



Investigating the Performance of Data Mining Models in Rainfall Forecasting and Drought Analysis of Bandar Abbas Synoptic Station

Emad Mahjoobi¹ | Hamid Abdolabadi² | Javad Mahjoobi³ | Ehsan Ghafoori⁴

1. Corresponding Author, Department of Water and Environmental Engineering, Faculty of Civil Engineering, Shahrood University of Technology, Shahrood, Iran. E-mail: emahjoobi@shahroodut.ac.ir
2. Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: h.abdolabadi@scu.ac.ir
3. Faculty of Civil Engineering-Hydraulic Structures, Plan and Development Deputy, Yazd Regional Water Company, Yazd, Iran. E-mail: jmahjoobi@gmail.com
4. Faculty of Civil Engineering, Islamic Azad University, Science and Research Branch, Yazd, Iran. E-mail: ehsan_ghafoori@hotmail.com

Article Info

Article type:
Research Article

Article history:

Received: 10 February 2023
Received in revised form:
4 April 2023
Accepted: 5 June 2023
Published online: 2 July 2023

Keywords:

*Artificial neural network,
Decision tree,
Standard precipitation index,
Support vector machine.*

ABSTRACT

It is common to use different data mining methods in drought prediction. However, the selection of the best model is mainly based on the accuracy of the simulation, while most of the studies do not mention the features of the models. In this paper, the performance of the most common data mining models, including Multilayer Perceptron Artificial Neural Network (ANN-MLP), Radial Base Function Neural Network (ANN-RBF), Regression Decision Tree (CART), Model Tree (M5P), and Support Vector Machine (SVM) is evaluated in order to predict monthly one year ahead rainfall at Bandar Abbas synoptic station and then the characteristics of each of them are described. Calibration and validation of the models were done using raw data and a three-year moving average of climatic parameters from 1347 to 1396. The performance of the models has been evaluated using different statistical indices and comparative diagrams. The results showed that the SVM and M5P models have good prediction performance with RMSE of 7.93 and 8.31 mm, the MAE of 3.66 and 4.69 mm, and the CC of 0.83 and 0.82, respectively. Also, with the exception of the CART, the change in the data mining tool makes an eight to 11 percent difference in the accuracy of the estimates. Therefore, the most appropriate model should be selected based on other characteristics of the methods besides their accuracy. In addition, using the three-year moving average of the input parameters has increased the correlation coefficient by about 78 percent and reduced the RMSE by about 63 percent. The analysis of the long-term drought situation showed that with the increase in the period of the standard precipitation index, the separation of wet and dry years becomes more specific.

Cite this article: Mahjoobi, E., Abdolabadi, H., Mahjoobi, J., & Ghafoori, E. (2023). Investigating the Performance of Data Mining Models in Rainfall Forecasting and Drought Analysis of Bandar Abbas Synoptic Station. *Journal of Water and Irrigation Management*, 13 (2), 429-449. DOI: <https://doi.org/10.22059/jwim.2023.355159.1053>





بررسی عملکرد مدل‌های داده‌کاوی در پیش‌بینی بارش و تحلیل وضعیت خشک‌سالی ایستگاه سینوپتیک بندرعباس

عماد محجوبی^۱ | حمید عبدالآبادی^۲ | جواد محجوبی^۳ | احسان غفوری^۴

۱. نویسنده مسئول، گروه مهندسی آب و محیط‌زیست، دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود، شاهرود، ایران. رایانامه: emahjoobi@shahroodut.ac.ir
۲. گروه مهندسی محیط‌زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. رایانامه: h.abdolabadi@scu.ac.ir
۳. دانشکده مهندسی عمران - سازه‌های هیدرولیکی، معاون طرح و توسعه شرکت آب منطقه‌ای یزد، یزد، ایران. رایانامه: jmahjoobi@gmail.com
۴. دانشکده مهندسی عمران، دانشگاه آزاد اسلامی واحد علوم و تحقیقات یزد، یزد، ایران. رایانامه: ehsan_ghafoori@hotmail.com

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

چکیده

استفاده از روش‌های مختلف داده‌کاوی در پیش‌بینی خشک‌سالی متداول است. با این حال، به‌طور عمده انتخاب مدل برتر بر مبنای دقت شبیه‌سازی صورت می‌گیرد. در حالی که در اغلب مطالعات به ویژگی‌های ساختاری مدل‌ها کمتر توجه شده است. در این مقاله کارایی مجموعه‌ای از متداول‌ترین مدل‌های داده‌کاوی شامل شبکه عصبی مصنوعی چندلایه پرسپترون (ANN-MLP)، شبکه عصبی با تابع پایه شعاعی (ANN-RBF)، درخت تصمیم رگرسیونی (CART)، مدل درختی (MSP) و ماشین بردار پشتیبان (SVM) جهت پیش‌بینی بارش یک سال بعد ایستگاه سینوپتیک بندرعباس ارزیابی شده و ویژگی‌های هر یک از آن‌ها تشریح می‌شود. واسنجی و صحت‌سنجی مدل‌ها با استفاده از داده‌های خام و میانگین متحرک سه ساله پارامترهای اقلیمی در بازه آماری ۱۳۴۷ تا ۱۳۹۶ انجام شد. عملکرد مدل‌ها با استفاده از پارامترهای آماری مختلف و نمودارهای مقایسه‌ای ارزیابی شد. نتایج نشان داد مدل‌های SVM و MSP به ترتیب با مقادیر RMSE برابر ۷/۹۳ و ۸/۳۱ میلی‌متر، MAE برابر ۳/۶۶ و ۴/۶۹ میلی‌متر و ضریب همبستگی ۰/۸۳ و ۰/۸۲ کارایی مطلوبی در پیش‌بینی بارش دارند. همچنین، به‌استثنای مدل CART، تغییر در ابزار داده‌کاوی تفاوت هشت تا ۱۱ درصدی در دقت تخمین‌ها ایجاد می‌کند؛ بنابراین انتخاب مدل مناسب‌تر باید بر مبنای سایر ویژگی‌های روش‌ها در کنار میزان دقت آن‌ها صورت پذیرد. به‌علاوه، بهره‌گیری از میانگین متحرک سه ساله به‌طور متوسط ضریب همبستگی را حدود ۷۸ درصد افزایش و RMSE را حدود ۶۳ درصد کاهش داده است. تحلیل وضعیت درازمدت خشک‌سالی نشان داد با افزایش طول دوره شاخص بارش استاندارد، میزان تفکیک سال‌های مرطوب و خشک مشخص‌تر می‌شود.

کلیدواژه‌ها:

درخت تصمیم، شاخص بارش استاندارد، شبکه عصبی مصنوعی، ماشین بردار پشتیبان.

استناد: محجوبی، عماد؛ عبدالآبادی، حمید؛ محجوبی، جواد و غفوری، احسان (۱۴۰۲). بررسی عملکرد مدل‌های داده‌کاوی در پیش‌بینی بارش و تحلیل وضعیت خشک‌سالی ایستگاه سینوپتیک بندرعباس. *نشریه مدیریت آب و آبیاری*، ۱۳ (۲)، ۴۲۹-۴۴۹. DOI: <https://doi.org/10.22059/jwim.2023.355159.1053>



۱. مقدمه

خشک‌سالی در هر منطقه‌ای مانند سایر پدیده‌های اقلیمی در نتیجه تغییرات آب‌وهوایی ناشی از گردش عمومی اتمسفر می‌باشد که به‌صورت مستقیم و غیرمستقیم جوامع را در دسترسی به منابع آب تحت تأثیر قرار می‌دهد (Loukas *et al.*, 2007). پدیده خشک‌سالی در همه قلمروهای آب‌وهوایی صورت می‌پذیرد، اما در مناطق نیمه‌خشک و نیمه مرطوب مداری به‌دلیل حساسیت سامانه‌های سینوپتیک جوی فراوانی بیش‌تری دارد. کشور ایران به‌واسطه قرارگیری در این موقعیت جغرافیایی، خشک‌سالی‌هایی با شدت و تعداد فراوان را تجربه کرده است. پدیده خشک‌سالی، هزینه‌های زیست‌محیطی، اجتماعی و اقتصادی فراوانی را تحمیل می‌کند. از طرفی انسان‌ها با تمام پیشرفت‌ها در علوم مختلف هنوز نتوانسته‌اند مکان و زمان خشک‌سالی‌های شدید را به‌طور دقیق تعیین کنند (Bachmair *et al.*, 2016). جلوگیری از رخداد پدیده خشک‌سالی تحت هیچ شرایطی در هیچ منطقه‌ای امکان‌پذیر نیست، اما می‌توان با تحلیل و بررسی سوابق تاریخی منطقه موردنظر و اطلاعات در دسترس، دوره‌های بازگشت این رخداد را برآورد نمود و با طرح‌های مدیریتی، مشکلات و خسارات را کاهش داد. برای ایجاد طرح‌های مدیریتی، ضروری‌ترین ابزار، طراحی سامانه‌هایی برای پیش‌بینی خشک‌سالی می‌باشد که می‌توان از آن‌ها اطلاعات دریافت کرده و نحوه مقابله با این پدیده را تعیین نمود. این‌گونه سامانه‌ها با بهره‌گیری از شاخص‌های خشک‌سالی و روش‌های مبتنی بر داده‌کاوی که در دهه‌های اخیر بیش‌تر موردتوجه قرار گرفته‌اند، ایجاد و مورداستفاده قرار می‌گیرند (Mokhtar *et al.*, 2021). شاخص‌های متعددی جهت پیش‌بینی خشک‌سالی و استفاده در طرح‌های مقابله با خشک‌سالی ارائه شده است که از جمله آن‌ها می‌توان شاخص دهک‌ها، شاخص پالم، شاخص درصد نرمال، شاخص رطوبت گیاهی و شاخص استاندارد شده بارش را نام برد. با استفاده از این شاخص‌ها اطلاعات لازم درباره آغاز، خاتمه و شدت خشک‌سالی در مقیاس‌های زمانی و مکانی مختلف فراهم می‌آید. از بین این شاخص‌ها، شاخص استاندارد شده بارش (SPI)^۱ و شاخص پالم کاربرد بیش‌تری در درجه‌بندی خشک‌سالی‌های کشاورزی و هواشناسی دارند (Nikbakht Shahbazi *et al.*, 2012). از طرفی با توجه به گسترش روزافزون سامانه‌های پایگاه داده و ساخت ابزارهای متنوع برای ذخیره حجم زیادی از داده‌ها، داده‌کاوی به‌عنوان یکی از شاخه‌های علوم بین‌رشته‌ای توسعه یافته است. کشف دانش و ارتباط بین داده‌ها دارای هزینه‌های زیاد مادی و نیروی انسانی می‌باشد؛ در این میان کاربرد روش‌های خودکار برای کشف دانش با کم‌ترین دخالت کاربر ضرورت دارد. از این‌رو، کشف اطلاعات نهفته و الگوهای ناشناخته از درون حجم انبوه داده‌ها که از آن به داده‌کاوی تعبیر می‌شود، می‌تواند ابزاری مفید برای توصیف ارتباط میان مجموعه‌ای از داده‌ها جهت استفاده تحلیل‌گران پایگاه داده باشد. یکی از علل اصلی مقبولیت و استفاده روزافزون از روش‌های داده‌کاوی در حوزه مطالعات منابع آب، قابلیت، قدرت و سرعت بالای آن‌ها در بررسی و شبیه‌سازی مسائل اقلیمی - هیدرولوژیکی است؛ به‌ویژه آنجا که این روش‌ها، قادر هستند روند تغییرات داده‌ها، حتی داده‌های مفقوده و مشکل‌دار را استخراج نموده و از طرف دیگر نیازی به توضیح صریح طبیعت فرایندها به‌صورت ریاضی ندارند. از جمله پژوهش‌های بین‌المللی و داخلی صورت‌گرفته بر روی کارایی مدل‌های داده‌کاوی در زمینه پیش‌بینی خشک‌سالی و علوم طبیعی می‌توان به مطالعات ذیل اشاره نمود. Khoshhal Dastjerdi and Hosseini (2010) به‌منظور پیش‌بینی دوره خشک‌سالی در استان اصفهان از شبکه عصبی مصنوعی چندلایه پرسپترون (MLP-ANN)^۲ استفاده کرده و به دقت بالای ۹۵ درصد در پیش‌بینی دوره خشک‌سالی دست یافتند. Nikbakht Shahbazi *et al.* (2012) برای پیش‌بینی خشک‌سالی هواشناسی از مدل ماشین بردار پشتیبان (SVM)^۳ در حوضه سدهای طالقان و ماملو که از تأمین‌کنندگان آب تهران هستند استفاده کردند. نتایج آن‌ها نشان داد که مدل SVM در اغلب موارد پیش‌بینی‌هایی با دقت قابل قبول ارائه می‌دهد و می‌تواند در پیش‌بینی رفتارهای غیرخطی داده‌های هواشناسی

