Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5, 1628-1636
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i5.1881

Performance evaluation of housing price index prediction model using DNN and SVR

Chang-Ho An[1]*
[1]Department of Financial Information Engineering, Seokyeong University; choan@skuniv.ac.kr (C.H.A).

Abstract: This study used the base interest rate, mortgage interest rate, deposit interest rate, consumer price index, exchange rate, and composite stock price index as variables to predict the housing price index. The variables were normalized using the min-max normalization method. Machine learning methods, Deep Neural Networks (DNN) and Support Vector Regression (SVR), were employed to construct a prediction model, and the fit of the model was evaluated using the Root Mean Square Error (RMSE). The analysis results of the DNN model indicated that when applying the softplus activation function, the RMSE was 0.1298, resulting in the smallest prediction error, and the correlation coefficient between the predicted and actual values was highest at 0.9342. The analysis results of the SVR model using the Gaussian RBF kernel showed that the RMSE was smallest at 0.2781 when the parameters were set to C=100 and Gamma=0.1. Therefore, the DNN model was found to be more fit for predicting the housing price index. Additionally, it was confirmed that the consumer price index had a non-linear effect with a positive relationship, and the base interest rate had a linear effect with a negative relationship.
Keywords: *Activation function, Correlation coefficient, DNN, Min-max normalization, SVR.*

1. Introduction

The global economy continues to experience a slowdown due to major countries' monetary tightening policies, the contraction of Germany's manufacturing exports, deflation in China, the Russia-Ukraine war, and so on. In particular, economic instability has persisted since the Russia-Ukraine war, leading to growing concerns about the global economy, marked by high inflation due to high-interest rates and rising commodity prices. The global economic recession, characterized by high prices, high-interest rates, and high exchange rates, significantly impacts our country's economy, leading to decreased exports, a slowdown in economic growth, and a decline in the real estate market. Housing prices are fundamentally determined by supply and demand. Key factors affecting supply include unsold homes, changes in the construction market, and land prices. Key factors affecting demand include household income and assets, housing purchase costs, and mortgage interest rates. According to the 2024 KB Real Estate Report, interest rates are expected to be the most significant factor influencing both the rise and fall of housing prices, with a nationwide decline in housing prices projected. Additionally, a survey conducted among real estate market experts, licensed real estate agents, and private bankers (PBs) revealed that 74% of real estate market experts and 79% of licensed real estate agents and PBs anticipate a decrease in housing prices. The survey identified the following as key factors for potential housing price increases: expectations of lower interest rates, deregulation policies, supply shortages, and improved economic outlook. Conversely, the primary factors for potential housing price decreases included high interest costs due to rate hikes, real estate project financing (PF) defaults, economic uncertainty both domestically and internationally, and an increase in listings due to price declines [1]. In the real estate market, housing prices are influenced by the interaction of key factors of supply and demand through various channels. Housing, as a durable good, is essential for living and is also a crucial asset for household wealth accumulation due to its value as a real estate asset. Therefore, this study aims to evaluate the fit of housing

* Correspondence: choan@skuniv.ac.kr

price index (HPI) prediction models by constructing Deep Neural Network (DNN) and Support Vector Regression (SVR) models using economic variables that impact the HPI, such as base interest rate (BIR), mortgage interest rate (MIR), deposit interest rate (DIR), consumer price index (CPI), exchange rate (ER), and composite stock price index (CSI), provided by the Bank of Korea's Economic Statistics System (ECOS).

2. Previous Research

As concerns over the real estate market increase due to the global economic recession, changes in housing prices have become a national issue, and the need for developing prediction models has grown. There are various methods to predict housing prices, considering both domestic and international economic variables, particularly machine learning methods like neural network models and support vector regression models, which are widely used across different fields. The following summarizes various research methods and previous studies using neural network models and support vector regression models.

