

MACHINE LEARNING VERSUS DEEP MACHINE LEARNING WITH THE USE OF BIG DATA INFORMATION CONTROL SYSTEMS

1. Effective configurations of classical machine learning with the use of Big Data information control systems

In its nascent years, artificial intelligence (AI) largely focused on rule-based systems that would make predictions using predefined sets of rules that had to be provided by a subject matter expert. However, these systems were fragile and relied on these «expert opinions», which eventually led to their falling out of mainstream use. As the scale and volume of data increased, these methods were replaced by a more data-driven approach – machine learning (ML).

ML is a set of algorithms and tools that help machines understand patterns in data and use that underlying structure to make inferences about a particular task. There are many ways in which machines seek to understand these basic patterns [1, 2].

ML is a class of methods for the automated creation of predictive models based on data. Machine learning algorithms turn a dataset into a model. Which algorithm works best (supervised, unsupervised, classification, regression, etc.) depends on the type of problem to be solved, the available computing resources, and the nature of the data [3–5].

ML algorithms can be described as learning an objective function f that best correlates the input variables X and the output variable Y : $Y = f(X)$.

It is not known what the function f is. After all, if it were known, it would be used directly, and not tried to learn with the help of various algorithms.

The most common task in ML is to predict Y values for new X values. This is called predictive modeling, and our goal is to make as accurate a prediction as possible.

ML is a subset of general AI and a general term for computers learning from data. It describes the intersection of computer science and statistics, where algorithms are used to perform a specific task without explicit programming; instead, they recognize patterns in the data and make predictions as new data comes in. supervised learning, unsupervised learning, and reinforcement learning.

After all, it is the method of learning that is most often used to classify machine learning into large categories: supervised, unsupervised, reinforcement learning, and learning with partial involvement of the «teacher», each of them has its own tasks of practical application in real conditions and its own distinguishing features and limitations.

In general, the ML process is based on the following steps:

- 1) preliminary analysis of the data set and the set ML problem, cleaning and preparation of data for use by the corresponding ML algorithm,
- 2) selection of the algorithm and setting, selection of its parameters;
- 3) data loading and model training;
- 4) testing and interpretation of the model;
- 5) implementation and use of the model for the automated solution of predictive tasks.

Unsupervised learning

While supervised learning requires input-output pairs (or labeled data), unsupervised learning algorithms use only input data (unlabeled data). Although this learning method is less intuitive, it is suitable for problems where we have little or no idea what our results should look like. The goal is to gain knowledge and find structure in the data.

The unsupervised ML algorithm processes unlabeled data, including finding different ways of presenting complex data to organize them or describe their internal structure, which simplifies their interpretation.

In unsupervised learning, there may be no target variable, i.e. the training data set does not have labeled data belonging to a certain class of predetermined data, i.e. «Teacher» (target variable) is absent. The algorithm independently determines correlations and connections based on the analysis and interpretation of all available data and draws conclusions based on them. The algorithm tries to organize the data in any way and describe its structure. The algorithm observes some similarities between groups of objects and joins them to the appropriate clusters. Some objects may be very different from all the clusters, and thus the algorithm assumes that these objects are anomalies. As more data is added for analysis, the algorithm's ability to make decisions based on that data grows, as does the accuracy of those decisions.

With only some arbitrary data at their disposal, sometimes untrained learning algorithms can be able to detect some non-trivial dependencies or even some kind of complex laws.

Teaching methods without a teacher solve the following tasks:

- clustering;
- dimensionality reduction;
- search for association rules (buyer basket analysis, etc.);
- detection of data anomalies.

ML without a teacher is also used as an auxiliary tool for the preparation of data sets for the subsequent correct solution of predicative tasks by machine learning algorithms with a teacher.

Supervised machine learning

Supervised learning uses a known/explicit relationship between inputs and output. This is where labeled data comes into play: the goal of the algorithm is to learn from the «correct answers» in the training data and use the insights to make predictions as new input is received.

The term «supervised learning» comes from the fact that we initially provided the algorithm with a data set in which «correct answers» were provided. This is a key difference in unsupervised learning.

Learning with a «teacher» functions only on labeled data and is currently the most developed and applicable type of ML. To implement it in practice, you need a task that can be formulated as a classification or regression analysis problem, as well as a sufficient set of labeled data. Learning with a teacher is the task of training the system on a training set of data that contains the necessary input and output values. The learning algorithm with the «teacher» by iteratively «fitting» the learning results to the training data set, searches for the most optimal parameters of the model (for further prediction of possible answers on other / new objects / records). The process continues until the algorithm reaches a given level of accuracy / performance (stopping criterion).

Supervised ML algorithms are divided into two main groups:

- if the set of possible answers is a real number, then this is a regression task;
- if the set of possible answers has a limited number of values, where these values are unregulated, then this is a classification task; in classification tasks, the machine learning program must draw a conclusion based on the observed values and determine to which category the new observations belong.

However, it is worth noting that to solve certain classes of problems in the conditions of semi-structured data – traditional ML methods require considerable effort and thorough human expert experience/understanding in a specific subject area precisely to obtain a qualitative classification or regression model after iterative training.

For example, to recognize animal images using traditional supervised ML, you need to do the following:

- manually correctly label hundreds of thousands of animal images
- configure the controlled ML algorithm to build a classification model on these labeled images
- test the trained model on a set of unknown images
- to determine why some results of the obtained model are inaccurate.

– according to the results of the previous step – expand/increase the input data set by marking new images to improve the accuracy of the model.

That is why – in supervised learning, the accuracy of the results improves only when you have a wide and sufficiently diverse set of data. For example, an algorithm can accurately identify black cats but not white cats because the training dataset contains more images of black cats. In this case, you will need to label more images of white cats and train the machine-learning model again.

Weak machine learning or semi-supervised machine learning

Weak machine learning is similar to supervised machine learning but uses both labeled and partitioned data. Labeled data are, in fact, sets of information units with labels (tags, variables) assigned to them. Unmarked data has no such tags. By combining learning techniques, these algorithms can learn to label unlabeled Big New Data. That is, such machine learning combines both approaches described above. This is an excellent opportunity to solve predictive tasks on large sets of new data. This method also allows a significant increase in the accuracy of the forecast using Big Data from ICS.

Weak supervised learning, also called semi-supervised learning, is an ML paradigm that has grown in relevance and prominence with the advent of large language models due to the large amount of data required to train them. It is characterized by using a small amount of human-labeled data (exclusively for the more expensive and time-consuming supervised learning paradigm) and then a large amount of unlabeled data (used exclusively in the unsupervised learning paradigm). In other words, the desired output values are provided only for a small subset of the training data. The rest of the data is unmarked or inaccurately marked.

Technically, this can be thought of as first – performing clustering on new data and then labeling the clusters with old but labeled data, pushing the decision boundaries away from areas of high density.

Obtaining labeled data for a learning problem often requires the expertise of a skilled human expert or group of experts (e.g., to transcribe an audio segment, etc.) or to perform a series of physical experiments (e.g., determine the 3D structure of a protein or determine the presence of oil at a specific location, etc.). Thus, the cost associated with the labeling process can make large, fully labeled training sets infeasible, while obtaining unlabeled data is relatively inexpensive. In such situations, it is semi-supervised learning that can be of great practical importance.

