

## РОЗДІЛ 2

### Макроекономічні механізми

#### Data Mining Methods and Models for Social and Economic Processes Forecasting

YULIIA V. DEHTIAROVA<sup>i</sup>, YURI YEVDOKIMOV<sup>ii</sup>

---

Developing social and economic systems and ensuring efficiency of social and economic processes is one of the major tasks for the government of any country. Forecasting models used for analyzing large data sets allow more efficient enterprise management. Big Data is a key resource that provides a competitive advantage to many businesses. The widespread use of Data Mining in retail, marketing, finance, healthcare, industrial production, and other areas suggests that gathered and processed information not only provides useful business information, but also allows more accurate evaluation and development of detailed business plans and development strategies. The use of Data Mining methods and models in forecasting tasks of socio-economic processes provides more accurate predictive calculations. To select the best method for solving prediction problems it is necessary to clearly understand whether the macro series is stationary or non-stationary. We must clearly understand whether our macro series is a clear trend or seasonality. For prediction of stationary time series, the most popular models are autoregression and moving average models. ARIMA models cover a sufficiently wide range of time series, and small modifications of these models allow seasoning time series to be more accurately described. This article discusses the role of Data Mining in social and economic processes, as well as the potential of using Big Data in a business environment. This article demonstrates procedures for using Data Mining methods for practical implementation of Data Mining algorithms to forecast Ukraine's GDP with Python codes, using the statsmodels package. This article analyzes the possibilities of using ARIMA model and uses a double exponential smoothing model for forecasting Ukraine's GDP.

*Keywords:* Data Mining model, forecasting, socio-economic process, macro series, stationarity, trend, seasonality.

УДК 303.6:303.7:33.012.2:005.21:330.101.541

*JEL codes:* C82, C83, C87

---

**Introduction.** In conditions of market economy, uncertainty and risks of business environment entrepreneurs and analysts demand highly efficient processing and analysis of large volumes of information in order to make optimal managerial decisions. Big Data is a key resource that provides a competitive advantage to many businesses. Often, companies have large amounts of data at their disposal, but management is not able to adequately process this data to get the desired result.

Forecasting models used for analyzing large data sets allow more efficient enterprise management. To simulate future effective enterprise development, managers have to take into

---

<sup>i</sup> Yuliia V. Dehtiarova, Student of the Faculty of Information Technology Taras Shevchenko National University of Kyiv.

<sup>ii</sup> Yuri Yevdokimov, PhD, Professor of Department of Economics, University of New Brunswick, Fredericton, Canada.

© Yuliia V. Dehtiarova, Yuri Yevdokimov, 2018.

<https://doi.org/10.21272/mer.2018.80.03>



account not only the risks of external macro- and microenvironment (state and dynamics of competition, demand, institutional processes) and internal processes in a company, but also perspective forecasts that are the basis of the formation and implementation of relevant competitive strategies [12]. Consequently, basic forecasting models are considered to be the first priority.

**Problem statement.** OLAP (Online Analytical Processing), KDD (Knowledge Discovery in Databases) and Data Mining are the most commonly used technologies for statistical analysis and help in decision-making [2]. Some aspects of Data Mining technologies application in forecasting tasks are studied by A. Barsehian [1], A. Bekkauer [2], Ye. Hnitetskyi [8], V. Dyuk [9], R. Eksler [10], I. Kravets [14], O. Krivtsova [15], O. Marchenko [16], V. Minakova [17], V. Pleskach [20], L. Samoilenko [21], M. Beyer [28], R. Schroeder [32], J. Halsall [32], etc.

The research on the possibilities of using ARMA or ARIMA models for the forecast of GDP in Ukraine remains poorly elaborated and requires a more detailed analysis.

**The research goal** is the analysis of data mining methods and the application of models in socio-economic processes forecasting tasks as well as for the forecast of Ukraine's GDP with Python codes, using the statsmodels statistical package.

**Research results.** Forecasting socio-economic processes involves the forecasts of the development of the national economy and social sphere in the future, based on scientific knowledge of socio-economic phenomena and the use of the totality of methods, means and capabilities of prognosis [26, p.15].

Berzlev O. [3, p. 80–82] states that the main classical forecasting models that have been used for many years in a wide range of tasks both by managers for making managerial decisions and researchers in various fields are: models of exponential smoothing that are visible and simple are often used by analysts for short-term forecasting; regression models that are designed to predict or estimate the value of one dependent variable based on the values of other independent variables by establishing communication between them; autoregressive models are used to identify the dependence between values or observations of a time series by studying autocorrelation between observations that are selected with a fixed time interval.

Another popular area of interest is the Time series analysis – a set of innovative methods that includes intelligent data analysis approaches used to identify latent or associative rules of time series data, methods of nonlinear dynamics, optimization on the basis of genetic algorithms and programming of genetic expressions [3, 31], etc.

Currently Data Mining technology is mostly used in solving business problems. Some authors argue that the reason lies in the high efficiency of Data Mining tools, which may reach 1000 %, with a fairly short turnover period.

The widespread use of Data Mining in retail, marketing, finance, healthcare, industrial production, and other areas suggests that gathered and processed information not only provides useful business information, but also allows more accurate evaluation and development of detailed business plans and development strategies.

