نشریه زراعت و فناوری زعفران
جلد ۵، شماره ۴، زمستان ۱۳۹۶ ص. ۳۹۲-۳۷۹

Saffron Agronomy & Technology
Vol. 5, No. 4, Winter 2018, P. 379-392

طبقه‌بندی اقلیمی و تعیین مناطق مستعد کشت زعفران با استفاده از الگوریتم‌های داده‌کاوی در سطح استان خراسان رضوی

مهدی بشیری[۱]، علی ماروسی[۲*]، امیر سالاری[۳] و محمد قدوسی[۴]

تاریخ دریافت: ۱۶ شهریور ۱۳۹۵ **تاریخ پذیرش:** ۳ بهمن ۱۳۹۵

بشیری، م.، ماروسی، ع.، سالاری، ا.، و قدوسی، م. ۱۳۹۶. طبقه‌بندی اقلیمی و تعیین مناطق مستعد کشت زعفران با استفاده از الگوریتم‌های داده‌کاوی در سطح استان خراسان رضوی. زراعت و فناوری زعفران، ۵(۴): ۳۹۲-۳۷۹.

چکیده

پیش‌بینی عملکرد محصولات کشاورزی، نقش مهمی در برنامه‌ریزی صادرات، واردات، خرید تضمینی، قیمت‌گذاری، سود مطمئن و افزایش بهره‌وری کشاورزی دارد. عملکرد محصولات، تابع پارامترهای متعددی از جمله اقلیم است. در این تحقیق، عملکرد زعفران در استان خراسان رضوی توسط الگوریتم‌های طبقه‌بندی شامل شبکه‌های عصبی مصنوعی، مدل‌های رگرسیونی، درخت خطی محلی، درخت تصمیم، آنالیز تشخیص، جنگل تصادفی، ماشین بردار پشتیبان و آنالیز نزدیک‌ترین همسایه با استفاده از ۱۱ پارامتر اقلیمی طی دوره ۲۰ ساله زراعی (۸۸-۱۳۶۸) ارزیابی شد. نتایج نشان داد که تعداد کمی از پارامترهای اقلیمی، بر عملکرد زعفران تأثیر دارند. پارامترهای دمای حداقل، میانگین و حداکثر و بارش به ترتیب بیشترین همبستگی مثبت و پارامترهای رطوبت مطلق حداکثر، رطوبت نسبی ساعت ۶:۳۰، ساعات آفتابی، رطوبت نسبی ساعت ۱۸:۳۰، تبخیر، رطوبت نسبی ساعت ۱۲:۳۰ و رطوبت مطلق حداقل نیز به‌ترتیب بیشترین همبستگی منفی را با مناطق کشت زعفران داشتند. همچنین در طبقه‌بندی مناطق کشت زعفران، آنالیز تشخیص و ماشین بردار پشتیبان، از دقت بالاتری برخوردار بودند. بین مناطق کشت زعفران و میزان عملکرد محصول همبستگی نسبتاً مناسبی با ضریب همبستگی ۰/۳۸ به‌دست آمد. بین مناطق کشت و پارامترهای اقلیمی همبستگی بالایی برخوردار است و آنالیز نزدیک‌ترین همسایه با ضرایب تعیین برابر ۱ و ۰/۹۴ در مراحل آموزش و آزمون، با دقت بالایی مناطق کشت را طبقه‌بندی نماید، اما در پیش‌بینی میزان عملکرد محصول بر اساس پارامترهای اقلیمی، دقت مدل‌ها نسبتاً پایین بود (متوسط ضریب تعیین برابر ۰/۴۸ و ۰/۰۵ در مراحل آموزش و آزمون) و آنالیز نزدیک‌ترین همسایه، بالاترین دقت پیش‌بینی را در مراحل آموزش و آزمون (به ترتیب ضریب تعیین برابر ۱ و ۰/۱۷) نشان داد. طبق یافته‌های تحقیق، می‌توان با استفاده از پارامترهای اقلیمی و الگوریتم‌های داده‌کاوی، به‌طور مناسب اقدام به تفکیک مناطق کشت نمود و با شناسایی مناطقی که اقلیم آن‌ها مشابه مناطق با میزان عملکرد بالا است، مناطق مستعد کشت زعفران را شناسایی نمود.

کلمات کلیدی: پارامترهای اقلیمی، توسعه کشت، داده‌کاوی، شاخص آماری، عملکرد محصول.

۱- استادیار گروه مرتع و آبخیزداری، دانشکده کشاورزی و منابع طبیعی، دانشگاه تربت‌حیدریه
۲ – استادیار گروه کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه تربت‌حیدریه
۳- استادیار گروه تولیدات گیاهی، دانشکده کشاورزی و منابع طبیعی، دانشگاه تربت‌حیدریه
۴- مربی گروه صنایع، دانشکده فنی و مهندسی، دانشگاه تربت‌حیدریه
(*- نویسنده مسئول: ali.maroosi@torbath.ac.ir)

DOI: 10.22048/jsat.2017.60768.1189

مقدمه

ایران بزرگ‌ترین تولیدکننده زعفران جهان و استان‌های خراسان رضوی و جنوبی عمده‌ترین نواحی تولید این محصول در ایران می‌باشد. پیش‌بینی عملکرد از جمله مسائل مهم کشاورزی محسوب می‌شود و هر کشاورزی، علاقه‌مند به دانستن عملکرد نهایی محصول خود می‌باشد. عملکرد محصول تابع پارامترهای متعددی از جمله پارامترهای اقلیمی است، از این‌رو شناخت اثر متغیرهای مختلف اقلیمی بر عملکرد محصول، حائز اهمیت است. امروزه فرآیند پیش‌بینی با استفاده از روش داده‌کاوی که روشی نسبتاً جدید در زمینه کشاورزی محسوب می‌گردد، انجام می‌پذیرد. در روش داده‌کاوی و پیش‌بینی عملکرد مورد انتظار، اصل بر آموزش مدل بوده و این مهم توسط داده‌های گردآوری شده از گذشته انجام می‌پذیرد. تکنیک‌های داده‌کاوی به دو دسته کلاس‌بندی و خوشه‌بندی تقسیم می‌شوند. تکنیک‌های کلاس‌بندی در کلاس‌بندی نمونه‌های ناشناخته با استفاده از اطلاعات تهیه شده به‌کار می‌روند.

روش‌های داده‌کاوی بسیار متنوع بوده و در زمینه‌های مختلف به‌کار برده می‌شوند. از جمله پرکاربردترین این روش‌ها، شبکه‌های عصبی مصنوعی است که پژوهشگران علوم کشاورزی از این شبکه‌ها در شبیه‌سازی پارامترهایی چون میزان تبخیر و تعریق هفتگی (Landeras et al., 2009)، تبخیر روزانه (Piri et al., 2009)، پیش‌بینی دمای هوا (Smith et al., 2009)، تشعشع خورشیدی (Mubiru, 2008)، پیش‌بینی سیلاب (Mukerji et al., 2009) و قابلیت نگه‌داشت آب خاک (Merdun et al., 2006) استفاده نمودند. پیش‌بینی عملکرد ذرت و سویا با استفاده از روش‌های داده‌کاوی نشان داد شبکه‌های عصبی مصنوعی در مقایسه با مدل‌های رگرسیونی دارای دقت بالاتری می‌باشند (Kaul et al., 2005). همچنین در تعیین عوامل اقلیمی مؤثر بر عملکرد گندم، نسبت بارندگی به تبخیر و تعرق پتانسیل، مهم‌ترین فاکتور اقلیمی مؤثر بر عملکرد به‌دست آمد (Alvarez, 2009). با انتخاب سه پارامتر مقدار آب آبیاری، کود نیتروژن و درجه روز رشد به عنوان ورودی مدل‌های شبکه‌های عصبی مصنوعی، مدل می‌تواند عملکرد گیاه را با دقت نسبتاً بالا برآورد نماید و در بین این سه پارامتر، پس از پارامتر درجه روز رشد، مقدار آب آبیاری و کود نیتروژن، در برآورد میزان عملکرد محصول نقش به‌سزایی داشتند (Bagheri et al., 2012). در تحقیقات زیادی توانایی شبکه عصبی در بررسی روابط غیرخطی مورد ارزیابی قرار گرفته است (Coulibaly et al., 2001; Imrie et al., 2000; Islam & Kothart, 2000; Moradkhani, 2004; Wu et al., 2005; Zealand, 1999).

