



Energy Prediction of Wheat Production Using Data Mining Technique in Iran

Nasim Monjezi

Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Corresponding author-mail: n.monjezi@scu.ac.ir

Received 27 September 2020; Accepted 18 December 2020; Available online 5 February 2021

Abstract: Wheat is considered as one of the most important products in Iran. Concerning high cultivation area of wheat in Khuzestan, an instrument is required to process stored data in order to give information resulted from such processing to managers of agricultural sectors. Data mining technique is able to give essential information and models to producers of wheat for modelling energy consumption. One of the most practical algorithms is an artificial neural network. The main aim of this research is to predict output energy of wheat farms using a multilayer perceptron neural network. This is an analytic research and its database consists of 1240 records. Data required for the research was obtained from wheat farm during 2014-2018. Data analysis was done via IBM SPSS modeller 14.2 and standard CRISP. Concerning the model used in the research, it was found that variables of chemical fertilizers, machinery & diesel fuel with coefficients of 0.2987, 0.2064 and 0.1527 respectively had the highest effect on output variable (productive energy). Amount of prediction precision in neural network algorithm, meaning ratio of correctly predicted records to total records was 93.08%. Also, linear correlation between actual values and predicted values were 0.92 and 0.88 respectively, for training data and testing data suggesting strong correlation. The results obtained can be effective for wheat farmers in direction of evaluation and optimization of energy consumption in process of wheat production and reduction of consumption of energy inputs.

Keywords: Energy, Predict, Data mining, ANN, Wheat.

Introduction

One of factors that cause stable production of agricultural products is optimal energy consumption. In one hand, concerning increasing price of agricultural inputs, paying more attention to optimal management of energy consumption is very important. Wheat is not exceptional as well. Energy in process of wheat production has been always paid attention due to its importance in decision

making on application of different inputs. Therefore, optimization and modeling of energy consumption are very important in production of agricultural products, especially wheat. On the other hand, collected data are various with complex relationship and it is difficult to analyze and manage them via empirical and statistical analyses and experiences. Data mining is a powerful technology in management and organization of

a large size of information. Data mining means searching in one database for finding inter-data patterns. Several researches have been conducted on application of data mining techniques in agriculture. Ferraro *et al.* (2009) analyze large production database describing crop yield patterns. They studied the influence of several factors controlling sugarcane productivity in one of the most important area of cane production in Argentina. They propose using a data mining technique called classification and regression tree to identify the dependence of sugarcane yield on the variation of both environmental and management factors. This paper provides useful information regarding Genotype \times Environment interaction as it was possible to detect multiple paths for defining different sugarcane yield groups of crop fields using genotype, management and environmental variables.

Ekasingh *et al.* (2005) used the C4.5 data mining algorithm to model farmers' crop choice in two watersheds in Thailand. Result showed that the decision trees produced in this study could be useful to analyze of water resource management as they can be integrated with biophysical models for sustainable watershed management. Ramesh & Vardhan (2013) predicted agricultural products yield using different data mining techniques such as k-means, k-nearest Neighbor, support vector machines and artificial neural network. They wanted to find a model with high accuracy and ability for prediction of yield of agricultural products.

Jeysenthil *et al.* (2014) designed and predicted a support system for database of soil using data mining clustering technique (k-means). Everingham *et al.* (2009) in Australia and Fernandes *et al.* (2011) in Brazil have estimated the yield of farms using data mining

techniques. Goktepe *et al.* (2005) classified soils of Antalya region in 2005 using 120 soil samples via k-means and fuzzy clustering. Results of the research showed that fuzzy clustering method had more optimal result.

Medar & Rajpurohit (2014) present the various crop yield prediction methods using data mining techniques. Different Data Mining techniques such as K-Means, K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for very recent applications of data mining techniques in agriculture field. Mucherino *et al.* (2009) present some of the most used data mining techniques in the field of agriculture.

Some of these techniques, such as the k-means, the k nearest neighbor, artificial neural networks and support vector machines, are discussed and an application in agriculture for each of these techniques is presented. In addition, the references included in brackets confirmed the ability and necessity of using data mining technique in agriculture (Raorane & Kulkarni, 2013; Kalpana *et al.*, 2014; Geetha, 2015; Khedr *et al.*, 2015; Raorane & Kulkarni, 2015; Almaliki *et al.*, 2016; Almaliki, 2017; Oliveira *et al.*, 2017; Thomas, 2017). Therefore, the energy used in wheat production was predicted and modelled using data mining techniques (Artificial Neural Network).

Materials & methods

The study site

The present research was conducted on wheat farms of Khuzestan province in Iran. Using data in 2014-2018. Khuzestan province with an area 64055 km² is located in southwest of Iran (within 29° 58' and 33° 04' N and 47° 41' and 50° 39' E). Fig. (1) showed the position of Khuzestan province in Iran.



Fig. (1): The position of Khuzestan province in Iran.

