Organization of the Economic and Business Activities Results of Multilevel Production Systems in the Context of a Digitized Economy

Dmytro Solokha[1] Olena Tanchyk[2] Oksana Bieliakova[3,*]

[1,2] *Donetsk State University of Management, Mariupol 87513, Ukraine*

[3] *Azov maritime institute National University "Odesa Maritime Academy", Mariupol, 87517, Ukraine*

* Corresponding author. E-mail: tttt66t@ukr.net

ABSTRACT

The article presents the main results of scientific research of representatives of three scientific schools in Ukraine, in particular Zaporizhzhia National University, Donetsk State University of Management, Azov maritime institute National University "Odesa Maritime Academy", which allowed to study the processes taking place in the digitalization of Ukraine's economic and business system from different perspectives.

The authors have conducted a retrospective analysis of the formation of conceptual and categorical tools for the data collection digitization, processing, and storage of databases, analyzing the results, and diagnosing prospects for multilevel production systems' economic and business activity in the long run.

Keywords: Analysis, information, digital economy, development, market, enterprise, diagnostics.

1. INTRODUCTION

Ukraine's chosen course towards European integration and entry into the globalized world market of goods and services requires the government, officials, and producers to create conditions for doing business and organizing export-import activities that are adequate to global requirements.

Integration processes that have covered almost all spheres of economic activities of domestic enterprises and sectors of the national economy will require fundamentally new approaches, measures, and mechanisms for effective management of multilevel production systems' economic and business activities, which is impossible without the use of modern information technologies.

The issue is becoming increasingly acute in Ukraine's national economic system's digitalization and the entire globalizing economic space.

Ensuring domestic enterprises' effective functioning is possible with the help of predictive models that combine the market's interests, its marketing tools development in the long term, which meet the digital economy's requirements.

In the era of information technology dominance, mass computerization, and digitalization, raise new demanding requirements for users asking them to work with data and databases [1].

Understanding the importance of these processes and the inevitability of their implementation (in fact, it is a requirement of time), the Government of Ukraine has established the Ministry of Digital Transformation that emphasizes the importance of the processes under study.

In the market segment, where data on the results of multilevel production systems' economic and business activities are not yet entirely generated, the question arises of how to implement it if producers want to remain competitive market participants.

Ensuring effective management of economic activity of multilevel production systems, in the long run, using analytical data of the achieved results is possible only based on their objective analysis, which in turn is impossible without obtaining objective and reliable information.

Therefore, the chosen topic of this scientific article is quite timely and relevant. The study's urgency is emphasized by the high interest of modern researchers [1-3] in these issues.

The study presented in this article results from scientists' fruitful cooperation from several scientific schools of Ukraine, including Zaporizhzhia National University, Donetsk State University of Management, Azov Maritime Institute National University "Odesa Maritime Academy".

The purpose of this study is to research the conceptual approaches to the organization of the analysis of the result of the economic and business activity of multilevel production systems in the context of a digitized economy in Ukraine.

2. MATERIALS AND METHODS

The basis of any analytical process in general and the analysis of the economic activity results of production systems of different levels, in particular, is the information that is manifested through specific data and their aggregates as databases.

Data science is a branch of information technology (IT), which deals primarily with transforming data into information and knowledge. The impetus for the rapid development of this scientific field was the economic systems digitalization in its entirety.

Next, from the standpoint of retrospective analysis, we examine the formation of this area's development in Ukraine. It is essential to distinguish between Data Science (DS) and computer science. Informatics is a broader concept that includes everything related to information (classification, storage, archiving, distribution, protection, collection, etc.), while DS focuses on analysis, collection, and formatting to turn information into knowledge. One can find outdated information that Data Science as science has just emerged and is still developing; this discrepancy (in our opinion) is being investigated by the authors of the monograph [1]. Nevertheless, some specialists work exclusively in this direction, namely, master's programs at universities, methods, and technologies inherent in DS, a sufficient number of books and publications on the field of Data Science [4-12].

