

توسعه یک روش کارآمد مبتنی بر آموزش ماشین و داده‌های زمینی و ماهواره‌ای برای پیش‌بینی و پهنه‌بندی خشکسالی کشاورزی در ایران

چکیده خشکسالی یکی از خطرناک‌ترین وقایع طبیعی برای هر کشور به حساب می‌آید. به همین علت نظارت و پیش‌بینی این پدیده در دنیای امروز از اهمیت بسیار بالایی برخوردار است. در این پژوهش، از شاخص خشکسالی SPEI که از معروف‌ترین شاخص‌های خشکسالی به شمار می‌رود، برای پیش‌بینی خشکسالی استفاده شده است. برای محاسبه این شاخص، ترکیبی از داده‌های زمینی و سنجش از دور مورد استفاده قرار گرفته است. با توجه به وزن مختلف هر کدام از این داده‌ها در محاسبه شاخص خشکسالی، ابتدا پارامترهای مؤثر توسط روش‌های انتخاب پارامتر مؤثر مانند روش فیلتر و روش لاسو برگزیده شده و به عنوان پارامترهای ورودی مدل در نظر گرفته شدند. سپس با استفاده از ۴ روش یادگیری ماشین شامل جنگل تصادفی تصحیح بایاس شده، جنگل تصادفی، ماشین بردار پشتیبان و پرسپترون چندلایه شاخص خشکسالی محاسبه شده است. برای اعتبارسنجی نتایج هر کدام از این مدل‌ها، از شاخص‌های $RMSE$ ، R^2 ، MSE و MAE استفاده شده و سپس بر اساس مقادیر این شاخص‌ها، مدل جنگل تصادفی تصحیح بایاس شده با مقادیر $R^2 = 0.9858$ و $RMSE = 0.1190$ برای خوشه ۱ و $R^2 = 0.9809$ و $RMSE = 0.1375$ برای خوشه ۲ به عنوان مدل دارای بهترین عملکرد انتخاب شده است. در نهایت از نتایج مدل بهینه برای تهیه نقشه‌های پهنه‌بندی خشکسالی به منظور شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی استفاده گردید. این نقشه‌ها می‌توانند اطلاعات ارزشمندی را در رابطه با توزیع و شدت خشکسالی در مناطق مختلف در اختیار ذینفعان و متخصصان قرار دهند.

واژه‌های کلیدی خشکسالی، سنجش از دور، یادگیری ماشین، شاخص خشکسالی SPEI، نقشه پهنه‌بندی خشکسالی

Developing a Robust Data-Driven Model Based on Ground and Satellite Measured Data for Agricultural Drought Prediction in Iran

Abstract Drought is considered one of the most hazardous natural phenomena for any country. Therefore, monitoring and forecasting this phenomenon are of critical importance in today's world. To this end, the Standardized Precipitation Evapotranspiration Index (SPEI), one of the most applied drought indices, was employed for drought prediction. To calculate this index, a combination of ground-based and remote sensing data was utilized. Given the different weights of each of these data types in calculating the drought index, the most relevant parameters were first selected using feature selection methods such as the Filter method and the LASSO method and were considered as input parameters for the model. Furthermore, artificial intelligence was utilized to apply various machine learning algorithms, resulting in the development of several models. These algorithms included Bias-Corrected Random Forest, Random Forest, Support Vector Machine, and Multilayer Perceptron. To validate the results of each of these models, indices such as RMSE, R^2 , MSE, and MAE were used. Based on the values of these indices, the Bias-Corrected Random Forest model with $R^2 = 0.9858$ and $RMSE = 0.1190$ for cluster 1 and $R^2 = 0.9809$ and $RMSE = 0.1375$ for cluster 2 was selected as the best-performing model. Finally, the results of the optimized model were used to create drought zoning maps to identify and classify areas prone to drought conditions. These maps can provide valuable information on the distribution and intensity of drought across different regions to stakeholders and experts.

Keywords Drought, Remote sensing, Machine learning, SPEI drought index, Drought maps

مقدمه

امروزه تغییر اقلیم یکی از معضلات بزرگ جهان به شمار می‌رود. در دهه‌های اخیر پیرو گرمایش زمین و افزایش میانگین دمای هوا، فجایع اقلیمی نظیر خشکسالی نیز به گونه‌ای شدت پیدا کرده‌اند که اثرات آن‌ها به وضوح بر اکوسیستم و زندگی انسان قابل مشاهده است. خشکسالی پدیده‌ای خرنده و زیان‌بار است که عموماً به حالتی اطلاق می‌شود که میزان بارندگی در یک مدت زمان مشخص از مقدار آستانه مشخص شده کمتر باشد. از آن‌جا که خشکسالی یک پدیده‌ی طبیعی است، مقابله با اثرات منفی آن تنها در صورتی امکان‌پذیر است که در مورد وقوع آن هشدار قبلی و به موقع‌ای داده شود [1]. به همین جهت وجود یک سیستم کارآمد برای پیش‌بینی این پدیده امری ضروری و اجتناب‌ناپذیر است. بدین منظور استفاده از شاخص‌های خشکسالی که رابطه پیچیده‌ای با اقلیم و محیط دارند، به عنوان روشی مؤثر برای پیش‌بینی و پایش خشکسالی شناخته می‌شود [2]. به طور کلی شاخص‌های خشکسالی ابزارهایی هستند که در تشخیص شروع، شدت و پایان خشکسالی به کمک انسان می‌آیند. نحوه‌ی عملکرد این شاخص‌ها بدین صورت می‌باشد که هزاران بیت داده در مورد بارندگی، برف، پوشش گیاهی و ... را دریافت کرده، آن‌ها را در بازه‌های زمانی مختلف تجزیه و تحلیل کرده و در نهایت به یک تصویر قابل درک تبدیل می‌کنند.

یکی از مهم‌ترین مراحل در نظارت و تحلیل خشکسالی، استفاده از شاخص‌های خشکسالی مناسب است. به‌طور کلی، شاخص خشکسالی یک متغیر اصلی برای تعیین مدت و شدت خشکسالی و ارزیابی تأثیرات آن است. با این حال، به دلیل پیچیدگی تعاریف خشکسالی، ارزیابی جامع و سیستماتیک ویژگی‌های خشکسالی در یک منطقه‌ی مشخص تنها با یک شاخص امکان‌پذیر نیست. در همین راستا، صدها شاخص خشکسالی توسط محققان در سراسر جهان توسعه یافته است.

این شاخص‌ها را می‌توان به دو دسته تقسیم کرد: شاخص‌های مبتنی بر زمین، و شاخص‌های سنجش از دوری. شاخص‌های خشکسالی مبتنی بر زمین معمولاً از اندازه‌گیری‌های زمینی متغیرهای هواشناسی، مانند بارندگی و دما، به دست می‌آیند. این شاخص‌ها قادرند به‌طور دقیق و مؤثر خشکسالی را در اطراف ایستگاه‌های اقلیمی نظارت کنند. برخی از آن‌ها شامل شاخص [3] PDSI، شاخص [4] SPI و شاخص [5] SPEI هستند.

اما برای پوشش گسترده‌ی مکانی شاخص‌های خشکسالی مبتنی بر زمین، شاخص‌های سنجش از دور بر اساس داده‌های ماهواره‌ای توسعه یافته‌اند. داده‌های سنجش از دور ماهواره‌ای شامل اطلاعاتی از قبیل بارندگی، دما، تبخیر و تعرق، و پوشش گیاهی هستند که همگی پوشش پیوسته و لحظه‌ای دارند. در دهه‌های اخیر، شاخص‌های خشکسالی سنجش از دور متعددی مانند شاخص [6] NDVI، شاخص [7] NDDI، و شاخص [8] NMDI توسعه یافته‌اند. با این حال، با وجود اینکه این شاخص‌ها کمبودهای شاخص‌های خشکسالی مبتنی بر زمین در پوشش مکانی را برطرف می‌کنند، اما از آن‌جا که داده‌های آن‌ها نسبتاً کوتاه مدت هستند، نمی‌توانند به طور کامل جایگزین مشاهدات مبتنی بر زمین در بحث نظارت بر خشکسالی شوند. علاوه بر این، کیفیت شاخص‌های خشکسالی سنجش از دور می‌تواند تحت تأثیر شرایط جوی (مانند ابرها) نیز قرار گیرد.

به طور کلی، در موضوع نظارت بر خشکسالی، شاخص‌های خشکسالی مبتنی بر زمین معمولاً دقیق‌تر از شاخص‌های خشکسالی سنجش از دور در نظر گرفته می‌شوند. این امر عمدتاً به این دلیل است که متغیرهای اقلیمی استخراج‌شده از اندازه‌گیری‌های زمینی معمولاً دقیق‌تر هستند. علاوه بر این، اکثر شاخص‌های خشکسالی مبتنی بر زمین به طور مستقیم تعادل میان آب و خاک را نظارت می‌کنند نه اطلاعات غیرمستقیم پوشش گیاهی. همچنین،

بسیاری از شاخص‌های خشکسالی سنجش از دور که به تازگی توسعه یافته‌اند نیز معمولاً نیاز به صحت‌سنجی با استفاده از شاخص‌های مبتنی بر زمین دارند تا اعتبار آن‌ها اثبات گردد [9].

در سال‌های اخیر، کاربردهای مدل‌های یادگیری ماشین پیشرفت چشمگیری در مدل‌سازی شاخص‌های خشکسالی و اقلیم‌شناسی از خود نشان داده‌اند [10]. لذا به‌کارگیری هم‌زمان مدل‌های یادگیری ماشین و شاخص‌های خشکسالی می‌تواند ابزار قدرتمندی برای نظارت، پیش‌بینی و مدیریت خشکسالی‌ها فراهم آورد. در همین راستا تلاش‌های بسیاری توسط محققین متعدد صورت پذیرفته تا بتوان داده‌ها را از منابع مختلف ترکیب کرده و شاخص‌های خشکسالی مبتنی بر زمین را با استفاده از مدل‌های داده‌محور مبتنی بر یادگیری ماشین مانند شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی و ... بازتولید کرد. به عنوان مثال کوثری و همکاران (۲۰۱۷) در مطالعه‌ای یک روش پیش‌بینی خشکسالی منطقه‌ای بر مبنای شبکه‌ی عصبی مصنوعی و شاخص SPI در سری‌های ۳، ۶، ۹، ۱۲، ۱۸ و ۲۴ ماهه در استان فارس ایران برای معرفی مدلی جهت پیش‌بینی خشکسالی ارائه نمودند. در این پژوهش از داده‌های بارندگی ۴۱ ایستگاه باران‌سنجی برای محاسبه مقادیر شاخص بارش استاندارد استفاده شد و آزمایش‌های مکرر و اعتبارسنجی آن‌ها برای به دست آوردن بهترین مدل شبکه‌ی عصبی مصنوعی در نظر گرفته شد. نتایج این پژوهش نشان داد مقادیر پیش‌بینی شده‌ی ۶ ماهه اول به طور میانگین ۷۳ درصد تطابق با مقادیر مشاهده شده را داراست [11]. مختارزاد و همکاران (۲۰۱۷) نیز در پژوهشی برای پیش‌بینی خشکسالی از شبکه عصبی مصنوعی، سیستم استنتاج عصبی فازی تطبیقی و ماشین بردار پشتیبان استفاده کرده و نتایج این روش‌ها را با یکدیگر مقایسه کردند. آن‌ها بدین منظور از تمام داده‌های هواشناسی (دما، رطوبت و بارندگی) ایستگاه بجنورد واقع در

ایران، میان سال‌های ۱۹۸۴ تا ۲۰۱۲ در مقیاس زمانی سه-ماهه استفاده گردید تا در نهایت شاخص SPI به عنوان خروجی به دست آید. نتایج نشان داد که الگوریتم ماشین بردار پشتیبان مدل دقیق‌تری را برای پیش‌بینی ارائه می‌دهد [12]. در پژوهشی دیگر عباسی و همکاران (۲۰۲۰) به بررسی وضعیت خشکسالی در ایستگاه سینوپتیک سقز در مقیاس‌های زمانی مختلف با استفاده از شاخص بارش-تبخیر و تعرق استاندارد شده و الگوریتم ماشین بردار پشتیبان با سه تابع هسته‌ای خطی، چند جمله‌ای و پایه شعاعی و شبکه بیزین پرداختند. برای این منظور از شاخص SPEI در مقیاس‌های زمانی کوتاه‌مدت یک و سه ماهه، میان‌مدت شش و ۱۲ ماهه و بلندمدت ۲۴ و ۴۸ ماهه در طی دوره آماری ۴۹ ساله برای پایش وضعیت خشکسالی در این ایستگاه استفاده شد. نتایج نشان داد که تابع هسته‌ای خطی در روش ماشین بردار پشتیبان نسبت به دو تابع دیگر دقت بیشتری داشته است [13]. توفانر و اوزبایز (۲۰۲۰) نیز در پژوهشی به تخمین شاخص شدت خشکسالی پالم در منطقه‌ی آدیامان واقع در کشور ترکیه پرداختند. در این پژوهش پارامترهای شاخص شدت خشکسالی پالم به کمک چهار روش رگرسیونی و یادگیری ماشین شامل رگرسیون خطی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و درخت تصمیم تخمین زده شدند. نتایج پژوهش نشان داد که الگوریتم‌های استفاده شده با شاخص شدت خشکسالی پالم سازگاری داشته و مدل‌های رگرسیونی محاسبات تخمین شاخص شدت خشکسالی پالم را تسهیل می‌کنند که این امر برای مطالعات آنالیز و پیش‌بینی خشکسالی بسیار کارآمد می‌باشد [14]. فنگ و همکاران (۲۰۱۹) نیز در مطالعه‌ای از سی عامل خشکسالی مبتنی بر روش‌های سنجش از دور برای بازتولید شاخص SPEI میان سال‌های ۲۰۰۱ تا ۲۰۱۷ برای منطقه‌ی نیو ساوت ولز در جنوب شرقی استرالیا استفاده کردند. در این پژوهش سه روش پیشرفته یادگیری ماشین شامل جنگل تصادفی تصحیح

پارامترهای مؤثر و چهار الگوریتم یادگیری ماشین متفاوت به منظور انتخاب دقیق‌ترین مدل ایجاد شده.

