DATA SCIENCE ANALYSIS FOR MANAGEMENT DECISIONS WITH MACRO- AND MICROECONOMIC UNCERTAINTY

Iryna Radionova[1], Yana Fareniuk[2]

[1]*Doctor of Economic Sciences, Prof. Habil., Kyiv National Economic University named after Vadym Hetman, KROK Economics and Law University, Kyiv, Ukaine, e-mail: irina.radionova@gmail.com, ORCID: https://orcid.org/0000-0002-0941-2867*

[2]*Graduate student, Taras Shevchenko National University of Kyiv, Kyiv, Ukaine, e-mail: yfareniuk@gmail.com, ORCID: https://orcid.org/0000-0001-6837-5042*

The relevance of the studied issue is explained by the fact that the phenomenon of the national economic uncertainty involves the use of special tools of analysis. These tools need to be more complex and sophisticated than those suitable for use in equilibrium economies with sustainable economic growth. The Ukrainian economy, by the nature of changes and the current mechanisms of reproduction, is an economics of uncertainty. Therefore, management decisions at the national level and at the level of individual economic entities provide an adequate response to uncertainties, identified using special tools.

One of the main *assumptions* of the study is the idea that Data Science tools are relatively more appropriate (relevant) for the analysis of the economics of uncertainty. This assumption is verified in this study and on the basis of verification certain management generalizations and conclusions have been made.

The purpose of the study is to identify specific examples of analysis and specification of Data Science capabilities in assessing economic uncertainty. This is uncertainty at the microeconomic and macroeconomic levels. This goal is achieved through the use of Data Science tools in the analysis of large arrays of economic information using available software.

The definition of Data Science as the science of working with large arrays (bases) of data with the aim to extract non-obvious (hidden behind a large number of events) information about existing relations and dependencies is quite common and undeniable. It is clear that the found non-obvious information should become the basis for making more substantiated, and therefore more effective, management decisions.

To understand the possibilities of Data Science in the analysis of uncertainty, it is advisable to detail the content of this science with its structural elements (components) extraction (Fig. 1).

According to Fig. 1, Data Science is an *integrated* analytical science. It is formed by the direct use of the achievements of at least four areas of research based on certain ideas and tools, namely:

– Cognitive Science, which is the science of general patterns of thinking and algorithms of knowledge of the world, which are implemented in various fields of knowledge, for example, in law, finance, medicine, art, economics, etc.;

– Machine Learning, which is the science of efficient ways of learning and of self-learning systems using computer programs which use algorithms for learning about the world, including the construction of artificial neural networks (Artificial Neural Network) [6];

– Big Data, which is the science of generating technologies and rules for organizing large databases, their visualization and representation;

– Data Mining, which is the science of extraction of connections (dependencies) and *regularities* from large databases according to the rules of software (machine) learning, not obvious, but objectively existing, therefore, useful for conscious decision-making.

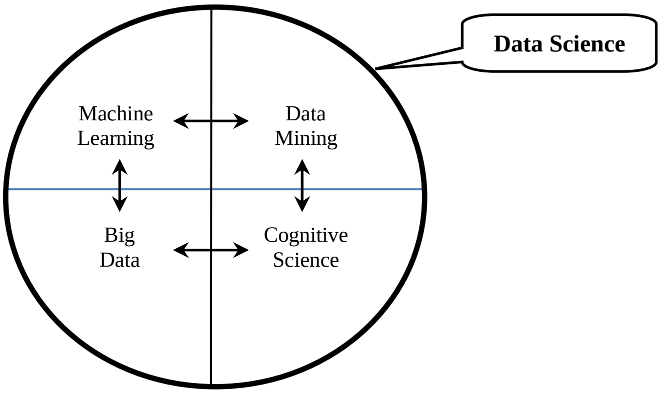

Fig. 1. The content of Data Science by structural elements

Source: authors' own based on [1-5]

The ability of Data Science to provide an appropriate (relevant) analysis of the economy in a state of uncertainty is primarily related to the identification of *non-obvious* relationships (dependencies).

In our opinion, non-obvious relationships (dependencies) have such typical criteria:

– they are not «on the surface» of phenomena, so they are not detected empirically or «in practice», as, for example, a fairly obvious relationship between demand and prices in a particular market, between production costs and output, etc. is identified;

– they do not follow from mass-conscious theoretical constructions, which are usually based on the logical (deductive) derivation of one phenomenon from another, such as the derivation of inflation from unsubstantiated (unexpected) growth of individual elements of aggregate demand or the impact of national currency devaluation on exporters;

– they are caused by discretely random factors («black swans»), which due to the limited intellectual capabilities of communities and ordinary economic entities, as well as due to the limited technical capabilities of the analysis, cannot be predicted with the necessary accuracy;

– they are not expected, in the sense that economic agents will not be able to respond adaptively, rationally, pseudo-rationally, or otherwise.;

– they cannot be detected by traditional methods of information processing and require more sophisticated tools, which are either not yet created or created, but are not used due to the lack of awareness of their benefits and capabilities.

There are grounds to believe that these signs of non-obviousness of relationships (dependencies) between processes and phenomena form the core of *the economics of uncertainty*. In the Ukrainian economy, the uncertainty of relationships caused by incompleteness of objectively necessary transformations (reforms), chaos in actions of government institutions, lack of heredity, therefore, discontinuity (discreteness) of positive traditions of public management of the economy, etc., are added to the fixed by us criterial signs of non-obviousness.

Data Science-analysis is based on the scientific foundation of various schools and areas of economic and mathematical research, within which the issue of detecting and assessing uncertainty is solved.

The retrospective analysis provides grounds for identifying certain stages in the evolution of research on the phenomenon of uncertainty:

– game theory, the ideas of which were formulated in the 1940s by J. von Neumann and O. Morgenstern, and the task was to analyze economic behavior and decision-making under the conditions of uncertainty and conflict situations;

– economic and mathematical modeling of the 1950s – 1980s with its new for that period tools, which cover at least six areas:

1) dynamic and stochastic models which take into account the factor of randomness and uncertainty (R. Frisch, J. Tinbergen, S. Kuznets, K. Arrow, L. Klein, etc.);

2) iterative methods for solving large-scale issues (N.N. Boholiubov, N.M. Krylov, M.F. Kravchuk, A. Yu. Luchka);

3) network planning methods (CPM critical path, PERT programs evaluation and analysis), which allow to minimize the duration of projects taking into account various factors of influence;

4) models of simultaneous equations, which provide for the coverage of a system of separate equations with different levels of identification and different dependent variables which are interrelated (T. Koopmans, T. Haavelmo, G. Gale, etc.);

5) improved models of economic distribution and resources evaluation, in particular evaluation of capital investments from the perspective of the theory of duality (G. Lemke, A.U. Tucker, and others);

6) optimization models of perspective planning, which are used in the development of more substantiated perspective plans for economic development based on knowledge of the most important economic proportions and ratios, production growth rates and consumption levels, rational industry structure (L.V. Kantorovych, V.S. Nemchynov, V.L. Makarov, A.B. Horstko);

– models of behavior of socio-economic systems and assessment of their manageability, which have been actively developed since the end of the twentieth century and involve the use of mathematical tools in at least two of these special forms:

1) detection of impacts of delay in systems responsiveness to stimuli using linear and nonlinear differential equations (V. Volterra, A.D. Mishkis, N.N. Boholiubov, D. Ya. Khusainov, V.I. Fodchuk);

2) expert-evaluated fuzzy logic with formalization in the form of fuzzy sets (L. Zadeh).

