بررسی کارایی الگوریتم داده‌کاوی (شبکه عصبی مصنوعی) در پیش‌بینی انرژی خروجی محصول ذرت

نسیم منجزی[1]، حسن ذکی دیزجی[2]

[1]استادیار، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران؛ n.monjezi@scu.ac.ir
[2]استادیار، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران؛ hzakid@scu.ac.ir

چکیده

ذرت یکی از مهم‌ترین محصولات زراعی است. استان خوزستان از بزرگترین مراکز کشاورزی تولید ذرت در کشور ایران است. این پژوهش، کاربرد تکنیک داده‌کاوی به منظور پیش بینی انرژی خروجی مزارع ذرت در استان خوزستان است. داده‌های مورد استفاده از ۱۷۳ مزرعه جمع‌آوری گردید. کشاورزان به روش نمونه‌گیری تصادفی انتخاب شدند. هدف از این پژوهش تعیین انرژی ورودی و خروجی و پیش‌بینی آنها در مزارع ذرت با استفاده از تکنیک داده‌کاوی (شبکه عصبی پرسپترون چند لایه) است. این پژوهش از نوع تحلیلی بوده و پایگاه داده آن شامل ۱۳۸۴ رکورد می‌باشد. داده‌های مورد نیاز این تحقیق، طی سال زراعی ۹۷-۱۳۹۶ بدست آمده است. تجزیه و تحلیل به کمک نرم افزار IBM SPSS modeler 14.2 با به کارگیری استاندارد CRISP انجام شده است. نتایج نشان داد، انرژی مصرفی در مزارع تولید ذرت ۴۰۴۳۳/۳۲ $MJ\ ha^{-1}$ است. حدود ۴۰ درصد این انرژی مربوط به نهاده کود شیمیایی و ۵۳ درصد آن مربوط به انرژی الکتریکی مصرفی برای آبیاری و سوخت دیزل می‌باشد. همچنین با توجه به مدل استفاده شده، مشخص گردید که به ترتیب متغیرهای انرژی الکتریسیته و آبیاری، نیروی انسانی و کودهای شیمیایی دارای بیشترین تاثیر روی متغیر خروجی (انرژی تولیدی) بودند. میزان صحت پیش‌بینی در الگوریتم شبکه عصبی، یعنی نسبت رکوردهای درست پیش‌بینی شده به کل رکوردها، برابر ۸۸/۵ درصد بود. همچنین همبستگی خطی میان مقادیر واقعی و مقادیر پیش‌بینی شده برای داده‌های آموزش و داده‌های آزمون به ترتیب برابر ۰/۹۴ و ۰/۸۸ درصد می‌باشد که بیانگر همبستگی قوی است. نتایج حاصل از این پژوهش می‌تواند برای کشاورزان ذرت کار استان خوستان در راستای ارزیابی و بهینه‌سازی مصرف انرژی در فرآیند تولید ذرت و کاهش مصرف نهاده‌های انرژی‌بر راه گشا باشد.

کلمات کلیدی: انرژی، پیش‌بینی، داده‌کاوی، شبکه عصبی مصنوعی، ذرت.

Evaluation of data mining algorithm (Artificial Neural Network) to predict output energy on maize cultivation

Nasim Monjezi[1], Hassan Zakidizaji[2]

[1]Assistant Professor, Department of Biosystems Engineering , Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, n.monjezi@scu.ac.ir
[2]Assistant Professor, Department of Biosystems Engineering , Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, hzakid@scu.ac.ir

ABSTRACT

Corn is one of Iran's most important food crops. Khuzestan province is the main center of maize production in Iran. This paper presents a data mining study of samples of maize obtained from producers in Khuzestan province. Data were collected from 173 producers, using face-to-face questionnaire method. The farms were chosen by random sampling method. The purpose of this study is to determine energy consumption of input and output used in maize production. In addition, Output energy of maize farms is

۱- نسیم منجزی، اهواز، بلوار گلستان، دانشگاه شهید چمران اهواز، تلفن: ۰۶۱۳۳۳۳۰۰۱۰.

predicted using data mining and multi-layer perceptron neural network. This is an analytic research and its database consists of 1384 records. Data required for this research was obtained during growing seasons in 2015-2017. Data analysis was done via IBM SPSS modeler 14.2 and standard CRISP. The results showed that the amount of energy consumed in maize production was 40433.32 MJ ha^{-1}. About 40% of this was generated by fertilizers and 53% from electricity (irrigation) and diesel fuel. Concerning the model used in the research, it was found that variables of electricity energy & irrigation, human labor and chemical fertilizer 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 88.5%. In addition, linear correlation between actual values and predicted values was 0.94 and 0.88 respectively for training data and testing data suggesting strong correlation. Results obtained in this research can be effective for maize farmers in Khuzestan in direction of evaluation and optimization of energy consumption in process of maize production and reduction of consumption of energy inputs.

Keywords: Energy, Predict, Data mining, Artificial Neural Networks (ANN), Maize.

۱- مقدمه

ذرت یکی از مهم‌ترین محصولات غذایی کشور ایران است. این محصول، از نظر سطح زیرکشت و تولید، بعد از گندم و برنج، در رتبه سوم قرار دارد. استان خوزستان یکی از مهم‌ترین تولیدکنندگان این محصول با سطحی برابر ۸۸۲۳۰ هکتار در سال زراعی ۹۷-۱۳۹۶می‌باشد (Anonymous, 2016). یکی از مسائل مهمی که کشاورزان بخصوص ذرت کاران با آن روبرو هستند، پیش بینی عملکرد یا انرژی محصول تولیدی است. رابطه بین کشاورزی و انرژی بسیار نزدیک است. تاکنون مطالعات قابل توجهی در زمینه مصرف انرژی در بخش کشاورزی در کشورهای مختلف انجام شده است (Namdari et al., 2011; Ozkan et al., 2011; Yilmaz et al., 2005; Sefeedpari et al., 2014). از طرفی، بخش کشاورزی هم مصرف‌کننده انرژی و هم تولیدکننده انرژی‌های زیستی می‌باشد (Namdari, 2011). به همین خاطر، استفاده موثر از منابع انرژی یکی از مهم‌ترین راه‌های پایداری در کشاورزی و محیط زیست است (Taheri-Garavand et al., 2010). لذا بهینه‌سازی و مدل‌سازی مصرف انرژی در تولید محصولات کشاورزی مانند ذرت بسیار حائز اهمیت می‌باشد. از طرف دیگر تولیدکنندگان در این بخش با حجم بسیار زیادی از داده‌های جمع‌آوری شده با خصوصیات بسیار متنوع و با روابط پیچیده در بین آن‌ها مواجه هستند که آنالیز و مدیریت آن‌ها به وسیله تجربه و تحلیل‌های تجربی و آماری، امری دشوار می‌باشد. داده کاوی[1] یک فناوری توانمند در مدیریت و سازمان‌دهی اطلاعات با حجم بالا می‌باشد. داده‌کاوی عبارت است از جستجو در یک پایگاه داده‌ها برای یافتن الگوهایی میان داده‌ها. در زمینه به کارگیری تکنیک‌های داده‌کاوی در کشاورزی تاکنون تحقیقات بسیاری انجام شده است.