مواد و روش‌ها

۱- منطقه مورد مطالعه

منطقه‌ی مورد مطالعه‌ی انتخابی در این پژوهش کشور ایران می‌باشد. ایران با مساحتی بیش از ۱/۶ میلیون کیلومتر مربع، کشوری پهناور محسوب می‌شود. به همین علت و نیز به دلیل گستردگی این مساحت بین عرض جغرافیایی ۲۵ الی ۴۰ درجه‌ی شمالی، مناطق مختلف کشور شرایط اقلیمی و آب و هوایی متنوعی را در فصول مختلف تجربه می‌کنند. به همین منظور، اهمیت کشاورزی و در پی آن اهمیت اعمال دقت دوچندان در پیش‌بینی خشکسالی در مناطق مختلف کشور، با توجه به تفاوت‌های چشمگیر میان این مناطق، بسیار حائز اهمیت است.

لذا در ابتدا با توجه به تقسیمات رسمی استانی در کشور ایران، که به تعداد ۳۱ استان تفکیک شده است، با در نظر گرفتن معیار سهم هر استان از کل مساحت اراضی زراعی کشور، یک یا چند ایستگاه هواشناسی برای هر استان معین و انتخاب شده و سپس تمامی محاسبات برای محل قرارگیری این ایستگاه‌ها صورت پذیرفت. بر اساس اطلاعات به دست آمده از نشریه‌ی سهم و رتبه استان‌ها بر اساس متغیرهای مهم بخش کشاورزی [18] که توسط مرکز آمار ایران منتشر گردیده است، می‌توان استان‌های کشور را بر اساس درصد مساحت اراضی کشاورزی استان از مساحت کل اراضی کشاورزی کشور، بین چهار دسته‌ی مختلف تقسیم‌بندی کرد. در ادامه، پس از تخصیص دادن تمامی استان‌ها به یکی از چهار دسته‌ی تعریف شده، از هر استان بنا به تعلق به هر یک از دسته‌های ذکر شده به ترتیب ۴، ۳، ۲ و ۱ ایستگاه هواشناسی انتخاب گردید. انتخاب این ایستگاه‌ها از میان مجموع ۴۵۵ ایستگاه سینوپتیک متعلق به سازمان هواشناسی کشور صورت گرفته

بایاس شده، ماشین بردار پشتیبان و شبکه عصبی پرسپترون چند لایه، به عنوان مدل‌های رگرسیون در این روش به کار گرفته شدند. نتایج نشان داد که مدل جنگل تصادفی تصحیح بایاس شده از دو مدل دیگر برای پیش‌بینی شاخص عملکرد بهتری دارد [15].

موضوع خشکسالی و سازگاری با آن در ایران نیز از اهمیت بالایی برخوردار است. مجموعه داده‌های شاخص PDSI از سال ۱۹۵۱ تا ۲۰۰۵ نشان داده است که شدت خشکسالی در ایران، به ویژه در شمال غرب و شمال شرق کشور افزایش یافته است [16]. به گونه‌ای که تنها در سال ۲۰۰۱، حدود هشت میلیون هکتار از اراضی کشاورزی ایران تحت تأثیر خشکسالی قرار گرفت و میلیون‌ها دلار خسارت به کشور وارد کرد [17]. لذا لزوم ایجاد مدلی کارآمد برای پیش‌بینی این مخاطره‌ی طبیعی در ایران، امری اجتناب‌ناپذیر به نظر می‌آید. بنابراین، می‌توان اهداف اصلی این پژوهش را بدین ترتیب بیان کرد: (۱) تهیه نقشه‌ی پهنه‌بندی خشکسالی کشاورزی؛ (۲) یافتن مناسب‌ترین پارامترها و شاخص‌های خشکسالی مؤثر در ایجاد مدل پیش‌بینی خشکسالی؛ (۳) اعمال الگوریتم‌های مختلف یادگیری ماشین و یافتن دقیق‌ترین و بهینه‌ترین الگوریتم.

ضمناً می‌توان به طور کلی مهم‌ترین نوآوری‌های این پژوهش را که حد تمایزی بین این مطالعه و سایر مطالعات پیشین انجام شده بر روی منطقه مورد مطالعه می‌باشند، در این موارد خلاصه کرد: (۱) بهره‌گرفتن همزمان از داده‌های زمینی و داده‌های سنجش از دوری به منظور افزایش دقت مدل‌سازی با توجه گستردگی منطقه‌ی مورد مطالعه؛ (۲) افزایش دقت مدل‌سازی در مناطق با اهمیت کشاورزی بالاتر از طریق تخصیص ایستگاه‌های جمع‌آوری داده‌ی بیشتر به استان‌های با سهم بالاتر از مجموع مساحت اراضی زراعی کشور؛ (۳) بهره‌گرفتن از دو روش متفاوت برای انتخاب

است. جدول ۱ نمایانگر جزئیات این دسته‌بندی می‌باشد. **Error! Reference source not found.** نیز محل-

های قرارگیری ایستگاه‌های منتخب را در سراسر کشور نشان می‌دهد. لازم به ذکر است که از ۷۵ ایستگاه پیش‌بینی شده برای این پژوهش، مجموعاً ۴ ایستگاه از استان‌های زنجان، خراسان شمالی، همدان و کرمان به علت عدم پوشش دهی کامل بازه‌ی زمانی مورد نظر در زمینه‌ی فراهم بودن داده از سوی سازمان هواشناسی کشور و نیز ۱ ایستگاه از استان سیستان و بلوچستان به علت عدم پوشش دهی منطقه‌ی موردنظر از سوی سنجنده‌ی مادیس، مورد تامین داده قرار نگرفته و به ناچار حذف گردیدند. مشخصات و اطلاعات کلی این ایستگاه‌ها در ادامه و در قالب جدول ۲ ارائه شده است.

در این پژوهش، بازه‌ی زمانی ۲۲ ساله از دی ماه ۱۳۷۹ هجری شمسی مصادف با ژانویه ۲۰۰۱ میلادی لغایت دی ماه ۱۴۰۱ شمسی مصادف با ژانویه ۲۰۲۳ میلادی انتخاب شده و مورد مطالعه قرار گرفته است.

۲- داده‌ها

خشکسالی به مانند اغلب رویدادهای اقلیمی پدیده‌ای پیچیده است که تحت تأثیر متغیرها و عوامل متعددی قرار دارد. به همین علت، درک این عوامل و متغیرها برای درک علل، اثرات و راهبردهای کاهش احتمالی اثرات چنین رویدادهایی بسیار حائز اهمیت است. در این میان، خشکسالی به تنهایی وابسته به طیف بسیار گسترده‌ای از این متغیرها و عوامل می‌باشد. لذا واضح است که مورد ارزیابی قرار دادن شرایط خشکسالی در حالی که تمامی عوامل تأثیرگذار بر آن در نظر گرفته شده و در پیش‌بینی آن لحاظ گردند، امری طولانی، بسیار هزینه‌بر و نیازمند انجام محاسبات بسیار سنگین است. با این حال، در این پژوهش به دلیل گستردگی منطقه‌ی مورد مطالعه و نیز به منظور افزایش دقت عملکرد مدل، از متغیرها و شاخص‌های خشکسالی مبتنی بر زمین و سنسجش از دوری متعددی استفاده شده است. تمامی این

متغیرها و شاخص‌ها برای بازه‌ی زمانی ۲۲ ساله‌ی مورد مطالعه، جمع‌آوری گردیده و مورد استفاده قرار گرفته‌اند. خلاصه‌ای از این داده‌ها و شاخص‌ها در قالب جدول ۳ ارائه شده است.

در ادامه به منظور دستیابی به مقیاسی واحد، تمامی متغیرها و شاخص‌های استخراج شده به تفکیک ماهانه تغییر مقیاس داده شده و نیز برای در نظر گرفتن پاسخ تأخیری ناشی از اثر تأخیری یک متغیر بر دیگری، علاوه بر مقادیر ماهانه، میانگین متحرک ۳ ماهه برای تمامی داده‌ها و شاخص‌های موجود در این مطالعه، محاسبه شده و نهایتاً از روش Z -score برای استانداردسازی تمامی این داده‌ها استفاده شده است. Z -score یک تکنیک پرکاربرد برای استانداردسازی داده‌ها است که مقادیر را به گونه‌ای تبدیل می‌کند که میانگین آن‌ها صفر و انحراف از معیار آن‌ها یک باشد. این روش به کمک رابطه‌ی ۱ قابل تعریف است. در این رابطه x_s نشان‌دهنده‌ی مقادیر استاندارد شده، x نشان‌دهنده‌ی مقادیر اولیه، μ میانگین تمام داده‌ها و σ انحراف از معیار داده‌ها می‌باشند.

$$x_s = \frac{x - \mu}{\sigma} \quad (1)$$

۳- مدل‌سازی

یکی از ملزومات ایجاد یک مدل دقیق و کارآمد در مدل‌سازی، فراهم کردن شرایطی است که پارامترهای دخیل در مدل دارای شرایط و خصوصیات مشابه باشند. خوشه‌بندی تکنیکی برای گروه‌بندی اشیاء مشابه یا گروه نقاط می‌باشد که در یادگیری ماشین و تجزیه و تحلیل داده‌ها مورد استفاده قرار می‌گیرد. بنابراین، در اولین گام از فرآیند مدل‌سازی، از میان موارد متعدد الگوریتم‌های خوشه‌بندی موجود، الگوریتم K-means [19] که از محبوبیت بالایی نیز برخوردار است، انتخاب شد و مورد استفاده قرار گرفت. در ادامه، پس از بررسی حالات مختلف، تعداد خوشه‌ها ۲ در نظر گرفته شده و سپس

از دو معیار میانگین بارش سالانه و میانگین دمای سالانه برای تخصیص هر یک از ایستگاه‌ها به یکی از خوشه‌ها استفاده شد. این الگوریتم ۶ ایستگاه را به خوشه‌ی دوم و ۶۴ ایستگاه را به خوشه‌ی اول اختصاص داد. در این خوشه‌بندی، ایستگاه‌های *بانه*، *مربوان*، *نوشهر*، *رشت*، *ساری* و *یاسوج* در خوشه‌ی دوم و مابقی ایستگاه‌ها در خوشه‌ی اول قرار گرفتند. همچنین مطابق با این دسته‌بندی، ایستگاه‌های متعلق به خوشه‌ی دوم دارای میانگین بارندگی سالانه‌ی ۹۴۵ میلی‌متر و میانگین دمای سالانه‌ی ۱۵/۷ درجه‌ی سانتی‌گراد بودند. این مقادیر برای خوشه‌ی اول به ترتیب ۲۸۹ میلی‌متر و ۱۷/۱ درجه‌ی سانتی‌گراد بود. این بدین معناست که ایستگاه‌های خوشه‌ی دوم به مراتب مرطوب‌تر و دارای دمای نسبتاً پایین‌تری نسبت به ایستگاه‌های خوشه‌ی اول بودند. از آنجایی که غالب مساحت ایران دارای اقلیم نسبتاً خشکی می‌باشد، تخصیص بیش از ۹۰ درصد ایستگاه‌های مورد مطالعه به یک خوشه، منطقی به نظر می‌رسد. نتایج این خوشه‌بندی در شکل ۲ قابل مشاهده است. در مرحله‌ی بعد، برای هر خوشه، چهار مدل مختلف یادگیری ماشین، یعنی پرسپترون چندلایه (MLP)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF) و جنگل تصادفی تصحیح بایاس شده (BRF) برای پیش‌بینی خشکسالی توسعه داده شدند و بهترین مدل تعیین شد. در نهایت، از نتایج بهترین مدل برای تولید نقشه‌های پهنه‌بندی خشکسالی استفاده گردید. شکل ۳ روند کلی مراحل مختلف این مدل-سازی را در قالب یک فلوجارت نشان می‌دهد. در ادامه به تفصیل به بررسی روش‌های استفاده شده خواهیم پرداخت.

۳-۱- شاخص بارش-تبخیر و تعرق استاندارد شده (SPEI)
شاخص SPEI یک شاخص خشکسالی است که اولین بار توسط سرانو و همکاران [5] مورد استفاده قرار گرفت. این شاخص که از تفاضل مقادیر دو پارامتر بارندگی و تبخیر و تعرق پتانسیل برای تعیین شدت و پایش شرایط خشکسالی استفاده می‌کند، یک شاخص استاندارد شده و بدون بعد است

که امکان مقایسه شدت خشکسالی در مکان‌ها و دوره‌های زمانی مختلف را فراهم می‌کند. از آنجایی که شاخص SPEI از روش Z-score برای استانداردسازی استفاده می‌کند، مقادیر این شاخص عموماً بین ۳- تا ۳ قرار می‌گیرد. در این بازه، مقدار صفر بیانگر شرایط عادی است، در حالی که مقادیر منفی و مثبت به ترتیب شرایط خشک و مرطوب را نشان می‌دهند و به طور کلی، مقادیر کوچک‌تر از ۱- به عنوان شرایط خشکسالی در نظر گرفته می‌شود [15]. از آنجایی که در پژوهش حاضر، شاخص SPEI به عنوان پارامتر مورد پیش‌بینی (متغیر وابسته) مدل یادگیری ماشین عمل می‌کند، محاسبه‌ی این پارامتر با استفاده از داده‌های ایستگاه‌های زمینی، که از دقت و قابلیت اعتماد بالاتری برخوردارند، صورت پذیرفته است. همانطور که پیش‌تر نیز اشاره شد، این شاخص از تفاضل مقادیر دو پارامتر بارندگی و تبخیر و تعرق پتانسیل حاصل می‌شود. در این پژوهش از روش تورنث‌ویت [20] برای محاسبه‌ی تبخیر و تعرق پتانسیل استفاده شده است. این روش تبخیر و تعرق پتانسیل را بر اساس داده‌های میانگین ماهانه‌ی دما و با در نظر گرفتن رابطه‌ی بین دما و تبخیر و تعرق محاسبه می‌کند.

