



# A New Iterative Neural Based Method to Spot Price Forecasting

Mehdi KHavaninzadeh<sup>1</sup>, Mohamad KHavaninzadeh<sup>2</sup>, Mohsen KHavaninzadeh<sup>3</sup>, Farshid Keynia<sup>4</sup>

<sup>1</sup>Tehran Area Operation Center, Tehran Regional Electric Company, Tehran, Iran

<sup>2</sup>MECO, MAPNA Company, Karaj, Iran

<sup>3</sup>Department of Electrical and Computer Engineering, WTIAU, Karaj, Iran

<sup>4</sup>Department of Energy Management and Optimization, Graduate University of Advanced Technology, Kerman, Iran

---

## Abstract

Electricity price predictions have become a major discussion on competitive market under deregulated power system. But, the exclusive characteristics of electricity price such as non-linearity, non-stationary and time-varying volatility structure present several challenges for this task. In this paper, a new forecast strategy based on the iterative neural network is proposed for Day-ahead price forecasting. For improved accuracy of prediction an intelligent two-stage feature selection is proposed here to remove the irrelevant and redundant inputs. In order to have a fast training the neural network normalization is vital, so in this paper the above technique is used. The proposed approach is examined in the Ontario electricity market and compared with some of the most recently published price forecast methods.

*Keywords:* Electricity price forecast; Artificial Neural Network; Feature selection; Normalization; Short term price forecasting;

© 2015 IAUCTB-IJSEE Science. All rights reserved

---

## 1. Introduction

All over the world in many countries, the power industry is moving towards a competitive market, and a pool based market environment is replacing the traditional centralized operation approach, a process that is known as restructuring. With the introduction of restructuring in the electric power industry, the price of electricity has become the most important factor of all activities in the power market.

The deregulated power market is a public sale market, and energy spot prices are volatile. Most users of electricity are, on short time scales, unaware of or indifferent to its price. These two facts drive the extreme price volatility or even price spikes of the electricity market [1-2].

Both the market regulators and players concern very much about the price evolution. Then, the market price prediction is an important information for the producers' production arrangement and bidding strategies, e.g.

For example, Unit commitment definition with regard to the minimum power constraints, or optimal

schedule for storage of hydro energy production with reference to the hydrology and the flexibility constraints of the thermal plants; as far as the bidding strategy, both the supply and the demand side need the price information to regulate their price submission to get more profits or hedge the bidding risk. Thus, the prediction accuracy greatly affects for the players' benefit. On the other hand, due to restructuring in electricity markets, price forecasting has become an important tool and price forecast is a challenging task and is very valuable in a competitive electricity market.

The complexity of electricity price forecasting on one hand, and it's important on the other hand, has motivated many research in this area. Stationary time series models such as auto-regressive (AR) [3], dynamic regression and transfer function [4-5], Auto-Regressive Integrated Moving Average (ARIMA) [6]. This approach can be very accurate, but it requires a lot of information, and the computational cost is very high [7]. Recently, non-stationary time series models

like generalized auto-regressive conditional heteroskedastic (GARCH), and wavelet transform and ARIMA models [8] have been proposed for this purpose. However, most of the time series models are linear forecasters, while electricity price is inherently a nonlinear function. So, the behavior of price series may not be completely captured by the time series techniques [9]. To solve this problem, some other research works have proposed Artificial Neural Networks (ANN) for price forecasting [10– 14], Input–Output Hidden Markov Models (IOHMM) [16], agents-based simulations [16] and Fuzzy Neural Network [17].

NNs have the Ability of modeling the non-linear input/output mapping functions. However, electricity price is a time variant signal and its functional relationships modify with time.

Between many available tools, ANN has received much attention because of its easy implementation, clear model and good performance. Also, there is not any such standards or/and rules to explain the relationship between price variations and other parameters such as weather conditions. Also, the data used in the training and testing of the price-forecasting model is usually uncertain and noisy, and the price forecast performance is sensitive to initial conditions such as historical temperature and load information [18-19].

In price forecasting applications, the main function of ANN is to forecast price for the next hour, day(s) or week(s).