Assumptions for semi-supervised learning: For any use of unlabeled data, there must be some relationship with the underlying distribution of the data. Semi-supervised learning algorithms use at least one of the following assumptions:

Assumption of continuity/smoothness. Points that are close to each other are more likely to have a common label. This is also commonly assumed in supervised learning and favors geometrically simple decision bounds. In the case of semi-supervised learning, the aforementioned smoothness assumption additionally gives the advantage of not constructing decision boundaries in partially labeled regions/areas with low density, so the training result can be valid when several points are close to each other but in different classes.

Clustering assumption: Data tend to form distinct clusters, and points in one cluster are likely to share a common label (although data that share a common label may span multiple clusters). This is a special case of the smoothness assumption and gives rise to feature learning using clustering algorithms.

Assumption of diversity: In reality, data typically have much less diversity than the theoretically possible space of values. In this case, studying diversity using both labeled and unlabeled data can avoid the curse of dimensionality. Training can then continue using the distances and densities defined on the particular dataset. The diversity assumption is also practically useful when multidimensional data are generated by some process that can be very difficult to simulate directly performing experiments, but this difficulty is overcome by the expert understanding that in a given set of input data when solving a specific analytical problem, the input dimensions have a limited number of degrees of freedom (category values). For example, the human voice is controlled by several vocal folds, and the image of different facial expressions is controlled by several muscles. In these cases, it is better to consider distances and smoothness in the natural space of the specific generating problem, rather than in the space of all possible acoustic waves or images, respectively.

Reinforcement machine learning

Reinforcement learning is not similar to any of the above types of machine learning tasks because it is used in tasks where there are neither conditioned labeled data sets nor unlabeled data sets.

In reinforcement learning, an algorithm (also often called an agent in this context) learns by trial and error, using feedback from its actions. Rewards and punishments act as signals for desirable and undesirable behavior. That is, agents/agents learn from past experiences and change their approach in response to a new situation, trying to achieve the best possible outcome.

In other words, in reinforcement learning, the algorithm chooses an action in response to each data point and based on the results of its previous predictions (the learning algorithm receives a success signal that lets it know how successful the past decision was). Based on this signal, the algorithm changes its strategy to achieve the highest reward.

That is, reinforcement learning is a type of ML in which software agents must take actions in some environment to maximize some notion of cumulative reward.

Reinforcement learning is widely used to implement non-generative AI, that is, this approach to ML is most common in robotics, the Internet of Things, and in the industry of computer games without a description of the rules of the game. Its use can be promising in the modeling of game theory problems.

The main difficulty in applying this approach to ML is that the real world is very difficult to model with the necessary accuracy and adequacy, as a result of which a trained AI can perfectly perform tasks in a virtual environment, but be practically useless in real complex, dynamic, stochastic conditions of BIG DATA and also under the influence of the «curse» of their dimensions.

Machine learning vs deep machine learning vs artificial intelligence

Consider the following definitions to understand deep learning and machine learning and artificial intelligence.

AI is:

- property of intelligent systems/algorithms to perform functions that are traditionally considered a human prerogative;
- the science and technology of creating intelligent machines, especially intelligent computer programs
- techniques and algorithms that allow solving complex predictive tasks on large data sets.

ML is a subset of artificial intelligence that uses techniques and algorithms that allow computers to take advantage of opportunities to improve tasks.

Deep learning is a subset of ML based on artificial neural networks. The learning process is deep because the structure of artificial neural networks consists of several input, output, and hidden layers. Each layer contains units that transform input data into information that the next layer can use for a specific predictive task. Through the structure of an artificial neural network, deep learning is designed to identify patterns in large volumes of unstructured, categorical data. For this reason, deep learning is now being actively implemented in the operational and tactical management of business processes in various industries (health care, energy, finance, transport). Table 1 provides a more detailed comparison of these two methods.

Configuring effective Machine learning in Big Data conditions

ML uses a wide range of algorithms to transform datasets into predictive models. Which algorithm will work better depends on the problem being solved.

There is such a thing as the «No Free Lunch» theorem.

The bottom line is that there is no single algorithm that is the best choice for a particular task/data set, especially tutored learning.

For example, it cannot be said that neural networks always perform better than decision trees, and vice versa. The effectiveness of algorithms is affected by many factors such as the size and structure of the data set. In addition, each ML algorithm has its own style of inductive shift.

For this reason, one has to try many different algorithms, check the performance of each on a test data set, and then choose the best one. Of course, you need to experiment and choose the best ML algorithm for your problem/dataset. For a particular problem, it may be appropriate to experiment with several ML algorithms and select the best algorithm and tune it accordingly. If we draw an analogy when cleaning the house, you will most likely use a vacuum cleaner, a broom, or a mop, but not a shovel.

The recommendations offered in this lecture are general rules. Some should be adapted to a specific situation, and some can be violated. Don't be afraid to use multiple ML algorithms simultaneously when processing and analyzing data.

Table 1

**COMPARISON OF MACHINE LEARNING
AND DEEP MACHINE LEARNING TECHNIQUES**

	Machine learning	Deep machine learning
Approach to learning	Divides the learning process into smaller steps. Then the results of each step are combined into one overall result.	Conducts a single end-to-end training process on the entire data set in turnkey mode.
The process of configuring features and functions	It is necessary that functions and features are precisely defined and/or created by users.	Automatic recognition of input features, automatic search for high-level features from data and generation of new features and/or features
Equipment requirements	Can work on low-end computers. Does not require a large amount of computing power.	Depends on high performance computers. It essentially performs a large number of matrix multiplication operations. The GPU can effectively optimize these operations.
Task completion time	It takes relatively little time to learn, from a few seconds to several hours.	Usually, the learning process takes a long time because the deep learning algorithm involves many layers and iterations.
Input and output data	This data must be structured, the output value is usually a numerical value such as a rating or classification.	Input and output data can have several formats, including unstructured: for example text, image or sound.
The volume of the input data set	Small amounts of data can be used to generate predictions, but the learning algorithm may require a priori criteria.	To create forecasts, it is necessary to use large volumes of training data.

Source: developed by the author

Choosing an effective and appropriate ML algorithm depends on many factors, including the size of the data, the number and type of attributes (features), the quality of the data, the presence and amount of anomalies in the data, and understanding what answers the business needs based on that data. Attention should also be paid to accuracy, training time, a priori parameters of the algorithm, representativeness of the sample, and many other things. Therefore, choosing the

appropriate algorithm is a complex expert decision based on business needs, specifications, experimental work, and analysis of available time.

Even the most experienced data scientists won't be able to tell you which algorithm will perform better without doing lots of experiments with different kinds of algorithms, their settings, and different data samples. Despite this, there are certain general methodological guidelines for machine learning algorithms that will help you find an effective algorithm suitable for your specific tasks and data.

Choosing an effective ML algorithm depends mainly on two different aspects of the data processing and analysis scenario:

- What should be done with the data? Specifically, what is the business question that needs to be answered, the type of problem of learning from past data (for example, is the problem categorical classification, predictive regression, cluster analysis, or anomaly detection?).