Cao et al. (2013) یک روش جدید برای حل معادلات اطلاعات مکانی موثر بر عملکرد ذرت ارائه دادند. این روش با استفاده از اصول و مبانی آمار، به پیش‌بینی عملکرد ذرت پرداخت. همچنین این روش تاثیر پارامترهای بارندگی و حاصلخیزی خاک منطقه بر عملکرد ذرت را با استفاده از شبکه عصبی محاسبه کرد. در این تحقیق، خطای آزمایش در سطح ۵٪ کنترل شد. Ureta et al. (2013) نیز با استفاده از تکنیک داده‌کاوی، روابط بین پارامترهای محیطی (ارتفاع، آب و هوا، شیب و خاک) و عوامل اجتماعی (آموزش و نژاد) را با توزیع مکانی کشت ذرت مکزیکی بررسی کردند. آن‌ها همچنین مناطقی که به طور بالقوه می‌تواند بیشترین پتانسیل برای کشت این محصول را داشته باشد، شناسایی کردند. Medar & Rajpurohit (2014) روش‌های مختلف پیش‌بینی عملکرد محصول را با استفاده از تکنیک‌های داده‌کاوی ارائه دادند. این روش‌ها شامل روش‌های خوشه‌بندی K-Means، K-Nearest Neighbor (KNN) و شبکه‌های عصبی مصنوعی (ANN) بود. در ادامه منابعی که در داخل پرانتز آورده شده‌اند نیز توانمندی و لزوم به کارگیری تکنیک‌های داده‌کاوی در کشاورزی را تصدیق می‌نمایند (Kalpana et al., 2014a; Kalpana et al., 2014b; Geetha, 2015; Raorane & Kulkarni, 2013; Raorane & Kulkarni, 2015; Jeysenthil et al., 2014).

هدف از تحقیق حاضر نیز، پیش‌بینی انرژی خروجی محصول ذرت در استان خوزستان با استفاده از رهیافت داده‌کاوی (شبکه عصبی مصنوعی) است.

[1]- Data mining

۲- بخش مواد و روش‌ها

۲-۱- منطقه مورد مطالعه و جمع‌آوری داده‌ها

استان خوزستان با وسعتی حدود ۶۴۲۸۲ کیلومترمربع در جنوب غربی کشور ایران واقع شده است. این منطقه در بین عرض‌های جغرافیایی ۲۹ درجه و ۵۸ دقیقه تا ۳۲ درجه و ۵۸ دقیقه شمالی و طول‌های جغرافیایی ۴۷ درجه و ۴۲ دقیقه تا ۵۰ درجه و ۳۹ دقیقه شرقی قرار گرفته است. داده‌های مورد نیاز تحقیق، در سال زراعی ۱۳۹۶-۹۷ جمع‌آوری شده است. برای برآورد حجم نمونه از فرمول کوکران (رابطه ۱) استفاده شد (Kizilaslan, 2009).

$$n = \frac{N(t.s)^2}{Nd^2 + (t.s)^2} \quad (۱)$$

که در آن که در آن N، اندازه جامعه آماری یا تعداد زارعین ذرت‌کار در منطقه مورد مطالعه، t ضریب اطمینان قابل قبول که با فرض نرمال بودن توزیع صفت مورد نظر از جدول t-استیودنت به‌دست می‌آید. S^2 برآورد واریانس صفت مورد مطالعه در جامعه، d، حداکثر خطای مجاز که بین ۰/۰۱ تا ۰/۰۵ و n، حجم نمونه است.

در پارامترهای بالا به‌خاطر اینکه واریانس صفت مورد مطالعه در ابتدا مشخص نیست، یک بررسی آزمایشی در مقیاس کوچک برای پی بردن به نواقص احتمالی و برآورد تقریبی صفت مورد مطالعه انجام شد. پس از تجزیه وتحلیل داده‌های نمونه‌گیری مقدماتی و به دست آوردن تقریبی آماره‌های فوق در صفت مورد مطالعه در جامعه مورد نظر با داشتن مقدار پارامترهای فوق و قرار دادن آن‌ها در فرمول کوکران حجم نمونه‌گیری اصلی بدست آمد. بدین‌ترتیب حجم نمونه برای مزارع ذرت ۱۷۳ مزرعه به دست آمد.

۲-۲- هم‌ارز انرژی نهاده‌ها و ستاده

انرژی‌های ورودی در تولید ذرت شامل انرژی مصرفی در عملیات و انرژی مصرف شده در تولید ماشین‌های کشاورزی، کود، سم، بذر، نیروی انسانی، سوخت مصرفی، الکتریسیته و آبیاری بودند. محاسبه انرژی مصرفی با استفاده از معادل انرژی مربوط به هر واحد از نهاده یا ستانده و ضرب آن در مقدار نهاده مصرف شده یا ستانده تولید شده، انجام گرفت البته در مطالعات، با توجه به شرایط حاکم، از ضرایب متفاوتی استفاده شده که در جدول ۱ به برخی از آنها اشاره شده است.

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم (ماشین‌های کشاورزی) و مکانیزاسیون ایران

جدول ۱- معادل نهاده ها و ستانده انرژی در تولید ذرت.

Table 1. Energy equivalent of inputs and output in maize production.

Item	Unit	Energy equivalent (MJ kg^{-1})	Reference
1. Human labor	h	1.96	Taheri-Garavand et al., 2010
2. Machinery	Kg		
a. Tractor		138	Taheri-Garavand et al., 2010
b. Plow		180	Taheri-Garavand et al., 2010
c. Disk harrow		149	Salami et al., 2010
d. Planter		133	
e. Equipment of fertilizing		129	Taheri-Garavand et al., 2010
f. Sprayer		129	Taheri-Garavand et al., 2010
g. Combine		116	
h. Other machinery		62.7	Mousavi-Avval et al.,2011
3. Diesel fuel	Lit	47.8	Banaeian et al., 2011
4. Chemicals	Kg		
a. Herbicide		238	Erdal et al., 2007
b. Pesticide		101.2	Erdal et al., 2007
c. Fungicide		216	Erdal et al., 2007
5. Fertilizers	Kg		
a. Nitrogen (N)		78.1	Pishgar Komleh et al., 2011a
b. Phosphate (P_2O_5)		17.4	Pishgar Komleh et al., 2011a
c. Potassium (K_2O)		13.7	Pishgar Komleh et al., 2011a
6. Water for irrigation	m^3	0.63	Pishgar Komleh et al., 2011a
7. Electricity	kWh	3.60	Heidari and Omid, 2011
8. Seed	Kg	14.7	Canakci et al., 2005

۳-۲- الگوریتم داده‌کاوی

کشف دانش در پایگاه داده، فرآیند شناسایی درست، ساده، مفید، و نهایتا الگوها و مدل‌های قابل فهم در داده‌ها می‌باشد. داده‌کاوی، مرحله‌ای از فرآیند کشف دانش می‌باشد و شامل الگوریتم‌های مخصوص داده‌کاوی است، به‌طوری‌که، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده کشف می‌کند. به بیان ساده‌تر، داده‌کاوی به فرآیند استخراج دانش ناشناخته، درست، و بالقوه مفید از داده اطلاق می‌شود. تعریف دیگر اینست که، داده‌کاوی گونه‌ای از تکنیک‌ها برای شناسایی اطلاعات و یا دانش تصمیم‌گیری از قطعات داده می‌باشد، به نحوی که با استخراج آن‌ها، در حوزه‌های تصمیم‌گیری، پیش‌بینی، پیشگویی، و تخمین مورد استفاده قرار گیرند (Maione, 2016).