It should be considered that the goal of this paper is to introduce a new forecasting method to minimize relevant market data and use the well - established forecasting methods to translate the from a practical point of view available market information into price signals.

Prediction strategy is proposed for day a head price forecasting of electricity markets. Also needs to know, an important task for forecasting methods based on neural networks is the optimal selection of inputs. Therefore, two stage feature selection for relevant input variables for electricity price forecast is proposed.

For training of selected data in NN in this paper iterative method is proposed. In an iterative method the output of each network is given to the input of other network until the training error becomes less than our desire value. And, in order to have a fast training the neural network, in this paper normalization is used. The remaining parts of the paper are organized as follows. In Section 2, Data-Driven Model Building is described. Section 3, is discussed feature selection technique. Section 4, concludes Normalization Procedures. Section 5, the proposed method is introduced and in Section 6, presents comparative results with conclusions. Finally, in section 7 references are listed.

## 2. Data-Driven Model Building

A data driven predictive model building has the three main steps: data preprocessing, feature selection, and model selection. This section will provide a brief description for each item:

### A) Data Pre-processing

Data preprocessing focuses on the initial treatment of data and includes the collection of information on data, statistics, missing values, anomalies and necessary data transformations. In the context of modeling electricity market price data, the reported studies highlight two aspects applicable to price data. First, is the problem of outliers, where prices do not follow the observed historical patterns [21]. Abnormal prices or outliers generally result from supply shortage or unexpected operating events like the forced outage of a generating unit.

Second, electricity prices are not changeless and show strong daily and weekly seasonality [22]. In order to get better stationary at the data, several data transformation approaches such as differencing, wavelet and Box-Cox transformations have been used [23-24]. However, stationary is not always a necessary condition, depending on the underlying assumptions of the used models; for example, neural networks are not limited to stationary data, but time series models are limited. In the present work, only data normalization is applied since it has been found to improve classification accuracy.

### B) Feature Selection

In this step, a proper subset of features (i.e., inputs or explanatory variables) is chosen from an initial feature set that efficiently captures patterns in the data. This will be discussed in detail in the next section. But it is noteworthy, In the context of forecasting electricity prices, the most popular features are historical load and price data. Other features such as hour and day indexes, load levels of neighboring systems, transmission constraints, temperature, variants of reserve margin, generator outages and availability of different types of generation resources, have also been offered with varying degrees of effectiveness [25-26].

### C) Model Selection

In the final step, a set of training instances is used to create a classification model that explains the available data and can be used to label future observations. Classification models can be categorized into logic-based, likelihood-based, perceptron-based and SVM-based approaches [27]. In logic-based models, forecasting is performed by setting some logical rules that are learned from a training set. Likelihood-based or statistical models, the prediction is done by constructing a probability

model based on the historical data. In perceptron-based models are driven on feed-forward neural networks in which the output is a follower of the weighted sum of the inputs. In SVM-based models, the fundamental idea is to determine separating hyper planes to identify different data class in a way that the hyper planes have the most possible distance from either of the data sets.

### 3. Feature Selection

Feature selection is a process generally used in machine learning, wherein a subset of the existing features of the data is chosen for the application of a learning algorithm [28]. Selecting the best set of input features is a crucial preprocessing for the successful application of neural networks. The main idea of feature selection is to select a subset of input variables by omission irrelevant features. Feature selection aims at identifying the most linked input variables within a data set [29]. It improves the performance of the forecasters by eliminating irrelevant inputs (and hence the noise), increased computational efficiency and achieves data reduction for accelerated training [30]. The consecutive set of features not only has a smaller dimension, but also has as much information as the original set. Feature selection can simplify the learning process of the prediction tool and raise its generalization capability for unseen data.

There are two common approaches for feature selection: wrapper and filter methods [31]. Wrapper methods are computationally expensive for data with a large number of features. In these methods, feature selection is wrapped around a learning method, the usefulness of a feature is directly measured by the estimated accuracy of the learning method [32], which is the case of electricity price forecast. Filter type methods are data filtering methods or essentially data preprocessing. In these methods, features are selected based on inherent characteristics, which specify their relevance to the target. In filters, the specifications of the feature selection are unrelated to those of the learning methods; so they have better generalization property [33].

