Novi Ekonomist
Vol 16(1), Year XVI, Issue 31, january - june 2022.
ISSN 1840-2313 (Print) 2566-333X (Online)
DOI: 10.7251/NOEEN2231088B

Submitted: 18.05.2022.
Accepted: 20.06.2022.
Review
UDK: 338.24.021.8:330.342(497.11)

FORECASTING GDP GROWTH USING DATA MINING ON THE EXAMPLE OF SERBIA

Rade Božić
University of East Sarajevo, Faculty of Business Economics Bijeljina.
Republic of Srpska, Bosnia and Herzegovina
rade.bozic@fpe.ues.rs.ba
ORCID: 0000-0001-6956-8049

Paper presented at the 10th International Scientific Symposium „EkonBiz: Economic challenges in the conditions of accelerated global changes", Bijeljina, 16 – 17th June 2022.

Abstract: *Predicting the outcome of various phenomena has always been an attractive research topic in a large number of scientific disciplines, especially in economics. As a scientific discipline, econometrics provides various models for predicting indicators such as GDP, inflation rate, interest rate, price of various goods and services, as well as many others at both micro and macro levels. The development of information technologies has made computational operations much faster and more precisely. However, a special contribution is reflected in the application of data mining for the purpose of extracting relevant information from a large data set. Models developed using data mining provide good results in predicting economic indicators, often more successfully than certain econometric models. This paper aims to forecast the growth of GDP through the application of time series mining on the example of the Republic of Serbia. The analysis was performed in two cases: in the first one models include independent attributes that additionally describe the dependent variable, while in the other case they do not contain these attributes. Three different mining methods were used in both cases (linear regression, multilayer perceptron and random forest) and the obtained results of model validation were presented and interpreted.*

Key words: *data mining, forecasting, GDP growth*

JEL classification: E47, O11

1. INTRODUCTION

Macroeconomic indicators reflect the state of the national economy and they represent research subject of a large number of papers from various scientific disciplines. The reason for this is their wide influence on various social areas and aspects, not only those in the economic sphere. Econometrics as a scientific discipline uses statistical techniques and economic theory to investigate the aforementioned indicators, often predicting their movement or outcome under the influence of selected factors (Stock & Watson, 2015). Information technologies provide help in a form of an tool for performing calculation operations. However, within them, a special contribution is provided by data mining. It is a process related to gathering, preparing, processing, analyzing and extracting useful information. It is applied in various domains and it represents a broad aspect that generally describes data processing (Aggarwal, 2015). It includes supervised (classification and regression) and unsupervised learning techniques (clustering and association rules) and finds application in various areas such as satellite image and text analysis, weather forecasting, medical diagnostics, electricity consumption forecasting, automatic abstraction, organic compound analysis and the dangers of pollution, etc. In addition to the above, the applications in the economy, such as, for example, marketing targeting, real estate valuation, product design, financial forecasting, credit card fraud detection, etc., are also popular (Brammer, 2016). Time series data mining is a special field and refers to the possibility of extracting information from a data set that has a time dimension. They are time-stamped and collected with a certain frequency. Examples of such data are sales on a monthly basis, stock trading on a weekly basis, website visits recorded during each hour, etc. This type of data plays an important role in the company's operations, especially in the analysis process. Another standard activity is forecasting. Time series data can be collected manually by humans, however, it is also common to store them through different machines or devices (Dean, 2014).

Gross domestic product (GDP) represents the value of final products and services produced in an economy during a certain period of time (Blanchard, 2008). This paper attempts to predict GDP growth as a macroeconomic variable using different time series data mining methods. This indicator represents the economic growth of the country and has high importance for international investors and creditors. It shows the change in GDP compared to the previous calendar year or other time period. The WEKA software tool was used in the analysis process and prediction in this paper was performed in two different cases. The first case included models with a time component and only one predictor variable (growth of BPD). A forecast outside the selected set for one quarter in future (the first quarter of 2022) was also performed here. The second case refers to models that made predictions within the collected set, including dependent variable (GDP growth), selected independent variables that additionally describe the predicted phenomenon were also included. They were selected from the originally collected set of 14 economic variables (overlay data in the WEKA tool) in the period from 2009 to 2021 on a quarterly basis. In both cases, 3 different data mining methods were used for mutual comparative analysis, and the obtained results are presented and interpreted in a separate chapter.