However, they are mainly aimed at practical application and do not give the usual presentation of the subject of science. Since DS is closely intertwined with statistics, the question often arises if DS is an evolution of statistics or an independent science.

The emergence of DS is closely linked to computer technology development, which is an integral tool for practical work with information and the Internet as a place with a large amount of diverse and unstructured information. DS is associated with the establishment of the Committee on Data for Science and Technology (CO-DATA) in 1966, which aimed to collect, critically evaluate, store, disseminate data important to science, the technology used by production systems as enterprises [7, 1 p.]. The process gave grounds to allocate work with data in an interdisciplinary field, but at that time, the definition of DS and its tools had not been formed.

In 1974, Naur published the research "Concise Survey of Computer Methods", in which he gave a comprehensive, scientifically sound definition of Data Science. The rapid development of DS as a separate application of computer science took place in the 2000s, which was more academic due to the rapid development of the Internet and the search engines (e.g., google.com), for which it is crucial to process much information and find keywords that belong to the DS toolkit.

In 2002, CODATA was the first to publish a "Data Science Journal", which addressed data systems description issues, their publication on the Internet, applications, and legal issues [4].

The next breakthrough in the DS development was in the 2010s due to the phenomenon of Big Data (hereinafter - BD), or in other words, a significant accumulation of data on the Internet, an increase in the number of private users, and within enterprises that caused by a significant increase in the number of computers in various fields and electronic document management. The data appeared many times more than in the 2000s and, accordingly, increased attention to their analysis because the more data available, the more opportunity for accurate prediction.

Currently, users independently publish information on social networks, voluntarily providing information about themselves valuable in marketing and sales. This phenomenon has led to the emergence of targeted advertising on the Internet, which adapts to a particular user's tastes and requests.

3. DATA SCIENCES DEVELOPMENT

Since 2013, the world's leading universities (The University of California, Berkeley; Imperial College London; University of Washington, etc.) are launching master's Data Science programs. There have been many events and conferences, but scientists have not described how important they are for DS's development, so the story is presented briefly to systematize a DS's nature.

Therefore, the authors will consider in detail the concepts that will be used in both this work and in the practical activities of production systems because they are mainly interpreted in various sources in slightly different ways, which affects the context. It is necessary to define the concepts used in research and reflected in this article. Scholars avoid defining the DS and instead describe the tools, principles, and areas of knowledge that are part of the DS. Here are some definitions to provide a comparative description:

- Data Science is a practical discipline that studies the methods of generalizing knowledge extraction from data. It consists of various parts, based on methods and theories from many fields of knowledge, including signal processing, mathematics, probability models, machine and statistical learning, programming, data technology, pattern recognition, learning theory, visual analysis, modeling uncertainties, organization, data storage, as well as highly efficient calculations in order to extract meaning from the data and create data processing products [8, 26 p.];
- Data Science is a set of fundamental principles that help and manage the process of extracting information and knowledge from data [11, 2 p.];
- Data Science is the science of obtaining knowledge from data; it is a continuation of Data Mining and Predictive Analytics.

This approach is interdisciplinary because it combines methods and models of such disciplines as mathe-

matics, statistics, probability theory, information technology, including signal processing, probabilistic models, machine learning, statistical learning, data mining, databases, object recognition, visualization, uncertainty modeling, data warehouses, data compression, computer programming and high-performance computing [3].

- Data Science is a computer science field that studies the problems of analysis, processing, and data presentation in digital form [12].

The following shared features can be identified from the above definitions: the transformation (or extraction) of knowledge from the data, methods used in the DS. Nevertheless, we believe that it is too early to name DS a science, but more an industry, direction, or discipline. The question may arise about how DS differs from the usual analysis of data or statistics because they solve similar problems in practice. Therefore, it is necessary to understand the analysis, analytics, data analysis, and statistics to determine the differences between them.