Energy equivalents of inputs and output

The data included the quantity of various energy inputs used per hectare of wheat production, including: human labor, machinery, diesel fuel, chemicals, fertilizer, water for irrigation and electricity and seed, and the production yield as output. In order to analyze the energy of farmers, all of inputs and output were then converted into energy equivalents by multiplying the quantity of input use with their corresponding energy equivalent coefficients. The energy coefficients of inputs are the energy used from primary production to the end user. The energy equivalents of inputs were presented in table (1). The energy equivalent of human labor is the muscle power used in field operations of crop production. Chemicals and chemical fertilizers energy equivalents means the energy consumption for producing, packing and distributing the materials and they are given on an active ingredient basis. Also, the energy sequestered in diesel fuel and electricity means their heating value (enthalpy) and the energy

needed to make their energy available directly to the farmers (Mousavi-Avval *et al.*, 2011).

The energy equivalent of water for irrigation input means indirect energy of irrigation consist of the energy consumed for manufacturing the materials for the dams, canals, pipes, pumps, and equipment as well as the energy for constructing the works and building the on-farm irrigation systems (Khan *et al.*, 2008). For calculating the embodied energy in agricultural machinery it was assumed that the energy consumed in the production of the tractors and agricultural machinery be depreciated during their economic life time (Beheshti Tabar *et al.*, 2010); therefore, the machinery energy input was calculated using the following Eq. (Srivastava *et al.*, 1993):

$$E_m = \frac{EI_m \cdot W \cdot T}{N} \quad (1)$$

Where E_m is the machinery energy per unit area (MJ ha^{-1}); W is the machine mass (kg); EI_m the production energy of machine (MJ kg^{-1}); T is the time that machine used per unit

Table (1). Energy coefficient equivalents of inputs and outputs.

Inputs/outputs	Unit	Energy coefficient equivalent (MJ unit ⁻¹)	Reference
Diesel Fuel	L	47.8	(Banaeian & Zangeneh, 2011)
Human Labor	h	1.96	(Mandal <i>et al.</i> 2002)
Chemical Fertilizers			
Nitrogen (N)	kg	47.1	(Kaltschmitt <i>et al.</i> 1997)
Phosphate (P ₂ O ₅)	kg	15.8	(Kaltschmitt <i>et al.</i> 1997)
Chemical herbicide	kg	238	(Kaltschmitt <i>et al.</i> 1997)
Machinery	Kg per year	9	(Khoshnevisan <i>et al.</i> 2013)
seed	kg	14.7	(Mandal <i>et al.</i> , 2002)
Electricity	kWh	11.93	(Kitani, 1999)

area (h ha⁻¹) and N is the economic life time of machine (h).

Data mining technique

Generally, data mining technique is divided into predictive and descriptive techniques (Fig. 2). The neural network is one of the most practical data mining techniques in prediction.

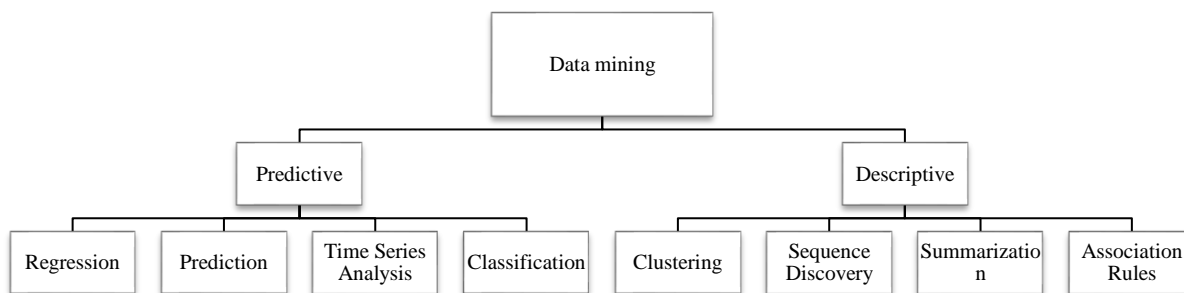


Fig. (2): Data mining techniques (Sharma 2006; Ayman *et al.* 2015)

Artificial neural networks

Neural networks can estimate approximately any continuous correlation between input and target even if such relationship is non-linear (Phillips-Wren *et al.* 2008). The most important feature of neural networks that is important in their application is their ability in learning samples (Efendigil *et al.* 2009). Learning method can be in three forms: with or without observer and both of them. Learning method with observer is conducted by pairs of input-output patterns while learning without

observer includes offering training samples to the network and it continues until the highest separation is obtained among classes of samples. The most famous neural networks are multilayer perceptron neural networks that are trained with observer or error-back propagation algorithm (Liao & Wen, 2007).

Standard CRISP-DM

This is a data-oriented research. Its main basis is to discover knowledge from database of wheat farms of Khuzestan province in Iran. Therefore, cross- industry standard process for

data mining (CRISP-DM) was used in the research. This method presents a process model for data mining that reviews life cycle of each data mining project. Life cycle of each data mining project includes six stages of problem perception, data perception, data

preparation, modelling, result evaluation and application of the model. Fig. (3) shows stages of this standard in the form of a process.