2. DATASET

To predict GDP growth in the second case, a set of 14 different macroeconomic variables was collected based on the analysis of previously written papers. They relate to the forecast of GDP growth or the analysis of economic indicators that affect it. In the analysis process, various combinations of these attributes were applied, and the best results were shown and presented. As the data for countries in transition are insufficiently available, only those variables that are most often used for the analysis of this problem were collected for this paper. According to the works of Carreiro, Galvao and Kapetanios (2019), Carreiro, Clark and Marcellino (2015) (2019), Schorfheide and Song (2014), Smets, Warne and Wouters (2014), the following indicators were selected and collected on a quarterly basis (table No. 1):

Table 1. *Dataset variables*

No.	Variable name	Units
1.	GDP growth	%
2.	Revenue in industry and construction compared to the previous quarter	%
3.	Revenue in transportation and storage compared to the previous quarter	%
4.	Revenue in information and telecommunications compared to the previous quarter	%
5.	Expenditures for household final consumption	mil. RSD
6.	Expenditures for the final consumption of the state	mil. RSD
7.	Gross investments in fixed assets	mil. RSD
8.	Import of goods and services	mil. RSD
9.	Export of goods and services	mil. RSD
10.	Changes in inventory	mil. RSD
11.	Average gross earnings	RSD
12.	Industrial production - chain index	%
13.	The exchange rate of the RSD in comparison with the USD	RSD
14.	Construction permits	numeric

Izvor: varijable odabrane od strane autora

GDP is most often broken down into the following categories: personal consumption, investment (non-residential and residential), government consumption, net exports and investment in inventories (Blanchard, 2008). In a given data set, these categories are covered by at least one parameter, which matches the variables selected by the previously mentioned authors. All data were taken from the website of the Republic of Serbia Statistical Office (2022, https://www.stat.gov.rs), except for the exchange rate, which was taken from the Investing website (2022, https://www.investing.com/currencies /usd-rsd-historical-data). Data related to monetary amounts are expressed in millions of dinars (RSD).

The data in this paper were collected on a quarterly basis due to insufficient availability of annual data. This type of problem was encountered by a large number of researchers who tried to apply forecasting of economic trends for countries in transition, especially in cases where macroeconomic variables were used. The data set covers the period from the first quarter of 2009 to the fourth quarter of 2021 and there are no missing data in it. The variables in the set are shown over a period of 13 years, which makes a total of 52 observations (instances). Data collected on a monthly basis are expressed on a quarterly basis.

The reason for such a short period is the lack of data related to the past period. As a dependent variable, GDP growth was chosen, which was expressed as a percentage. In the last two years of the included data, the world faced an economic crisis caused by the COVID-19 pandemic, which also affected the Republic of Serbia. This is important to state because the economy was affected by unforeseen factors that led to a 9.2% drop in GDP in the second quarter of 2020. Before the actual application of the selected methods, the data that are not expressed in percentages are normalized, (reduced to an interval from 0 to 1). In this way, their comparison is enabled even though they are not expressed in the same units of measure (Aggarwal, 2015).

3. TIME SERIES FORECASTING IN DATA MINING, METHODS AND EVALUATION PARAMETERS

Time series data is a time-stamped set collected according to a certain frequency (Dean, 2014). Unlike other sets, this type also includes a time dimension that must be considered. Some examples of time series are daily prices of securities in the stock market, monthly number of visits to a website, weekly traffic of cars on a highway, daily emissions of greenhouse gases, etc. Time series can be stationary or non-stationary. A stationary stochastic process is one whose parameters (such as mean and variance) do not change over time, while non-stationary ones are those in which changes are present. Time series that contain the movement of only one variable over a period of time are called univariate, while those that include more variables are called multivariable (Aggarwal, 2015).

