

The Application of Neural Networks in Balancing Production of Crude Sunflower Oil and Meal

Bojan Ivetić¹, Dragica Radosav²

¹ Oil factory Banat, Serbia

² University of Novi Sad, Tehnical faculty "Mihajlo Pupin", Zrenjanin, Serbia

Abstract - The aim of the research is to predict specific output characteristics of half finished goods (crude sunflower oil and meal) on the basis of specific input variables (quality and composition of sunflower seeds), with the help of artificial neural networks. This is an attempt to predict the amount much more precisely than is the case with technological calculations commonly used in the oil industry. All input variables are representing the data received by the laboratory, and the output variables except category % of oil which is obtained by measuring the physical quantity of produced crude sunflower oil and sunflower consumed quantity of the processing quality. The correct prediction of the output variables contributes to better sales planning, production of sunflower oil, and better use of storage. Also, the correct prediction of technological results of the quality of crude oil and meal provides timely response and also preventing getting rancid and poor-quality oil, timely categorizing meal, which leads to proper planning and sales to the rational utilization of storage space, allows timely response technologists and prevents the growth of microorganisms in the meal.

Keywords – neural networks, artificial intelligence, crude sunflower oil, balancing production,

1. Introduction

The tendency of man to create intelligent machines is several decades old. In all of the attempts that have been made until today the only available model that has been used was the nervous system of living beings, particularly humans. Because of that, it is naturally that all methods of artificial intelligence by solving certain problems in fact are trying to simulate the way in which the same problems are solved by a man, human, namely by his nervous system [1].

The development of artificial intelligence is certainly not motivated by a desire to make some kind of imitation of the man himself, but the reason lies in the superiority of the nervous system in solving the so-called poorly defined and ambiguous problems, even compared to the most advanced super computers. The reason probably is that the organization of the nervous system, actually of its processing units, is such that it is more appropriate to

the nature of these problems but problems such as numerical calculations. Unlike conventional computers nervous system contains a huge number of simple processing units (neurons) working simultaneously. The massive parallelism just makes the strength of the nervous system. Besides the aforementioned advantages, massive parallelism makes the nervous system less vulnerable to failures of individual neurons. Neurons in the brain die constantly, but globally speaking, it does not interfere with their functioning. Computer components must function flawlessly so the whole system could be operational. A collection of information, knowledge, is spread across the brain. Failure of one neuron does not mean that any of the information is lost, or at least not completely. In this way, it provides a certain redundancy of the nervous system, [2].

At the present time when the lifepace is faster than ever, the management of production systems is required to make more reliable decisions as soon as possible. The importance of making timely and valid decisions can result from inadequate models of management that are based on unreliable and incomplete data [3].

The application of artificial neural networks is exactly the subject of this paper. The goal is to see the possibilities of development of artificial neural networks in predicting the quantitative and qualitative indicators of crude sunflower oil and sunflower meal.

2. Problem, subject and purpose of the research

Probably most of the people who do not have much in common with the production of crude sunflower oil and meal, in terms of material balance considers the derivation of material balance irrelevant. In fact, at the entrance of the production facilities we have sunflower just as raw material, and as output products and by-products we get crude sunflower oil, sunflower meal and hulls. If we tried to make a material balance of processing we would be kept to the basic equation:

Amount of sunflower = amount of sunflower oil + amount of sunflower meal + amount of hulls + loss in water

This raises the question of the amount of produced sunflower and sunflower meal, and how much hulls do we get?

So, if everything was so simple, technology calculations would solve the problem on the basis of laboratory data on the quality of raw materials and oil and provide the information about the ratio between obtained quantities. The experience of the process shows that it does not go so easily.

In the following 2 tables the percentage of obtained quantities of oil and meal is shown by technological calculations and it is compared with the obtained physical quantities. This data is from processed crop in 2012 in Oil Factory Banat AD.

