A brief comparative study of the potentialities and limitations of machine-learning algorithms and statistical techniques

Li Eckart[1], *Sven* Eckart[2], and *Margit* Enke[3]

[1]Department of Marketing and International Trade, Technische Universität Bergakademie Freiberg, Freiberg, Germany
[2]Institute of Thermal Engineering, Technische Universität Bergakademie Freiberg, Freiberg, Germany
[3]Department of Marketing and International Trade, Technische Universität Bergakademie Freiberg, Freiberg, Germany

Abstract: Machine learning is a popular way to find patterns and relationships in high complex datasets. With the nowadays advancements in storage and computational capabilities, some machine-learning techniques are becoming suitable for real-world applications. The aim of this work is to conduct a comparative analysis of machine learning algorithms and conventional statistical techniques. These methods have long been used for clustering large amounts of data and extracting knowledge in a wide variety of science fields. However, the central knowledge of the different methods and their specific requirements for the data set, as well as the limitations of the individual methods, are an obstacle for the correct use of these methods. New machine learning algorithms could be integrated even more strongly into the current evaluation if the right choice of methods were easier to make. In the present work, some different algorithms of machine learning are listed. Four methods (artificial neural network, regression method, self-organizing map, k-means al-algorithm) are compared in detail and possible selection criteria are pointed out. Finally, an estimation of the fields of work and application and possible limitations are provided, which should help to make choices for specific interdisciplinary analyses.

1 Introduction

"How to use the power of such big data" becomes a central topic across various research domains in the nowadays big data world. Machine-learning algorithms can provide practicable solutions to analyze and leverage the massive volume of available data. The roots of machine learning can be found in computer science, engineering, mathematics, physics, and astronomy. These technologies also find extensive practical use in the field of neuroscience and medicine [1-13]. In that field, artificial intelligence (AI) was used for very topical issues such as the spread of pandemics [14] and as an effective cancer prediction method [15]. Various problems were also investigated in the field of engineering sciences. Salehi et al. [16] use AI for the prediction of structures in the construction sector.

Whereas Tabor et al. [17] used these approaches to achieve a better synergy of different renewable energy areas and the chemical industry, Ali and Frimpong [18] applied these methods for the automation of mining processes and mining systems. In contrast, these methods are used in the social sciences to make risk analyses fairer [19] and to better understand learning processes and emotions [20]. However, merely a few studies discussed the application and perspectives of machine learning algorithms for the marketing research area [21,22].

Since the selection of the appropriate method represents a complex issue, this work attempts to introduce four novel machine-learning techniques: the artificial neural network (ANN), regression method, k-mean-algorithm, and self-organizing map (SOM), to compare them with similar approaches, as well as to point out the trends and perspectives for further research issues. Finally, an assessment of the fields of use and potential limitations will be provided to assist in the selection for specific interdisciplinary analyses.

Two main types of machine learning algorithms are considered in the present work: supervised and unsupervised learning. The supervised machine-learning approaches can solve prediction and regression problems, whereas the unsupervised learning process is involved in the explorative analysis with little or no prior knowledge, for instance, clustering problems. This work confines attention to ANN as representative of the supervised learning algorithms and to SOM as representative of the unsupervised learning algorithms. Furthermore, machine-learning approaches name the statistical terms differently from the conventional statistics. For instance, the input and output of machine learning algorithms equate respectively the independent and dependent variables in the conventional statistics. The target and error in machine learning can be viewed as the actual value and residuals of the conventional techniques. Moreover, the training cases or samples are applied to conduct the machine-learning process. Such training cases are called observations in the conventional framework.

In order to allow a general comparison of the four selected forecast methods, publications since 1980 were evaluated and shown in Figure 1. All publications examined were in the scientific database Scopus. Keywords, abstracts, and titles were included in the clustering and filtered according to the four methods used in this work. Subsequently, the publications were assigned to the research areas and were investigated in detail. It can be seen that the first rise in usage can be observed, in particular, after the release of the world wide web and its commercial use around 1990. The second rise appeared in the mid-2000s, where the electronic data sum quadrupled in only four years. [23]. An additional rapid increase in machine learning, which can be related to the topics of smart city [24,25], smart home [26], industry 4.0 [27,28] and digitization [29,30], can be observed in the last few years after 2014 [31]. Overall, the ANN method is mentioned and applied most often, followed by K-means, regression method, and SOM.

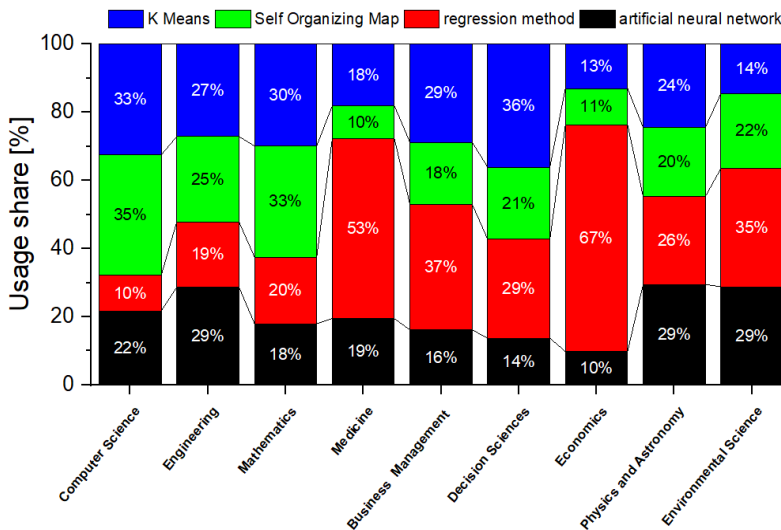**Fig. 1.** Number of annual publications using the methods studied.