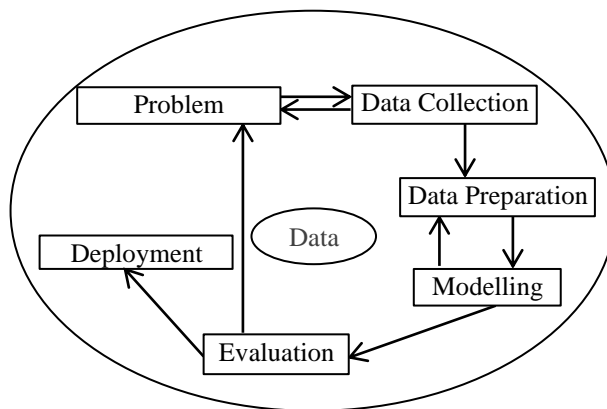


Fig. (3): Data mining process model (Chapman *et al.*, 2011).

Data analysis

Data mining software of IBM SPSS Modeler 14.2 has been used in the research for modelling neural network and result validation. The data were divided into training and testing data. 70% of data were training and 30% of them were testing data. Research stages have been indicated in fig. (4).

Model assessment

In this stage, output model of training data was used for testing data. In other words, testing data (30% of data in the research) are predicted by derived model and results of prediction are compared with reality using this model.

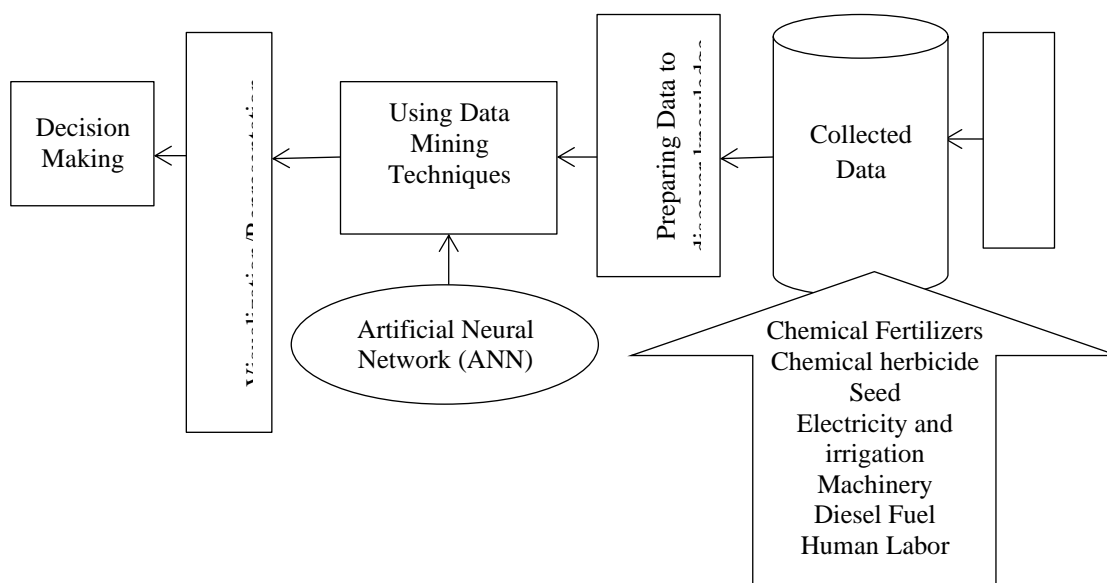


Fig. (4): Proposed framework.

Results & Discussion

Analysis of energy input and output in wheat production

In the first research phase, wheat farms should be studied in order to investigate energy consumed in farms.

Input energies based on inputs consumed in farms

Table (2) showed input energies in production of wheat farms. As seen in table (2), between consumed inputs, chemical Fertilizers 40723.67 MJ ha⁻¹ (62.92%) energy has been determined as the first energy input for production of wheat in farms. On average, total consumed energy in process of wheat production in farms was 64720.53MJ ha⁻¹ of which 62.92% associate with inputs of

chemical fertilizer. In another study in Iran, reported by Houshyar *et al.* (2010), total energy input and chemical fertilizer, in wheat farming were 38817.82 MJ ha⁻¹ and 12347.44 MJ ha⁻¹, respectively. In another study in Iran, Shahin *et al.* (2008) found that total energy input and chemical fertilizer in wheat production were 38356.39 MJ ha⁻¹ and 14749.51 MJ ha⁻¹, respectively. In another study in India, Nassiri *et al.* (2009) noted that total energy input and chemical fertilizer in wheat production were 15261 MJ ha⁻¹ and 8004 MJ ha⁻¹, respectively.

Output energy of wheat farms

Product yield for farms has been summarized in table (3).

Table (2): Input energy in wheat farms.

Input	Average (MJ ha ⁻¹)	Percent (%)
Human Labor	2078.03	3.21
Diesel Fuel	9304.44	14.38
Electricity and irrigation	3741.05	5.78
Machinery	4532.11	7.00
Chemical Fertilizers	40723.67	62.92
Chemical herbicide	2390.42	3.69
Seed	1950.81	3.01
Total input energy	64720.53	100

Table (3): Yield of farms.