یکی دیگر از روش‌های داده‌کاوی، روش نزدیک‌ترین همسایه می‌باشد. تخمین رطوبت خاک در مزرعه ذرت بدون تخریب ساختمان خاک با استفاده از الگوریتم نزدیک‌ترین همسایه، نشان داد که این مدل از دقت بالایی در این زمینه برخوردار است. ضمن آنکه دقت این روش جهت شبیه‌سازی بارش روزانه و سایر متغیرهای اقلیمی و همچنین تخمین متغیرهای جنگل جهت آنالیز تصاویر ماهواره‌ای قابل قبول گزارش گردیده است (Meng et al., 2014). اخیراً الگوریتم‌های گروهی مدل‌سازی داده‌ها (GMDH) نیز به‌طور موفقیت‌آمیزی در سیستم‌های خطی و غیرخطی و در دامنه وسیعی از زمینه‌های پزشکی، سیستم‌های کنترل، فرآیند سیگنال، اقتصادی، اکولوژی و کشاورزی مورد استفاده قرار گرفته است (Tamura & Kondo, 1980). الگوریتم GMDH برای اولین بار جهت تولید مدل‌های ریاضی سیستم‌های پیچیده با مدیریت داده‌های مشاهداتی، پیشنهاد گردید (Madala & Ivakhnenko, 1974). استفاده از مدل GMDH جهت پیش-

بینی عملکرد برنج نشان داد که این مدل به مراتب بهتر از شبکه‌های عصبی مصنوعی، عملکرد برنج را پیش‌بینی می‌نمایند (Samsudin et al., 2008).

با توجه به مرور منابع موجود و گستره کاربرد الگوریتم‌های داده‌کاوی در علوم مختلف، پژوهش حاضر با هدف شناسایی مناطق کشت زعفران و عملکرد محصول بر مبنای پارامترهای اقلیمی و روش‌های داده‌کاوی صورت پذیرفت تا دقت این الگوریتم‌ها را در علوم کشاورزی و محصول زعفران با توجه به فقدان پیشینه پژوهشی ارزیابی کرده و به توصیه‌های کاربردی در این راستا دست یافت.

مواد و روش‌ها

منطقه مورد مطالعه

پژوهش حاضر بر اساس عملکرد زعفران و پارامترهای اقلیمی استان خراسان رضوی انجام پذیرفت. این استان بین ۵۶ درجه و ۱۹ دقیقه تا ۶۱ درجه و ۱۶ دقیقه طول شرقی و ۳۳ درجه و ۵۲ دقیقه تا ۳۷ درجه و ۴۲ دقیقه عرض شمالی قرارگرفته و ۱۱۸۸۵۴ کیلومترمربع مساحت دارد. میانگین بارندگی آن ۲۰۸ میلی‌متر است و بیشترین بارش در اواخر زمستان و اوایل بهار رخ می‌دهد. مناطق پرباران استان در نواحی مرتفع قرار دارند و میزان بارش از شمال به جنوب کاهش می‌یابد (Hoseini et al., 2008). جهت تعیین بهترین الگوریتم داده‌کاوی در طبقه‌بندی مناطق و عملکرد بر مبنای پارامترهای اقلیمی، داده‌های عملکرد زعفران (برحسب کیلوگرم در هکتار) در سطح ۸ شهرستان استان شامل قوچان، گناباد، کاشمر، مشهد، سبزوار، نیشابور، سرخس و تربت‌حیدریه، طی دوره ۲۰ ساله زراعی ۸۸-۱۳۶۸ بر اساس بانک اطلاعات مرکز تحقیقات کشاورزی و منابع طبیعی خراسان رضوی وارد تحلیل شد. مدل‌سازی نیز با انتخاب ۱۱ پارامتر اقلیمی از محل آمار سازمان هواشناسی کشور و مؤثر بر رشد و عملکرد زعفران شامل تبخیر، ساعات آفتابی، بارندگی، رطوبت‌های مطلق حداکثر و حداقل، رطوبت نسبی ساعت‌های ۶:۳۰، ۱۲:۳۰ و ۱۸:۳۰و دماهای میانگین، حداقل و حداکثر انجام گرفت. جدول ۱ میانگین مقادیر پارامترهای اقلیمی و عملکرد در واحد سطح زعفران را نمایش می‌دهد. همان‌گونه که در جدول ۱ ارائه شده، نواحی کشت (شهرستان) بر اساس افزایش میانگین عملکرد زعفران کدبندی شده‌اند، به‌طوری‌که برای شهرستان با کمترین میانگین عملکرد زعفران (نیشابور) شماره ۱ و شهرستان با بیشترین میانگین زعفران (گناباد) شماره ۸ در نظر گرفته شده است.

روش پژوهش

در ابتدا با استفاده از تحلیل همبستگی به روش پیرسون[1] ارتباط بین پارامترهای اقلیمی و عملکرد محصول و مناطق کشت زعفران بررسی گردید. این ضریب که بین دو متغیر تصادفی برابر با کوواریانس آن‌ها تقسیم بر انحراف معیار آن‌ها است، میزان همبستگی خطی بین دو متغیر تصادفی را می‌سنجد. مقدار این ضریب بین ۱- تا ۱ تغییر می‌کند و ۱ به معنای همبستگی مثبت کامل، صفر به معنی نبود همبستگی و منفی ۱ به معنی همبستگی منفی کامل است. جهت برازش الگوریتم‌های مختلف طبقه‌بندی[2]، از نرم‌افزار متلب[3] استفاده گردید، بدین منظور روش‌های مختلف داده‌کاوی با برازش بر داده‌ها، مورد ارزیابی قرار گرفت. برای همه الگوریتم‌های استفاده شده سعی شده است پارامترهای پیش‌فرض که در متلب ارائه شده و معمولاً مقادیر متداولی می‌باشند رعایت شود.

در برخی از موارد که پارامترهای متفاوتی استفاده شده است پس از توضیح آن روش پارامترهای آن نیز آورده شده است. برای مقایسه روش‌های مختلف، داده‌های ورودی (شرایط اقلیمی) و داده‌های خروجی (میزان عملکرد و ناحیه کشت) نرمال‌سازی شدند، سپس آموزش مدل با ۹۰ درصد داده‌ها انجام و ۱۰ درصد باقیمانده داده‌ها در آزمون مدل به‌کار گرفته شد.