شبکه‌های عصبی

شبکه‌های عصبی از تکنیک‌های پیشرفته داده‌کاوی محسوب می‌شوند (Ayman, 2015). شبکه‌های عصبی می‌توانند هر گونه وابستگی پیوسته بین ورودی و هدف را به تقریبی برآورد نمایند، حتی اگر این رابطه غیر خطی باشد (Phillips-Wren et al., 2008). مهم‌ترین ویژگی شبکه‌های عصبی که در کاربرد آن‌ها حائز اهمیت است، توانایی یادگیری نمونه‌ها است (Efendigil et al., 2009). روش یادگیری می‌تواند به سه صورت: با ناظر یا بدون ناظر و یا هر دو باشد. روش یادگیری با ناظر با زوج‌هایی از الگوهای ورودی–خروجی انجام می‌شود، در صورتی که یادگیری بدون ناظر صرفا شامل ارائه نمونه‌های آموزشی به شبکه است و تا زمانی ادامه می‌یابد که شبکه به بیشترین تفکیک بین طبقه‌هایی از نمونه‌ها دست یابد (Arockiaraj, 2013). معروف‌ترین شبکه‌های عصبی، شبکه عصبی پرسپترون چند لایه[1] هستند که عموما از نوع آموزش دیده با ناظر و با الگوریتم پس انتشار خطا[2] می‌باشند (Liao and Wen, 2007).

[1]- Multilayer perceptron
[2]- Error Back- propagation Algorithm

۲-۴- تحلیل داده‌ها

در این تحقیق از نرم افزار داده کاوی IBM SPSS Modeler 14.2 برای مدل‌سازی شبکه عصبی و اعتبار سنجی نتایج بهره گرفته شد. داده‌ها به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم شدند که سهم داده‌های آموزشی ۷۰ درصد داده‌ها و سهم داده‌های آزمایش ۳۰ درصد داده‌ها است. مراحل تحقیق در شکل ۱ نمایش داده شده است.

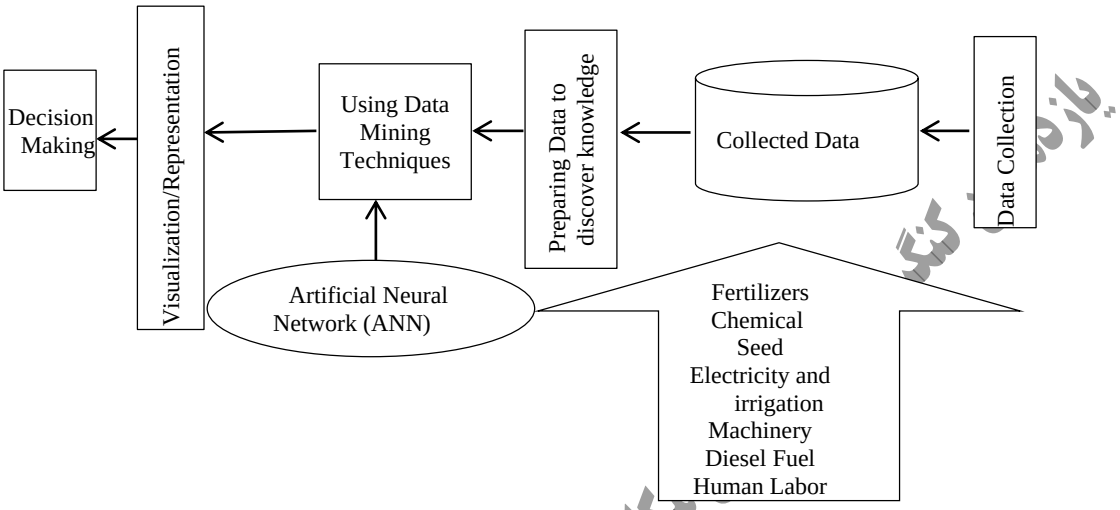

Fig. 1. Proposed Framework.

شکل ۱- مراحل تحقیق.

۲-۵- ارزیابی مدل

در این مرحله، مدل خروجی از داده‌های آموزش برای داده‌های آزمون به کار گرفته می‌شود، به بیان دیگر در این قسمت داده‌های آزمون-که حدود ۳۰ درصد از داده‌های این پژوهش را تشکیل می‌دهند- با مدل استخراجی پیش‌بینی می‌شوند و نتایج پیش‌بینی به وسیله این مدل با واقعیت مقایسه می‌گردد.

۳- نتایج و بحث

۳-۱- تحلیل انرژی‌های ورودی و خروجی در تولید ذرت

مقادیر انرژی‌های ورودی و خروجی در تولید ذرت در جدول ۲ نشان داده شده است. کل انرژی ورودی برابر ۴۰۴۳۳/۳۲ $MJ\ ha^{-1}$ می‌باشد. نهاده کود بیشترین مقدار انرژی مصرفی را به خود اختصاص داده است و پس از آن نهاده‌های الکتریسیته و سوخت دیزل در رتبه‌های بعدی قرار دارند. مقدار انرزی خروجی تولید ذرت نیز برابر ۹۵۳۵۷/۵۴ $MJ\ ha^{-1}$ است. در تحقیقی مشابه، Pishgar Komleh et al. (2011b) الگوی مصرف انرژی در ذرت علوفه‌ای را در استان تهران بررسی کردند. نتایج این تحقیق نشان داد که نهاده‌های ماشین و کود شیمیایی به ترتیب ۴۲٪ و ۲۸٪ انرژی مصرفی را به خود اختصاص می‌دهند. کل انرژی خروجی برابر ۶۸۹۲۸ $MJ\ ha^{-1}$ برآورد گردید. همچنین مقادیر انحراف معیار، بیشترین و کمترین مقدار انرژی های ورودی و انرژی خروجی مورد استفاده در مدل شبکه عصبی مصنوعی ANN نیز در جدول ۲ نشان داده شده است.

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم و مکانیزاسیون ایران

جدول ۲- میزان نهاده‌های ورودی و خروجی در تولید ذرت.

Table 2. Amounts of energy inputs and output for maize production.

Item	Total energy Equivalent (unit ha^{-1})	Percentage (%)	standard deviation (SD) (unit ha^{-1})	Max (unit ha^{-1})	Min (unit ha^{-1})
Inputs					
Human labor (MJ)	119.58	0.30	17.06	148.98	80.50
Machinery (MJ)	1398.91	3.46	1270.64	17437.98	700.98
Diesel fuel (MJ)	8181.65	20.23	1165.59	10546.9	6004.65
Chemicals (MJ)	546.85	1.35	242.33	1800.87	120.44
Fertilizers (MJ)	16489.10	40.78	2480.72	21087.49	11077.33
Electricity and irrigation (MJ)	13332.11	32.97	2679.83	16970.70	7065.98
Seed (MJ)	365.12	0.90	8.90	382.20	323.40
Total energy input (MJ)	40433.32	100	5299.30	59570.74	26431.18
Output					
Maize yield (kg)	6486.91	-	748.22	7600	5000
Total energy output (MJ)	95357.54	-	10998.89	111720	73500

۲-۳- داده‌کاوی

در این تحقیق، اطلاعات جمع آوری شده از مزارع ذرت استان خوزستان به عنوان پایگاه داده استفاده شده‌اند. مراحل اجرایی فاز دوم پژوهش (داده کاوی) به شش قسمت مرحله شناخت و انتخاب داده‌ها، توصیف داده‌ها، مرحله پیش پردازش و آماده‌سازی داده‌ها، مرحله مدل‌سازی و تجزیه و تحلیل داده‌ها، ارزیابی و اعتبار سنجی مدل و توسعه مدل، تفکیک شده‌اند.