Farm	Yield (ton ha ⁻¹)	Output energy (MJ ha ⁻¹)
Wheat	3.12	45864

Product yield has been 3.12 tons of wheat per hectare in farms of Khuzestan province. Concerning degree of harvested wheat and

energy of wheat per kilogram, output energy of farms was obtained and it is equal to 45864 MJ ha⁻¹.

Data mining

Information used in this research was collected from farms of Khuzestan province. 155 farms (where wheat was produced between 2014 and 2018) were used as database. Executive stage of the second phase of the research (data mining) was divided into six stages of data identification and selection, data description, data pre-processing and preparation, data analysis and modeling, model evaluation and validation and model development.

Stage of data identification and selection

In the first stage, information available about consuming inputs into process of wheat production has been studied. Concerning collected information, seven effective variables required for modeling energy consumption in process of wheat production were defined and stored in an informational bank of farms as separate fields. These seven variables are independent variables of the research and they are as follows: human labor,

diesel fuel, electricity and irrigation, machinery, chemical fertilizers, chemical herbicide and seed. Variable of output energy of the farm is dependent variable of the research.

Data description and definition of variables used in the model

Among seven input variables in neural network, seven variables were previously defined in energy analysis (the first stage of research).

Data used in this research is entered in IBM SPSS modeler 14.2 in form of an Excel file with 8 columns including seven independent variables (input) and one dependent variable (target) and descriptive information of variables has been shown in table (4). Table (4) includes type of variable, minimum and maximum amounts of the variable, variable mean, standard deviation and number of valid records in database of each variable.

Table (4): Description of all wheat variables used for this study.

Variable	Variable type	minimum amount	maximum amount	Average	Standard deviation	number of valid records
Chemical Fertilizers	Continuous	30093.720	317759.29	40606.37	22582.104	155
Electricity and irrigation	Continuous	3004.32	4987.35	3586.95	339.29	155
Chemical herbicide	Continuous	1444.38	12400.70	2289.67	863.93	155
Seed	Continuous	2940	3675	3180.89	320.87	155
Machinery	Continuous	1000.76	41397.80	4283.78	3123.98	155
Diesel Fuel	Continuous	6004.65	96347.11	9254.20	7108.81	155
Human Labor	Continuous	1100.28	2299.56	2045.24	272.47	155
Output Energy	Continuous	54235.00	64668.00	59864.17	1701.35	155

Stage of data pre-processing and preparation

Before modeling, raw data collected from wheat farms were prepared and pre-processed for improvement of data quality. Data pre-processing and preparation include standardization of unique values in different data sources, correction or deletion of inconsistent and irrational values and addition of different sources based on the number of standardized farm. Also, presence of unconventional records in database increased error factor in results of data mining. Therefore, identification and management of such unconventional data is very important in data preparation stage. Concerning statistical data description, there is no missing data among records in this stage.

Stage of data analysis and modeling

Data are modelled in this stage and necessary analyses are presented regarding model validation. In order to predict output energy of wheat farms and to enter the software, final database had 1240 records and 8 fields of input and output variables. In order to model data, field of output energy variable was introduced to the software as output and other variables (fields) were introduced to the software as input variables. The main algorithm used in this research for energy of wheat farms is artificial neural network. The network used here is multilayer perceptron which is one of the most practical artificial neural networks. Number of hidden layers of the network is one layer with three neurons (Fig. 5).

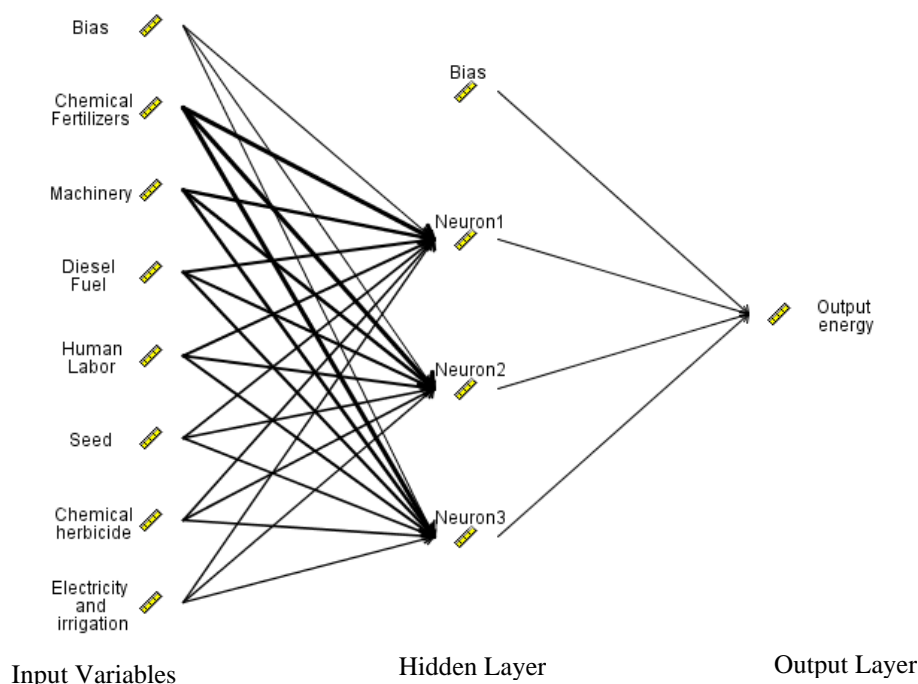


Fig. (5): Architecture of a multilayer perceptron.

