نشریه تحقیقات کاربردی علوم جغرافیایی سال چهاردهم، شماره ۳۴، پاییز ۹۳

ارزیابی کارایی مدل درخت تصمیم در پیش بینی بارش ایستگاه سینوپتیک کرمانشاه

دریافت مقاله:۹۲/۱۰/۲ پذیرش نهایی: ۹۳/۴/۱۱
صفحات:۱۱۰-۸۹

کمال امیدوار: استاد آب و هواشناسی، دانشگاه یزد[۱]
Email: komidvar@yazd.ac.ir
شهاب شفیعی: دانشجوی دکتری آب و هواشناسی، دانشگاه یزد
Email: shafieshahab@gmail.com
زهرا تقی زاده: دانشجوی دکتری آب و هواشناسی، دانشگاه یزد
Email:ztaghizade95@stu.yazd.ac.ir
مهرداد علی پور: دانشجوی دکتری صنایع، دانشگاه یزد
Email: mhrdd_alipoor@yahoo.com

چکیده

بارش یکی از اجزای اصلی بیلان منابع آب بوده و پیش بینی آن می‌تواند در مدیریت تامین آب کشاورزی مدیریت منابع آب موجود در مخازن سدها و ... مفید باشد. درخت تصمیم به عنوان یکی از مدل‌های پیش بینی، کارایی زیادی در این زمینه دارد و به تولید قانون می‌انجامد. در این پژوهش جهت رسیدن به اهداف از ارزیابی کارایی مدل درخت تصمیم جهت پیش بینی بارش در ایستگاه سینوپتیک کرمانشاه و از الگوریتم CART (Classification And Regression Tree) به عنوان یکی از انواع درختان تصمیم رگرسیونی جهت پیش بینی بارش ۳۰ ماه بعد استفاده شده است. داده‌های مورد استفاده این پژوهش مربوط به آمار ماهیانه بارندگی، تبخیر، رطوبت نسبی، دمای ماکزیمم، دمای متوسط و سرعت باد در دوره آماری(۱۳۸۹- ۱۳۴۹) می‌باشد. سپس جهت ارزیابی درخت‌های ایجاد شده در این پژوهش از معیارهای آماری مختلف استفاده شده است که در نهایت نتایج نشان می‌دهد در ایستگاه سینوپتیک کرمانشاه درخت تصمیم گیری رگرسیونی، مدلی نسبتاً کارا در پیش بینی بارش می‌باشد که استفاده از میانگین متحرّک نسبت به سایر حالات منجر به افزایش چشمگیر کارایی مدل درخت تصمیم می‌شود و در صورت تعدیل دامنه تغییرات داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان بارش را ۳۰ ماه قبل از وقوع برآورد نماید که در شبیه سازی‌های صورت گرفته، زمانی که از میانگین متحرّک پنج ساله داده‌ها برای اجرای مدل استفاده گردیده، ترکیب بارش قبلی، دمای ماکزیمم به عنوان مناسب‌ترین حالت شناسایی شده است.

کلید واژگان: درخت تصمیم، پیش بینی، بارش، الگوریتم CART، ایستگاه سینوپتیک کرمانشاه

---

[۱]. **نویسنده مسئول:** یزد- صفاییه – بلواردانشگاه- دانشکده علوم انسانی- گروه جغرافیا