۳-۲- انتخاب پارامترهای مؤثر
انتخاب پارامترهای مؤثر یک گام مهم در ساخت یک مدل پیش‌بینی است چرا که حذف پارامترهای نامربوط یا زائد می‌تواند تفسیرپذیری و عمومیت‌پذیری مدل را بهبود بخشد. در این پژوهش از دو روش فیلتر و لاسو [21] برای تعیین پارامترهای مؤثر استفاده شد. شکل ۴ و شکل ۵ نشان‌دهنده‌ی ماتریس همبستگی پارامترهای انتخاب شده به ترتیب در خوشه‌های ۱ و ۲ با روش فیلتر می‌باشند. در روش فیلتر همبستگی خطی بین هر پارامتر و پارامتر هدف اندازه‌گیری شده و پارامترها بر اساس مقادیر همبستگی آن‌ها با هدف، امتیازدهی می‌شوند. بدین ترتیب با در نظر گرفتن حد آستانه-ی ۰/۴ برای ضریب همبستگی، پارامترهایی که همبستگی

کمتر از این مقدار با پارامتر هدف دارند، حذف می‌شوند. در ادامه برای جلوگیری از افزونگی، هر پارامتری که همبستگی بالای ۰/۷ با پارامترهای دیگر دارد نیز از لیست خارج می‌شود. اما در روش لاسو به هریک از پارامترها ضرایبی اختصاص داده می‌شود و بر اساس آن‌ها سعی به حذف پارامترهای کم اهمیت‌تر و انتخاب پارامترهای مهم‌تر می‌شود. این موضوع بدین صورت انجام می‌پذیرد که به کم اهمیت‌ترین پارامترها ضریب صفر داده شده و از مدل حذف می‌گردند و تنها پارامترهای با ضرایب غیر صفر باقی می‌مانند. لازم به ذکر است که درجه‌ی سخت‌گیری مدل در روش لاسو به کمک پارامتر تنظیم λ صورت می‌پذیرد. مقدار بالاتر λ منجر به منظم‌سازی قوی‌تر و در نتیجه نسبت دادن تعداد ضرایب بیشتری به عدد صفر می‌شود. این در حالی است که، مقدار کمتر λ میزان منظم شدن را کاهش داده و به ضرایب بیشتری اجازه می‌دهد که مقادیر غیر صفر داشته باشند. در این پژوهش به کمک روش جست‌وجوی شبکه‌ای بازه‌ی ۰/۱ تا ۱ به ۱۰۰ قسمت تقسیم شده و بهینه‌ترین حالت برای λ به دست آمده است. لازم به ذکر است که مقادیر کمتر از ۰/۱ برای λ منجر به انتخاب تعداد بسیار زیادی از پارامترها خواهد شد که به هیچ عنوان بهینه نیست. مقدار λ در این پژوهش برابر با ۰/۱ به دست آمده است. نتایج نهایی پارامترهای انتخاب شده در جدول ۴ قابل مشاهده است. لازم به ذکر است که در این جدول، Pre بیانگر پارامتر بارندگی و LST بیانگر پارامتر دمای سطح زمین می‌باشند. ضمناً پسوند‌های ۱ و ۳ به ترتیب بیانگر میانگین ۱ ماهه و میانگین متحرک ۳ ماهه‌ی آن پارامتر می‌باشند.

۳-۳- پرسپترون چندلایه

پرسپترون چند لایه (MLP) نوعی شبکه عصبی مصنوعی است که اولین بار توسط رازن‌بالت [22] معرفی گردید. MLP ها که از چندین لایه گره به هم پیوسته به نام نورون تشکیل شده‌اند، به نوعی شبکه عصبی پیشخور بحساب

می‌آیند، به این معنی که اطلاعات در یک جهت و ابتدا از لایه ورودی و سپس از طریق یک یا چند لایه پنهان به لایه خروجی جریان می‌یابند. به طور کلی شبکه‌های عصبی مصنوعی با پردازش داده‌های تجربی قانون نهفته بین متغیرهای یک مدل را شناسایی کرده و به نحوی مناسب و با سرعت بالا به ساختار شبکه انتقال می‌دهند [23]. در این شبکه‌ها هر نورون سیگنال‌های ورودی را دریافت کرده، یک تابع فعال‌سازی غیرخطی را به مجموع وزنی ورودی‌ها اعمال کرده و خروجی را به لایه‌ی بعدی ارسال می‌کند. ساختار کلی این شبکه‌ها در شکل ۶ مشاهده می‌شود. به طور کلی MLP ها قادر به یادگیری روابط پیچیده غیر خطی هستند و به طور گسترده برای کارهایی مانند طبقه بندی، رگرسیون و تشخیص الگو مورد استفاده قرار می‌گیرند. در این پژوهش تعداد لایه‌های پنهان برابر با ۱ و تعداد نورون‌های لایه‌ی پنهان برابر با ۱ می‌باشند. ضمناً از واحد خطی اصلاح شده به عنوان تابع فعال‌سازی استفاده شده است.

۳-۴- ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) که اولین بار توسط کورتس و واپنیک [24] مورد استفاده قرار گرفت، یک الگوریتم یادگیری ماشین قدرتمند و پرکاربرد است و زیرمجموعه‌ی یادگیری نظارت شده به حساب می‌آید. این الگوریتم ابتدا داده‌های ورودی و خروجی را دریافت کرده و سپس اگر داده‌ها به صورت خطی در فضای ویژگی اصلی مسئله قابل تفکیک نباشند، از یک تابع کرنل برای تبدیل داده‌ها به یک فضای ویژگی با ابعاد بالاتر استفاده می‌کند. هدف از این تبدیل یافتن یک رابطه خطی بین متغیرهای ورودی و مقادیر هدف است. به همین منظور در مدل‌سازی رگرسیونی، SVM به یافتن ابرصفحه‌ای می‌پردازد که بهترین تطابق با داده‌ها را داشته باشد و در عین حال خطاهای پیش‌بینی را به حداقل برساند. این ابر صفحه با تنظیم پارامترهای مدل برای به حداقل رساندن مجموع مقادیر خطا

بین مقادیر هدف پیش بینی شده و واقعی تعیین می‌شود. اصطلاح بردارهای پشتیبان، اشاره به نزدیک‌ترین نقاط داده به ابرصفحه را دارد که این نقاط بر موقعیت و جهت ابرصفحه تاثیر گذاشته و در تعیین پیش‌بینی‌های مدل نقش دارند. هنگامی که ابرصفحه‌ی بهینه تعیین شد، می‌توان از آن برای پیش‌بینی مقادیر هدف و نیز برای داده‌های جدید و نادیده استفاده کرد. مقدار هدف پیش‌بینی شده برای داده‌های جدید بر اساس موقعیت آن‌ها نسبت به ابرصفحه محاسبه خواهد شد. ساختار این الگوریتم در شکل ۷ نمایش داده شده است. در این پژوهش از تابع پایه شعاعی به عنوان تابع هسته استفاده شده و مقیاس آن $2/6$ در نظر گرفته شده است.

۳-۵- جنگل تصادفی

الگوریتم جنگل تصادفی (RF) که اولین بار توسط بریمن [25] مورد استفاده قرار گرفت، یک الگوریتم یادگیری ماشین قدرتمند است که می‌تواند برای مسئله‌های رگرسیونی مورد استفاده قرار گیرد. این الگوریتم یک روش یادگیری گروهی است که برای پیش‌بینی، چندین درخت تصمیم را ترکیب می‌کند. جنگل تصادفی پس از دریافت داده‌های ورودی و متغیر هدف، مجموعه‌ای از درختان تصمیم را می‌سازد که در آن هر درخت تصمیم با استفاده از یک زیرمجموعه تصادفی از مجموعه داده‌های مربوط به آموزش ساخته می‌شود. علاوه بر این، یک زیرمجموعه‌ی تصادفی از پارامترها در هر گره از درخت انتخاب می‌شوند تا بهترین تقسیم را انجام داده و تنوع را در بین درختان فراهم کند. سپس هر درخت تصمیم به صورت بازگشتی و با پارتیشن بندی داده‌ها در هر گره بر اساس یک معیار تقسیم انتخابی، که معمولاً کاهش واریانس یا میانگین مربعات خطاست، رشد می‌کند. این تقسیم تا زمانی که حداقل یکی از معیارهای توقف برآورده شود، ادامه می‌یابد. این معیارها می‌توانند شامل رسیدن به حداقل عمق یا حداقل تعداد نمونه‌های مورد نیاز برای یک گره برگ باشند. در ادامه، الگوریتم با ترکیب

پیش‌بینی‌های همه درخت‌های تصمیم، یک مجموعه ایجاد می‌کند. در رگرسیون، پیش‌بینی‌های فردی از هر درخت برای به دست آوردن پیش‌بینی نهایی میانگین‌گیری می‌شوند. این فرآیند میانگین‌گیری به کاهش تأثیر نقاط پرت و نویز در داده‌ها کمک می‌کند. نحوه‌ی عملکرد این الگوریتم در شکل ۸ قابل مشاهده است.

۳-۶- جنگل تصادفی تصحیح بایاس شده

همانطور که پیش‌تر نیز اشاره شد، در الگوریتم جنگل تصادفی هر درخت تصمیم در مجموعه‌ی درختان، پیش‌بینی‌های مستقلی انجام می‌دهد و پیش‌بینی نهایی با میانگین‌گیری پیش‌بینی‌های همه‌ی درختان به دست می‌آید. با این حال، با وجود اینکه این میانگین‌گیری به نحوی از تولید مقادیر پیش‌بینی شده‌ی غیرمنطقی جلوگیری می‌کند، ممکن است پیش‌بینی‌های این الگوریتم، به‌ویژه در شرایطی که به عنوان مثال داده‌های آموزشی ناهمگون هستند یا روابط غیرخطی در داده‌ها وجود دارد و یا زمانی که مجموعه‌ی مشاهدات بسیار بزرگ یا بسیار کوچک هستند، مقادیری بایاس از خود نشان دهند [26].

الگوریتم جنگل تصادفی تصحیح بایاس شده (BRF) به منظور در نظر گرفتن همین اثر، توسط ژنگ و لو [27] معرفی شده است. این الگوریتم ابتدا مطابق با رابطه‌ی ۲ یک مدل RF ساده را بر روی داده‌های آموزشی اعمال کرده و نتایج پیش‌بینی شده را به دست می‌آورد.

$$Y_{train} = RF(X_{train}) \quad \text{Error! No text of specified style in document.}$$

در این رابطه، X_{train} بیانگر متغیرهای مستقل داده‌های آموزش و Y_{train} بیانگر متغیر وابسته داده‌های آموزش می‌باشند.

در ادامه مقادیر باقی مانده با توجه به مقادیر پیش‌بینی شده‌ی مدل RF توسط رابطه‌ی ۳ محاسبه می‌گردد.

$$r_{train} = Y_{train} - \hat{Y}_{train} \quad \text{Error! No text of specified style in document.}$$

در این رابطه، $Y_{bias-corrected}$ بیانگر متغیر وابسته به صورت اصلاح بایاس شده، Y_{test} بیانگر متغیر وابسته داده‌های آزمون و r_{test} بیانگر باقی‌مانده‌ی داده‌های آزمون می‌باشند. در این پژوهش، تعداد درختان تصمیم در هر دو الگوریتم RF و BRF ۵۰۰ در نظر گرفته شده است.

۴- اعتبارسنجی مدل

هدف نهایی در مدل‌سازی با الگوریتم‌های یادگیری ماشین، توسعه مدلی است که بتواند پیش‌بینی‌های دقیقی بر روی داده‌های جدید و دیده نشده انجام دهد. لذا به هنگام فرآیند آموزش مدل، ارزیابی عملکرد آن مدل و نیز اطمینان حاصل کردن از توانایی تعمیم مدل به داده‌هایی فراتر از داده‌های آموزشی، امری ضروری به حساب می‌آید. برای این منظور از مفهوم اعتبارسنجی استفاده می‌شود. یکی از پرکاربردترین روش‌های اعتبارسنجی، روش اعتبارسنجی متقابل K-fold می‌باشد. در پژوهش حاضر، از این روش برای اعتبارسنجی تمامی الگوریتم‌های استفاده شده، بهره گرفته شده است. ضمناً در تمامی الگوریتم‌ها $K=5$ در نظر گرفته شده است.

۵- معیارهای ارزیابی عملکرد مدل

در هر فرآیند مدل‌سازی، عملکرد نهایی یک مدل برای پیش‌بینی متغیر هدف نیاز به ارزیابی دارد. این وظیفه بر عهده‌ی معیارهای ارزیابی است. این معیارها، روش‌های عددی هستند که بسته به نوع مسئله، میزان عملکرد مدل را از نظر دقت، یادآوری، خطا یا سایر معیارهای مرتبط ارزیابی کرده و امکان انتخاب بهترین و دقیق‌ترین مدل را فراهم می‌آورند. در این پژوهش از معیارهای ارزیابی MAE ، $RMSE$ ، MSE و R^2 برای ارزیابی عملکرد مدل استفاده شده است. روابط محاسبه‌ی این معیارها در ادامه قابل مشاهده است.

در این رابطه، r_{train} بیانگر باقی‌مانده‌ی داده‌های آموزش و \hat{Y}_{train} بیانگر مقادیر پیش‌بینی شده‌ی متغیر وابسته توسط RF می‌باشند.

در ادامه مطابق با رابطه‌ی ۴، یک مدل RF دیگر به باقی‌مانده‌ها به عنوان متغیر وابسته و کل داده‌های آموزشی به عنوان متغیرهای مستقل برازش داده می‌شود.

$$r_{train} = RF_{res}(X_{train}, Y_{train}) \quad (۴)$$

سپس مقادیر پیش‌بینی شده‌ی متغیر وابسته داده‌های آزمون توسط همان مدل RF استفاده شده در رابطه‌ی ۲، مطابق رابطه‌ی ۵ تولید می‌شوند.

$$Y_{test} = RF(X_{test}) \quad (۵)$$

در این رابطه، X_{test} بیانگر متغیرهای مستقل داده‌های آزمون و Y_{test} بیانگر متغیر وابسته داده‌های آزمون می‌باشند.

سپس درست به مانند گام سوم، مقادیر باقی‌مانده‌ی داده‌های آزمون توسط همان مدل RF موجود در گام سوم، و مطابق رابطه‌ی ۶ محاسبه می‌شوند.

$$r_{test} = RF_{res}(X_{test}, Y_{test}) \quad (۶)$$

در این رابطه، r_{test} بیانگر باقی‌مانده‌ی داده‌های آزمون می‌باشد.

در انتها مقادیر متغیر وابسته به صورت اصلاح بایاس شده، از طریق رابطه‌ی ۷ قابل محاسبه خواهند بود.

$$Y_{bias-corrected} = Y_{test} + r_{test} \quad (۷)$$

پارامترهای ورودی با پارامتر هدف به طور قابل توجهی بالاست. مثلاً برای خوشه‌ی شماره ۱ پارامتر میانگین متحرک سه ماهه‌ی دمای سطح زمین در شب (LST_night-3) و برای خوشه‌ی شماره ۲ پارامتر میانگین متحرک سه ماهه‌ی میانگین دمای سطح زمین در روز و شب (LST_mean-3) همبستگی بالای ۹۰ درصد با شاخص SPEI به عنوان پارامتر هدف از خود نشان داده‌اند.