The condition for stopping modeling is that if error is not reduced, optimization will be stopped. In present model, data prediction precision was 93.08%. Concerning description of input and output variables, it is noteworthy

that some variables are more effective on prediction of the model. In the graph of neural network in fig. (6), descending trend of effect of variables has been determined by target variable. The first variable with the highest

effect is chemical fertilizers. Concerning Table (5), variables of chemical fertilizers, machinery & diesel fuel with coefficients of 0.2987, 0.2064 and 0.1527 respectively have

the highest effect on output variable (productive energy). Other variables are effective as well but their effectiveness is less than the three variables.

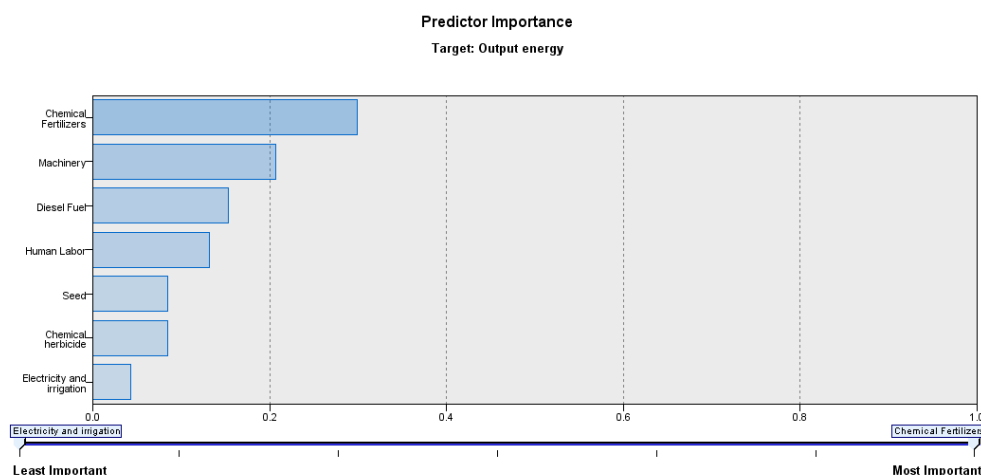


Fig. (6): Predictor importance (Target: output energy)

Table (5): The importance order of variables in prediction of target variable.

Variable	coefficient
Human labor	0.1315
Seed	0.0842
Chemical herbicide	0.0842
Machinery	0.2064
Electricity and irrigation	0.0423
Diesel fuel	0.1527
Chemical fertilizers	0.2987

Model evaluation and validation

After preparation of model, it should be evaluated. Evaluation results cause improvement of model and they make it practical. Since the method presented in every research should be measured based on validity and concerning that the research method is data-oriented, validation method is as follows: data are divided into two training and testing data collections. Training data makes the model and testing data evaluates the model. In other words, testing data were predicted by extracted model and prediction results are compared with reality by this model.

Data were divided randomly by software. Regarding number of data in each set, number of training data is always more than testing data. In present research, number of training sets is 70% (109 farms) and the rest (30%) has been considered as testing data (46 farms) (Fig. 7). Validity value is tested by results of new data and training data are entered the algorithm as observer and evaluate results of its accuracy. Amount of prediction precision in neural network algorithm meaning the ratio of correct predicted records to total records was 93.08% (Fig. 8). Therefore, the model made by neural network has high estimative precision. Another argument is how precision of

prediction methods can be measured. In order to evaluate precision of prediction models (the difference between the real value and predicted value of dependent variable), index of mean absolute value of prediction error is used. Mean absolute error value (MAE) of training data and testing data is 1368.09 and 1650.39 respectively which is an optimal value (Table

6). It is clear that in order to increase precision of a predictive method, the values of above mentioned indices should be small. Also, linear correlation between real and predicted values is 0.92 and 0.88% for training data and testing data respectively and it suggests a strong correlation.

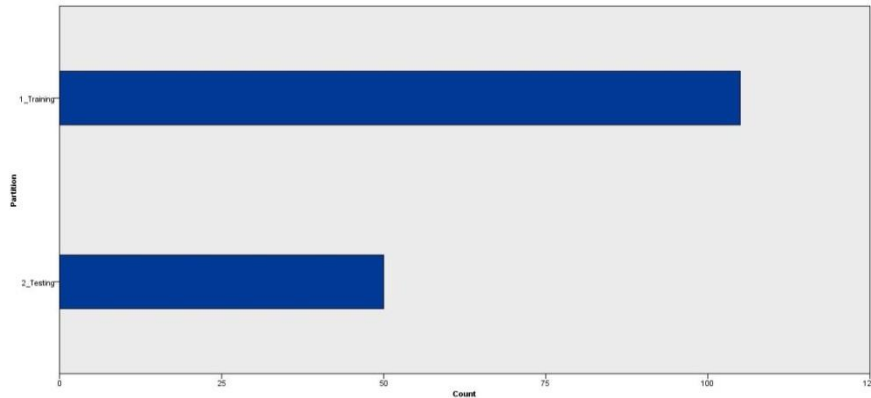


Fig. (7): Ratio of training and testing data in the neural network model.