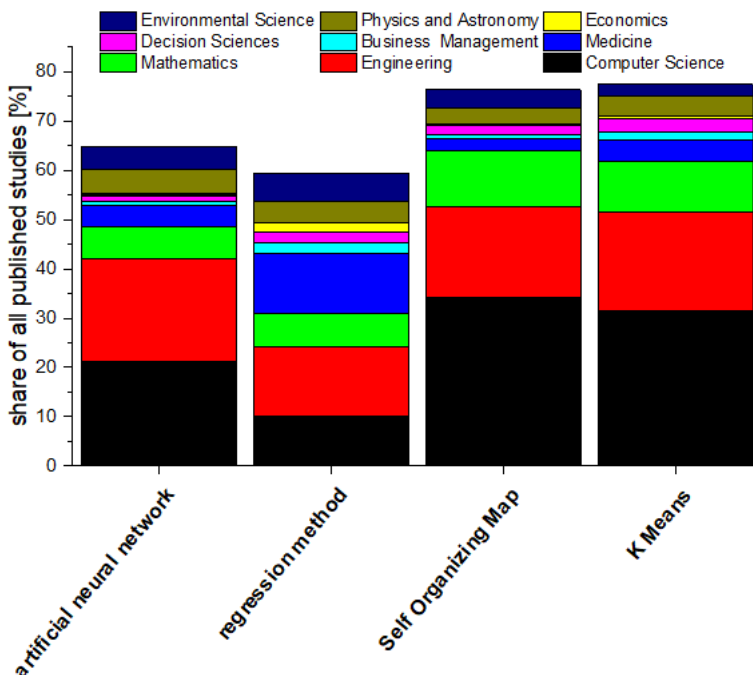**Fig. 2.** Use of the four selected methods investigated in different subject areas.

Fig. 3. Proportion of publications per field of research for the different methods examined.

The total number of publications was spread mainly among nine scientific fields, covering over 60% of all papers on this topic. This data set is shown in Figure 2. The four fields (computer science, engineering, mathematics, and medicine) account for at least 40% of each method. In the ANN, SOM, and K-means, the fields of engineering and computer science are represented by more than 20%. The usual disciplines play a rather small role,

sometimes less than 1%. It can be shown that the dissemination of the method could be significantly expanded, especially in the business, management, and economic sectors. Furthermore, the scientific areas in which the methods are already being used have been identified. They are shown in Figure 3. Specific fields of science are not uniformly distributed. While regression methods are much more commonly used in medicine and economics, neural networks are mainly used in engineering, physics, and environmental sciences. The K means method is being used increasingly in the decision sciences, as well as in mathematics. The SOM method is also a common approach in computational science and mathematics.

2 Methods

In the present work, four different mathematical approaches were used, namely Artificial Neural Network (ANN), Self-Organizing Map (SOM), Regression Method, and K-mean method. A short description of the classifiers used and their current applications can be found in the following section.

2.1 Artificial neural network

As a special type of supervised machine learning, the artificial neural network (ANN) is inspired by the human neural systems, first introduced in the 1950s. Due to the massive scale of big data and novel supervised learning algorithms, this approach has lately resurfaced.

An ANN is viewed as a system of connected directed artificial neurons. In an ANN, the process of a biological neural network will be reproduced with connected artificial neurons. Each network layer consists of artificial neurons and bias. The network signals will pass through various network layers and finally arrive at the output layers. Mathematical functions (propagation function and activation function) represent the strengths of neural connections and activation levels of neurons. The connecting weights will be stored in the propagation function. The received signal will be transformed into network input using the activation function, also called the transfer function. Bias can be seen as a weight, of which input always equals one. Adding bias into an the ANN enables the model to perform an affine transformation (nonlinear) instead of a linear transformation. The application of bias prevents the problem that a neuron without a bias will always transfer zero when the network inputs are zero.

The magnitude and direction of the neural connections show the different contributions of inputs referring to the output. Inputs with larger connection weights possess greater intensities in the signal transfer process, compared to inputs with smaller connection weights. The connection directions can be inhibitory (negative) or excitatory (positive). The inhibitory connections reduce the intensity of the received signal and the excitatory connections, oppositely, intense the received signal. It indicates that the connection weights can provide insight into the relative importance of input variables. Therefore, the information about the weights can be used to assess the robustness of the observed system.

Based on the universal approximation theorem [32], an ANN with one hidden layer can approximate any continuous function arbitrarily well, regardless of the activation function and the dimension of the input space [33]. An ANN with bias, a nonlinear, and a linear output layer is capable of approximating any function with a finite number of discontinuities without any previous justification. This shows the high predictive ability of ANNs and gives the theoretical support for the great performance of ANNs in nonlinear statistical modeling.

Given input and target data, the weights of the ANN can be trained by various learning algorithms, for instance, backpropagation. In each learning step of the backpropagation, the weights in the propagation function will be adapted to minimize learning errors (differences between outputs and targets). The adaptive changes of weights follow the learning rule determined by learning algorithms. The knowledge of the trained ANN stored in weights and bias can be used for further prediction.