Time series often appear in the economic domain, especially in business operations. Such types of information are the basis for decision-making by management. However, these series also play an important role in macroeconomics, especially when it comes to predicting the movement of a certain phenomenon. Forecasting is one of the most common applications of time series analysis and is often used in retail, economic indicators, stock markets, weather forecasting, and in other cases. The aim of this paper is to predict the future value of a certain variable based on its previous values (Aggarwal, 2015). The methods used in this work are linear regression, multilayer perceptron and random forest.

Linear Regression: when the output from the model is numerically expressed, as well as all the input attributes, it is natural to consider linear regression as the method to apply. This is a basic method in statistics. The idea is to express the dependent variable as a linear combination of attributes with predetermined coefficients:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \cdots + w_k a_k$$

where x represents dependent variable, $a_1, a_2 \ldots a_k$ attribute values, while w_0, $w_1 \ldots w_k$ represents coefficients whose calculation is based on the part of dataset for model training.

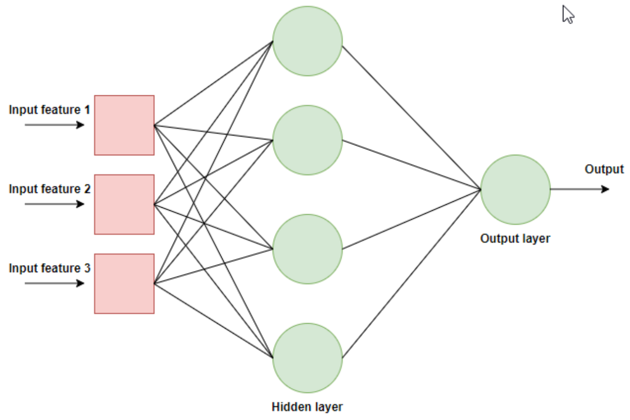
It is important to note that the predicted value of the dependent variable is not its actual value. The difference between the actual and predicted values indicates the success of the model. The method of least squares is used to select the coefficients. The best coefficients are those whose sum of squared deviations from the predicted value in relation to the actual value is the smallest compared to the other coefficients. The performance of this method may be degraded if all input attributes are highly correlated. It is often used as an initial basis for the development of other methods (Witten, Frank, Hall, & Pal, 2017).

Multilayer perceptron (MLP): represents the artificial neural network most often used in practice. MLP has three basic layers (Figure 1): the first layer contains the input data for processing, the second layer is a "hidden" layer where neurons create weight and use an activation function to obtain an output (third) layer that depends on the activity in the previous hidden layer (Tsai & Wu, 2008).

Perceptrons with more than one layer that picks weights are called multilayer perceptrons (Kriesel, 2007). This method belongs to the category of supervised learning because the output values are known to the MLP. It has two basic properties: it has nonlinear characteristics, and strong interdependence determines strong connections between layers (Erkam, Kayakutlu, & Daim, 2011). The MLP neural network is based on an irreversible architecture, and is trained using the back propagation algorithm (Nemes & Butoi, 2013).

The paper used MLP with one hidden layer and a learning coefficient of 0.3. The momentum was 0.2 and the maximum number of learning epochs is 500. In the case where additional attributes are excluded, the input has 11 variables, while the hidden layer has one neuron. In the case where additional attributes are included, there were 15 input variables, while in the hidden layer there was one neuron.

Figure 1. MLP scheme



Source: (Menzies, Kocagüneli, Peters, & Turhan, 2015)

Random forest method: represents a combination of decision trees such that each tree depends on the value of a random vector independently sampled with the same distribution for all trees in the forest. The generalization error for the forest converges to a limit as the number of trees in the forest becomes larger. The error depends on the strength of the individual trees and the correlation between them (Breiman, 2001).