با دوره کوتاه آماری مورد توجه و استفاده قرار گیرد. *Ghorbani et al.* (2015) پیش‌بینی خشک‌سالی سالانه را براساس مدل درختی M5P با توجه به شرایط خشک‌سالی فصل بهار منطقه اراز کوسه استان گلستان مورد ارزیابی قرار دادند. نتایج آن‌ها نشان داد که خشک‌سالی سالانه با توجه به مقادیر بارش در فصل بهار با مقدار ضریب همبستگی ۰/۸۹ و مجذور میانگین مربعات خطا ۰/۵۷ قابل پیش‌بینی است. *Dehbozorgi et al.* (2015) عملکرد مدل MLP-ANN را در پیش‌بینی خشک‌سالی هواشناسی مورد ارزیابی قرار دادند. نتایج حاکی از آن بود که MLP-ANN در روابط غیرخطی خشک‌سالی و بارش به‌خوبی قادر به پیش‌بینی می‌باشد و نتایج به‌دست‌آمده با مقادیر واقعی به‌طور قابل‌ملاحظه‌ای انطباق دارند. *Mozafari et al.* (2016) برای پیش‌بینی خشک‌سالی، کارایی مدل درخت تصمیم رگرسیونی (CART)^۴ را در ایستگاه سینوپتیک سنندج مورد ارزیابی قرار دادند. نتایج آن‌ها نشان داد که مدل CART در پیش‌بینی خشک‌سالی در ایستگاه مذکور کارایی مناسب دارد و استفاده از میانگین متحرک پنج ساله داده‌ها مناسب‌ترین حالت پیش‌بینی را از خود نشان می‌دهد. *Deo et al.* (2017) برای پیش‌بینی خشک‌سالی با شاخص SPI از روش خطوط رگرسیون تطبیقی چندمتغیره (MARS)^۵، ماشین بردار پشتیبان حداقل مربعات (LSSVM)^۶ و مدل درختی M5P در شرق استرالیا استفاده کردند. نتایج آن‌ها نشان داد که روش M5P و MARS نتایج بهتری برای پیش‌بینی ارائه می‌دهند. *Mokhtarzad et al.* (2017) از مدل‌های شبکه عصبی مصنوعی (ANN)^۷، سیستم استنتاج فازی-عصبی تطبیقی (ANFIS)^۸ و SVM به‌منظور پیش‌بینی شاخص SPI در شهر بجنورد استفاده کردند. نتایج آن‌ها نشان داد که مدل SVM عملکرد دقیق‌تری برای پیش‌بینی خشک‌سالی نسبت به دو مدل دیگر دارد. *Zhang et al.* (2020) توانایی روش پیش‌بینی میانگین متحرک یکپارچه خودهمبسته (ARIMA)^۹، مدل شبکه عصبی موجک (WNN)^{۱۰} و SVM را برای پیش‌بینی خشک‌سالی در دشت سانجیانگ کشور چین مورد بررسی قرار دادند. نتایج آن‌ها نشان داد که مدل ARIMA مزیت آشکاری نسبت به دو مدل دیگر در پیش‌بینی خشک‌سالی دارد. *Nivedika et al.* (2021) از الگوریتم MLP-ANN برای پیش‌بینی خشک‌سالی‌های بلندمدت و کوتاه‌مدت در ۱۷ ایستگاه هواشناسی واقع در ایالت‌های راجستان کشور هند استفاده کردند. نتایج آن‌ها حاکی از پیش‌بینی مطلوب سری زمانی SPI با استفاده از دو ماه اول متغیرهای آب‌وهواشناسی بود. *Malik et al.* (2021) مدل‌های MLP-ANN و سیستم استنتاج فازی عصبی (ANFIS) و رگرسیون خطی چندگانه (MLR)^{۱۱} را برای پیش‌بینی خشک‌سالی هواشناسی در هفت ایستگاه کشور هند مورد بررسی قرار دادند. نتایج آن‌ها نشان داد که مدل‌های ANFIS و MLP-ANN عملکرد قابل توجهی داشته‌اند.

همان‌طور که مشاهده می‌شود، پژوهش‌گران زیادی بر توانایی روش‌های مختلف داده‌کاوی در پیش‌بینی خشک‌سالی اشاره داشته‌اند. با این حال، سؤال اساسی هم‌چنان این است که تغییر در ابزار داده‌کاوی چه اثری بر روی شرایط مدل‌سازی و نتایج خواهد داشت؛ چرا که تاکنون به‌طور عمده انتخاب مدل برتر بر مبنای دقت شبیه‌سازی صورت گرفته است، درحالی‌که در اغلب مطالعات به ویژگی‌های مدل‌های برتر اشاره‌ای نشده است. به‌عنوان مثال، شبکه‌های عصبی مصنوعی امکان تحلیل وزن‌های محاسبه‌شده در الگوریتم، یا به‌عبارت دیگر فرایندهای حاکم بر مدل را فراهم نمی‌نماید و تنها دقت در تخمین خروجی مدل مورد اهمیت است. مسئله مهم دیگر، تعداد زیاد پارامترهای مورد نیاز جهت کالیبره کردن مدل بدون توجه به فیزیک مسئله است. ماشین‌های بردار پشتیبان تعداد پارامترهای کم‌تری برای اجرای مدل نسبت به شبکه‌های عصبی مصنوعی دارند، با این وجود، نمایش مناسبی از چگونگی تحلیل و پیش‌بینی ارائه نمی‌دهند. این در حالی است که مدل‌های درختی بر خلاف شبکه‌های عصبی مصنوعی و ماشین‌های بردار پشتیبان، به‌راحتی قابل توضیح و تفسیر می‌باشند. بر این اساس هدف از این پژوهش ارزیابی کارایی مجموعه‌ای از متداول‌ترین مدل‌های داده‌کاوی و بررسی ویژگی‌های هر یک از آن‌ها شامل شبکه عصبی مصنوعی پرسپترون چندلایه (ANN-MLP)،

شبکه عصبی مصنوعی با تابع پایه شعاعی (ANN-RBF)، درخت تصمیم‌گیری رگرسیونی (CART)، مدل درختی (M5P) و ماشین بردار پشتیبان (SVM) در پیش‌بینی خشکسالی در ایستگاه سینوپتیک بندرعباس طی دوره ۱۳۴۷ تا ۱۳۹۶ با استفاده از میانگین متحرک سه ساله داده‌های اقلیمی و سپس تحلیل وضعیت خشکسالی این ایستگاه بر مبنای شاخص بارش استاندارد (SPI) می‌باشد. بدین ترتیب، این مطالعه تحقق اهداف زیر را دنبال می‌نماید: ۱- ارزیابی متداول‌ترین مدل‌های داده‌کاوی در مطالعات خشکسالی؛ ۲- بررسی ویژگی‌های مدل‌های اجرا شده بر روی دو مجموعه داده خام و میانگین متحرک سه ساله؛ ۳- تحلیل نتایج بر مبنای ویژگی‌های مدل و ارزیابی کارایی مدل‌ها و ۴- تحلیل وضعیت خشکسالی بر مبنای جعبه ابزار تجزیه و تحلیل خشکسالی استاندارد (SDAT).

۲. مواد و روش‌ها

در این بخش پس از معرفی منطقه مطالعاتی و مشخصات داده‌ها، مدل‌های شبکه عصبی مصنوعی چند لایه پرسپترون، شبکه عصبی با تابع پایه شعاعی، درخت تصمیم رگرسیونی، درخت مدل و ماشین بردار پشتیبان به صورت اجمالی تشریح می‌شود. سپس، معیارهای سنجش کارایی مدل‌ها و شاخص ارزیابی خشکسالی معرفی می‌گردد. برای به‌کارگیری مدل‌های داده‌کاوی از آمار و اطلاعات مربوط به برخی پارامترهای اقلیمی شامل بارش (P)، دمای متوسط (T_{mean})، دمای حداکثر (T_{max})، دمای حداقل (T_{min})، رطوبت نسبی (RH)، سرعت باد (WS) و جهت باد (WD) که در مقیاس ماهانه در طول سال‌های ۱۳۴۷ لغایت ۱۳۹۶ در ایستگاه سینوپتیک بندرعباس ثبت شده است، استفاده گردید. همگنی داده‌ها با استفاده از روش آزمون توالی مورد بررسی قرار گرفت. نواقص آماری در این ایستگاه تنها مربوط به مقادیر بارش در تعداد محدودی از ماه‌هاست که با استفاده از روش نسبت نرمال تکمیل شد. با توجه به نوسانات شدید پارامترهای اقلیمی در طول دوره مطالعاتی، میانگین متحرک سه ساله داده‌ها در کنار مجموعه داده‌های خام نیز به کار گرفته شد. جهت بررسی روند داده‌ها از آزمون من‌کنندال با توالی سالانه استفاده شد. نتایج نشان داد که برای داده‌های خام، روند کلی بارش نزولی بوده، لیکن پس از استفاده از میانگین متحرک این روند مشخص‌تر شده و به لحاظ آماری با سطح معنی‌داری پنج درصد دارای روند کاهشی است. جهت آموزش مدل‌ها در مرحله واسنجی از ۸۰ درصد داده‌ها در مقیاس ماهانه در بازه ۱۳۴۷ تا ۱۳۸۶ و برای آزمایش آن‌ها در مرحله صحت‌سنجی از ۲۰ درصد داده‌ها مربوط به سال‌های ۱۳۸۷ تا ۱۳۹۶ استفاده شد. شکل (۱) فرایند این پژوهش را نشان می‌دهد.

۲-۱. منطقه مورد مطالعه

شهرستان بندرعباس واقع در استان هرمزگان در حال حاضر شامل پنج بخش بوده و ۴۷ درصد از سطح استان هرمزگان را شامل می‌شود. شهر بندرعباس در بخش مرکزی شهرستان و در موقعیت ۲۷ درجه و ۱۰ دقیقه عرض جغرافیایی و ۵۶ درجه و ۱۱ دقیقه طول جغرافیایی در کنار خلیج فارس و دهانه تنگه هرمز واقع شده است. این شهر دارای هوایی بسیار گرم بوده و با ارتفاع از سطح دریا بین سه تا ۴۰ متر از ویژگی‌های اقلیم بحری برخوردار است. هم‌چنین دارای تابستان‌های گرم و طولانی و زمستان‌های کوتاه و معتدل می‌باشد. متوسط حداکثر و حداقل دمای سالانه شهر بندرعباس به ترتیب حدود ۳۳ و ۲۱ درجه سانتی‌گراد می‌باشد. میانگین بارش درازمدت سالانه شهر بندرعباس در حدود ۱۷۰ میلی‌متر است. شکل ۲، موقعیت ایستگاه سینوپتیک بندرعباس را نشان می‌دهد. این ایستگاه در ارتفاع ۱۰ متری از سطح دریا در سال ۱۳۳۵ تأسیس شده است. جدول (۱)، مشخصات آماری پارامترهای اقلیمی مورد استفاده در این مطالعه را برای سال‌های ۱۳۴۷ لغایت ۱۳۹۶ نمایش می‌دهد.