1- Pearson
2- Classification
3- MATLAB

جدول ۱- میانگین پارامترهای اقلیمی و عملکرد زعفران در شهرستان‌های استان طی دوره مطالعاتی

Table 1- Average of climatic parameters and saffron yield in Khorasan-Razavi province during the study period

شهرستان County	عملکرد Yield(kg.he^{-1})	میانگین دما Mean temperature (°c)	دمای مطلق حداکثر Maximum absolute temperature (°c)	دمای مطلق حداقل Minimum absolute temperature (°c)	درصد رطوبت نسبی Relative humidity percent (18:30 o'clock)	درصد رطوبت نسبی Relative humidity percent (12:30 o'clock)	درصد رطوبت نسبی Relative humidity percent (6:30 o'clock)	رطوبت مطلق حداکثر Maximum absolute humidity (g.m^{-3})	رطوبت مطلق حداقل Minimum absolute humidity (g.m^{-3})	بارندگی Rainfall (mm)	ساعت آفتابی Sunny hours	تبخیر Evaporation (mm)
قوچان (کد ۲) Ghuchan (Code 2)	2.88	12	19	6	50	41	71	77	35	306	2814	1612
گناباد (کد ۸) Gonabad (code 8)	4.26	17	23	10	34	30	51	28	34	141	3182	2769
کاشمر (کد ۴) Kashmar (code 4)	3.30	17	23	11	37	33	49	52	29	196	3151	2724
مشهد (کد ۳) Mashhad (code 3)	3.01	15	21	8	46	39	66	73	36	242	2907	2157
نیشابور (کد ۱) Neyshabour (code 1)	2.58	14	21	6	42	35	65	70	28	230	3054	2414
سبزوار (کد ۶) Sabzevar (code 6)	3.42	18	24	11	36	31	53	59	25	181	3081	26.7
سرخس (کد ۵) Sarakhs (code 5)	3.33	12	25	11	44	36	62	67	32	192	2909	2437
تربت حیدریه (کد ۷) Torbat Heydarieh (code 7)	3.79	14	20	7	40	35	58	56	25	253	3193	2068

به‌منظور مقایسه و تعیین مدل برتر، از آماره‌های متوسط مربع خطا و ضریب تعیین استفاده گردید. اعتبارسنجی مدل‌ها نیز با استفاده از روش ارزیابی متقاطع[1] با پارامتر K-fold برابر با ۱۰ انجام گردید. روش‌های مورد استفاده شامل موارد زیر می‌باشند.

روش‌های رگرسیونی

اساس مدل‌های رگرسیونی بر اساس این نظریه استوار است که اگر دو عامل با یکدیگر همبستگی داشته باشند، تغییر یکی، با تغییر دیگری همراه است و با دانستن ضریب همبستگی، می‌توان اندازه تغییر عامل وابسته را با توجه به تغییر عوامل مستقل مرتبط با آن، اندازه‌گیری و پیش‌بینی نمود. روش‌های رگرسیونی بکار رفته در این پژوهش عبارت است از رگرسیون خطی[2]، رگرسیون درجه دوم خالص[3]، رگرسیون اثر متقابل[4] و رگرسیون درجه دوم[5] (Siemsen et al., 2010).

روش شبکه عصبی[6]

یک شبکه عصبی مصنوعی شامل تعداد زیادی نرون می‌باشد که به‌صورت واحد برای حل یک مسئله مخصوصاً مسائل طبقه‌بندی استفاده می‌شوند (Abraham, 2005; Damangir, 2001; Azadeh et al., 2006). نرون‌ها می‌توانند از توابع محرک متفاوتی جهت تولید خروجی استفاده کنند که رایج‌ترین آن ها توابع لگاریتم سیگموئیدی، تانژانت سیگموئیدی و تابع محرک خطی است. الگوریتم‌های آموزش شبکه عصبی به‌کار رفته در این تحقیق شامل الگوریتم لونبرگ -مارکوآرت، شبکه با سرعت یادگیری متغیر، شبکه با الگوریتم گرادیان توأم مقیاس شده و الگوریتم با تنظیم به‌روش بیز هستند (Golabi et al., .

1- Cross Validation
2- Linear Regression
3- Pure Quadratic Regression
4- Interaction Regression
5- Quadratic Regression
6- Neural Network

2013). پارامترهای شبکه عصبی پیش‌فرض متلب می‌باشد. از آنجا که تعداد تکرار آموزش[1] خیلی زیاد با اینکه زمان شبیه‌سازی را زیاد می‌کند، باعث بهبود محسوس طبقه‌بندی نمی‌شود مقدار این پارامتر از ۱۰۰۰ به ۳۰۰ تغییر داده شد. مقدار اولیه نرخ آموزش[2] ۰/۰۰۵ در نظر گرفته شد و میزان خطای هدف به‌عنوان یکی از معیارهای پایان آموزش، صفر در نظر گرفته شد. نسبت-های مختلفی از داده برای آموزش و آزمون استفاده شد و به‌منظور یکسان‌سازی برای کل روش‌های استفاده‌شده در این پژوهش، ۹۰ درصد داده‌ها برای آموزش و ۱۰ درصد برای آزمون در نظر گرفته شد.

روش تشخیص الگو[3]

روش شناسایی الگو یکی از مؤثرترین روش‌های شبکه عصبی است و جهت آموزش دسته‌بندی داده‌های ورودی به کار می‌رود. در این مدل از تابع گرادیان توأم مقیاس شده[4] استفاده شده است (Basu et al., 2010). پارامترهای سیگما[5] نیز که تغییرات وزن را برای تقریب مشتق دوم تعیین می‌کند برابر با $5e^{-5}$ در نظر گرفته شد و پارامتر لامبدا[6] که نامحدود بودن ماتریس هسین[7] را کنترل می‌کند برابر با $5e^{-7}$ در نظر گرفته شد. تعداد تکرار آموزش ۳۰۰، مقدار اولیه نرخ آموزش ۰/۰۰۵ و همچنین میزان خطای هدف به‌عنوان یکی از معیارهای پایان آموزش، صفر در نظر گرفته شد.

روش مدل درخت خطی محلی[8]

مدل درخت خطی محلی جزء مدل‌های فازی-عصبی خطی محلی دسته‌بندی می‌گردد. این الگوریتم فضای مسئله را به تعدادی مدل خطی محلی یا LLM تقسیم می‌نماید و پس از پیدا کردن بدترین LLM (LLM با خطای بیشتر) با تقسیم آن به دو LLM، الگوریتم را ادامه می‌دهد (Hajian et al., 2012; Fink et al., 2005). از آنجا که توابع این روش در متلب وجود ندارد در این تحقیق از جعبه‌ابزار شبکه‌های مدل محلی (Hartmann et al., 2012a) که کد متلب آن موجود است (Hartmann et al., 2012b) در نرم‌افزار استفاده شده است. پارامترهای آن پارامترهای پیش‌فرض این جعبه‌ابزار می‌باشد و معیار خطای استاندارد نسبی[9] به‌عنوان تابع جریمه محلی[10] و خطای جذر مربع متوسط نرمالیزه شده[11] به‌عنوان تابع جریمه در نظر گرفته شده است.

روش تحلیل ممیزی (آنالیز تشخیصی[12])

تحلیل تشخیصی ترکیب خطی دو یا چند متغیر مستقل را که به بهترین وجه تفاوت بین دو یا چند گروه را تبیین می‌کند، نشان می‌دهد (Kalantari, 2008; Johnson & Wichern, 2008; Pardoe et al., 2007; Pinto et al., 2012). در این تحقیق نوع خطی تحلیل ممیزی استفاده شده و پارامترهای آن پارامترهای پیش‌فرض متلب در نظر گرفته‌شده است.