The mean value of the random forest prediction is obtained using bootstrap aggregation and random variable selection. The random forest method turned out to be a robust predictor for small sample sizes and high dimensional data (Gérard & Scornet, 2016).

The usual parameters were used to evaluate the model: mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), root mean squared error (RMSE). The formulas are shown below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\bar{y}_i - y_i|$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - y_i)^2$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\bar{y}_i - y_i}{y_i}\right|$$

$$RMSE = \sum\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - y_i)^2}$$

In formulas above $\bar{y}_i$ is predicted value while y_i is real value for N number of observations. MAE represents the mean absolute difference between the values predicted by the model and the observed historical data, while MSE indicates the mean of the squares of their differences.

MAPE is the average absolute percentage difference between model output and actual values, while RMSE represents the square root of the MSE indicator.

4. RESULTS

The analysis process was conducted using the WEKA tool, i.e. its extension for forecasting time series data. This software was developed by the University of Waikato (New Zealand) and it contains a set of machine learning algorithms. The version used for this paper is 3.8.6.

The analysis process was conducted in two different cases. The first way involves forecasting without using additional variables that explain GDP growth, while the second way involves their inclusion in the models. The first analysis case was conducted due to the possibility of predicting the future value of the dependent variable outside the the test part of dataset. In this paper, one quarter in future (the first quarter in 2022) was selected. When additional variables are included in the model, then in the WEKA tool it is not possible to make a forecast for a future period outside of the collected dataset. The reason for this is that additional independent variables must be known in the future in order for the prediction to be made.

However, they are incorporated into the model by simulating predictions within the existing set, thereby excluding predictions outside of it. If this option is not disabled, then the program will report an error. In both cases, the same methods were used, and the following settings are common: the period is set to a quarterly interval, while the confidence coefficient of 95% is set by default in the software settings and has not been changed. For model training was used 80% of instances, while the last 20% were used for model evaluation. It should be noted that 20% of the last instances include 10 quarters where the COVID-19 virus pandemic was active in 8 of them (since 2020).

4.1. Prediction of BPD growth without additional variables

For the time stamp, the use of an artificial time index (artificial time index) was chosen, where WEKA automatically assigns time stamps to the rows, while one quarter was chosen as the time interval for which the prediction will be made outside of the set.

Table 2 shows the predicted, actual values and their difference (error), while the obtained evaluation results of all three applied methods are shown in Table 3.

Table 2. Presentation of actual and predicted values of applied methods

Quarter	Actual value	LR		MLP		RF	
		Predicted value	Difference	Predicted value	Difference	Predicted value	Difference
43	2,1	2,2452	0,1452	1,4259	-0,6741	1,1558	-0,9442
44	1,8	1,9853	0,1853	1,2905	-0,5095	0,96	-0,84
45	-0,5	1,4892	1,9892	0,5021	1,0021	0,572	1,072
46	-9,2	1,1762	10,3762	0,27	9,47	0,3806	9,5806
47	7,2	1,2644	-5,9356	1,4306	-5,7694	0,2284	-6,9716
48	2,2	2,7704	0,5704	1,4581	-0,7419	0,5543	-1,6457
49	2	8,6321	6,6321	1,4562	-0,5438	1,2295	-0,7705
50	1,7	4,1812	2,4812	1,0395	-0,6605	0,7495	-0,9505
51	1,7	-0,5497	-2,2497	-0,8383	-2,5383	0,3022	-1,3978
52	1,7	2,0963	0,3963	1,3076	-0,3924	0,3812	-1,3188
53*	-	2,5828	-	1,4075	-	0,4173	-