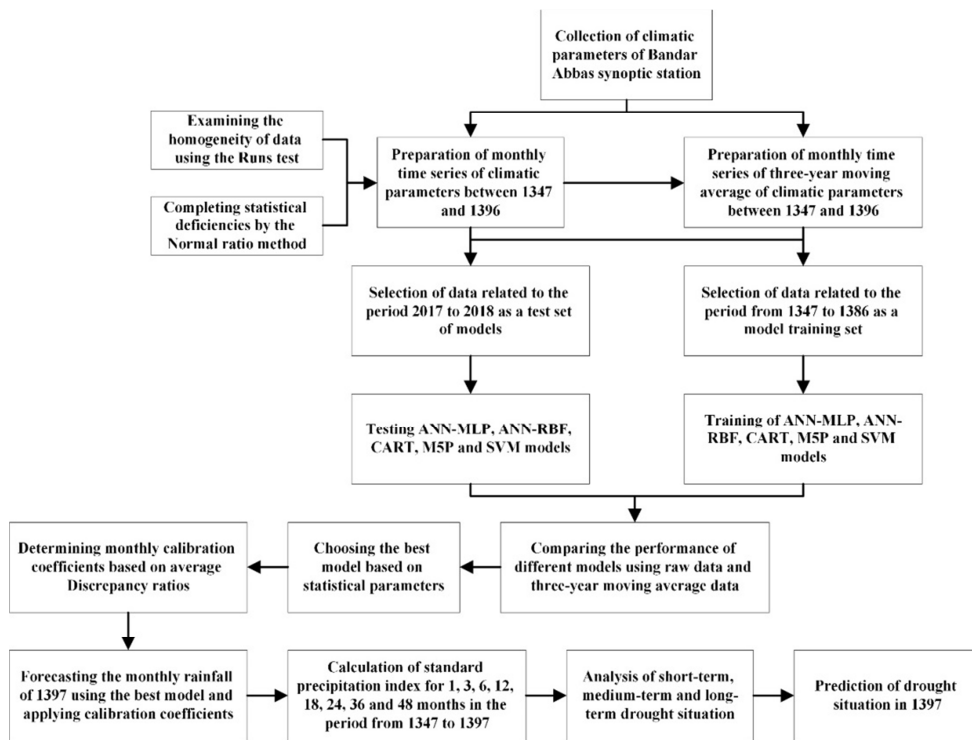


Figure 1. Research Process

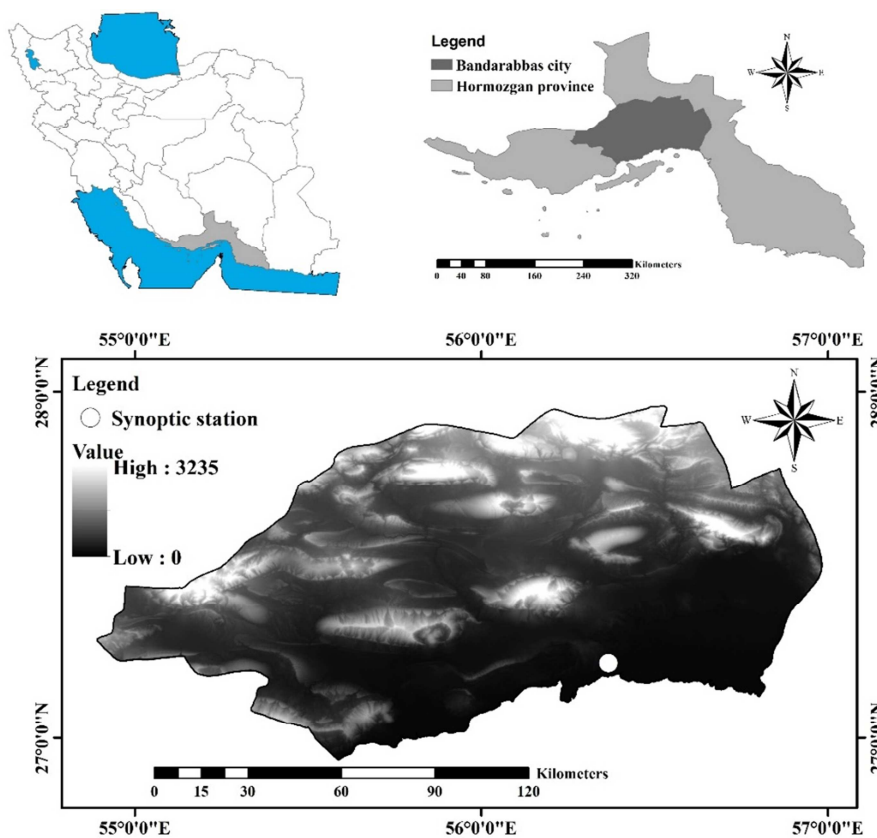


Figure 2. Geographical location of the study area

Table 1. Statistical characteristics of climatic parameters

Parameter	Symbol	Unit	Max	Min	Mean
Precipitation	P	mm	263	0	14.2
Average Temperature	T _{mean}	°C	35.8	14.2	27
Maximum Temperature	T _{max}	°C	51	23	36.6
Minimum Temperature	T _{min}	°C	30	2	17.7
Relative Humidity	RH	%	83	39	64.5
Wind Speed	WS	m/s	70	4	11.2
Wind Direction	WD	Degree	360	10	168.1

۲-۲. شبکه عصبی مصنوعی پرسپترون چندلایه (ANN-MLP)

الگوریتم پرسپترون چندلایه در اوایل دهه ۱۹۷۰ توسعه یافت و یکی از کاربردی‌ترین انواع شبکه‌های عصبی مصنوعی است که از سیستم بیولوژیکی سلول‌های مغز انسان و اتصالات آن الهام گرفته است. شبکه پرسپترون چندلایه یک سامانه پیچیده و غیرخطی است که در آن گره‌های ورودی، مقادیر داده را دریافت کرده و آن‌ها را به اولین گره‌های لایه پنهان منتقل می‌کند (Jahani and Mohammadi, 2018). مقدار گره‌های لایه پنهان با استفاده از رابطه زیر محاسبه می‌شوند.

$$X = \left(\sum_{i=1}^n W_{ij} X_i \right) + b_j \quad \text{(رابطه ۱)}$$

به طوری که، W_{ij} وزن‌های به هم پیوسته داده‌های ورودی X_i ، b_j مقدار انحراف برای نورون و n تعداد داده‌های ورودی است. سپس نتایج از طریق یک تبدیل غیرخطی مانند تابع انتقال سیگموئید در لایه خروجی گزارش می‌شود (روابط ۲ و ۳).

$$Y = F(x) = F \left[\left(\sum_{i=1}^n W_{ij} x_i \right) + b_j \right] \quad \text{(رابطه ۲)}$$

$$F(x) = \frac{1}{1 + e^{-x}} \quad \text{(رابطه ۳)}$$

البته اگر لایه‌های ورودی و خروجی مقادیر منفی داشته باشند، تابع انتقال سیگموئید به صورت رابطه (۴) تغییر می‌یابد. در نهایت مقدار خروجی با مقدار مشاهداتی مقایسه شده و خطای موجود طبق الگوریتم پس انتشار با تنظیم وزن‌ها و مقادیر انحراف به حداقل می‌رسد.

$$F(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad \text{(رابطه ۴)}$$

به طور کلی شبکه‌های عصبی مصنوعی بسیار پارامتریک هستند. مهم‌ترین این پارامترها، معماری شبکه شامل تعداد لایه‌ها و نورون‌های هر لایه، نوع تابع انتقال، ضریب مومنتوم و تعداد تکرار است. یافتن این مقادیر در هر مسئله مستلزم سعی و خطای فراوان کاربر می‌باشد.

۲-۳. شبکه عصبی مصنوعی با تابع پایه شعاعی (ANN-RBF)

شبکه تابع پایه شعاعی تقریباً همیشه از سه لایه ورودی، پنهان و خروجی تشکیل می‌شود. تبدیل از گره‌های ورودی به گره‌های پنهان، غیرخطی بوده و از تابع پایه متقارن شعاعی به عنوان تابع فعال‌سازی گره‌های پنهان استفاده می‌شود (Wu, 2018). مقادیر گره‌های لایه پنهان طبق رابطه (۵) محاسبه می‌شوند:

$$h_j = EXP \left[- \frac{\|X - c_j\|^2}{\sigma_j} \right] \quad \text{(رابطه ۵)}$$

که در آن، h_j خروجی نورون X ، c_j بردار ورودی، c_j مرکز نورون و σ_j پارامتر گسترش مرکز است. یکی از توابع پایه

شعاعی مرسوم تابع انتقال گوسی است. تبدیل از گره‌های پنهان به گره‌های خروجی با استفاده از تابع انتقال خطی انجام می‌شود. می‌توان به سادگی گفت که مقدار گره خروجی برابر جمع وزنی مقادیر تمام نورون‌های پنهان متصل به آن است. در مرحله آموزش بدون نظارت، از روش خوشه‌بندی مبتنی بر فاصله پائو اقلیدسی برای انتخاب تعداد نورون‌های پنهان و مراکز عصبی استفاده می‌شود. در این روش الگوهای ورودی با توجه به شباهت‌های کشف‌شده بین ویژگی‌های ورودی خوشه‌بندی می‌شوند. فرایند خوشه‌بندی توسط آستانه‌ای به نام پارامتر هوشیاری و تابع متریک اقلیدسی اداره می‌شود. در خوشه‌بندی، الگوی اول به‌عنوان مرکز اولین خوشه انتخاب می‌شود. سپس، الگوی بعدی با الگوی مرکز خوشه اول مقایسه می‌شود. اگر فاصله کم‌تر از پارامتر هوشیاری باشد، در خوشه اول قرار می‌گیرد. در غیر این صورت، مرکز یک خوشه جدید است. این روند برای همه الگوها تکرار می‌شود. سپس در مرحله نظارت‌شده، وزن‌های بین لایه‌های پنهان و خروجی به‌طور مستقیم با استفاده از تکنیک شبه‌معکوس محاسبه می‌شود (Karami, 2010).

$$y_k = \sum_{j=1}^m W_{kj} h_j \quad \text{رابطه ۶}$$

۲-۴. درخت تصمیم رگرسیونی (CART)

مسائل پیچیده مدل‌سازی را می‌توان به زیرمسائل کوچک‌تر تقسیم و نتایج را ترکیب کرده و از این طریق راحت‌تر حل کرد. این امر ایده ایجاد و توسعه درخت‌های تصمیم‌گیری بوده است. الگوریتم CART، به‌عنوان یکی از متداول‌ترین روش‌های درخت تصمیم‌گیری توسط Breiman (Breiman et al., 2017)، زمانی که خروجی‌های درخت به‌صورت اعداد حقیقی باشند به آن درخت رگرسیونی و زمانی که خروجی‌های درخت، به‌صورت گسسته باشند به آن درخت دسته‌بندی اطلاق می‌شود (Mitchell, 1997). شاخص‌های مختلفی برای تولید و ایجاد شاخه در درختان تصمیم‌گیری وجود دارد. این شاخص در درختان تصمیم‌گیری رگرسیونی، شاخص حداقل انحراف مربعات (LST) است که به‌صورت زیر تعریف می‌شود:

$$SS(t) = \sum_{i=1}^{N_t} (y_i(t) - \bar{y}(t))^2 \quad \text{رابطه ۷}$$

$SS(t)$ به‌عنوان شاخصی برای ایجاد و تولید درخت تصمیم رگرسیونی، N_t تعداد رکوردها در گره برگ t ، $y_i(t)$ متغیر هدف (مقدار خروجی) در گره برگ t و $\bar{y}(t)$ میانگین مقادیر متغیر هدف برای کل گره‌ها می‌باشد. زمانی متغیر ورودی S مناسب‌ترین متغیر برای ایجاد شاخه در گره t می‌باشد که مقدار $Q(S,t)$ بیشینه شود:

$$Q(S, t) = SS(t) - SS(t_R) - SS(t_L) \quad \text{رابطه ۸}$$

که در آن، $SS(t_L)$ میزان $SS(t)$ در شاخه سمت چپ و $SS(t_R)$ میزان $SS(t)$ در شاخه سمت راست گره می‌باشند.