روش درخت تصمیم

درخت تصمیم با مرتب نمودن نمونه‌ها در درخت، آن‌ها را از گره ریشه به سمت گره‌های برگ دسته‌بندی می‌نماید (Coppersmith et al., 1999). از مزایای درخت تصمیم این است که نسبت به نویز داده‌های ورودی مقاوم است (Toloei-Ashlaghi et al., 2013). درخت تصمیم از چند الگوریتم

1- Number of epochs (epochs)
2- Initial learning rate (lr)
3- Pattern Recognition
4- Scaled Conjugate Gradient
5- Sigma
6- Lambda
7- Hessian
8- Local Linear Model Tree (LOLIMOT)
9- Relative standard error
10- Local lost function
11- Normalized Root Mean Squared Error
12- Discriminant Analysis

تشکیل یافته که در این پژوهش از الگوریتم CHAID[1] استفاده شد. این روش در هر گام، متغیر مستقلی که بیشترین اثر متقابل با متغیر هدف دارد را انتخاب می‌کند. حداقل تعداد والد در درخت ۱۰ و حداقل تعداد فرزند ۱ در نظر گرفته شده است؛ یعنی هر گره حداقل مبین ۱۰ داده متعلق به خود باشد و هر گره به جز برگ‌ها حداقل یک گره فرزند داشته باشند. از شاخص تنوع جینی[2] برای تعیین شاخه‌های درخت طبقه‌بندی و متوسط مربع خطا[3] برای درخت رگرسیونی استفاده می‌شود (Loh, 2011).

روش جنگل تصادفی[4]

در روش جنگل‌های تصادفی روش کیسه‌ای[5] به‌کار گرفته‌شده و حاوی چندین درخت تصمیم است که خروجی آن از خروجی‌های درخت‌های انفرادی به‌دست می‌آید (Ebrahimkhani et al., 2012; Khoshgoftaar et al., 2007; Breiman, 2001). برای اجرای آن از جعبه‌ابزار ارائه‌شده توسط کارپسی[6] که در متلب فراخوانی شده، استفاده‌شده است. پارامترهای روش جنگل تصادفی شامل عمق[7] بهینه الگوریتم برابر ۹، تعداد درخت‌های بهینه برابر ۱۰۰ و تعداد قطعات[8] بهینه برابر ۵ به‌دست آمد.

روش K نزدیک‌ترین همسایه[9]

در این روش، بر اساس تشابه متغیرها با یکدیگر، الگوی داده‌ها بدون نیاز به الگوهای از پیش مشخص، طبقه‌بندی می‌گردد. این طبقه‌بندی، نمونه آزمون را متعلق به کلاسی می‌داند که بیشترین آرا را در بین K نزدیک‌ترین همسایگان آن داشته باشد. برای به‌دست آوردن نزدیک‌ترین همسایگان یک نمونه، از فاصله اقلیدسی[10] استفاده‌شده است (Zhan et al., 2006). مقدار k (تعداد همسایگان) ۴ در نظر گرفته شد و روش جستجوی نزدیک‌ترین همسایه[11]، روش فراگیر[12] است که همه فاصله داده‌های مورد آزمون را از همه داده‌ها برای یافتن نزدیک‌ترین همسایه بررسی می‌کند.

ماشین بردار پشتیبان[13]

ماشین بردار پشتیبان، جزء روش‌های هسته‌ای[14] یا توابع دستوری در یادگیری ماشین محسوب می‌شود. این روش چندین هسته مختلف را به‌صورت پیش‌فرض پشتیبانی می‌کند که در این تحقیق از هسته‌های خطی[15]، چندجمله‌ای[16]، شعاع مبنا[17] و مربع افت لولا[18] استفاده شد (Christopher & Bures, 1998). الگوریتم استفاده‌شده برای آموزش ماشین بردار پشتیبان در این تحقیق الگوریتم بهینه حداقل متوالی[19] می‌باشد (Platt, 1998) و درجه چندجمله‌ای برابر با سه و فاکتور مقیاس شده[20] برای هسته تابع شعاع مبنا برابر با یک در نظر گرفته‌شده است.

روش گروهی مدل‌سازی داده‌ها (GMDH[21])

ویژگی خاص آن، قابلیت شناسایی و غربال‌کردن متغیرهای کم اثر ورودی در دوره آموزشی شبکه و حذف آن‌ها از روند شبیه‌سازی در دوره آزمون می‌باشد (Assaleh et al., 2013).

1- Chi-squared Automatic Interaction Detection
2- Gini's Diversity Index(gdi)
3- Mean square error
4- Random Forest
5- Bagging
6- Karpathy's toolbox
7- Depth
8- Splits
9- K Nearest Neighbor (KNN)
10- Euclidean Distance
11- Nearest neighbor search method (NSmethod)
12- Exhaustive
13- Support Vector Machine
14- Kernel
15- Linear
16- Polynomial
17- Radial Basis
18- Squared Hinge-Loss
19- Sequential Minimal Optimization
20- Scaling factor
21- Group Method of Data Handling

جدول ۲- نتایج همبستگی بین پارامترهای اقلیمی و عملکرد و نواحی کشت زعفران

Table 2- The results of correlation between climatological parameters, saffron yield and cultivation areas

	EVP	SH	RF	MAH	MIH	RH6.3	RH12.3	RH18.3	MIT	MAT	MT	SY	CN
EVP	1												
SH	0/468***	1											
RF	-0/529***	-0/591***	1										
MAH	-0/504***	-0/412***	0/506	1									
MIH	-0/191**	-0/363***	0/300***	0/215**	1								
RH6.3	-0.642***	-0/633***	0.690***	0.748***	0.524***	1							
RH12.3	-0.641***	-0.629***	0.718***	0.689***	0.553***	0.951***	1						
RH18.3	-0.657***	-0.563***	0.605***	0.765***	0.440***	0.929***	0.852***	1					
MIT	0.671***	0.420***	-0.539***	-0.565***	-0.243***	-0.750***	-0.688***	-0.800***	1				
MAT	0.732***	0.488***	-0.630***	-0.523***	-0.293***	-0.771***	0.755***	-0.747***	0.957***	1			
MT	0.585***	0.343***	-0.445***	-0.566***	-0.170**	-0.689***	-0.594***	-0.803***	0.969***	0.861***	1		
SY	-0.082	-0.156*	0.043	-0.268***	-0.028	-0.150*	-0.071	-0.163**	0.108	0.044	0.156*	1	
CN	0.36***	0.32***	-0.34***	-0.77***	-0.16*	-0.60***	-0.55***	-0.67***	0.55***	0.46***	0.59***	0.38***	1

*، ** و ***: به ترتیب معناداری در سطح ۰/۱، ۰/۰۵ و ۰/۰۱.

*,**and *** indicate significance levels at 0.1, 0.05 and 0.01, respectively.

RF: بارندگی (میلی‌متر) Rainfall (mm)

MAH: رطوبت مطلق حداکثر (گرم بر مترمکعب)Maximum Absolute Humidity (g m^{-3})

RH6.3: رطوبت نسبی(ساعت۶:۳۰)Relative humidity (6:30 o'clock)

RH12.3: رطوبت نسبی(ساعت۱۲:۳۰)Relative humidity (12:30 o'clock)

RH18.3: رطوبت نسبی(ساعت۱۸:۳۰)Relative humidity (18:30 o'clock)

MT: میانگین دما (سانتی‌گراد) Mean Temperature(°c)

EVP: تبخیر (میلی‌متر) Evaporation (mm)

MIH: رطوبت مطلق حداقل (گرم بر متر مکعب)Minimum Absolute Humidity (g m^{-3})

SH: ساعت آفتابی Sunny hours

MIT: دمای مطلق حداقل Min Absolute Temperature (°c)

MAT: دمای مطلق حداکثر Max Absolute Temperature (°c)

SY: عملکرد (کیلوگرم در هکتار) Yield ($kg.ha^{-1}$)

CN: کد شهرستان County Number

این مدل شامل لایه پنهان و نرون‌هایی در این لایه می‌باشد که بیشترین مقدار نرون در هر لایه ۱۵، بیشترین تعداد لایه چهار و پارامتر فشار انتخاب[1] ۰/۶ در نظر گرفته شد (Tauser & Burya, 2011).