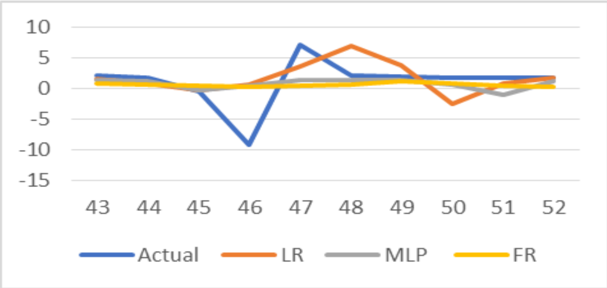
Source: author research

Table number 2 shows the actual and predicted values according to the methods, as well as their difference, or error. In addition to the 10 quarters on which the model was evaluated (quarters 43-52), the 53rd quarter (marked with an asterisk) is also shown, which represents the first quarter of 2022, that is, the predicted quarter outside the collected original set.

Since the actual data for this quarter is not yet publicly available, that value has been omitted as well as the difference between it and the forecasted value.

All three methods predicted GDP growth in this quarter, the highest was by LR (2.5828%), while the lowest growth was predicted by the RF algorithm in the amount of 0.4173%. The tabular data are also shown on graph 1, i.e. the predicted and actual values of GDP growth:

Chart 1. Actual and predicted GDP growth



Source: Author research

Table 3. Errors in applied methods

Errors	Methods		
	LR	MLP	RF
MAE	3,0961	**2,2302**	2,5492
MAPE	126,9408	**71,6045**	83,6026
RMSE	4,501	**3,6409**	3,8863
MSE	20,2593	**13,2564**	15,1036

Source: Author research

Table 3 shows the data related to the evaluation of the model's performance, calculated based on data from the previous table. The smallest amounts of errors indicate that the selected method provides the best results compared to other methods. All four selected indicators have the MLP algorithm at the lowest level, while the linear regression has the highest.

4.2. Prediction with included independent variables

In this case, the models also include independent variables. When forecasting with added variables that subsequently explain the phenomenon, it is not possible to make a forecast for a period outside the set (future period). The reason for this is that in the WEKA tool additional variables must have their value in the period for which the dependent variable is forecast. In the test part of the set (20%) these values exist, so it is possible to make a forecast. Before the algorithm implementation process, different filters related to the selection of attributes were applied, however the manually selected combination provided better results. They are presented in table number 4 and graph 2. The linear regression had the best results provided by the model with included revenue from industry and construction, transport and storage, information and telecommunications compared to the previous period and the export of goods and services. With the included attributes, all evaluation indicators achieved better results. When it comes to the MLP algorithm, with the included attributes revenue from industry and construction, transport and storage, information and telecommunications compared to the previous period and expenditures for state consumption, it provided better results with the MAPE indicator, while the other evaluation parameters were slightly higher rather than without additional attributes included.

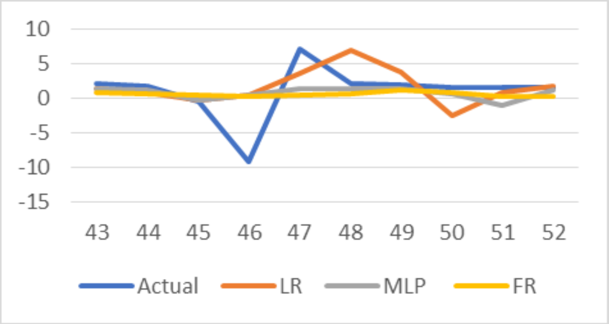
Table 4. Actual and predicted values by quarters