Analytics (from the ancient Greek - "art of analysis") - the science and part of the art of reasoning - logic, which considers the doctrine of analysis, operations of imaginary or absolute dismemberment of the whole (things, properties, processes, or relationships between objects) into components, performed in the process of cognition or subject-practical human activity, originated in ancient Greece [13]. This definition is somewhat outdated because today, it is understood as the application of all analytical techniques, not just analysis, and aims to obtain the result of such work in the form of knowledge (or an information product). Thus, modern analytics is:

- The basis of intellectual, logical activities aimed at solving practical cases. It is based not only on the fact-finding principle but also on the principle of "anticipation of events", which allows the organization or individual to predict the future state of the object of analysis;
- A holistic set of principles of methodological, organizational, and technological support of an individual and collective mental activity allows effectively processing information to improve the quality of existing and new knowledge and preparing an information base for optimal management decisions [2].

Statistics is a science that develops and uses mathematical methods to describe data and form judgments about them [14]. According to the above mentioned, the authors concluded that data analysis (hereinafter - *analysis*) is a scientific field that solves data processing problems regardless of their nature and type and develops data processing methods.

Currently, data analysis is an identical concept to analytics. In its essential characteristics, data analysis is similar to statistics. However, the concepts are not identical since data analysis covers the processing of both quantitative and qualitative data (expert opinions, conclusions, etc.), i.e., the use of mathematical methods is optional, as for statistics. Moreover, *analysis* is often identified with DS while DS is based not only on statistical tools but also on statistics, machine learning, and artificial intelligence, expertise, and computer science; therefore, it is a new evolutionary stage in the development of classical mathematical statistics, without actually replacing it. Simultaneously, statisticians will improve their DS skills as vacancies at the DS require knowledge of statistics.

Data analysis (analytics) is not identical to DS because DS can be considered a modern productive analytical work component.

Approaches to determining the stages and procedure for analyzing the indicators of multilevel production systems' economic activity using DS's tools and methods differ among authors depending on the direction of their research interests. For instance, if the author focuses on machine learning algorithms and neural networks through DS, it will be possible to find such principles as selecting parameters and learning algorithms, model learning, and quality assessment of model learning that does not reflect the whole picture.

The best approach to the organization of DS processes is given in the book "Introducing Data Science" (Cielen, Meysman, Ali) [15]. The paper provides block diagrams (process diagrams) of each stage of analysis; although it should be noted that the book focuses on Big Data, the approaches are described as existing universal [15].

Analyzing the information [15, 30 p.], we distinguish the stages of the DS process:

1) Statement of tasks of the analysis, definition of its depth and scales;

2) Search and data collection (internal and external);

3) Data preparation (cleaning, classification, transformation, and combination);

4) Data research (expert assessment and graphical methods);

5) Data modeling (selection, testing, diagnosis of different models, and comparison of their results);

6) Display (conclusions) and automation of analysis.

It shall be understood that the process is iterative and cyclical, i.e., during the analysis, it is possible to return from the 5th stage to the 4th, or the 3rd or 2nd, depending on how close the analyst (researcher) is to the set tasks. If the set tasks have not been achieved, it is appropriate to say that such an analysis has failed and a new one begins, with new tasks or data.

The first stage (setting the analysis tasks) involves choosing the questions to which the analyst will seek answers. The questions should be quite specific, if possible, indicating the time intervals for the analyzed data. The

result of such searches is a technical (or design) specification (TS). The TS is significant if the contractor performs analysis because it will describe what they paid for and the specifics of the relationship between the customer and the contractor regarding the responsibility for maintaining data confidentiality, communications, deadlines, etc.

To set goals correctly and understand them requires at least a superficial knowledge of the subject area of research or the business's specifics.