نتایج و بحث

خروجی همبستگی بین متغیرها همراه با سطح معناداری آن‌ها در جدول ۲ آمده است. نتایج نشان داد که عملکرد زعفران با عوامل اقلیمی ساعات آفتابی، رطوبت مطلق حداکثر، رطوبت نسبی ساعت۱۸:۳۰، رطوبت نسبی ساعت ۶:۳۰ و دمای مطلق حداقل همبستگی معنی‌داری دارد، لذا عوامل اقلیمی مورداستفاده را می‌توان در عملکرد محصول، مهم و مناسب تلقی نمود.

میزان بارش، دما و ساعات آفتابی از جمله مهم‌ترین پارامترهای تأثیرگذار در عملکرد زعفران در خراسان رضوی می‌باشد (Bashiri et al., 2014؛ Shirdeli & Tawassoli,

1- Selection pressure parameter

2015). نتایج همبستگی نشان داد که بین شرایط اقلیمی و میزان عملکرد زعفران و همچنین بین شرایط اقلیمی و ناحیه زیر کشت، ارتباط خوبی وجود دارد. بر این اساس می‌توان مدل مناسبی جهت پیش‌بینی عملکرد محصول و تعیین مناطق مستعد کشت ارائه نمود. در این صورت می‌توان بررسی نمود که در سایر مناطق با شرایط اقلیمی متفاوت و خارج از ورودی‌های تحقیق که تحت کشت زعفران قرار ندارد، می‌توان کشت زعفران را پیشنهاد نمود یا خیر.

پارامترهای روش جنگل تصادفی به‌صورت عمق[1] بهینه الگوریتم برابر ۹، تعداد درخت‌های بهینه برابر ۱۰۰ و تعداد قطعات[2] بهینه برابر ۵ به‌دست آمد. در روش شبکه عصبی هم با آزمون تعداد لایه‌های پنهان مختلف و تعداد نورون‌های متفاوت، شبکه عصبی با دو لایه پنهان، شامل ۶ نورون در لایه پنهان اول و ۵ نورون در لایه پنهان دوم، به‌دلیل دقت بالاتر، به‌عنوان گزینه برتر انتخاب گردید (جدول ۳).

به منظور به‌دست آوردن مدل مناسبی که امکان پیش‌بینی عملکرد در واحد سطح و همچنین تعیین ناحیه از روی فاکتورهای اقلیمی را داشته باشد، روش‌های مختلفی ارائه شد و دقت و عملکرد هر کدام از روش‌ها نیز در تعیین مدل بررسی گردید که نتایج به تفکیک پیش‌بینی منطقه کشت از روی پارامترهای اقلیمی و پیش‌بینی میزان عملکرد از روی این پارامترها در جدول‌های ۴ و ۶ ارائه شده است.

در پیش‌بینی مناطق کشت زعفران بر مبنای عوامل اقلیمی، الگوریتم‌های مورد استفاده، دقت مناسب‌تری در مراحل آموزش و آزمون داشتند. بهترین طبقه‌بندی مربوط به روش‌های آنالیز تشخیص (ضریب تعیین به ترتیب ۰/۹۹ و ۰/۹۶ در مرحله آموزش و آزمون) و پس از آن K نزدیک‌ترین همسایه (ضریب تعیین به ترتیب ۱ و ۰/۹۴ در مرحله آموزش و آزمون) و ماشین‌های بردار پشتیبان خطی (ضریب تعیین به ترتیب ۰/۹۹ و ۰/۹۲ در مرحله آموزش و آزمون) بود. اما در پیش‌بینی میزان عملکرد محصول بر مبنای عوامل اقلیمی، دقت الگوریتم‌ها کاهش چشمگیری پیدا کرد و مناسب‌ترین الگوریتم‌ها در این مرحله، K نزدیک‌ترین همسایه (ضریب تعیین به ترتیب ۱ و ۰/۱۸ در مرحله آموزش و آزمون) و آنالیز تشخیص (ضریب تعیین به ترتیب ۰/۷۴ و ۰/۰۹ در مرحله آموزش و آزمون) به‌دست آمد.

برای بررسی مهم‌ترین فاکتورهای تعیین‌کننده ناحیه کشت زعفران از روش رگرسیون گام به گام[3] نیز استفاده شد که حداکثر رطوبت مطلق با همبستگی ۰/۷۷- ، حداقل دمای مطلق با همبستگی ۰/۵۵ و حداکثر دمای مطلق با همبستگی ۰/۴۶ به‌عنوان مهم‌ترین عوامل تأثیرگذار بر روی تعیین ناحیه کشت تعیین شدند. جدول ۵ پیش‌بینی الگوریتم‌های مختلف با این سه فاکتور را نشان می‌دهد. همان‌طور که در جدول ۵ مشاهده می‌شود، در مقایسه با جدول ۴ که همه پارامترهای اقلیمی در پیش‌بینی ناحیه مشارکت دارند تفاوت چندانی وجود ندارد و این سه عامل توانسته‌اند به‌خوبی ناحیه کشت زعفران را پیش‌بینی کنند.

در پیش‌بینی مناطق کشت زعفران، میزان میانگین مربع خطا در مرحله آموزش پایین‌تر از مرحله آزمون بود و در هر دو مرحله اخیر، آنالیز نزدیک‌ترین همسایه و ماشین‌های بردار پشتیبان، کمترین مقادیر خطا را دارا بودند. در پیش‌بینی میزان عملکرد محصول، خطای کلی الگوریتم‌ها کاهش یافت و در مرحله آموزش، آنالیز نزدیک‌ترین همسایه و درخت رگرسیونی دارای پایین‌ترین مقادیر بودند. در مرحله آزمون مقادیر خطا نسبت به مرحله آموزش افزایش نشان داد و روش‌های مدل‌سازی گروهی داده‌ها و درخت خطی محلی کمترین مقادیر را دارا بودند.

1- Depth
2 Splits

3- stepwise

جدول ۳- نتایج آزمون شبکه‌های مختلف عصبی با تعداد نورون‌های متفاوت

Table 3- The evaluation results of various neural networks with different neuron numbers

الگوریتم آموزش شبکه Algorithm for network Training		لونبرگ -مارکوآرت Levenberg-Marquardt		سرعت یادگیری متغیر Variable learning rate		تنظیم به روش بیز Bayes regulation method		گرادیان توأم مقیاس شده Scaled conjugate gradient	
تعداد نورون لایه اول Neuron number in the first layer	تعداد نورون لایه دوم Neuron number in the second layer	میانگین مربع خطا (MSE)	ضریب تعیین (R^2)	میانگین مربع خطا (MSE)	ضریب تعیین (R^2)	میانگین مربع خطا (MSE)	ضریب تعیین (R^2)	میانگین مربع خطا (MSE)	ضریب تعیین (R^2)
2	0	1.10	0.40	3.97	0.13	2.56	0.27	3.44	0.29
2	3	2.31	0.31	4.53	0.16	2.44	0.27	1.83	0.36
2	5	1.36	0.41	3.46	0.19	1.70	0.34	3.51	0.20
4	0	1.50	0.39	3.20	0.19	1.71	0.36	1.91	0.26
4	1	1.74	0.30	4.70	0.16	1.71	0.30	1.84	0.30
4	4	1.24	0.44	4.87	0.21	2.17	0.23	2.79	0.20
4	5	1.84	0.36	3.26	0.13	2.04	0.39	1.76	0.29
6	0	1	0.47	3.70	0.19	1.61	0.30	1.31	0.37
6	3	0.66	0.51	3.27	0.09	1.81	0.24	3.63	0.17
6	5	0.80	0.53	4.43	0.14	1.23	0.43	2.07	0.34
8	0	1.11	0.44	5.09	0.14	1.10	0.37	2.23	0.24
8	1	1.13	0.41	4.17	0.11	3.07	0.20	2.26	0.26
8	4	1.33	0.47	3.94	0.13	2.26	0.30	1.17	0.36
8	5	1	0.51	3.84	0.11	1.59	0.41	1.76	0.31
10	0	1.87	0.43	3.51	0.17	1.79	0.31	1.79	0.36
10	2	1.31	0.34	4.14	0.19	1.64	0.36	1.93	0.33
10	3	0.94	0.50	4.30	0.19	2.39	0.23	1.76	0.33
10	5	1.64	0.40	3.20	0.17	1.64	0.41	1.40	0.36