Table 1. Material balance of obtained products based on technological calculations

%	Drying plant	Hull section	Pressing section	Extraction section	total
oil			35.35	7.64	42.99
meal				37.66	37.66
water	2.69		1.86	-0.78	3.77
impurities	0.25				0.25
hulls		15.32			15.32
total	2.94	15.32	37.22	44.52	100.00

Table 2. Physical (real) material balance

Oil	40,95%
Meal	38,49%
water	4,19%
impurities	0,32%
hulls	16,05%

Comparing data presented in the previous two tables we could see significant differences. By now, the model of Oil Factory Banat, when it comes to sales and production plans, has been based on technological calculations. Results have been taken

with a grain of salt because the prediction of resulting amount has been bad enough. The market position of Oil Factory Banat AD has been significantly and negatively affected by this miscalculation. Also, orders arranged in advance had to be cancelled due to lack of oil. This situation has demanded a different approach to the forecast of the amount of products and by-products.

Another problem is to predict the quality of crude oil and meal. Namely, sales plans and further production are generated without an oil production plan and with no indications that meal sometimes does not meet all the requirements defined in standards and laws. The commercial service often plans out the sale of certain quantities of goods, excluding a potential quality problem that requires further refinement of crude sunflower oil and includes delivery for customers delayed. It also happens that customer requires certain quality which is above the prescribed standards and the management of Oil Factory Banat can not give the answer if these conditions are acceptable in terms of quality to be achieved for the simple reason-there is not a good prediction of quality of obtained product.

The solution to this situation is to predict the quality of sunflower oil and meal as precisely as possible based on the input parameters obtained by laboratory analysis. Neural networks are used for this purpose.

Based on everything said in the previous section, the case study is to predict the specific output characteristics of semi-finished products (crude sunflower oil and meal) on the basis of specific input variables (quality and composition of sunflower seeds), with the help of artificial neural networks.

This is an attempt to predict the quantity more precisely than is the case with technological calculations commonly used in the oil industry. All input variables are representing the data obtained through laboratory, and also the output variables except categories % of oil which is obtained by physical measurement of quantity of produced crude sunflower oil and expended quantity of sunflower, the processing quality. Categories were selected as the most important categories by leading technologists of Oil Factory Banat AD. Next sketch illustrates the subject of this paper the best:

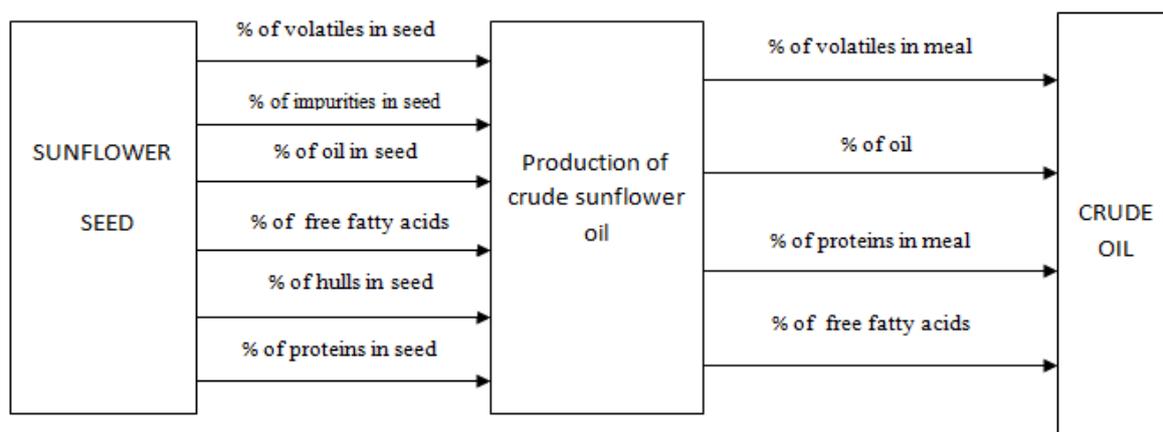


Figure 1. Purpose of research

Table 2. Reasons necessary to predict the output parameters

OUTPUT CHARACTERISTIC OF THE MODEL	WHY IS IT IMPORTANT TO PREDICT PROPERLY?
% of crude sunflower oil obtained in relation to the input quantity of processing quality of sunflower	<ul style="list-style-type: none"> • provides sales planning • planning further production • allows the rational use of storage space
% of free fatty acids (FFA) in sunflower crude oil	<ul style="list-style-type: none"> • provides timely response and preventing getting rancid and poor-quality oil
% of proteins in sunflower meal, which occurs as a by-product in the production of crude oil	<ul style="list-style-type: none"> • allows timely categorizing the meal, which leads to a proper sales planning and to a rational utilization of storage space
% of volatiles (moisture) in sunflower meal as an important category of material balance	<ul style="list-style-type: none"> • provides timely response and thus prevent the development of microorganisms in meal

