for accurately training.

Analytical cost techniques (sometimes called bottom-up estimating and engineering build-up) provide a detailed decomposition of costs, and the total cost is usually computed as a sum of such parts as labor, material, machine-related things multiplied by the time required to perform manufacturing tasks. Analytical approaches depict the relevant processes of product creation in detail, and derive the costs incurred, aggregating them properly [13]. The analytical approach requires decomposing a product into elementary units, operations, and activities that represent different resources consumed during the production cycle and expressing the cost as a summation of all these components [8].

Moreover, some *analogical* techniques can be assigned to quantitative one, when numerical similarity measure is used to identify the analogous product [16]. Analogy estimates are based on a comparison of the new product to similar ones. They can be applied when the new product is similar to a number of others that were carried out earlier. In this case, the total cost of the product is determined on the basis of accumulated experience. Cost estimation using analogy reasons from functional or geometrical similarity to a similar product considering the similarity measures describing the level of correspondence of the relevant characteristics [13].
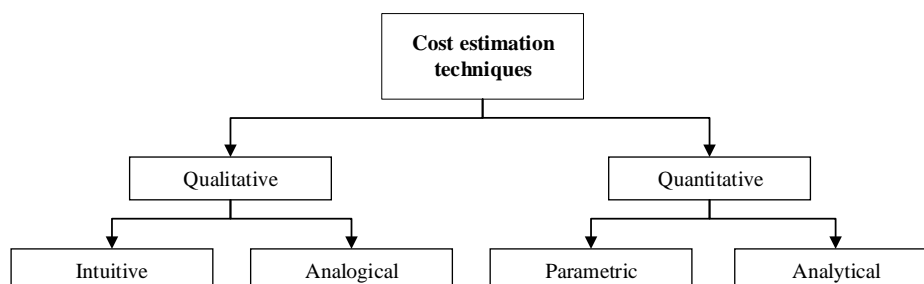

Fig. 1. Cost estimation techniques.

III. MACHINE LEARNING-BASED COST ESTIMATION

Customized furniture manufacturing companies can be categorized into two groups. Companies of one group are responsible for the design, thus, all manufacturing process includes from design creation to final manufacturing. In these cases, the final cost depends on the complexity of the design, and the manufacturer and the customer can negotiate the price, reducing or complicating the design.

Other type companies obtain orders with a design sketch (quotation), and it is necessary to estimate the cost of the furniture. The cost estimation from sketch analysis would be accurate because sketches reveal the complexity of items that influences manufacturing cost. However, if we want to involve machine learning techniques, a problem arises due to historical data. A lot of various sketches need because an image recognition problem should be solved, and the training process requires many data/images to obtain accurate results. It is necessary to have many similar sketches to train image recognition algorithms accurately. Furthermore, it is another problem in customized furniture manufacturing.

We propose to solve the cost estimation problem by analyzing the numerical data of manufacturing. Different manufacturing companies use different tools for data storage as well as different methodologies for cost estimation. It would be appropriate to store the data in Enterprise Resource Planning (ERP) or other systems, however, a typical case, especially small companies, when the companies do not use ERPs or unable to exploit all the functionality. Thus, data extraction is also a relevant problem in this domain. Moreover, if data are of a large volume, we face the challenge to analyze them using machine learning techniques [17]. It is important to set a proper data set for cost estimation in customized furniture manufacturing. Some categories could be identified:

1) item measurement (length, height, width, weight, the volume of the bounding box);
2) material data (the material used for production, material cost);
3) operational data (operation list and the time required to complete the operation process);
4) labor data (much manual work, expensive machine tools are used); production time (the customer's requirement for production time);
5) batch size (more the same products reduce the cost of one product);
6) manufacturing complexity (a qualitative parameter indicative of the uniqueness of the item and complexity of the work).