The application of Data Mining techniques helps to solve business problems including increasing profitability of the department or the whole enterprise; understanding desires and needs of consumers; identification of profitable customers; attracting new customers; preserving regular customers and increasing their loyalty; increasing return on investment; reduction of expenses for the promotion of goods and services; increasing the sales of additional goods and services to existing customers; detection of fraud cases, inappropriate and inefficient spending of funds; credit risk assessment; increasing the effectiveness of using the website; increasing output of a store; optimization of goods location to increase sales; business efficiency

monitoring, etc. [27].

Kravets I. O. and Uzun T. F. [14 p. 123–124] prove that Data Mining algorithms are effective in the analysis of socio-economic indicators. They help to solve classification and regression problems, search for associative groups, clustering and forecasting. For socio-economic information analysis, it is most appropriate to use such intellectual analysis algorithms as: algorithm for finding a decision tree for classification problems (covering algorithm); non-hierarchical Fuzzy C-Means algorithm for addressing the problem of cluster analysis; forecasting of time series with automatic choice of forecasting and taking into account seasonal change of indicators; association rules search algorithm. Using Data Mining methods, management of the enterprise can predict future trends, while taking into account current and past trends.

There are numerous Data Mining methods, but the priority is gradually being shifted to logical search algorithms in the if-then rules of data. These help to solve forecasting and classification tasks, pattern recognition, database segmentation (DB), extraction of 'hidden' knowledge from data, data interpretation, establishment of associations in the database, etc. The results of such algorithms are efficient and easy to interpret [1, 7]. They allow management of the company to choose the most effective competitive strategies while considering all alternatives. In particular, such strategies may be: competitive strategies based on the use of neural networks and strategies for the development of "Internet of Things" [19, p. 369–372].

Data Mining technology combines strictly formalized methods and methods of informal analysis, as well as quantitative and qualitative data analysis [21, p. 19].

Most Data Mining tools are based on two technologies: machine learning and visualization (visual representation of information). The quality of visualization is determined by the possibilities of graphical representation of data values. Variation of graphic representation by changing colors, shapes and other elements simplifies the search for hidden dependencies.

Both technologies complement each other in the process of data mining analysis. Visualization is used to find exceptions, general tendencies and dependencies, and helps in obtaining data at the initial stage of a project. Machine learning is used further to find dependencies in an established project. Table 1 presents the characteristics of intellectual data analysis based on research [7, 11, 16, 22, 24].

The potential of using Big Data in business is extremely high, mostly to increase efficiency, productivity and economic growth. This refers to almost all aspects of business: from research and development to sales and marketing, supply chain management, and company's value. At the same time, the challenges associated with the use of large data in business are growing [5, 25, 28].

To select the method for solving prediction problems it is necessary to understand whether the macro series is stationary or non-stationary. We must clearly understand whether our macro series is a clear trend or seasonality. For prediction of stationary time series, the most popular models are autoregression and moving average models. An ARMA model (autoregressive moving average) is very suitable for stationary time series forecasting. ARIMA models cover a sufficiently wide range of time series, and small modifications of these models allow seasoning time series to be more accurately described [18]. These models will be used to solve different research problems.

In this article, we obtain data related to Ukraine's GDP from the Ministry of Finance with reference to the World Bank data [4]. Recommendations regarding the minimum number of points for forecasting are not available through this data. In addition, we do not use neural networks for training and forecasting, so the available dataset points should be considered

enough to build the prognosis model. We use Python codes for Ukraine’s GDP forecast as Python is inherently readable and simple to use. Data analysts in almost every sector will find Python libraries already tailored to their needs. Moreover, Python has a large range of powerful visualization libraries, such as Matplotlib, Plot.ly, or Seaborn which will help us to build relevant and modern graphics. In order to fulfill the task of forecasting, we need to import the necessary modules and analytical packages. We will need libraries pandas, numpy, matplotlib, scipy, as well as the statsmodel module. To understand which model is best suited, we visualize the data using the matplotlib library. We will build Ukraine’s GDP graphs in UAH and USD (Figure 1-2).

*Table 1*  
Methods of intellectual data analysis [7, 11, 16, 22, 24]

Method	Characteristics
<i>Statistical</i>	Preliminary analysis of the nature of statistical data (verification of hypotheses stationarity, normality, independence, homogeneity, assessment of the type of distribution function, its parameters, etc.). Explore the relation and regularities (linear and nonlinear regression analysis, correlation analysis, etc.). Multivariate statistical analysis (linear and nonlinear analysis, component analysis, factor analysis, etc.). Dynamic models and forecast based on time series.
<i>Cybernetics:</i>  <i>Decision Trees</i>	Classification of data using weighting factors for distribution of data elements into smaller groups. Identify causative relations and determine probabilities or certainty coefficients which allow making appropriate conclusions. Can be used to predict or evaluate unknown parameters (values).
<i>Associative rules</i>	Classify data based on a set of rules, similar rules in expert systems. Rules can be generated using the search and checking of rule combinations, or rules can be generated from a decision tree.
<i>Genetic algorithms</i>	Operate with a set of ‘individuals’, representing rows, each of which encodes one of the task solutions. The suitability of an individual is assessed with the help of a special function. Used to search the global extremum of a multiparameter function; function optimization; dynamical systems optimization; tasks of parameters optimization; adjusting an artificial neural network; training of intellectual models; combinatorial tasks.
<i>Artificial neural networks</i>	Recognition, clustering, forecast – knowledge is presented in the form of links that connect a set of nodes; the strength of relation determines the relation between data factors.
<i>Evolutionary programming</i>	Designed for numerical optimization tasks, in particular, to optimize actual functions. The method of group account of arguments allows providing an acceptable quality of the model in multifactority conditions of a controlled object and a training sample scarcity.