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

مقدمه

بارش از مهم‌ترین عناصر اقلیمی می‌باشد که در تعیین و پراکندگی دیگر عناصر اقلیمی نیز موثر است(عساکره، ۱۳۸۶: ۵). به عبارتی بارش به عنوان یکی از عناصر مهم هواشناسی، مطرح می‌باشد که شناخت لازم از میزان این عنصر، تغییرات و پیش بینی آن، از یک سو به جهت داشتن برنامه ریزی دقیق‌تر در مدیریت بخش‌های کشاورزی، اقتصادی و اجتماعی ضروری می-نماید(عرفانیان و همکاران، ۱۳۹۲: ۵۴). و از سوی دیگر مطالعه و اندازه گیری آن در اکثر موارد برای مطالعات رواناب، خشکسالی، آب‌های زیرزمینی، سیلاب، رسوب و ... لازم و ضروری است(حلبیان، ۱۳۸۸: ۷). در عین حال در مناطق شهری پیش بینی بارش تأثیر بسیار زیادی بر کنترل ترافیک، فاضلاب و فعالیت‌های دیگر انسان دارد. اما اساساً پیش بینی بارش به سبب رفتار بسیار نامنظم و آشوب‌مندی که از خود نشان می‌دهد؛ دشوار، پیچیده و در عین حال حائز اهمیت است(حلبیان و دارند، ۱۳۹۱: ۵۹- ۴۸) به طوریکه در طی چند دهه اخیر یکی از تحقیقات جالب توجه محافل علمی، مطالعه رفتار بارش در سطح محلی، ناحیه‌ای، منطقه‌ای و جهانی بوده است(پور اصغری و همکاران، ۱۳۹۲: ۲۸). کشور ایران نیز به علت قرار گرفتن در کمربند خشک جغرافیایی جهان، در زمره مناطق کم باران جهان محسوب می‌گردد. علاوه بر آن نوسانات شدید بارش در مناطق مختلف کشور، وقوع خشکسالی‌های ضعیف تا شدید را به موضوعی گریز ناپذیر تبدیل نموده است. بر اساس مطالعات مختلف کمبود بارش نسبت به میانگین دراز مدت یک منطقه به عنوان مؤلفه اصلی رخداد خشکسالی محسوب می‌گردد(دستورانی و همکاران،۱۳۹۱: ۱۵). و مطالعه اقلیم شناسی بارش ضمن شناخت رفتار بارش، امکان برنامه ریزی مبتنی بر آن را مهیا می‌سازد. این گونه توجه به بارش به ویژه در نواحی مختلف سرزمین کم آبی نظیر ایران، که منابع آب آن متکی به بارش می‌باشد لازم و ضروری است. در این راستا برنامه ریزی‌های صحیح و کسب اطلاعات از طریق مطالعات علمی ضروری می‌باشد(فیروزی و همکاران، ۱۳۹۱: ۷۸). اما با توجه به اینکه فرایند بارش، فرایندی کاملاً غیرخطی و از نظر زمانی و مکانی تصادفی و عوامل مختلف اقلیمی مانند شکل، شدت، مدت، دوره بارش، پراکنش شدت در مدت بارش، تبخیر و غیره در آن نقش دارند و تشریح آن با مدل‌های ساده و خطی به راحتی امکان پذیر نیست و با خطای زیادی همراه است(امیدوار و اژدرپور، ۱۳۹۱: ۱۴۰). با توجه به اهمیت این موضوع در سال‌های گذشته روش‌ها و مدل‌های گوناگونی جهت ارزیابی، پیش بینی و تخمین بارندگی ارائه شده است. از جمله این روش‌ها می‌توان به روش شبکه عصبی مصنوعی و روش نوین درختان تصمیم گیری(مدل‌های درختی)

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

اشاره کرد. درختان تصمیم[1] به همراه قوانین تصمیم یکی از روش‌های داده کاوی به شمار می-آیند(کانتاردزیک، ۱۳۸۵: ۱۱۲). درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته بندی و پیش بینی می‌باشد. مدل درخت تصمیم بر خلاف مدل شبکه عصبی مصنوعی به تولید قانون می‌پردازد. در ساختار درخت تصمیم، پیش بینی به دست آمده در قالب یک سری قواعد توضیح داده خواهد شد. همچنین در درخت تصمیم بر خلاف شبکه عصبی مصنوعی، ضرورتی وجود ندارد که داده‌ها لزوماً به صورت عددی باشند(مشکانی و ناظمی، ۱۳۸۸: ۷۳). درختان تصمیم ابزاری هستند که قابلیت پاسخگویی به مسائل پیچیده و غیر خطی را دارند و از آنجایی که در علوم دیگر از جمله برق و الکترونیک که اساسا دارای مسائل غیر خطی هستند به خوبی از عهده حل مسائل بر آمده‌اند، لذا کم کم پای این ابزار جدید به علوم مهندسی آب و منابع طبیعی که جزء پدیده های غیر خطی و پیچیده هستند نیز باز شده است(طالبی و اکبری، ۱۳۹۲: ۱۱۰). از جمله تحقیقات صورت گرفته بر روی کارایی درخت تصمیم در زمینه علوم طبیعی می‌توان به موارد ذیل اشاره نمود.