It is significant that a certain potential for estimating uncertainty is formed by fairly simple tools of econometric analysis. This applies in particular to multiple regression, a method of multidimensional analysis in which a dependent variable is associated with a set of independent variables (regressors). The study of the tools gives grounds to assert that the actual evolution of regression analysis took place in the direction of finding ways to estimate uncertainty. It is a question of the transition from simple linear regression to nonlinear, and also to autoregression with lag values of variables for estimation of the duration of impulse «responses» and to the use in regression equations, so-called «fictitious variables».

Each of the mentioned tools of econometric analysis, providing certain advantages, has its limitations. Such a typical limitation is that with excessive amounts of information there is a need to limit and simplify it. This causes a distorted model reflection of reality and, consequently, imperfection of solutions based on model constructions. This limitation is partially overcome, for example, with the use of artificial neural networks (Neural Networks). Econometric models and artificial neural networks are tools of Machine Learning, and, therefore, a component of Data Science (Fig. 1).

The input signals are added and, after going through the transmitting (transmission) function, the output is generated, according to the logic of artificial neural network models. The input signals define particular layers (or «nodes») of the interaction, resulting in a multilevel system.

Artificial neural networks are capable of solving a wide range of tasks and have more analytical capabilities than traditional technologies. The most important, in terms of the goal of our research, is the ability to use them for analysis with partial data when assigning a risk level to each operation [7]. Additional alternatives for uncertainty analysis are developed in this approach.

The use of Data Science tools in economic research, in general, and macroeconomic research, in particular, has a long and illustrious history. The following objects were frequently the subject of data science research:

– exchange rate volatility in bilateral relations between countries [8];

– profitability of the financial market, the dynamics of stock indices [9];

– GDP dynamics, economic shocks and «turning points» in the process of economic fluctuations [10; 11];

– factors of price fluctuations and inflation dynamics [12; 13];

– efficiency of certain types of economic policy, in particular with the use of constructions focused on agents and policy makers *(agent-based modeling in economics and policy-making)* [14; 15].

The application of Data Science **at the macroeconomic level** will be illustrated by the example of building neural networks in the study of two macroeconomic phenomena – *economic growth and income differentiation*. Like all other processes and phenomena of the Ukrainian economy, the variables we selected for the study have a high level of uncertainty. In particular, the trend of economic growth in 2016 – 2018 is unexpectedly interrupted by the slowdown in 2019 and a significant decline in 2020. The latter is not explained solely by Covid-19 factors and the global crisis, but has other special significant internal causes in the Ukrainian economy and society.

The study of the phenomena of economic growth and the level of income differentiation used data on the 31 indicators of the Ukrainian economy for the period 1992 – 2018. To ensure the homogeneity of statistical information, the data on Ukraine available in the IMF databases [16] has been used to the maximum. In order to continue the series of data for the 1990s which the IMF information didn’t contain, two other sources have been used for six years [17; 18].

All available indicators are divided into endogenous, i.e. those explained in the models, and five groups of exogenous indicators, i.e. those used to explain. Economic growth rate and Gini coefficient relate to such endogenous indicators. The selection of these indicators is consistent with our goal to explain economic growth and income differentiation in the Ukrainian economy using Data Science tools. Exogenous variables are divided into five groups:

1) the achieved level of development of the national economy;
2) social;
3) financial;
4) monetary;
5) resource potential of the economy.

The division of exogenous model variables into five groups is quite conditional and does not have a significant technical load when building models and explaining the results.

The conditionality of the division of indicators from the database of the studied data is manifested, in particular, in the fact that some of them can be attributed to one or another group. For example, we assume that two indicators of education funding – G_{educ}/G та $Exp_{G/st}/Y_N$ – are indicators of the achieved development level. The reason for this is the idea that the level and quality of education funding, in particular higher (university) education, most reflects the level of development of both the economy and society. Instead, another approach is possible, according to which the indicators of education financing will be presented as belonging exclusively to the group of financial ones. The state budget distribution proportions reflection through them is the basis for such their affiliation. Thus, the division of indicators into groups can be based on various *theoretical assumptions*, so any existing division is «conditionally correct».

However, the division of exogenous indicators into certain groups, despite its conditionality, is necessary for the so-called premodel analysis. Within its limits, at the level of well-known and conscious theoretical constructions, the question of what «in principle» the studied variable may depend on is solved. In particular, according to the new classical theory, economic growth primarily depends on the resource potential of the economy, the achieved level of productivity, savings and investment in technological progress and so on. Thus, these are the indicators covered by our classification, of groups 1 and 5. Instead, in Keynesian theory, economic growth is explained in terms of financial, credit and income indicators, which affect the total costs. Thus, we are talking about the indicators of our selected 2nd, 3rd, 4th groups.

The list of indicators on the basis of which two neural networks were built is given in Table 1.