Antipov and Pokryshevskaya (2012) estimated Russian apartment prices using machine learning algorithms and showed that the artificial neural network model had better predictive power than the multiple regression model [2]. Shynkevich et al. (2017) developed an artificial intelligence model for stock price prediction using technical analysis indicators such as moving averages (MA), rate of change, and relative strength index as input variables [3]. Bae and Yu (2017) estimated apartment prices using various machine learning methods and showed that the predictive power of machine learning methods such as support vector machine (SVM) and deep neural network (DNN) models was superior to that of multiple regression models [4], and Shin et al. (2018) applied the methodology of Jurado et al. (2015) to Korean economic data to measure uncertainty and compare it with the movement of macro variables [5]. Kim et al. (2020) analyzed the linkage between economic uncertainty and the change rate of housing prices in the metropolitan area and local areas using the linkage index. The results showed that the Economic Policy Uncertainty Index (EPU Index) of Korea affected the national housing price change rate, and the regional analysis also showed a high correlation between the Economic Policy Uncertainty Index (EPU Index) and the housing price change rate [6]. Lee (2021) constructed a multilayer perceptron neural network model and adjusted the activation function and learning rate to predict pile bearing capacity. The validity of the neural network model was demonstrated by estimating the minimum mean absolute error of the model [7]. Kim, Kim, and Ham (2021) compared the prediction models by applying the artificial neural network model and the support vector regression model using daily data. The two models were found to provide high predictive power in predicting drought [8]. Lee and Ruy (2021) demonstrated the superiority of the artificial neural network model by predicting housing prices using 12 macroeconomic and microeconomic variables [9]. Lee, Youn, and Lee (2023) analyzed apartment sale prices as dependent variables and the money supply, interest rates, consumer price index, real estate policy, and housing loan amount as independent variables. The analysis results showed that policy variables and consumer price index affect apartment sale prices [10].

As seen from the results of previous studies, housing prices are greatly affected by domestic and foreign economic variables. As uncertainty in the real estate market increases due to the global economic downturn, research on prediction models using economic variables is an important task at hand.

3. Research Model

3.1. Data Normalization

The data used were provided by the Economic Statistics System (ECOS) of the Bank of Korea [11]. Since the ranges of the variables in the data were different, min-max normalization was performed by adjusting the data range to 0 and 1, as shown in (Equation 1).

$$X = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{1}$$

3.2. Deep Neural Network (DNN)

Deep neural network (DNN) consists of an input layer, hidden layers (two or more), and an output layer. Each layer consists of multiple nodes, and the nodes of each layer process and transform data with weights and biases. The output elements of each layer are calculated as follows. First, the input value is calculated as the sum of the input data (normalized input data) and the weights, as in (Equation 2) [12],

$$I_i = \sum_{i=1}^{n} x_i w_{ji} + b_j \tag{2}$$

where I_j is the input value, x_i is the data of the input layer, w_{ji} is the weight, and b is the threshold. Then, the calculated input value is sent to the activation function to calculate the input data of the hidden layer as in (Equation 3),

$$y_d = f(I_j) \tag{3}$$

where y_d is the input data of the hidden layer, and f is the activation function.

Second, in the case of the hidden layer, it is calculated as the sum of the input data of the hidden layer and the weights as in (Equation 4), and the calculated value is sent to the output layer [13],

$$L_k = \sum_{i=1}^{n} w_{kj} y_d + b_k \tag{4}$$

where L_k is the response value, w_{kj} is the weight of the hidden layer, and b_k is the threshold value in the hidden layer. In this study, soft plus, hyperbolic tangent (tanh), and rectified linear unit (ReLU) activation functions were used.

$$\begin{gathered} softplus : f(x) = log\,(1 + e^x) \\ \tanh : f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \\ ReLU : f(x) = max\,(0, x) \end{gathered} \tag{5}$$

3.3. Support Vector Regression (SVR)

Support vector regression (SVR) constructs a regression equation using an ε-insensitive loss function in the regression model of SVM. The objective function of SVR based on the ε-insensitive loss function is defined as follows [14],

$$L_{SVR} = min\frac{|w|^2}{2} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{6}$$

where $\frac{|w|^2}{2}$ maximizes the margin, $C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$ minimizes the loss, C is a hyperparameter that controls the loss and smoothness, ξ_i and ξ_i^* are slack variables that represent the error. The objective function of (Equation 6) has the following conditions.

$$(w^T x_i + b) - y_i \leq \varepsilon + \xi_i$$
$$y_i - (w^T x_i + b) \leq \varepsilon + \xi_i^* \quad (7)$$
$$\xi_i \xi_i^* \geq 0$$

Applying the Lagrange multiplier and the KKK(Karush-Kuhn-Tucker) condition to the objective function, the general support vector regression function is as follows,

$$f(x, v) = f(x, \alpha, \alpha^*) = \sum_{i}^{n} (\alpha - \alpha^*) K(x_i, x_j) + b \quad (8)$$

where $K(x_i, x_j)$ is the kernel function.