ماهش و مادر(۲۰۰۳) از مجموعه داده‌های آموزش و آزمون در دو منطقه جغرافیایی و با دو سنجنده مختلف جهت ارزیابی توانایی درخت تصمیم تک متغیره و چندمتغیره در طبقه بندی پوشش زمین استفاده نموده‌اند نتایج نشان داد که درخت تصمیم تک متغیره در مقایسه با دیگر طبقه بندی کننده‌ها، به جز برای داده‌های با ابعاد بالا، به طور قابل قبولی بهتر اجرا شده است. باتاچاریا و همکاران(۲۰۰۷) از دو روش یادگیری ماشینی شامل شبکه عصبی مصنوعی و مدل-های درخت تصمیم، جهت مدل سازی انتقال بار کل و بار کف رودخانه استفاده نموده‌اند. طبق نتایجی که آنها گرفتند مدل‌های یادگیری ماشینی دقت بیشتری نسبت به دیگر روش‌ها دارد و از بین مدل‌های مذکور، مدل درخت تصمیم کمترین خطا را نشان داده است.

گیسن و همکاران(۲۰۰۷) فرسایش آبی سطحی و زیرسطحی را با مدل درخت تصمیم در تاباسکو (جنوب شرقی مکزیک) مورد ارزیابی قرار داده‌اند. نتایج نشان داده است که این مدل وقوع فرسایش‌های تونلی و گودالی را با موفقیت پیش بینی می‌کند. محجوبی و اعتماد شهیدی(۲۰۰۷) به تخمین ارتفاع امواج ناشی از باد در نکاء به کمک درختان تصمیم رگرسیونی مصنوع پرداخته‌اند. نتایج حاکی از دقت بیشتر مدل‌های درختی نسبت به شبکه‌های عصبی بوده و نشان دهنده نیرومند بودن این مدل‌ها می‌باشند.

[1] Decision Trees

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

روس جان و میکوس(۲۰۰۸) الگوریتم درخت تصمیم رگرسیونی M5 برای مشاهده نقش متقابل وضعیت‌های فصلی و هیدرولوژیکی که آب شویی نیترات را در طول وقایع هیدرولوژیکی در یک حوزه جنگلی واقع در جنوب غربی اسلوانی کنترل می‌کند استفاده کردند. کوسو و همکاران(۲۰۰۹) از درختان تصمیم تک منظوره و چندمنظوره برای پیش بینی وضعیت یا کیفیت گیاهان بومی باقی مانده در منطقه وسیعی در شمال شرق استرالیا -ایالت ویکتوریا استفاده کرده‌اند. بررسی نتایج این مطالعه نشان داده است که مدلسازی چند منظوره بر تک منظوره برتری دارد. ایوب لو و همکاران(۲۰۱۰) از مدل درخت تصمیم و شبکه عصبی مصنوعی برای تخمین عمق آب شستگی پایپل‌ها استفاده نموده‌اند. آن‌ها از دو مدل مزبور برای پیش بینی عمق کنش نرمال شده به عنوان یک تابع از دو سری جداگانه پارامترها (دارای بعد و بدون بعد) استفاده نموده‌اند آن‌ها بر اساس نتایج تحقیق خود یک مدل ترکیبی مبتنی بر مفاهیم ژئومتریک ارائه نموده‌اند. باتاچاریا و همکاران (۲۰۰۷) مدل درختان تصمیم را با چندین معادله برای برآورد رسوب در کف رودخانه استفاده نموده‌اند و نتیجه گرفتند که مدل درخت تصمیم خطای بسیار کمتری را نسبت به مدل‌های تجربی دارند. یوریک لی و همکاران (۲۰۱۲) از تکنیک درخت تصمیم جهت ارزیابی و پیش بینی طبقات خشکسالی در حوزه‌ای در ترکیه استفاده کرده و نتایج حاصله را با روش طبقه بندی و تحلیل خشکسالی بر اساس شاخص بارش استاندارد (SPI) مقایسه نموده‌اند. مقایسه نتایج طبقه بندی خشکسالی بر اساس این دو روش تفاوت قابل ملاحظه‌ای نشان نداده است. از جمله مطالعات داخلی در این زمینه نیز می‌توان به موارد ذیل اشاره نمود. محجوبی و صمغی(۱۳۸۸)، به پیش بینی پارامترهای امواج ناشی از باد در دریای خزر با استفاده از روش درختان تصمیم رگرسیونی و شبکه‌های عصبی مصنوعی پرداخته‌اند نتایج حاصل از درخت های تصمیم گیری با روش شبکه‌های عصبی مصنوعی مورد مقایسه قرار گرفته است. نتایج حاکی از نزدیکی دقت دو روش در پیش بینی پارامترهای امواج ناشی از باد می‌باشد .همچنین سرعت مدل سازی و اجرای درخت های تصمیم گیری به مراتب بیشتر از شبکه‌های عصبی می‌باشد. طالبی و اکبری(۱۳۹۲)، به بررسی کارایی مدل درخت تصمیم گیری در برآورد رسوبات معلق رودخانه‌ای در حوضه سد ایلام پرداخته‌اند. نتایج نشان داده است به طور کلی و براساس تمامی معیارهای آماری ذکر شده در پژوهش مدل درخت تصمیم گیری در مقایسه با روش منحنی سنجه رسوب تطابق بیشتری با مقادیر اندازه گیری شده داشته است. اما برابر بررسی‌های انجام شده، تاکنون تحقیقات خیلی کمی در زمینه پیش بینی بارش به عنوان یکی از پارامترهای اقلیمی با استفاده از مدل درخت‌های تصمیم گیری صورت گرفته است که از جمله این مطالعات می‌توان به مطالعات شایق(۱۳۹۰) از مدل درخت