- What are the requirements for the data processing and analysis scenario? In particular, what is the accuracy, training time, linearity, number of parameters, and number of functions supported by the solution? That is, various accompanying requirements should be taken into account when choosing a machine learning algorithm. So, a list of the main groups of requirements for an effective machine learning scenario (and their details are given below in the text):

- Accuracy
- Study time
- Linearity
- Number of parameters
- Number and type of target functions

Accuracy in ML measures the effectiveness of the model in the form of the ratio of valid results to the total number of variants, that is, a set of industry metrics for accuracy assessment is calculated.

It is not always necessary to get the most accurate model possible. Sometimes an approximate answer is enough, depending on what the model is used for. If this is indeed the case – it is possible to significantly reduce the processing time by replacing with more approximate methods. Approximate methods also usually avoid excessive overtraining.

There are three scenarios for evaluating the accuracy of a trained model:

- creation of scores for training data to evaluate the model
- generating estimates for the model, but comparing these estimates to the values in the reserved test set

– comparing estimates for two different but related models using the same data set.

Study time. In supervised learning, learning refers to the use of historical data to create an error-free ML model. The number of minutes or hours required to train a model varies greatly for different algorithms. Learning time is often closely related to accuracy; one of them is usually an obstacle to the other.

Also, some algorithms are more sensitive to the number of data points than others. The analyst can choose a specific algorithm, as there are time constraints, especially if the data set is large.

Creating and using an ML model is usually done in three stages:

1. Configure the model by choosing a specific type of algorithm and then determining its optimal parameters.

2. Download a prepared dataset compatible with the selected algorithm. Connect the data and model to the model training module.

3. After training is complete, use the trained model with one of the estimation modules to make predictions based on new data.

Linearity in statistics and ML means that there is a linear relationship between a variable and a constant in a data set. For example, linear classification algorithms assume that classes can be separated by a straight line (or its approximation). Linearity is used in many classic ML algorithms. For example, linear regression algorithms assume that data trends follow a straight line. This assumption is only suitable for some problems, but reduces accuracy for others. Despite their shortcomings, linear algorithms are popular as a first strategy. They are usually algorithmically simple and quickly mastered. It should be remembered that using the linear regression method for data with a non-linear trend will lead to significantly more errors than necessary.

The number of ML algorithm parameters. Parameters are the controls that the analyst chooses when setting up the ML algorithm. These are certain settings that affect the behavior of the algorithm, for example: the permissible learning error or the number of iterations, options for the behavior of the algorithm in case of falling into a local extremum, etc. Training time and algorithm accuracy are often difficult to balance to obtain ideal algorithm settings. As a rule, algorithms with a large number of parameters take more time to experiment to find an effective combination of them.

It should be noted that in many applications for ML, a separate Model Setup Module is available. The purpose of this module is to select the optimal parameters for the machine learning algorithm. The module automatically creates and tests several models using different combinations of parameters. It compares metrics across all learning

outcomes to find a good combination. However, the time required to train the model increases exponentially with the number of parameters. However, having a large number of parameters usually means that the algorithm has a lot of flexibility. And then it is possible to achieve very good accuracy of the trained model, provided that it is possible to find the right combination of ML algorithm parameter settings.

Number and type of target variables/functions. In ML, the target variable is the variable to be analyzed. For some data types and industries, the number of features/target variables can be relatively large compared to the number of data points. This is often the case with genetic or textual data.

A large number of features can slow down some learning algorithms, making the learning time very long. Support vector computers are particularly well suited for problems with a large number of functions. For this reason, they have been used in many information analysis applications for text and image classification. Vector computers can also be used to speed up the solving of classification and regression problems for ordinary data sets.

The selection of target variables occurs in the process of performing statistical tests on the input data taking into account the given output data. The purpose of such tests is to determine which columns are more predictive of the original data. In many applications for ML – The filter-based component selection module provides several options for selecting target variables and includes correlation methods such as Pearson correlation, chi-square value, etc. It is also worth using the service for automatic calculation of a set of evaluations of the importance of features for a specific data set if the functionality is available.

Data extraction and cleaning (ETL) technology for machine learning. Directly in Big Data, initially «clean» data does not exist. To be useful for ML, data must be heavily prepared and transformed. For example, you can:

- Review the data and exclude all columns that have a lot of missing data.
- View the data again and select the columns to be used for forecasting. (If you repeat this operation several times, you can select different columns.)
- Exclude all rows in which there is no data in the last columns.
- Correct obvious errors and combine equivalent values. For example, GB and Great Britain can be lumped into one category.
- Exclude rows with data that are outside the required range.

There is much more that can be done, but it will depend on the particular Data Set. This can be time-consuming for an expert, but if

the data cleansing step in the machine learning pipeline is configured, it can be repeated using triggers later on if desired.

Coding and normalization of data for machine learning. To use categorical text data for machine classification, the text labels must be transformed into another form. There are two encoding modes.

The first is manual label numbering, each text label value is replaced by a number. The second is one-hot encoding, each value of the text label is converted into a column with a binary value (1 or 0). However, most machine learning platforms have features that do the coding themselves. Automatic one-hot coding is generally preferred, as manual label numbering can sometimes confuse the ML algorithm into thinking that the codes are ordered.

Also, to use numerical data for machine regression, the data usually needs to be normalized. This process is also often called scaling. Without normalization when calculating Euclidean distances, sets of numbers with large ranges may start to «dominate» unnecessarily, and «steepest» descent optimization may not converge. There are several ways to normalize data for ML, including minimax normalization, centering, standardization, and scaling to unit length.

Features in machine learning are individual measurable properties or characteristics of an observed phenomenon. The concept of «sign» is related to the concept of an independent variable, which is used in statistical methods such as linear regression. Feature vectors combine all features of one line into a numeric vector.

A major part of the art of effective feature selection is to select a minimal but sufficient set of independent variables that explain the task. If two variables are highly correlated, they should be combined into one or deleted. Some data transformations used to construct new features or reduce the dimensionality of their vectors are simple. For example, by subtracting the year of birth from the year of death, the age of death is obtained, which is the main independent variable for the analysis of mortality and life expectancy. In other cases, the preparation of signs may not be so obvious.

Hyperparameters for machine learning algorithms

The algorithms themselves have variables called hyperparameters. They are called hyperparameters because, unlike parameters, they control the operation of the algorithm, not defined weights.

The most important hyperparameter is often the learning rate, which determines the step size used when finding the next set of weights to optimize. If the learning rate is too high, the steep descent may quickly

plateau or lead to a suboptimal local extremum. If the learning rate is too slow, the descent may stall and never fully converge.

Many other common hyperparameters depend on the algorithms used. For example, most ML algorithms have a stopping hyperparameter, such as the maximum number of epochs, maximum execution time, or minimum epoch-to-epoch improvement. Certain algorithms have hyperparameters that control the shape of their search. For example, a random forest classifier has hyperparameters for minimum samples to generate a leaf/branch, maximum depth, minimum samples when splitting, minimum mass fraction for a leaf, etc. Some ML frameworks provide automatic tuning of hyperparameters. Essentially, you tell the system which hyperparameters to tune, and possibly which metric to tune, and the system looks at those hyperparameters for as many runs as it's allowed (for example, Google Cloud hyperparameter tuning pulls the relevant metric from the TensorFlow model, so you don't need to specify it.)