A deficiency of the forenamed approaches is that the selected features could be correlated among themselves. This enhances the issue of redundancy of the selected feature set.

In [28] suggested performing the feature selection task in two stages: the first stage tries to recover all the relevant features, and the second stage, examining a much smaller feature subset, removes redundant features. In this paper, this idea has been used.

#### A) Statistical correlation analysis

This is a statistical technique that can show whether and how strongly pairs of variables are related. The most popular of these are the Pearson

correlation coefficient, which is sensitive only to a linear relationship between two variables. It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The population correlation coefficient  $\rho_{A,B}$  between two random variables A and B with expected values  $\mu_A$  and  $\mu_B$  and standard deviations  $\sigma_A$  and  $\sigma_B$  is defined as:

$$\rho_{A,B} = \text{corr}(A,B) = \frac{\text{COV}(A,B)}{\sigma_A \sigma_B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B} \quad (1)$$

Where E is the expected value operator, cov means covariance, and, Corr a widely used symbol replacement for Pearson's correlation.

The Pearson correlation is +1 for a perfect positive (increasing) linear relationship (correlation), -1 in the case of a perfect decreasing (negative) linear relationship (anti-correlation) [22], and some value between -1 and 1 in all other cases, expressing the degree of linear dependence between the variables. As it approaches zero, there is less of a communication (closer to uncorrelated). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient finds only linear dependencies between two variables. If we have a series of x measurements of A and B written as  $a_i$  and  $b_i$  where  $i = 1, 2, \dots, x$ , then the sample correlation coefficient can be used to estimate the population Pearson correlation r between A and B. The sample correlation coefficient is written:

$$r_{ab} = \frac{\sum_{i=1}^x (a_i - \bar{a})(b_i - \bar{b})}{(x-1)s_a s_b} = \frac{\sum_{i=1}^x (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^x (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^x (b_i - \bar{b})^2}} \quad (2)$$

Where a and b are the sample means of A and B, and  $s_a$  and  $s_b$  are the sample standard deviations of A and B. This can also be written as:

$$r_{ab} = \frac{\sum_{i=1}^x a_i b_i - x \bar{a} \bar{b}}{(x-1)s_a s_b} = \frac{x \sum_{i=1}^x a_i b_i - \sum_{i=1}^x a_i \sum_{i=1}^x b_i}{\sqrt{x \sum_{i=1}^x a_i^2 - (\sum_{i=1}^x a_i)^2} \sqrt{x \sum_{i=1}^x b_i^2 - (\sum_{i=1}^x b_i)^2}} \quad (3)$$

If a and b are results of measurements that contain measurement error, the realistic limits on the correlation coefficient are not -1 to +1 but a smaller range [34].

#### B) Mutual information

Mutual information (MI) technique without making any assumption about the nature of their

underlying relationships can measure the interdependency of random variables [34].

MI is one of many quantities that measure how much one accidental variables tells us about another. It is a dimensionless quantity with (generally) units of bits, and can be thought of as the decrease in uncertainty about one accidental variable given knowledge of another. High mutual information demonstrates a large reduction in uncertainty; low mutual information demonstrates a small reduction; and zero mutual information between two accidental variables means the variables are independent. Generally, the mutual information of two discrete accidental variables A and B can be defined as:

$$I(A,B) = \sum_{b \in B} \sum_{a \in A} p(a,b) \log\left(\frac{p(a,b)}{p(a)p(b)}\right) \quad (4)$$

Where  $p(a, b)$  is the joint probability distribution function of A and B, and  $p(a)$  and  $p(b)$  are the marginal probability distribution functions of A and B respectively.

In the case of continuous accidental variables, the summation is replaced by a definite double integral:

$$I(A,B) = \int_B \int_A p(a,b) \log\left(\frac{p(a,b)}{p(a)p(b)}\right) da db \quad (5)$$

Where  $p(x, y)$  is now the joint probability density function of A and B, and  $p(a)$  and  $p(b)$  are the marginal probability density functions of A and B respectively.