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

تصمیم رگرسیونی جهت ارزیابی پروژه‌های باروری ابرها در استان فارس استفاده نموده‌اند. دستورانی و همکاران(۱۳۹۱) به بررسی کارایی مدل درخت تصمیم در پیش بینی بارش ایستگاه سینوپتیک یزد پرداخته‌اند. در این تحقیق از مدل درخت تصمیم به عنوان یکی از روش‌های داده کاوی جهت پیش بینی بارش و ارزیابی وضعیت خشکسالی درایستگاه سینوپتیک یزد استفاده شد. نتایج نشان داده است که در ایستگاه یزد، مدل درخت تصمیم گیری خصوصاً در شرایطی که از میانگین متحرک ۵ ساله داده‌ها استفاده گردد، دارای توانایی مناسبی در پیش بینی میزان بارش می‌باشد.

به طور کلی نتایج این تحقیقات به طور عمده مدل درخت تصمیم را به عنوان ابزاری دقیق و قابل اعتماد معرفی می‌کنند بنابراین استفاده از این روش به عنوان روشی نو و ابزاری قدرتمند و قابل اعتماد می‌تواند یاریگر مسئولان جهت استفاده از منابع کشور باشد. هدف پژوهش حاضر نیز ارزیابی کارایی مدل درخت تصمیم در پیش بینی بارش به عنوان یکی از داده‌های غیر خطی در ایستگاه سینوپتیک کرمانشاه می‌باشد که برای رسیدن به اهداف پژوهش از آمار و داده‌های ایستگاه سینوپتیک کرمانشاه استفاده شده است.