There are three approaches in automatic optimization of hyperparameters: Bayesian optimization, grid search, evolutionary optimization, and random search. As a rule, the most effective approach turns out to be Bayesian optimization.

It can be assumed that setting as many hyperparameters as possible will give a better result. However, this can be very expensive/time-consuming. With experience comes an understanding of which hyperparameters are most important for the data used in each specific case/problem and the algorithms selected.

The human factor in machine learning

Although data analysis and computational analysis may make us think that we are obtaining objective information, this is not the case; ML results are not always neutral. The shift in results introduced by the human factor affects not only the collection and systematization of data but also the algorithms themselves, which determine how ML will interact with this data.

Suppose there is a group of people collecting and labeling images for ML. She gets a task to collect images of flowers. If most people chose a rose for this image, the computer would not be able to classify a lily, sunflower, or violet as a flower.

Another example. Photographs of scientists were used as training data in ML. The computer received a set of images where the overwhelming majority of white men were represented. As a result, the machine could not properly classify people with different skin colors

and women. Recent studies have shown that AI and ML programs have inherited racial and gender biases.

As ML is increasingly used in business today, undetected biases can cause systemic problems: based on prejudice, programs can prevent a person from getting a loan, finding an ad for a high-paying job, or even ordering a delivery.

In ML, the human factor can negatively affect other people, and it is now extremely important to be aware of this and strive to eliminate it. To do this, different people should be involved in the development, testing, and analysis of ML projects. Some developers request model monitoring and auditing from objective regulatory third parties. Others are creating alternative systems that can detect ethical biases in algorithms. Fighting «superstitions» in the field of ML is very important. To do this, you need to constantly raise awareness of possible problems, remember your own unconscious biases, and take them into account when collecting, preparing, and structuring data.

Automated machine learning

There is only one way to find out which algorithm or ensemble of algorithms will produce the best model in each case – to try all algorithms and their combinations. But if you also try all possible variants of data normalization and variants of functions, there will be a combinatorial «explosion».

Trying to do everything by hand is impractical, so ML tool vendors have put a lot of effort into releasing AutoML-class systems (so-called automated ML). The best of them combine parameter setting, algorithm selection, and data normalization. Hyperparametric tuning of the best model (or models) is often performed at later stages of the project.

Consider, for example, automated feature extraction, which is part of the complete AutoML workflow that provides optimized models. The workflow includes three simple steps that automate feature selection, model selection, and hyperparameter tuning. New high-level methods of automatic feature extraction from signals have appeared. Autoencoders, wavelet scattering, and deep neural networks are commonly used for feature extraction and data dimensionality reduction. Wavelet diffusion networks automate the extraction of low-variance features from real-time series and image data. This approach creates a representation of the data that minimizes within-class variation while maintaining the ability to distinguish between classes. Wavelet scattering works well when you don't have a lot of data.

Hence, autoML is used to reduce human interaction and automate all tasks to solve real-world problems. This functionality includes the entire process from raw data to the final ML model. The goal of AutoML is to offer extensive learning techniques and models for non-ML experts. Remember, even though AutoML doesn't require human interaction, that doesn't mean it's completely superior to it.

Thus, the ML algorithms themselves are only part of the mosaic of machine learning. In addition to algorithm selection (manual or automatic), you will need to deal with optimizers, data cleaning, feature selection, normalization, and algorithm hyperparameter tuning.

Once you've completed the above procedures and then built a quality model that will work with your data, it's time to use the model and then update it as conditions change. However, the management of machine learning models is a modern scientific and practical problem, and the results of its solution will be presented in future publications.

2. Deep machine learning parameters with the use of Big Data information control systems

In its nascent years, AI largely focused on expert systems based on knowledge in the form of production rules that solved mainly diagnostic problems, but also design-type problems (using previously manually collected and formalized knowledge/experience of human experts in a specific subject area). However, this type of intelligent systems had numerous shortcomings, including the subjectivity of experts' opinions, which ultimately led to their loss of mass popularity [6–10].

As the scale and volume of data increased, these methods were replaced by a more controlled and objective data-oriented approach – ML. ML is a set of algorithms and tools that help machines understand hidden patterns in data and use the structure and essence of these hidden patterns in the data/heuristics to perform logical inference/prediction for a specific task. Currently, there is a diverse range of such methods/algorithms by which machines seek to understand these basic patterns such as association, sequence, classification, clustering, regression prediction, and anomaly detection in data.

If we systematically consider the history of the development of computational analytics and analyze its perspective, deep learning is a further evolution and subdomain of machine learning. Thanks to the emergence of architectures with increased computing power and large sets of semi-structured and unstructured data, specialized architectures and corresponding deep learning algorithms can independently study

hidden patterns in BIG DATA and even perform generative functionality.

A key technology in deep ML is artificial neural networks. Such connectionist architectures in cybernetics have existed for more than 70 years, but new cloud hardware architectures, specialized processors (GPU and TPU), and the rapid distribution and accumulation of BIG DATA have once again brought them to the forefront of artificial intelligence and information control systems in particular.

However, recently there has been a growing misconception that deep learning is a competing technology to classical ML. Deep learning is not the only possible approach, but rather a specialized separate class of algorithms and topologies that should be applied to a certain range of scientific problems and practical tasks, in particular in the field of information control systems.

In this subsection of the monograph, the results of research and a comparative analysis are given not only in relation to the above hypothesis but also the results of thorough research on the advantages, problems, and features of effective deep ML in the conditions of BIG DATA information control systems are presented.

Deep learning is a machine learning technique that:

- automatically extracts features (Feature extraction automatically identifies/selects the most discriminating characteristics of signals, which are more easily exploited by machine learning or a deep learning algorithm. Feature extraction is a very difficult to automate and time-consuming process of converting raw data into numerical features (for example, recognizing speech recordings) that can be processed while retaining the information in the original dataset.)

- learns hidden patterns directly from big data.

Such Big Data for deep ML can be not only texts, speech recordings or other signals that are tied to time, place, and frequency (for example, sound, time series, signals from various microphones and other sensors).

That is why deep learning is often also called end-to-end learning.

Although deep ML was first formulated in the last century, there are the following main reasons why the use of deep learning has grown dramatically in recent years:

First, deep learning methods are now more accurate and more objective than human experts, particularly in information control systems.

Second, deep learning requires significant computing power. That is why the emergence of high-performance GPUs and TPUs with

parallel architectures are effective for deep learning. When combined with clusters or cloud computing, this allows development teams to reduce training time for a deep learning network from weeks to hours or less.

Third, in the past few years, large amounts of labeled data have become available for deep learning.

fourthly, the booming interest and, therefore, investments from the corporate and state scientific communities also contributed to the advancement in this area and the effective continuation of relevant research.

The vast majority of deep learning methods use architectures and learning algorithms of conventional artificial neural networks. Artificial neural networks attempt to mimic the human brain using a combination of inputs, weights, and biases. These elements work together to accurately recognize, classify, and describe objects in data.