جدول ٤- میزان میانگین مربع خطا (MSE) و ضریب تعیین (R^2) برای روش‌های مختلف داده‌کاوی در پیش‌بینی منطقه کشت زعفران

Table 4- The values of mean squared error and coefficient of determination for different data mining methods in the estimation of saffron cultivation area

مدل Model	MSE (آموزش) (Training)	MSE (آزمون) (Testing)	R^2 (آموزش) (Training)	R^2 (آزمون) (Testing)
رگرسیون خطی Linear regression	2.43	2.73	0.24	0.17
رگرسیون درجه دوم خالص Pure quadratic regression	1.10	1.64	0.42	0.34
رگرسیون اثر متقابل Interaction regression	0.61	2.71	0.51	0.40
رگرسیون درجه دوم Quadratic regression	0.54	3.73	0.54	0.31
شبکه عصبی Neural network	0.66	0.81	0.61	0.53
تشخیص الگو Pattern recognition	1.07	1.14	0.44	0.34
درخت خطی محلی Local linear model tree	0.73	0.99	0.48	0.43
آنالیز تشخیص Discriminant analysis	0.01	0.51	0.99	0.96
درخت طبقه‌بندی Classification tree	0.28	2.15	0.96	0.74
درخت رگرسیونی Regression tree	0.28	1.65	0.93	0.73
جنگل تصادفی Random forest	0.26	0.84	0.96	0.89
K نزدیک‌ترین همسایه K nearest neighbor	0.00	0.31	1.00	0.94
ماشین بردار پشتیبان (خطی) Support vector machine (Linear)	0.01	0.42	0.99	0.93
ماشین بردار پشتیبان (چندجمله‌ای) Support vector machine (Polynomial)	0.00	0.96	1.00	0.88
ماشین بردار پشتیبان (شعاع مبنا) Support vector machine (Radial basis)	0.01	0.51	0.99	0.94
ماشین بردار پشتیبان (مربع افت لولا) Support vector machine (Squared Hinge-Loss)	0.50	0.74	0.68	0.46
مدل‌سازی گروهی داده‌ها Group method of data handling	1.34	2.21	0.38	0.34

جدول ۵- میزان میانگین مربع خطا (MSE) و ضریب تعیین (R^2) برای روش‌های مختلف داده‌کاوی در پیش‌بینی منطقه کشت زعفران با استفاده از سه پارامتر مؤثر

Table 5- The values of mean squared error and coefficient of determination for different data mining methods in the estimation of saffron cultivation area using three effective factors

مدل Model	MSE (آموزش) (Training)	MSE (آزمون) (Testing)	R^2 (آموزش) (Training)	R^2 (آزمون) (Testing)
رگرسیون خطی Linear regression	2.13	2.82	0.22	0.16
رگرسیون درجه دوم خالص Pure quadratic regression	1.40	1.54	0.37	0.32
رگرسیون اثر متقابل Interaction regression	1.28	1.82	0.40	0.38
رگرسیون درجه دوم Quadratic regression	1.23	1.73	0.41	0.39
شبکه عصبی Neural network	1.40	1.62	0.43	0.37
تشخیص الگو Pattern recognition	1.50	2.20	0.31	0.28
درخت خطی محلی Local linear model tree	1.10	1.30	0.47	0.41
آنالیز تشخیص Discriminant analysis	0.36	0.41	0.92	0.89
درخت طبقه بندی Classification tree	0.41	1.50	0.92	0.75
درخت رگرسیونی Regression tree	0.39	0.43	0.84	0.70
جنگل تصادفی Random forest	0.32	0.45	0.92	0.85
K نزدیک‌ترین همسایه K nearest neighbor	0.00	0.86	1.00	0.83
ماشین بردار پشتیبان (خطی) Support vector machine (Linear)	0.15	0.31	0.93	0.91
ماشین بردار پشتیبان (چندجمله‌ای) Support vector machine (Polynomial)	0.27	0.36	0.94	0.90
ماشین بردار پشتیبان (شعاع مبنا) Support vector machine (Radial basis)	0.38	0.41	0.92	0.90
ماشین بردار پشتیبان (مربع افت لولا) Support vector machine (Squared Hinge-Loss)	0.90	1.20	0.50	0.37
مدل‌سازی گروهی داده‌ها Group method of data Handling	1.30	2.10	0.32	0.23

جدول ٦- میزان میانگین مربع خطا (MSE) و ضریب تعیین (R^2) برای روش‌های مختلف داده‌کاوی در پیش‌بینی میزان عملکرد

Table 6- The values of mean squared error and coefficient of determination for different data mining methods in estimation of yield

مدل Model	MSE (آموزش) (Training)	MSE (آزمون) (Testing)	R^2 (آموزش) (Training)	R^2 (آزمون) (Testing)
رگرسیون خطی Linear regression	0.03	0.05	0.28	0.02
رگرسیون درجه دوم خالص Purequadratic regression	0.03	0.05	0.41	0.01
رگرسیون اثر متقابل Interaction regression	0.02	0.12	0.66	0.02
رگرسیون درجه دوم Quadratic regression	0.01	0.26	0.72	0.01
شبکه عصبی Neural network	0.03	0.05	0.25	0.01
تشخیص الگو Pattern recognition	0.03	0.06	0.24	0.03
درخت خطی محلی Local linear model tree	0.03	0.05	0.39	0.05
آنالیز تشخیص Discriminant analysis	0.01	0.07	0.74	0.09
درخت طبقه‌بندی Classification tree	0.06	0.09	0.32	0.07
درخت رگرسیونی Regression tree	0.01	0.08	0.80	0.04
جنگل تصادفی Random forest	0.05	0.05	0.17	0.14
K نزدیک‌ترین همسایه K nearest neighbor	0.00	0.08	1.00	0.18
مدل‌سازی گروهی داده‌ها Group method of data handling	0.03	0.04	0.30	0.04

گرچه اختلاف چندانی بین مقادیر خطا در الگوریتم‌های مختلف طبقه‌بندی مشاهده نشد. لازم به ذکر است که در این مقاله مقادیر عملکرد زعفران بین صفر و یک نرمالیزه شده‌اند به‌طوری‌که داده های اولیه دارای حداقل میزان عملکرد برابر ۰/۳ کیلوگرم بر هکتار، حداکثر آن ۷/۱۷ و میانگین آن ۳/۳۲ کیلو-گرم بر هکتار می‌باشند و میانگین مربع خطا بر اساس داده‌های نرمالیزه شده محاسبه گردیده است.