3. Systematic approach to creating a neural model for prediction

Selecting data: In order to select data properly from a set of available data and eventually to form a neural model, first it is necessary to determine the purpose of the model. Determine the purpose of the model means to determine which variable the system wants, and also is able to predict considering the available data.

- Selection of input variables: Earlier it was thought that the artificial neural network were capable of determining the importance of specific variables independently for which the selection of

optimum input variables has not been paid much attention [3]. However, it was shown that the use of large number of input variables results in larger and more complex model, in addition to computational inefficiency has a tendency of "learning by heart", and it is easy to overtrain, [4]. That is the reason why at the entrance of the model are only six features, although the laboratory of Oil Factory Banat examines over 15 categories of quality of input seeds. Selecting the most important features is developed in cooperation with leading technologists of Oil Factory Banat AD.

- Data processing: data are used from information system of Oil Factory Banat. Data are stored in an Oracle database 11g Release 2 and are obtained by creating a number of SQL-s, as well as a base view.

Processing is done in the SQL Navigator tool version 6.1. Data prepared like this are exported in xls file suitable for importing into the tool for training the neural network. The prepared data could be found in the appendix of this paper. To process the input values in the form of preparation, training, testing the artificial neural network, a statistical software tool Statistica version 7.0. is used. Beside these operations, all the descriptive statistics, correlations, and all the charts and tables presented in this paper in the presenting research section are made by these tools.

- **Distribution of data:** Although the importance of correctly divided data due to the direct influence to the quality of the model emphasized earlier, according to the report, [3] the division of data is usually conducted by arbitrary or random distribution, not by considering the statistical properties of the sub-sets created division. This practice is not changed much even in the last decade, as can be seen in [5]. In this paper, data are most often distributed in the following structure: Number of samples to train: validation samples No. : No. of test samples = 50%: 25%: 25%

- **Choosing the network architecture:** Determine the neural network architecture means to determine the type of network that will be used to form the neural model. This step is probably the most important step in the neural model formation, because the other steps depend on this step and it is good to decide at the outset which architecture will be used. MLP two-layer neural network with the universal approximator capacity, [6], or with non-linear (sigmoidal) activation function in the hidden layer and linear activation function in the output layer, is the most used architecture for modeling this system. Although, according to [7] there are reports on the profitability of the use of more hidden layers due to the greater flexibility of the model and a smaller total number of parameters, there are also reports of network training with difficulties with two hidden layers due to more expressive problem of local minimum, [8] and is still proposed to MLP neural networks with one hidden layer. Beside MLP architecture, the RBFNN is also been used for flow modeling. The aim is to detect which one of two architectures provides better performance.

- **Adjusting the structure of the model:** As the number of input neurons is determined by dimensions of the input vectors in the third step, and a number of output neurons by dimension of the output vectors in the first step when using the MLP architecture with the status of universal approximator, determining the structure of the network is reduced

to determining the number of hidden neurons layer. The optimal number of hidden layer neurons is very important for proper operation of the model because it directly affects the general properties. An excessive number of neurons will form a network with too many parameters and will be inclined to overtraining (*overfitting*). On the contrary, the network with an insufficient number of neurons, or the parameters, will be unable to approximate the set of non-linear relations in the correct manner, so there is a problem called underfitting. Generally, the optimal number of hidden layer neurons is the minimum number of neurons, which enables the network to model the ratio between input and output data correctly. However, a method for accurate and reliable determination of minimum number of neurons required is not determined yet. What is possible to determine, in some extent, is the upper limit, or the maximum number of neurons in the hidden layer, which could be used for the modeling the system represented by a given set of data. In the operation software Statistica 7.0 modeled a number of models and chooses five of them to display. Practically, this work is solved by software. It is only necessary to determine the upper limit of the number of hidden neurons. In this paper 10 is the maximum number of neurons of the hidden layer.