Artificial neural networks consist of multiple layers of interconnected nodes, each layer building on the previous layer to refine and optimize prediction or categorization. This flow of computation through the network is called forward propagation. The input and output layers of a deep neural network are called visible layers. The input layer is where the learning model receives data to process, and the output layer is where the final prediction or classification is made. Each neuron in the hidden layer is connected to many others. Each synapse has a weight property that controls how much activation of that neuron affects others attached to it.

Another process, called backpropagation, uses algorithms such as gradient descent to calculate the errors in the predictions, then adjusts the weights and biases of the feature by moving back through the layers to train the model. Together, forward and back propagation allow the neural network to make predictions and correct any errors accordingly. Over time, the algorithm becomes more accurate.

Thus, in the learning process, the ANN performs iterative nonlinear transformations of the input data and uses the obtained results to create a black box model, and the learning iteration continues for many epochs with different experimental parameters until the result reaches an acceptable level of accuracy.

Considering the common features of shallow neural net and deep neural net described above, it is worth noting that deep learning is an evolution of classical machine learning. Although a conventional shallow artificial neural network with one or more layers can still make approximate predictions, an additional increase in the number of internal layers in a DNN can greatly help to optimize and improve its

performance, even on unstructured big data. The term «deep» usually refers to the number of hidden layers of such a neural network. Traditional shallow neural networks contain only 2–3 hidden layers, while deep networks can have up to 150. Therefore, the term «deep» in deep learning is associated with these deep hidden layers, and this is where its effectiveness comes from. The choice of the number of hidden layers depends on the nature and complexity of the problem and the size and characteristics of the data set.

Thus, deep learning programs have many layers of interconnected layers of artificial neurons, with each layer building on the previous one to refine and optimize regression or classification predictions. This is similar to the way our human brain works to solve problems – it runs queries through various hierarchies of concepts and related questions to find an answer. It is the number of layers of data processing in the hidden layers through which the input data must pass that gave rise to the term «deep».

The simplest type of shallow neural network was briefly analyzed above. However, both deep learning algorithms are incredibly complex, and deep neural network architectures are very specific and variable for solving a range of very different problems. Therefore, it is appropriate to highlight six types of DNN architectures that have gained popularity and proven their effectiveness over the past 20 years. Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are the two oldest approaches on this list, but also the two most used in a variety of applications.

Convolutional Neural Networks (CNNs), used primarily in computer vision and image classification applications, can detect features and patterns in an image, enabling tasks such as object detection or recognition.

Recurrent Neural Networks (RNNs) are commonly used in natural language and speech recognition applications because they use sequential or time series data.

The current research classifies deep learning architectures into supervised and unsupervised learning and introduces several popular deep learning architectures: Convolutional Neural Networks, Recurrent Neural Networks (RNN), Long Short-Term Memory/Supervised Recurrent Unit (GRU), Self-Organizing Map (SOM), autoencoders (AE) and restricted Boltzmann machine (RBM), deep belief networks (DBN) and deep stacking networks (DSN).

Machine and deep learning models are capable of different types of learning, commonly classified as supervised learning, unsupervised learning, and reinforcement learning. As outlined above, both classical

machine learning and deep learning technologies use data to train, but the key difference is what data they use, how they process that data, and how they learn from it using which architecture configuration.

The fundamental questions of the theory of classical machine learning were revealed in their works by such scientists as Nils J. Nilsson, Trevor Hastie & Robert Tibshirani & Jerome H. Friedman, Pedro Domingos, Ian H. Witten & Eibe Frank, Ethem Alpaydin, David J.C. MacKay, Richard O. Duda and Peter E. Hart, Christopher Bishop, Stuart Russell & Peter Norvig, Ray Solomonoff, Kevin P. Murphy.

The main principles of the theory of deep machine learning were revealed in their works by such scientists as: Schulz Hanne & Behnke Sven, LeCun Yann & Bengio Yoshua & Hinton Geoffrey, Marblestone Adam H. & Wayne Greg & Kording, Konrad P., Deng L. & Yu, D., Zhang W.J. & Yang G. & Ji C. & Gupta, M.M., Bengio Yoshua, Bengio Y. & Courville A. & Vincent, P. LeCun Yann & Bengio Yoshua & Hinton, Geoffrey, Schmidhuber J., Hinton G.E. and other.

But the urgent question of an effective choice between machine and deep machine learning in information control systems, and the question of analyzing the advantages, problems, and features of deep machine learning on big data information control systems, remained unresolved.

Considering the above, the actual and initial goal of the current of this part of complex interdisciplinary collective scientific research was not only to propose recommendations for the choice between machine and deep ML in information control systems but also to propose methodological and practical recommendations regarding the advantages, problems and features of deep ML in scientific research and practical projects/tasks of information control systems.

Below are the results of the study of important differences between classic machine learning and deep learning technologies:

1. ML involves using thousands of data points, while deep learning uses millions of data points. Machine learning algorithms typically work well with relatively small sets of labeled and structured data. Machine learning algorithms are also better when data is scarce. Conversely, deep learning needs really very large amounts of unlabeled or partially labeled data that have semi-structured or unstructured data format to understand the data and perform better than traditional machine learning algorithms.

2. ML algorithms solve problems using explicit programming (algorithms for building decision trees and forests, finding association rules, finding regression equation coefficients, kNN, SVM and Naïve Baies classifiers, hierarchical and non-hierarchical cluster analysis, etc.). That is, classical machine learning offers a wide variety of

methods and algorithms that can be chosen depending on the specifics of the subject area/industry, your specific task, the nature of the structured data being processed, and the type of problem to be solved. However, deep learning algorithms solve all problems after a complex process of learning hidden deep layers of neural networks of special architectures and their further use in black box mode.

3. ML algorithms take relatively less time to train, from a few seconds to a few hours. On the other hand, deep learning algorithms take a long time to train an artificial neural network, from a few days to many months.

4. Classical ML provides the possibility of using very different methods/algorithms (algorithms for building decision trees and forests, finding association rules, finding coefficients of regression equations, kNN, SVM and Naïve Baies classifiers, hierarchical and non-hierarchical cluster analysis, etc.); while deep machine learning involves the use of exclusively artificial neural networks of special architectures.

5. Classical ML requires an interactive process of interaction with the analyst at all stages of the project, unlike deep machine learning. Deep learning networks do not require human intervention, as multi-level layers in neural networks contain data in a hierarchy of different concepts that eventually learn from their own mistakes, i.e. are capable of self-learning over time. However, even these can be misleading if the data quality is not good enough.

6. In classical machine learning and in deep machine learning, the quality and representativeness of the input data significantly affect the quality of such learning, however, given the minimal intervention of expert analysts in the auto ML mode of deep learning, the risks for it due to low-quality and incomplete data are much greater.

7. Classic ML with time and/or the appearance of significant changes (either systemic or force majeure) in the studied subject area requires re-training of the model – but exclusively on new sets of collected data, in which such changes should be reflected.

8. Since machine learning algorithms can learn only according to pre-programmed set criteria and require labeled data, they are not suitable for solving innovative complex scientific queries in free search mode, which involves a significant amount of unlabeled semi-structured and unstructured data (for example, deciphering and understanding completely unknown In other words, cases, where deep learning excels, include situations where there is a large amount of unlabeled data, a lack of domain understanding for feature discovery/extraction, or more complex problems such as speech recognition and NLP.