با این حال، می‌توان از نتایج به دست آمده این‌گونه استنباط کرد که برای ایستگاه‌های متعلق به خوشه‌ی ۱، الگوریتم جنگل تصادفی تصحیح بایاس شده با ثبت بالاترین مقدار در معیار ضریب تعیین ($R^2 = 0.9858$) و پایین‌ترین مقدار در معیار ریشه‌ی میانگین مربعات خطا ($RMSE = 0.1190$) بهترین عملکرد را برای پیش‌بینی شاخص SPEI و در نتیجه پیش‌بینی خشکسالی کشاورزی داراست. پس از الگوریتم جنگل تصادفی تصحیح بایاس شده، به ترتیب الگوریتم‌های جنگل تصادفی، پرسپترون چندلایه و ماشین بردار پشتیبان بهترین عملکرد را از خود نشان دادند. جزئیات عملکرد و نتایج این الگوریتم‌ها در جدول ۵ قابل مشاهده است.

همچنین به مانند خوشه‌ی ۱، برای ایستگاه‌های متعلق به خوشه‌ی ۲ نیز الگوریتم جنگل تصادفی تصحیح بایاس شده با ثبت بالاترین مقدار در معیار ضریب تعیین ($R^2 = 0.9809$) و پایین‌ترین مقدار در معیار ریشه‌ی میانگین مربعات خطا ($RMSE = 0.1375$) بهترین عملکرد را برای پیش‌بینی شاخص SPEI و در نتیجه پیش‌بینی خشکسالی کشاورزی از خود نشان داد. در این خوشه پس از الگوریتم جنگل تصادفی تصحیح بایاس شده، به ترتیب الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و پرسپترون چندلایه بهترین عملکرد را از خود به جای گذاشتند. جزئیات عملکرد و نتایج این الگوریتم‌ها در جدول ۶ قابل مشاهده است.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

در این روابط n تعداد داده‌ها، y_i مقادیر واقعی، \hat{y}_i مقادیر پیش‌بینی شده و \bar{y} میانگین مقادیر واقعی می‌باشند.

نتایج و بحث

۱- ارزیابی عملکرد مدل‌ها

پس از اعمال پردازش‌های مربوطه بر روی داده‌ها و نیز تنظیم فرآیندهای الگوریتم‌های یادگیری ماشین، فرآیند مدل‌سازی با استفاده از چهار الگوریتم یاد شده و با توجه به نتایج انتخاب پارامترهای موثر صورت گرفت. نتایج این مدل‌سازی برای خوشه‌ی ۱ در جدول ۵ و برای خوشه‌ی ۲ در جدول ۶ قابل مشاهده است. لازم به ذکر است که فرآیند مدل‌سازی برای هر یک از ترکیب‌های پارامترهای ورودی به تعداد ۱۰ مرتبه تکرار شد و میانگین نتایج برای هر مدل در جدول‌های مربوطه ثبت گردید.

همان‌طور که از جدول ۵ و جدول ۶ پیداست، عملکرد تمامی مدل‌ها به طور قابل توجهی بالا می‌باشد. این بدین معنی است که متغیرهای مستقل (پارامترهای ورودی) به شدت با متغیر وابسته (پارامتر هدف) مرتبط بوده و می‌توانند به طور مؤثر مقادیر آن را پیش‌بینی کنند. به عنوان مثال، همان‌طور که از ماتریس‌های همبستگی روش فیلتر در شکل-های ۴ و ۵ پیداست، ارتباط و وابستگی میان برخی از

نشان‌دهنده‌ی گراف مقایسه‌ی نتایج مدل‌ها بر اساس معیار میانگین خطای مطلق می‌باشد.

۲- تهیه نقشه‌های پهنه‌بندی خشکسالی کشاورزی

پس از ایجاد مدل‌های پیش‌بینی خشکسالی و انتخاب مدل با بهترین عملکرد، می‌توان از نتایج این مدل برای تهیه‌ی نقشه‌های پهنه‌بندی خشکسالی استفاده کرد. این نقشه‌ها ابزاری هستند که برای شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی مورد استفاده قرار می‌گیرند. برای تهیه‌ی این نقشه‌ها، از مقادیر پیش‌بینی شده‌ی مدل ایجاد شده با الگوریتم جنگل تصادفی تصحیح بایاس شده برای شاخص SPEI، استفاده شده و این مقادیر در ۴ ماه منتخب از سال که هرکدام نمونه‌ای برای یکی از چهار فصل سال می‌باشند، در تمامی ایستگاه‌های مورد مطالعه محاسبه شده و پس از آن بر روی سطح کل کشور درونیایی می‌شوند.

همان‌طور که پیش‌تر نیز مطرح شد، عموماً خشکسالی به حالتی اطلاق می‌شود که شاخص SPEI مقادیری کمتر از ۱- داشته باشد. این در حالی است که مقادیر بزرگتر از یک برای این شاخص نشان‌دهنده‌ی شرایط مرطوب، مقادیر بین صفر و منفی یک نشان‌دهنده‌ی شرایط نسبتاً خشک و مقادیر بین صفر و یک نشان‌دهنده‌ی شرایط نسبتاً مرطوب می‌باشند. به این ترتیب و با توجه به شکل ۱۷ که نشان‌دهنده‌ی نقشه‌ی پهنه‌بندی خشکسالی کشاورزی در ایران برای فروردین ۱۴۰۰ می‌باشد، مشاهده می‌شود که مدل پیش‌بینی خشکسالی برای اغلب مناطق کشور حالت نسبتاً خشک را پیش‌بینی کرده و تنها بخش‌هایی از استان‌های گلستان و مازندران را به حالت نسبتاً مرطوب نسبت داده است. در ادامه همان‌طور که از شکل ۱۸ پیداست، مدل برای تمام مناطق کشور در تیر ماه ۱۴۰۰ شرایط خشکسالی را پیش‌بینی کرده است. بررسی دقیق‌تر این نقشه نشان می‌دهد که قسمت‌هایی از استان‌های مازندران، گیلان، اردبیل و آذربایجان شرقی، که همگی از استان‌های پر بارش کشور به حساب می‌آیند، تحت شرایط

همان‌طور که از نتایج نیز پیداست، به‌طور کلی می‌توان عملکرد الگوریتم جنگل تصادفی را در مقایسه با ماشین بردار پشتیبان و پرسپترون چندلایه بهتر دانست. این امر به این دلیل می‌باشد که جنگل تصادفی دارای چندین مزیت بوده که آن را در چنین سبک مسائل به انتخابی بهتر تبدیل می‌کند. به‌عنوان مثال این الگوریتم به دلیل استفاده از چندین درخت تصمیم‌گیری و بوت‌استرپ معمولاً نسبت به دو الگوریتم دیگر مقاومت بیشتری در برابر بیش‌برازش دارد. همچنین می‌توان به تفسیرپذیری بهتر این الگوریتم نسبت به دو الگوریتم دیگر اشاره کرد. با این حال، جنگل تصادفی معمولاً در بازتولید مشاهدات شدید عملکرد ضعیفی دارد. به طوری که گاهی مقادیر بزرگ کمتر از حد و مقادیر کوچک بیشتر از حد برآورد می‌شوند [25]. لذا از آنجایی که خشکسالی معمولاً به‌عنوان شرایط خشک بیش از اندازه و غیرعادی تلقی می‌شود، استفاده از جنگل تصادفی ممکن است منجر به ایجاد بایاس در فرآیند پایش خشکسالی گردد. به همین جهت، ترکیب روش تصحیح بایاس با الگوریتم جنگل تصادفی می‌تواند به حذف بایاس ایجاد شده و حصول نتایج بهتر در مدل‌سازی منجر گردد. دستیابی به مقادیر ضریب تعیین بالاتر و ریشه‌ی میانگین مربعات خطای پایین‌تر برای روش جنگل تصادفی تصحیح بایاس شده نسبت به جنگل تصادفی ساده، مهر تاییدی بر این ادعا در این پژوهش می‌باشد.

در ادامه شکل ۹ و شکل ۱۰ نشان‌دهنده‌ی مقایسه‌ی بین مقادیر پیش‌بینی شده و مقادیر واقعی SPEI در خوشه‌ی ۱ و شکل ۱۱ و شکل ۱۲ بیانگر مقایسه‌ی بین مقادیر پیش‌بینی شده و مقادیر واقعی SPEI در خوشه‌ی ۲ می‌باشند. همچنین شکل ۱۳ نشان‌دهنده‌ی گراف مقایسه‌ی نتایج مدل‌ها بر اساس معیار ضریب تعیین، شکل ۱۴ نشان‌دهنده‌ی گراف مقایسه‌ی نتایج مدل‌ها بر اساس معیار ریشه‌ی میانگین مربعات خطا، شکل ۱۵ نشان‌دهنده‌ی گراف مقایسه‌ی نتایج مدل‌ها بر اساس معیار میانگین مربعات خطا و شکل ۱۶

میلی‌متر می‌باشد، نیز الگوریتم جنگل تصادفی تصحیح بایاس شده با دقت ۹۸/۰۹ درصد بهترین عملکرد را نشان داد. پس از آن الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و پرسپترون چندلایه به ترتیب در رده‌های بعدی قرار گرفتند. (۳) بررسی‌ها نشان داد که مدل پیش‌بینی خشکسالی که با الگوریتم جنگل تصادفی تصحیح بایاس شده ایجاد شده است، قابلیت پیش‌بینی بسیار دقیقی هم در اقلیم پر باران و هم در اقلیم خشک و نیمه‌خشک دارد. (۴) از نتایج به دست آمده در این مدل‌سازی اقدام به تهیه نقشه‌های پهنه‌بندی خشکسالی شد. این نقشه‌ها که ابزاری مناسب برای شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی می‌باشند، می‌تواند اطلاعات ارزشمندی را در رابطه با توزیع و شدت خشکسالی در مناطق مختلف در اختیار ذینفعان، از جمله سیاست‌گذاران، کشاورزان و برنامه‌ریزان منابع آب قرار می‌دهد.

واژه‌نامه

PDSI	شاخص شدت خشکسالی پالمر
SPI	شاخص بارندگی استاندارد شده
SPEI	شاخص بارش-تبخیر و تعرق استاندارد شده
NDVI	شاخص تفاوت پوشش گیاهی نرمال شده
NDDI	شاخص تفاوت خشکسالی نرمال شده
NMDI	شاخص خشکسالی نرمال شده‌ی چندباندی
MLP	پرسپترون چندلایه
SVM	ماشین بردار پشتیبان
RF	جنگل تصادفی
BRF	جنگل تصادفی تصحیح بایاس شده
MSE	میانگین مربعات خطا
RMSE	ریشه‌ی میانگین مربعات خطا
MAE	میانگین خطای مطلق
R ²	ضرب تعیین

خشکسالی شدیدتری نسبت به سایر مناطق کشور قرار دارند. این بدین خاطر است که به علت برقراری شرایط مرطوب در سایر اوقات سال در این مناطق، کاهش نسبی میزان بارندگی و گرمای هوا در تابستان منجر به افت شدیدتر مقادیر شاخص SPEI نسبت به سایر مناطق در این فصل می‌گردد. شکل ۱۹ نیز به مانند شکل ۱۷ نشان‌دهنده‌ی حالت نسبتاً خشک برای اغلب مناطق کشور و حالت نسبتاً مرطوب برای تنها بخش‌هایی از استان‌های مازندران و گیلان در مهر ۱۴۰۰ می‌باشد. در نهایت، شکل ۲۰ نشان از این دارد که مدل حالت کاملاً مرطوب را برای تمام نقاط کشور در دی ۱۴۰۰ پیش‌بینی کرده است. به مانند تفسیر شکل ۱۸، علت برقراری شرایط مرطوب‌تر برای استان‌های جنوبی کشور که غالباً آب و هوایی گرم و خشک دارند، شرایط بارش و دمایی بسیار متفاوت برای این استان‌ها در این بازه از سال نسبت به سایر اوقات سال می‌باشد که منجر به افزایش نسبی مقادیر شاخص SPEI برای این مناطق در این ماه شده است.

نتیجه‌گیری

در پژوهش حاضر سعی بر ایجاد راهکاری مناسب برای پیش‌بینی خشکسالی به عمل آمده است. در این مطالعه از چهار الگوریتم یادگیری ماشین به نام‌های جنگل تصادفی تصحیح بایاس شده، جنگل تصادفی، ماشین بردار پشتیبان و پرسپترون چندلایه به منظور ایجاد مدل پیش‌بینی خشکسالی استفاده گردید که نتایج آن به شرح زیر به دست آمد: (۱) برای ایستگاه‌های متعلق به خوشه‌ی ۱، که میانگین بارش سالانه‌ای در حدود ۹۵۰ میلی‌متر دارند، الگوریتم جنگل تصادفی تصحیح بایاس شده با دقت ۹۸/۵۸ درصد بهترین عملکرد را نشان داد. پس از آن الگوریتم‌های جنگل تصادفی، پرسپترون چندلایه و ماشین بردار پشتیبان به ترتیب و با فاصله‌ی اندک نسبت به دقت الگوریتم جنگل تصادفی تصحیح بایاس شده، در رده‌های بعدی قرار گرفتند. (۲) برای ایستگاه‌های متعلق به خوشه‌ی ۲، که میانگین بارش سالانه‌ی آن‌ها کمتر از ۳۰۰

- [1] D. Dutta, A. Kundu, and N. R. Patel, "Predicting agricultural drought in eastern Rajasthan of India using NDVI and standardized precipitation index," *Geocarto Int.*, vol. 28, no. 3, pp. 192–209, Jun. 2013, doi: 10.1080/10106049.2012.679975.
- [2] N. M. Alam *et al.*, "Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India," *Phys. Chem. Earth*, vol. 100, pp. 31–43, 2017, doi: 10.1016/j.pce.2017.02.008.
- [3] W. C. Palmer, *Meteorological drought*, vol. 30. US Department of Commerce, Weather Bureau, 1965.
- [4] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Eighth Conference on Applied Climatology*, 1993, vol. 17, no. January, pp. 17–22, [Online]. Available: <https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf> https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf.
- [5] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index," *J. Clim.*, vol. 23, no. 7, pp. 1696–1718, 2010, doi: <https://doi.org/10.1175/2009JCLI2909.1>.
- [6] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA Spec. Publ.*, vol. 351, no. 1, p. 309, 1974.
- [7] Y. Gu, J. F. Brown, J. P. Verdin, and B. Wardlow, "A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States," *Geophys. Res. Lett.*, vol. 34, no. 6, Mar. 2007, doi: <https://doi.org/10.1029/2006GL029127>.
- [8] L. Wang and J. J. Qu, "NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing," vol. 34, no. July, pp. 1–5, 2007, doi: 10.1029/2007GL031021.
- [9] J. Rhee, J. Im, and G. J. Carbone, "Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2875–2887, 2010, doi: <https://doi.org/10.1016/j.rse.2010.07.005>.
- [10] Z. M. Yaseen, M. Ali, A. Sharafati, N. Al-Ansari, and S. Shahid, "Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh," *Sci. Rep.*, vol. 11, no. 1, p. 3435, 2021, doi: 10.1038/s41598-021-82977-9.
- [11] M. R. Kousari, M. E. Hosseini, H. Ahani, and H. Hakimelahi, "Introducing an operational method to forecast long-term regional drought based on the application of artificial intelligence capabilities," *Theoretical and Applied Climatology*, vol. 127, no. 1–2, pp. 361–380, 2017, doi: 10.1007/s00704-015-1624-6.
- [12] M. Mokhtarzad, F. Eskandari, N. Jamshidi Vanjani, and A. Arabasadi, "Drought forecasting by ANN, ANFIS, and SVM and comparison of the models," *Environ. Earth Sci.*, vol. 76, no. 21, p. 729, 2017, doi: 10.1007/s12665-017-7064-0.
- [13] A. Abbasi, K. Khalil, J. Behmanesh, and A. Shirzad, "Application of Support Vector Machine and Bayesian Network method in agricultural drought prediction," *Sci. Res. J. Watershed Manag. Eng.*, vol. 12, no. 1, pp. 107–124, 2020, doi: 10.22092/ijwmse.2019.122619.1519.
- [14] F. Tufaner and A. Özbeyaz, "Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms," *Environ. Monit. Assess.*, vol. 192, no. 9, 2020, doi: 10.1007/s10661-020-08539-0.
- [15] P. Feng, B. Wang, D. L. Liu, and Q. Yu, "Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia," *Agricultural Systems*, vol. 173, pp. 303–316,