3.4. Model Performance Evaluation

There are measures to evaluate the performance of the prediction model, such as MAE, MAPE, and MSE. In this study, the root mean square error (RMSE) was used,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

where y_i is the ith actual value, and $\hat{y}_i$ is the ith predicted value by the model.

4. Research Results

4.1. Data features

(Table 1) presents the distribution characteristics of economic variables using descriptive statistics. The distributions of the Housing Price Index (HPI) and Consumer Price Index (CPI) variables have negative skewness values, so the data are concentrated on the left, while the distributions of the Base Interest Rate (BIR), Mortgage Interest Rate (MIR), Deposit Interest Rate (DIR), Exchange Rate (ER), and Composite Stock Price Index (CSI) variables have positive values, so the data are concentrated on the right. And the minimum and maximum values of the Composite Stock Price Index (CSI) variable range from 1063.03 to 3296.68, so the Composite Stock Price Index has a greater influence on predicting the Housing Price Index (HPI) than other variables.

Table 1. Data features.

Descriptive statistics						
Variables	**Mean**	**Std. dev.**	**Median**	**Skewness**	**Min.**	**Max.**
HPI	85.43	9.70	84.85	-0.14	61.90	104.80
BIR	2.41	1.26	2.00	0.55	0.50	5.25
MIR	4.21	1.30	3.89	0.57	2.39	7.58
DIR	2.80	1.35	2.64	0.51	0.80	6.28
CPI	93.53	9.44	94.57	-0.12	75.23	113.26

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1628-1636, 2024
DOI: 10.55214/25768484.v8i5.1881

ER	1133.94	109.68	1130.76	0.23	915.86	1461.98
CSI	2043.72	435.24	2006.54	0.38	1063.03	3296.68

The correlation between variables is as shown in (Fig.1). In this result, the variables were found to be related to each other. Looking specifically at the results in row 1 in (Fig.1), the housing price index and the consumer price index, the housing price index and the exchange rate, and the housing price index and the comprehensive stock index have positive correlations, and the housing price index and the base interest rate, the housing price index and the mortgage loan interest rate, and the housing price index and the deposit interest rate have negative correlations. In particular, the correlation coefficient between the housing price index and the consumer price index was 0.95, and the correlation coefficient between the housing price index and the comprehensive stock index was 0.84, showing a strong positive correlation.

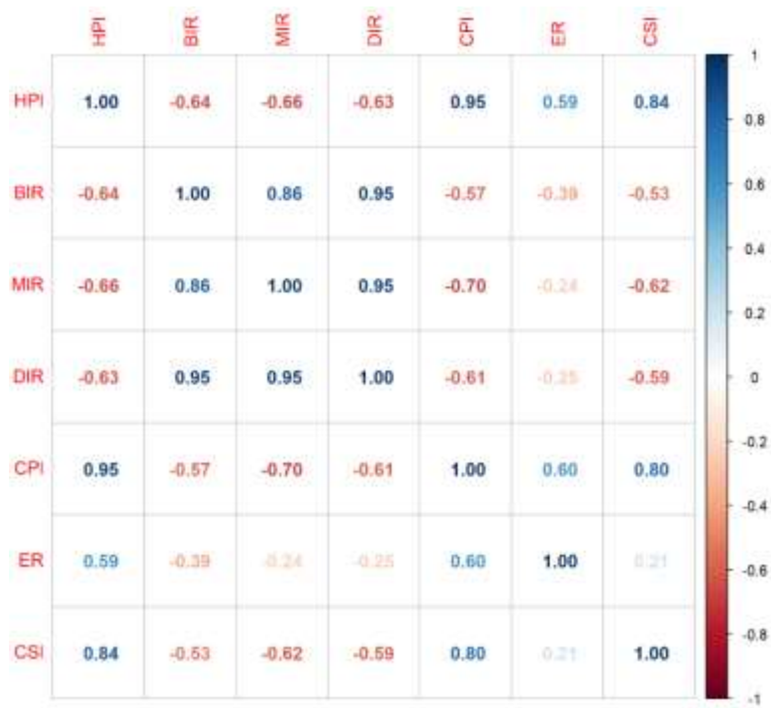

Figure 1. Pearson's multiple correlation coefficient matrix.