مرحله شناخت و انتخاب داده‌ها

در مرحله‌ی نخست اطلاعات موجود در زمینه نهاده‌های مصرفی در تولید ذرت مورد بررسی قرار گرفته است. با توجه به اطلاعات جمع‌آوری شده، ۸ متغیر تاثیرگذار و مورد نیاز برای مدل‌سازی انرژی مصرفی و انرژی تولیدی در تولید ذرت شناسایی و به صورت فیلدهای جداگانه‌ای در بانک اطلاعاتی مزارع تعریف و ذخیره‌سازی گردید. ۷ متغیر مستقل پژوهش عبارتند از:

- سوخت - کود شیمیایی - الکتریسیته و آبیاری - ماشین - بذر - سم - نیروی انسانی

متغیر انرژی خروجی مزرعه نیز متغیر وابسته تحقیق است.

توصیف داده‌ها و تعریف متغیرهای مورد استفاده در مدل

داده‌های مورد استفاده در این پژوهش در قالب یک فایل Excel با ۸ ستون، شامل ۷ متغیر مستقل (ورودی) و یک متغیر وابسته (هدف) به نرم افزار IBM Spss modeler 14.2 وارد شدند که اطلاعات توصیفی متغیرها در جدول ۳ نشان داده شده است. این جدول شامل نوع متغیر، مقدارهای حداکثر و حداقل متغیر، میانگین متغیر، انحراف میانگین و تعدادهای رکورد معتبر موجود در پایگاه داده برای هر متغیر می‌باشد.

جدول ۳- توصیف متغیرهای ورودی مدل‌ها.

Table 3. Description of maize variables used for this study.

Field	Measurement	Min (MJ ha^{-1})	Max (MJ ha^{-1})	Mean (MJ ha^{-1})	standard deviation(MJ ha^{-1})	Valid
Fertilizers	Continuous	11077	21087	16489.14	2487.91	173
Electricity and irrigation	Continuous	7066	16971	13332.09	2687.61	173
Chemical	Continuous	120	1801	546.90	243.04	173
Seed	Continuous	323	382	365.54	9.05	173
Machinery	Continuous	701	17438	1398.94	1274.33	173
Diesel Fuel	Continuous	6005	10547	8181.68	1168.98	173
Human Labor	Continuous	80	149	119.58	17.09	173
Output energy	Continuous	73500	111720	95357.54	11030.82	173

مرحله پیش‌پردازش و آماده‌سازی داده‌ها

قبل از مدل‌سازی، فرآیند آماده‌سازی و پیش‌پردازش برای بهبود کیفیت داده‌ها بر روی داده‌های خام جمع‌آوری شده از مزارع ذرت، انجام شد. همچنین وجود رکوردهای غیرمتعارف و پرت در پایگاه داده موجب افزایش ضریب خطا در نتایج داده‌کاوی می‌گردد، لذا تشخیص و مدیریت این داده‌های غیرمتعارف در مرحله آماده‌سازی داده‌ها از اهمیت بسزایی برخوردار می‌شود. در این مرحله در بین رکوردهای ثبت شده، با توجه به توصیف آماری داده‌ها داده گمشده وجود نداشت.

مرحله مدل‌سازی و تجزیه و تحلیل داده‌ها

در این بخش داده‌ها را مدل‌سازی کرده و تحلیل‌های لازم در خصوص اعتبارسنجی مدل ارائه می‌شود. پایگاه داده نهایی برای پیش بینی انرژی خروجی مزارع ذرت به منظور ورود به نرم افزار، دارای ۱۳۸۴ رکورد و ۸ فیلد از متغیرهای ورودی و خروجی بود. برای فراخوانی مدل‌سازی، فیلد متغیر انرژی خروجی به عنوان خروجی و بقیه متغیرها (فیلدها) به عنوان متغیر ورودی به نرم افزار معرفی شد. الگوریتم اصلی اعمال شده در این پژوهش به منظور مدل‌سازی انرژی تولیدی مزارع ذرت، شبکه عصبی مصنوعی بود. شبکه مورد استفاده از نوع پرسپترون چند لایه است که در زمره‌ی کاربردی‌ترین شبکه‌های عصبی مصنوعی می‌باشد. تعداد لایه‌های پنهان شبکه، یک لایه با ۳ نرون می‌باشد (شکل۲). شرط توقف مدل‌سازی این است که اگر خطا کاهش بیشتری نداشت، بهینه‌سازی متوقف می‌شود. در مدل حاضر، صحت پیش‌بینی داده‌ها ۸۸/۵ درصد بدست آمد.

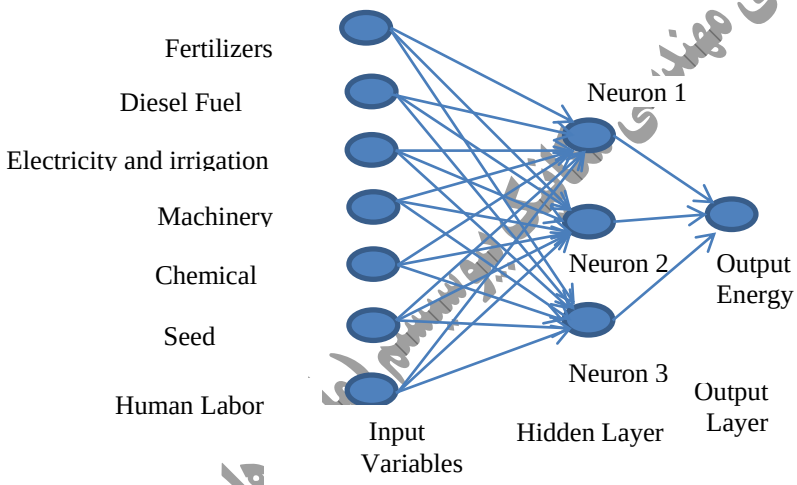

Fig. 2. Architecture of a multilayer perceptron network.

شکل ۲- معماری یک شبکه پرسپترون چند لایه.

با توجه به شرح متغیرهای ورودی و خروجی، باید به این نکته اشاره کرد که برخی از متغیرها تاثیر بیشتری در مدل پیش‌بینی دارند. در نمودار مدل شبکه عصبی شکل ۳، سیر نزولی تاثیر متغیرها روی متغیر هدف مشخص شده است. اولین متغیر با بیشترین تاثیر، متغیر الکتریسیته و آبیاری می‌باشد. نیروی انسانی و کودهای شیمیایی در رتبه‌های بعدی تاثیر گذاری روی متغیر خروجی (انرژی تولیدی) قرار دارند. سایر متغیرها نیز تاثیرگذار بوده ولی میزان اثرگذاری آن‌ها کمتر از سه متغیر مذکور است.