- 2019, doi: 10.1016/j.agry.2019.03.015.
- [16] M. Zoljoodi and A. Didevarasl, "Evaluation of Spatial-Temporal Variability of Drought Events in Iran Using Palmer Drought Severity Index and Its Principal Factors (through 1951-2005)," *Atmos. Clim. Sci.*, vol. 03, no. 02, pp. 193–207, Jan. 2013, doi: 10.4236/acs.2013.32021.
- [17] A. Darvishi, S. Arkhi, and A. Ebrahimi, "Risk and disaster management to mitigate the effects of droughts in Iran," in *Proceeding of the Conference on Drought in Charmahal-Bakhtiari, Shahrekord, The University of Shahrekord*, 2008, pp. 22–23.
- [18] Iran's Statistical Centre, "Shares and rankings of provinces based on important variables in the agricultural sector," 2023, [Online]. Available: https://www.amar.org.ir/LinkClick.aspx?fileticket=rzp_8zPLtu8%3D&portalid=0.
- [19] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *J. R. Stat. Soc. Ser. C (Applied Stat.)*, vol. 28, no. 1, pp. 100–108, Jan. 1979, doi: 10.2307/2346830.
- [20] C. W. Thornthwaite, "An Approach toward a Rational Classification of Climate," *Geogr. Rev.*, vol. 38, no. 1, pp. 55–94, Jan. 1948, doi: 10.2307/210739.
- [21] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, Jan. 1996, [Online]. Available: <http://www.jstor.org/stable/2346178>.
- [22] F. Rosenblatt, *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, vol. 55. Spartan books Washington, DC, 1962.
- [23] A. Aghasi, M. Torbatinejad, and A. Bagherzadeh, "Modelling of rutting behavior of modified polymer asphalt mixture using artificial neural network," *J. Ferdowsi Civ. Eng.*, vol. 35, no. 3, pp. 69–88, 2022, doi: 10.22067/jfeci.2022.76444.1135.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
- [25] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [26] J. Song, "Bias corrections for Random Forest in regression using residual rotation," *J. Korean Stat. Soc.*, vol. 44, no. 2, pp. 321–326, 2015, doi: <https://doi.org/10.1016/j.jkss.2015.01.003>.
- [27] G. Zhang and Y. Lu, "Bias-corrected random forests in regression," *J. Appl. Stat.*, vol. 39, no. 1, pp. 151–160, Jan. 2012, doi: 10.1080/02664763.2011.578621.
- [28] Iran Meteorological Organization, "Weather data acquisition system," [Online]. Available: <https://data.irimo.ir/>.
- [29] G. J. Huffman E.F. Stocker, D.T. Bolvin, E.J. Nelkin, Jackson Tan, "GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06." NASA Goddard Earth Sciences Data and Information Services Center, Jan. 07, 2019, doi: 10.5067/GPM/IMERG/3B-HH/06.
- [30] Z. Wan, S. Hook, and G. Hulley, "MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061." NASA EOSDIS Land Processes Distributed Active Archive Center, Jan. 07, 2021, doi: 10.5067/MODIS/MOD11A1.061.
- [31] E. Vermote and R. Wolfe, "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061." NASA EOSDIS Land Processes Distributed Active Archive Center, Jan. 07, 2021, doi: 10.5067/MODIS/MOD09GA.061.
- [32] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, "Investigation and improvement of multi-layer perception neural networks for credit scoring," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3508–3516, 2015, doi: 10.1016/j.eswa.2014.12.006.
- [33] A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha, "Chapter 6 - Machine learning for soil moisture assessment," in *Cognitive Data Science in Sustainable Computing*, R. C. Poonia, V. Singh, and S. R. B. T.-D. L. for S. A. Nayak, Eds.

- [34] A. Afzal *et al.*, “Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows,” *Aerosp. Sci. Technol.*, vol. 107, p. 106318, 2020, doi: <https://doi.org/10.1016/j.ast.2020.106318>.

جداول

جدول ۱- جزئیات دسته‌بندی استان‌های کشور بر اساس اهمیت کشاورزی

نام دسته	استان	سهم ایستگاه‌های استان از مجموع ایستگاه‌های مورد مطالعه (درصد)
الف (سهم بیش از ۵ درصد از کل اراضی زراعی کشور)	خراسان رضوی	۵/۷۱٪
	خوزستان	۵/۷۱٪
	آذربایجان شرقی	۵/۷۱٪
	کردستان	۵/۷۱٪
	فارس	۵/۷۱٪
	اردبیل	۵/۷۱٪
	همدان	۴/۲۹٪
	زنجان	۴/۲۹٪
ب (سهم بین ۳ تا ۵ درصد از کل اراضی زراعی کشور)	لرستان	۴/۲۹٪
	آذربایجان غربی	۴/۲۹٪
	کرمانشاه	۴/۲۹٪
	مرکزی	۴/۲۹٪
	گلستان	۴/۲۹٪
	قزوین	۲/۸۶٪
	اصفهان	۲/۸۶٪
	بوشهر	۲/۸۶٪
پ (سهم بین ۱ تا ۳ درصد از کل اراضی زراعی کشور)	ایلام	۲/۸۶٪
	مازندران	۲/۸۶٪
	کهگیلویه و بویراحمد	۲/۸۶٪
	چهارمحال و بختیاری	۲/۸۶٪
	خراسان شمالی	۱/۴۳٪
	کرمان	۱/۴۳٪
	سیستان و بلوچستان	۱/۴۳٪
	گیلان	۱/۴۳٪
ت (سهم کمتر از ۱ درصد از کل اراضی زراعی کشور)	خراسان جنوبی	۱/۴۳٪
	سمنان	۱/۴۳٪
	تهران	۱/۴۳٪
	هرمزگان	۱/۴۳٪
	البرز	۱/۴۳٪
	یزد	۱/۴۳٪
	قم	۱/۴۳٪

جدول ۲- مشخصات کلی ایستگاه‌های سینوپتیک مورد استفاده در این پژوهش

استان	نام ایستگاه	طول جغرافیایی	عرض جغرافیایی	میانگین بارندگی سالانه (میلی‌متر)	میانگین دمای سالانه (درجه‌ی سانتی‌گراد)
خراسان رضوی	مشهد	۵۹/۶	۳۶/۲	۲۲۷	۱۵/۹
	سبزوار	۵۷/۶	۳۶/۲	۱۷۶	۱۸/۲
	قوچان	۵۸/۴	۳۷/۱	۳۲۴	۱۲/۵
	تربت حیدریه	۵۹/۲	۳۵/۳	۲۲۱	۱۴/۷
خوزستان	اهواز	۴۸/۷	۳۱/۳	۱۹۳	۲۶/۷
	آبادان	۴۸/۲	۳۰/۴	۱۳۴	۲۶/۶
	ایذه	۴۹/۹	۳۱/۸	۵۷۰	۲۳/۸
	صفی‌آباد (دزفول)	۴۸/۴	۳۲/۳	۲۸۶	۲۴/۴
آذربایجان شرقی	تبریز	۴۶/۲	۳۸/۱	۲۵۷	۱۳/۴
	جلفا	۴۵/۶	۳۸/۹	۲۲۴	۱۵/۵
	میانه	۴۷/۷	۳۷/۴	۲۶۲	۱۴/۴
	مراغه	۴۶/۱	۳۷/۳	۲۵۲	۱۳/۷
کردستان	سنندج	۴۷/۰	۳۵/۳	۳۵۰	۱۴/۶
	بانه	۴۵/۹	۳۶/۰	۶۹۰	۱۴/۳
	بیجار	۳۷/۶	۳۵/۹	۳۲۰	۱۲/۰
	مریوان	۴۶/۲	۳۵/۵	۸۶۱	۱۳/۶
فارس	شیراز	۵۲/۶	۲۹/۵	۲۸۴	۱۹/۱
	لار	۵۴/۴	۲۷/۷	۱۸۸	۲۴/۸
	آباده	۵۲/۶	۳۱/۲	۱۳۷	۱۴/۹
	فسا	۵۳/۷	۲۸/۹	۲۵۶	۱۹/۹
اردبیل	اردبیل	۴۸/۳	۳۸/۲	۲۵۹	۱۰/۱
	مشکین‌شهر	۴۷/۵	۳۸/۴	۳۷۶	۱۱/۲
	خلخال	۴۸/۵	۳۷/۶	۳۷۱	۹/۲
	پارس‌آباد	۴۷/۹	۳۹/۶	۲۷۱	۱۵/۱
همدان	فرودگاه همدان	۴۸/۵	۳۴/۹	۳۰۲	۱۲/۴
	نهایند	۴۸/۴	۳۴/۲	۳۶۵	۱۴/۹
	ملایر	۴۸/۸	۳۴/۳	۳۴۰	۱۳/۹
زنجان	زنجان	۴۸/۵	۳۶/۷	۲۸۶	۱۱/۹
	خرم‌دره	۴۹/۲	۳۶/۲	۲۹۷	۱۲/۷
	خدابنده	۴۸/۶	۳۶/۱	۳۸۴	۱۱/۵
لرستان	خرم‌آباد	۴۸/۳	۳۳/۴	۴۶۱	۱۷/۵
	بروجرد	۴۸/۸	۳۳/۹	۴۶۲	۱۵/۳
	الیگودرز	۴۹/۷	۳۳/۴	۳۹۱	۱۳/۲
آذربایجان غربی	ارومیه	۴۵/۱	۳۷/۷	۳۰۳	۱۱/۹
	مهاباد	۴۵/۷	۳۶/۸	۳۷۸	۱۳/۸
	ماکو	۴۴/۴	۳۹/۴	۳۲۲	۱۱/۲

۱۶/۱	۳۸۹	۳۴/۴	۴۷/۲	کرمانشاه	کرمانشاه
۲۰/۶	۳۸۷	۳۴/۵	۴۵/۹	سرپل ذهاب	
۱۳/۶	۳۷۲	۳۴/۵	۴۸/۰	کنگاور	
۱۴/۶	۳۰۱	۳۴/۱	۴۹/۸	اراک	مرکزی
۱۹/۹	۱۸۴	۳۵/۱	۵۰/۴	ساوه	
۱۵/۶	۲۹۲	۳۳/۶	۵۰/۱	خمین	
۱۷/۷	۴۹۰	۳۶/۹	۵۴/۴	گرگان	گلستان
۱۹/۳	۴۳۰	۳۷/۳	۵۵/۲	گنبد کاووس	
۱۸/۰	۳۶۳	۳۷/۸	۵۵/۹	مراوه تپه	
۱۴/۴	۳۰۱	۳۶/۲	۵۰/۰	قزوین	قزوین
۱۱/۹	۳۷۳	۳۵/۶	۴۹/۲	آوج	
۱۷/۵	۱۳۱	۳۲/۵	۵۱/۷	اصفهان	اصفهان
۱۹/۸	۱۲۳	۳۴/۰	۵۱/۵	کاشان	
۲۵/۷	۲۲۴	۲۹/۰	۵۰/۸	فرودگاه بوشهر	بوشهر
۲۶/۱	۳۱۳	۲۷/۸	۵۲/۴	جم	
۱۷/۰	۵۰۹	۳۳/۶	۴۶/۴	ایلام	ایلام
۲۶/۹	۲۶۱	۳۲/۷	۴۷/۳	دهلران	مازندران
۱۷/۸	۷۴۷	۳۶/۵	۵۳/۰	ساری	
۱۶/۸	۱۳۱۸	۳۶/۷	۵۱/۵	نوشهر	
۱۵/۱	۷۷۱	۳۰/۷	۵۱/۶	یاسوج	کهگیلویه و بویراحمد
۲۳/۴	۴۰۹	۳۰/۳	۵۰/۸	دوگنبدان	
۱۱/۸	۳۰۹	۳۲/۳	۵۰/۸	شهرکرد	چهارمحال و بختیاری
۱۶/۰	۵۴۸	۳۱/۵	۵۰/۸	لردگان	
۱۳/۵	۲۴۳	۳۷/۵	۵۷/۳	بجنورد	خراسان شمالی
۱۷/۴	۱۱۱	۳۰/۳	۵۷/۰	کرمان	کرمان
۲۶/۷	۱۲۰	۲۵/۳	۶۰/۷	چابهار	سیستان و بلوچستان
۱۶/۵	۱۲۸۳	۳۷/۳	۴۹/۶	رشت	گیلان
۱۷/۰	۱۳۲	۳۲/۹	۵۹/۳	بیرجند	خراسان جنوبی
۱۸/۸	۱۳۰	۳۵/۶	۵۳/۴	سمنان	سمنان
۱۸/۶	۲۳۵	۳۵/۷	۵۱/۳	فرودگاه مهرآباد	تهران
۲۷/۳	۱۳۱	۲۷/۲	۵۶/۴	بندرعباس	هرمزگان
۱۶/۱	۲۶۱	۳۵/۸	۵۰/۹	کرج	البرز
۲۰/۷	۵۱	۳۱/۹	۵۴/۳	یزد	یزد
۱۹/۰	۱۴۱	۳۴/۸	۵۰/۹	قم	قم