#### 4. Normalization Procedures

Neural network training could be made more efficient by doing certain preprocessing steps on the network inputs and targets. Network inputs, processing functions transform inputs into good form for the network user. The normalization process for the raw inputs have excellent effect on preparing the data to be appropriate for the training. Without this normalization, training the neural networks would have been very slow. There are many kinds of data normalization. It can be applied to scale the data in the same range of values for each input feature in order to diminish bias within the neural network for one feature to another. Data normalization can also accelerate training time by starting the training process for each feature within the same scale. It is especially effective for modeling application where the inputs are commonly on widely different scales. Different techniques can use different rules such as sum rule, min rule, max rule, product rule and so on.

In this paper, to normalize the price and the load the following formula is used [35-36], Which train is Input matrix that is obtained from the output of the

feature selection process and n train is a normalized train matrix.

$$n \text{ train} = \frac{\text{train}}{\text{Max}|\text{train}|} \quad (6)$$

#### 5. Proposed Method

The iterative is our main method in this paper. In this method that shown in "Fig.1," selected features are given to two neural networks. The output of each network is given to the other network as a set of input. The stopping criteria are the Mean Square Error (MSE). When MSE becomes less than  $\epsilon$  or the two outputs become almost the same as each other networks are trained.

A two stage Feature selection is used in this paper. At each stage different threshold for mutual coefficient can be selected. Different threshold selection results in a different number of input data to be selected. For example, in the first stage the lower the mutual coefficient the higher the number of input data, while in the second stage the vice versa is right. In order to have a better understanding of Mutual feature selection is recommended. As "Fig. 2," depicts a different threshold selection can cause a different input candidate to be selected. For example, by choosing 0.2 as first threshold the number of input data will be 16, but how the optimum number of input data can be selected. In order to overcome this difficulty different MAPE should be earned while different thresholds are examined. The best way to find the optimal threshold in the presence of minimum error is to plot three-dimensional shape. When the error is in its minimum value of three-dimensional plot both thresholds are gained. The optimum threshold can be seen in "Fig. 3".

#### 6. Case Studies

The Ontario electricity market is interconnected with the New England, Midwest, New York and PJM electricity markets. Generation companies and wholesale electricity consumers in the region can choose to sell or buy electricity either bilateral contracts or through the interconnected markets. Further, the demand-side entities may choose to supply their energy needs through on-site generation facilities. Given such a wide diversity of options, forecasting the electricity markets, prices are critical and an essential function of market participants to optimize their operations.

The proposed neural network approach is used to forecast Electricity hourly prices in the Ontario electricity market. Price forecasting is computed using historical data of year 2014. To evaluate the accuracy of the neural network approach in forecasting electricity prices, different scales are used. This accuracy is computed in function of the actual

market prices that occurred. The mean absolute percentage error — MAPE criterion, the sum squared error — SSE criterion, and the standard deviation of the error — SDE criterion.

In this paper, the following formula is used to check for errors:

$$MAPE = \frac{1}{N} \sum_{k=1}^N \frac{|L_{Actual}(k) - L_{Forecasted}(k)|}{L_{Actual}(k)} \quad (7)$$

Where, N is the forecast horizon,  $L_{Actual}(k)$  is the actual load of hour k, and  $L_{Forecasted}(k)$  is the load forecast of hour k. Also, as be shown in TABLE 1, we compare our proposed method with other methods in 6 weeks.