If we compare Ukraine’s GDP using the national currency and USD, we see that there is noticeable exponential growth in the national currency, while in USD we see an explicit trend at some intervals, however, as a whole, we do not distinguish an explicit trend from the graph. We execute the forecast of Ukraine’s GDP in hryvnia using the model of double exponential smoothing, and we will also check if we can use an ARMA or ARIMA model for forecasting.

We check stationarity of the macro series by the Dickey Fuller criterion. The Dickey-Fuller criterion shows that the macro series both in USD and UAH is not stationary. It suggests that

the ARIMA class model is not best suited for predicting this macro series, since it is not stationary. We try to make the macro series stationary using the Box-Cox transformation. If it is successful and the null hypothesis of stationarity is rejected, then the macro series can be predicted with an ARIMA model.

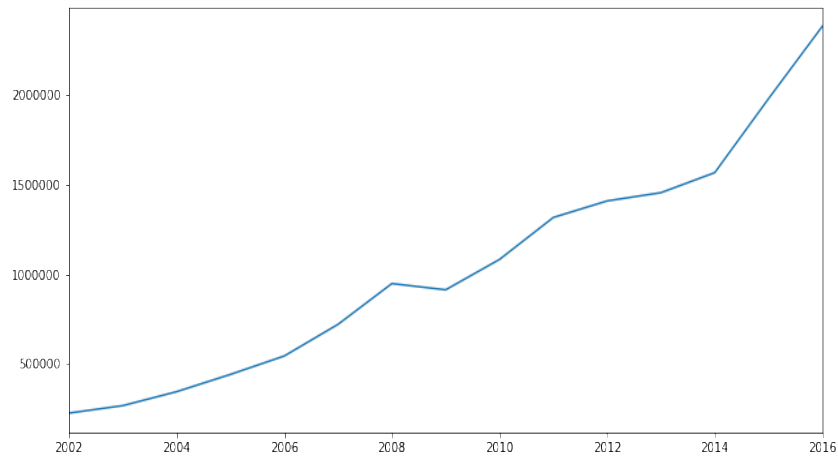


Figure 1. Ukraine's GDP data graph, UAH million

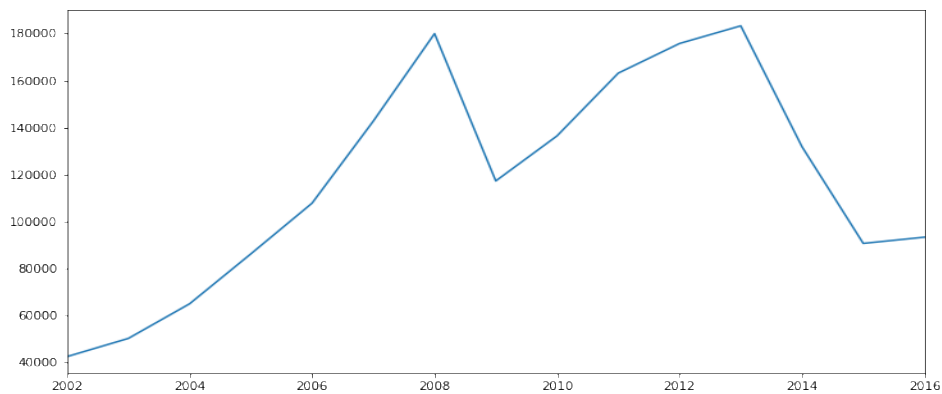


Figure 2. Ukraine's GDP data graph, USD million

If we compare the tests that do not make the normality assumption, we will see that the assumption of normality often leads to simple, mathematically tractable, and powerful tests. We understand that, many real data sets are in fact not approximately normal. As stated in [29] 'an appropriate transformation of a data set can often yield a data set that does follow approximately a normal distribution. This increases the applicability and usefulness of statistical techniques based on the normality assumption'. A measure of the normality of the resulting transformation can be defined through a particular transformation such as the Box-Cox transformation [29]. First, we will try to make a first-order differentiation. We get the following result: The Dickey-Fuller criterion has not changed, which means that our macro series cannot be stable; therefore,

it is necessary to use non-stationary forecasting models for forecasting. The Dickey-Fuller criterion showed the explicit trend in the macro series, and the graph in Figure 3 shows the growing trend. Also, non-parametric methods are most suitable when the sample size is small ( $n < 100$ ). Since seasonality has not been traced and the sample size is 15, we will use double exponential smoothing model (The Holt's model) for time series forecasting, which uses non-parametric methods.