The cost estimation is a time-consuming problem for many furniture manufacturing companies, especially for companies that produced customized items. When a quotation for a new product is received, it needs to estimate if this product is new for the company. If the product has been manufactured before, the cost is known. However, if the ordered product is new, it is necessary to estimate the components of the cost such as material quantities and cost, operation and labor cost, packing cost. After that, overhead is added, and the quotation answer is ready for sending to the customer. This process is depicted in Fig. 2(a). The usage of machine learning can significantly shorten the time of this process. We propose to change the estimation of operations and labor as well as

packing costs by machine learning-based estimation so that the final cost would be estimated only from material data (see Fig. 2(b)).
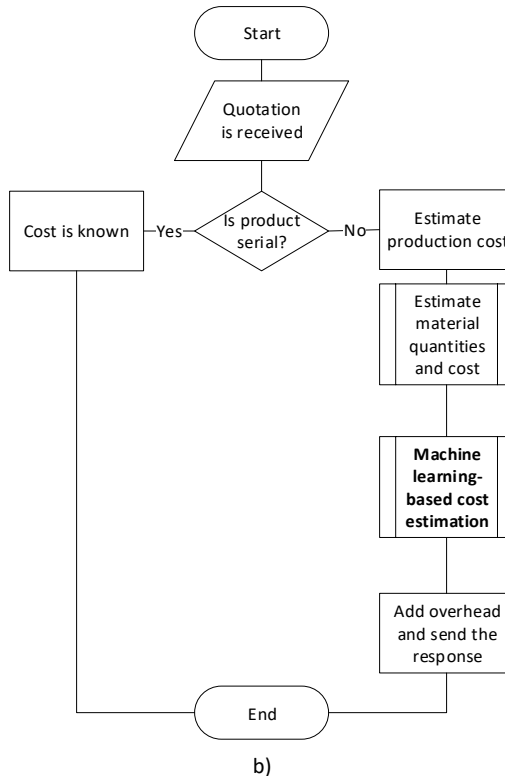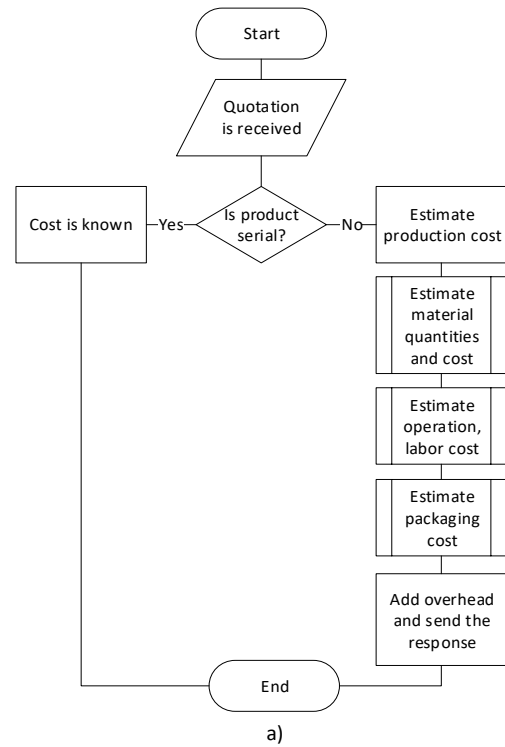

Fig. 2. Cost estimation: a) labour intensive process; b) proposed approach using machine learning.

Various prediction approaches can be used to solve the cost estimation problem:

1) Linear regression. It is the simplest regression model, where the relationship between a dependent variable and one or more independent variables is linear. The advantage of this type of regression is that the obtained results are easily interpreted. Thus, the linear regression

is widely used, despite the fact that there are more accurate prediction approaches.

2) Decision tree-based regression. There are plenty of methods that build regression in the form of a tree structure. The following methods are currently the most popular and give the most accurate prediction results: Decision tree, Random forest, Extra tree, AdaBoost, Gradient boosting.

3) K-neighbors regression. It a non-parametric method used for regression, where the input consists of the k closest training examples in the feature space, the output is the predicted value, which is obtained by averaging the values of k nearest neighbors. Because of its simplicity, k-neighbors-based methods are often used for big data analysis.