دقت کاربرد روش آنالیز تشخیص در طبقه‌بندی داده‌های اقلیمی بر اساس فاکتور بارش به اثبات رسیده است (Ashrafi, 2011). اخیراً توجه خاصی به کاربرد تکنیک‌هایی مبتنی بر روش K نزدیک‌ترین همسایه برای تولید داده‌های هواشناسی پدید آمده است و پژوهش‌های مختلف، نشان‌دهنده کاربرد موفقیت‌آمیز این روش‌ها در مدل‌سازی داده‌های هواشناسی می-باشد (Bashiri & Salari, 2016؛ Golabi et al., 2013؛ Rajagopalan & Lall, 1999). ماشین‌های بردار پشتیبان نیز از جمله روش‌های نسبتاً جدیدی محسوب می‌گردند که در تحقیقات اخیر، در مقایسه با روش‌های قدیمی‌تر از دقت بالاتری برخوردار بوده است (Christopher & , 1998 Burges).

نتیجه‌گیری

نتایج پژوهش حاضر نشان داد که از بین پارامترهای اقلیمی، رطوبت مطلق حداکثر، بیشترین همبستگی را با عملکرد زعفران دارد، ضمن آنکه با استفاده از پارامترهای سهل‌الوصول و الگوریتم‌های داده‌کاوی، می‌توان با دقتی مناسب (با ضرایب تعیین نزدیک به ۱ در مراحل آموزش و آزمون الگوریتم طبقه-بندی) اقدام به تفکیک مناطق کشت نمود. عمده مناطق با عملکرد بالای محصول، دارای آب‌وهوای گرم و خشک با زمستان‌های سرد است. با توجه با اینکه این مناطق با محدودیت شدید منابع آبی و بارش کم مواجه‌اند و زعفران نیاز آبی کمی دارد، ضمن آنکه این بخش‌ها، از نظر اقلیمی مناسب کشت تشخیص داده‌شده است، با حرکت به سمت مناطق شمالی استان، میزان تناسب اقلیمی اراضی به‌طور قابل‌ملاحظه‌ای کاهش می‌یابد. در نتیجه بر مبنای تشابه اقلیمی مناطق، می‌توان کشت محصول را توسعه داد.

در طبقه‌بندی مناطق کشت زعفران می‌توان از آنالیز تشخیص و یا ماشین بردار پشتیبان که به‌ترتیب از دقت بالاتری در پیش‌بینی عملکرد برخوردار بوده‌اند استفاده نمود. هر چند در پیش‌بینی میزان عملکرد محصول بر اساس پارامترهای اقلیمی، دقت مدل‌ها کاهش محسوسی پیدا کرد اما در بین مدل‌های واسنجی شده، کاربرد روش نزدیک‌ترین همسایه به دلیل دقت پیش‌بینی بالاتر در هر دو مرحله آموزش و آزمون، نتایج قابل‌اعتمادی در پیش‌بینی عملکرد محصول ارائه خواهد نمود. نهایتاً در راستای پژوهش حاضر پیشنهاد می‌شود با سایر روش-های داده‌کاوی، اقدام به تکرار و مقایسه نتایج گردد. همچنین ورود متغیرهای دیگر اقلیمی، خاک و کیفیت آب با هدف مدل‌سازی، می‌تواند در بهبود روش‌های تشخیص مناطق مستعد کشت زعفران مؤثر واقع شود.

منابع

Abraham, A. 2005. Artificial neural networks, Presented in book: Handbook of Measuring System Design, Oklahoma State University, Stillwater, USA. pp. 902-908.

Alvarez, R. 2009. Predicting average regional yield and production of wheat in the argentine pampas by an artificial neural network approach. European Journal of Agronomy 30 (2): 70-77.

Ashrafi, S. 2010. Precipitation classification of northwest of Iran using cluster analysis and discriminate analysis. Journal of Climate Research 1 (3-4): 27-44. (In Persian).

Assaleh, K., Shanableh, T., and Kheil, Y.A. 2013. Group method of data handling for modeling magnetorheological dampers. Intelligent Control and Automation 4: 70-79.

Azadeh, A., Ghaderi, S.F., Anvari, M., and Saberi, M. 2006. Measuring performance electric power generations using artificial neural networks and fuzzy clustering. Proceedings of the 32nd Annual Conference of the IEEE Industrial Electronics Society (IECON'06), Conservatoire National des Arts and Metiers, Paris, France.

Bagheri, S., Gheysari, M., Ayoubi, Sh., and Lavaee, N. 2012. Silage maize yield prediction using artificial neural networks. Journal of Plant Production 19 (4): 77-96. (In Persian).

Bashiri, M., and Salari, A. 2016. Using geostatistics for zoning prone areas of saffron cultivation in Razavi-Khorasan province based on climatological parameters. Journal of Saffron Agronomy and Technology 4 (2): 155-167. (In Persian).

Bashiri, M., Seyedi, S.M., and Tosan, M. 2014. The effect of sunny hours numbers on saffron yield in Khorasan-Razavi province. 3rd National Congress on Saffron Scientific-Research Achievements, Torbat Heydarieh, 26-27 November 2014, 3pp. (In Persian).

Basu, J.K., Bhattacharyya, D., and Kim, T.H. 2010. Use of artificial neural network in pattern recognition. International Journal of Software Engineering and its Applications 4 (2): 23-34.

Breiman, L. 2001. Random forests. Machine Learning Journal 45 (1): 5-32.

Christopher, J., and Burges, C.A. 1998. Tutorial on support vector machines for pattern recognition. Journal of Data Mining and Knowledge Discovery 2: 121-167.

Coppersmith, D., Hong, S.J., and Hosking, J.R. 1999. Partitioning nominal attributes in decision trees. Journal of Data Mining and Knowledge Discovery 3 (2): 197–217.

Coulibaly, P., Anctil, F., Aravena, R., and Bobee, B. 2001. Artificial neural networks modelling of water table depth fluctuations. Water Resources Research 37 (4): 885-896.

Damangir, H. 2001. Training artificial neural network dynamics and its application in real time prediction of floods. M.Sc thesis, Shiraz University. (In Persian).

Ebrahimkhani, S., Afzali, M., and Shokoohi, A. 2012. Weather and road accidents factors using data mining algorithms. Zanjan Law Students Journal 1 (1): 112-125. (In Persian).

Golabi, M.R., Akhondali, A.M., and Radmanesh, F. 2013. Comparison of the performance of different neural networks algorithm functions in simulation of seasonal precipitation, Case study: Selected stations of Khuzestan province. Journal of Geographical Sciences 13 (30): 151-169. (In Persian).

Hajian, A., Zomorrodian, H., Styles, P., Greco, F., and Lucas, C. 2012. Depth estimation of cavities from microgravity data using a new approach: the local linear model tree (LOLIMOT). Near Surface Geophysics Journal 10 (3): 221-234.

Hartmann, B., Ebert, T., Fischer, T., Belz, J.,

Kampmann, G., and Nelles, O. 2012a. LMNtool - Toolbox zum automatischen Trainieren lokaler Modellnetze. 22nd Workshop of Computational Intelligence, Dortmund, Germany, p. 341-355.

Hartmann, B., Ebert, T., Fischer, T., Belz, J., Kampmann, G., and Nelles, O. 2012b. LMN-Tool: Matlab-Toolbox for Local Model Networks. Siegen University, Deutschland.

Hosseini, M., Mollafilabi, A., and Nassiri, M. 2008. Spatial and temporal patterns in Saffron (*Crocus sativus* L.) yield of Khorasan province and their relationship with long term weather variation. Journal of Iranian Field Crop Research 6 (1): 79-88. (In Persian).

Imrie, C.E., Dean, S., and Korre, A. 2000. River flow using artificial neural network: generalization beyond the calibration range. Journal of Hydrology 233 (1-4): 138-153.