۲-۵. مدل درختی (M5P)

مدل درختی M5P توسط Quinlan جهت پیش‌بینی داده‌های پیوسته ارائه شد (Quinlan, 1992). اساس کار این مدل قانون تئوری اطلاعات است. در این الگوریتم فضای مسئله به زیرفضاهای کوچک‌تری تقسیم شده و برای هر کدام از این زیرفضاها یک مدل رگرسیون خطی چندمتغیره اختصاص داده می‌شود. این الگوریتم در فضای چندمتغیره توسط یک شاخص کلی کیفی، تقسیم‌بندی‌های ممکن را انجام داده و مدل‌هایی را برای این فضاها به‌صورت خودکار می‌سازد و در نهایت یک درخت مرتبه‌ای با قوانین تقسیم‌بندی‌شده در گره‌های داخلی و خروجی در برگ‌ها ارائه می‌دهد. مدل‌های درختی، خروجی‌ها را به‌صورت رده‌های گسسته نتیجه می‌دهند، به‌طوری‌که یک مدل چندمتغیره خطی را برای هر گروه

از داده‌ها در هر گره ارائه می‌دهند. ساختار این مدل‌ها شامل دو مرحله ایجاد و هرس کردن درخت است (Quinlan, 1992). در مرحله اول برای ساخت درخت تصمیم از یک الگوریتم استقرایی یا معیار انشعاب استفاده می‌شود. معیار انشعاب در الگوریتم M5P بر مبنای ارزیابی انحراف معیار استاندارد خطای مقادیر ذخیره‌شده در هر گره است. همچنین مقدار کاهش موردانتظار در این خطا را تحت عنوان نتیجه آزمون هر صفت در گره محاسبه می‌کند. کاهش انحراف استاندارد از رابطه زیر محاسبه می‌شود:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} * sd(T_i) \quad \text{رابطه ۹}$$

T مجموعه‌ای است که به یک گره می‌رسد، sd نشان‌دهنده انحراف استاندارد و T_i زیرمجموعه‌هایی هستند که ممکن است حاصل از شکست گره باشند. به دلیل انشعابات، انحراف استاندارد داده‌هایی که در گره‌های فرزند قرار می‌گیرند نسبت به داده‌های گره مادر کم‌تر بوده و بنابراین این داده‌ها خالص‌تر هستند.

مرحله دوم ساخت مدل درختی، هرس کردن درخت توسعه یافته و جایگزین کردن درختان فرعی به جای درخت اصلی، با توابع رگرسیونی می‌باشد. پس از بیشینه‌کردن کل انشعابات ممکن، M5P صفتی که کاهش موردانتظار را حداکثر کند، انتخاب می‌کند. این تقسیم باعث ایجاد ساختار شبه‌درختی بزرگی می‌شود که بیش‌برازش مدل را در پی دارد. برای غلبه بر این مسئله از مرحله دوم هرس کردن که به‌صورت جایگزین کردن یک درخت فرعی به جای یک برگ است، استفاده می‌شود.

در خصوص درختان تصمیم، دو تفاوت شاخص مدل CART و مدل درختی M5P در معیار شاخه‌زدن و نحوه محاسبه خروجی در گره‌های برگ می‌باشد. خروجی برگ در الگوریتم CART، میانگین حسابی مقادیر داده‌های رسیده به آن برگ است، اما در مدل درختی M5P، خروجی براساس توسعه یک رابطه چندمتغیره خطی به‌عنوان تابعی از پارامترهایی که از ریشه درخت تا آن برگ سهیم بوده‌اند، محاسبه می‌شود.

۲-۶. ماشین بردار پشتیبان (SVM)

ماشین‌های بردار پشتیبان در سال‌های ۱۹۹۲، ۱۹۹۵ و ۱۹۹۸ توسط Vapnik ساخته و تکمیل و توسعه داده شد. این الگوریتم می‌تواند برای برازش غیرخطی و کلاس‌بندی الگوها مورد استفاده واقع شود (Hykin, 2004). ماشین بردار پشتیبان یک روش یادگیری نظارت شده است که از تئوری یادگیری آماری و به حداقل رساندن ریسک ساختاری مشتق می‌شود (Boser et al., 1992; Vapnik, 2000) و از یک سطح تصمیم برای جداسازی کلاس‌های هدف از طریق به حداکثر رساندن حاشیه بین آن‌ها استفاده می‌کند. با این حال، اغلب ماشین‌های بردار پشتیبان نمونه‌ای از الگوریتم‌های دو کلاس خطی هستند که هدفشان به حداکثر رساندن حاشیه بین دو کلاس است، لذا می‌توان از آن برای طبقه‌بندی یک کلاس استفاده کرد، جایی که سعی می‌شود یک کلاس شناسایی شده و بقیه رد شوند (Roodposhti et al., 2017). سطح تصمیم را معمولاً ابرصفحه بهینه و نزدیک‌ترین نقاط داده به این ابرصفحه را بردارهای پشتیبان می‌نامند. در صورتی که دو کلاس تفکیک‌ناپذیر باشند، مدل تلاش می‌کند که ابرصفحه‌ای که حاشیه بیشینه و احتمال خطای میانگین کلاس‌بندی شده کمینه را دارد، بیابد. این انتخاب توسط یک مقدار مثبت و ثابت که از قبل مشخص است، کنترل می‌شود. این روش می‌تواند برای سطوح غیرخطی نیز استفاده و تعمیم داده شود. Vapnik (1995) با تعریف یک بردار اتلاف غیرحساس، بردار پشتیبان رگرسیونی را پیشنهاد کرد. این بردار پیشنهادی تابعی را دارد که در حد امکان مسطح و دارای انحراف فراوان از بردار هدف واقعی، برای همه داده‌های آموزشی باشد. به بیانی ساده‌تر برای هر داده

آموزشی خطا باید از ε کم‌تر باشد. برای مجموعه داده آموزشی $(x_1, y_1), \dots, (x_k, y_k)$ ، یک تابع خطی به صورت زیر خواهیم داشت:

$$f(x) = \langle w, x \rangle + d \quad \text{رابطه (۱۰)}$$

که در آن، (w, x) ضرب نقطه‌ای در فضای RN، w عضو RN، d عضو R و N بعد فضای ورودی می‌باشد. هر چقدر مقدار w کوچک‌تر باشد، رابطه مذکور مسطح‌تر است. این کار با کمینه کردن فاصله اقلیدسی، $\|w\|^2$ ، امکان‌پذیر می‌باشد که به صورت زیر قابل تعریف است (Smola, 1996).

$$\text{Minimize } \frac{1}{2} \|W\|^2 \quad \text{رابطه (۱۱)}$$

$$\text{s. t. } \begin{cases} y_i - \langle W, x_i \rangle - d \leq \varepsilon \\ \langle W, x_i \rangle + d - y_i \leq \varepsilon \end{cases}$$

مسئله بهینه‌سازی بالا براساس وجود تابعی که خطایی کم‌تر از ε را بر کل زوج‌های آموزشی مشخص می‌کند، استوار است. بنابراین با تعریف ξ تحت عنوان یک متغیر کمکی خطا که شاخصی برای میزان انحراف یک نقطه داده از شرایط ایده‌آل جداسازی الگو می‌باشد، به صورت زیر تعریف می‌شود:

$$\text{Minimize } \frac{1}{2} \|W\|^2 + C \sum_{i=1}^K (\xi_i + \xi'_i) \quad \text{رابطه (۱۲)}$$

$$\text{s. t. } \begin{cases} y_i - \langle W, x_i \rangle - d \leq \varepsilon + \xi_i \\ \langle W, x_i \rangle + d - y_i \leq \varepsilon + \xi'_i \\ \xi_i, \xi'_i \geq 0 \quad \text{for all } i = 1, 2, \dots, k \end{cases}$$

انتخاب بین مسطح‌بودن و میزان انحراف مجاز از خطای ε توسط پارامتر C کنترل می‌شود. شکل مسئله بهینه‌سازی بالا با روش انتقال مسئله به فضای دوتایی و با جایگذاری نامساوی‌ها با یک فرم ساده‌تر و با ضرایب لاگرانژی $\lambda_i, \lambda'_i, \eta_i, \eta'_i$ قابل تعریف و حل می‌باشد (Luenberger and Ye, 1984).

$$f(x, \alpha) = \sum_{i=1}^K (\lambda'_i - \lambda_i) \langle x_i, x \rangle + d \quad \text{رابطه (۱۳)}$$

این روش را می‌توان با تعریف تابع Kernel به بردار رگرسیون پشتیبان غیرخطی تعمیم داد. این کار با اعمال رگرسیون خطی و نگاشت به فضایی با ابعاد بالاتر در فضای مشخصه، امکان‌پذیر است. در نهایت تابع رگرسیون در فضای مشخصه با جایگذاری x_i, x_j با $\varphi(x_i), \varphi(x_j)$ که در آن $K(x_i, x_j) = \varphi(x_i) \times \varphi(x_j)$ است، حاصل می‌شود (Vapnik, 2000).

$$f(x, \alpha) = \sum_{i=1}^K (\lambda_i - \xi_i) K(x_i, x_j) + d \quad \text{رابطه (۱۴)}$$

۲-۷. شاخص‌های ارزیابی عملکرد مدل‌ها

در این مطالعه، کارایی مدل‌های مختلف از طریق محاسبه پارامترهای آماری مجذور میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE)، ضریب کارایی نش-ساتکلیف (NSE)، ضریب انحراف (Bias) و ضریب همبستگی (CC) مورد ارزیابی قرار گرفت.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad \text{رابطه (۱۵)}$$

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n} \quad \text{رابطه ۱۶}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \text{رابطه ۱۷}$$

$$Bias = \bar{P} - \bar{O} \quad \text{رابطه ۱۸}$$

$$CC = \sqrt{1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad \text{رابطه ۱۹}$$

در روابط بالا، P مقادیر شبیه‌سازی شده، O مقادیر مشاهده شده، \bar{P} میانگین مقادیر شبیه‌سازی شده، \bar{O} میانگین مقادیر مشاهده شده و n تعداد داده‌ها است.