Among different structures of ANNs, a multilayer feed-forward neural network (MFNN) consisting of three layers (input, hidden, and output layer) is favored among the scholars (see Figure 4). An ANN with n $(, \vec{x_i}, i = 1, \cdots, n)$ inputs is considered. $\vec{w_{i,j}}$ implies the weights of the connections between the input layer and the hidden layer, which consists of m neurons $(j = 1, \cdots, m)$. $\vec{b_j}$ describes the bias of the neuron j. The network signal net_j is represented by the propagation function in terms of inputs, weights, and bias

$$net_j = \sum (\vec{x_i} \cdot \vec{w_{i,j}} + \vec{b_j}) \qquad (1)$$

Depending on the activation level of neurons, the network signal is transmitted with the help of the activation function

$$a_j = f(net_j) = f(\sum (\vec{x_i} \cdot \vec{w_{i,j}} + \vec{b_j})). \qquad (2)$$

After the activation process, the signal a_j is passed to the output layer. For simplicity, it is assumed that merely one neuron is contained in the output layer. The output value of the neuron in the output layer is Y

$$Y = \sum \vec{w_{j,r}} f(net_j) + \vec{b_r} = \sum \vec{w_{j,r}} f(\sum (\vec{x_i} \cdot \vec{w_{i,j}} + \vec{b_j})) + \vec{b_r} \qquad (3)$$

where b_r is the bias of the output neuron. The frequently employed learning goal includes squared error $\sum (Y - y)^2$ and absolute error $|Y - y|$ between the output Y and the target y[13].
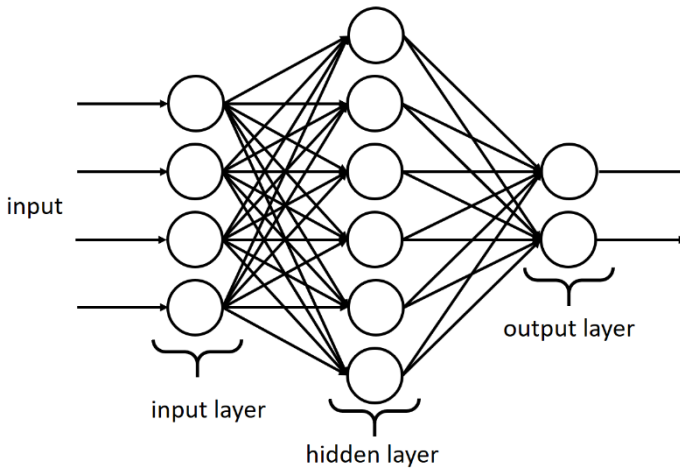

Fig. 4. Schematic illustration of ANNs.

2.2 Regression method

Similar attempts to examine the relationship between the independent and dependent variables were also identified by the idea of the regression. Wold's decomposition theorem [34] contributes to the foundation of the linear forecasting issue of Kolmogorov in 1941. In this paper, the regression method can be subdivided into generalized linear models (MLM) and support vector regression (SVR).

MLM employs the linear function to describe the relationship, considering parallelly the error distribution. Thus, the predicted value can be given by

$$Y_i = g\left(x_i^j, \beta_j\right) + E_i. \qquad (4)$$

E_i represents the error term and g is the linear regression function, which must be predefined. Previous studies have shown the easy usage and the efficiency of the MLM. However, the main limitation of the MLM, its incompetence to handle nonlinear problems has also been pointed out.

In contrast to the MLM, SVR can be applied to a nonlinear context. Additionally, SVR defines the acceptable area in the model to fit the data in a better way. The input is mapped through a nonlinear function (so-called kernel function) [35]. To predict the dependent variable, a linear regression function is computed in high-dimensional feature space [36]. The determination of the kernel function is crucial for the result accuracy of the SVR. However, there are still no common guidelines for the selection of the kernel function .

SVR optimizes the prediction based on the principle of structural risk minimization. Derived from this principle, the upper bound of the general error which includes the training error will be minimized [37].

Table 1. Comparison of advantages and disadvantages of the SVR regression method.

Advantage	Disadvantage
● Consideration of the general error, not only the training error ● Better balance between the result accuracy and computational cost	● No common guidelines for the selection of the kernel function

2.3 Self-organizing map

As part of unsupervised learning, the self-organizing map (SOM) acts as the cluster approach in machine learning. Kohonen[38] first introduced the SOM. The unsupervised learning rules refer to the machine-learning rules of the artificial neural networks by giving the input but without revealing the output, and it assumes no special data structure. Hence, this approach is especially appropriate for explorative data analysis in terms of identifying hidden patterns in unknown structured data sets. Indeed, the SOM is a single-layer neural network with a pre-specified number of neurons spread on a two-dimensional lattice. The purpose of this approach is to classify objects and to reduce the complexity of high-dimensional problems.

The input data will be transformed into a low-dimensional discrete map by topological mapping. The nonlinear statistical relationships between high-dimensional data can be converted into geometric relations on a low-dimensional display [38]. The SOM follows the "winner-takes-all" rule. Its process generally comprises four steps: initialization, competition, cooperation, and adaption.

1)   In the first step of initialization, all the connection weights are initialized with small random values.

2)   For a given input vector $x_i, i = 1, \cdots, n$, its similarity to the neurons $w_i, i = 1, \cdots, n$ can be depicted as the Euclidean distance between them. The closest neuron w_k to the input vector is the so-called winning neuron, mathematically: $\|x_i - w_k\| = min\|x_i - w_i\|, \ i = 1, \cdots, n$. The values computed by the discrimination function (e.g., the Euclidean distance function) provide the basis for the next step.

3)   In the last step, the connection weights of the winning neuron and its connected neighborhood neurons will also be adapted according to the rule proposed by Kohonen[38,39]: $w_i(t+1) = w_i(t) + \alpha(t) \bullet g_i(t) \bullet [x_i(t) - w_i(t)]$.

t presents the number of iteration steps. The connection weights in the $t+1$ iteration step depend on the previous outcome $w_i(t)$ and the learning rate $\alpha(t)$, as well as the neighborhood function $g_i(t)$. $g_i(t)$ determines how fast the connection weights of the neighborhood are allowed to be adapted. One of the most commonly appropriated neighborhood functions of the SOM approach is the Gaussian function. Based on the Gaussian function, the neurons near the winning neuron will "learn" faster than those far away. For more details about this approach, see the work of Kohonen[38,39].