Determination of Alpha and Beta parameters for maximum accuracy of the forecast in the Holt's model is a very important process in the analysis. We used a bunch of values in the model to finalize the result. The setup showed out that Alpha = 0.8 and Beta = 0.03 are best suited for maximum approximation to the actual data model. During the use of the model solution, the predictive model showed the following result:

- the Holt's model forecasting for 2018 – 22465344.6491496265 million.
- the Holt's model forecasting for 2019 – 22546810.7927002399 million.

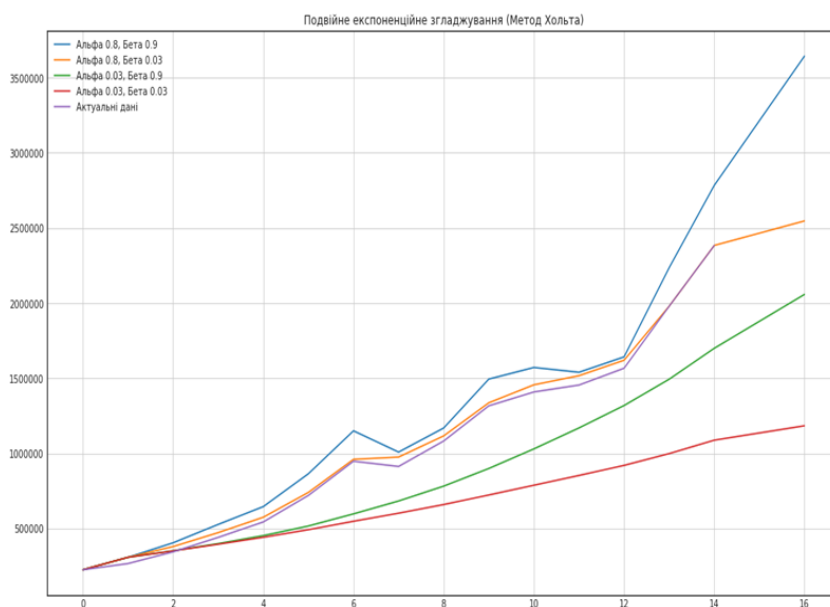


Figure 3. Ukraine's GDP forecasts

**Conclusions and further research.** The widespread use of Data Mining in forecasting of socio-economic processes shows that extracted and processed information generally allows more accurate assessment of enterprises' activities and development of detailed business and country development strategies. Using Data Mining in predicting economic performance is extremely important for our country and helps to track possible trends in the future in order to form effective development strategies.

The use of Data Mining methods and models in forecasting tasks of socio-economic processes provides more accurate predictive calculations. To select the best forecasting model parameters such as seasonality, trend and stationary data must be taken into account. In particular, to find the best forecasting model, special attention must be paid to whether or not the model is stationary. The use of the Dickey-Fuller criterion to test stationarity helped to determine the best forecasting model.