Table 1. A complete list of indicators used to build two artificial neural networks

Designation of the indicator in Latin letters	The content of the indicator	Sequence number of the indicator in the model
Endogenous (explained) variable models		
g_Y	Economic growth rate (%)	6
k_{Gin}	Gini income differentiation coefficient	2
Indicators of the achieved level of economic development		
Y/N	GDP per capita (USD)	1
Im/Y	Share of imports in GDP (%)	16
Ex/Y	Share of exports in GDP (%)	17
$S_{N/Y}$	The share of national savings in GDP (%)	15
G_{educ}/G	Expenditures on education in government expenditures (%)	26
$Exp_{G/st}/Y_N$	The share of public spending per student (higher education) in GDP per capita (%)	27
Social indicators		
Sh_{lov20}	The share of the poorest quintile group in income	3
Sh_{hig20}	The share of the richest quintile group in income	4
k_{dif}	Quintile income differentiation coefficient	5
Sh_{Pov}	The share of the population below the poverty line (share of those consuming less than 1,9 USD)	31
Financial indicators		
$D_{Ext/GNI}$	The share of external debt in gross national income (%)	7
D_{Ext}	External debt (USD)	8
T/Y	The share of income taxes in GDP (%)	11
G/Y	The share of expenditures in GDP (%)	12
D_G/Y	The share of public debt in GDP (%)	18
Cr_{IMF}	IMF loans used (in USD)	20
Monetary indicators		
π_{CPI}	Inflation rate according to the consumer price index (%)	10
q_{USD}	Exchange rate (UAH per USD)	13
π_{defl}	Inflation rate in terms of GDP deflator (%)	14
Res_{CB}	National Bank reserves (in US dollars)	19
i^r	Real interest rate (%)	21
i_{dep}	Interest rate on deposits (%)	22
g_{BM}	The growth of broad money (%)	23

Designation of the indicator in Latin letters	**The content of the indicator**	**Sequence number of the indicator in the model**
	Indicators of resource potential	
L	Labor force (people)	24
u_{ILO}	Unemployment rate, according to international evaluation methods (%)	25
$u^{/}$	Unemployment rate according to national statistics bodies (%)	9
N	Population of the country (people)	28
n	Population growth rate (%)	29
$migr_N$	Net migration (people)	30

Source: author's own

The relationships verification between the variables of the selected database using regression analysis methods revealed that the following list of variables appeared to be the best to explain *economic growth* g_Y (6): π_{CPI} (10), q_{USD} (13), S_N/Y (15), Im/Y (16), Res_{CB} (19), Cr_{IMF} (20), i^r (21). The Table 2 below discloses the quality characteristics of the regression model.

Table 2. Technical characteristics of model quality

Регрессионная статистика	
Множествен	0,91
R-квадрат	0,84
Нормирован	0,78
Стандартная	4,10
Наблюдения	28,00

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	7,00	1714,49	244,93	14,60	0,00
Остаток	20,00	335,48	16,77		
Итого	27,00	2049,98			

	Коэффициенты	*дартная ои*	*-статистик*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечен	-71,20	12,39	-5,75	0,00	-97,04	-45,36	-97,04	-45,36
10 πCPI Inflat	0,01	0,00	4,86	0,00	0,00	0,01	0,00	0,01
13 qUSD Offic	0,32	0,16	1,92	0,07	-0,03	0,66	-0,03	0,66
15 SN/Y Gross	0,61	0,17	3,51	0,00	0,25	0,97	0,25	0,97
16 Im/Y Impo	1,08	0,25	4,32	0,00	0,56	1,61	0,56	1,61
19 ResCB Tota	0,00	0,00	4,55	0,00	0,00	0,00	0,00	0,00
20 CrIMF Use	0,00	0,00	-2,77	0,01	0,00	0,00	0,00	0,00
21 ir Real inte	0,29	0,05	6,42	0,00	0,20	0,39	0,20	0,39

Source: authors' own

The given technical characteristics of the model testify to the following:

– it is possible to observe a high quality of the model, as the variables presented in it explain the economic growth of 84% (R^2 = 0,84);

– the probability that the model variables are selected incorrectly is zero (F-criterion = 0), i.e. the variables are selected correctly;

– the impact of all model variables on economic growth, represented by the coefficients at the variables, is significant (P-values <0.1 for all variables).

Despite the high technical characteristics of the model, the signs at the coefficients («+» or «-») testify to the contradictory influences and such which are difficult to explain from the macroeconomic perspective. The following are examples of contradicting effects of variables, particularly in terms of the signs at the coefficients:

– a strong positive impact of inflation on economic growth has been revealed (coefficient at π_{CPI} = +0,01), despite the traditional approach to excessive inflation as a factor with an inhibitory effect on growth;

– a strong positive impact of increasing the share of imports in GDP on economic growth has been identified (coefficient at Im/Y = +1,08), despite the traditional approach to imports as a phenomenon of «withdrawal» from the economic cycle, and therefore as a factor that should inhibit growth;

– a strong positive effect of the real interest rate on economic growth has been revealed (coefficient at i^r = +0,29), despite the notion of the deterrent effect of raising interest rates on investment and, consequently, on economic growth.

The clarified contradictions between the character (direction) of the influence of indicators, according to the qualitative regression model, on the one hand, and according to the logic of theoretical explanatory models on these influences, on the other, give grounds at least for such assumptions:

– the effects of econometrically selected indicators are more complex than it is predicted in macroeconomic theory;

– it is probable that these effects are nonlinear, cross, and based on multilevel dependencies.

If our assumptions are correct, then in this case to model and explain economic growth it is appropriate to use Data Science tools with the construction of an artificial neural network.

The artificial neural network model has been created in the Deductor environment. The results of network construction are presented in Fig. 2.

The graph of the neural network shows the nonlinear influence of the seven variables on economic growth, their cross-interaction, and the formation of certain «nodes» of such interaction. The strongest, when forming one of the «nodes», was the impact of the real interest rate and used IMF loans, and the weakest – the impact of the share of imports in GDP. When forming another «node», the influence of all variables was approximately the same.

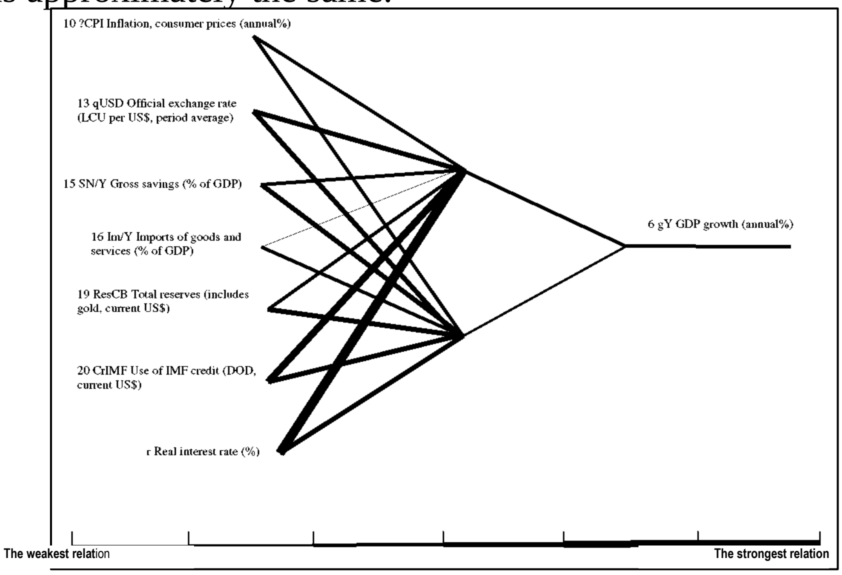

Fig. 2. A neural network graph illustrating the relationship between selected variables and economic growth

Source: authors' own

The verification of the neural network in the test sample is given in Fig. 3.