4)   Repeat steps 2 and 3 until the convergence criteria are fulfilled.

As a result, the information of the initial high-dimensional problems will be visualized by different color shades on a 2-dimensional lattice in SOM weight planes. After the SOM learning process, each training sample is connected to all neurons on the lattice and assigned to a neuron according to connection weights (similarities). It implies that all the input samples in a certain neuron share similar characteristics. The following illustration (Figure5) shows an exemplary visualization of the SOM using MATLAB:
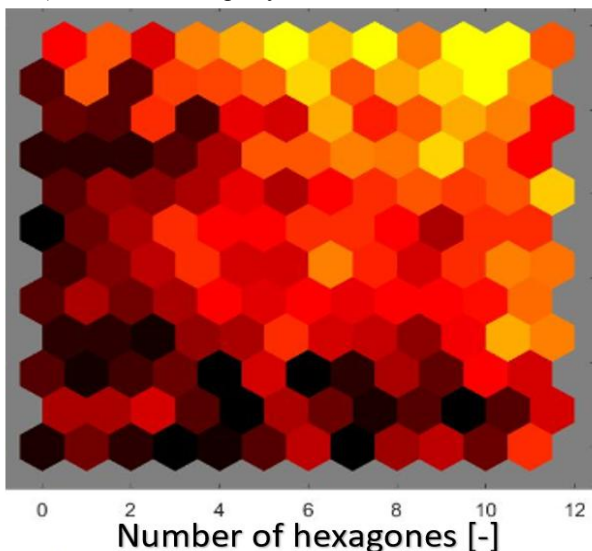

Fig. 5. Visualization of a self-organizing map.

2.4 K-means algorithm

Like the SOM, the conventional approach k-means algorithm belongs to the squared error-based (vector quantization) clustering algorithms. Except for the squared error-based approaches, hierarchical, density-based, matrix optimization-based and mixture algorithms have also been reviewed in the previous studies [40,41]; see the following table.

Table 2. Comparisons of squared error-based clustering algorithms.

Category	Clustering algorithm
Square error- based algorithms	- k-means [42] - self-organizing analysis [43,44] - genetic k-means algorithm [45] - PAM (Partitioning Around Medoids) [46]

Hierarchical algorithms	- Single linkage [47] - Complete linkage [48] - Group average linkage [34] - BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies) [49] - CURE (Clustering Using Representatives) [50] - ROCK (Robust Clustering using LinKs) [51]
Density-based algorithms	- DBSCAN [52] - OPTICS [53] - HDBSCAN [54] - MeanShift[55]
Matrix-optimization algorithms	Affinity propagation [56]
Mixture algorithms	Gaussian Mixture Modeling [57]

Due to the limited scope of this paper, the k-means algorithm is selected to conduct the comparison with the SOM. The cross-categorical observations could be interesting fields for future research.

The K-means algorithm is classified as a non-hierarchical approach. Therefore, the number of clusters should be predefined. The aim of this method is to assign objects with similar characteristics to the same cluster, and different clusters should be heterogeneous. The K-means algorithm is a greedy algorithm. Due to the characteristics of the greedy algorithm, the best solution can converge into a local minimum, and the global optimum cannot be guaranteed. Besides, the initial centroid choice influences the accuracy of the result of the k-means algorithm.

Based on the comparison of the Euclidean distances between one object and the centroid of each cluster, the object is assigned to the closest cluster (with minimal Euclidean distance). After the update of cluster members, the centroids will be recalculated. The stop criteria of this algorithm are 1) the marginal modification of the centroids between the iteration steps and 2) every object is assigned to one cluster [58].

The basis iteration steps of the k-means algorithm:

1)   Determine the Cluster number (K)
2)   Initialization of the centroids ($c_j, j = 1, \cdots, K$)
3)   Allocation of the training data to each cluster based on the minimal Euclidean distance criterion:$\|x_i - c_l\| = min\|x_i - c_j\|, j = 1, \cdots, K, 1 \leq l \leq K$. x_i is the i-th sample of input data set and is assigned to the l-th cluster.
4)   The regarded centroid will be recalculated after the assignment of x_i

$$c_{l\_new} = \frac{1}{N_{l\_new}} \sum x_n \qquad (5)$$

c_{l_new} is the updated centroid considering the new cluster member x_i. N_{l_new} is the number of the cluster members including x_i.
5)   Repeat steps 3 and 4 till convergence occurs.

3   Discussion and classification

In the following, the methods described above are examined and validated for different possible applications. Based on this, concrete recommendations for their application are derived. Two of the four methods are compared with each other for a specific application area. First, the use of the ANN and regression is investigated, and then the SOM and k-means are compared in detail.

3.1 Comparison between artificial neural network and regression

In this section, two approaches predicting the relationship between the independent and the dependent variables are discussed. While the ANN algorithm has been successfully applied as a powerful prediction tool to various research disciplines [3,6,59,60], the regression method is viewed as one of the standard tools to deal with prediction problems [35,59,36]. The ANN algorithm and the regression method are selected to conduct the comparative analysis owing to the following considerations: 1) the ANN algorithm, as well as the regression method, is proposed to examine the relationship between the independent and dependent variables; 2) both techniques can be used with respect to minimizing the mean squared error between the predicted value and the actual value.

Based on the previous considerations, it can be summarized:

Table 3. Comparison between artificial neural network and regression.