4) Artificial neural networks (ANN). Recently, this is a rapidly developing group of artificial intelligence techniques that trained by machine learning paradigms and is able to handle various complex data analysis problems. ANNs can be successfully employed as prediction approaches, too. In most cases, they outperform other methods, however, in order to achieve the most accurate results, a large amount of historical data for training is required.

All the mentioned methods have their advantages and disadvantages; therefore, it makes sense to use them all and choose the best results as the final ones. Moreover, all algorithms are implemented in popular machine learning libraries (e.g., *NumPy*, *scikit-learn*, etc.), thus, their use does not cause any technical difficulties.

IV. Use Case Analysis

The real historical data for 1026 products provided by a Lithuanian furniture manufacturing company are used in the use case analysis and experimental investigation. The data gathered over the last five years include the real costs of these products. A set of products includes items of various sizes and complexity (from small pieces of furniture to large furniture kits).

TABLE I: Data Features Analyzed

x_i	Notation	Short descriptions
$x_1 - x_5$	m10-m50	metal processing times
$x_6 - x_{10}$	w60-w100	wood processing times
x_{11}	m	total meters of materials
x_{12}	m^2	total square meters of materials
x_{13}	qty	total amount of materials
x_{14}	kg	total weight of materials
x_{15}	qty_parts	number of parts
x_{16}	qty_unique_parts	number of different parts
x_{17}	qty_materials	number of different materials
x_{18}	qty_order	quantities (order size)

Thus, the data set consists of values of some features of products (Table I):

1) y is a cost (prediction/dependent variable),
2) $x_1, \ldots, x_n$ are operational times (exploratory/independent variables),
3) $x_{n+1}, \ldots, x_m$ are quantitative material features (exploratory variables).

Commonly, solving data analysis problems requires to identify the most important features. Some feature selection approaches can be employed, e.g., the principal component analysis (PCA) [18], methods of univariate feature selection (e.g., *SelectKBest*), and others. Applying *SelectKBest*, five of the most important features (from Table I) have been identified: 'm^2', 'qty_parts,' 'm,' 'qty_unique_parts,' 'qty.' As we see, all these features are from a set $x_{n+1}, \ldots, x_m$ and assigned to the material features.

In practical applications, the cost is estimated by analogous. Usually, it performs intuitively, when the person responsible for costing determines which products are similar. However, when machine learning techniques are applied, it is necessary to evaluate products according to mathematical metrics of similarity. Some similarity metrics can be used: Euclidean, standardized Euclidean, Cityblock (Manhattan), and Hamming distances.

The exploration process is as follows:

1) Similarities are computed by various metrics between all pair of items when the inputs are $x_1, \ldots, x_m$.
2) Similarities are also computed by the same metrics between all pairs of items when the inputs are y (cost).
3) Correlation coefficients are computed between similarities obtained at Steps 1 and 2.

The high values of the correlation coefficients show that the items that are similar in similarity metrics will be similar in cost, too. The experimental investigation shows that the data are the most similar according to types of Minkowski distances (Cityblock, Euclidean, standardized Euclidean distance).

Solving the cost estimation problem, a relationship between y and x_i needs to be indicated. Thus, cost estimation can be formulated as a prediction problem. Statistic-based methods find this relationship via some mathematical functions. The other machine learning techniques create a more sophisticated model from historical data.

TABLE II: Prediction Result Evaluation

Regression algorithm	R^2 avg	R^2 std	RMSE avg	RMSE std
Linear (all)	0.793	0.125	136.84	55.47
Linear (5-features)	0.739	0.11	155.97	50.52
Decision tree (all)	0.726	0.118	159.5	42.18
Decision tree (5-features)	0.698	0.096	167	39.7
Random forest (all)	0.842	0.079	122.15	44.64
Random forest (5-features)	0.806	0.094	134.14	47.24
Extra tree (all)	0.738	0.098	157.87	44.99
Extra tree (5-features)	0.72	0.074	161.88	34.57
AdaBoost (all)	0.809	0.075	134.06	38.36
AdaBoost (5-features)	0.775	0.097	144.78	46.08
Gradient boosting (all)	0.84	0.074	123.02	40.48
Gradient boosting (5-features)	0.802	0.094	136.54	48.99
K neighbors (all)	0.811	0.087	133.76	43.29
K neighbors (5-features)	0.799	0.099	138.61	52.51
ANN (all)	0.805	0.077	133.97	36.46
ANN (5-features)	0.764	0.075	150.48	42.39