۲-۷. تحلیل وضعیت خشک‌سالی

خشک‌سالی به دوره‌ای با کمبود بارش غیرعادی در مقایسه با میانگین بلندمدت در یک منطقه اطلاق می‌شود. شاخص SPI انحراف بارش کل مشاهده شده برای هر مکان معین و دوره زمانی مشخص را نشان می‌دهد. بزرگی انحراف از میانگین، یک اندازه‌گیری احتمالی شدت رویداد مرطوب یا خشک است. از آنجایی که SPI را می‌توان در دوره‌های مختلف تجمع بارندگی (معمولاً بین یک تا ۴۸ ماه) محاسبه کرد، شاخص‌های مختلف SPI به‌دست‌آمده، امکان برآورد آثار مختلف خشک‌سالی را فراهم می‌کنند. زمانی که SPI برای دوره‌های تجمعی کوتاه‌تر محاسبه می‌شود (مثلاً یک تا سه ماه)، می‌توان از آن به‌عنوان شاخصی برای ارزیابی اثرات فوری مانند کاهش رطوبت خاک، ذخیره برف و جریان در رودهای کوچک استفاده کرد. هنگامی که SPI برای دوره‌های تجمعی متوسط (مثلاً سه تا ۱۲ ماه) محاسبه می‌شود، می‌تواند به‌عنوان شاخصی برای ارزیابی دبی جریان و میزان ذخیره‌سازی آب سطحی در پیکره‌های آبی باشد و زمانی که SPI برای دوره‌های طولانی‌تر (مثلاً ۱۲ تا ۴۸ ماه) محاسبه می‌شود، می‌تواند جهت بررسی تغذیه آبخوان‌ها و آب مخازن استفاده شود. در این مقاله از جعبه ابزار تجزیه و تحلیل خشک‌سالی استاندارد (SDAT) که یک چارچوب تعمیم‌یافته برای استخراج شاخص‌های استاندارد تک‌متغیره و چندمتغیره پارامتریک است، استفاده شد (Farahmand and AghaKouchak, 2015). بیش‌تر شاخص‌های خشک‌سالی براساس یک تابع توزیع احتمال پارامتری که با داده‌ها مطابقت دارد، محاسبه می‌گردند. با این حال، یک تابع توزیع پارامتریک ممکن است با داده‌ها انطباق مناسب را نداشته باشد. SDAT براساس یک چارچوب غیر پارامتریک می‌تواند برای متغیرهای آب‌وهوایی مختلف از جمله بارش، رطوبت خاک و رطوبت نسبی بدون نیاز به فرض توزیع پارامتری نماینده اعمال شود.

شاخص بارش استاندارد (SPI) بر مبنای توزیع فراوانی بارش و با استفاده از تابع چگالی احتمال دو پارامتری گاما محاسبه می‌شود. می‌توان به جای استفاده از تابع گاما از احتمال تجربی جهت محاسبه شاخص غیر پارامتریک استاندارد استفاده نمود. رابطه (۲۰)، تابع احتمال تجربی گرینگورتن را نمایش می‌دهد (Gringorten, 1963).

$$p(x_i) = \frac{i - 0.44}{n + 0.12} \quad \text{رابطه ۲۰}$$

که در آن، n تعداد نمونه‌ها و i شماره جایگاه بارش‌ها از کوچک به بزرگ است. با محاسبه مقدار p می‌توان مقدار شاخص استاندارد (SPI) را محاسبه نمود (Farahmand and AghaKouchak, 2015).

$$SPI = \varphi^{-1}(p) \quad \text{رابطه ۲۱}$$

که در آن، φ تابع توزیع استاندارد نرمال است. آستانه‌های طبقه‌بندی خشک‌سالی براساس این شاخص در جدول (۲) نمایش داده شده است.

Table 2. Drought classification thresholds based on standard precipitation index

Index Value	Qualitative Description	Index Value	Qualitative Description
$SPI \leq -2$	Extreme Drought	$0 < SPI \leq 1$	Mildly Wet
$-2 < SPI \leq -1.5$	Severe Drought	$1 < SPI \leq 1.5$	Moderately Wet
$-1.5 < SPI \leq -1$	Moderate Drought	$1.5 < SPI \leq 2$	Severely Wet
$-1 < SPI \leq 0$	Mild Drought	$2 < SPI$	Extremely Wet

۳. نتایج و بحث

در این مقاله با توجه به این که هر شبکه عصبی مصنوعی سه لایه قادر به تقریب هر نوع تابع غیرخطی است، از شبکه پرسپترون سه لایه استفاده شد. تعداد نرون‌های لایه پنهان نیز طبق پیشنهاد Deo *et al.* (2017) حداکثر تا دو برابر تعداد ورودی‌ها به علاوه یک لحاظ شد. بدین ترتیب معماری شبکه با سعی و خطای کم‌تری، بهینه شد. تعداد خوشه‌ها در شبکه عصبی مصنوعی با تابع پایه شعاعی نیز با سعی و خطا مشخص شد. مهم‌ترین پارامترهای روش ماشین بردار پشتیبان نیز، نوع توابع هسته‌ای و ضریب پیچیدگی است که در این مقاله از توابع پایه شعاعی و ضریب پیچیدگی برابر ۱۰ بر مبنای روش Cross Validation برای مدل‌سازی بهره گرفته شد.

پس از آموزش مدل‌ها و واسنجی ضرایب و پارامترهای درونی آن‌ها، صحت‌سنجی مدل‌ها انجام شد. جدول ۳ مقادیر محاسبه‌شده پارامترهای آماری خطا را در مرحله صحت‌سنجی مدل‌های مختلف در شرایط استفاده از سری داده‌های خام و سری داده‌های میانگین متحرک سه ساله نشان می‌دهد. همان‌طور که مشاهده می‌شود بهره‌گیری از میانگین متحرک سه ساله داده‌ها باعث ارتقای چشم‌گیر عملکرد مدل‌ها شده است. به گونه‌ای که به‌طور متوسط ضریب همبستگی حدود ۷۸ درصد افزایش و مجذور میانگین مربعات خطا حدود ۶۳ درصد کاهش یافته است. در مقایسه عملکرد مدل‌های استفاده‌شده با میانگین متحرک سه ساله داده‌ها، مقدار ضریب انحراف برای تمام مدل‌ها به غیر از مدل SVM مثبت است که به معنای بیش‌برآورد مقادیر بارش می‌باشد. مقدار این ضریب برای مدل SVM منفی و قدر مطلق آن کم‌تر از ضریب انحراف سایر مدل‌ها است. از طرف دیگر، مدل CART دارای بیش‌ترین بیش‌برآوردی است. خروجی مدل‌های ANN-MLP، M5P، SVM همبستگی بالای ۰/۸ با داده‌های مشاهداتی داشته و مدل CART دارای کم‌ترین مقدار ضریب همبستگی است. بررسی پارامترهای RMSE و MAE نشان می‌دهد که مدل SVM کم‌ترین خطا و مدل CART بیش‌ترین خطا را در شبیه‌سازی بارش داشته است. از طرفی طبق شاخص NSE، مدل‌های SVM، M5P، و ANN-MLP کارایی بالاتر از ۰/۶ و مطلوبی را داشته‌اند. البته مدل ANN-RBF نیز کارایی بسیار نزدیکی به این محدوده دارد. به هر حال، بررسی پارامترهای آماری مختلف حاکی از آن است که تفاوت قابل ملاحظه‌ای در عملکرد مدل‌های SVM، M5P و ANN-MLP دیده نشده و هر سه رفتار قابل‌قبولی دارند.

در همین رابطه Zainudin *et al.* (2016) نیز پس از بررسی عملکرد پنج مدل بیزین، درخت تصمیم، ماشین بردار پشتیبان، شبکه عصبی و جنگل تصادفی در پیش‌بینی بارش در چند ایستگاه در مالزی اعلام کردند که حداکثر اختلاف دقت مدل‌ها حدود نه درصد می‌باشد. در ضمن، بر ویژگی‌های خاص مدل‌های درخت تصمیم و جنگل تصادفی تأکید کردند. همچنین Murad and Salih (2020) با بررسی عملکرد سه مدل شبکه عصبی مصنوعی، ماشین بردار پشتیبان و بیزین در پیش‌بینی بارش ایستگاه سلیمانیه عراق، بر نزدیکی دقت هر سه مدل (اختلاف دقت حداکثر دو درصد) صحه گذاشتند. Anwar *et al.* (2020) نیز عملکرد چهار مدل درخت تصمیم، جنگل تصادفی، بیزین و پرسپترون چندلایه را در پیش‌بینی بارش ایستگاهی در اندونزی ارزیابی کرده و نشان دادند که اختلاف دقت مدل‌ها در شرایط مختلف مدل‌سازی حداکثر در محدوده سه درصد قرار دارد. بر این اساس توجه به سایر ویژگی‌های روش‌ها در کنار میزان دقت آن‌ها می‌تواند مبنای انتخاب مدل برتر باشد.

Table 3. Errors of statistical parameters in validation of different models with raw and three-year moving average datasets at a monthly time step

Model	Three-year Moving Average Datasets					Raw Dataset	
	Bias	CC	NSE	RMSE	MAE	CC	RMSE
ANN-MLP	1.42	0.813	0.649	8.39	5.27	0.460	24.42
ANN-RBF	1.46	0.794	0.595	9.00	5.33	0.459	25.10
CART	2.90	0.764	0.314	11.71	5.92	0.403	26.17
M5P	1.24	0.828	0.655	8.31	4.39	0.467	23.89
SVM	-0.34	0.833	0.686	7.93	3.66	0.481	22.22

شکل (۳)، عملکرد مدل‌ها در شبیه‌سازی بارش ماهانه بر مبنای میانگین متحرک سه ساله داده‌ها را نسبت به خط دقیق نشان می‌دهد. همان‌طور که مشاهده می‌شود بیش‌ترین انطباق میان نتایج مدل و مقادیر مشاهداتی در SVM و به‌ویژه در بارش‌های کم‌تر از ۲۰ میلی‌متر دیده می‌شود.

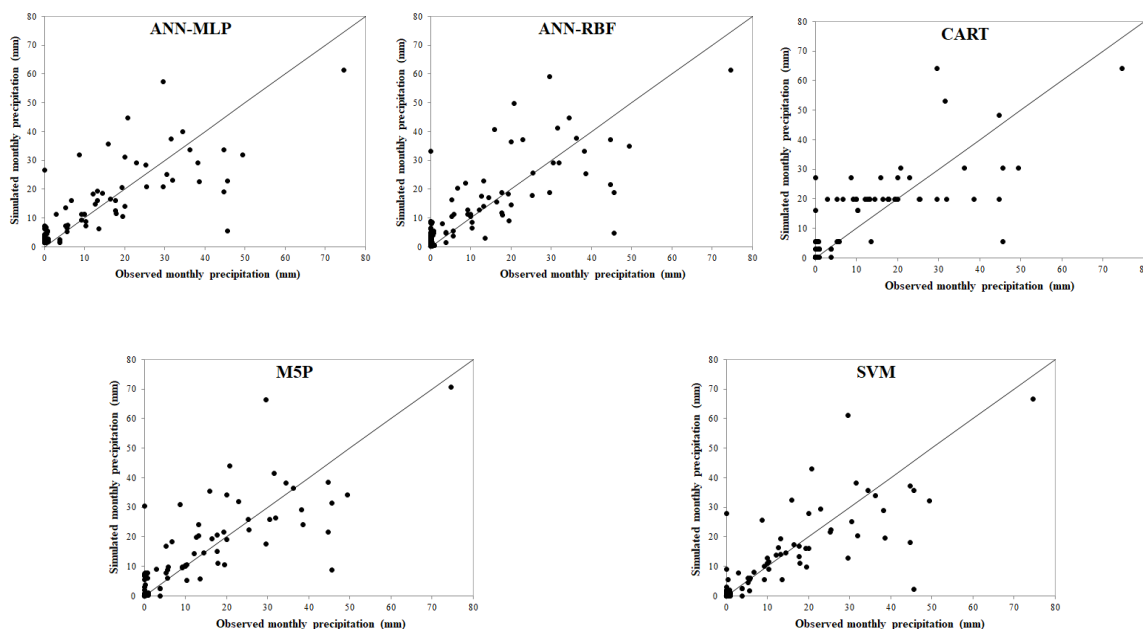


Figure 3. Performance of different models in monthly one year ahead rainfall simulation based on three-year moving average dataset

شکل (۴)، سری زمانی مقادیر مشاهداتی و شبیه‌سازی در مدل‌های مختلف استفاده‌شده بر مبنای میانگین متحرک سه ساله داده‌ها را نشان می‌دهد. با وجود این‌که کلیه مدل‌ها روند تغییرات بارش را به‌صورت قابل‌قبولی شبیه‌سازی کرده‌اند، مشاهده می‌شود که مدل ANN-MLP خطای قابل‌توجهی در شبیه‌سازی بارش‌های کم‌تر از پنج میلی‌متر در ماه دارد. درحالی‌که عملکرد مدل M5P و SVM در این محدوده بسیار مطلوب است. از طرفی مدل SVM توانایی بالایی در انطباق با رفتار بارش‌ها در مقادیر پیک نیز دارد.