## References

1. Barsehian, A. A. Kupryianov, M. S., Stepanenko, V. V., et al. (2004). *Metody i modeli analiza danyih: OLAP i DatanMining [Methods and data analysis models]*. SPb.: BHV-Peterburg.
2. Bekkauer, A. O. (2016). Vykorystannia tekhnolohii Data Mining dlia avtomatyzatsii biznes-protsesiv na vyrobnytstvi [Using Data Mining Technologies to Automate Business Processes] *Systemy obrobky informatsii*, 1(138), 161–163.
3. Berzlev, O. Iu. (2013). Suchasnyi stan informatsiinykh system prohnouzuvannia chasovykh riadiv [The current state of information systems for forecasting time series]. *Upravlinnia rozvytkom skladnykh system*, 13, 78–82. Retrieved from <http://urss.knuba.edu.ua/files/zbirnyk-13/78-82.pdf>.
4. Valovyi vnutrishnii produkt. (2018). [Gross Domestic Product]. Minfin. Retrieved from <https://index.minfin.com.ua/ua/economy/gdp/>.
5. “Velyki dani”: mozhlyvosti i vyklyky [“Big Data”: opportunities and challenges]. Akademiia “Maibutnie bukhhalterii”. Retrieved from [https://www.nctbpu.org.ua/userfiles/file/analytics/big\\_data\\_its\\_power\\_and\\_perils\\_ua.pdf](https://www.nctbpu.org.ua/userfiles/file/analytics/big_data_its_power_and_perils_ua.pdf).
6. Vigers, K., Bitti, D. (2014). *Razrabotka trebovaniy k programmomu obespecheniyu [Development of software requirements]*. Moscow: Izdatelstvo “Russkaya redaktsiya”. Saint Petersburg: BHV-Peterburg.
7. Hlybovets, M. M. & Hulaieva, N. M. (2013). Evoliutsiine prohramuvannia [Evolutionary programming]. *Problemy prohramuvannia*, 4, 3–13. Retrieved from [http://nbuv.gov.ua/UJRN/Progr\\_2013\\_4\\_2](http://nbuv.gov.ua/UJRN/Progr_2013_4_2).
8. Hnitetskyi, Ye. V. (2017). Big Data v marketynhu: oriientatsiia na spozhyvacha [Big Data in Marketing: Consumer Orientation] *Ekonomichnyi visnyk NTUU “KPI”*, 14. Retrieved from <http://ev.fmm.kpi.ua/article/view/108730>.
9. Dyuk, V. & Samoylenko, A. (2001). “Data Mining”: uchebnyi kurs [“Data Mining”: study course]. SPb.: Piter, 368.
10. Eksler, R. (2017). Big Data: velyki dani, bezmezni mozhlyvosti [Big Data: large data, infinite possibilities]. *Biznes*. Retrieved from [https://biz.nv.ua/ukr/experts/exler\\_ron/big-data-veliki-dani-bezmezni-mozhlyvosti-1565514.html](https://biz.nv.ua/ukr/experts/exler_ron/big-data-veliki-dani-bezmezni-mozhlyvosti-1565514.html).
11. Kalinina, I. V. & Lisovychenko, O. I. (2015). Vykorystannia henetychnykh alhorytmiv v zadakh optimizatsii [Use of genetic algorithms in optimization problems]. *Mizhvidomchyi naukovotekhnichnyi zbirnyk “Adaptyvni systemy avtomatychnoho upravlinnia”*, 1(26), 48–61.
12. Karlberg, K. (2013). *Biznes-analiz s ispolzovaniem Excel. Resheniye biznes-zadach [Business analysis using Excel. Solution of business problems]*. *Business Analysis: Microsoft Excel*. M.: “Viliams”, 576.
13. Kovalchuk, K. F. & Nykytenko, O. K. (2013). Spetsyfika prohnouzuvannia finansovykh ryinkiv na osnovi tekhnolohii Knowledge Mining [Specifics of financial risks forecasting on the basis of Knowledge Mining technology]. *Ekonomichnyi visnyk*, 4, 139–146.
14. Kravets, I. O. & Uzun, T. F. (2017). Vybir ta doslidzhennia efektyvnosti alhorytmiv Data Mining stosovno analizu sotsialno-ekonomichnykh pokaznykiv [Selection and research of the efficiency of Data Mining algorithms in relation to socio-economic indicators analysis] *Kompiuterni tekhnolohii. Naukovi pratsi*, 44, 114–125.
15. Krivtsova, E. (2015). Big Data: Vliyaniye na biznes. Obzor i perspektivy rynku [Big Data: Impact on Business. Overview and market prospects]. *DataReview*. Retrieved from <http://datareview.info/article/big-data-vliyanie-na-biznes-obzor-i-perspektivy-ryinka/>.
16. Marchenko, O. O. & Rossada, T. V. (2017). *Aktualni problemy Data Mining [Actual problems of Data Mining]*. Kyiv: KNU imeni T.Shevchenka, 150.
17. Minakova, V. P. & Shikovets, K. O. (2017). Aktualnist vykorystannia modeli Big Data v biznes-protsesakh [Relevance of Big Data model use in business processes]. *Ekonomika i suspilstvo*, 10, 892–896.
18. Model avtoregresii i kovznogo serednoho (ARMA) [Autoregressive and Moving Average Model (ARMA)]. Retrieved from <http://ekon.in.ua/modele-avtoregresiyi-i-kovznogo-seredneogo-arma.html>.
19. Sotnyk, I. M. & Taranyuk, L. M. (Eds.). (2018). *Pidpriemnytstvo, torhivlia ta birzhova diialnist [Entrepreneurship, trade and stock exchange activities]*. Sumy: VTD “Universytetska knyha”, 572.

20. Pleskach, V. L. & Zatonatska, T. H. (2011). *Informatsiini systemy i tekhnologii na pidpriemstvakh* [Information systems and technologies for enterprises]. Kyiv: Znannia.
21. Samoilenko, L. B. (2018). Mozhlyvosti ta problemy zastosuvannya tekhnologii Big Data vitchyznianymy kompaniiamy [Opportunities and problems of using Big Data technologies by domestic companies] *Efektivna ekonomika*, 1. Retrieved from [http://www.economy.nayka.com.ua/pdf/1\\_2018/59.pdf](http://www.economy.nayka.com.ua/pdf/1_2018/59.pdf).
22. Skakalina, O. V. (2015). Alhorytmy metodu hrupovoho vrakhuvannya arhumentiv pry korotkostrokovomu prohnouzuvanni [Algorithms for the method of group consideration of arguments for short-term forecasting]. *Visnyk KrNU imeni Mykhaila Ostrohradskoho*, 1(90). 18–26.
23. Kharynovych-Yavorska, D. O. (2017). Zastosuvannya neironerezhevnykh tekhnologii dlia prohnouzuvannya konkurentnoi stratehii torhovelnykh pidpriemstv [Application of neural network technologies for forecasting of competitive strategy of trading enterprises]. *Mizhnarodnyi naukovyi zhurnal "Internauka" Seriya: "Ekonomichni nauky"*, 2(2), 25–30.
24. Chto takoye Data Mining [What is Data Mining] (2003). *Journal BPM World Intersoft Lab*. Retrieved from <http://iso.ru/ru/press-center/journal/1948.phtml>.
25. Shakhovska, N. B. (2015). Model Velykykh danykh "sutnist-kharakterystyka" [Big Data Model "essence-characteristic"]. *Visnyk Natsionalnoho universytetu "Lvivska politehnika". Seriya: Informatsiini systemy ta merezhi*, 814, 186–196.
26. Shumska, S. S. (2015). *Makroekonomichne prohnouzuvannya* [Macroeconomic forecast]. Kyiv: Vydavnychiy dim "Kyievo-Mohylianska akademiia", 176.
27. Shuriga, L. (2014). Intellektualnyy analiz dannykh – "zolotaya zhila" bolshogo biznesa [Intelligent data analysis is the "gold mine" of big business]. *DataReview*. Retrieved from <http://datareview.info/article/data-mining-zolotonosnaya-zhila-bolshogo-bizn/>.
28. Beyer, M. A. (2012). The Importance of "Big Data": A Definition Gartner. Retrieved from <https://www.gartner.com/id=2057415>.
29. Box-Cox Normality Plot. NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved from <https://www.itl.nist.gov/div898/handbook/eda/section3/eda336.htm>.
30. Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: John Wiley and Sons, Ltd, 288.
31. Peters, E. E. (1994). *Fractal market analysis: applying chaos theory to investment and economics*. John Wiley & Sons, Inc, 336.
32. Schroeder, R., Halsall, J. (2016). Big data business models: Challenges and opportunities, *Cogent Social Sciences*, 2:1. Retrieved from <https://www.tandfonline.com/doi/pdf/10.1080/23311886.2016.1166924?needAccess=true>.