Quarter	Actual value	LR		MLP		RF	
		Predicted value	Difference	Predicted value	Razlika	Predvidena vrednost	Predicted value
43	2,1	1,5194	-0,5806	1,4293	-0,6707	0,945	-1,155
44	1,8	0,9239	-0,8761	1,1559	-0,6441	0,745	-1,055
45	-0,5	-0,2579	0,2421	-0,2661	0,2339	0,4703	0,9703
46	-9,2	0,5813	9,7813	0,4159	9,6159	0,3801	9,5801
47	7,2	3,7005	-3,4995	1,4621	-5,7379	0,525	-6,675
48	2,2	7,0178	4,8178	1,4735	-0,7265	0,6218	-1,5782
49	2	3,9144	1,9144	1,4654	-0,5346	1,214	-0,786
50	1,7	-2,541	-4,241	0,7345	-0,9655	0,8693	-0,8307
51	1,7	0,7904	-0,9096	-0,982	-2,682	0,3959	-1,3041
52	1,7	1,8471	0,1471	1,1879	-0,5121	0,3419	-1,3581

Source: author research

The model with the RF method achieved better results on all evaluation parameters, but on a small level with the inclusion of exports of goods and services and the exchange rate of USD and RSD. The obtained results are presented on chart number 2.

Chart 2. Actual and predicted values by quarter



Source: author research

The evaluation of the model is presented in table number 5.

Table 5. Errors in applied methods

Greške	Metode		
	LR	MLP	RF
MAE	2,701	**2,2323**	2,5293
MAPE	90,6011	**60,3155**	82,1012
RMSE	3,9344	**3,6812**	3,8349
MSE	15,4795	**13,5509**	14,7067

Source: author research

Table 5 shows that the MLP algorithm again has the best results for all evaluation parameters compared to the other applied methods, although 3 out of 4 indicators had a slightly higher level of error. Table number 6 shows the differences in the evaluation of models with and without included attributes.

Tabela 6. Error difference in applied methods

	LR	MLP	RF
MAE	0,3951	-0,0021	0,0199
MAPE	36,3397	11,289	1,5014
RMSE	0,5666	-0,0403	0,0514
MSE	4,7798	-0,2945	0,3969

Source: author research

Looking at table number 6, it can be seen that LR and RF improved all evaluation parameters by introducing new attributes that additionally explain the dependent variable. RF made almost negligible progress while LR's inclusion of new attributes had a significant impact on the model. MLP improved only the MAPE indicator, while the others had slightly lower results compared to the model without additional variables.

CONCLUSION

Forecasting macroeconomic indicators is a challenging task, primarily due to unforeseen impacts. In this paper, GDP growth forecasting was performed over a time series of data using three different data mining methods in two cases. The first case included only GDP growth and forecasting based on it, while in the second case additional variables were included in order to better explain the dependent variable. Part of set (80%) was used for model training, while the last 20% of instances (10 quarters) were used for evaluation. The dat set was collected on a quarterly basis and covered the period from January 1, 2009 to January 1, 2022. The original data set had 14 additional variables (independent variables) where only certain were selected to build the model. Additional variables with LR and RF algorithms provided better results on all evaluation parameters. However, in the MLP method MAPE indicator had a lower value, while the others had slightly higher. The results indicate that the MLP algorithm was the most successful in both cases. Although he provided the best results, they are not at an good level. One of the reasons for this is the COVID-19 pandemic that occurred in the quarters where model prediction and evaluation was performed. This unforeseen impact had a global character, so that a large number of countries recorded turbulent trends in GDP growth. Continuation of research can be based on improving the results of prediction by applying new methods or changing existing, new additional attributes (especially those that show the development of the COVID-19 pandemic), collecting more instances in a data set and etc.

REFERENCES

[1] Aggarwal, C. C. (2015). *Data Mining - The Textbook.* London: Springer International Publishing .

[2] Blanchard, O. (2008). *Macroeconomics.* Massachusetts: Pearson College Div.

[3] Brammer, M. (2016). *Principles of Data Mining.* Oxford, UK: Springer-Verlag London.

[4] Breiman, L. (2001). Random Forest. *Machine learning*, 5-32.

[5] Carreiro, A., Clark, T., & Marcellino, M. (2015). Bayesian vars: specification choices and forecast accuracy. *Jurnal of applied econometrics*, 46-73.