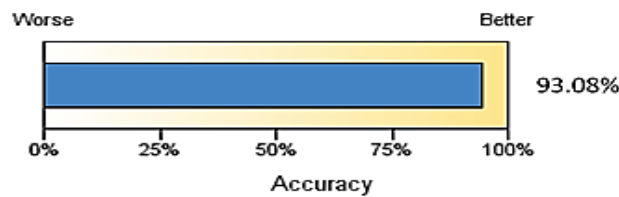


Fig. (8): Amount of predictive precision in neural network algorithm.

Table (6): Comparing predicted output energy with actual output energy.

Partition	1_Training	2_Testing
Minimum Error	-2390.774	-4870.99
Maximum Error	3782.924	5291.23
Mean Error	109.245	217.730
Mean Absolute Error	1368.098	1650.398
Standard Deviation	1023.45	1458.93
Linear Correlation	0.923	0.885
Occurrences	109	46

Model development

In this stage, the model and its results can be presented to farmers to be used in future for

prediction of energy production during the process of wheat production. Therefore, after necessary reports, it was explained the most effective variables on prediction of output

energy of wheat farms based on this model. Concerning that variables of chemical fertilizers, machinery & diesel fuel are the most effective variables on output energy of wheat farms, energy efficiency will be increased in farms if input variables are optimized.

Conclusion

Data of wheat farms of Khuzestan province was used in present study. In the first research phase, input and output energies of farms were analyzed. In the second research phase (data mining), database was divided into training and testing parts. Then, a model was created based on artificial neural network technique of data mining in order to estimate energy in process of wheat production using IBM modeler 14.2 and training dataset. As a result, the model was assessed using testing datasets and could reach 93.08% precision for estimation of output energy. It is clearly evident that this model has a good accuracy for estimation of amounts of output energy. Also, results of the study indicate that prediction methods provide proper source allocation and increase efficiency of inputs by presenting more accurate picture of energy status in wheat farms. It is suggested to use other techniques available for data mining in future studies and they are compared with results of the present research, exploration of rules for including input and output variables in database via farma can help considerably data mining in order to extract more real results from them. Concerning high potential of data storage and advanced analysis in wheat farms, it is suggested to collect systematically and mechanize data by creating proper software infrastructures, data warehouse with aim of using reporting and analytic tools. Finally, concerning that data of this research belongs to one province, it is suggested that a similar research is conducted on other provinces and

results are compared with results of the present research and other methods.

Acknowledgement

This paper was supported by Shahid Chamran university of Ahvaz, Iran.

Conflict of interest: The authors declare that they have no conflict of interest.

ORCID:

N. Monjezi: 0000000182297706

References

- Almaliki, S., Alimardani, R., & Omid, M. (2016). Artificial neural network based modeling of tractor performance at different field conditions. *Agricultural Engineering International: CIGR Journal*, 18, 262-274. <https://cigrjournal.org/index.php/Ejournal/article/view/3880>
- Almaliki, S. (2017). *Development and evaluation of models for MF-285 tractor performance parameters using computational intelligence Techniques*. Ph. D. Thesis. University of Tehran, 215pp.
- Ayman, E.K.; Kadry, M., & Walid, G. (2015). Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector. *Procedia Computer Science*, 65, 633-642. <https://doi.org/10.1016/j.procs.2015.09.007>
- Banaeian, N., & Zangeneh, M. (2011). Modeling energy flow and economic analysis for walnut production in Iran. *Research Journal of Applied Sciences, Engineering and Technology*, 3, 194-201. <https://www.researchgate.net/publication/268341453>
- Beheshti Tabar, I., Keyhani, A., & Rafiee, S. (2010). Energy balance in Iran's agronomy (1990–2006). *Renewable and Sustainable Energy Reviews*, 14, 849-55. <https://doi.org/10.1016/j.rser.2009.10.024>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2011). *CRISP-DM 1.0: Step-by-step data mining guide*. Viewed 22 October 2011. <http://www.crisp-dm.org/CRISPwP-0800.pdf>;
- Efendigil, T., Onut, S., & Kahraman, C. (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A