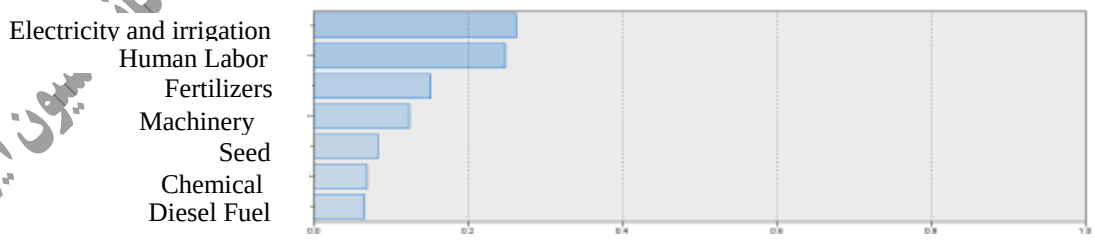

Fig. 3. Predictor Importance (Target: output energy).

شکل ۳- اهمیت متغیرها (هدف: انرژی خروجی).

در شکل‌های ۴ و ۵ نمودارهای هیستوگرام مقادیر واقعی و مقادیر پیش‌بینی شده انرژی خروجی مزارع ذرت، نمایش داده شده است. در این نمودارها، نسبت بین داده‌های آموزش و آزمون نیز مشخص شده است.

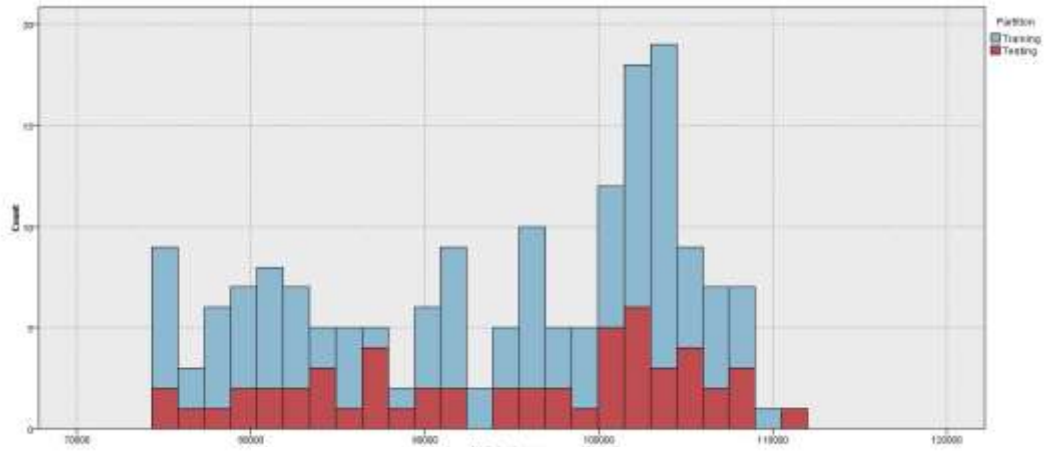

Fig. 4. Histogram of predicted values of output energy of maize farms.

شکل ۴- هیستوگرام مقادیر پیش‌بینی شده انرژی خروجی مزارع ذرت.

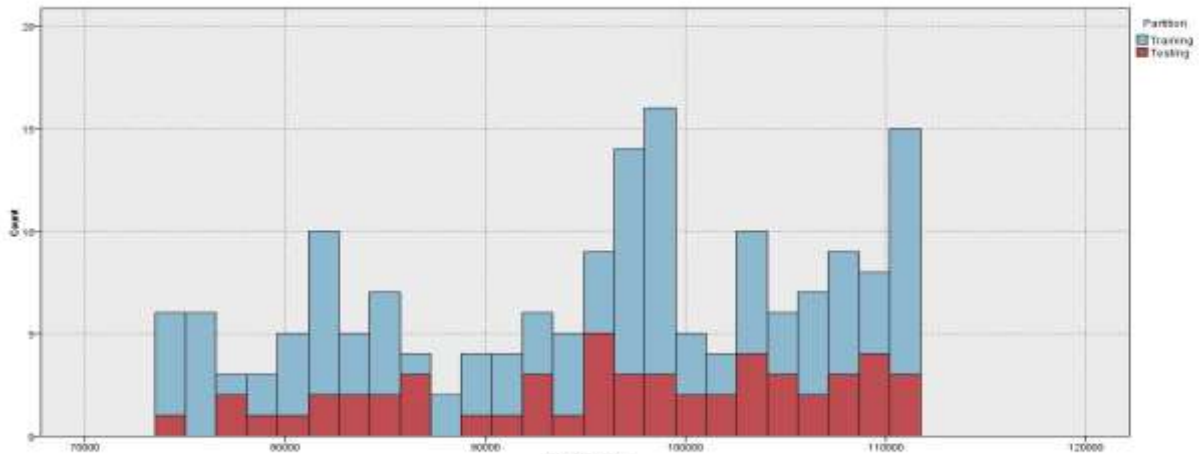

Fig. 5. Histogram of actual values frequency of output energy of maize farms.

شکل ۵- هیستوگرام فراوانی مقادیر واقعی انرژی خروجی مزارع ذرت.

ارزیابی و اعتبارسنجی مدل

پس از ایجاد مدل بایستی به ارزیابی مدل ایجاد شده پرداخت. نتایج ارزیابی باعث بهبود مدل شده و مدل را قابل استفاده می‌کند. از آنجا که روش ارائه شده در هر پژوهشی باید به لحاظ اعتبار، مورد سنجش قرار گیرد، بنابراین در این پژوهش نیز با عنایت به اینکه روش پژوهش از نوع داده محور است، روش اعتبار سنجی به این صورت است که داده‌ها به دو مجموعه داده‌های آموزشی و داده‌های آزمایشی تقسیم می‌شوند، داده‌های بخش آموزش، مدل را می‌سازند و داده‌های بخش آزمون، مدل ایجاد شده را مورد ارزیابی قرار می‌دهند. به بیان دیگر داده‌های آزمون با مدل استخراجی پیش‌بینی شدند و نتایج پیش‌بینی به وسیله این مدل با واقعیت مقایسه شد.

تقسیم داده‌ها توسط نرم افزار مورد استفاده و به صورت تصادفی صورت می‌گیرد. از لحاظ تعداد داده‌های هر مجموعه، همیشه تعداد داده‌های آموزشی بیشتر از داده‌های آزمایشی در نظر گرفته می‌شود. در این پژوهش تعداد مجموعه‌های آموزش ۶۸/۷۹ درصد از کل مزارع بوده (۱۱۹ مزرعه) و ۳۱/۲۱ درصد باقی مانده (مزرعه۵۴) به عنوان داده‌های آزمایش در نظر گرفته شدند (شکل ۶). میزان اعتبار با نتایج داده‌های جدید آزمون می‌شود و داده‌های آزمایشی به منزلهٔ ناظر به الگوریتم وارد شده و نتایج میزان صحت آن را ارزیابی می‌کند.