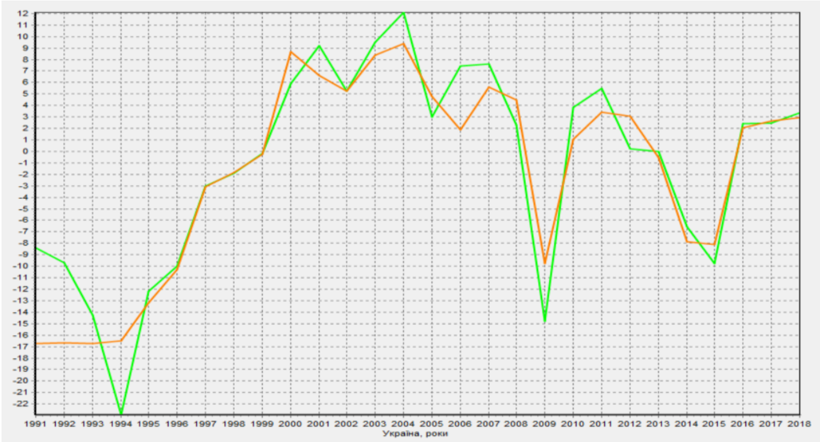

Fig. 3. The verification of the neural network quality

Source: authors' own

Fig. 3 illustrates the degree of approximation between economic growth actual values and values simulated using a neural network. In the test sample (2017 – 2018), the forecast error averaged less than 1%. This indicates the high quality of the neural network and the possibility of its use for prediction and analysis.

The verification of the relationships between the variables to build a neural network which simulates the level of income differentiation, according to the Gini coefficient, regression analysis methods revealed the best variables to explain. Such variables have become as follows: $u^{/}$ (9), π_{CPI} (10), q_{USD} (13), S_N/Y (15), i^r (21), i_{dep} (22), n (29).

The quality characteristics of the regression model of the relationship between the Gini coefficient and other variables are given below in Table 3.

Table 3. Technical characteristics of the model

Регрессионная статистика	
Множественны	0,98
R-квадрат	0,97
Нормированны	0,95
Стандартная оц	0,86
Наблюдения	19,00

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	7,00	267,20	38,17	51,13	0,00
Остаток	11,00	8,21	0,75		
Итого	18,00	275,42			

	Коэффициенты	*андартная ошибі*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*рхние 95,(*
Y-пересечение	23,62	1,84	12,87	0,00	19,58	27,66	19,58	27,66
9 u Unemploym	-0,92	0,27	-3,41	0,01	-1,52	-0,33	-1,52	-0,33
10 πCPI Inflatioı	-0,01	0,00	-7,66	0,00	-0,01	0,00	-0,01	0,00
13 qUSD Official	0,11	0,05	2,09	0,06	-0,01	0,22	-0,01	0,22
15 SN/Y Gross sa	0,16	0,06	2,65	0,02	0,03	0,30	0,03	0,30
21 ir Real intere:	0,06	0,03	1,95	0,08	-0,01	0,13	-0,01	0,13
22 idep Deposit	0,24	0,03	7,17	0,00	0,17	0,32	0,17	0,32
29 n Population	-6,89	1,44	-4,78	0,00	-10,06	-3,72	-10,06	-3,72

The considered technical characteristics of the model in this case indicate the following:

– it is possible to observe a very high quality of the model, because the variables presented in it explain the income differentiation, estimated by the Gini coefficient by 97% (R^2 = 0,97);

– in fact, there is no probability that the model variables are selected incorrectly (F-criterion = 0), i.e. the variables are selected correctly;

– the influence of all model variables on income differentiation, represented by the coefficients at these variables, is significant (P-value < 0.1 for all variables).

In terms of coefficients signs for variables («+» or «-»), i.e. the character of the dependencies, the impact on income differentiation of selected variables looks even more controversial than in the previous model of economic growth. The nature of the influence of only one variable (out of 7) can be explained at least «from the standpoint of common sense». This is the inverse relationship between population growth rate and income differentiation (coefficient at n = -6.89). This inverse relationship can be explained at least by the fact that a decrease in population growth, other things being equal, increases the share of older people, whose incomes are usually lower. The inverse relationship between income differentiation and two variables – unemployment (coefficient at $u^{/}$ = -0,92) and inflation (coefficient at π_{CPI} = -0,01) – contradicts the theory. After all, the assertion regarding inflation as a factor which contributes to the stratification of citizens in society and unemployment, which causes the growth of poverty, is well-established. The positive effect of devaluation on the increase of income differentiation (coefficient at q_{USD} = +0,11) can be explained by the peculiarities of mass storage of savings by Ukrainian citizens in the currency of other countries. The relationships of income differentiation with three more variables – real interest rate (coefficient at i^r = +0,24), interest rate on deposits (coefficient at i_{dep} = -0,01) and the share of savings in GDP (coefficient at S_N/Y = +0,16) – are the most secretive and little understood. In our opinion, they show uncertainty to the greatest extent in the sense of what cannot be explained, realized and expected.

All these features of the income differentiation model give grounds to assume nonlinearity and multilevel character of dependencies, hence, the expediency of using artificial neural networks in modeling tools.

An illustration of a model of an artificial neural network built in the Deductor environment is presented in Fig. 4.

The neural network graph in Fig. 4 testifies to the nonlinear, cross influence of seven variables on income differentiation and the fact of interaction «nodes» formation. The impact of the share of national savings in GDP in the formation of the second «node» appeared to be the strongest of all impacts. When forming the first «node» of interaction, the influence of the exchange rate of the national currency was the strongest. Instead, population growth rate had the weakest impact on income differentiation in both nodes.

The neural network quality verification in the test sample is given in fig. 5. Figure 5 illustrates the high degree of approximation of actual and simulated with the use of neural network values of income differentiation, which is estimated by the Gini coefficient. In the test sample (2017 – 2018), the forecast error was, on average, less than 1%. This indicates a very high quality of neural network and that it can be used for forecasting, analysis and management decisions taking.