The main points that are usually included in the TS are [15, 27 p.]:

- A clear overall goal of the study;
- Purpose and context of the project;
- Preliminary description of analysis methods;
- Resources needed to perform the analysis;
- Proof of the ability to implement research in practice (based on similar projects or expert judgments);
- Possible results of the analysis and criteria of its success (estimated indicators, for example, the accuracy of the forecast above 90%);
- Calendar plan.

The set tasks are the basis for analytical work, as the answer to them will be conclusions and forecast information.

The second stage is data collection, meaning "raw" data that needs to be processed and prepared for further analysis. During this stage, the analyst can see whether it will be possible to achieve the objectives based on their expertise. Here, it is important to gather as many indicators as possible that are somehow related to the issues under study. Data sources are divided (classified) on the following grounds:

1. Datatype:

- Numerical;
- Text;
- Images;
- Audio;
- Video.
- Reliable or unreliable;

2. With high authority, authoritative and non-authoritative.

3. Relative place of formation (relative to the customer of the study):

- External;
- Internal.

4. Placement of formation:

- Official (represented by government services in a particular form, or directly by the company about its activities);

- Informal (usually a counterweight to the official, if there are no official, then such a classification is not mentioned).

5. Form of automation:

- Automated (collected with the help of programs in information networks without direct human intervention);
- Non-automated (collected and introduced into the database directly by a person).

6. Frequency:

- Do not have a time dimension (such that influenced the study);
- Annual;
- Quarterly;
- Monthly;
- Daily (often such data in the economy are given on working days, for example, exchange rates);
- Hourly (the nuance is the same as with daily);
- Others (even smaller measurements or another calendar are possible).

7. Open or closed;

8. Paid or free;

9. Legal (official) or insider (disclosure of which is prohibited by law);

10. Universal or specific.

Universal, in the case of marketing or economic research, include the following indicators that affect the living standards of the country's population (export-import balance, population, and migration, inflation, bank interest rate, population distribution by various criteria, exchange rates, etc.), while specific in this case will be data on the industry or specific enterprises. For the computer systems, all data is represented as numbers, but for the analysis of images, audio, or video, it will be necessary to convert them into text or numeric, if possible, in any way while preserving the meaning of this data.

It is crucial when compiling the database to compile a file (document) with metadata by classification and indicating the unit of measurement. Metadata will help (or even be needed) during the next step. Important and interesting data may appear in regular news or trade resources (in the economy, for instance) but not aggregated in an easy-to-analyze form in open sources or enterprises. Therefore, such data will have to be searched and entered into the database manually or refer to paid statistical data portals, for example, statista.com [16], Nielsen, GFK, or order data search from marketing research firms, also data may not be provided for easy copying. It is reflected in the fact that the site has tools for displaying data for any particular day in recent decades, and there is

no monthly submission. If we need daily data for five years to solve the analysis problems, we will need to manually process approximately 1825 web pages of the site, which is very resource-intensive (human resource time) and may be impractical.

The software to retrieve data from web pages, so-called site scraping or web scraping, is used in this case. However, resources protect themselves from this with various tools, including the most well-known to users, the CAPTCHA - a test for human use or work.

In the third stage, data is prepared, or "raw" data is converted into a dataset, which should be suitable for further convenient analysis. This stage includes several sub-stages [15, 30 p.]:

1) Data cleaning:

a) Input or copying errors;

b) Physically impossible values;

c) Missing and empty values;

d) Disconnection;

e) Deviations from the general rules (for example, the country's name is coded in different formats and languages).

2) Data conversion:

a) Aggregation;

b) Extrapolation;

c) Derived metrics (new indicators that can be calculated from available data);

d) Ancillary indicators;

e) Reduction of the number of indicators (variables).

3) Combining data:

a) Connection of data sets;

b) Creation of views (temporary table with data, without removing base tables from the database).

These three sub-steps are inconsistent for all data in general but iteratively (sequentially) for each source or data set of the project because, in the middle of the project, the same units of measurement and formatting rules are stored.