جدول ۳- خلاصه‌ی داده‌ها و شاخص‌های مورد استفاده

داده‌ها	منبع دسترسی	نام محصول	تفکیک زمانی	تفکیک مکانی
دمای هوا	سازمان هواشناسی کشور [۲۸]	میانگین دمای ماهانه	ماهانه	ایستگاه
بارندگی	سازمان هواشناسی کشور [۲۸]	مجموع بارش ماهانه	ماهانه	ایستگاه
	GPM [29]	GPMv6	نیم ساعته	۱۱۱۳۲ متر
دمای سطح زمین	MODIS [30]	MOD11A1	روزانه	۱۰۰۰ متر

۵۰۰ متر	روزانه	MOD09GA	MODIS [31]	NDVI
۵۰۰ متر	روزانه	MOD09GA	MODIS [31]	EVI
۵۰۰ متر	روزانه	MOD09GA	MODIS [31]	NMDI
۵۰۰ متر	روزانه	MOD09GA	MODIS [31]	NDWI
۵۰۰ متر	روزانه	MOD09GA	MODIS [31]	NDDI

جدول ۴- نتایج انتخاب پارامترهای مؤثر

نام روش	خوشه	تعداد پارامترهای انتخاب شده	پارامترهای انتخاب شده
فیلتر	شماره ۱	۷	LST_night-3, NDDI6-1, NDVI-3, NDDI5-1, NDWI5-3, Pre-1, Pre-3
	شماره ۲	۶	LST_mean-3, NDDI7-1, NDDI6-3, Pre-1, NMDI-1, Pre-3
لاسو	شماره ۱	۶	Pre-3, LST_day-1, LST_mean-3, LST_night-3, NDDI6-1, NDDI7-3
	شماره ۲	۷	Pre-3, LST_day-1, LST_day-3, LST_mean-3, LST_night-3, NDDI7-3, NDVI-3

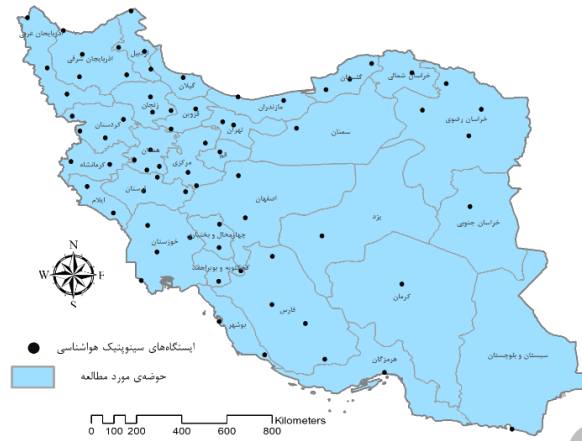
جدول ۵- نتایج مدل سازی پیش بینی خشکسالی در خوشه ۱ به کمک الگوریتم های یادگیری ماشین

نام الگوریتم یادگیری ماشین	روش انتخاب پارامتر مؤثر	R ²	RMSE	MSE	MAE	زمان آموزش مدل (ثانیه)
پرسپترون چندلایه	فیلتر	۰/۹۸۳۴	۰/۱۲۸۶	۰/۰۱۶۵	۰/۰۸۹۴	۴۰
	لاسو	۰/۹۸۵۲	۰/۱۲۱۰	۰/۰۱۴۷	۰/۰۸۵۳	۳۸
ماشین بردار پشتیبان	فیلتر	۰/۹۸۲۲	۰/۱۳۳۰	۰/۰۱۷۷	۰/۰۹۱۵	۱۹
	لاسو	۰/۹۸۴۸	۰/۱۲۲۹	۰/۰۱۵۱	۰/۰۸۳۹	۱۲
جنگل تصادفی	فیلتر	۰/۹۸۳۲	۰/۱۳۰۲	۰/۰۱۷۰	۰/۰۹۱۱	۲۰۴
	لاسو	۰/۹۸۵۸	۰/۱۱۹۱	۰/۰۱۴۲	۰/۰۸۲۰	۲۰۵
جنگل تصادفی تصحیح بایاس شده	فیلتر	۰/۹۸۴۱	۰/۱۲۶۳	۰/۰۱۶۰	۰/۰۸۷۹	۲۱۴
	لاسو	۰/۹۸۵۸	۰/۱۱۹۰	۰/۰۱۴۲	۰/۰۸۳۱	۲۱۸

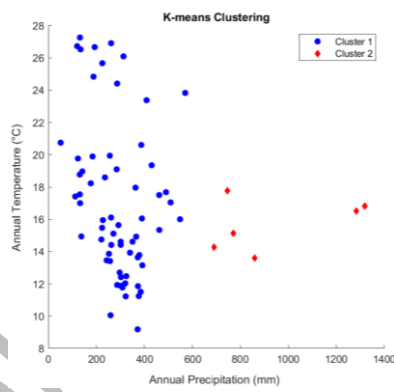
جدول ۶- نتایج مدل سازی پیش بینی خشکسالی در خوشه ۲ به کمک الگوریتم های یادگیری ماشین

نام الگوریتم یادگیری ماشین	روش انتخاب پارامتر مؤثر	R ²	RMSE	MSE	MAE	زمان آموزش مدل (ثانیه)
پرسپترون چندلایه	فیلتر	۰/۹۷۷۰	۰/۱۵۱۲	۰/۰۲۲۹	۰/۱۱۴۱	۵
	لاسو	۰/۹۷۸۶	۰/۱۴۵۷	۰/۰۲۱۳	۰/۱۰۸۸	۶
ماشین بردار پشتیبان	فیلتر	۰/۹۷۶۳	۰/۱۵۳۴	۰/۰۲۳۵	۰/۱۱۴۴	۲
	لاسو	۰/۹۸۰۳	۰/۱۴۰۱	۰/۰۱۹۷	۰/۱۰۴۵	۲
جنگل تصادفی	فیلتر	۰/۹۷۶۱	۰/۱۵۷۸	۰/۰۲۵۰	۰/۱۱۶۴	۳
	لاسو	۰/۹۸۰۰	۰/۱۴۰۴	۰/۰۱۹۸	۰/۱۰۵۹	۳
جنگل تصادفی تصحیح بایاس شده	فیلتر	۰/۹۷۷۹	۰/۱۴۹۶	۰/۰۲۲۵	۰/۱۱۰۹	۳
	لاسو	۰/۹۸۰۹	۰/۱۳۷۵	۰/۰۱۸۹	۰/۱۰۳۱	۴

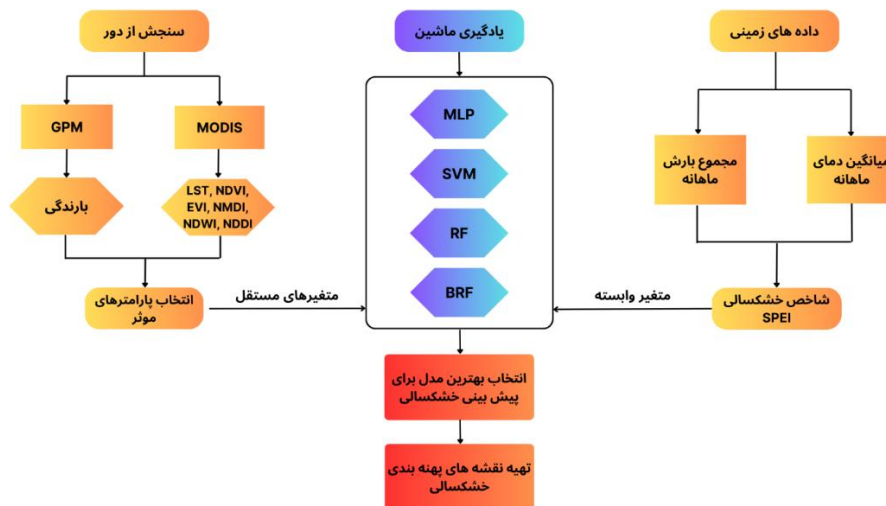
شکل‌ها



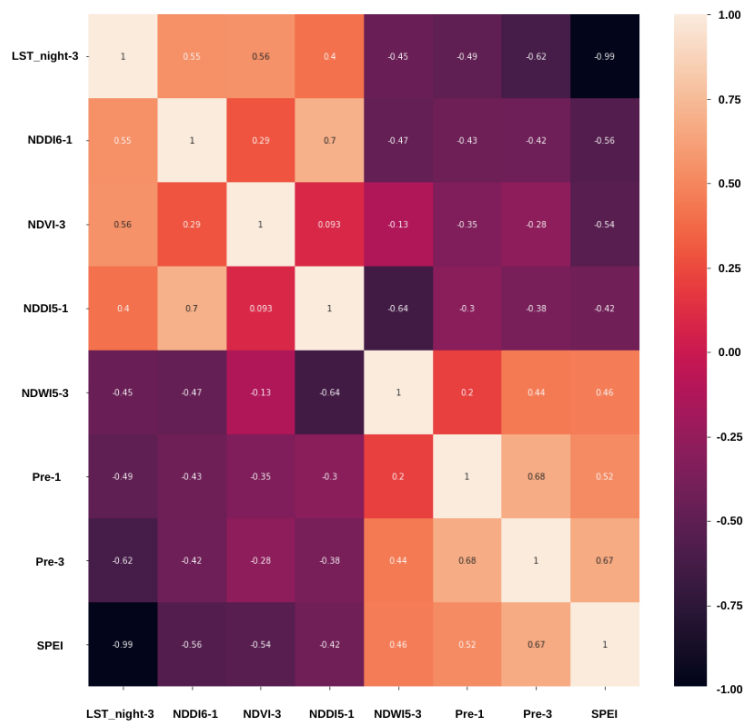
شکل ۱- منطقه‌ی مورد مطالعه به همراه محل قرارگیری ایستگاه‌های هواشناسی مورد استفاده



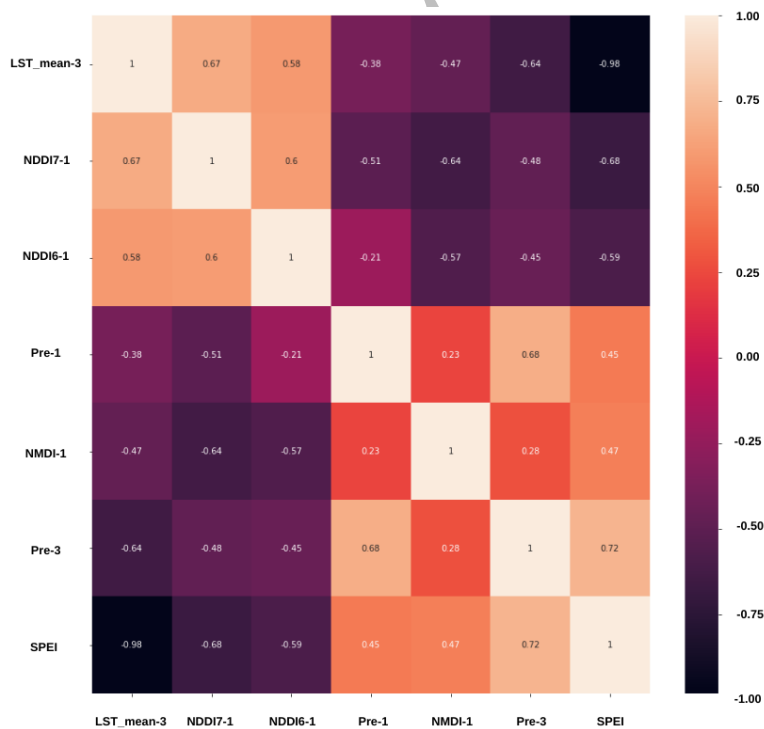
شکل ۲- نتایج خوشه‌بندی به روش K-means



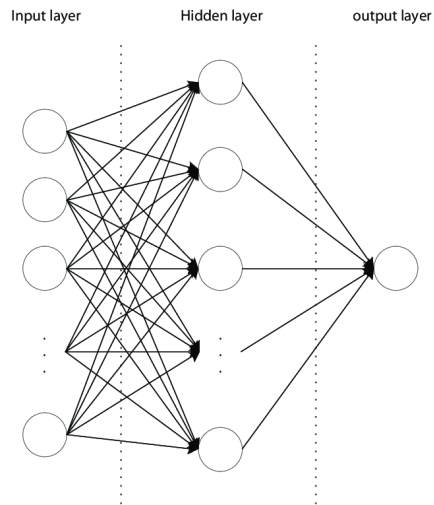
شکل ۳- فلوچارت مراحل انجام این پژوهش



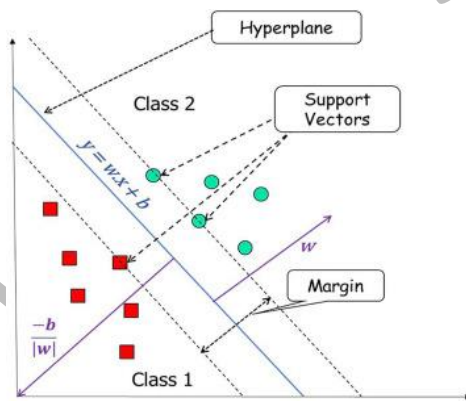
شکل ۴- ماتریس همبستگی پارامترهای انتخاب شده در خوشه‌ی ۱ با روش فیلتر



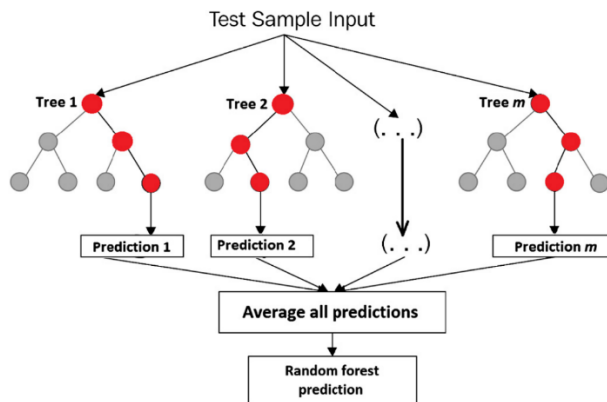
شکل ۵- ماتریس همبستگی پارامترهای انتخاب شده در خوشه‌ی ۲ با روش فیلتر