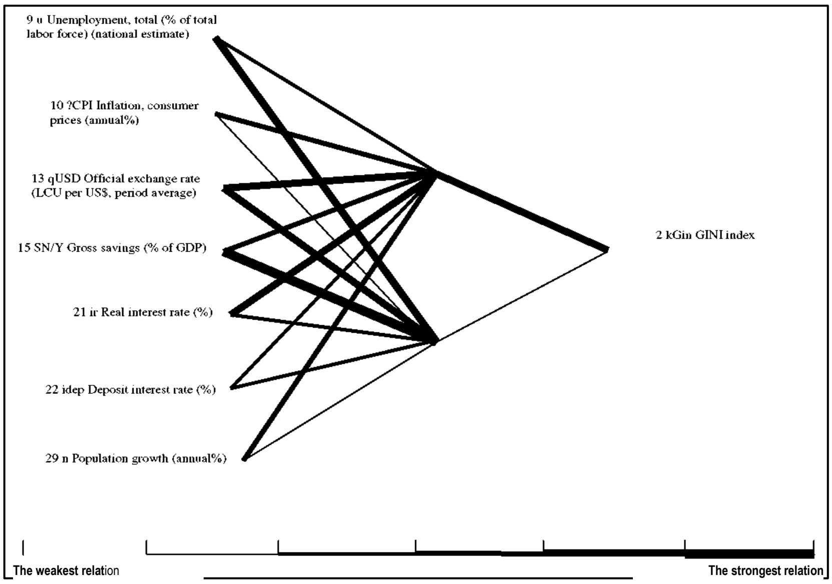

Fig. 4. The graph of the neural network of the relationship between the selected variables and income differentiation, by the Gini coefficient

Source: authors' own

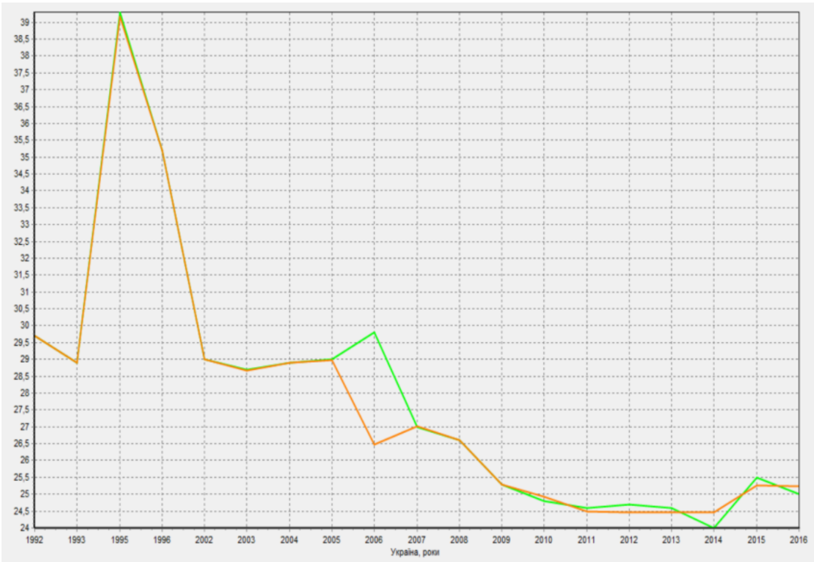

Fig. 5. The neural network quality verification

Source: authors' own

Summarizing the results of using Data Science tools in the analysis at the macroeconomic level, we draw the following intermediate conclusions:

– in the process of econometric analysis of macroeconomic phenomena – economic growth and income differentiation – significant, but «unexplained» and unforeseen relations (dependencies) are revealed. According to our assumption, this reveals economic ambiguity, the so-called «hidden meanings», hence, economic uncertainty;

– the use of Data Science tools makes it possible to draw conclusions about the actual existing dependencies and qualitatively predict the future course of events, even in the presence of vaguely identified effects. Despite the vague identification, the automatic formation of «nodes» of interaction between variables indicates the existence of complex dependencies.

The purpose of this study is to find an answer to the question of how the use of Data Science tools can contribute to taking more substantiated management decisions. At the macroeconomic level, it is evident that we are discussing decisions made by public authorities and so-called national regulators. The example we have considered can be used in the context of at least the following recommendations regarding management decisions:

– forecasting the dynamics of variables that may be the target of the influence of national regulators (economic growth and income differentiation are just such targets), cannot be based only on obvious factors of influence. Nonlinear multilevel relations between variables should be considered in forecasting;

– to substantiate the specific results (responses) of impulses going into the economy from public authorities, it is advisable to form a more complete and perfect idea of the impulses transmission. The latter covers both obvious and non-obvious elements of the transmission, i.e. the relations between variables.

Both recommendations as concerns management decisions improvement at the macroeconomic level can be achieved using Data Science tools.

The relevance of using Data Science at the **microeconomic level** – the level of individual economic entities – in the modern Ukrainian economy is determined by at least the following circumstances. First, there is a low efficiency of investment in general and investment in the banking sector in particular. Therefore, there is a need to analyze and use such research tools which would help identify not only obvious, but also deep, hidden (non-obvious) causes of low efficiency. Second, in management decisions, at present, a tiny fraction of all information related to the activities of individual economic entities is actually used. According to expert estimates, this share is only 0,5%. Therefore, expanding the database of research data may contribute to a more substantiated answer to the question «Why efficiency is low and how to increase the return on investment» for each case.

We will illustrate the practical use of Data Science technologies at the microeconomic level with a specific example, which makes it possible to identify the benefits of this toolkit.

The object of the study was a specific bank[*], which was one of the top 10 banks in Ukraine in 2018. The task (purpose) of the analysis was to find and substantiate ways to increase investment efficiency. It was about media investments in advertising banking services. The purpose of optimizing investment costs for advertising was to increase sales of banking services and increase the profitability of the bank, i.e. increase ROMI (Return of Marketing Investments).

The logic of the model construction in this case was as shown in fig. 6 and reflects the so-called «sales funnel» of the bank through one of the sales channels – the call center.

[*] Due to the obligation not to disclose information, the name of the bank is not made public

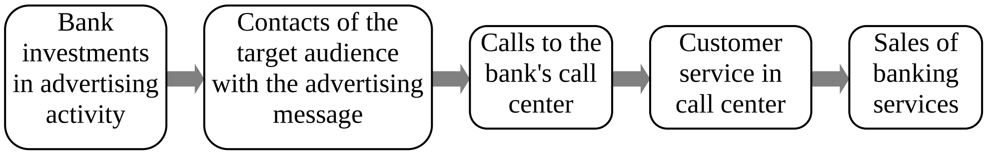

Fig. 6. Model construction logic when substantiating decisions for a commercial bank using Data Science tools

Source: authors' own

The information base of the study covered data on more than 30 indicators in different time intervals, which, according to the source (origin) of the data, were divided into three groups, namely:

- indicators of the bank's activity in relation to the «sales funnel»;
- indicators of advertising activity of the bank and its competitors;
- data on the macroeconomic and social environment of entrepreneurial activity.