In general, the stage can be quite a long process (up to 80% of the study) [15, 29 p.], as it requires the most attention. Thus, in the first stage of data cleansing, the adequacy test will require the analyst to identify the accurate range of the indicator and eliminate unrealistic ones (for example, a person's age may not exceed 200 years and be a negative number).

In the second sub-stage (data transformation), it is interesting to reduce to a single periodization because the selected method can change the result, so the user can apply several at once and in subsequent stages to choose the most appropriate. For example, it is necessary to analyze the US dollar exchange rate by months against the hryvnia, but the NBU publishes data every working day, i.e., it is necessary to aggregate the data. There are several ways to reduce the dimensionality of data:

- Take the maximum rate for the month;
- Take the minimum rate for the month;
- Arithmetic mean;
- Average median value.

Moreover, if we have quarterly data that needs to be converted to monthly data, we need to find out more about what the data is, whether it changes gradually or not. For example, the minimum wage in Ukraine is set in grades, like most utility tariffs, while the price of sugar or other food products is set solely through market mechanisms, i.e., prices are continually floating. For floating indicators, one may apply smooth extrapolation using a linear trend, while for step indicators, it is better to duplicate the data.

Accordingly, it is possible to use derived metrics to build graphs to search for inaccuracies (errors), disconnections, and unnatural deviations. Reducing the number of variables (indicators) means cutting (without physically removing from the database) those that are not had to be analyzed (more relevant for large databases) or have a robust correlation with another indicator. For example, if there is a database of the population of regions, then knowing the total population and the number of one sex, the other can be counted, of course, provided that there are only two, i.e., has no other values (the other may be: undefined or otherwise).

When combining data (third sub-step), it is necessary to exclude recurring indicators from different sources and at the same time check whether they match because there are different approaches to calculating the same indicators. If they differ nominally, it is possible to compare with graphs if they show a similar pattern and then, comparing the metadata by their origin and authority, choose one source. Both options can be applied, but then, there will be too many options for checks during further modeling because one will need to choose different combinations of indicators for the models.

The next step is to study the data. It includes graphical methods (also called graph) and non-graphical (expert). Graphic ones, in turn, are divided into methods of simple and combined graphs. There are many types of graphs, and it is difficult to identify fundamentally important characteristics because the main task is to display

data, which is sometimes quite helpful. After all, seemingly chaotic data for mathematical (statistical) research can take a meaningful form on the graph.

Overall, the task of analytics in terms of economics can be divided into two interrelated types: evidence-based research and forecasting (extrapolation) using modeling. Forecasting is impossible without current data, and modeling is not always appropriate, so sometimes the first type of task is performed, and the researcher stops at this stage.

Based on the available data, it is possible to analyze the current situation (identifying product demand and trends in the industry or society) and perform retrospective analysis.

Such cases have more methods and approaches to solving because there is no effect of uncertainty compared to forecasting, which is based on assumptions and probability theory. For example, the task of identifying demand can be solved not only by the actual sales data of the company and competitors but also by analyzing the texts of reviews and product reviews for keywords and frequencies and identifying which characteristics and phrases are most mentioned by the authors.

On the other hand, this analysis is unreliable and has to be applied based on expertise, and it should be understood that articles on certain products do not always describe the characteristics of interest to consumers, as, for example, the reliability of household appliances can be determined only after prolonged use. Therefore, it does not make sense to try to extrapolate such data, although this is possible in the case of a sufficient number of time points of the data. Data modeling (fifth stage of analysis) consists of 3 substages:

1) Choice of model and variables (indicators);

2) Execution of the model and selection of parameters of the algorithm;

3) Diagnostics and comparison of models.

They are repeated cyclically until the tasks' achievement at the beginning of the analysis; respectively, the previous step is repeated if they are not achieved. The first sub-stage involves understanding all the main algorithms and models, their scope, strengths, and weaknesses.