In this investigation, the various prediction algorithms are used. During the training process, cross-validation with 10 folds is applied. The average values (avg) and standard

deviations (std) of determination coefficient R^2 and root mean square error (RMSE) are presented in Table II and Fig. 3.

Two cases are analyzed:
1) when all features $x_1, \ldots, x_m$ are used for prediction,
2) and five features which have been identified as the most important using the *SelectKBest* method.

The experimental investigation shows that the highest value of R^2 reaches by Random forest algorithm (0.842) and Gradient boosting algorithm (0.84). Both algorithms are assigned to decision-based methods. As expected, ANNs did not show the best results as the training data set is not large. The results also demonstrate the fact that the reduction of the number of features affects accuracy slightly. Moreover, the prediction results even worsen when the number of features is reduced to five.
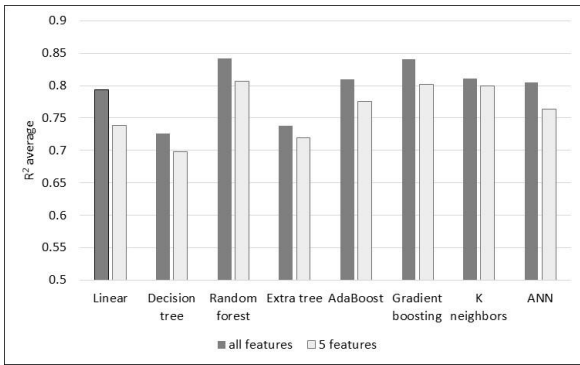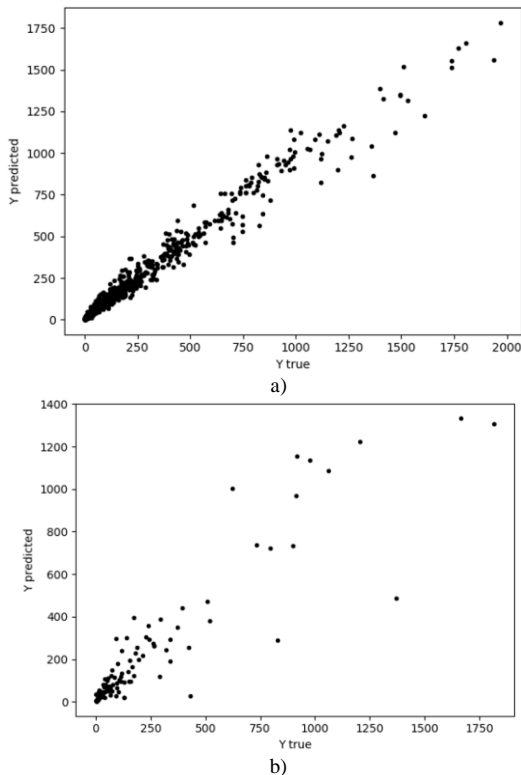

Fig. 3. Average values of determination coefficient R^2.

a)

b)

Fig. 4. Graphical representation of cost estimation (all features): a) training (R^2=0.970, RMSE=53.61); b) testing (R^2=0.823, RMSE=148.14).