شکل (۵)، نمودارهای ویالونی مربوط به عملکرد مدل‌ها در پیش‌بینی بارش در مقایسه با مقادیر مشاهداتی بر مبنای میانگین متحرک سه ساله داده‌ها را نشان می‌دهد. بر این اساس، بیشینه مقادیر در M5P کم‌ترین اختلاف (در حدود پنج درصد) و در ANN-RBF بیش‌ترین اختلاف (در حدود ۱۸ درصد) را با بیشینه مقادیر مشاهداتی دارد. کمینه مقادیر در M5P و SVM معادل کمینه مقادیر مشاهداتی برآورد شده است، اما در ANN-MLP و CART کلیه بارش‌ها غیر صفر

و در ANN-RBF بارش‌های منفی نیز تخمین زده شده‌اند. سری شبیه‌سازی شده توسط مدل SVM نزدیک‌ترین میانه و میانگین را نسبت به سری مشاهداتی دارد. در مجموع بیش‌ترین شباهت در توزیع داده‌ها در مدل M5P و SVM و کم‌ترین شباهت در مدل CART دیده می‌شود.

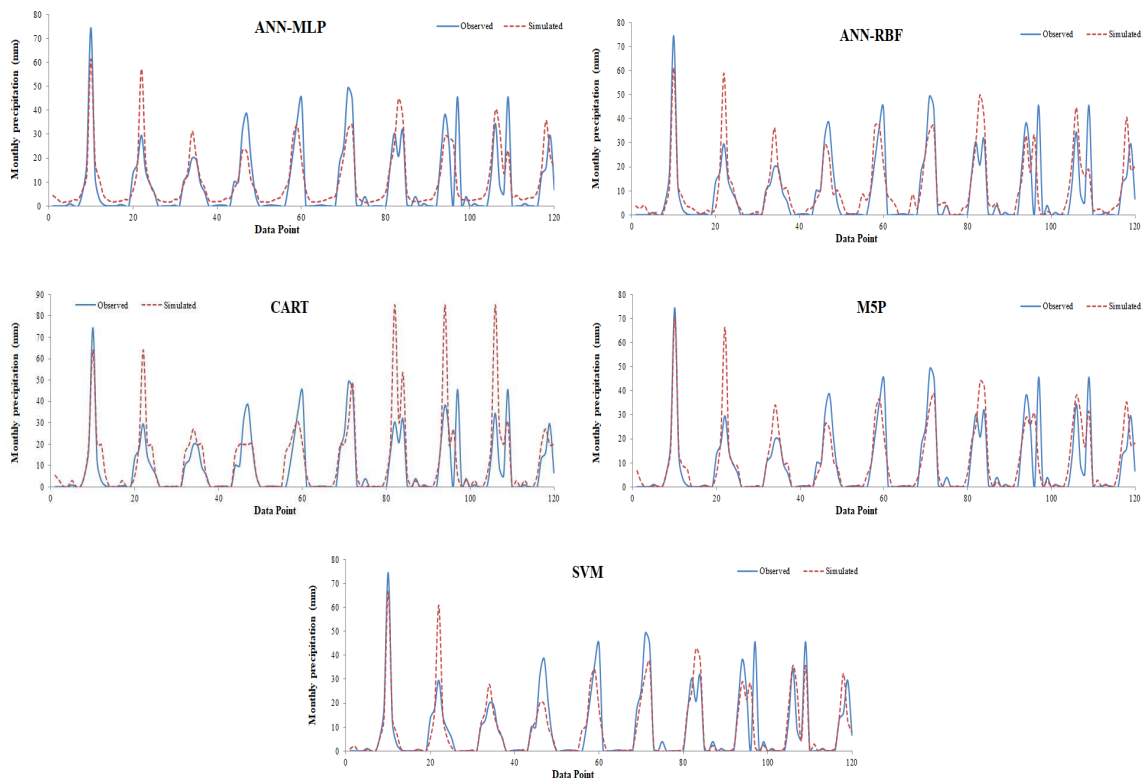


Figure 4. Observed and simulated monthly precipitation time series in validation of different models with three-year moving average dataset

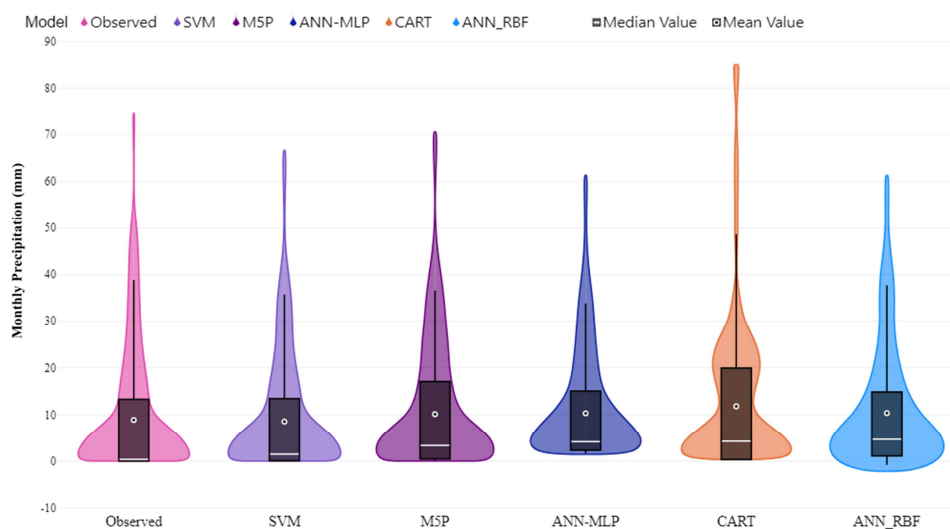


Figure 5. Violin diagram of observed and simulated values of monthly precipitation in validation of different models with three-year moving average dataset

شکل (۶)، دیاگرام تیلور مقایسه عملکرد مدل‌ها بر مبنای میزان همبستگی، مجذور میانگین مربعات خطا و انحراف معیار داده‌ها را نشان می‌دهد. بیش‌ترین همبستگی و کم‌ترین خطا مربوط به مدل SVM است. مدل M5P نیز نزدیک‌ترین انحراف معیار را به داده‌های مشاهداتی دارد. لذا براساس بررسی‌ها از منظرهای مختلف، این گونه استنباط می‌گردد که بهترین عملکرد مربوط به مدل SVM و پس از آن M5P است. نکته قابل‌توجه این است که با عنایت به اختلاف بسیار ناچیز در عملکرد این دو مدل به‌نظر می‌رسد انتخاب مدل برتر باید بر مبنای میزان تفسیرپذیری نتایج مدل صورت پذیرد. خروجی مدل SVM قابلیت تفسیر ندارد، درحالی‌که در شکل ۷ قوانین خروجی مدل M5P برای پیش‌بینی بارش ۱۲ ماه بعد ارائه شده است.

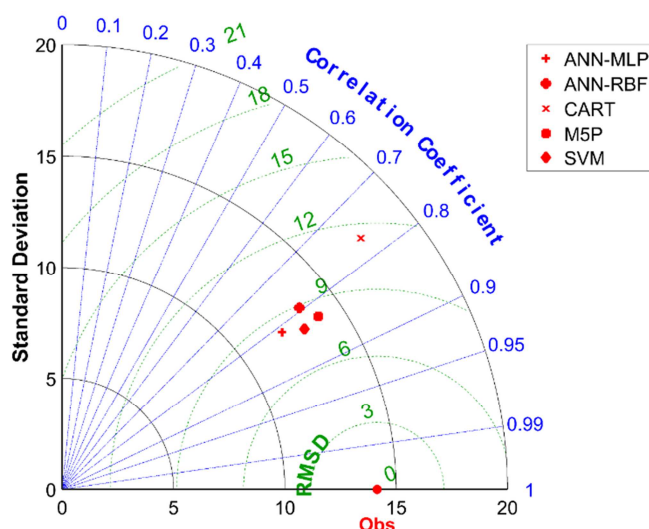


Figure 6. Taylor diagram comparing the performance of different models developed based on the three-year moving average of the data

$$P \leq 12.815$$

$$T_{min} \leq 19.8$$

$$\text{Rule\#1: } P_{Lead12} = 0.0786P + 0.0326T_{min} + 0.9618T_{max} - 2.1889T_{mean} + 0.5659RH + 0.5861WS - 13.294$$

$$T_{min} > 19.8$$

$$\text{Rule\#2: } P_{Lead12} = 0.6093P + 0.0187T_{min} - 0.1206T_{mean} + 0.0365RH + 0.0227WS + 1.2497$$

$$P > 12.815$$

$$P \leq 49.415$$

$$\text{Rule\#3: } P_{Lead12} = 0.5802P - 0.1641T_{min} - 1.2479T_{mean} + 0.8059RH - 11.468$$

$$P > 49.415$$

$$T_{min} \leq 6.5$$

$$\text{Rule\#4: } P_{Lead12} = 0.171P + 7.0675T_{min} + 4.5903T_{mean} + 0.2067RH + 0.061WS - 76.67$$

$$T_{min} > 6.5$$

$$\text{Rule\#5: } P_{Lead12} = 0.171P - 7.8058T_{min} + 6.937T_{mean} + 0.2067RH + 0.0317WS - 37.8917$$

Figure 7. M5P model tree output rules in monthly one year ahead rainfall simulation based on three-year moving average dataset