4.2. Model Performance Evaluation

The housing price index (HPI) prediction model was analyzed by fitting a deep neural network (DNN) model and a support vector regression (SVR) model. The deep neural network (DNN) model has two hidden layers, three nodes in hidden layer 1, and two nodes in hidden layer 2. The dataset used was a min-max normalized dataset, with 80% training data and 20% validation data. And the activation functions applied were soft plus, hyperbolic tangent (tanh), and rectified linear unit (ReLU). The results are as shown in (Table 2). In these results, when the soft plus activation function was applied, the correlation coefficient between the predicted value and the actual value of the housing price index was the highest at 93.42%, and the root mean square error (RMSE) was the lowest at approximately 0.1298. In deep neural network (DNN) models, the correlation coefficient is a very useful measure for predicting numerical data between predicted values and actual values.

Table 2. Deep neural network model evaluation.

Model performance evaluation: DNN		
Activation function	**Correlation coefficient**	**RMSE**
soft plus	0.9342424	0.129856
tanh	0.8362614	0.173967
ReLU	0.2048159	0.285159

In the support vector regression (SVR) model, the kernel function used was the Gaussian radial basis function RBF kernel, and the parameter Gamma, which adjusts the standard deviation of the Gaussian function for model tuning, and the parameter Cost, which adjusts the allowable limit of error in the model, were adjusted. The results of presenting the error between the predicted value and the actual value of the housing price index as the root mean square error (RMSE) are as shown in (Table 3). In these results, when C=100 and Gamma=0.1, the root mean square error was the smallest at approximately 0.2781. Comparing the root mean square error (RMSE) of the Deep Neural Network (DNN) model and the Support Vector Regression (SVR) model, the DNN model has a smaller RMSE, indicating that the DNN model is more suitable for predicting the housing price index.

Table 3. Support vector regression model evaluation.

Model performance evaluation: SVR			
Kernel function	**C**	**Gamma**	**RMSE**
RBF kernel	1	0.1	0.2824552
	10	0.1	0.2881567
	50	0.1	0.2808265
	100	0.1	0.2781816
	1	0.5	0.4219979
	10	0.5	0.4219119
	50	0.5	0.4167615
	100	0.5	0.4127387
	1	1.0	0.4443944
	10	1.0	0.4425561
	50	1.0	0.4396452
	100	1.0	0.4396452

4.3. Network Topology of Deep Neural Network (DNN) Model

The network topology of the fitted Deep Neural Network (DNN) model, including the input layer, hidden layers, output layer, and the weights and bias terms of each connection, is shown in (Fig. 2).

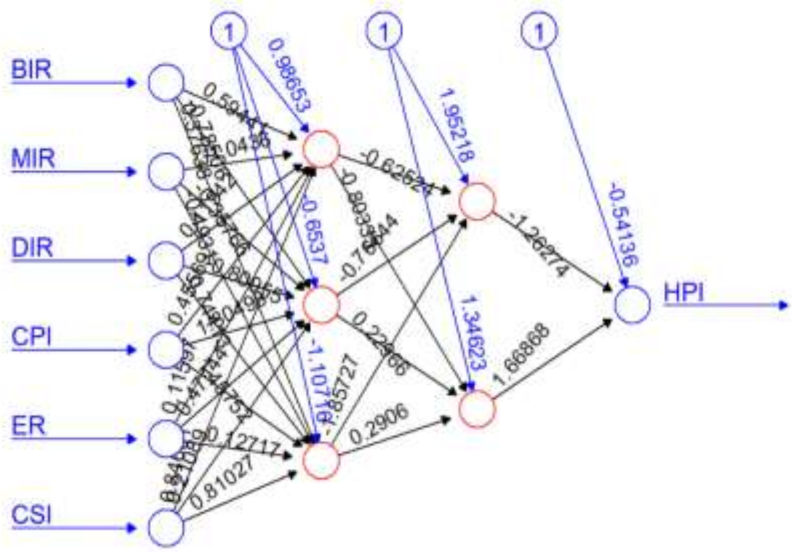

Figure 2. The network topology of the fitted Deep Neural Network (DNN) model.