Grid search strategies are used to search over specified hyper-parameter values for algorithms. In Fig. 4 and Fig. 5, the best results are depicted considering all features and only five, respectively, when the prediction problem is solved. The x-axis corresponds to true cost values, and the y-axis corresponds to the predicted cost values. The results demonstrate that, in the best case, R^2 reaches 0.977 for training data and 0.839 for testing data. Moreover, we see that the smaller costs are predicted more accurately than in the cases of higher costs. Thus, further research lines can be related to cost estimation, taking into consideration cost groups, machine learning-based clustering methods can be employed.
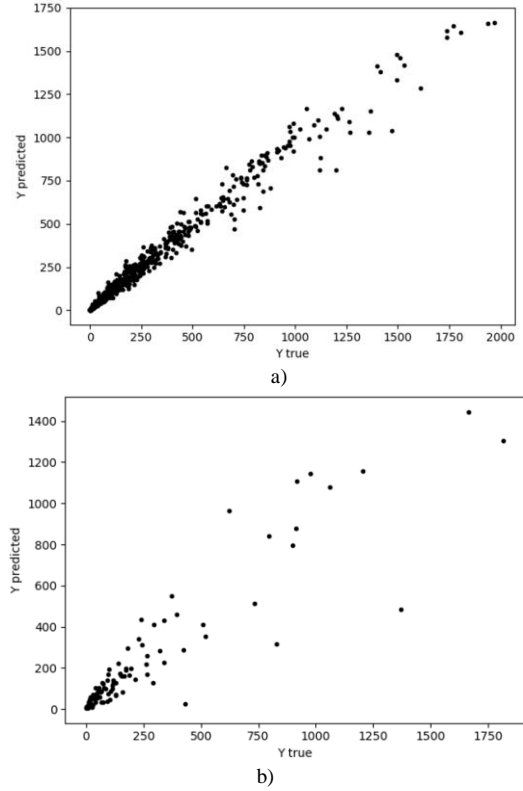

a)

b)

Fig. 5. Graphical representation of cost estimation (5-features): a) training (R^2=0.977, RMSE=47.39); b) testing (R^2=0.839, RMSE=142.82).

V. CONCLUSIONS

The paper has investigated an early cost estimation problem facing customized furniture manufacturing. Usually, the cost estimation is a complicated and time-consuming process due to the need to evaluate many components in the early design stage, when information is most limited. Moreover, a lot of human resources are required. Application of the machine learning techniques allows simplifying and accelerating this process by providing accurate and effective cost estimation. However, using machine learning techniques, it is crucial to have a proper set of historical data for the training process and to identify essential data features to obtain accurate results.

In the experimental investigation, the real historical data for 1026 products provided by a Lithuanian furniture manufacturing company are used. The study has shown that the cost of the customized furniture can be estimated quite precisely when only data on the quantities of materials are used. The use of machine learning techniques can reduce the time required to estimate the cost in the early design phase and to accelerate product entry to the market.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENT

REFERENCES

[1] B. Mikulskienė, V. Medvedev, T. Vedlūga, and O. Navickienė, "Behaviour patterns in expert recognition by means of structured expert judgement in price estimation in customized furniture manufacturing," in *Proc. Modeling and Simulation of Social-Behavioral Phenomena in Creative Societies. MSBC 2019. Communications in Computer and Information Science*, 2019, vol. 1079.

[2] H. D. Budiono, G. Kiswanto, and T. P. Soemardi, "Method and model development for manufacturing cost estimation during the early design phase related to the complexity of the machining processes," *International Journal of Technology*, vol. 2, pp. 183-192, 2014.

[3] D. D. Clercq, D. Jalota, R. Shang, K. Ni, Z. Zhang, Z. A. Khan, A. Z. Wen, L. Caicedo, and K. Yuan, "Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data," *Journal of Cleaner Production*, vol. 218, pp. 390-399, 2019.

[4] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel, "A review of machine learning for the optimization of production processes," *International Journal of Advanced Manufacturing Technology*, vol. 104, pp. 1889-1902, 2019.

[5] D. Lieber, M. Stolpe, B. Konrad, J. Deuse, and K. Morik, "Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning," *Procedia Cirp*, vol. 7, pp. 193-198, 2013.

[6] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144-156, 2018.