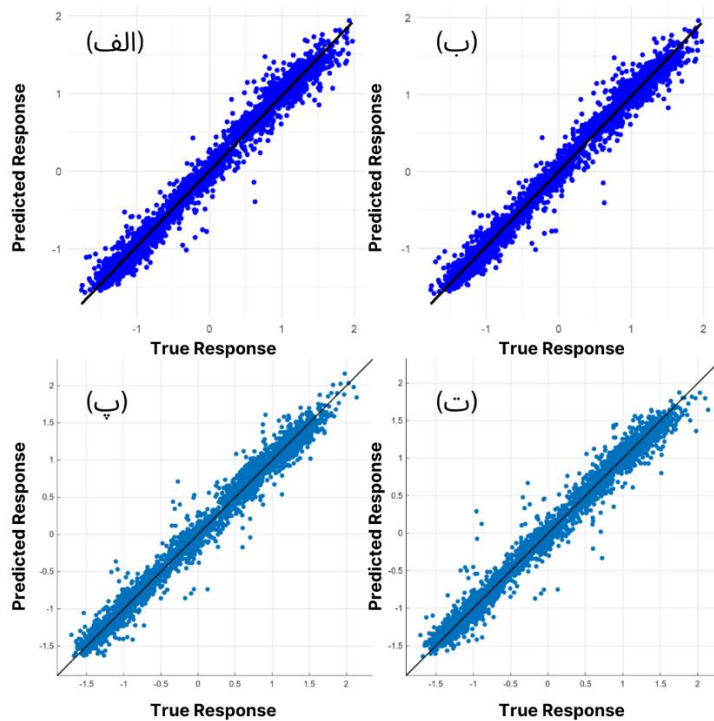
شکل ۶- ساختار شبکه‌ی پرسپترون چند لایه [32]



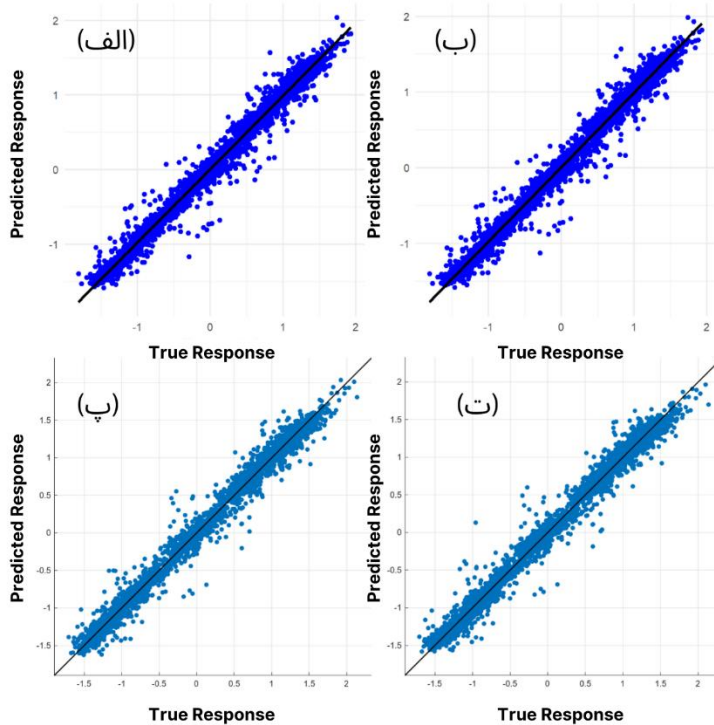
شکل ۷- ساختار الگوریتم ماشین بردار پشتیبان [33]



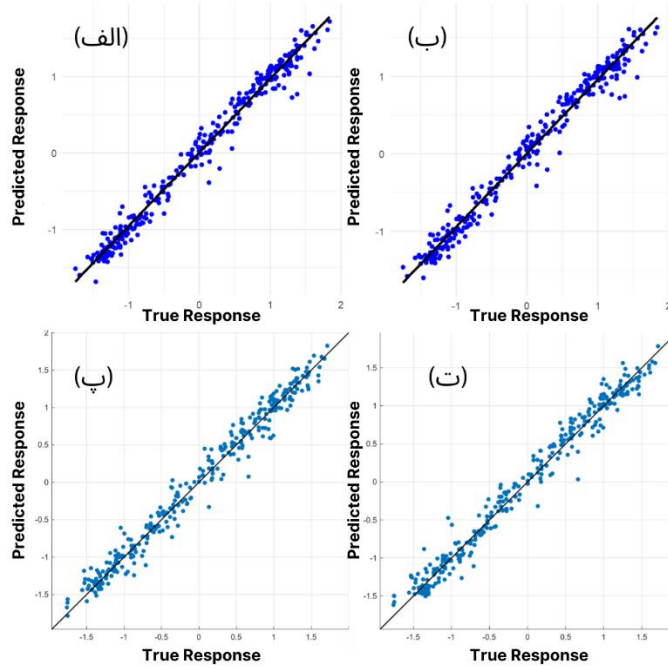
شکل ۸- نحوه‌ی عملکرد الگوریتم جنگل تصادفی [34]



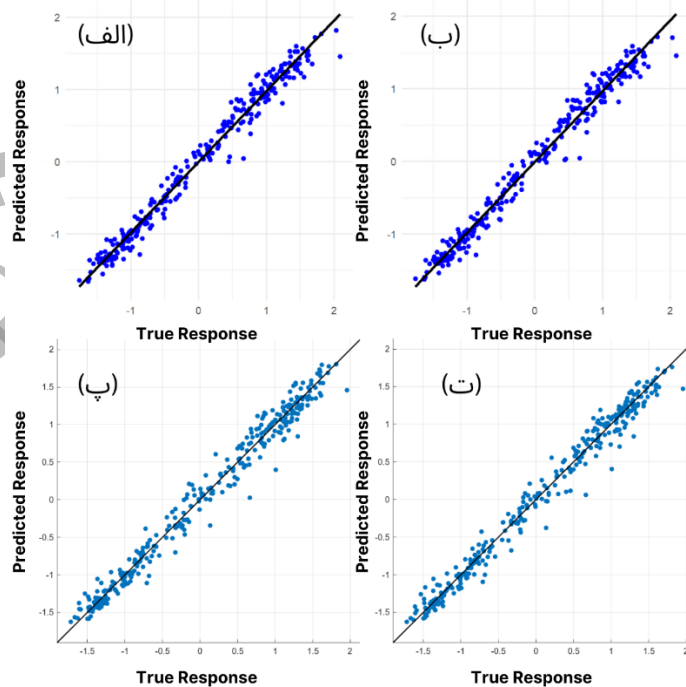
شکل ۹- مقایسه‌ی مقادیر پیش‌بینی شده و واقعی SPEI با روش فیلتر در خوشه‌ی ۱: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده؛ ب) الگوریتم جنگل تصادفی؛ پ) الگوریتم پرسپترون چندلایه؛ ت) الگوریتم ماشین بردار پشتیبان



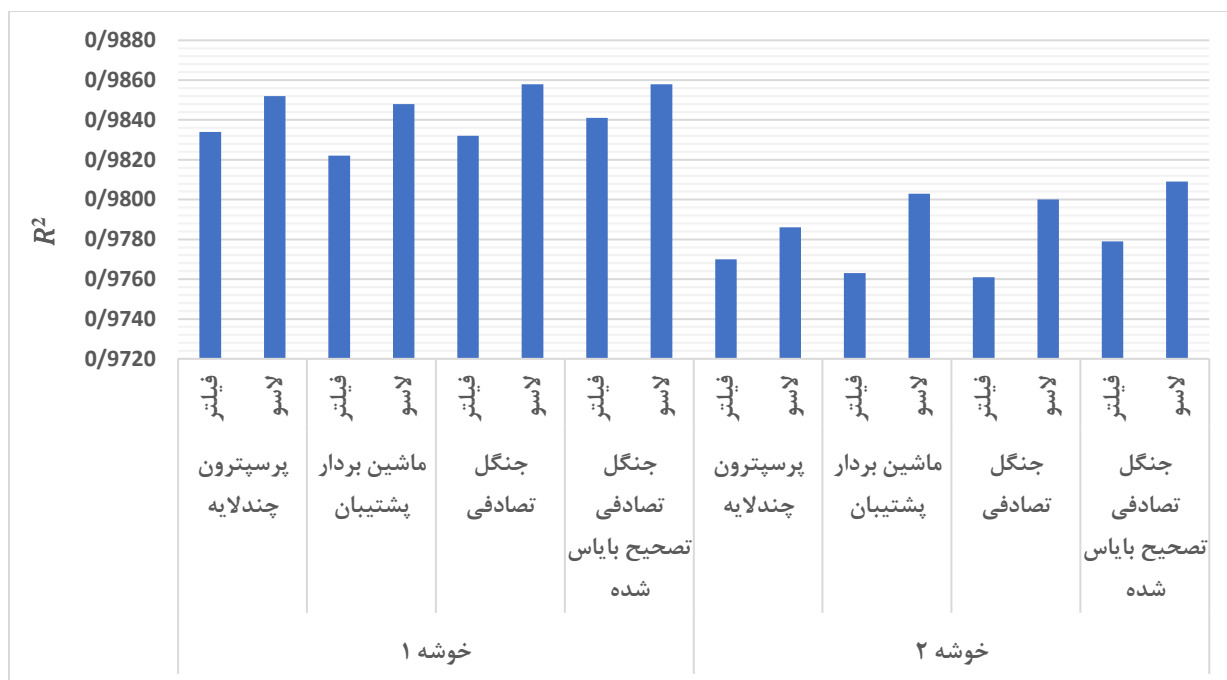
شکل ۱۰- مقایسه‌ی مقادیر پیش‌بینی شده و واقعی SPEI با روش لاسو در خوشه‌ی ۱: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده؛ ب) الگوریتم جنگل تصادفی؛ پ) الگوریتم پرسپترون چندلایه؛ ت) الگوریتم ماشین بردار پشتیبان



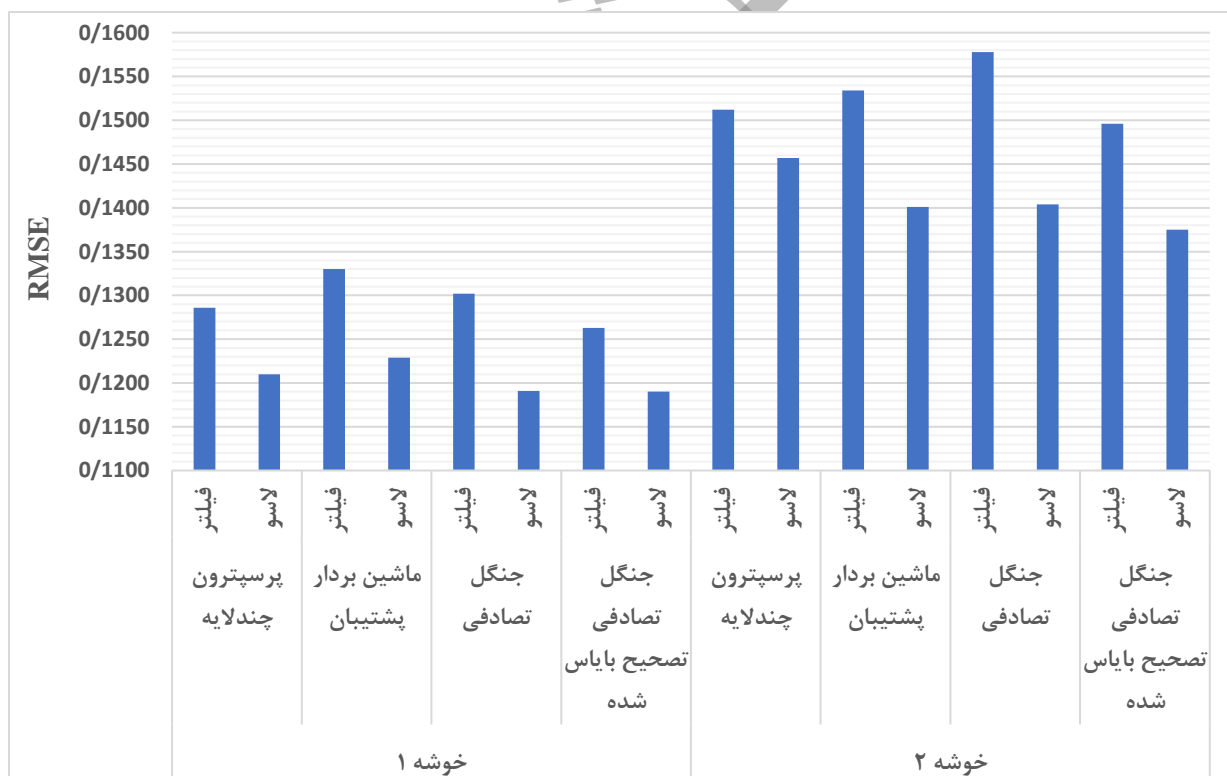
شکل ۱۱- مقایسه‌ی مقادیر پیش‌بینی شده و واقعی SPEI با روش فیلتر در خوشه‌ی ۲: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده؛ ب) الگوریتم جنگل تصادفی؛ پ) الگوریتم پرسپترون چندلایه؛ ت) الگوریتم ماشین بردار پشتیبان



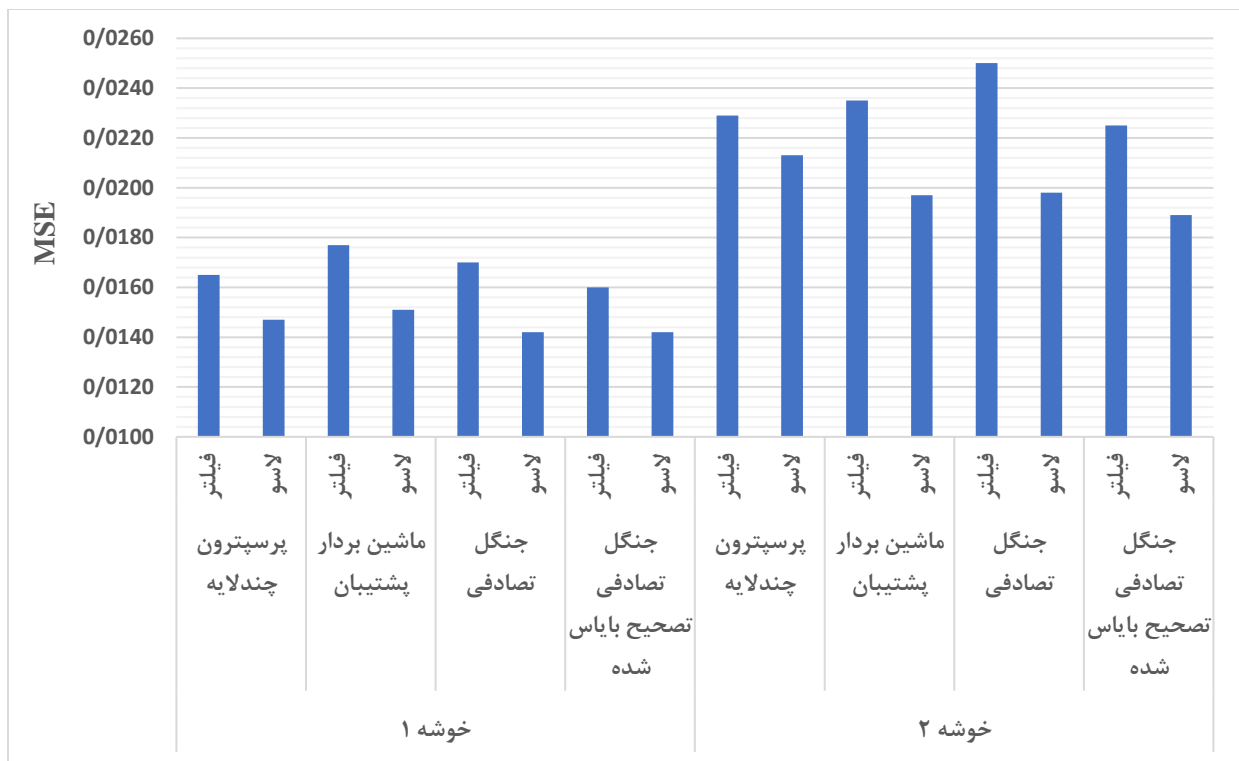
شکل ۱۲- مقایسه‌ی مقادیر پیش‌بینی شده و واقعی SPEI با روش لاسو در خوشه‌ی ۲: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده؛ ب) الگوریتم جنگل تصادفی؛ پ) الگوریتم پرسپترون چندلایه؛ ت) الگوریتم ماشین بردار پشتیبان