میزان صحت پیش‌بینی در الگوریتم شبکه عصبی، یعنی نسبت رکوردهای درست پیش‌بینی شده به کل رکوردها، برابر ۸۸/۵ درصد بود (شکل ۷). بنابراین مدل ساخته شده با شبکه عصبی از دقت تخمینی بالایی برخوردار بود. بحث بعدی آن است که چگونه می‌توان صحت روش‌های پیش‌بینی را اندازه گیری کرد. برای ارزیابی صحت روش‌های پیش‌بینی (اختلاف بین مقدار واقعی و مقدار پیش‌بینی شده متغیر وابسته)، شاخص

میانگین قدرمطلق خطای پیش‌بینی مورد استفاده قرار می‌گیرد. میانگین قدرمطلق خطای (MAE) آن برای داده‌های آموزشی برابر ۲۹۲۱/۹۶ و برای داده‌های آزمایش برابر ۳۹۳۵/۳۲ می‌باشد که مقدار مطلوبی است (جدول ۴). واضح است که برای بالابردن صحت یک روش پیش‌بینی لازم است که مقادیر دو شاخص یاد شده تا حد مقدور کوچک باشند. همچنین همبستگی خطی میان مقادیر واقعی و مقادیر پیش‌بینی شده برای داده‌های آموزش و داده‌های آزمون به ترتیب برابر ۰/۹۴ و ۰/۸۸ درصد می‌باشد که بیان‌گر همبستگی قوی است (شکل ۸).

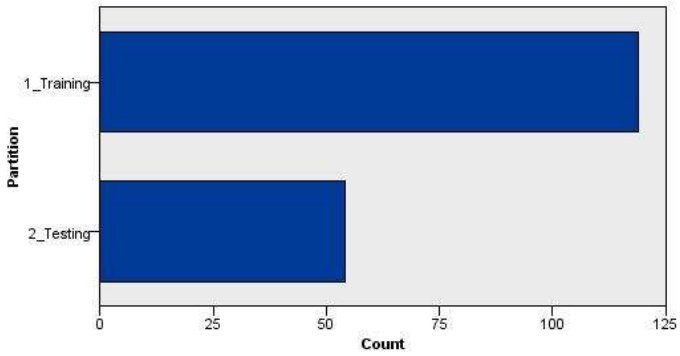

Fig. 6. Ratio of training and testing data in the neural network model.

شکل ۶- نسبت داده‌های آموزش و آزمایش در مدل شبکه عصبی.

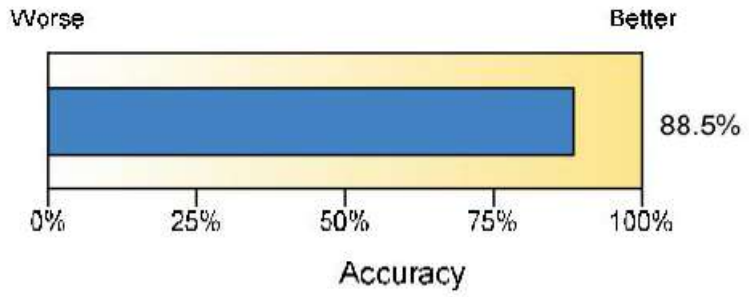

Fig. 7. Amount of predictive precision in neural network algorithm.

شکل ۷- میزان صحت پیش‌بینی در الگوریتم شبکه عصبی.

جدول ۴- مقایسه انرژی خروجی پیش بینی شده با انرژی خروجی واقعی.

Table 4. Comparing predicted output energy with actual output energy.

Partition	1_Training	2_Testing
Minimum Error	-6840.49	-6422.70
Maximum Error	12123.88	16727.86
Mean Error	929.76	1848.45
Mean Absolute Error	2921.96	3935.32
Standard Deviation	3719.43	5031.36
Linear Correlation	0.94	0.88
Occurrences	119	54

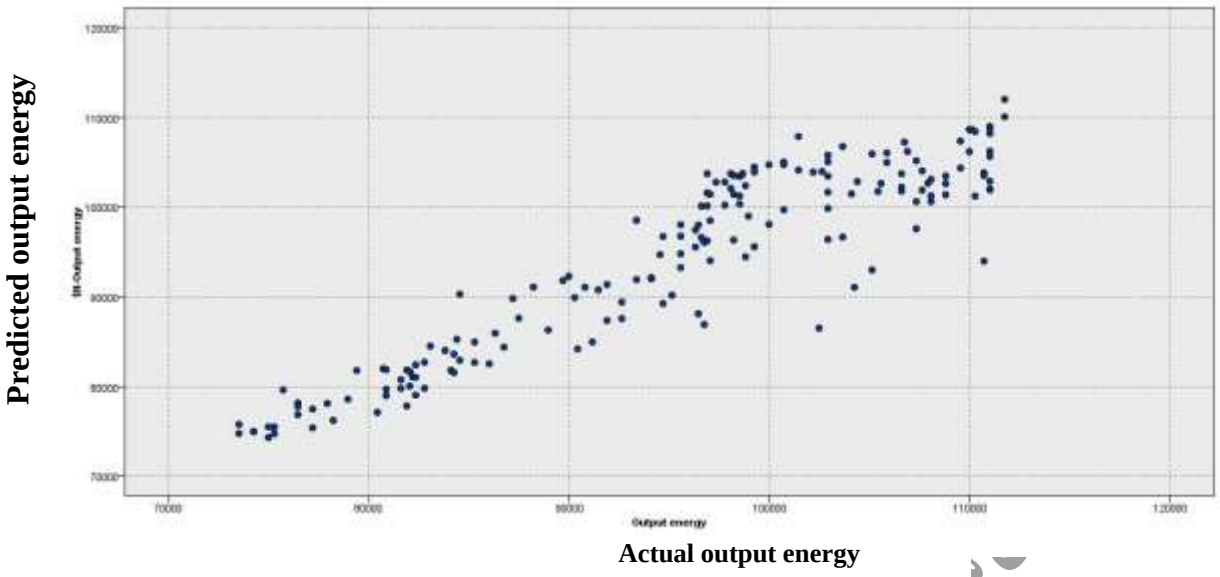

Fig. 8. Actual output energy versus predicted output energy.

شکل ۸- نمودار همبستگی میان داده‌های واقعی و مقادیر پیش‌بینی شده.

نمودار ترفیع[1]

برای نمایش نحوه عملکرد مدل و پیش‌بینی آن از نمودار ترفیع استفاده شد. مقادیر نمودار ترفیع از رابطه (۲) محاسبه شده است.

(۲) (تعداد مجموع رکوردها/تعداد مجموع اصابت‌ها)/(رکوردها در افزایش/اصابت‌ها در افزایش)

محور افقی نمودار، بیانگر درصد مجموعه داده‌های آزمون و آزمایش مورد بررسی است. این نمودار بدین صورت است که در یک نمودار به عنوان حالت ایده آل (شکل نمودار در صورتی که کلیه حالت‌ها صحیح پیش‌بینی شده باشد) رسم می‌شود و همچنین نمودار دیگری که بیان کننده حالت مدل پیش‌بینی شده می‌باشد را در همان گراف رسم می‌کند (هان و کامبر[2]، ۲۰۰۶). در نتیجه با مقایسه این دو نمودار می‌توان دریافت مدل در چه قسمت‌هایی توانسته پیش‌بینی خوبی داشته باشد و یا اینکه در کدام قسمت‌ها ضعیف عمل کرده است. شکل ۹ نشان‌دهنده ترفیع مدل ایجاد شده است.

همانطور که در نمودار مشاهده می‌شود سه شاخص رسم شده است که خط ۱ (قرمز ضخیم) نشان‌دهنده نمودار ایده آل، خط ۲ (قرمز نازک)، نمودار ترفیع مدل و خط ۳ (آبی)، شاخص بهترین نمودار ترفیع است.