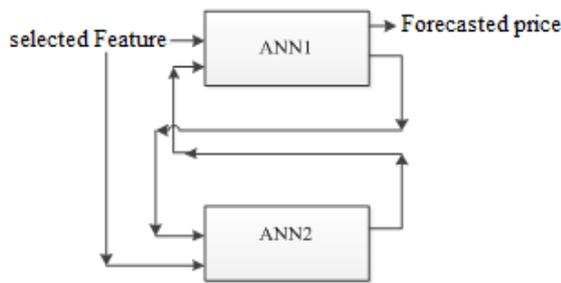


Fig. 1. Iterative structure of forecaster

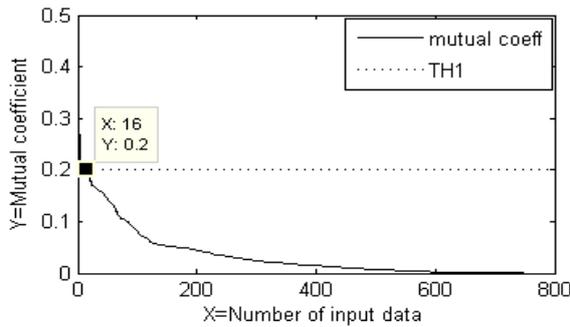


Fig. 2. Sorted mutual coefficient

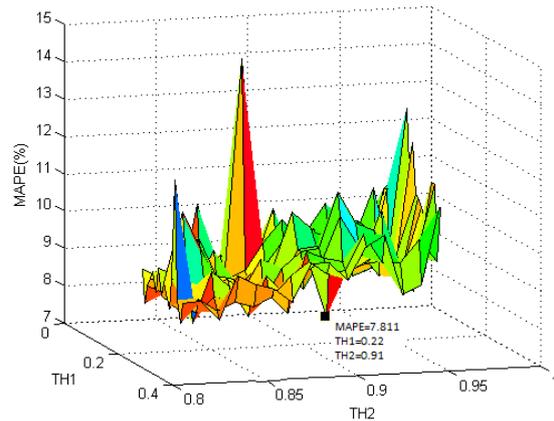


Fig. 3. Best threshold values for minimum test error

### 7. Conclusion

As a result of deregulation of the electricity market, the knowledge of the electricity prices is fundamental to manage properly the energy systems. This paper proposes a neural network approach to forecast next-week prices in the electricity markets in Ontario. The Iterative algorithm is used to train the network. The average errors with replacement of data are 10.90% in the weeks under study. The result shows that the NNs learned by normalization and iterative algorithm have far better forecasting methods that other proposed methods in TABLE 1.

Table.1.  
Weekly MAPE (%) for HOEP forecast in the Ontario electricity market

| Test week         | ARIMA        | Multivariate transfer function | Multivariate dynamic regression | Proposed Method |
|-------------------|--------------|--------------------------------|---------------------------------|-----------------|
| 26 April – 2 May  | 15.9         | 15.6                           | 15.9                            | 10.1            |
| 3 – 9 May         | 18.6         | 18                             | 18.1                            | 11.59           |
| 26 Jul – 1 August | 13.6         | 13                             | 13                              | 10.28           |
| 2 – 8 August      | 21.5         | 19                             | 19                              | 12.76           |
| 13 – 19 December  | 15.4         | 14.7                           | 14.7                            | 9.54            |
| 20 – 26 December  | 17.8         | 18.5                           | 18.5                            | 11.14           |
| <b>Average</b>    | <b>17.13</b> | <b>16.46</b>                   | <b>16.53</b>                    | <b>10.90</b>    |

### References

[1] K. Bhattacharya, M. H. J. Bollen and J. E. Daalder, "Operation of Restructured Power Systems", Kluwer Academic Publishers, Boston, 2001

[2] Website: <http://www.omip.pt/Downloads/SpotPrices/tabid/296/language/en-GB/Default.aspx>.

[3] Fosso OB, Gjelsvik A, Haugstad A, Birger M, Wangenstein I. Generation scheduling in a deregulated system The Norwegian case. IEEE Trans Power Syst 1999;14:75–81.