The use of the first two groups of data sources in this study is a priori understandable. After all, this is a project as regards the use of advertising activity to increase banking products sales. The third group of sources needs a more detailed explanation.

The banking sector in general, and this bank in particular, incurred enormous losses as a result of changes in the macroeconomic environment during the economic downturn caused by the political crisis and the start of the Russian-Ukrainian war in 2014 – 2015. During this period, the national currency devalued 3-4 times (from 8 to 26-32 UAH per US dollar), the index of confidence in banks decreased significantly [19], deposits in currencies of other countries decreased by 57%, deposits in hryvnia – by 8%, loans – by 38% [20]. Therefore, it is evident, that the recommendations for the bank in 2018 had to be based on the assumption regarding the existence of certain trends in expectations and reactions of consumers of banking services, formed in 2014 – 2015. It was necessary to take into account new macroeconomic trends in economic growth in 2016 – 2018.

The decision to invest in advertising was influenced by the fact that during the national economic crisis, banks in the country reduced their advertising activity. Only a few significant banks were mentioned in the media as attempting to maintain Ukrainians' trust [21; 22]. In 2016, the surveyed bank only ran one advertising campaign, with a low response rate. As a result, the campaign was scaled back and no media advertising was used for the next two years.

Investing in media advertising and designing a media plan to get the best response in the form of calls to the bank's call center emerged as a general business task during the investigated bank's advertising campaign. The investment project's specific goal was to create the «best media mix» of diverse advertising techniques with the optimum distribution of the advertising budget. The achievement of a good return on investment (ROI) is the condition for optimization. As a result, the Data Science project's goal was to research, model, and estimate the call center's load in relation to the bank's media advertising operations. The Data Mining technique was applied, as well as machine learning technology.

The project was carried out using the CRISP-DM intersectoral data sharing standard [23], which is the most often used analytical model in similar investigations [24]. Information on the indicators of past bank advertising campaigns from 2013 to 2016, as well as statistics on socioeconomic development, were collected in order to generate the necessary database. The classical practice of media planning was considered (taking into account the share of voice in the media channel, the frequency of contacts with consumers and the level of coverage of the target audience). However, the main focus was on business indicators (calls to the bank's call center, sales of bank services and conversion levels). In this way, a new integrated approach to decision making based on Data Science, machine learning technologies and maximum use of data was developed.

With the help of Excel and R-Studio software, an econometric model has been developed with the key metric «Incoming calls to the call center». To implement the approach, mathematical methods of analysis and forecasting based on a database with historical data of the bank, media agency and the external environment (social and macroeconomic indicators) have been used. Parameters which affected the conversion (transmission mechanism) from media activity to calls, as well as from calls to orders and sales had been added to the model. The parameters have been optimized to get the best conversion coefficient.

The general econometric model was divided into submodels for monitoring business tasks at each stage, namely:

1) a model of short-term weekly planning, which allowed to record and assess the impact of advertising on the achievement of the bank's performance (both positive and negative) at any time;

2) a model of tactical planning which allows you to plan the hourly traffic intensity of the bank's call center in response to changes in media activity. In particular, the relationship between incoming calls of potential customers of the bank to the call center and the amount of advertising on television during the day was revealed. This made it possible to determine the efficiency of television activity at every hour of the day and throughout the week. The model created the basis for operational optimization of the bank's call center work.

The multiple regression model with more than 30 factors with daily and hourly characteristics looks like this:

$$\begin{gathered} Calls\_by\_hours = hours\_coefficient \times day\_coefficient \times (Constant + a_1 \times \\ \times Adstock(TV_1) + a_2 \times Adstock(TV_2) + \ldots + a_n \times Adstock(TV_n) + b \times \\ \times Radio\_activity + c_i \times billboards\_i + d_i \times Integrated\_economic\_indicator\_i), \end{gathered}$$

where Calls_by_hours – number of calls to the call center with hourly information; hours_coefficient – the efficiency of television activity for a certain hour; day_coefficient – the efficiency of television activity for a particular day of the week;

Constant – basic (organic) call level for a certain hour; a_i – efficiency of television activity of the i-th type, i = 1…n; Adstock – accordingly, the immediate, long-lasting and lagging effect of television advertising on the behavior of banking customers over certain time; $Adstock(TV)_t = TV_t + a \times Adstock(TV)_{t-1}$; b – efficiency of activity on the radio; Radio_activity – activity on the radio; c_i – the efficiency of activity in outdoor advertising of the i-th type, i = 1…n; billboards_i – activity in outdoor advertising of the i-th type, i = 1…n; Integrated_economic_indicator – an integrated indicator that simultaneously reflects the level of GDP, income and dynamics of banking products.

The model is quite complex, from a technical point of view, because it embodies a combination of submodels taking into account the course of events during each day and each week. Data from the data source*. Became the basis for calculations. The same data was used to construct the figures below.

Technical characteristics of one of the submodels of the general model are offered in Table 4.

Table 4. Technical characteristics of the submodel

Indicator	Coefficient	Stand. Error	t-statistics	P-value
Constant	19,78	5,97	3,31	0,0017
Economic indicator	-3,82	0,08	-50,44	0,0000
Billboard	32,98	0,42	77,77	0,0000
Radio	65,24	4,45	14,67	0,0000
TV1	158,53	0,75	211,77	0,0000
TV2	140,34	1,08	130,09	0,0000
TV3	178,96	1,45	123,61	0,0000
TV4	110,27	7,70	14,32	0,0000
Multiple R-squared	0,97		Adjusted R-squared	0,97
F-statistics	11894,423		p-value	0,0000

Source: authors' own

The main criteria of technical optimization of the model were: increase of R2, avoidance of autocorrelation, of heteroskedasticity, and of multicollinearity. The achieved modeling results are as follows: the model estimates the influence of factors with a probability of 97% (R2 = 97%), there is homoskedasticity (i.e. no heteroskedasticity), no autocorrelation.

The developed econometric model allowed to determine the influence of each significant factor and to develop recommendations for the most efficient use of media resources. Here are the top five recommendations.

First, the advisability of adhering to certain duration of the media campaign to minimize the «wear-out effect» based on the data that after reaching the so-called weight of the flight in X TRPs (the main measure of television activity), within Y weeks the efficiency of television activity decreases due to the wear-out effect (Fig.

* Internal database of the Ukrainian bank (confidential information).

7). It was recommended to maintain the flight duration at the required level of TRPs for maximum advertising efficiency.

Second, the importance of rotating (replacing) the advertising videoclips during the flight to further increase the number of calls to the call center of the bank and partially reduce the wear-out effect. The conclusion is based on the assessment that the replacement of the advertisement allows to increase the number of calls by 19%, but it does not compensate for the wear effect. Because there are a lot of short flights, it's a good idea to use a variety of advertising materials.