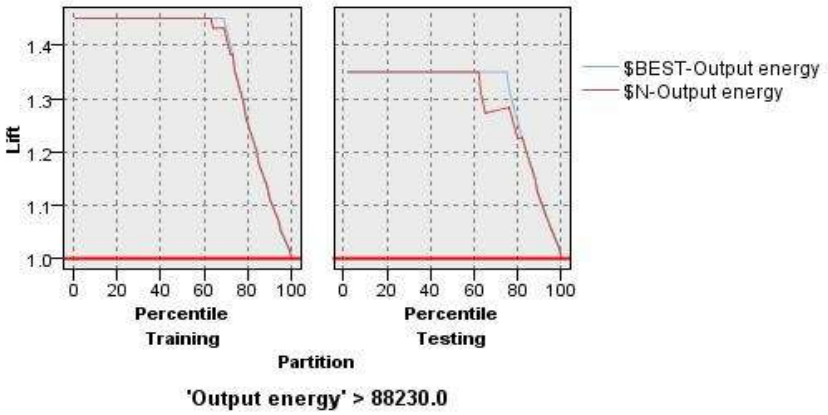

Fig. 9. Lift chart of neural network model.

شکل ۹- نمودار ترفیع مدل شبکه عصبی.

[1]- Lift chart
[2]- Han and Kamber

نمودار بهره[1]

بهره برابر است با درصد رکوردهای طبقه هدف نسبت به کل پایگاه داده، که در این گره قرار می‌گیرند. بهره به صورت رابطه (۳) تعریف می‌شود:

(۳) ۱۰۰٪*(تعداد کل اصابت‌ها/اصابت در افزایش)

این چارت نشان می‌دهد که چقدر باید شبکه را گسترش داد تا بتوان این درصد اصابت را در شبکه عصبی به دست آورد (شکل ۱۰).

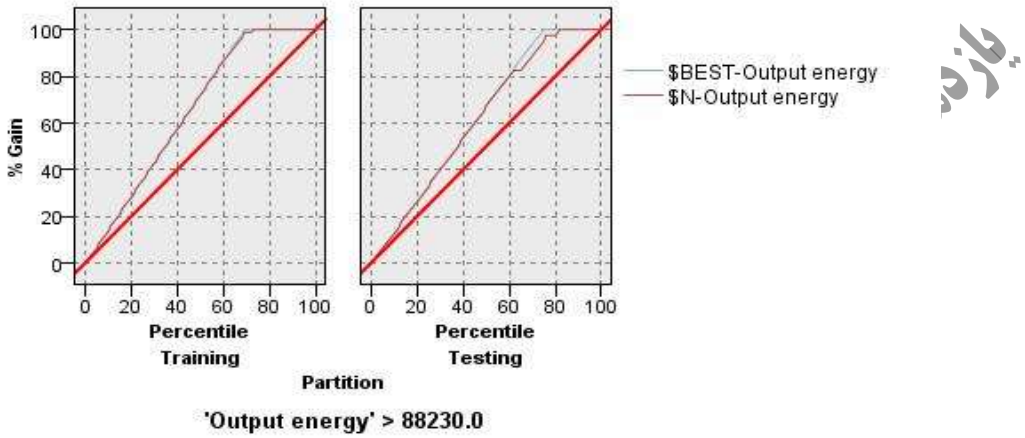

Fig. 10. Chart of neural network gain.

شکل ۱۰- نمودار بهره مدل شبکه عصبی.

نمودار پاسخ[2]

پاسخ برابر است با درصد رکوردهای موجود در گره جاری که در طبقه هدف قرار می‌گیرند. نمودار پاسخ شبکه عصبی مدل در شکل ۱۱ نمایش داده شده است. این مقدار از رابطه (۴) محاسبه می‌شود:

(۴) ۱۰۰٪*(رکوردها در افزایش/پاسخ‌ها در افزایش)

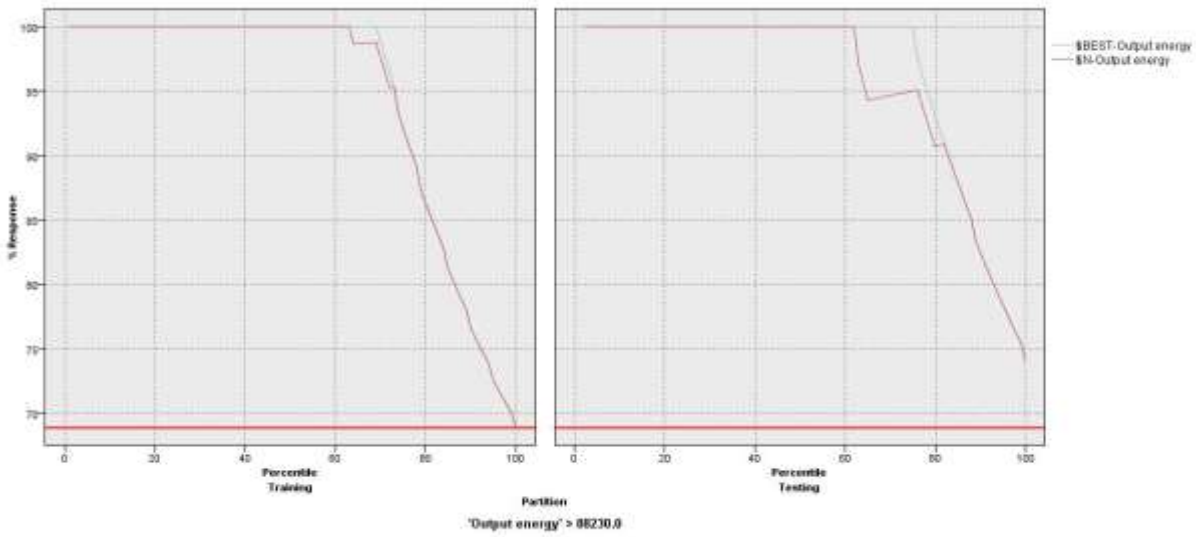

Fig. 11. Chart of neural network response.

شکل ۱۱- نمودار پاسخ مدل شبکه عصبی.

[1]- Gain
[2]- Response

توسعه مدل

در این مرحله مدل به دست آمده به همراه نتایج آن قابل ارائه به کشاورزان است تا در آینده برای پیش‌بینی تولید انرژی در طی فرآیند تولید ذرت به کار گرفته شود. بنابراین بعد از ایجاد گزارشات لازم سعی شد، توضیح داده شود که بر اساس مدل ایجاد شده تاثیرگذارترین متغیرها بر پیش‌بینی انرژی خروجی مزارع ذرت کدامند. با توجه به اینکه متغیرهای الکتریسیته و آبیاری، نیروی انسانی و کودهای شیمیایی از تاثیرگذارترین متغیرها بر انرژی خروجی مزارع ذرت می‌باشند، چنانچه استفاده از این نهاده‌های ورودی بهینه گردد، کارآیی انرژی در مزارع نیز افزایش خواهد یافت.