Comparison between artificial neural network and regression		
Advantages of ANN	1)	The ANN algorithm is a data-driven approach. The underlying relationship between inputs and outputs is based on the given data sets. To develop a regression model, this underlying relationship is predefined by a (linear or nonlinear) regression function. The connection weights between the neurons are similar to the beta coefficients in a least-squares regression model. However, the effect of an independent variable on the dependent variable is depicted by different neurons on different layers instead of by only one beta coefficient in the regression approach.
	2)	The ANN algorithm does not require the absence of multicollinearity. Since multicollinearity impacts the performance of the regression method, in particular, the relevant information in highly collinear data sets may not be discovered by the regression approach [61]. The ANN can outperform the regression method in the case of multicollinearity.
	3)	The ANN algorithm can investigate implicitly any complex nonlinear relationship between independent and dependent variables [62,32,63]. Compared with ANNs, the conventional regression approach employs merely one layer to describe the relationship between independent and dependent variables. The multilayer structure enables ANNs to model more complex nonlinear relationships [12]. Hence, ANNs can be tailored to the idiosyncratic marketing research problems owing to various potential learning algorithms and activation functions.
	4)	An ANN including hidden layers considers all the interplays between the independent variables. Due to the particular structure of ANNs, all the interactions between the independent variables are taken into account in the learning process of the ANN algorithm, whereas those interactions should be explicitly added to the regression model.
Disadvantages of ANN:	1)	Due to lack of transparency, the ANN model has been criticized as a "black box" model. Based on data sets, an ANN is trained to learn the relationship between the independent and dependent variables by itself. There still are no standard guidelines to interpret the connection weights of ANNs , even if prior scholars attempted to reveal more insight into the connection weights of the ANN system [64-66].
	2)	Compared with regression methods, the computational cost of ANNs is relatively high. The development of the ANN model is an intensive process, which requires more computational resources than the conventional regression method.
	3)	The ANN requires a large number of training samples.
	4)	The ANN algorithm is susceptible to overfitting. The hidden layer structure of an ANN enables the depiction of all possible interactions of

	independent variables. However, the superfluous high-order consideration can lead to overfitting [12]. To solve this problem, prior research suggested to reduce the number of hidden nodes. Adding a penalty term or limiting the amount of training, as well as using cross-validation, may also help avoid the overfitting problem in ANNs [67].

The ANN is a powerful and versatile tool to conduct data analysis and prediction. This data-driven approach uses the given historical or real-time data to perform a non-linear mapping between the input and output arrays. Despite the potential overfitting problem, the ANN as a universal approximator is preferable to the conventional techniques. In particular, the ANN outperforms other methods in terms of the result accuracy by dealing with complex and nonlinear relationships, for instance, in the area of medical imaging analysis [68,69], fundamental flame parameters [70], times series forecasting [13], and demand analysis [71]. To ensure high accuracy, the ANN requires usually a huge amount of data sets. Consequently, the application of the ANN is connected with high computational costs. Therefore, the ANN is less attractive for small-sized samples with simple relationships. In this case, the conventional regression method is beneficial for forecasting, for instance, in the business administration research domain [72,73].

3.2 Comparison between the self-organizing map and k-means algorithm

In this part of the work, the focus is put on the approaches dealing with the classification problems. Such problems are the main concerns of the pattern and language recognitions, image data compression [74], and robotics [75-80]. SOM and k-means algorithms are both acknowledged to handle the clustering tasks well. For both techniques, the Euclidean distance plays an important role in the clustering process. Based on the previous consideration, it can be summarized:

Table 4. Comparison between the self-organizing map and k-means algorithm.

	Comparison between self-organizing map and k-means algorithm	
Advantages of SOM:	1)	The SOM does not predefine the number of clusters, whereas the number of clusters should be determined in advance in the k-means algorithm. The initial settings (i.e., the choices of the initial centroids) will impact the accuracy of the result of the k-means approach [81]. The result of the k-means approach is sensitive to the previous model setting.
	2)	The SOM takes into consideration the neighborhood of the winning neuron . In the k-means algorithm, neighborhood samples are not taken into account in the clustering process.
	3)	The method provides new visualization possibilities.
	4)	Additionally, the SOM can project a high-dimensional data space into a low-dimensional data space [80].
	5)	The SOM is proven to outperform other methods, , data sets in particular, e.g., the skewed data [22,82,83].
	6)	The SOM is less prone to local optima [84]
Disadvantages of SOM:	1)	No standard guidance of the interpretation of the SOM visualization results.
	2)	Overfitting problem of the SOM. Due to the structure, the SOM is also prone to overfitting problems, like the ANN algorithm [60].
	3)	The theoretical groundwork in terms of network parameter settings is still incomplete. The decision about the structure of ANNs, such as the number of layers and neurons shuld be made by the scholars themselves [80].
	4)	The disadvantage of the k-means: this approach is based on the Euclidean distance and inclines to generate hyperspherical formed

	clusters. For other geometric forms, the k-means has low efficiency.

The superiority of the SOM in the treatment of multiple optima problems is shown. In principle, the SOM offers the possibility of early exploration of the search space, and as the search progresses, it gradually narrows the search. At the end of the search process (if the neighborhood radius decreases to zero), the SOM is exactly equal to the k-mean, which allows minimizing the distances between the observations and the cluster centers. Therefore, it can be concluded that self-organizing maps are a time-efficient method for data sets with both high numbers of samples and high dimensionality and could be further used in the fields of, for example, customer behavior, energy consumptions diagnostics, additive manufacturing, or investigations of high-speed diffusion flames.

4   Conclusion

This paper provides a brief overview of various approaches to conducting the data analysis, ranging from conventional statistics to machine-learning algorithms, evolving from different research disciplines. The focus of this work was on the novel machine-learning algorithms and comparing them with other similar approaches, including their basic ideas, strengths, and weaknesses. In particular, the ANN vs. the regression methods and the SOM vs. the k-means algorithms are compared. The central knowledge of the different methods, their specific requirements, as well as the limitations of the individual methods, can often be an obstacle for the correct use of the methods.