9. Classical machine learning does not require the same expensive high-quality computing machines and high-performance processors (GPU and TPU) as deep learning.

10. After all, many data scientists choose traditional ML over deep ML because of its superior interpretability or ability to understand solutions. On the contrary, deep learning models are complex, they work like a «black box», that is, it is difficult to explain the course of the decision and prove the result.

11. Deep learning is a subset of machine learning that is distinguished by the way it solves problems. Most ML application functions require an expert in the field to define. Deep learning, on the other hand, understands hidden functions incrementally, in an offline mode, thus eliminating the need for domain expertise and the systematic intervention of a human expert analyst. Because of this, deep learning algorithms take much longer to train than machine learning algorithms, which only take a few seconds to a few hours. However, the opposite happens during testing. Deep learning algorithms take much less time to run tests than machine learning algorithms, whose testing time increases with the size of the data.

12. Retraining for classical machine learning may turn out to be a more pressing problem on modern data sets, given the automatic ability of shallow hidden layers of a deep neural network to level the impact of noise and anomalies in big data and the ability of subsequent, deeper hidden layers to generalize to really big data.

The benefits of deep learning include the following:

1. Automatic learning/extraction of features/features from big data.

Deep learning systems can perform feature extraction from data automatically, meaning they do not need supervision to add new features. Deep learning algorithms can save time by not requiring humans to manually extract features from raw data. Deep learning will perform in auto ML mode most of the previous interactive «manual» data processing that is usually associated with machine learning. These algorithms can accept and process unstructured data, such as text and images, and automate feature extraction, removing some reliance on human experts.

2. Discovery/search for hidden regularities/patterns in big data.

Deep learning systems can intelligently and deeply analyze large volumes of data and detect complex patterns in images, text, and audio recordings of conversations, other signals, in order to extract hidden, non-trivial, and useful regularities/patterns.

3. Ability to effectively process and analyze highly variable/unstable/variable features in big data.

Deep learning systems can classify and sort data sets that have very large variations. Volatile/volatile data sets have large variations and/or noise.

4. Automatic marking/marketing of new data.

Deep learning requires labeled/labeled data for training. After training, a DNN can independently label new data and identify different types, categories, or values of new chunks of data.

5. Effective processing of unstructured data.

Classical ML techniques find unstructured data, such as conversation recordings, difficult for automated processing and analysis because the training dataset can have infinite variations. On the other hand, deep learning models can take in unstructured data and make general observations without manually extracting features. For example, a neural network can recognize that these two different input sentences have the same meaning:

- Tell me how to make the payment.
- How should I transfer the money?

This does not necessarily mean that classical ML does not use unstructured data; it simply means that in this case, unstructured data requires a rather long, complex, and time-consuming procedure of pre-processing the input unstructured data for the purpose of manual extraction of features, their quantification, normalization, and organization into a structured format (which is acceptable for classical ML methods).

In turn, it is worth noting that simple architectures of deep neural networks are also able to process structured data if there is an appropriate amount of it.

6. Accuracy.

Adding additional hidden layers to DNNs helps optimize deep learning models to provide greater accuracy.

7. Faster debugging of AutoML mode compared to classical machine learning.

Compared to conventional machine learning processes, deep learning requires much less human intervention and manual experimentation and can automatically analyze types and volumes of data that classical ML cannot handle and analyze.

8. More efficient and functional unsupervised machine learning.

Deep learning models can learn and improve over time based on user behavior and their own experience. They do not require large variations of labeled datasets. For example, consider a neural network

that automatically corrects or suggests words by analyzing your typing behavior. Suppose that a DNN has been trained on the English language and can check the spelling of English words. However, if the user often types/speaks non-English words like *danke*, a trained (according to your behavior and characteristics) deep neural network will automatically identify such common typos/spellings and errors and automatically and correctly correct them.

While deep learning has gained popularity over the past few years, it comes with its own set of challenges that the community is working hard to solve.

Deep learning projects and cases may also have systemic flaws and certain limitations, in particular:

The problem of biases and biases in input data for training and its verification. If a model is trained on insufficiently large and representative data that is obtained from different and independent sources, the data may contain biases and biases, and the trained model will reproduce these biases in its subsequent predictions. This is especially a threat to convolutional DNNs with an unreasonably large number of hidden layers – which can often be prone to overgeneralization.

Optimal configuration of DML hyperparameters, in particular learning speed. Choosing the optimal learning rate also becomes a serious challenge for deep learning models. If the speed is too high, the model converges too quickly, producing a not-always optimal and accurate model. If the learning rate is too low, the learning process may take too much time (and therefore exceed the operating budget). In addition, a suboptimal learning rate can lead the model to fall into local or sub-global optima. Another problem with slow learning rates is the threat of overfitting. However, even in the case of a deep learning model transfer strategy (which is currently a popular practice and seems to be able to solve the above threats), retraining a pre-trained deep neural network on new data will still require much more time and computing power than in the case of using classical ML.

High requirements for the global financial and operational parameters of the Data Center and its equipment. Multi-core high-performance specialized processors (GPU and TPU) and other related high-performance hardware are necessary to ensure the comprehensive cost-effectiveness of the DML project. However, this hardware not only requires significant investment but is currently in short supply. Complex conditions of a data processing center with such equipment should also take into account the main item of the operating cost of such projects – the price of electricity.

The need for large volumes of large-dimensional data, including and verified marked. To obtain high-quality DNNs in real conditions of big data, there is a need for a large amount of prepared data of the maximum dimension. And to solve regression forecasting problems in such conditions, it is even more difficult to satisfy the need for a large number of prepared and checked correctly/correctly marked/marked data of the maximum dimension. Obtaining which will be an expensive, difficult task that requires considerable time and labor costs.

The need for large amounts of high-quality data. Deep learning algorithms produce better results if they are trained on large volumes of high-quality data. Outliers, errors, lack of representativeness and adequacy of input big data, and other errors in big data – as a rule, will significantly affect the deep learning process. To avoid such problems, in the case of deep learning, data engineering, namely ETL/ELT, is very important and responsible. Such pre-processing of input big data for deep machine learning requires a lot of time, labor, and computing power.

Lack of multitasking/versatility of trained DNN in real conditions of its use over time. In order to obtain quality results from DML – machine learning problem statement, the relevant data for training and testing, as a rule, should be as specific/targeted/aimed at solving one specific narrow problem/problem as possible. This can lead to the loss of even the limited/conditional universality of the obtained model – and, therefore, to further limitations in the mode of using the trained DNN in real, current conditions (which will contain/reflect objective needs due to even a slight change in the macro environment of the organization/project/company with the passage of time).

Lack of adequacy of trained DNN with real new actual data. To obtain quality results DML – machine learning problem statement, suitable big data for training and testing must be cleaned and prepared. This can lead to the loss of even the limited/conditional universality/adequacy of the model trained on such data – and therefore to further errors in the mode of using the trained DNN on new, real, and relevant data, especially in the conditions of big data (which naturally will definitely contain anomalies and information noise that were filtered out in the ETL/ELT step before training the DNN).