There are many types of models, and each subtype can have its competitive features. We will study only such types which are applied in the research:

1) Regressions (for example, simple linear regression);

2) Classifiers (method of k-nearest neighbors);

3) Clustering (k-means method);

4) Decision trees;

5) Neural networks (basic algorithms).

The regression includes the least-squares method, a conditional subtype of linear regression, familiar to economists in econometrics and statistics. Linear regression is often used in graphical analysis because it is a trend line that is easily constructed by automated means and suitable for forecasting.

Among the classifiers, the most well-known algorithm is the k-nearest neighbors algorithm (k-NN) method, which essentially divides the set of points into two classes, drawing a conditional "boundary" between them [17, pp. 43-44].

That is, the algorithm is used where there are two different entities (classes). For example, consumers who have purchased an additional warranty and those who have not, or the same division of potential consumers in the industry into those who use the services and those who do not. It is possible to apply it for forecasting, but, of course, it is not possible to employ for one class and time series, as it is impossible to predict the exchange rate without transformations that can be used in the analysis of several currencies and their division into two classes: the rate rose and the rate fell.

Clustering is used to divide a set of values into groups according to certain shared features or determine the number of such groups, especially if dealing with Big data. For instance, we have a consumer base and need to be divided into significant groups to apply different marketing approaches to each; then, we apply clustering algorithms. It can be used for forecasting, but the application's logic is more to study the available data.

Decision trees are more commonly used for classifications, most often in expert computer systems. For example, medical software for making diagnoses from patient examinations, based on knowledge from the database of decisions of human doctors, i.e., the algorithm gives a solution: ill, non-ill, or suspected disease (the average probability that requires additional examination), in contrast to the method of k-nearest neighbors can manipulate several classes. It can deal with forecasting discrete quantities, but the forecast power is relatively small compared to other algorithms; accordingly, it is not suitable for forecasting continuous quantities (GDP, import or export of the country, etc.).

Neural networks, in particular, controlled as machine learning, work on a different principle than the algorithms described before. The principle is to input data and answers to them, and the neural network tries to randomly select such weights for indicators and neurons that would satisfy a given predictive accuracy during algorithm startup. If there is not enough data to train the neural network, then each run will create different weights

for various indicators and neurons so that the readings will be unstable. When there is not enough data, it is called *undertraining*; the reverse is when there is too much data - *overtraining*.

The 2nd sub-stage (model execution and selection of algorithm operation parameters) will be described more in the practical part.

The 3rd sub-stage (diagnostics and comparison of models) will also be considered in more detail in the practical part, on specific examples. However, the main criterion for comparing models in economic analysis is their predictive ability because the study's primary purpose is the entrepreneur's efforts to predict the future and reflect it in plans with a division into pessimistic, optimal, and optimistic forecasts.

The last 6th stage of analysis is conclusions or presentation of research results. It selects interesting and key trends identified from the entire study and answers the first stage's questions - setting goals. If possible and necessary, such an analysis project can be automated because to refine the analysis, one will need to collect more relevant data and add to the dataset without changing algorithms and their parameters, which requires 95-99% fewer resources than a repeat project study.

However, automation and digitization can significantly expand the quantity and quality of data and their databases at the enterprise's disposal, depending on the degree of automation and computerization of all production processes.

4. CONCLUSIONS

Thus, the authors can draw certain conclusions and generalizations obtained based on the results of the study.

Reliable information allows creating forecasts and assessing their reality using available expert knowledge and giving a subjective assessment of the forecasts.

In general, the forecast indicates that the current economic situation will continue, and no apparent improvement and growth of the economy are expected. However, these are the results of models that perform in a bounded environment, creating a forecast based on the data provided, and it is possible to insert limited data compared to all those data flows that occur in the real physical world.

Therefore, such a forecast is of value for formal operations, where quantitative expression of phenomena and processes are required. The authors did not receive accurate plans for the efficient operation of production systems - a specific enterprise. However, we researched the methodology and algorithm to obtain and process data to create high-quality plans and long-term forecasts

of effective development of different levels and different business scales.