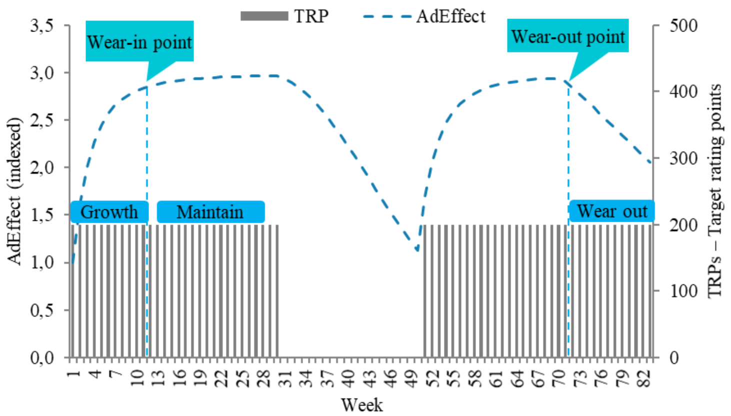

Fig. 7. «Wear-out» effect («wear effect» of the advertising message)

Source: authors' own

Third, the expediency of placing only specific media resources, for example, only the video X" (Fig. 8), avoiding the use of others. It was found that, given the price, advertising, for example, with an X" video, is the most efficient. Therefore, to implement KPIs, it is recommended to use 100% only a certain version of video advertising.

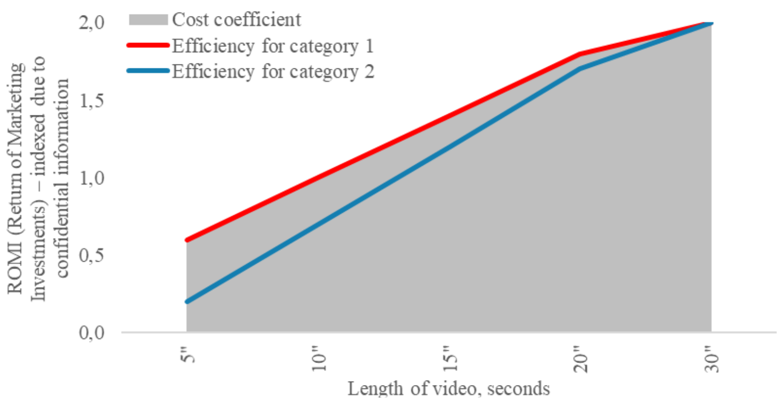

Fig. 8. The efficiency of different durations of videos

Source: authors' own

Fourth, the substantiation of additional activity in other (except already used) communication channels in the last weeks of the TV campaign for potential customers additional coverage and increase of the number of calls to the bank's call center (Fig. 9). It has been proven that another media channel helps to generate additional calls, and the simultaneous use of several communication channels gives an increase in calls for each day of activity of + 20% compared to calls provided by only one TV commercial.

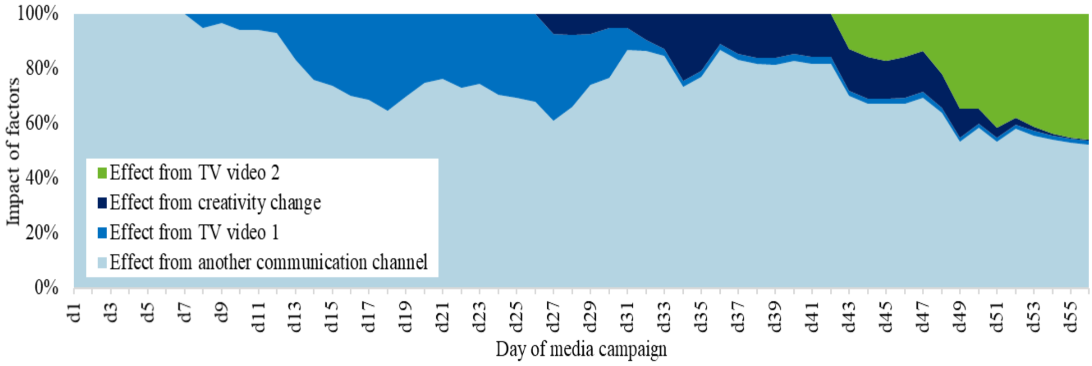

Fig. 9. Decomposition of the model using different communication channels

Source: authors' own

Fifth, the feasibility of implementing tactical organizational changes, namely:

a) restriction of advertising on weekends and holidays on television, as evidenced by the schedule (Fig. 10);

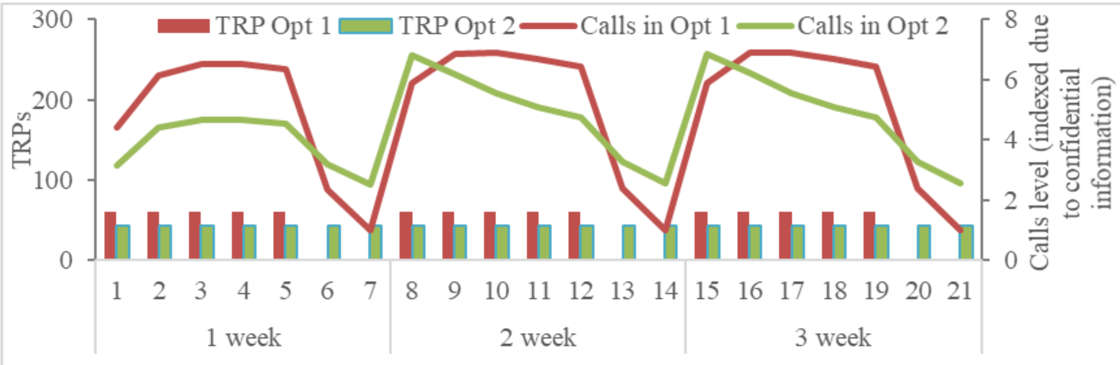

Fig. 10. Calls to the bank's call center according to different scenarios of television activity distribution during the week

Source: authors' own

b) even distribution of activity throughout the day, with the share of evening placements kept to a minimum. It has been established that the efficiency of evening placement is lower than that of daytime and morning placement, and that the effect of evening prime time advertising is similar to that of daytime advertising (Fig. 11).