- comparative analysis. *Expert Systems With Applications*, 36, 6697-6707. <https://doi.org/10.1016/j.eswa.2008.08.058>
- Ekasingh, B., Ngamsomsuke, K., Letcher, R., & Spate, J. (2005). A data mining approach to simulating farmers' crop choices for integrated water resources management. *Journal of Environmental Management*, 77, 315-325. <https://doi.org/10.1016/j.jenvman.2005.06.015>
- Everingham, Y. L., Smyth, C. W., & Inman-Bamber, N. G. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production. *Agricultural and Forest Meteorology*, 149: 689-696. <https://doi.org/10.1016/j.agrformet.2008.10.018>
- Fernandes, J. L., Rocha, J. V., & Lamparelli, R. A. C. (2011). Sugarcane yield estimates using time series analysis of spot vegetation images. *Scientia Agricola*, 68, 139-146. <https://doi.org/10.1590/S0103-90162011000200002>
- Ferraro, D. O., Rivero, D. E., & Ghersa, C. M. (2009). An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. *Field Crops Research*, 112, 149-157. <https://doi.org/10.1016/j.fcr.2009.02.014>
- Geetha, M. C. S. (2015). A Survey on data mining techniques in Agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, 3, 887-892. <https://link.springer.com/article/10.1007/s12351-009-0054-6>
- Goktepe, A. B., Altun, S., & Sezar, A. (2005). Soil clustering by fuzzy C-Means algorithm. *Advances in Engineering Software*, 36, 691-698. <https://doi.org/10.1016/j.advensoft.2005.01.008>
- Houshyar, E., Sheikh Davoodi, M. J., & Nassiri, S. M. (2010). Energy efficiency for wheat production using data envelopment analysis (DEA) technique. *Journal of Agricultural Technology*, 6, 663-672. <https://www.researchgate.net/publication/235654571>
- Jeysenthil, K. M. S., Manikandan, T., & Murali, E. (2014). Third generation agricultural support system development using data mining. *International Journal of Innovative Research in Science, Engineering*, 3, 9923-9930. http://www.ijirset.com/upload/2014/march/6_Third.pdf
- Kalpana, R., Shanthi, N., & Arumugam, S. (2014). A survey on data mining techniques in agriculture. *International Journal of Advanced Computer Science and Technology*, 3, 426-431. <http://warse.org/pdfs/2014/ijacst05382014.pdf>
- Kaltschmitt, M., Reinhardt, G. A., & Stelzer, T. (1997). Life cycle analysis of befouls under different environmental aspects. *Biomass. Bioenergy*, 12, 121-134. <https://doi.org/10.1002/ese3.315>
- Khan, M. A., Zafar, J., & Bakhsh, A. (2008). Energy requirement and economic analysis of sugarcane production in Dera Islamic Khan district of Pakistan. *Gomal University Journal of Research*, 24, 72-82. <https://www.researchgate.net/publication/262048156>
- Khedr, A. E., Kadry, M., & Walid, G. (2015). Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector applied case on food security information center ministry of agriculture, Egypt. *Procedia Computer Science*, 65, 633-642. <https://doi.org/10.1016/j.procs.2015.09.007>
- Khoshnevisan, B., Rafiee, S., Omid, M., Mousazadeh, H., & Sefeedpari, P. (2013). Prognostication of environmental indices in potato production using artificial neural networks. *Journal of Cleaner Production*, 5, 402-409. <https://doi.org/10.1016/j.jclepro.2013.03.028>
- Kitani, O. (1999). *Energy and Biomass Engineering*. In: *CIGR Handbook of Agricultural Engineering Vol. 5*, ASAE Publication, St. Joseph, M. I., 330pp. <http://cigr.org/documents/CIGRHandbookVol5.pdf>
- Liao, S., & Wen, C. (2007). Artificial neural networks classification and clustering of methodologies and applications- literature analysis from 1995 to 2005. *Expert Systems With Applications*, 32, 1-11. <https://doi.org/10.1016/j.eswa.2005.11.014>
- Mandal, K. G., Saha, K. P., Ghosh, P. K., Hati, K. M., & Bandyopadhyay, K. K. (2002). Bioenergy and economic analysis of soybean-based crop production systems in central India. *Biomass. Bioenergy*, 23, 337-345. [https://doi.org/10.1016/S0961-9534\(02\)00058-2](https://doi.org/10.1016/S0961-9534(02)00058-2)
- Medar, R. A., & Rajpurohit, V. S. (2014). A survey on data mining techniques for crop yield prediction. *International Journal of Advance Research in Computer Science and Management Studies*, 2, 59-64. <https://www.academia.edu/9421293>
- Mousavi-Avval, S. H., Rafiee, S., Jafari, A., & Mohammadi, A. (2011). Optimization of energy