۴- نتیجه‌گیری

در این مقاله داده‌های مزارع ذرت استان خوزستان مورد استفاده قرار گرفت. در فاز اول تحقیق، آنالیز انرژی های ورودی و خروجی مزارع انجام شد. کل انرژی مصرفی در تولید ذرت برابر ۴۰۴۳۳/۳۲ MJ ha^{-1} شد. نهاده کود شیمیایی بیشترین سهم را در بین نهاده های مصرفی داشته است. انرژی الکتریسیته و آبیاری در رتبه های بعدی قرار دارند. کل انرژی خروجی ذرت نیز برابر ۱۹۵۳۵۷/۵۶ MJ ha^{-1} شد. در فاز دوم تحقیق (داده کاوی)، پایگاه داده تهیه شده به دو بخش آموزش و آزمایش تقسیم گردید. در ادامه، مدلی بر اساس تکنیک شبکه عصبی مصنوعی داده کاوی، برای تخمین انرژی در فرآیند تولید ذرت با استفاده از نرم افزار IBM modeler 14.2 و مجموعه داده‌های آموزشی، تولید گردید. در نهایت مدل تولید شده با استفاده از مجموعه داده‌های تست، مورد ارزیابی قرار گرفت و به دقت حدود ۸۸/۵٪ برای تخمین انرژی خروجی رسید. به خوبی مشاهده می‌شود که مدل تولید شده از دقت بسیار خوبی در تخمین مقادیر انرژی خروجی برخوردار است. همچنین نتایج این مطالعه نشان می‌دهد بهره‌گیری از روش‌های پیش‌بینی، با ارائه تصویری دقیق‌تر از وضعیت انرژی در مزارع ذرت، امکان تخصیص مناسب منابع و افزایش بهره‌وری از نهاده‌ها را فراهم می‌نماید.

۵- تقدیر و تشکر

این مقاله مستخرج از طرح تحقیقاتی می‌باشد. بدین وسیله نویسندگان از معاونت پژوهشی دانشگاه شهید چمران اهواز به خاطر تامین بخشی از هزینه‌های تحقیق، سپاسگزاری می‌نمایند.

۶- مراجع

- Anonymous. (2016). General report of mechanization in Ahvaz Township. Management of Jahad Keshavarzi of Ahvaz Township Publication.
- Arockiaraj, M. (2013). Applications of Neural Networks in Data Mining. *International Journal of Engineering and Science*, 3(1), 8-11.
- 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.
- Banaeian, N., Omid, M., & Ahmadi, H. (2011). Energy and economic analysis of greenhouse strawberry production in Tehran province of Iran. *Energy Conversion and Management*, 52, 1020-1025.
- Canakci, M., Topakci, M., Akinci, I., & Ozmerzi, A. (2005). Energy use pattern of some field crops and vegetable production: Case study for Antalya Region, Turkey. *Energy Conversion and Management*, 46, 655- 666.
- Cao, L., Sanb, X., Zhaob, Y., & Chenb, G. (2013). The application of the spatio-temporal data mining algorithm in maize yield prediction. *Mathematical and Computer Modelling*, 58, 507–513.
- 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.
- Erdal, G., Esengun, K., Erdal, H., & Gunduz, O. (2007). Energy use and economical analysis of sugar beet production in Tokat province of Turkey. *Energy*, 32, 35-41.
- Geetha, M. C. S. (2015). A Survey on Data Mining Techniques in Agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2), 887-892.
- Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*, San Morgan, Francisco Kaufman.
- Heidari, M. D., & Omid, M. (2011). Energy use patterns and econometric models of major greenhouse vegetable productions in Iran. *Energy*, 36, 220-225.
- 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 and*

Technology, 3(3), 9923- 9930.

- Kalpana, R., Shanthi, N., & Arumugam, S. (2014a). Data Mining – An Evolutionary View of Agriculture. *International Journal of Application or Innovation in Engineering & Management*, 3(3), 102- 105.

- Kalpana, R., Shanthi, N., & Arumugam, S. (2014b). A Survey on Data Mining Techniques in Agriculture. *International Journal of Advances in Computer Science and Technology*, 3(8), 426-431.

- Kizilaslan, H. (2009)|. Input–output energy analysis of cherries production in Tokat Province of Turkey. *Applied Energy*, 86, 1354–1358.

- Liao, S., & Wen, Ch. (2007). Artificial neural networks classification and clustering of methodologies and applications – literature analysis from 1995 to 2005. *Expert Systems with Applications*, 32, 1–11.

- Maione, C., Batista, B. L., Campiglia, A. D., Barbosa Jr, F., & Barbosa, R. M. (2016). Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. *Computers and Electronics in Agriculture*, 121, 101–107.

- 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(9), 59-64.

- Mousavi-Avval, S. H., Rafiee, S., Jafari, A., & Mohammadi, A. (2011). Improving energy use efficiency of canola production using data envelopment analysis (DEA) approach. *Energy*, 36, 2765-2772.

- Namdari, M. (2011). Energy use and cost analysis of watermelon production under different farming technologies in Iran. *International Journal of Environmental Sciences*,1(6), 1144-1153.

- Namdari, M., Mohammadi, A., & Ghasemi Mobtaker, H. (2011). Assessment of energy requirements and sensitivity analysis of inputs for watermelon production in Iran. *International Journal Plant, Animal Environment Science*, 1(1), 102-110.

- Ozkan, B., Ceylan, R. F. & Kizilay, H. (2011). Energy inputs and crop yield relationships in greenhouse winter crop tomato production. *Renewable Energy*, 36, 3217-3221.

- Phillips-Wren, G., Sharkey, Ph., & MorssDy, S. (2008). Mining lung cancer patient data to assess healthcare resource utilization. *Expert Systems with Applications*, 35, 1611–1619.

- Pishgar Komleh, S. H., Omid, M., & Keyhani, A. (2011a). Study on Energy use Pattern and Efficiency of Corn Silage in Iran by using Data Envelopment Analysis (DEA) Technique. *International Journal of Environmental Sciences*, 1(6), 1094-1106.

- Pishgar Komleh, S. H., Keyhani, A., Rafiee, Sh., & Sefeedpary, P. (2011b). Energy use and economic analysis of corn silage production under three cultivated area levels in Tehran province of Iran. *Energy*, 36, 3335-3341.

- Raorane, A. A. & Kulkarni R. V. (2013). Review- Role of Data Mining in Agriculture. *International Journal of Computer Science and Information Technologies*, 4 (2), 270 – 272.

- Salami, P., Ahmadi, H., & Keyhani, A. R. (2010). Estimating the Energy Indices and Profitability of StrawberryProduction in Kamyaran Zone of Iran. *Energy Research Journal*,1(1), 32-35.

- Sefeedpari, P., Shokoohi, Z. & Behzadifar, Y. (2014). Energy use and carbon dioxide emission analysis in sugarcane farms: a survey on Haft-Tappeh Sugarcane Agro Industrial Company in Iran. *Journal of Cleaner Production*, 83, 212-219.

- Taheri-Garavand, A., Asakereh, A. & Haghani, K. (2010). Energy elevation and economic analysis of canola production in Iran a case study: Mazandaran province. *International Journal of Environmental Sciences*, 1(2), 236-242.

- Ureta, C., González-Salazara, C., Gonzálezc, E. J., Álvarez-Buyllad, E. R., & Martínez-Meyera, C. (2013). Environmental and social factors account for Mexican maize richness and distribution: A data mining approach. *Agriculture, Ecosystems and Environment*, 179, 25– 34.

- Yilmaz, I., Akcaoz, H., & Ozkan, B. (2005). An analysis of energy use and input costs for cotton production in Turkey. Renew. *Energy*, 30, 145–55.

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم و مکانیزاسیون ایران