[6] Carreiro, A., Clark, T., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 137-154.

[7] Carriero, A., Galvao, A. B., & Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 1226-1239.

[8] Chatfield, C. (2016). *The Analysis of Time Series: An Introduction.* CRC Press.

[9] Dean, J. (2014). *Data mining and machine learning, Value Creation for Business Leaders and Practitioners.* New Yersey: SAS Institute.

[10] Erkam, G., Kayakutlu, G., & Daim, T. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 10389-10397.

[11] Frank, E., Hall, M. A., Witten, I. H., & Pal, C. J. (2016). *The WEKA Workbench.* Retrieved from Waikato: https://waikato.github.io/weka-wiki/citing_weka/

[12] Gérard, B., & Scornet, E. (2016). A random forest guided tour. *Springer Nature SharedIt*, 197-227.

[13] Hall, M. (2014). Time Series Analysis and Forecasting with Weka - Pentaho Data Mining. Retrieved from https://pentaho-community.atlassian.net/wiki/spaces/DATAMINING/pages/293700841/Time+Series+Analysis+and+Forecasting+with+Weka#TimeSeriesAnalysisandForecastingwithWeka-3.2.2Lagcreation

[14] Kriesel, D. (2007). *A Brief Introduction to Neural Networks.* Retrieved from https://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf

[15] Limited, F. M. (2022). *investing.com*. Retrieved from https://www.investing.com/currencies/usd-rsd-historical-data

[16] Menzies, T., Kocagüneli, E., Peters, F., & Turhan, B. (2015). Using Goals in Model-Based Reasoning. *Sharing Data and Models in Software Engineering*, 321-353.

[17] Nemes, M., & Butoi, A. (2013). Data Mining on Romanian Stock Market Using Neural Networks for Price. *Informatica Economică*, 125-136.

[18] Schorfheide, F., & Song, D. (2014). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*.

[19] Shmueli, G., Bruce, C. P., & Patel, R. N. (2016). *Data Mining For Business Analytics, Concepts, Techniques, and Applications with XLMiner.* New Yersey: John Wiley & Sons.

[20] Smets, F., Warne, A., & Wouters, R. (2014). Professional forecasters and real-time forecasting with a DSGE model. *International Journal of Forecasting*, 981-995.

[21] R. z. (2022). *Republički zavod za statistiku*. Retrieved from https://www.stat.gov.rs

[22] Stock, H. J., & Watson, W. M. (2015). *Introduction to econometrics - third edition.* New Jersey: Pearson Education.

[23] Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy

prediction and credit scoring. *Expert Systems with Applications*, 2639-2649.

[24] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining - Practical Machine Learning Tools and Techniques.* Cambridge: Elsevier.

SUMMARY

Forecasting macroeconomic indicators is a challenging task, primarily due to unforeseen impacts. In this paper, GDP growth forecasting was performed over a time series of data using three different data mining methods in two cases. The first case included only GDP growth and forecasting based on it, while in the second case additional variables were included in order to better explain the dependent variable. Part of set (80%) was used for model training, while the last 20% of instances (10 quarters) were used for evaluation. The dat set was collected on a quarterly basis and covered the period from January 1, 2009 to January 1, 2022. The original data set had 14 additional variables (independent variables) where only certain were selected to build the model. Additional variables with LR and RF algorithms provided better results on all evaluation parameters. However, in the MLP method MAPE indicator had a lower value, while the others had slightly higher. The results indicate that the MLP algorithm was the most successful in both cases. Although he provided the best results, they are not at an good level. One of the reasons for this is the COVID-19 pandemic that occurred in the quarters where model prediction and evaluation was performed. This unforeseen impact had a global character, so that a large number of countries recorded turbulent trends in GDP growth. Continuation of research can be based on improving the results of prediction by applying new methods or changing existing, new additional attributes (especially those that show the development of the COVID-19 pandemic), collecting more instances in a data set and etc.