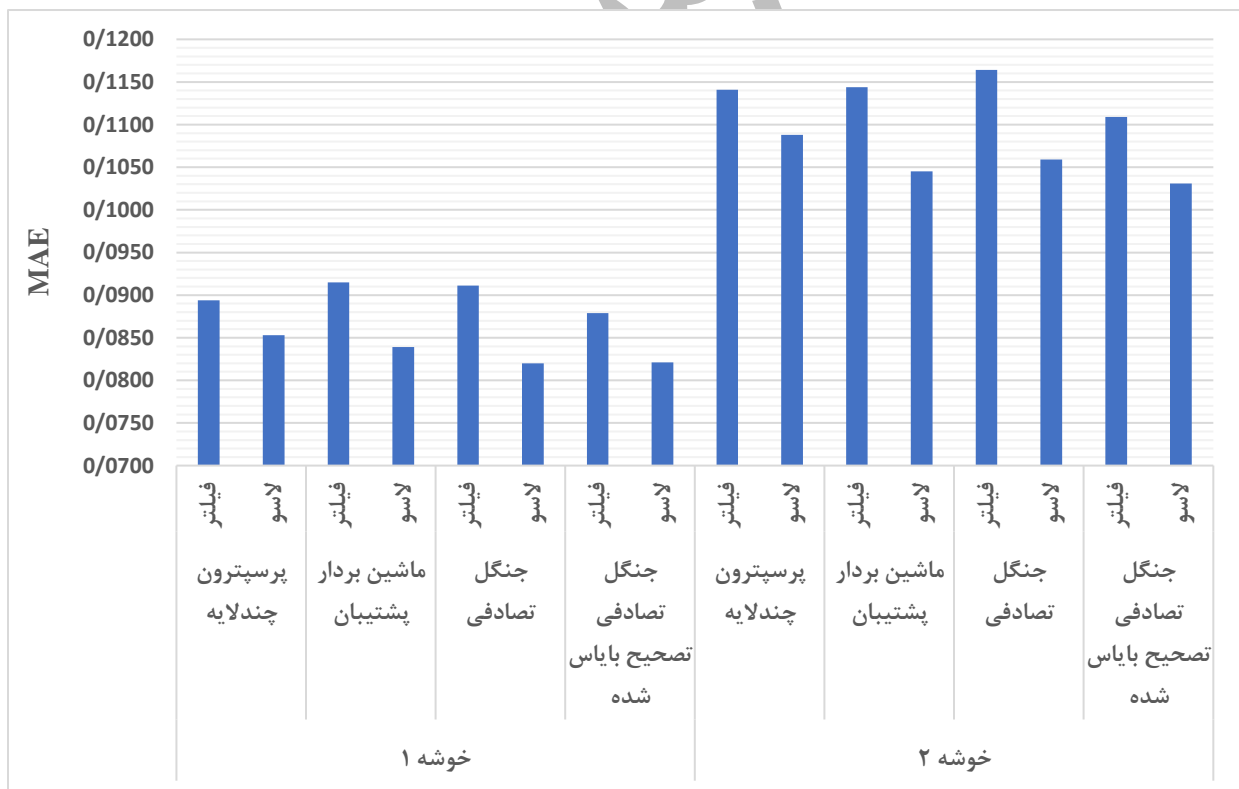
شکل ۱۳- مقایسه‌ی نتایج مدل‌ها بر اساس معیار ضریب تعیین



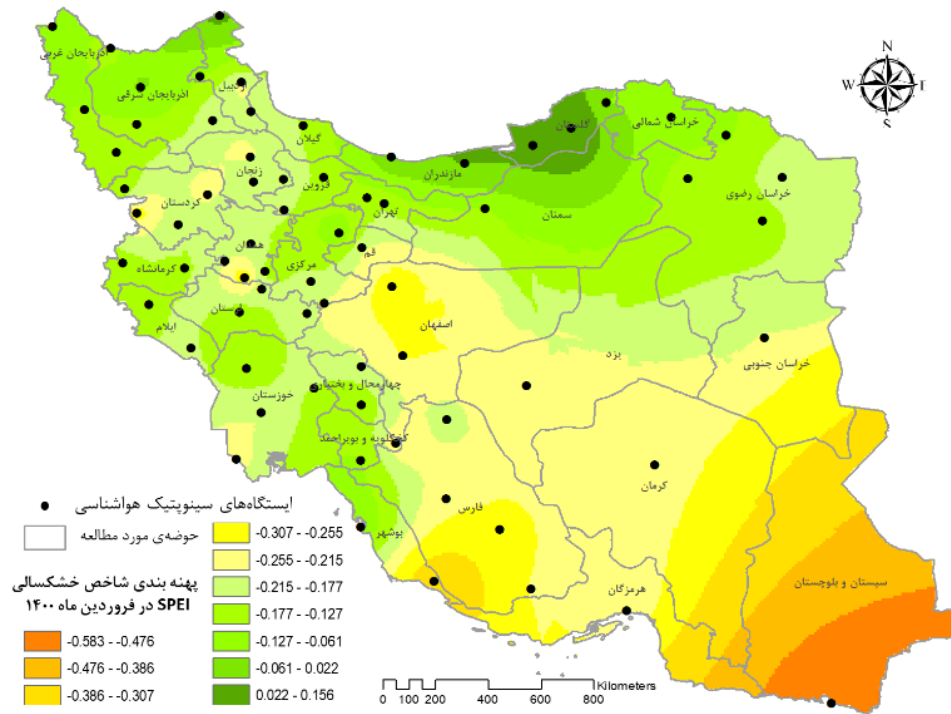
شکل ۱۴- مقایسه‌ی نتایج مدل‌ها بر اساس معیار ریشه‌ی میانگین مربعات خطا



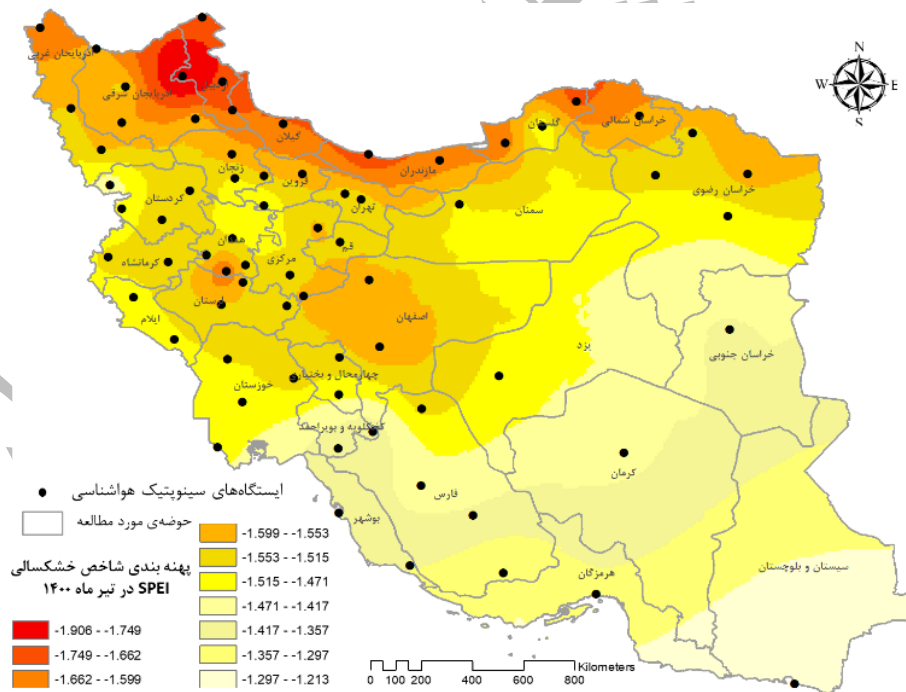
شکل ۱۵- مقایسه‌ی نتایج مدل‌ها بر اساس معیار میانگین مربعات خطا



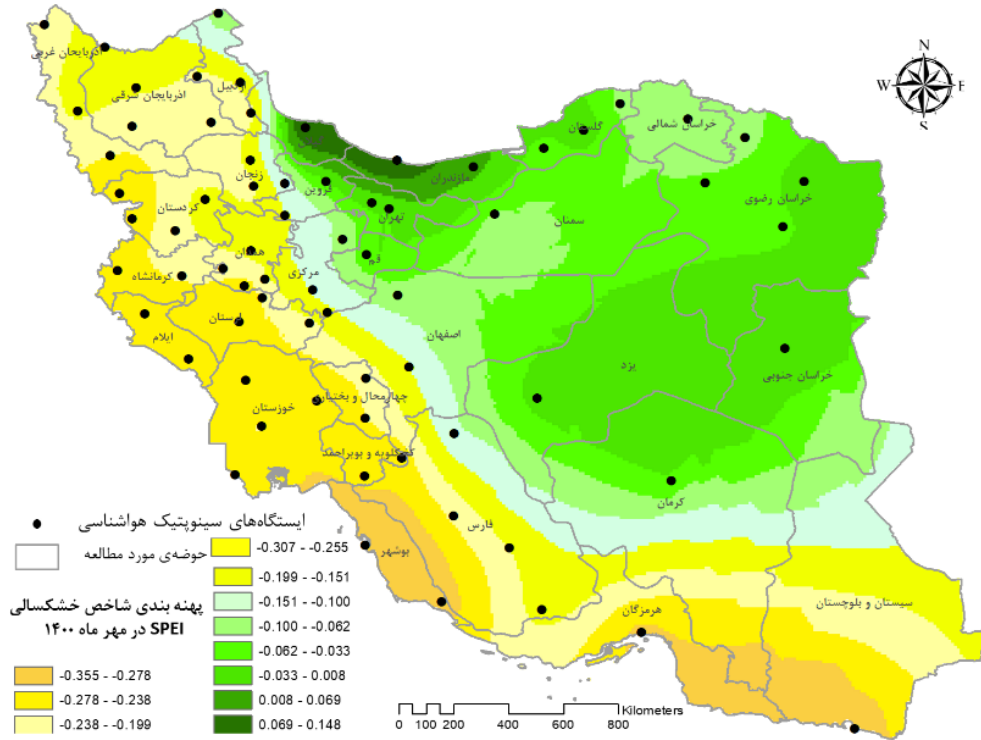
شکل ۱۶- مقایسه‌ی نتایج مدل‌ها بر اساس معیار میانگین خطای مطلق



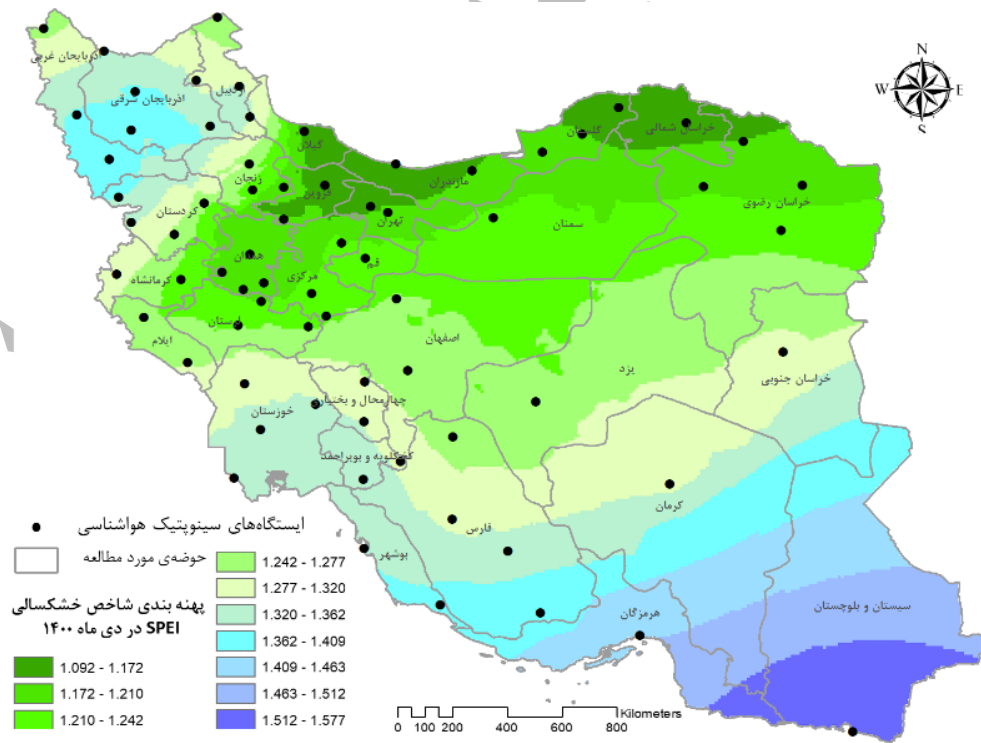
شکل ۱۷- نقشه‌ی پهنه‌بندی خشکسالی کشاورزی در فروردین ماه ۱۴۰۰



شکل ۱۸- نقشه‌ی پهنه‌بندی خشکسالی کشاورزی در تیر ماه ۱۴۰۰



شکل ۱۹- نقشه‌ی پهنه‌بندی خشکسالی کشاورزی در مهر ماه ۱۴۰۰



شکل ۲۰- نقشه‌ی پهنه‌بندی خشکسالی کشاورزی در دی ماه ۱۴۰۰