As a data-driven prediction tool, ANNs are already widely applied for a large number of research areas ranging from engineering and environmental science to physics and astronomy. The flexibility of ANNs as universal approximation makes them preferable to the regression method. The challenges in this approach are to determine the appropriate structure and learning algorithm. Compared with the regression approach, ANNs can provide more accurate predictions, especially for the short-term prediction. Moreover, the regression method will easily run into its limits, if a wide range of predictors should be considered. As a data-driven cluster tool, the SOM can transfer the high-dimensional problems into a one- or two-dimensional topographic map [85]. Such dimension-reduced results can be employed for further cluster analysis using other algorithms. It has been noted that a good SOM result requires the appropriate data pretreatment [86]. Compared with the k-means algorithm, the SOM does not necessitate a priori specifications of the cluster number and the initial positions of the cluster centers.

As the computational cost of both machine-learning algorithms will increase rapidly with the increasing data size, the tradeoff between the result accuracy and the computational effort should be always taken into account. In general, data-driven techniques may pose the risk of over fitting and case-dependence. To address this issue, additional research is required. Further, future research should extend the application areas of the machine-learning algorithms, which are evidently underrepresented, such as marketing, sociology, and political science.

The results of the comparisons indicate that there is no universal approach to solve all research problems. Thus, the selection of the appropriate technique is based on the specific applications and experiment conditions. The new technology enables scholars to deal with high-complexity and high-dimensional problems. However, there remain questions referring to the efficiency of the novel algorithms:

1)   How to conduct data analysis with high-dimensional features within the scope of the acceptable computational cost?
2)   How to handle the outliers in a better way?
3)   How to provide more insight into the analysis process of black-box models?

In summary, the aforementioned methods of machine-learning discipline enable the science to communicate to analyze complex data sets of enormous size and solve challenging big data problems on applicable time scales. Despite the complicacy, it is widely distributed and implemented due to the lack of clear and reliable guidelines for model settings and the burden of interpreting results. . Nevertheless, to pursue such state-of-the-art methods, scientists must be motivated by a deeper understanding of the problems in their areas of research.

References

1.   A.V. Boikov, R.V. Savelyev, V.A. Payor, O.O. Erokhina, Evaluation of bulk material behavior control method in technological units using DEM. Part 1. *CIS Iron and Steel Review*,**19**:4–7(2020).

2.   E.N. Grishchenkova, Development of a Neural Network for Earth Surface Deformation Prediction. *Geotech Geol Eng*. **36(4)**: 1953–1957(2018).

3.   G.E. Hinton, R.R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks. *Science*. **313(5786)**: 504–507(2006).

4.   S. Hochreiter, J. Schmidhuber Long Short-Term Memory. *Neural Computation*. **9(8)**: 1735–1780(1997)

5.   D.T.Jones, Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *Journal of Molecular biology*. **292(2)**: 195–202(1999).

6.   Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature*. **521(7553)**:436–444(2015).

7.   R. Milo, S.S. Shen-Orr, S. Itzkovitz, N.Kashtan, D.M. Alon, U. Chklovskii, Network Motifs: Simple Building Blocks of Complex Networks. *Science*. **298(5594)**: 824–827(2002).

8.   H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of Their Cleavage Sites. *Protein Engineering Design and Selection*. **10(1)**: pp. 1–6(1997).

9.   J.D.Olden, D.A.Jackson, Illuminating the "Black Box": a Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks. *Ecological Modelling*. **154(1-2)**: 135–150(2002).

10. M.Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J.Denzler, N.Prabhat Carvalhais, Deep learning and process understanding for data-driven Earth system science. *Nature*. **566(7743)**: 195–204(2019).

11. M. Rubinov, O.Sporns, Complex Network Measures of Brain Connectivity: Uses and Interpretations. *NeuroImage*. **52(3)**: 1059–1069(2010).

12. J.V.Tu, Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *Journal of Clinical Epidemiology*. **49(11)**: 1225–1231(1996).

13. C.Voyant, G.Notton, S.Kalogirou, M.-L.Nivet, C.Paoli, F.Motte, A Fouilloy Machine Learning methods for solar radiation forecasting. A review. *Renewable Energy*. **105**: 569–582(2017).

14. RVaishya, MJavaid, IHKhan, A Haleem Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & metabolic syndrome*. **14(4)**: 337–339(2020)

15. K.Potdar, R.Kinneerkar, A Comparative Study of Machine Learning Algorithms Applied to Predictive Breast Cancer Data. *International Journal of Science and Research*. **9**: 1550–1553(2016).

16. H.Salehi, R.Burgueño, Emerging artificial intelligence methods in structural engineering. *Engineering Structures*. **171**: 170–189(2018).

17. D.P.Tabor, L.M.Roch, S.K.Saikin, C.Kreisbeck, D.Sheberla, J.H.Montoya, S.Dwaraknath, M.Aykol, C.Ortiz, H.Tribukait, C.Amador-Bedolla, C.J.Brabec, B.Maruyama, K.A.Persson, A.Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater*. **3(5)**: 5–20(2018).

18. D.Ali, S.Frimpong, Artificial intelligence, machine learning, and process automation: existing knowledge frontier and way forward for the mining sector. *ArtifIntell Rev*. **53(8)**: 6025–6042(2020).

19. R.Berk, H.Heidari, S.Jabbari, M.Kearns, A Roth Fairness in Criminal Justice Risk Assessments. *Sociological Methods & Research*:004912411878253(2018).