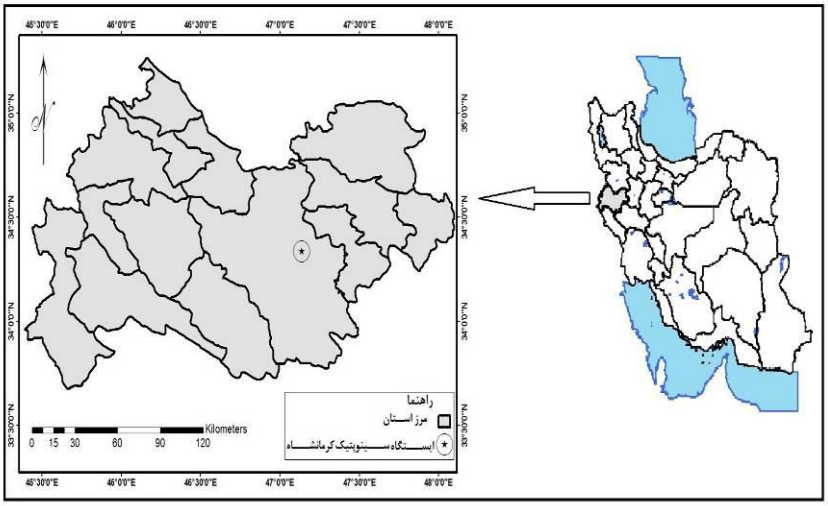
داده‌ها و روش شناسی پژوهش

معرفی منطقه مورد مطالعه

ایستگاه سینوپتیک کرمانشاه دارای بلندترین طول دوره آماری در بین ایستگاه‌های استان می‌باشد. داده‌های مورد استفاده در این تحقیق مربوط به آمار ماهیانه بارندگی، تبخیر، رطوبت نسبی، دمای ماکزیمم، دمای متوسط و سرعت باد در دوره آماری(۱۳۸۹- ۱۳۴۹) در یک دوره ۴۰ ساله در ایستگاه سینوپتیک کرمانشاه می‌باشد. جهت کنترل کیفیت آمار و اطلاعات موجود از آزمون توالی استفاده شده است. نتایج نشان داد که داده‌های مزبور از همگنی لازم برخوردار می‌باشند و سپس از نرم افزار مت لب[1] جهت تحلیل و بررسی استفاده شده است. جدول(۱) و شکل(۱) موقعیت و مشخصات ایستگاه سینوپتیک کرمانشاه را نشان می‌دهد.

[1]. MATLAB

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]



شکل (۱) موقعیت جغرافیایی ایستگاه سینو پتیک کرمانشاه

جدول(۱) مشخصات ایستگاه سینوپتیک کرمانشاه

ایستگاه	دوره آماری	ارتفاع	عرض جغرافیایی	طول جغرافیایی
کرمانشاه	۱۳۴۹-۱۳۸۹	۱۳۱۸/۶	۳۴° ۲۱ ′	۴۷° ۹ ′

روش پژوهش

درختان تصمیم گیری از نسل جدید تکنیک‌های داده کاوی به شمار می‌آیند که در دو دهه اخیر توسعه زیادی یافته‌اند. از این تکنیک‌ها هم می‌توان برای کشف و استخراج دانش از یک پایگاه داده‌ها و هم برای ایجاد مدل‌های پیش بینی استفاده نمود Kazeminezhad et al, 2005: (1711). درخت تصمیم یکی از روش های داده کاوی و ابزارهای قوی و متداول برای دسته بندی و پیش بینی می باشد که برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد(دستورانی و همکاران، ۱۳۹۱: ۱۸). یعنی درخت تصمیم پیش بینی خود را در قالب یکسری قوانین توضیح می‌دهد(شکل: ۲). در حالی که در شبکه‌های عصبی تنها پیش بینی بیان می‌شود و چگونگی آن در خود شبکه پنهان باقی می‌ماند. علاوه بر آن در درخت تصمیم گیری برخلاف شبکه‌های عصبی، می‌توان از داده‌های غیر عددی نیز استفاده نمود به عبارتی این ساختار تصمیم گیری می‌تواند به شکل تکنیک‌های ریاضی و محاسباتی که به توصیف، دسته بندی و عام سازی یک مجموعه از داده‌ها کمک می‌کنند نیز معرفی شوند (Hand et al, 2001). درختان تصمیم گیری به دو نوع درختان طبقه بندی شده و رگرسیونی تقسیم می‌شوند به طوری که اگر متغیر پاسخ

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

مقداری گسسته داشته باشد به آن درخت طبقه بندی شده و زمانی که درخت مقادیر پیوسته را پیش بینی کند نوع رگرسیونی را به آن نسبت می‌دهند (طالبی و اکبری،۱۳۹۲: ۱۱۱). در این پژوهش از الگوریتم داده‌کاوی درخت تصمیم رگرسیونی[1] به عنوان یکی از انواع درختان تصمیم جهت پیش بینی بارش ۳۰ ماه قبل از وقوع استفاده شده است. ساختار این درختان بر سه اصل استوار است:

۱) مجموعه‌ای از سوالات به شکل $x \leq d$? که در آن x یک متغییر مستقل و d یک مقدار ثابت است و جواب هر سوال بله / خیر است.