4.4. Variable Importance of Deep Neural Network (DNN) Model

(Fig. 3) presents the effects of variables affecting the Housing Price Index (HPI) in the fitted deep neural network (DNN) model and whether they have a linear or nonlinear relationship, using generalized weights. In this result, the exchange rate (ER) has a small influence because the variance of the weights is close to 0, the consumer price index (CPI) has a nonlinear effect with a positive increase/decrease relationship because the variance of the weights is all greater than 1, and the base interest rate (BIR) has a linear effect with a negative increase/decrease relationship because the variance of the weights is all less than 1.

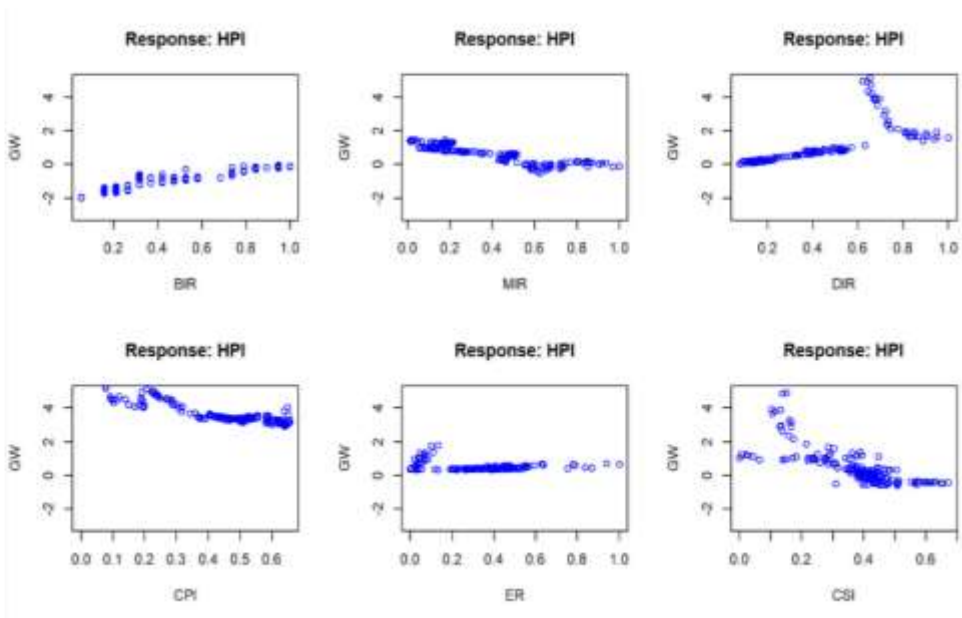

Figure 3. Generalized weights by variable.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1628-1636, 2024
DOI: 10.55214/25768484.v8i5.1881

5. Conclusion

In this study, we applied the machine learning methods of Deep Neural Network (DNN) and Support Vector Regression (SVR) models, using macroeconomic and microeconomic variables, to evaluate the fit of housing price index (HPI) prediction models based on the root mean square error (RMSE). The results of the model fitting are as follows.

The analysis results of the data features showed that the data did not follow a normal distribution, and the housing price index and the consumer price index had a strong positive correlation, and so did the housing price index and the comprehensive stock price index. In order to apply the machine learning method, min-max normalization was performed on the variables, and when the soft plus, hyperbolic tangent (tanh), and rectified linear unit (ReLU) activation functions were applied to the deep neural network model for prediction, the soft plus activation function yielded the lowest root mean square error, approximately 0.1298. And when using the Gaussian RBF kernel and adjusting the parameters C and Gamma in the Support Vector Regression model, the lowest root means square error, approximately 0.2781, was achieved with C=100 and Gamma=0.1. Therefore, based on the evaluation of model performance using the root mean square error, the Deep Neural Network prediction model, which exhibited a relatively smaller error, is more accurate. Additionally, the correlation coefficient between the predicted and actual housing price index values in the Deep Neural Network model was also high, at 93.42%. The results of this study demonstrate the prediction of the housing price index and the evaluation of model fit using machine learning models that exhibit excellent predictive performance, even with unstable data, by utilizing six domestic and international economic variables. Since housing, as an asset, holds significant financial value, these findings are expected to assist real estate market participants in their decision-making processes.