[7] N. Fang and P. S. Pai, "A new computational intelligence approach to predicting the machined surface roughness in metal machining," *International Journal of Machine Learning and Computing*, vol. 8, no. 6, pp. 524-529, 2018.

[8] A. Niazi, J. S. Dai, S. Balabani, and L. Seneviratne, "Product cost estimation: Technique classification and methodology review," *Journal of Manufacturing Science and Engineering*, vol. 128, no. 2, pp. 563-575, 2006.

[9] G. Molcho, A. Cristal, and M. Shpitalni, "Part cost estimation at early design phase," *CIRP Annals*, vol. 63, no. 1, pp. 153-156, 2014.

[10] E. Shehab and H. Abdalla, "Manufacturing cost modelling for concurrent product development," *Robotics and Computer-Integrated Manufacturing*, vol. 17, no. 4, pp. 341-353, 2001.

[11] NASA cost estimating handbook. (April 2019). [Online]. Available: https://www.nasa.gov/o_ces/ocfo/nasa-cost-estimating-handbook-ceh

[12] F. K. Cheung, J. Rihan, J. Tah, D. Duce, and E. Kurul, "Early stage multi-level cost estimation for schematic BIM models," *Automation in Construction*, vol. 27, pp. 67-77, 2012.

[13] A. Layer, E. T. Brinke, F. V. Houten, H. Kals, and S. Haasis, "Recent and future trends in cost estimation," *International Journal of Computer Integrated Manufacturing*, vol. 15, no. 6, pp. 499-510, 2002.

[14] I. Mahamid, "Early cost estimating for road construction projects using multiple regression techniques," *Construction Economics and Building*, vol. 11, no. 4, pp. 87-101, 2011.

[15] G. H. Kim, S. H. An, and K. I. Kang, "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning," *Building and Environment*, vol. 39, no. 10, pp. 1235-1242, 2004.

[16] J. L. Loyer, E. Henriques, M. Fontul, and S. Wiseall, "Comparison of machine learning methods applied to the estimation of manufacturing cost of jet engine components," *International Journal of Production Economics*, vol. 178, pp. 109-119, 2016.

[17] O. Kurasova, V. Marcinkevičius, V. Medvedev, A. Rapečka, and P. Stefanovič, "Strategies for big data clustering," in *Proc. IEEE 26th International Conference on Tools with Artificial Intelligence*, 2014, pp. 740-747.

[18] G. Dzemyda, O. Kurasova, and J. Žilinskas, *Multidimensional Data Visualization: Methods and Applications*, Berlin: Springer, 2013.

Olga Kurasova received a doctoral degree in computer science (PhD) from the Institute of Mathematics and Informatics jointly with Vytautas Magnus University in 2005. He is now at Mykolas Romeris University as a principal researcher as well as at the Institute of Data Science and Digital Technologies of Vilnius University as a principal researcher and professor. Her research interests include data mining methods, optimization theory and applications, artificial intelligence, neural networks, visualization of multidimensional data, multiple criteria decision support, parallel computing, and image processing. She is the author of more than 70 scientific publications.

Virginijus Marcinkevičius got the PhD in computer science from Vytautas Magnus University. He is a senior researcher from Mykolas Romeris University. His research interest includes machine learning, artificial intelligence, cybersecurity, and natural language processing.

Viktor Medvedev received a doctoral degree in computer science (PhD) from Institute of Mathematics and Informatics jointly with Vilnius Gediminas Technical University in 2008. His research interests include artificial intelligence, visualization of multidimensional data, dimensionality reduction, neural networks, data mining and parallel computing.

Birutė Mikulskienė is a professor of management in the Faculty of Public Governance in Mykolas Romeris University (Lithuania), where she teaches Theory of Decision Making and R&D Project Management Master and PhD students. Her research interest cover quality and performance management in both the public and private sector with special focus on expert judgment, social network analyses, and stakeholder management. She also conducts public management and policy analysis research in various sectors such as health, education, science, and business with special focus on public participation.