20. A.Chatterjee, U.Gupta, M.K.Chinnakotla, R.Srikanth, M.Galley, P.Agrawal Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior*. **93**: 309–317(2019).

21. D.Bejou, B.Wray, T.N.Ingram Determinants of relationship quality: An artificial neural network analysis. *Journal of Business Research*. **36(2)**: 137–143(1996).

22. M.Y.Kiang, A.Kumar, An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications. *Information Systems Research*. **12(2)**: 177–194(2001).

23. C.Stutzman, D.Cooperstein, C.Munchbach,*Measure and Manage Brand Health*. *How To Diagnose And Treat Your Brand's Resilience And Responsiveness*(CMO Professionals. Forthcoming. 2012)

24. I.Khajenasiri, A.Estebsari, M.Verhelst, G.Gielen, A Review on Internet of Things Solutions for Intelligent Energy Control in Buildings for Smart City Applications. *Energy Procedia*. **111**: 770–779(2017).

25. J.Winkowska, D.Szpilko, S.Pejić, Smart city concept in the light of the literature review. *Engineering Management in Production and Services*. **11(2)**: 70–86(2019).

26. M.Alaa, A.A.Zaidan, B.B.Zaidan, M.Talal, M.Kiah, A review of smart home applications based on Internet of Things. *Journal of Network and Computer Applications*. **97**: 48–65(2017).

27. V.Alcácer, V.Cruz-Machado, Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science and Technology, an International Journal*. **22(3)**: 899–919(2019).

28. K.-D.Thoben, S.Wiesner, T.Wuest "Industrie 4.0" and Smart Manufacturing – A Review of Research Issues and Application Examples. *Int*. *J*. *Automation Technol*.**11(1)**: 4–16(2017).

29. J.Reis, M.Amorim, N.Melão, P.Matos, Digital Transformation: A Literature Review and Guidelines for Future Research. *World Conference on Information Systems and Technologies*: 411–421(2018).

30. G.Vial Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*. **28(2)**: 118–144(2019).

31. P.Asghari, A.M.Rahmani, H.H.S.Javadi, Internet of Things applications: A systematic review. *Computer Networks*. **148**: 241–261(2019).

32. A.N.Kolmogorov, On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. *Dokl*. *Akad*. *Nauk SSSR*. **114**, 5(1957).

33. K.Hornik, M.Stinchcombe, H.White, Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*. **2(5)**: 359–366(1989).

34. H.O.A.Wold, On Prediction in Stationary Time Series. *Annals of Mathematical Statistics*. **19(4)**: 558–567(1948).

35. D.Basak, S.Pal, D.C.Patranabis, Support Vector Regression. *Neural Information Processing – Letters and Reviews*. **11(10)**: 203–224(2007).

36. Z.Wang, R.S.Srinivasan A review of artificial intelligence-based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*. **75**: 796–808(2017).

37. A.Foucquier, S.Robert, F.Suard, L.Stéphan, A.Jay, State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*. **23**: 272–288(2013).

38. T. Kohonen, The Self-organizing Map. *Proceedings of the IEEE*. **78(9)**:1464–1480 (1990).

39. T.Kohonen, Physiological interpretation of the self-organizing map algorithm. *Neural Networks*. **6(6)**: 895–905(1993).

40. A.Saxena, M.Prasad, A.Gupta, N.Bharill, O.P.Patel, A.Tiwari, M.J.Er, W.Ding, C.-T.Lin, A Review of Clustering Techniques and Developments. *Neurocomputing*. **267**: 664–681(2017).

41. R.Xu, D. Wunsch, Survey of clustering algorithms. *IEEE transactions on neural networks*. **16(3)**: 645–678(2005).

42. *James B*. MacQueen, Some methods for classification and analysis of multivariate observations, *5th Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley*(1967).

43. T.Kohonen, The self-organizing map. *Neurocomputing*. **21(1-3)**: 1–6(1998).

44. J.Vesanto, E.Alhoniemi, Clustering of the self-organizing map. *IEEE transactions on neural networks*. **11(3)**: 586–600(2000).

45. D.E.Goldberg,*Genetic algorithms in search, optimization, and machine learning* (Addison-Wesley, Reading, Mass., Wokingham, 1989).

46. L.Kaufman, P.J.Rousseeuw,*Finding groups in data*. *An introduction to cluster analysis* (Wiley, New York, 1989).

47. P.H.A.Sneath, R.R.Sokal,*Numerical taxonomy*. *The principles and practice of numerical classification*.(Forthcoming, 1973)

48. B King, Step-Wise Clustering Procedures. *Journal of the American Statistical Association*. **62(317)**: 86 (1967)

49. T.Zhang, R.Ramakrishnan, M.Linvy, BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Record*. **25(2)**: 103–114(1996).

50. S.Guha, R.Rastogi, K.Shim, CURE. An efficient clustering algorithm for large databases. *ACM SIGMOD Record*. **27(2)**: 73–84(1998).

51. S.Guha, R.Rastogi, K.Shim, Rock: A robust clustering algorithm for categorical attributes. *Information Systems Research*. **25(5)**: 345–366(2000).

52. *Martin* Ester, *Hans-Peter*Kriegel, *Jörg* Sander, *Xiaowei* Xu, A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*: 226–231(1996).

53. M.Ankerst, M.M.Breunig, H.-P.Kriegel, J.Sander OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Record*. **28(2)**: 49–60(1999).

54. R.J.Campello, D.Moulavi, A.Zimek, J.Sander, Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Transactions on Knowledge Discovery from Data*. **10(1)**: 1–15(2015).