قوانین ارائه شده توسط مدل M5P نشان می‌دهد که پارامترهای مؤثر بر مقدار بارش یک سال بعد در شرایط بارش‌های کم‌تر از ۱۲/۸۱۵ میلی‌متر و حداقل دمای کم‌تر از ۱۹/۸ درجه سانتی‌گراد به ترتیب دمای متوسط، حداکثر دما، سرعت باد، رطوبت نسبی، بارش و حداقل دما و در شرایط بارش‌های کم‌تر از ۱۲/۸۱۵ میلی‌متر و حداقل دمای بیش‌تر از ۱۹/۸ درجه سانتی‌گراد به ترتیب بارش، دمای متوسط، رطوبت نسبی، سرعت باد و حداقل دما هستند. در بارش‌های بزرگ‌تر از ۱۲/۸۱۵ تا ۴۹/۴۱۵ میلی‌متر به ترتیب دمای متوسط، رطوبت نسبی، بارش و حداقل دما بر مقدار بارش سال بعد مؤثر هستند. درحالی‌که در بارش‌های بزرگ‌تر از ۴۹/۴۱۵ میلی‌متر، پارامترهای مؤثر به ترتیب دمای متوسط، حداقل دما، رطوبت نسبی، بارش و سرعت باد هستند. قوانین یک تا سه نشان می‌دهند که در بارش‌های کم‌تر از ۴۹/۴۱۵ میلی‌متر، مقدار بارش یک سال بعد با دمای متوسط رابطه عکس دارد؛ یعنی با افزایش دما، مقدار بارش سال بعد کاهش می‌یابد. از طرفی حداکثر دما تنها در شرایط بارش‌های کم‌تر از ۱۲/۸۱۵ میلی‌متر و حداقل دمای کم‌تر از ۱۹/۸ درجه سانتی‌گراد بر مقدار بارش سال بعد اثرگذار است. جهت باد نیز در هیچ‌یک از قوانین ارائه شده اثری بر خروجی مدل ندارد. به هر حال، با توجه به مقادیر پارامترهای آماری مختلف، مدل SVM بیش‌ترین دقت را در پیش‌بینی بارش یک سال بعد داشته است، بنابراین برای دقیق‌تر کردن مقدار پیش‌بینی، نسبت به تعیین ضرایب کالیبراسیون ماهانه براساس نتایج این مدل اقدام شد. به این ترتیب که ابتدا برای هر ماه از سال نسبت ناهمخوانی $^{12}(Dr)$ طبق رابطه (۲۲) محاسبه شد. سپس ضریب کالیبراسیون برای هر ماه از میانگین نسبت‌های ناهمخوانی ماه‌های مشابه در طول دوره آماری به دست آمد.

$$Dr_i = \frac{O_i}{P_i} \quad \text{رابطه (۲۲)}$$

که در آن، P_i مقادیر پیش‌بینی شده و O_i مقادیر واقعی می‌باشد. بر این اساس از اردیبهشت‌ماه تا مردادماه و همچنین مهرماه ضریب کالیبراسیون معادل صفر اعمال می‌شود. ضریب کالیبراسیون ماه‌های آذر و دی نیز بسیار نزدیک به یک است (جدول ۴).

Table 4. Calibration coefficients of SVM model results by different months

Month	Calibration Coefficient	Month	Calibration Coefficient	Month	Calibration Coefficient	Month	Calibration Coefficient
January	0.98	April	0.86	July	0	October	0
February	1.2	May	0	August	0	November	0.58
March	1.15	June	0	September	0.2	December	1.03

در ادامه، بارش ماه‌های مختلف سال ۱۳۹۷ با استفاده از مدل SVM و ضرایب کالیبراسیون پیش‌بینی شد. سپس با محاسبه شاخص بارش استاندارد (SPI) در دوره‌های یک، سه، شش، ۱۲، ۱۸، ۲۴، ۳۶ و ۴۸ ماهه، وضعیت خشک‌سالی در کل دوره آماری ۱۳۴۷ تا ۱۳۹۷ مورد بررسی قرار گرفت (شکل ۸).

مطابق شکل مشاهده می‌شود که شاخص SPI برای دوره‌های کوتاه‌تر (یک و سه ماه) تغییرات زیادی را در بارش هشت ماه ابتدایی سال نسبت به میانگین بلندمدت نشان نمی‌دهد. بدین ترتیب، مشاهده می‌گردد که سال ۱۳۹۷ در وضعیت نرمال با تغییرات خفیف بارش قرار دارد. با افزایش طول دوره بررسی شاخص (شش و ۱۲ ماه) مشاهده می‌شود که وضعیت سال ۱۳۹۷ در محدوده خشک، خشکی خفیف و خشکی شدید قرار می‌گیرد. همچنین، شاخص ۱۲ ماهه بیش‌ترین شدت کمبود بارش را در بین تمامی دوره‌های مورد بررسی برای سال ۱۳۹۷ نشان می‌دهد. نکته بسیار قابل توجه این است که برای دوره‌های طولانی‌تر (۱۸، ۲۴، ۳۶ و ۴۸ ماه) مشاهده می‌گردد که سال‌های قرار گرفته در محدوده‌های خشک و مرطوب قابل تشخیص و تفکیک هستند، به نحوی که سال‌های ۱۳۴۷ تا ۱۳۵۴ (دوره هفت ساله) و

سال‌های ۱۳۸۱ تا ۱۳۹۳ (۱۲ سال) در وضعیت کم‌بارشی قرار دارد. از طرف دیگر از سال ۱۳۵۴ تا ۱۳۸۰ بارش در وضعیت بهتری قرار دارد. جدول ۵، اطلاعات مربوط به بررسی شاخص استاندارد بارش با طول دوره‌های مختلف را براساس طبقه‌بندی خشکی برای ۶۱۲ ماه ارائه می‌دهد.

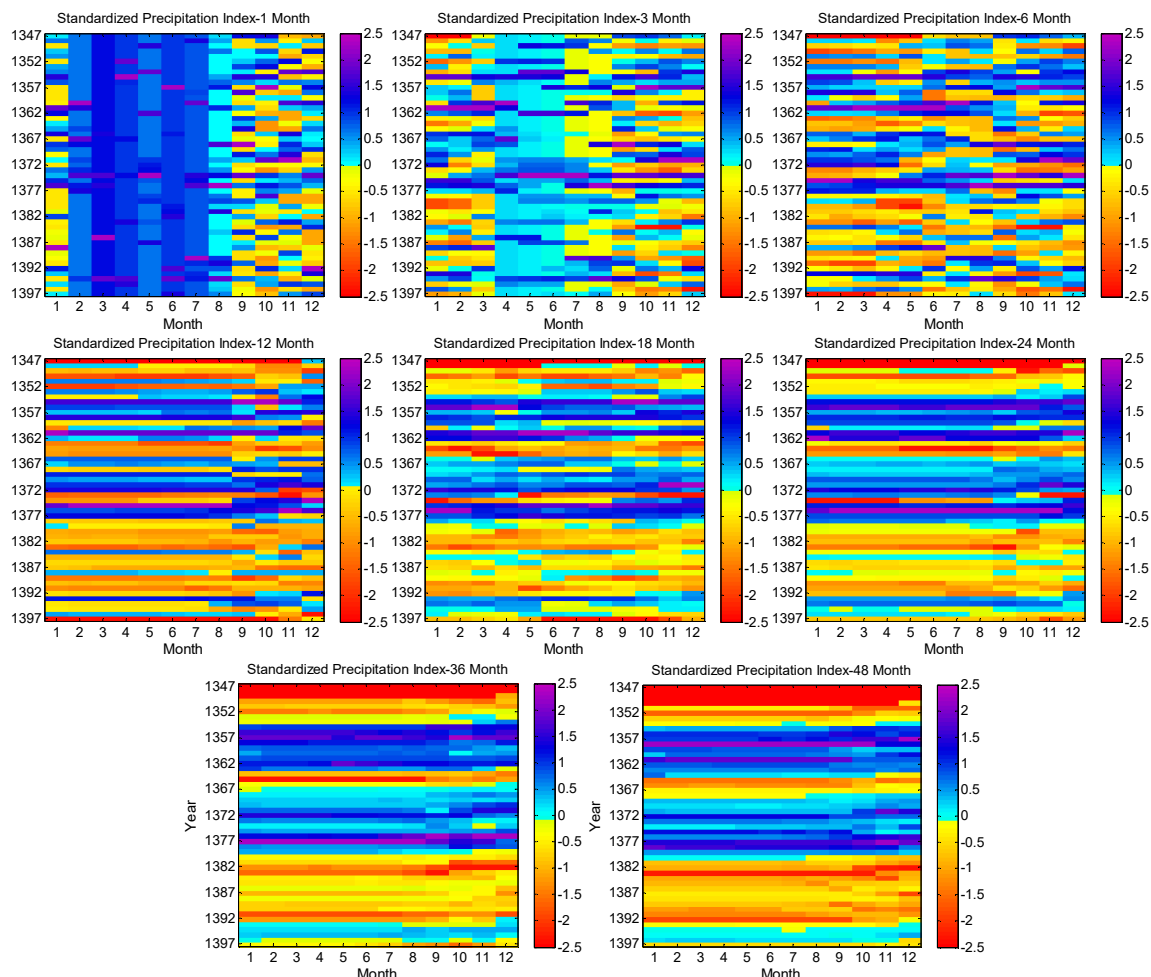


Figure 8. Standard precipitation index chart for 1, 3, 6, 12, 18, 24 and 36 months from 1347 to 1397

Table 5. Standard precipitation index based on drought classification for 612 months (51 years)

SPI	Normal	Moderately Wet	Extremely Wet	Severely Wet	Moderate Drought	Extreme Drought	Severe Drought
SPI-1	447	105	24	12	24	0	0
SPI-3	468	60	24	12	33	12	3
SPI-6	436	60	24	12	56	17	7
SPI-12	420	60	24	12	60	24	12
SPI-18	420	60	24	12	61	23	12
SPI-24	420	60	24	12	60	24	12
SPI-36	420	60	24	12	60	24	12
SPI-48	442	49	24	12	49	24	12

براساس شاخص SPI یک‌ماهه، تعداد ۴۴۷ ماه در وضعیت نرمال و پس از آن ۱۰۵ ماه در وضعیت مرطوب قرار دارند. همچنین، هیچ‌کدام از ماه‌ها در وضعیت خشکی شدید یا بسیار شدید قرار نگرفته است. در شاخص سه‌ماهه، تعداد

ماه‌های خشک تقریباً ۱۰۰ درصد افزایش یافته است. با افزایش دوره شاخص، شاهد افزایش تعداد ماه‌های قرار گرفته در شرایط خشک هستیم، به نحوی که در شاخص ۱۲ ماهه، مجموع تعداد ماه‌های طبقات مرطوب و خشک برابر می‌باشد. این وضعیت تقریباً در تمام ماه‌های بعد ثابت بوده و در نهایت، با وجود افزایش تعداد ماه‌های دارای وضعیت نرمال در دوره ۴۸ ماهه، نسبت ماه‌های خشک به مرطوب همچنان یکسان مانده است.

جدول (۶)، تعداد ماه‌های خشک در بازه‌های ده ساله را نمایش می‌دهد. بر این اساس، سال ۱۳۸۰ تا سال ۱۳۹۰ خشک‌ترین دهه بوده است، به نحوی که با افزایش مقیاس شاخص SPI تمام ماه‌های این دهه در وضعیت خشک قرار دارند. همچنین، نتایج حاکی از آن است که تعداد ماه‌های خشک از سال ۹۱ تا ۹۷ افزایش یافته است. دهه ۱۳۶۹ تا ۱۳۷۹ بهترین وضعیت از لحاظ شاخص SPI را داشته است. این موضوع در پژوهش Abadeh and Khosroshahi (2021) که به بررسی ارزیابی و پایش پدیده خشک‌سالی براساس دو شاخص SPI و SPEI در استان هرمزگان برای پنج ایستگاه سینوپتیک بندرعباس، بندر لنگه، حاجی‌آباد، میناب و بندر جاسک پرداخته است، تأیید می‌شود. نتایج این پژوهش نشان می‌دهد که براساس هر دو شاخص، با افزایش بازه‌های زمانی، فراوانی دوره‌های خشک کاهش یافته اما دوره‌ها دارای تداوم بیش‌تری می‌گردند.