The neural network algorithm's efficiency has not been fully disclosed, as no economic-mathematical model has been created that would include all the data from the existing set of databases. Perhaps it would provide a better forecast and be much closer to objective reality; such a model could be worthy of new research and could be useful for further study.

In the digital economy, this acquires additional opportunities through the branching of neural networks, creating their sites, and so on. The site provides information for the community, raising potential consumers' educational level, which is especially important and significant in the Coronavirus pandemic context [13].

The lack of information, or its inaccuracy, can be corrected by contacting specialized statistical agencies, most of the data they will have available and will not require additional surveys, but such services are paid and will directly affect the final profitability of the enterprise.

REFERENCES

[1] D. Solokha, V. Zagariy, O. Belyakova, C. Marfenko, The effectiveness of the use of Data Science in planning the economic activity of small businesses, KNUCA, 2020, 156 p.

[2] N. Dyachenko, Methodological support of information-analytical activity of state authorities and local self-government bodies, Theory and practice of public administration 4 (2013) 1-10.

[3] W. Strielkowski, I. Gryshova, S. Kalyugina, Modern Technologies in Public Administration Management: A Comparison of Estonia, India and United Kingdom. Administration & Public Management Review 28 (2017), 174-185.

[4 CODATA. About Data Science Journal. https://codata.org/publications/data-science-journal/ Accessed 12 Jan 2021

[5] F. Brockhaus, I. Efron, Encyclopedic dictionary. http://www.vehi.net/brokgauz. Accessed 12 Jan 2021

[6] I. Chernyak (2012) Data Science is a science to be born. https://www.osp.ru/os/2012/05/13016245 Accessed 10 Jan 2021

[7] D. R. Lide, G. H. Wood (2012) CODATA@45 years: The Story of the ICSU Committee on Data for Science and Technology (CODATA) From 1966 to 2010. http://www.codata.org/uploads/CODATA@45years.pdf Accessed 10 Jan 2021

[8] J. Grasse, Data Science. Data Science from Scratch, O'Reilly, 2017, 336 p.

[9] P. Naur, Concise Survey of Computer Methods, Petrocelli Books, 1975, 397 p.

[10] W. Strielkowski, M. Krejcí, I. Čabelková, Factors that influence the success of small and medium enterprises in ICT: a case study from the Czech Republic, Business: Theory and Practice 16(3) (2015), 304-315. DOI: https://doi.org/10.3846/btp.2015.521

[11] F. Provost, T. Fawcett, Data Science and Its Relationship to Big Data and Data-Driven Decision Making, Big Data 1(1) (2013) 51-59. DOI: https://doi.org/10.1089/big.2013.1508

[12] A. Zotov (2018) What is Data Science and why Ukraine needs data science? https://nv.ua/ukr/techno/technoblogs/shcho-take-data-science-i-navishcho-ukrajini-nauka-pro-dani-bloh-antona-zotova-2450168.html Accessed 10 Jan 2021

[13] S. Drobyazko, D. Solokha, O. Bieliakova, Strategic policy of companies in the area of social responsibility: Covid-19 challenges, in: Proceedings of the 1st International Symposium of Earth, Energy, Environmental Science and Sustainable Development (JESSD), E3S Web of Conferences, vol. 211, 2020, 04011. DOI: https://doi.org/10.1051/e3sconf/202021104011

[14] S. Boslaugh, Statistics in a Nutshell, O'Reilly, 2012, 594 p.

[15] D. Cielen, A. Meysman, M. Ali, Introducing Data Science: Big data, machine learning, and more, using Python tools, Shelter Island, 2016, 253 p.

[16] STATISTA, GlobalNo.1 Business Data Platform. www.statista.com Accessed 20 Jan 2021

[17] A. C. Muller, S. Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly, 2016, 392 p.