*Manuscript received 22 February 2018*

**Методы и модели Data Mining в задачах  
прогнозирования социально-экономических процессов**

**Юлия Викторовна Дегтярёва\***  
**Юрий Валентинович Евдокимов\*\***

\* студентка факультета информационных технологий  
Киевского национального университета имени Тараса Шевченко,  
ул. Ванды Василевской, 24, г. Киев, 04116, Украина,  
тел.: 00-380-5847820, e-mail: [yulidehtiarova@gmail.com](mailto:yulidehtiarova@gmail.com)

\*\* доктор экономики, профессор кафедры экономики,  
Университет Нью-Брансвик, Фредриктон, Канада  
465 Singer Hall, Фредриктон, Канада,  
тел.: 00-1-506-447-3221, e-mail: [yuri@unb.ca](mailto:yuri@unb.ca)



Повышение уровня развития социально-экономических систем, обеспечение эффективности процессов, которые в них происходят – одна из главных задач для руководства любой страны. Модели прогнозирования, используемые для анализа больших массивов данных, позволяют более эффективно управлять предприятиями. Достаточно широкое применение методов Data Mining в розничной торговле, маркетинге, финансах, здравоохранении, промышленном производстве и других областях свидетельствует о том, что добытая и обработанная информация не только предоставляет полезную информацию о бизнесе, но и дает возможность более точно оценить работу и составить детальные бизнес-планы и стратегии развития. Использование методов и моделей Data Mining в процессах прогнозирования социально-экономических процессов обеспечивает более точные прогнозные расчеты. Для подбора метода решения задач прогнозирования необходимо четко понимать является ли макроряд стационарным или нестационарным. Для прогнозирования стационарных временных рядов самыми популярными являются модели авторегрессии, скользящего среднего. ARIMA-модели охватывают достаточно широкий спектр временных рядов, а небольшие модификации этих моделей позволяют достаточно точно описывать и временные ряды с сезонностью. В статье рассматривается роль, значение и содержание методов Data Mining в социально-экономических процессах, потенциал использования Big Data в бизнес среде. Выполнены процедуры применения методов Data Mining для практической реализации алгоритмов Data Mining и создания прогнозного ВВП Украины на языке программирования Python с использованием статистического пакета statsmodels. Выполнен прогноз ВВП Украины с помощью модели двойного экспоненциального сглаживания. Проведен анализ возможностей использования модели ARIMA для прогноза ВВП Украины.

*Ключевые слова:* модель Data Mining, прогнозирование, социально-экономический процесс, макроряд, стационарность, тренд, сезонность.

*Mechanism of Economic Regulation, 2018, No 2, 34–44  
ISSN 1726-8699 (print)*

**Методи та моделі Data Mining в задачах  
прогнозування соціально-економічних процесів**

**Юлія Вікторівна Дегтярєва\***  
**Юрій Валентинович Євдокімов\*\***

*\* студентка факультету інформаційних технологій  
Київського національного університету імені Тараса Шевченка,  
вул. Ванди Василевської, 24, м. Київ, 04116, Україна,  
тел.: 00-380-5847820, e-mail: yulidehtiarova@gmail.com*

*\*\* доктор економіки, професор кафедри економіки,  
Університет Нью-Брансвік, Фредріктон, Канада,  
465 Singer Hall, Фредріктон, Канада,  
тел.: 00-1-506-447-3221, e-mail: yuri@unb.ca*