- **Calibration of the model:** Model calibration is the process of setting up the model parameters and network training and carrying out one of the algorithms described in the theoretical part of the paper. In general, the goal of training is to find a set of values for adjustable parameters of the network for which the error or the objective function is minimal. As the objective function a different criteria of quality are used, and square errors MSE and SSE are commonly used.

- **Testing the model:** When it comes to examining the model or testing, actually the validation, it should distinguish testing the model for the time validation process, the identification of independent tests and the procedure of identification. Testing the model during the identification, validation, is usually implemented using single quality criteria and measures that are used as the objective function for model calibration, while for the model evaluation several different measures could and must be used. The most commonly used quality measures that are used as the objective function is the sum of squared errors and mean square error, while the assessment model uses a number of measures for evaluating the absolute and relative accuracy of the model. The following table shows the basic characteristics of the input and output parameters. Beside the number of samples, the average value is shown too, and also the minimum and maximum value, or the standard

deviation of the mean. Then histograms of output variables are given, in order to introduce the movements of individual categories. Display

indicates that the analysis was performed on 198 valid series. One series represented the average results within one day of work in production order.

Table 4. Descriptive statistics of input and output parameters

	Valid N	Mean	Minimum	Maximum	Std.Dev.
% of free fatty acids (FFA) in seed	197	0.80015	0.43000	2.31000	0.201316
% of oil in seed	198	46.09581	44.33000	47.99000	0.730974
% of impurities in seed	198	3.16384	2.00000	5.68000	0.517992
% of hulls in seed	198	27.08121	13.57000	28.44000	1.407219
% of volatiles in seed	198	6.05672	4.91000	7.38000	0.473297
% of proteins in seed	198	14.10247	12.18000	15.69000	0.621098
% of free fatty acids (FFA) in oil	198	0.90308	0.51000	1.60000	0.177650
% of obtained oil	198	42.98374	36.27000	48.93000	2.894969
% of proteins in meal	198	31.19944	27.66000	34.79000	1.328202
% volatiles in meal	198	10.92182	7.89000	12.69000	0.856198

4. A neural network for estimate the percentage of free fatty acids in crude sunflower oil

Summary results of learning are presented in the following table:

Table 5. Summary analysis of learning neural networks percentage SMK oil

Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/Members	Inputs	Hidden(1)
RBF 1:1-11-1:1	0.702852	1.52189	0.75539	4.56995	14.9216	5.67231	KM,KN,PI	1	11
Linear 3:3-1:1	0.859914	0.83580	0.87283	0.14877	0.2111	0.17322	PI	3	0
Linear 4:4-1:1	0.859808	0.83510	0.87089	0.14872	0.21104	0.17282	PI	4	0
MLP 3:3-3-1:1	0.800031	0.75505	0.74710	0.11073	0.15288	0.11852	BP100,CG20,CG0	3	3
MLP 3:3-4-1:1	0.823351	0.75391	0.73761	0.11396	0.15279	0.11706	BP100,CG20,CG0	3	4

The table above shows that the neural network is tested using five different algorithms. Common to all of it is that in the learning phase were 99 samples in the validation stage 49, as well as in the testing phase. The lowest level of the error was on the level of 0.117063, while the correlation was on the level of 75.39%. Training time was limited to 60s, but immediately after 5 seconds, the results achieved were similar to the final. Different algorithms have been tested. The best results showed the MLP 3:3-4-1:1. MSP is a general feed forward neural network and one of the most widely used NN algorithms. In order to optimize the error function used in conventional propagation backwards which was originally developed by Paul Williams, [9] it is a network with four neurons in the hidden layer. Graphical network representation is given by the following picture:

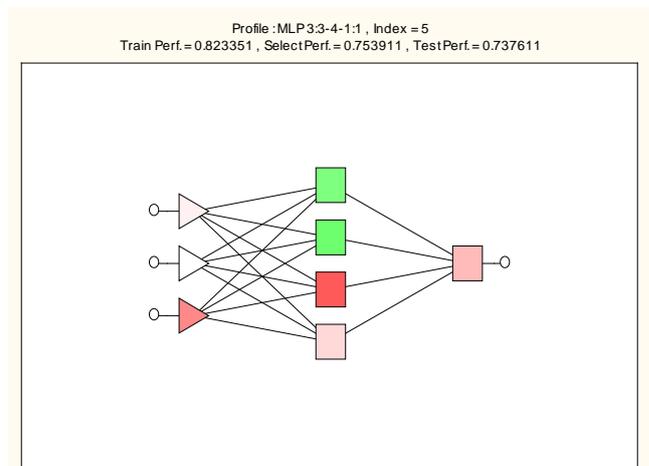


Figure 2. Show the accepted neural networks for the evaluation of ffa oil.

The achieved level of correlation accepted model MLP 3:3-4-1:1 of 75.39% is shown by the scatter diagram and the following figure:

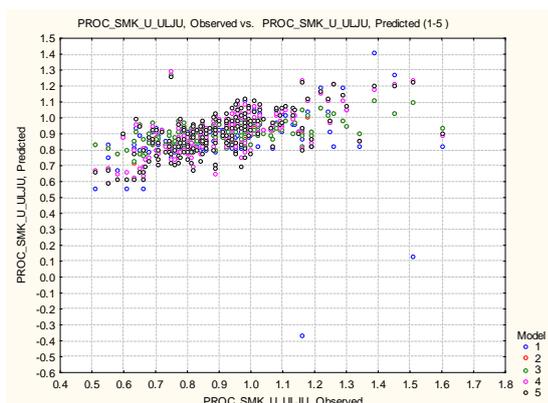


Figure 3 Diagram of scattering correlation between actual output and actual percentage of ffa oil

The results achieved are satisfactory, and the network created like that could be used to predict the category of the *percentage of free fatty acids in the oil*. This estimation could help greatly with further adjustment of the production process. Specifically it is important to know and predict the level of free fatty acids in the oil because they affect the rancidity or poor-quality of oil. If the level of free fatty acids is greater than 3% then it comes to a lower quality of crude sunflower oil.

5. Neural networks for estimating the percentage of generated amount of crude sunflower oil

The following neural network is created to estimate the amount of produced crude oil. The entrance to network consists of 6 standard input parameters. 5 different algorithms were tested, and again the MLP network showed the best results. Training time was limited to 1 minute, but after only a short period of time the error level was stabilized.

The best result was shown by the network MLP 6:6-8-1 where the level of error was 0.160857, unlike for example the RBF network where the level of error was almost 30%. Network MLP 6:6-8-1 was performed with 6 neurons in the hidden layer and achieved a high level of correlation with the right result in the amount of 93.18%.

The network showed the best results when the ratio of the samples was as follows:

The number of samples to train: the number of samples for the validation: number of test samples = 99:49:49

Reliability of the model is slightly lower and when use the data should be interpreted with caution.

Data are of great importance according to manipulation of the oil obtained and further plans for higher levels of processing.

Table 6. Summary analysis of learning neural networks for the percentage volume of crude oil produced.

Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/Members	Inputs	Hidden(1)
RBF 3:3-9-1:1	0.936765	0.944444	0.899990	0.315147	0.302636	0.296798	KM,KN,PI	3	9
Linear 2:2-1:1	0.980440	0.972105	0.965965	0.229506	0.212457	0.221706	PI	2	0
Linear 3:3-1:1	0.952136	0.972346	0.986815	0.222880	0.212021	0.226415	PI	3	0
MLP 6:6-7-1:1	0.874201	0.938441	0.886496	0.164933	0.163514	0.162703	BP100,CG20,CG0b	6	7
MLP 6:6-8-1:1	0.834007	0.931857	0.871915	0.156260	0.163201	0.160857	BP100,CG20,CG15b	6	8

6. Neural networks for estimating the percentage of protein in sunflower meal

Next created neural network is to estimate protein in sunflower meal. The entrance to the network consists of 6 standard input parameters. 5 different

algorithms were tested, and again the MLP network showed the best results. Training time was limited to 1 minute, but after only a short period the error level was stabilized.

Table 7. Summary analysis of learning neural networks for the percentage of protein in sunflower meal.

Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/Members	Inputs	Hidden(1)
RBF 5:5-11-1:1	0.889345	0.902434	0.831536	0.674545	0.676459	0.689139	KM,KN,PI	5	11
Linear 3:3-1:1	0.954915	0.986859	0.923879	0.175376	0.179401	0.183553	PI	3	0
Linear 1:1-1:1	0.975689	0.974582	0.918994	0.179191	0.179062	0.183911	PI	1	0
MLP 3:3-7-1:1	0.911686	0.890540	0.873342	0.133951	0.129572	0.140445	BP100,CG20,CG21b	3	7
MLP 4:4-5-1:1	1.089430	0.880875	0.862773	0.160252	0.127025	0.138617	BP100,CG20,CG0b	4	5

The best result was shown by the MLP network 4:4-5-1 where the level of error was 0.138617, unlike for example the RBF network where the error level was extremely high. MLP Network 4:4-5-1 was performed with 5 neurons in the hidden layer and achieved a high level of correlation with the right result in the amount of 88.09%.

The network showed the best results when the ratio of the samples was as follows:

The number of samples to train: the number of samples for the validation: number of test samples = 147:25:25

Graphical display of the optimal network is shown in the following figure:

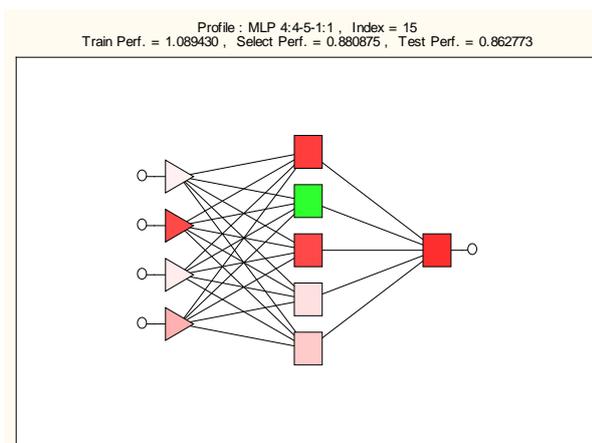


Figure 4. Graphical representation of neural networks for the percentage of protein in the meal with the most optimal

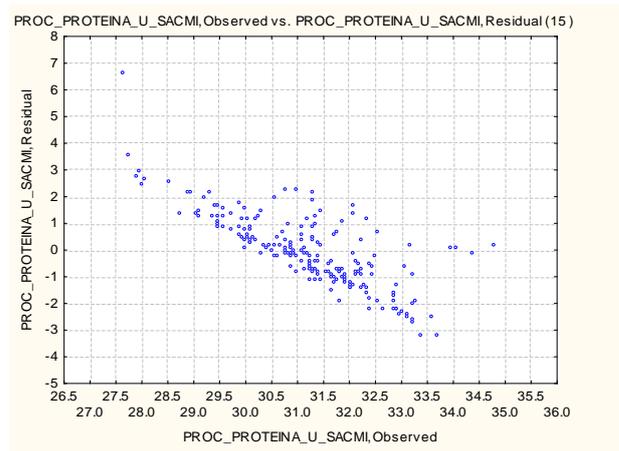


Figure 5 Diagram of scattering correlation between actual output and actual percentage of protein in the meal

Previous diagram shows a reliable level of correlation between the predicted output values and the true output values. That correlation level is higher in the case of low protein values in the meal.

The sensitivity analysis of output was also done here related to certain categories. The analysis showed that for the category % of proteins in meal the most important category is % of impurities in the processing quality of sunflower, and then % of free fatty acids in the processing quality sunflower.

The previous method of prediction could play a key role in the organization of production in the Oil Factory Banat AD. The category % of proteins in meal is crucial for further processing sunflower meal. If the meal has a higher protein level, a meal leakage is planned through sieves in order to eject the hulls and get meal with much higher protein value, and to be sold as a separate product.