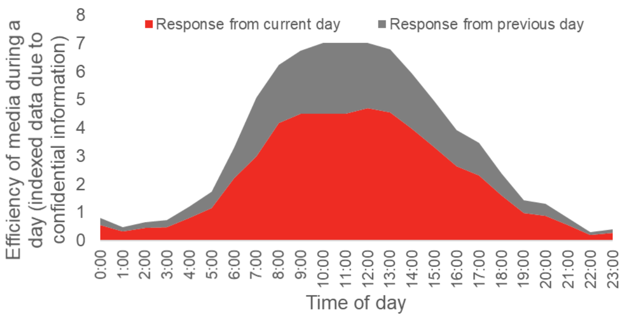

Fig. 11. The number of calls with even distribution of ratings on television during the day (hourly efficiency of television activity)

Source: authors' own

Since the results are a collection of various aspects and conditions that change over time, such suggestions cannot be made simultaneously for all enterprises in the market. This necessitates a unique approach in each scenario.

The main obtained findings of the Data Science project using machine learning technologies are recommendations for improving advertising investment in relation to bank business performance. As previously stated, these recommendations addressed the creation of the most efficient combination (mix) of media, the optimal period (time of day or week) for using mass media to advertise the bank, the duration of specific advertising instruments, and so on. The following specific economic results have been achieved:

– the cost of the advertising campaign is reduced by 14%;

– 58% higher conversion rate (ROMI) – i.e. the link between investing in media activity (advertising campaigns) and the performance indicators of the advertising bank – compared to the standard market level of conversion;

– the activity of the bank's call center is optimized and the probability of «loss» of customer calls due to the impossibility of efficient contact is minimized;

– the possibilities of forecasting the results of the bank's advertising campaign were expanded: on average, the deviation from the forecast was not more than 11% for daily forecasting of results and not more than 8% – for weekly.

It is notable, that the use of Data Science helped to identify the essential impact of factors which could not have been predicted at the beginning of the study and assessed using other research tools. As a result, we conclude: Data Science is an efficient tool for detecting and assessing economic uncertainty at the microeconomic level.

Conclusions: The following generalizations can be drawn based on the study's findings:

1. Data Science is an integrated analytical science which guarantees the best results when using large databases and solving the problem of identifying and evaluating economic uncertainty.

2. The use of Data Science tools in the analysis at the macroeconomic level has helped to identify the following manifestations of uncertainty:

– the significant influence of previously unidentified, i.e. hidden, «non-surface» factors of influence (variables);

– the existence of unnatural, i.e. those that are not explained by either deductive or empirical relationships;

– the formation of random «nodes» of interaction between variables which can mutually strengthen or weaken.

3. The use of Data Science tools in the analysis at the microeconomic level helped to identify the following manifestations of uncertainty:

– the influence on the processes and phenomena of factors from other related areas which change moods, expectations, reactions;

– the nonlinear conversion (transmission) of impulses, which probably has its optimums, or periods of best values;

– the unpredictable distribution of the same events in time when it comes to time peaks and declines of activity;

– the unexpected combinations (mixes) of events and instruments which either strengthen or weaken impulses;

– management decisions improvement based on the applied Data Science tools can relate to more reliable forecasting, selection of management methods, more substantiated tactical planning, organizational changes in the interaction of individual units of economic entities, etc.

References:

1. Provost F., Fawcett T. Data Science for business. How to collect, analyze and use data / Transl. from English by A. Dudchenko. Kyiv: Nash Format, 2019. 400 c.
2. Cherniak O., Zakharchenko P. Data intellectual analysis: textbook. Kyiv, 2014. 599 p.
3. Brin S. et al. Dynamic Itemset Counting and Implication Rules for Market Basket Data. N. Y. : ACM Press, 1997.
4. Einav L., Levin J. The Data Revolution and Economic Analysis. Stanford University and NBER. URL: http://www.fas.org/irp/agency/dod/ jason/data.pdf.
5. Blum A., Hopcroft J., Kannan R. Foundations of Data Science. 2018. URL: http://deeplearning.net/tutorial/deeplearning.pdf.
6. Tom M. Mitchell Machine Learning. 1997. URL: http://profsite.um. ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf.
7. Cherednichenko A.A., Shura N. Application of artificial neural networks as an effective mechanism for making effective management decisions. *Global and national economic issues*. 2015. Issue 4. P. 628-630.
8. Davis J. T., Episcopos A., Wettimuny S. Predicting direction shifts on Canadian–US exchange rates with artificial neural networks. *International Journal of Intel ligent Systems in Accounting, Finance & Management*. 2010. №10. P. 83-96.
9. Armano G., Marchesi M., Murru A. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*. 2005. №170(1). P. 3-33.
10. Tkacz G. Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*. 2001. №17. P. 57-69.
11. Vishwakarma K. P. Recognizing business cycle turning points by means of a neural network. *Computational Economics*. 1994. №7. P. 175-185.
12. Moshiri S., Cameron N. Neural network versus econometrics models in forecasting inflation. *Journal of Forecasting*. 2000. №19. P. 201-217.
13. Mitra S., Rupak T., Seema B., Mondal S. Artificial Neural Network Based Model for Forecasting of Inflation in India. *Fuzzy Information and Engineering*. 2016. Vol. 8. Issue 1. P. 87-100.
14. Hoog S. Deep Learning in (and of) Agent-Based Models: A Prospectus / Sander van der Hoog. 2017. URL: https://arxiv.org › pdf.
15. Chakraborty C., Joseph A. Machine learning at central banks / Chiranjit Chakraborty and Andreas Joseph. *Bank of England. Staff Working Paper*. 2017. №674. URL: https://www.bankofengland.co.uk › working-paper.
16. Official website of the World bank database. URL: https://data. worldbank.org/indicator/.
17. Kravchuk R. S. Ukrainian Political Economy: The First Ten Years. *Palgrave Macmillan*. NY. 2002. 352 p. URL: https://books.google. com.ua/books?id=9BuBDAAAQBAJ&pg=PA147&lpg=PA147&dq.
18. Official site of the State Statistics Service of Ukraine. URL: http://www.ukrstat.gov.ua.
19. GFK Ukraine Website. URL: https://www.gfk.com/uk-ua/.
20. The Website of the National Bank of Ukraine. URL: https://bank.gov.ua/.
21. Nielsen Ukraine Website. URL: https://www.nielsen.com/ua/uk/.
22. Website of the Industrial Television Committee. URL: http://itk.ua/en.
23. Brown M. S. What IT Needs To Know About The Data Mining Process. Forbes. 2015.
24. Shearer C. The CRISP-DM model: the new blueprint for data mining. 2000. №5. P. 13-22.
25. Radionova, I. & Fareniuk, Ya. (2021). Data Science Analysis for Management Decisions with Macro- and Microeconomic Uncertainty. *Economics of uncertainty: content, assessment, regulation*: a collective monograph. Kyiv: KROK University. 287 p. DOI: https://doi.org/10.31732/EU-CER-5