Lack of transparency of deep machine learning results. While the algorithm looks at millions of data points to find hidden patterns during this training, it will likely be very difficult to understand how the neural network arrives at its decision. The lack of transparency in how DNNs process data in both training mode and usage mode makes it much more difficult for human experts to detect unwanted biases and explain predictions. Despite the fact that deep learning algorithms surpass the

accuracy of human thinking, there is no clear and clear way to go back (conduct reverse reasoning – deduction) and justifications to explain how exactly, step by step, each made prediction was obtained on the basis of a new portion of input data. This makes it difficult to use DML in some types of scientific research, where the obtained result must be clearly explained and justified.

Behavioral fairness of reinforcement learning agents. Another aspect that tends to be problematic when using reinforcement learning models is the presence of meaningful/behavioral/systematic biases/distortions in the input data itself, which can lead to poor/suboptimal performance of the model trained on them due to its suboptimal learning trajectory. Software agents learning using a reward-based mechanism sometimes stop behaving ethically/fairly/constructively because all they need to do to minimize the aggregate error is to maximize the reward they receive. In the literature, such an example is mentioned, when the software agent simply stopped playing the game and got into an endless cycle of collecting prize points. While this may be partially acceptable in a game scenario, wrong or unethical decisions can have profound negative consequences in the real world. Thus, it is relevant to have additional control of training results with reinforcement for obtaining balanced/constructive/ethical results. To this end, there are specialized open-source toolkits for detecting, investigating, and eliminating such systematic errors in deep learning algorithms and data. It is important for deep learning researchers and users to keep these potential issues in mind when planning and conducting reinforcement learning deep learning experiments.

Running deep learning algorithms on cloud infrastructure can overcome many DML problems. It is effective to use deep learning in the cloud to more effectively design, develop and train deep learning applications. Cloud computing services help make deep learning more accessible by making it easier to manage large data sets and train algorithms on distributed hardware.

Cloud services contribute to effective deep learning in the following aspects:

- provide access to large-scale computing power on demand, enabling model training to be distributed across multiple machines.
- provide access to special hardware configurations, including GPUs, FPGAs, and massively parallel high-performance computing (HPC) systems.
- do not require significant initial investments — because, in the conditions of using cloud technologies, analysts get access to modern equipment without buying it.

- help manage deep learning workflows — cloud services provide advanced features and templates for managing data sets and algorithms, training models, and their efficient deployment in production.

- increase the speed of DML. Experts and analysts can train deep learning models faster by using clusters of GPUs and CPUs to perform the complex mathematical operations required for your neural networks. Analysts can then deploy these models to process large volumes of data and produce increasingly relevant results.

- improve the accuracy of DML. High accuracy: Deep learning algorithms in the cloud computing mode can achieve state-of-the-art performance in various tasks in the field of information control systems, in particular, such as semi-structured text recognition and natural language signal processing.

- facilitate DML scalability.

With a wide range of on-demand resources available through the cloud, professionals gain access to virtually unlimited hardware resources to work with deep learning models of any size. Configurable deep neural network architectures can take advantage of multiple processors to smoothly and efficiently distribute workloads among different types and numbers of processors.

**The main areas of application of deep learning
in the field of information control systems can be divided
into three areas: computer vision, natural language processing, /
and reinforcement learning**

In computer vision, deep learning models can enable machines to identify and understand visual data. Some of the major applications of deep learning in computer vision include:

- Object detection and recognition: A deep learning model can be used to identify and locate objects in images and videos, enabling machines to perform tasks such as self-driving cars, surveillance, and robotics.

- Image classification: Deep learning models can be used to classify images into categories such as animals, plants, and buildings. This is used in applications such as medical imaging, quality control, and image retrieval.

- Image segmentation. Deep learning models can be used to segment images into different regions, enabling the identification of specific features in images.

In NLP, deep learning can enable machines to understand and generate human language. Some of the main applications of deep learning in machine linguistics include:

- Automatic Text Generation – A deep learning model can learn a corpus of text and new text like resumes, and essays can be automatically generated using these trained models.

- Language translation: deep learning models can translate text from one language to another, making it possible to communicate with people from different linguistic backgrounds.

- Sentiment analysis: Deep learning models can analyze the sentiment of a piece of text, allowing it to determine whether the text is positive, negative, or neutral. This is used in applications such as customer service, social media monitoring, and political analysis.

- Speech recognition: Deep learning models can recognize and transcribe spoken words, enabling tasks such as speech-to-text, voice search, and voice-controlled devices.

In reinforcement learning, deep learning works by teaching agents to act in an environment to maximize reward. Some of the main applications of deep learning in reinforcement learning include:

- Games: Reinforcement deep learning models have been able to beat human experts in games such as Go, Chess, and Atari.

- Robotics: Deep reinforcement learning models can be used to train robots to perform complex tasks such as object grasping, navigation, and manipulation.

- Control systems: Deep reinforcement learning models can be used to control complex systems such as power grids, traffic management, and supply chain optimization.

Conclusions

For years, all individuals and legal entities have generated, registered and stored (in databases, repositories and data showcases) huge volumes of heterogeneous structured and semi-structured information (quantitative, qualitative, text, hypertext, transactional, geo-information, multimedia, meta-information, etc.) regarding all aspects of their business, technological and management activity. In addition, the rapid development and spread of Big Data, Web 4.0 and Web 5.0, IOT, augmented reality, blockchain in recent years caused an additional avalanche-like increase in stored data.

This big data contains a significant potential for finding and formalizing hidden new regularities (knowledge, patterns), which are the basis for making optimal and effective managerial and technological decisions, including in the field of marketing.

From these regularities and patterns (after their verification and interpretation), the corporate knowledge base is mainly formed,

including and for expert systems and classic artificial intelligence systems, which should become an integral part of the innovative marketing management system of any company or enterprise.

In the modern conditions of the development of the global economy, and in connection with the emergence of new branches of economic activity in the field of informatization, the application of innovative technologies for the preparation, processing, analysis, and analytics of extremely large arrays of heterogeneous data leads to the obtaining of additional competitive advantages by users at the state, regional, branch and corporate levels of management, which is especially relevant in times of crisis [11–15].

Summing up the following interim summary of this work, it is worth noting that there are dozens of classical machine learning algorithms, varying in complexity from linear regression to deep neural networks. Until recently, these classical approaches using classical machine learning allowed us to achieve acceptable accuracy and quality, but in the context of structured ICST data.

The influence of big data ICST – has urgently actualized for classical methods of machine learning the need for their combination and their joint sequential/parallel use – in a hybrid-scenario mode, including using assembly technologies.

But even an experienced ICST expert will not say which algorithm (or their ensemble) in which configuration of parameters and variant of hybridization will work most effectively in a specific subject area for a specific task, without a complex of additional studies, a series of tests and tests.

To summarize from Chapter 2, deep learning is a subfield of machine learning that involves the use of deep neural networks to model and solve complex problems. Deep learning has made significant strides in a variety of fields, and its use is expected to continue to grow as more data and more powerful computing resources become available.