- consumption for soybean production using Data Envelopment Analysis (DEA) approach. *Applied Energy*, 88, 3765-3772. <https://doi.org/10.1016/j.apenergy.2011.04.021>
- Mucherino, A., Papajorgji, P., & Pardalos, P. M. (2009). A survey of data mining techniques applied to agriculture. *Operational Research*, 9, 121-140. <https://doi.org/10.1007/s12351-009-0054-6>
- Nassiri, S. M., & Singh, S. (2009). Non-parametric energy use efficiency, energy ratio and specific energy for irrigated wheat crop production. *Iran Agricultural Research*, 28, 27-38. http://iar.shirazu.ac.ir/article_151_3e06ddd6ccfdfeb5f9b2aaca5c75d.pdf
- Oliveira, M. P. G., Bocca, F. F., & Rodrigues, L. H. A. (2017). From spreadsheets to sugar content modeling: A data mining approach. *Computers and Electronics in Agriculture*, 132, 14-20. <https://doi.org/10.1016/j.compag.2016.11.012>
- Phillips-Wren, G., Sharkey, P. H., & MorssDy, S. (2008). Mining lung cancer patient data to assess healthcare resource utilization. *Expert Systems With Applications*, 35, 1611-1619. <https://doi.org/10.1016/j.eswa.2007.08.076>
- Ramesh, D., & Vardhan, B. V. (2013). Data mining techniques and applications to agricultural yield data. *Journal of Advanced Research in Computer and Communication Engineering*, 2, 3477-3480. <https://www.ijarccce.com/upload/2013/september/31-h-dantam%20ramesh%20-data%20mining%20techniques%20and%20applicat ion%20to.pdf>
- Raorane, A.A. & Kulkarni, R.V. (2013). Review- role of data mining in agriculture. *International Journal of Computer Science and Information Technology*, 4, 270-272. <http://cloud.politala.ac.id/politala/1.%20Jurusan/Te knik%20Informatika/19.%20e-journal/Jurnal%20Internasional%20TI/IJCSIT/Vol %204/ISSUE%202/ijcsit20130402018.pdf>
- Raorane, A. A., & Kulkarni, R. V. (2015). Application of data mining tool to crop management system. *Russian Journal of Agricultural and Socio-Economic Sciences*, 37, 3-16. <https://ideas.repec.org/a/scn/031261/16082009.html>
- Shahin, S., Jafari, A., Mobli, H., Rafiee, S., & Karimi, M. (2008). Effect of farm size on energy ratio for wheat production: A case study from Ardabil Province of Iran. *American-Eurasian Journal of Agricultural and Environmental Sciences*, 3, 604-608. <https://www.researchgate.net/publication/237334154>
- Sharma, A. (2006). Spatial data mining for drought monitoring: An approach using temporal NDVI and rainfall relationship. M. Sc. Thesis. Geo-information Science Earth Observation, University of Twente, 75pp. <https://www.iirs.gov.in/node/254>
- Srivastava, A. K., Goering, C. E., & Rohrbach, R. P. (1993). *Engineering Principles of Agricultural Machines*. ASAE Publication, 604pp. <https://www.amazon.com/Engineering-Principles-Agricultural-Machines-2nd/dp/1892769506>
- Thomas, E. (2017). An artificial neural network for real-time hardwood lumber grading. *Computers and Electronics in Agriculture*, 132, 71-75. <https://doi.org/10.1016/j.compag.2016.11.018>

تنبؤات الطاقة لإنتاج القمح باستخدام تقنية استخراج البيانات في إيران

نسليم منجزي

قسم هندسة النظم الحيوية، كلية الزراعة، جامعة شهيد چمران، الأهواز، إيران

المستخلص: يعتبر القمح من أهم المحاصيل الزراعية في إيران. وبسبب كبر المساحة المزروعة بالقمح في محافظة خوزستان لذلك لابد من وجود أداة لمعالجة البيانات المستحصلة من أجل إعطاء المعلومات الدقيقة لمديري القطاعات الزراعية. وتعد تقنية استخراج البيانات ذات كفاءة عالية في تقديم المعلومات والنماذج الأساسية لمنتجي القمح لنمذجة استهلاك الطاقة. تعد الشبكة العصبية الاصطناعية واحدة من أكثر الخوارزميات دقة. الهدف الرئيسي من هذا البحث هو التنبؤ بالطاقة الناتجة من مزارع القمح باستخدام شبكة عصبية متعددة الطبقات. وهو بحث تحليلي وقاعدة بياناته تتكون من 1240 سجلاً. تم الحصول على البيانات المطلوبة للبحث من مزرعة قمح خلال الاعوام 2014-2018. تم إجراء تحليل البيانات عبر IBM SPSS 14.2 و CRISP القياسي. بالنسبة للنموذج المستخدم في البحث وجد أن متغيرات الأسمدة الكيماوية والآلات ووقود الديزل ذات المعاملات 0.2987 و 0.2064 و 0.1527 على التوالي كان لها التأثير الأعلى على متغير الإنتاج (الطاقة الإنتاجية). مقدار دقة التنبؤ في خوارزمية الشبكة العصبية والتي تمثل نسبة السجلات المتوقعة بشكل صحيح إلى إجمالي السجلات كانت 93.08%. كان الارتباط الخطي بين القيم الفعلية والقيم المتوقعة 0.92 و 0.88 على التوالي ، لبيانات التدريب وبيانات الاختبار التي تشير إلى ارتباط قوي. يمكن أن تكون النتائج التي تم الحصول عليها فعالة لمزارعي القمح في اتجاه تقييم وتحسين استهلاك الطاقة في عملية إنتاج القمح وتقليل استهلاك مدخلات الطاقة.

الكلمات المفتاحية: الطاقة، التنبؤ، التنقيب في البيانات، ANN، القمح.