Islam, S., and Kothart, R. 2000. Artificial neural networks in remote sensing of hydrologic processes. Journal of Hydrologic Engineering 5 (2): 138-144.

Johnson, R.A., and Wichern, W.D. 2008. Applied multivariate statistical analysis (6th Edition), Upper Saddle River, NJ, USA, 800 p.

Kalantari, K.h. 2008. Processing and analysis of data on socio-economic research using SPSS software. Tehran, Saba Farhang, 388 p. (In Persian).

Kaul, M., Hill, R.L., and Walthall, C. 2005. Artificial neural networks for corn and soybean yield prediction. Journal of Agricultural System 85 (1): 1-18.

Khoshgoftaar, T.M., Golawala, M., and Van Hulse, J. 2007. An empirical study of learning from imbalanced data using random forest. Proceedings of the 19th IEEE Conference on Tools with Artificial Intelligence 2: 310-317.

Landeras, G., Ortiz-Barredo, A., and Lopez, J.J. 2009. Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. Irrigation and Drainage Engineering, American Society of Civil Engineers 135 (3): 323-334.

Loh, W.Y. 2011. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (1):14-23.

Madala, H.R., and Ivakhnenko, A.G. 1974. Inductive learning algorithms for complex system modeling. Boca Raton: CRC, 373 pp.

Meng, X., Zhang, Z., and Xu, X. 2014. A novel K-Nearest neighbor algorithm based on I-Divergence with application to soil moisture estimation in maize field. Journal of Software 9 (4): 841-846.

Merdun, H., Çınar, O., Meral, R., and Apan, M. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil and Tillage Research 90 (1-2): 108-116.

Moradkhani, H., Hsu, K.L., Gupta, H.V., and Sorooshian, S. 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. Journal of Hydrology 295 (1-4): 246-262.

Mubiru, J. 2008. Predicting total solar irradiation values using artificial neural networks. Renewable Energy journal 33 (10): 2329-2332.

Mukerji, A., Chatterjee, C., and Raghuwanshi, N.S. 2009. Flood forecasting using ANN, Neuro-fuzzy and Neuro-GA models. Journal of Hydrologic Engineering. American Society of Civil Engineers 14 (6): 647-652.

Pardoe, I., Yin, X., and Cook, R.D. 2007. Graphical tools for quadratic discriminant analysis. Technometrics Journal 49 (2): 172-183.

Pinto, U., Maheshwar, B., Shrestha, S., and Morris, C. 2012. Modeling eutrophication and microbial risks in peri-urban river systems using discriminant function analysis. Water Research Journal 4 (6): 6476- 6488.

Piri, J., Amin, S., Moghaddamnia, A., Keshavarz,

A., Han, D., and Remesan, R. 2009. Daily pan evaporation modeling in a hot and dry climate. Journal of Hydrologic Engineering, American Society of Civil Engineers 14 (8): 803-811.

Platt, J. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Microsoft Research TechReport, Number MSR-TR-98-14: p. 21.

Rajagopalan, B., and Lall, U. 1999. A k-nearest neighbor simulator for daily precipitation and other variables. Water Resources Research 35 (10): 3089–3101.

Samsudin, R., Saad, P., and Shabri, A. 2008. The GMDH model and its application to forecasting of rice yields. Journal Teknologi Maklumat 20 (4): 113-124.

Shirdeli, A., and Tawassoli, A. 2015. Predicting yield and water use efficiency in saffron using models of artificial neural network based on climate factors and water. Saffron Agronomy and Technology 3 (2): 121-131. (In Persian).

Siemsen, E., Roth, A., and Oliveira, P. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. Organizational Research Methods 13 (3): 456-476.

Smith, B.A., Hoogenboom, G., and McClendon, R.W. 2009. Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture 68 (1): 52-61.

Tamura, H., and Kondo, T. 1980. Heuristics free group method of data handling algorithm of generating optimal partial polynomials with application to air pollution prediction. International Journal of Systems Science 11 (9): 1095-1111.

Tauser, J., and Buryan, P. 2011. Exchange Rate Predictions in International Financial Management by Enhanced GMDH Algorithm. Prague Economic Papers 20 (3): 232-249.

Toloei Ashlaghi, A., Poorebrahimi, A., Ebrahimi, M., and Ghasemahmad, L. 2013. Using data mining techniques for prediction breast cancer recurrence. Iranian Journal of Breast Diseases 5 (4): 23-34. (In Persian).

Zealand, C.M., Bum, D.H., and Simonovic, S.P. 1999. Short term stream flow forecasting using Artificial Neural Networks. Journal of Hydrology 214 (1-4): 32-48.

Zhan, Y., Chen, H., and Zhang, G.C. 2006. An optimization algorithm of K-NN classification. IEEE International Conference on Machine Learning and Cybernetics, Dalian, China, pp. 2246-2251.

Climatic zonation and land suitability determination for saffron in the Khorasan-Razavi province using data mining algorithms

Mahdi Bashiri[1], Ali Maroosi[2*], Amir Salari[3] and Mohamad Ghodoosi[4]

Submitted: 6 September, 2016 **Accepted:** 22 January, 2017

Bashiri, M., Maroosi, A., Salari, A., and Ghodoosi, M. 2018. Climatic zonation and land suitability determination for saffron in Khorasan-Razavi province using data mining algorithms. Saffron Agronomy & Technology 5(4):379-392.

Abstract

Yield prediction for agricultural crops plays an important role in export-import planning, purchase guarantees, pricing, secure profits and increasing agricultural productivity. Crop yield is affected by several parameters especially climate. In this study, saffron yield in the Khorasan-Razavi province was evaluated by different classification algorithms including artificial neural networks, regression models, local linear trees, decision trees, discriminant analysis, random forest, support vector machine and nearest neighbor analysis. These algorithms analyzed data for 20 years (1989-2009) including 11 climatological parameters. The results showed that a few number of climatological parameters affect saffron yield. The minimum, mean and maximum of temperature have the highest positive correlations and the relative humidity of 6.5h, sunny hours, relative humidity of 18.5h, evaporation, relative humidity of 12.5h and absolute humidity have the highest negative correlations with saffron cultivation areas, respectively. In addition, in classification of saffron cultivation areas, the discriminant analysis and support vector machine had higher accuracies. The correlation between saffron cultivation area and saffron yield values was relatively high (r=0.38). The nearest neighbor analysis had the best prediction accuracy for classification of cultivation areas. For this algorithm, the coefficients of determination were 1 and 0.944 for the training and testing stages, respectively. However, the algorithms accuracy for prediction of crop yield from climatological parameters was low (the average coefficients of determination equal to 0.48 and 0.05 for training and testing stages). The best algorithm i.e. nearest neighbor analysis had coefficients of determination equal to 1 and 0.177 for saffron yield prediction. The results showed that using climatological parameters and data mining algorithms can classify cultivation areas. Using this approach, it is possible to identify areas that have similar climate to prone areas and recognize suitable areas for cultivation.

Keywords: Separation, Singulation, Saffron flower, Aligning.

1 -Assistant Professor, Department of Range and Watershed Management, Faculty of Agriculture and Natural Resources, University of Torbat Heydarieh

2- Assistant Professor, Department of Computer Engineering and Information Technology, Faculty of Technical and Engineering, University of Torbat Heydarieh

3-Assistant Professor, Department of Plant Production, Faculty of Agriculture and Natural Resources, University of Torbat Heydarieh

4- Lecturer, Department of Industrial Engineering, Faculty of Technical and Engineering, University of Torbat Heydarieh

(Corresponding Author Email: ali.maroosi@torbath.ac.ir;ali.maroosi@gmail.com)

DOI: 10.22048/jsat.2017.60768.1189