۲) بهترین معیار شاخه زدن جهت انتخاب بهترین متغیر مستقل برای ایجاد شاخه جدید.

۳) ایجاد آمار خلاصه برای گره انتهایی(Breiman et al: 1984).

معیارهای متفاوتی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، اما از آنجا که پژوهش حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل که انحراف حداقل مربعات[2](LSD) نام دارد که این معیار به صورت زیر تعریف می‌شود:

$$SS(t) = \sum_{i=1}^{N_{tt}} \left(y_i(t) - \bar{y}(t)\right)^2 \quad (۱)$$

Ntt: تعداد رکوردها (داده‌ها) در گره برگ t.

Yi (t): مقدار خروجی (متغیر هدف در گره برگ).

$\bar{y}(t)$: میانگین مقادیر متغیر هدف برای همه گره‌ها

حال متغیر ورودی S زمانی بهترین متغیر برای ایجاد شاخه در گره t می‌باشد که مقدار Q(s,t) را بیشینه نماید.

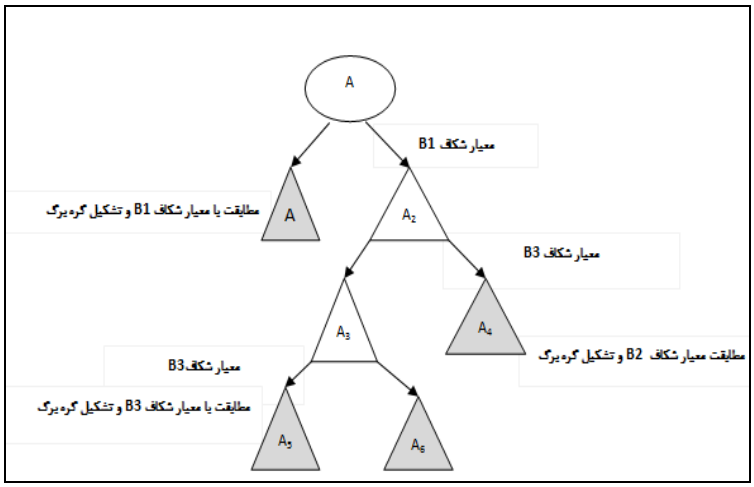
$$Q(s,t) = ss(t) - ss(t\ R) - SS(t\ L) \quad (۲)$$

که در آن ss(t R) و SS(t L) به ترتیب میزان ss(t) در شاخه سمت راست و سمت چپ گره t می‌باشد.

[1]. Classification and regression tree

[2] Least squares deviation

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]



شکل(۲) نمونه‌ی از یک درخت تصمیم ساده

تقسیم بندی داده‌ها برای مدل درخت تصمیم در ایستگاه سینوپتیک کرمانشاه

جهت اجرای مدل درختان تصمیم از آمار و اطلاعات مربوط به برخی عناصر اقلیمی در مقیاس ماهانه طی سال‌های(۱۳۴۹-۱۳۸۹) در ایستگاه سینوپتیک کرمانشاه استفاده گردیده است. در این فرایند داده‌ها به دو دسته تقسیم شده‌اند، داده‌های آموزشی[1] و داده‌های آزمایشی[2] به طوری‌که در این پژوهش ۷۰ ٪ از کل داده‌ها به آموزش مدل تعلق دارد و ۳۰ ٪ باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردیده است. انتخاب داده‌های آموزش و آزمون به صورت سیستماتیک و توسط کاربر انجام شده است و بدیهی است که داده‌های آزمون در مرحله آموزش مورد استفاده قرار نگرفته‌اند.

عناصر مورد استفاده در شبیه سازی

در شبیه سازی با استفاده از مدل درخت تصمیم گیری از عناصر مختلفی استفاده شده است. این عناصر به عنوان متغیر مستقل به مدل معرفی و شبیه سازی‌ها برای پیش بینی متغیر هدف صورت گرفته است. متغیرهای مستقل مورد استفاده در این تحقیق و علائم اختصاری آن‌ها در جدول(۲) نمایش داده شده است.