Table 6. Number of dry months over periods of 10 years

SPI	1347-1357	1358-1368	1369-1379	1380-1390	1391-1397
SPI-1	21	25	23	37	19
SPI-3	43	50	38	61	31
SPI-6	47	62	47	95	47
SPI-12	42	61	49	107	41
SPI-18	57	48	52	110	26
SPI-24	58	40	46	115	27
SPI-36	57	44	5	120	50
SPI-48	47	46	0	120	59

۴. نتیجه‌گیری

در این مقاله به بررسی کارایی برخی از مدل‌های داده‌کاوی شامل ANN-MLP، ANN-RBF، CART، M5P و SVM در پیش‌بینی بارش یک سال بعد با استفاده از میانگین متحرک سه ساله داده‌های اقلیمی با هدف بررسی وضعیت خشک‌سالی بر مبنای شاخص بارش استاندارد پرداخته شد. در این راستا از ۵۰ سال آمار ایستگاه سینوپتیک بندرعباس بهره گرفته شد.

نتایج این مطالعه نشان داد که به‌جز مدل CART، سایر مدل‌های داده‌کاوی کارایی مطلوبی در پیش‌بینی بارش در این منطقه داشته‌اند. از طرفی تعدیل نوسانات شدید داده‌های ورودی از طریق استفاده از میانگین متحرک سه ساله آن‌ها، با نمایان نمودن روند سری زمانی، امکان آموزش و واسنجی بهتر مدل‌ها را فراهم کرده است. به این ترتیب با افزایش دقت در پیش‌بینی بارش، مطالعه دقیق‌تر وضعیت خشک‌سالی میسر شده است. بررسی پارامترهای آماری مختلف در مرحله صحت‌سنجی مدل‌ها نشان از توانمندی بالای مدل SVM در پیش‌بینی بارش سال بعد را دارد، به‌طوری‌که استفاده از این مدل در مقایسه با مدل ANN-MLP و M5P به‌ترتیب باعث بهبود ۵/۴ و ۴/۵ درصدی در شاخص NSE، ۵/۸ و ۴/۸ درصدی در شاخص RMSE و ۲/۴ و ۰/۶ درصدی در شاخص CC شده است. اختلاف عملکرد مدل SVM با دو مدل ANN-RBF و CART بیش‌تر و قابل توجه است. بر این اساس به‌نظر می‌رسد در انتخاب مدل برتر از میان سه مدل SVM، M5P و ANN-MLP باید به سایر ویژگی‌های آن‌ها نیز توجه داشت. چرا که نتیجه آموزش شبکه‌های عصبی مصنوعی ایجاد یک سری وزن‌های داخلی است که در شبکه توزیع می‌شوند. این وزن‌ها امکان مناسبی از چگونگی تحلیل و پیش‌بینی ایجاد نمی‌کنند و مانند این است

که از افراد خبره و متخصص پرسش شود که چرا چنین تصمیمی درست است و آن‌ها بگویند که فقط می‌دانند که درست است. در واقع فهم این وزن‌ها کار آسانی نیست، بنابراین به‌نظر می‌رسد شبکه‌های عصبی مصنوعی بیش‌تر برای مواقعی مناسب هستند که نیازی به شناسایی چگونگی رخ‌دادن اتفاقات نباشد و تنها نتیجه و خروجی، اهمیت داشته باشد. مسئله مهم دیگر این است که این مدل‌ها برای انجام تخمین‌های با دقت بالا از پارامترهای وابسته، نیازمند تنظیم پارامترهای فراوانی هستند که در اکثر موارد باعث به‌وجودآمدن فرایند سعی و خطا در رسیدن به جواب بهینه می‌شود. در مقابل در روش ماشین‌های بردار پشتیبان تعداد پارامترهای کم‌تری برای اجرای مدل نسبت به شبکه‌های عصبی مصنوعی وجود دارد. لذا به سعی و خطای کم‌تری جهت رسیدن به دقت موردنظر نیاز خواهند داشت، اما هم‌چنان مانند شبکه‌های عصبی مصنوعی نمایش مناسبی از چگونگی تحلیل و پیش‌بینی ارائه نمی‌دهند. این در حالی است که مدل‌های درختی برخلاف شبکه‌های عصبی مصنوعی و ماشین‌های بردار پشتیبان، به‌راحتی قابل توضیح و تفسیر می‌باشند. چرا که پیش‌بینی درخت در قالب یک سری قوانین و روابط کاربردی و مفید ارائه می‌شود.

نتایج این مقاله نشان می‌دهد که هر سه دسته ابزارهای داده‌کاوی شامل شبکه‌های عصبی مصنوعی، درختان تصمیم و ماشین‌های بردار پشتیبان قادر به شبیه‌سازی و پیش‌بینی وضعیت بارش و خشک‌سالی ماه‌های سال بعد براساس اطلاعات میانگین متحرک سه ساله داده‌های اقلیمی منطقه مورد مطالعه با دقت قابل قبولی هستند. ضمن این‌که تغییر در ابزار داده‌کاوی در اغلب موارد تفاوت چشم‌گیری در دقت نتایج ایجاد نمی‌کند.

تحلیل وضعیت خشک‌سالی از سال ۱۳۴۷ تا ۱۳۹۷ با استفاده از جعبه ابزار تجزیه و تحلیل خشک‌سالی استاندارد نشان داد که تفاوت قابل‌توجهی در نتایج برای دوره‌های زمانی مشخص بارش در شاخص ایجاد می‌شود. با افزایش طول دوره SPI، میزان تفکیک سال‌های مرطوب و خشک مشخص‌تر شده و در نهایت می‌توان این‌گونه بیان نمود که در حدود ۶۰ درصد بازه زمانی وضعیت نرمال حاکم بوده و شرایط خشک و مرطوب نیز دارای سهم یکسان (حدود ۲۰ درصد) هستند. البته، سال ۱۳۹۷ براساس شاخص‌های با دوره کوتاه‌مدت، بیش‌تر در شرایط نرمال و مرطوب قرار داشته؛ در صورتی‌که بر مبنای دوره‌های بارش طولانی‌تر در شرایط خشک قرار می‌گیرد. تحلیل وضعیت خشک‌سالی در بازه‌های ده‌ساله نیز حاکی از تداوم وضعیت خشکی طولانی‌مدت در دو دهه اخیر است.

۵. پی‌نوشت‌ها

1. Standard Precipitation Index
2. Multilayer Perceptron Artificial Neural Network
3. Support Vector Machine
4. Classification And Regression Trees
5. Multivariate Adaptive Regression Splines
6. Least Square Support Vector Machine
7. Artificial Neural Network
8. Adaptive Neuro Fuzzy Inference System
9. Autoregressive Integrated Moving Average
10. Wavelet Artificial Neural Networks
11. Multiple Linear Regression
12. Discrepancy Ratio

۶. تعارض منافع

هیچ‌گونه تعارض منافی توسط نویسندگان وجود ندارد.

۷. منابع

- Abadeh, M., & Khosroshahi, M. (2021). Assessment and drought monitoring using Standardized Precipitation (SPI) and Standardized Precipitation Evapotranspiration (SPEI) Indices in Hormozgan province., *Iranian Journal of Rangeland and Desert Research*, 28(4), 718-732. (In Persian).
- Anwar, M. T., Hadikurniawati, W., Winarno, E., & Widiyatmoko, W. (2020, December). Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia. In 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 83-88). IEEE.
- Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., ... & Overton, I. C. (2016). Drought indicators revisited: the need for a wider consideration of environment and society, *Wiley Interdisciplinary Reviews: Water*, 3(4), 516-536.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers, In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152).
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees, Routledge.
- Dehbozorgi, M., Malekian, A., & Ehsani, A.H. (2015). Evaluation the efficiency of using artificial neural networks in predicting meteorological droughts in north-west of Iran. *Journal of Geographical Sciences*, 15(36), 139-156. (In Persian).
- Deo, R. C., Kisi, O., & Singh, V. P. (2017). Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model, *Atmospheric Research*, 184, 149-175.
- Farahmand, A., & AghaKouchak, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators, *Advances in Water Resources*, 76, 140-145.
- Ghorbani, K., Salari Jazi, M., & Abdolhosseini, M. (2015). Feasibility study of the prediction of annual drought based on drought conditions in the spring season. *Iranian Journal of Irrigation & Drainage*, 9(4), 636-645. (In Persian).
- Gringorten, I. I. (1963). A plotting rule for extreme probability paper, *Journal of Geophysical Research*, 68(3), 813-814.
- Haykin, S. (2004). Neural networks: A comprehensive foundation, 2, 41.
- Jahani, B., & Mohammadi, B. (2019). A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran, *Theoretical and Applied Climatology*, 137(1), 1257-1269.
- Karami, A. (2010). Estimation of the critical clearing time using MLP and RBF neural networks, *European Transactions on Electrical Power*, 20(2), 206-217.
- Khoshhal Dastjerdi, J., & Hosseini, S.M. (2010). Application of artificial neural network in climatic elements simulation and drought cycle predication (case study: Isfahan province). *Geography and Environmental Planning*, 21(3 (39)), 107-120. (In Persian).
- Loukas, A., Vasiliades, L., & Tzabiras, J. (2007). Evaluation of climate change on drought impulses in Thessaly, Greece, *European Water*, 17(18), 17-28.
- Luenberger, D. G., & Ye, Y. (1984). Linear and nonlinear programming, (Vol. 2). Reading, MA: Addison-wesley.
- Malik, A., Kumar, A., Rai, P., & Kuriqi, A. (2021). Prediction of multi-scalar standardized precipitation index by using artificial intelligence and regression models, *Climate*, 9(2), 28.
- Mitchell, T. M. (1997). Machine learning. 1997. Burr Ridge, IL: McGraw Hill, 45(37), 870-877.
- Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., ... & Rodrigo-Comino, J. (2021). Estimation of SPEI meteorological drought using machine learning algorithms, *IEEE Access*, 9, 65503-65523.

- Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., & Arabasadi, A. (2017). Drought forecasting by ANN, ANFIS, and SVM and comparison of the models, *Environmental Earth Sciences*, 76(21), 1-10.
- Mozafari, Gh., Shafie, Sh., & Taghizade, Z. (2016). Evaluate the performance regression decision tree model in predicting drought (case study: synoptic station in Sanandaj). *Journal of Natural Environment Hazards*, 4(6), 1-19. (In Persian).
- Murad, S. H., & Salih, Y. M. M. (2020). Comparable Investigation for Rainfall Forecasting using Different Data Mining Approaches in Sulaymaniyah City in Iraq. *International Journal of Advances in Life Science and Technology*, 4(1), 11-18.
- Nikbakht Shahbazi, A., Zahraie, B., & Nasser, M. (2012). Seasonal meteorological drought prediction using support vector machine. *Journal of Water and Wastewater*, 23(2), 73-85. (In Persian).
- Nivedika, M., Meghwal, M., & PV, R. (2021). Forecasting Drought via Soft-Computation Multi-layer Perceptron Artificial Intelligence Model, *International Research Journal on Advanced Science Hub*, 3, 30-36.
- Quinlan, J.R. (1992) Learning with Continuous Classes, Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart 16-18 November 1992, 343-348.
- Roodposhti, M. S., Safarrad, T., & Shahabi, H. (2017). Drought sensitivity mapping using two one-class support vector machine algorithms, *Atmospheric Research*, 193, 73-82.
- Smola, A. J. (1996). Regression estimation with support vector learning machines, Doctoral dissertation, Master's thesis, Technische Universität München.
- Vapnik Vladimir, N. (2000). The nature of statistical learning theory, Second Edition. Springer.
- Wu, J. (2018, August). Co-Evolution Algorithm for Parameter Optimization of RBF Neural Networks for Rainfall-Runoff Forecasting, In International Conference on Intelligent Computing (pp. 195-206). Springer, Cham.
- Zainudin, S., Jasim, D. S., & Bakar, A. A. (2016). Comparative analysis of data mining techniques for Malaysian rainfall prediction. *International Journal on Advanced Science Engineering Information Technology*, 6(6), 1148-1153.
- Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2020). Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China, *Natural Resources Research*, 29(2), 1447-1464.