Table 8 Summary analysis of learning neural networks for the percentage of moisture in sunflower meal

Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Training/Members	Inputs	Hidden(1)
RBF 1:1-4-1:1	0.875765	0.925767	0.93282	1.060099	1.321394	1.14587	KM,KN,PI	1	4
Linear 4:4-1:1	0.910496	0.872832	0.91248	0.162165	0.180676	0.16367	PI	4	0
Linear 5:5-1:1	0.909172	0.872150	0.93007	0.161929	0.180570	0.16674	PI	5	0
MLP 4:4-4-1:1	0.815169	0.808229	0.95811	0.116205	0.135292	0.13828	BP100,CG20,CG46b	4	4
MLP 5:5-7-1:1	0.812297	0.810510	0.95108	0.115746	0.135103	0.13732	BP100,CG20,CG21b	5	7

7. Neural networks for estimating the percentage of moisture in sunflower meal

Next created neural network is to estimate moisture content in sunflower meal. The entrance to network consists of 6 standard input parameters. 5 different algorithms were tested, and again the MLP network showed the best results. Training time was

limited to 1 minute, but after only a short period the error level was stabilized.

The best result was shown by the MLP network 5:5-7-1 where the level of error wa0.135103, unlike for example the RBF network where the error level was extremely high. Network MLP 5:5-7-1 was performed with 7 neurons in the hidden layer and

achieved a high level of correlation with the right result in the amount of 81.05%.

The network showed the best results when the ratio of the samples was as follows:

The number of samples to train: the number of samples for the validation: number of test samples = 157:20:20

8. Conclusion

Neural networks and artificial intelligence in general could contribute to better data processing. Nowadays, practice is to analyze data using statistical regression methods, but as shown in previous studies in most of the cases data could be analyzed much faster and better and eventually could reach more accurate results. The current success of neural network so far is expected to be used more in future and the statistical methods will lose the advantage of forecasting.

This paper is one of many that demonstrates important role of artificial neural networks in a situation where it is necessary to predict the output of a process that has not been fully explained or where classical mathematics can not give a good answer.

This study confirmed the hypothesis that it is possible to create a model with the support of artificial neural networks that can provide sufficient accuracy in the prediction of qualitative and quantitative properties of crude sunflower oil and sunflower meal during the production. It is also confirmed information that MLP neural network, in this case more accurately predict output of RBF neural network. The results may have an important role in organizing and planning of production, storage and sales of sunflower oil and meal, as well as in the prevention of getting poor-quality products. In order to improve the accuracy of the observation model, it would be necessary to revise the input variables and slightly increase their number, to test other intelligent and statistical methods to compare their accuracy in data, and to analyze the possibilities of an integrated approach by combining multiple methods for obtaining accurate predictive models, and in that purpose we could use different algorithms, artificial neural networks, decision trees, genetic algorithms, fuzzy logic or other intelligent methods.

References

- [1]. Milenković, Dr Srđan.(1997) *Veštačke neuronske mreže*. Beograd : Zadužbina Andrejević.
- [2]. Knight, K. (1990). Connectionist ideas and algorithms. *Communications of the ACM*, 33(11), 58-74.
- [3]. Lujić, R., Šarić, T., & Heffer, G. (2009). Application of expert system for determination of the most beneficial suppliers in single production. *Tehnički vjesnik*, 16(4), 81-86.
- [4]. Press, AFCEA International.(1988) *Neurall network*. pages 60.
- [5]. *Inteligentni sistemi*. [Online] [Accessed: 06 15, 2013.] <http://eris.foi.hr/11neuronske/nn-predavanje2.html>
- [6]. Koprivica, Marko (2010). *Supervizorni algoritami za učenje neuronskih mreža sa prostiranjem signala unapred*. Fakultet organizacioih nauka.
- [7]. Novaković, B., Majetić, D., & Široki, M. (1998). *Umjetne neuronske mreže.FSB, Zagreb*.
- [8]. Lojanica, Vladimir. *Uloga veštačkih neuronskih mreža u predviđanju*. [Online] 2001. [Accessed: 08 06, 2013.] <http://www.loya.5u.com/strane/13.html>.
- [9]. Tonković, Z., Zekić-Sušac, M., & Somolanji, M. (2009). Predviđanje potrošnje prirodnog plina pomoću neuronskih mreža. *Tehnički vjesnik*, 16(3), 51-61.

Corresponding author: Dragica Radosav

Institution: Technical faculty "Mihajlo Pupin" Zrenjanin

E-mail: radosav@tfzr.uns.ac.rs