55. Y.Cheng, Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **17(8)**: 790–799(1995).

56. B.J.Frey, D.Dueck, Clustering by passing messages between data points. *Science*. **315(5814)**: 972–976(2007).

57. G.J. McLachlan, E.B. Kaye, *Mixture models: Inference and applications to clustering*. Vol. 38. (New York: M. Dekker, 1988).

58. J.A. Hartigan, M.A. Wong, A. K-Means Clustering Algorithm. *Applied Statistics*. **28(1)**: 100–108 (1979)

59. N.Vásquez, C.Magán, J.Oblitas, T.Chuquizuta, H.Avila-George, W.Castro, Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of Swiss-type cheese using spectral profiles. *Journal of Food Engineering*. **219**: 8–15(2018).

60. I.M.Yusri, Majeed A.Abdul, R.Mamat, M.F.Ghazali, O.I.Awad, W.H.Azmi, A Review on the Application of Response Surface Method and Artificial Neural Network in Engine Performance and Exhaust Emissions characteristics in alternative fuel. *Renewable and Sustainable Energy Reviews*. **90**: 665–686(2018).

61. R.H.Myers *Classical and Modern Regression With Application* (Duxbury Press, Boston, MA, 1986).

62. G.E.Hinton, How Neural Networks Learn from Experience. *Scientific American*. **267(3)**: 144–151(1992).

63. H.White, Learning in Artificial Neural Networks: A Statistical Perspective. *Neural Computation*. **1(4)**: 425–464(1989)

64. G.D. Garson Interpreting Neural-network Connection Weights. *Artificial Intelligence Expert*. **6(4)**: 46–51(1991).

65. J.D.Olden, M.K.Joy, R.G.Death, An Accurate Comparison of Methods for Quantifying Variable Importance in Artificial Neural Networks Using Simulated Data. *Ecological Modelling*. **178(3-4)**: 389–397(2004).

66. J.D.Olden, D.A.Jackson, Illuminating the ''Black Box'': A Randomization Approach For Understanding Variable Contributions in Artificial Neural Networks. *Ecological Modelling*,**154**: 135–150(2002).

67. M.Smith,*Neural networks for statistical modeling* (Van Nostrand Reinhold, New York, 1993).

68. J.-G.Lee, S.Jun, Y.-W.Cho, H.Lee, G.B.Kim, J.B.Seo, N.Kim, Deep Learning in Medical Imaging: General Overview. *Korean journal of radiology*. **18(4)**:570–584(2017).

69. K.Suzuki, Overview of deep learning in medical imaging. *Radiological physics and technology*. **10(3)**: 257–273(2017).

70. S.Eckart, C.Penke, S.Voss, H.Krause, Laminar burning velocities of low calorific and hydrogen containing fuel blends. *Energy Procedia*,**120**: 149–156(2017).

71. I.Antonopoulos, V.Robu, B.Couraud, D.Kirli, S.Norbu, A.Kiprakis, D.Flynn, S.Elizondo-Gonzalez, S.Wattam, Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renewable and Sustainable Energy Reviews*. **130**:109899(2020).

72. F.Kujur, S.Singh, Emotions as a predictor for consumer engagement in YouTube advertisement. *JAMR*,**15(2)**: 184–197(2018)

73. C.Yoo, K.C.Cha, S.-H.Kim, A quantile regression approach to gaining insights for reacquisition of defected customers. *Journal of Business Research*. **120**: 443–452(2020)

74. J.M.Wandeto, B.Dresp-Langley, The quantization error in a Self-Organizing Map as contrast and color-specific indicator of single-pixel change in large random patterns. *Neural networks: the official journal of the International Neural Network Society*. **119**: 273–285(2019)

75. A.Delbimbo, L.Landi, S.Santini Three-dimensional planar-faced object classification with Kohonen maps. *Opt. Eng*.**32(6)**: 1222–1234(1993)

76. H.Jangid, S.Jain, B.Teka, R.Raja, A.Dutta, Kinematics-based end-effector path control of a mobile manipulator system on an uneven terrain using a two-stage Support Vector Machine. *Robotica*. **38(8)**: 1415–1433(2020)

77. S.Kuramoto, H.Sawada, P.Hartono, Visualization of the topographical internal representation of learning robots. *International Joint Conference on Neural Networks*: 1–7(2020)

78. L.Leinonen, T.Hiltunen, K.Torkkola, J.Kangas Self-organized acoustic feature map in the detection of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*. **93(6)**: 3468–3474(1993)

79. J.A.Walter, K.I.Schulten, Implementation of self-organizing neural networks for visuomotor control of an industrial robot. *IEEE transactions on neural networks*. **4(1)**: 86–96(1993)

80. Y.Wei, X.Zhang, Y.Shi, L.Xia, S.Pan, J.Wu, M.Han, X.Zhao, A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*,**82**: 1027–1047(2018)

81. G.W.Milligan, M.C.Cooper A study of standardization of variables in cluster analysis. *Journal of Classification*. **5(2)**: 181–204(1988)

82. P.Mangiameli, S.K.Chen, D.West A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*. **93(2)**: 402–417(1996)

83. N.G.Waller, H.A.Kaiser, J.B.Illian, M.Manry A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. *Psychometrika*,**63(1)**: 5–22(1998).

84. F.Bação, V.Lobo, M.Painho, Self-organizing Maps as Substitutes for K-Means Clustering. *Computational Science – ICCS*, 476-483(2005)

85. Çs.F.Magoule *Data Mining and Machine Learning in Building Energy Analysis. Towards High-Performance Computing*. (Wiley-ISTE, UNITED STATES, 2014).

86. I.P.Panapakidis, T.A.Papadopoulos, G.C.Christoforidis, G.K.Papagiannis, Pattern recognition algorithms for electricity load curve analysis of buildings,*Energy and Buildings*,**73**: 137–145(2014)