As major countries around the world maintain high-intensity monetary policies and high interest rates, and exports continue to slump, uncertainty in the domestic economy is becoming more apparent. Economic uncertainty is closely related to changes in national housing prices. Research on housing price prediction models using accumulated indicators that take into account domestic and international environments should continue, and the application of deep learning models, one of the methods for predicting complex problems more accurately, is left as a future research task.

Acknowledgments:

This Research was supported by Seokyeong University in 2024.

References

[1] 2024 KB Real Estate Report, KB Financial Group Management Research Center, (2024).

[2] Antipov, E. A. and Pokryshevskaya, E. B., Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics, Expert Systems with Applications, 39, (2012), 1772-1778. https://doi.org/10.1016/j.eswa.2011.08.077

[3] Shynkevich, Y., McGinnity, T. M., Sonya, A., Belatreche, A. and Li, Y., Forecasting price movements using technical indicators: Investigating the impact of varying input window length, Neurocomputing 264 (2017), 71-88. https://doi.org/10.1016/j.neucom.2016.11.095

[4] Bae, S. W. and Yu, J. S, Estimation of the Apartment Housing Price Using the Machine Learning Methods, Journal of the Korea Real Estate Analysts Association, 24 (2018), 69-85. http://dx.doi.org/10.19172/KREAA.24.1.5

[5] Shin, M., Zhang, B., Zhong, M. and Lee, D. J., Measuring international uncertainty: The case of Korea, Economics Letters,162 (2018), 22-26. https://doi.org/10.1016/j.econlet.2017.10.014

[6] Kim, J. Y., Connectedness between EPU Index and Korean Housing Market Returns, Journal of the Korea Real Estate Analysts Association, 26 (2020), 7-24, http://dx.doi.org/10.19172/KREAA.26.1.1

[7] Lee, B. S., Evaluation of Neural Network Model for Estimating Pile Load Capacity, Journal of Korean Society of Hazard Mitigation, 21 (2021), 221-228. https://doi.org/10.9798/KOSHAM.2021.21.5.221

[8] Kim, K. S., Kim B. H. and Han, K. Y., Performance Evaluation of Effective Drought Prediction Using Machine Learning, Journal of Korean Society of Hazard Mitigation, 21 (2021), 195-204.https://doi.org/10.9798/KOSHAM.2021.21.2.195
[9] Lee, J. y. and Ruy, J. P., Prediction of Housing Price Index Using Artificial Neural Network, Journal of the Korea Academia-Industrial cooperation Society, 22 (2021), 228-234. https://doi.org/10.5762/KAIS.2021.22.4.228
[10] Lee, S. W., Toun, J. Y. and Lee, J. M., An analysis of the effects of macroeconomic variables and real estate policy changes on apartment sales prices, Journal of the Korea Real Estate Management Review, 27 (2023), 31-54. https://doi.org/10.37642/JKREMR.2023.27.2
[11] http://ecos.bok.or.kr
[12] Ryu, J. P. and Shin, H. J., A Methodology for Hedging Equity Linked Warrant Using Artificial Neural Network, The Korea Academia-Industrial cooperation Society, 13 (2012), 1091-1098. https://doi.org/10.5762/KAIS.2012.13.3.1091
[13] Lee, J. H., Kim, J. S., Jang, H. W. and Lee, J. C., Drought Forecasting Using the Multi- Layer Perceptron (MLP) Artificial Neural Network Model, Journal of Korea Water Resources Association, 46 (2013), 1249-1263. http://dx.doi.org/10.3741/JKWRA.2013.46.12.1249
[14] Vapnik V., The Nature of Statistical Learning Theory. Springer, New York, (1995).