Підвищення рівня розвитку соціально-економічних систем, забезпечення ефективності процесів, що в них відбуваються – одна з головних задач для керівництва будь-якої країни. Моделі прогнозування, що використовуються для аналізу великих масивів даних, дозволяють більш ефективно управляти підприємствами. Досить широке застосування методів Data Mining в роздрібній торгівлі, маркетингу, фінансах, охороні здоров'я, промислового виробництва та інших областях свідчить про те, що видобута та оброблена інформація не тільки надає корисну інформацію про бізнес, а й дає змогу більш точно оцінити свою роботу та розробити детальні бізнес-плани та стратегії розвитку. Використання методів та моделей Data Mining у процесах прогнозування соціально-економічних процесів забезпечує більш точні прогнозні розрахунки. Для

підбору методу вирішення задач прогнозування необхідно розуміти чи є макроряд стаціонарним чи нестаціонарним, чи характеризується макроряд чітким трендом або сезонністю. Для прогнозування стаціонарних часових рядів найпопулярнішими моделями є модель авторегресії та рухомого середнього. ARIMA-моделі охоплюють достатньо широкий спектр часових рядів, а невеликі модифікації цих моделей дозволяють достатньо точно описувати і часові ряди з сезонністю. У статті розглядається роль, значення і зміст методів Data Mining в соціально-економічних процесах, потенціал використання Big Data в бізнес середовищі. Виконано процедури застосування методів Data Mining для практичної реалізації алгоритмів Data Mining і створення прогнозного ВВП України на мові програмування Python з використанням статистичного пакета statsmodels. Виконано прогноз ВВП України за допомогою моделі подвійного експоненціального згладжування. Проведено аналіз можливостей використання моделі ARIMA для прогнозу ВВП України.

*Ключові слова:* модель Data Mining, прогнозування, соціально-економічний процес, макроряд, стаціонарність, тренд, сезонність.

*JEL Codes:* C82, C83, C87

*Table: 1; Figures: 3; References: 32*

*Language of the article:* English

*Література*

1. Барсегян, А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко та ін. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2004. – 336 с.
2. Беккауер, А. О. Використання технологій Data Mining для автоматизації бізнес-процесів на виробництві / А. О. Беккауер // Системи обробки інформації. – 2016. – Випуск 1 (138). – С. 161–163.
3. Берзлев, О. Ю. Сучасний стан інформаційних систем прогнозування часових рядів [Електронний ресурс] / О. Ю. Берзлев // Управління розвитком складних систем. – 2013. – С. 78–82. – Режим доступу : <http://urss.knuba.edu.ua/files/zbirnyk-13/78-82.pdf>.
4. Валовий внутрішній продукт. Мінфін [Електронний ресурс]. Режим доступу : <https://index.minfin.com.ua/ua/economy/gdp/>.
5. «Великі дані»: можливості і виклики [Електронний ресурс] // Академія «Майбутнє бухгалтерії». – Режим доступу : [https://www.nctbpu.org.ua/userfiles/file/analytics/big\\_data\\_its\\_power\\_and\\_perils\\_ua.pdf](https://www.nctbpu.org.ua/userfiles/file/analytics/big_data_its_power_and_perils_ua.pdf).
6. Вигерс, К. Разработка требований к программному обеспечению / К. Вигерс, Д. Битти. – 3-е изд., дополненное / пер. с англ. – М. : Издательство «Русская редакция». – СПб. : БХВ-Петербург, 2014. – 736 с.
7. Глибовець, М. М. Еволюційне програмування / М. М. Глибовець, Н. М. Гулаєва // Проблеми програмування. – 2013. – № 4. – С. 3–13. – Режим доступу : [http://nbuv.gov.ua/UJRN/Progr\\_2013\\_4\\_2](http://nbuv.gov.ua/UJRN/Progr_2013_4_2).
8. Гнітецький, С. В. Big Data в маркетингу: орієнтація на споживача [Електронний ресурс] / С. В. Гнітецький // Економічний вісник НТУУ «КПІ». – 2017. – № 14. – Режим доступу : <http://ev.fmm.kpi.ua/article/view/108730>.
9. Дюк, В. «Data Mining»: учебный курс / В. Дюк, А. Самойленко. – СПб. : Питер, 2001. – 368 с.
10. Екслер, Р. Big Data: великі дані, безмежні можливості. 28 липня 2017 [Електронний ресурс] / Р. Екслер // Бізнес. – Режим доступу : [https://biz.nv.ua/ukr/experts/exler\\_roun/big-data-veliki-dani-bezmezchni-mozhливosti-1565514.html](https://biz.nv.ua/ukr/experts/exler_roun/big-data-veliki-dani-bezmezchni-mozhливosti-1565514.html).
11. Калініна, І. В. Використання генетичних алгоритмів в задачах оптимізації / І. В. Калініна, О. І. Лісовиченко // Міжвідомчий науково-технічний збірник «Адаптивні системи автоматичного управління». – 2015. – № 1(26). – С. 48–61.
12. Карлберг, К. Бизнес-анализ с использованием Excel. Решение бизнес-задач / К. Карлберг // Business Analysis: Microsoft Excel. – 4-е издание. – М. : «Вильямс», 2013. – 576 с.
13. Ковальчук, К. Ф. Специфіка прогнозування фінансових ринків на основі технології Knowledge