However, the emergence of new challenges associated with large semi-structured and unstructured data in information control systems does not allow classical methods/algorithms of machine learning to effectively and efficiently respond to the above-mentioned challenges.

In summary, we can say that it is the hybrid use of deep machine learning (which involves the use of deep neural networks of various specialized architectures) and classical machine learning methods in various ensemble modes (which will allow not only to solve modern problems of machine linguistics productively and quickly but also correctly) is appropriate for responding to the above-mentioned

challenges and problems of big data analytics of information control systems [16–19].

All of the above results are becoming more relevant due to the already existing specialization and recent successes of hardware manufacturers for fast and more affordable deep machine learning (including in experimental modes to determine optimal machine learning hyperparameters and optimal machine learning configurations).

References

1. Krasnyuk, M., & Krasniuk, S. (2020). Comparative characteristics of machine learning for predictive financial modeling. *Scientific Bulletin ΛΟΓΟΣ*, 55–57. <https://doi.org/10.36074/26.06.2020.v1.21>
2. Krasnyuk, M., Tkalenko, A., & Krasniuk, S. (2021). Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. *Scientific Bulletin ΛΟΓΟΣ*. <https://doi.org/10.36074/logos-09.04.2021.v1.07>
3. Krasnyuk, M., & Krasniuk, S. (2021). Modern practice of machine learning in the aviation transport industry. *Scientific Bulletin ΛΟΓΟΣ*. <https://doi.org/10.36074/logos-30.04.2021.v1.63>
4. Krasnyuk, M., & Krasniuk, S. (2020). Application of artificial neural networks for reducing dimensions of geological-geophysical data sets for the identification of prospective oil and gas deposits. *Scientific Bulletin ΛΟΓΟΣ*, 18–19. <https://doi.org/10.36074/24.04.2020.v2.05>
5. Krasnyuk, M., Krasniuk, S., Goncharenko, S., Roienko, L., Denysenko, V., & Natalia, L. (2023). Features, problems, and prospects of the application of deep machine learning in linguistics. *Bulletin of Science and Education*, 11(17), 19–34. <http://perspectives.pp.ua/index.php/vno/article/view/7746/7791>
6. Sytnyk, V.F., & Krasniuk, M.T. (2002). Polityka upravlinia znanniamy naftohazovoi kompanii yak kliuchovyi faktor pidvyshchennia yii efektyvnosti [Oil and gas company's knowledge management policy as a key factor in increasing its efficiency]. In *Problemy formuvannia rynkovoi ekonomiky – Problems of the formation of a market economy*, 10. Kyiv National Economic University (KNEU) [in Ukrainian].
7. Tuhaienko, V., & Krasniuk, S. (2022). Effective application of knowledge management in current crisis conditions. *Grail of Science*, 16, 348–358.
8. Krasnyuk, M.T. (2006). Problemy zastosuvannia system upravlinnia korporatyvnymy znanniamy ta yikh taksonomiia [Problems of applying corporate knowledge management systems and their taxonomy]. *Modeliuvannia ta Informatsiini Systemy v Ekonomitsi [Modeling and Information Systems in the Economy]*, 73, 256 [in Ukrainian].

9. Kulynych, Y., Krasnyuk, M., & Krasniuk, S. (2022). Knowledge discovery and data mining of structured and unstructured business data: Problems and prospects of implementation and adaptation in crisis conditions. *Grail of Science*, 12(13), 63–70. <https://doi.org/10.36074/grail-of-science.29.04.2022.006>

10. Tuhaienko, V., & Krasniuk, S. (2022). Effective application of knowledge management in current crisis conditions. *Grail of Science*, 1(16), 348–358.

11. Krasnyuk, M.T., Hrashchenko, I.S., Kustarovskiy, O.D., & Krasniuk, S.O. (2018). Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. *Economies' Horizons*, 3(6), 121–136.

12. Krasnyuk, M., & Krasniuk, I. (2024). Big data analysis and analytics for marketing and retail. *Zbirnyk tez Mizhnarodnoi naukovoï konferentsii «Shtuchnyi intelekt u nautsi ta osviti» (AISE)*, 3, 1–2.

13. Krasnyuk, M.T., Hrashchenko, I.S., Kustarovskiy, O.D., & Krasniuk, S.O. (2018). Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. *Economies' Horizons*, 3(6), 121–136.

14. Krasnyuk, M., Nevmerzhytska, S., & Tsalko, T. (2024). Processing, analysis & analytics of big data for innovative management. *Grail of Science*, 5(38), 75–83. <https://www.journal-grail.science/issue38.pdf>

15. Krasnyuk, M., & Elishys, D. (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. *Science and Technology Today*, 4(32), 833–857.

16. Krasnyuk, M., Goncharenko, S., & Krasniuk, S. (2022). Intelktualni tekhnolohii v hibrýdnyi korporatyvnyi SPPR (na prykladi Ukrainskoi naftohazovydobuvnoi kompanii) [Intelligent technologies in hybrid corporate DSS (on the example of Ukraine oil & gas production company)]. In O.L. Haltsovoi, *Innovation and investment mechanism for ensuring the country's competitiveness: Collective monograph* (pp. 194–211). League-Press. [in Ukrainian].

17. Krasnyuk, M., Hrashchenko, I., Goncharenko, S., & Krasniuk, S. (2022). Hybrid application of decision trees, fuzzy logic, and production rules for supporting investment decision-making (on the example of an oil and gas producing company). *Access to Science, Business, Innovation in the Digital Economy*, ACCESS Press, 3(3), 278–291. DOI: [https://doi.org/10.46656/access.2022.3.3\(7\)](https://doi.org/10.46656/access.2022.3.3(7))

18. Hrashchenko, I.S., Krasniuk, M.T., & Krasniuk, S.O. (2019). Hibrýdno-stsenarne zastosuvannia intelektualnykh, oriientovanykh na znannia tekhnolohii, yak vazhlyvyi antykrýzovyi instrument lohistychnykh kompanii v Ukraini [Hybrid-scenario application of intellectual, knowledge-oriented technologies as an important anti-crisis tool of logistics companies in Ukraine]. *Vcheni zapysky Tavriiskoho Natsionalnoho Universytetu imeni V.I. Vernadskoho. Serii: Ekonomika i upravlinnia – Scientific notes of Tavri National University named after V.I. Vernadskyi. Series: Economics and Management*, 30(69), 121–129 [in Ukrainian].

19. Krasnyuk, M. (2014). Hibrydyzatsiia intelektualnykh metodiv analizu biznesovykh danykh (rezhym vyivlennia anomalii) yak skladovyi instrument korporatyvnoho audytu [Hybridization of intelligent methods of business data analysis (anomaly detection mode) as a standard tool of corporate audit]. *Stan i perspektyvy rozvytku oblikovo-informatsiinoi systemy v Ukraini – Stan i perspektyvy rozvytku oblikovo-informatsiinoi systemy v Ukraini: Materialy III Mizhnar. Nauk.-Prakt. Konf. [m. Ternopil, 10–11 zhovt. 2014 r.] – The state and prospects of the development of the accounting and information system in Ukraine: materials of the III International science and practice conf. [m. Ternopil, October 10–11. 2014].* TNEU, 211–212. [in Ukrainian].