- [4] Zareipour H, Canizares CA, Bhattacharya K, Thomson J. Application of public domain market information to forecast Ontario's wholesale electricity prices. *IEEE Trans Power Syst* 2006;21:1707–17..
- [5] Nogales FJ, Conejo AJ. Electricity price forecasting through transfer function models. *J Oper Res* 2006;57:350–6.
- [6] Contreras J, Espinola R, Nogales FJ, Conejo AJ. ARIMA models to predict next day electricity prices. *IEEE Trans Power Syst* 2003;18:1014–20.
- [7] C.P. Rodriguez, G.J. Anders, Energy price forecasting in the Ontario competitive power system market, *IEEE Trans. Power Syst.* 19 (1) (2004) 366–374.
- [8] Conejo AJ, Plazas MA, Espinola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Trans Power Syst* 2005;20:1035–42.
- [9] Garcia RC, Contreras J, Akkeren M, Garcia JBC. A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Trans Power Syst* 2005;20:867–74.
- [10] Amjady N, Hemmati M. Energy price forecasting – problems and proposals for such predictions. *IEEE Power Energy Mag* 2006;4:20–9.
- [11] Yamin HY, Shahidehpour SM, Li Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *J Electr Power Energy Syst* 2004;26:571–81.
- [12] Hong YY, Hsiao CY. Locational marginal price forecasting in deregulated electricity markets using artificial intelligence. *Inst Electr Eng Gen Transm Distrib* 2002;149:621–6.
- [13] Mandal P, Senjyu T, Funabashi T. Neural networks approach to forecast several hours ahead electricity prices and loads in deregulated markets. *Energy Convers Manage* 2006;47:2128–42.
- [14] Szkuta BR, Sanabria LA, Dillon TS. Electricity price short-term forecasting using ANN. *IEEE Trans Power Syst* 1999;14:851–7.
- [15] Rodriguez CP, Anders GJ. Energy price forecasting in the Ontario competitive power system market. *IEEE Trans Power Syst* 2004;19:366–74.
- [16] Bunn DW. Forecasting loads and prices in competitive power markets. *IEEE* 2000;88:163–9.
- [17] Gonzalez AM, Roque AMS, Georgia-Gonzalez J. Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Trans Power Syst* 2005;20:13–24.
- [18] Amjady N. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Trans Power Syst* 2006;21:887–96.
- [19] Pavlos S. Georgilakis, “Artificial Intelligence Solution to Electricity Price Forecasting Problem”, *International Journal of Applied Artificial Intelligence*, Vol. 21, Iss. 8, 2007
- [20] YanBin Xu and Ken Nagasaka, “A research on the spike jump of Electricity price in the Deregulated Power markets”, *International Journal of Electrical and power Engineering*, vol. 3 (2), pp. 99-104, 2009 *Trans. Power Syst.* 18 (4) (2003) 1547–1555.
- [21] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, “Forecasting next-day electricity prices by time series models,” *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.
- [22] A. Lora, J. Santos, A. Exposition, J. Ramos, and J. Santos, “Electricity market price forecasting based on weighted nearest neighbor techniques,” *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1294–1301, Aug. 2007.
- [23] R. Weron and A. Misiorek, “Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models,” *Int. J. Forecast.*, vol. 24, no. 4, pp. 744–763, 2008.
- [24] R. Garcia, J. Contreras, M. Van Akkeren, and J. Garcia, “A GARCH Forecasting model to predict day-ahead electricity prices,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, May 2005.
- [25] C. P. Rodriguez and G. J. Anders, “Energy price forecasting in the Ontario competitive power system market,” *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 366–374, Feb. 2004.
- [26] G. Li, C. -C. Liu, C. Mattson, and J. Lawarree, “Day-ahead electricity price forecasting in a grid environment,” *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 266–274, Feb. 2007.
- [27] S. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
- [28] Nima Amjady a, Ali Daraeepour ‘Design of input vector for day-ahead price forecasting of electricity markets’ *Expert Systems with Applications* 36 (2009) 12281-12294
- [29] J.Y. Yang, S. Olafsson, Optimization-based feature selection with adaptive instance sampling, *Computers & Operations Research* 33 (11) (2006) 3088–3106.
- [30] S. Piramuthu, Evaluating feature selection methods for learning in data mining applications, *European Journal of Operational Research* 156 (2) (2004) 483–494.
- [31] K ohavi, R, & John, G. H. (1997). Wrappers for subset selection. *Journal of Artificial Intelligence*, 9 7 (1-2), 273 • 324.
- [32] [32] Guyon, I, & Weston, J. (2003). An introduction to the feature and variable selection. *Journal of Machine Learning Research*, 1157-1182.
- [33] Francis, DP; Coats AJ, Gibson D (1999). "How high can a correlation coefficient be?". *IntJCardiol* 69:185–199. .
- [34] Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley and Sons, Inc.]
- [35] Amjady, N., 2002. *Introduction to Intelligent Systems*. Semnan University Press, Semnan, Iran.