- Mining / К. Ф. Ковальчук, О. К. Никитенко // Економічний вісник. – 2013. – № 4. – С. 139–146.
14. *Кравець, І. О.* Вибір та дослідження ефективності алгоритмів Data Mining стосовно аналізу соціально-економічних показників / І. О. Кравець, Т. Ф. Узун // Комп'ютерні технології. Наукові праці. Том 57. – 2006. – Випуск 44. – С. 114–125.
  15. *Кривцова, Е.* Big Data: Влияние на бизнес. Обзор и перспективы рынка. 14.12.2015 [Електронний ресурс] / Е. Кривцова // DataReview. – Режим доступу : <http://datareview.info/article/big-data-vliyanie-na-biznes-obzor-i-perspektivy-ryinka/>.
  16. *Марченко, О. О.* Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики / О. О. Марченко, Т. В. Россада. – Київ : КНУ імені Т.Шевченка. – 2017. – 150 с.
  17. *Мінакова, В. П.* Актуальність використання моделі Big Data в бізнес-процесах / В. П. Мінакова, К. О. Шіковець // Економіка і суспільство. – 2017. – Випуск № 10. – С. 892–896.
  18. *Модель авторегресії і ковзного середнього (ARMA)* [Електронний ресурс]. – Режим доступу : <http://ekon.in.ua/modele-avtoregresiyi-i-kovznogo-seredneogo-arma.html>.
  19. *Підприємництво, торгівля та біржова діяльність : підручник / за заг. ред. д.е.н., проф. І. М. Сотник, д.е.н., проф. Л. М. Таранюка.* – Суми : ВТД «Університетська книга», 2018. – 572 с.
  20. *Плескач, В. Л.* Інформаційні системи і технології на підприємствах : підручник / В. Л. Плескач, Т. Г. Затоначка. – К. : Знання, 2011. – 718 с.
  21. *Самойленко, Л. Б.* Можливості та проблеми застосування технологій Big Data вітчизняними компаніями [Електронний ресурс] / Л. Б. Самойленко // Ефективна економіка. – 2018. – № 1. Режим доступу : [http://www.economy.nauka.com.ua/pdf/1\\_2018/59.pdf](http://www.economy.nauka.com.ua/pdf/1_2018/59.pdf).
  22. *Скакаліна, О. В.* Алгоритми методу групового врахування аргументів при короткостроковому прогнозуванні / О. В. Скакаліна // Вісник КрНУ імені Михайла Остроградського. – 2015. – Випуск 1/2015 (90). Частина 2. – С. 18–26.
  23. *Харинович-Яворська, Д. О.* Застосування нейромережових технологій для прогнозування конкурентної стратегії торговельних підприємств / Д. О. Харинович-Яворська // Міжнародний науковий журнал «Інтернаука» Серія: «Економічні науки». – 2017. – № 2(2). – С. 25–30.
  24. *Что такое Data Mining* [Електронний ресурс] // Журнал BPM World 31 июля 2003. Intersoft Lab – Режим доступа : <http://iso.ru/ru/press-center/journal/1948.phtml>.
  25. *Шаховська, Н. Б.* Модель Великих даних «сутність-характеристика» / Н. Б. Шаховська, Ю. Я. Боллобаш // Вісник Національного університету «Львівська політехніка». Серія: Інформаційні системи та мережі: збірник наукових праць. – 2015. – № 814. – С. 186–196.
  26. *Шумська, С. С.* Макроекономічне прогнозування : навч. посібник / С. С. Шумська : у двох ч. – К. : Видавничий дім «Києво-Могилянська академія», 2015. – Ч. 1. – 176 с.
  27. *Шурига, Л.* Интеллектуальный анализ данных – «золотая жила» большого бизнеса/DataReview [Електронний ресурс] / Л. Шурига, 17.06.2014. – Режим доступу : <http://datareview.info/article/data-mining-zolotonosnaya-zhila-bolshogo-bizn/>.
  28. *Beyer, M. A.* The Importance of 'Big Data': A Definition Gartner [Electronic resource] / M. A. Beyer, D. Laney. – 2012. – Accessed mode : <https://www.gartner.com/id=2057415>.
  29. *Vox-Cox Normality Plot.* NIST/SEMATECH e-Handbook of Statistical Methods [Electronic resource]. – Accessed mode : <https://www.itl.nist.gov/div898/handbook/eda/section3/eda336.htm>.
  30. *Knaflig Cole Nussbaumer.* Storytelling with Data: A Data Visualization Guide for Business Professionals / Knaflig Cole Nussbaumer. – Hoboken, NJ : John Wiley and Sons, Ltd. – 2015. – 288 p.
  31. *Peters, E. E.* Fractal market analysis: applying chaos theory to investment and economics / E. E. Peters. – John Wiley & Sons, Inc, 1994. – 336 p.
  32. *Schroeder, R.* Big data business models: Challenges and opportunities, Cogent Social Sciences, 2:1 [Electronic resource] / R. Schroeder, J. Halsall. – 2016. – Accessed mode : <https://www.tandfonline.com/doi/pdf/10.1080/23311886.2016.1166924?needAccess=true>.