1. Training data
2. Testing data

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

جدول(۲) معرفی عناصر ورودی مدل (دوره آماری ۱۳۴۹ تا ۱۳۸۹)

میانگین	حداقل	حداکثر	عناصر ورودی	علامت اختصاری
۳۱/۷	۰	۱۵۳	بارش(میلی متر)	P
۱۴/۴	-۱/۷	۲۶	دمای متوسط (درجه سانتیگراد)	T_{mean}
۲۲/۶	-۱/۲	۲۲/۶	دمای حداکثر (درجه سانتیگراد)	T_{max}
۴۶	۱۸	۷۷	رطوبت نسبی (درصد)	H
۳/۷	۱/۹	۶/۵	سرعت باد غالب (متر بر ثانیه)	W_S

حالات مختلف در اجرای مدل

الف) استفاده از اصل داده‌ها بدون نرمال سازی: در این حالت بر روی داده‌ها هیچ گونه عملیات تغییر مقیاس انجام نشد. ورودی و خروجی همان داده‌های خام بودند که به مدل معرفی شده-اند. به عبارت دیگر در این حالت، عناصر درج شده در جدول(۲) به عنوان ورودی در اختیار مدل قرار گرفته و بارش ۲/۵ سال بعد پیش بینی گردیده است.

ب) استفاده از میانگین متحرک سه ساله داده‌های خام: بعد از انجام چندین شبیه سازی جهت بالا بردن ضریب کارایی در پیش بینی بارش، به جای داده‌های ماهانه از میانگین متحرک سه ساله استفاده شد. به این ترتیب میانگین متحرک سه ساله برای کل عناصر ورودی مدل محاسبه و با در نظر گرفتن این عناصر به عنوان متغیر مستقل، میانگین متحرک سه ساله بارش برای یک سال بعد پیش بینی می‌گردد.

ج) استفاده از میانگین متحرک سه ساله نرمال شده: در این حالت میانگین متحرک سه ساله داده‌ها، با استفاده از رابطه ۳ نرمال شد و سپس درخت تصمیم برای انجام پیش بینی ایجاد و مورد ارزیابی قرار گرفته است.

$$Xn = \left(\frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \right) \qquad (۳$$

X0 و Xn به ترتیب داده‌های نرمال شده و داده‌های اولیه را نشان می‌دهد.

Xmin ، Xmax به ترتیب مینیمم و ماکزیمم داده‌های اولیه را نشان می‌دهد.

در این حالت میانگین متحرک سه ساله نرمال شده داده‌ها، به عنوان ورودی به مدل معرفی شده و میانگین متحرک سه ساله نرمال شده است بارش برای یک سال بعد پیش بینی گردیده است.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

د) استفاده از میانگین متحرک پنج ساله داده‌های خام: در این حالت میانگین متحرک پنج ساله برای کل عناصر ورودی مدل محاسبه شد و با در نظر گرفتن این پارامترها به عنوان متغیر مستقل، میانگین متحرک پنج ساله بارش برای یک سال بعد پیش بینی گردیده است.

ارزیابی کارایی مدل درخت تصمیم

جهت ارزیابی کارایی مدل درختان تصمیم در این پژوهش از چهار شاخص آماری به شرح ذیل استفاده گردیده است.

$$\mathrm{RMSE}=\left[\frac{\sum_{i-1}^{n}(p_i-o_i)^2}{n}\right]^{1/2} \quad (۴)$$

$$\mathrm{MAE} = \frac{\sum|o_i-p_i|}{N} \quad (۵)$$

$$\mathrm{Bias} = \overline{P}-\overline{O} \quad (۶)$$

$$\mathrm{R}=\sqrt{1-\frac{\sum_{i-1}^{n}(o_{i-}p_i)}{\sum_{i-1}^{n}(o_i-\overline{o})}} \quad (۷)$$

در این روابط Pi و Oi به ترتیب مقادیر مشاهده شده و شبیه سازی شده، $\overline{o}$ و $\overline{p}$ میانگین مقادیر مشاهده شده و شبیه سازی شده و N تعداد نمونه‌ها می‌باشد.

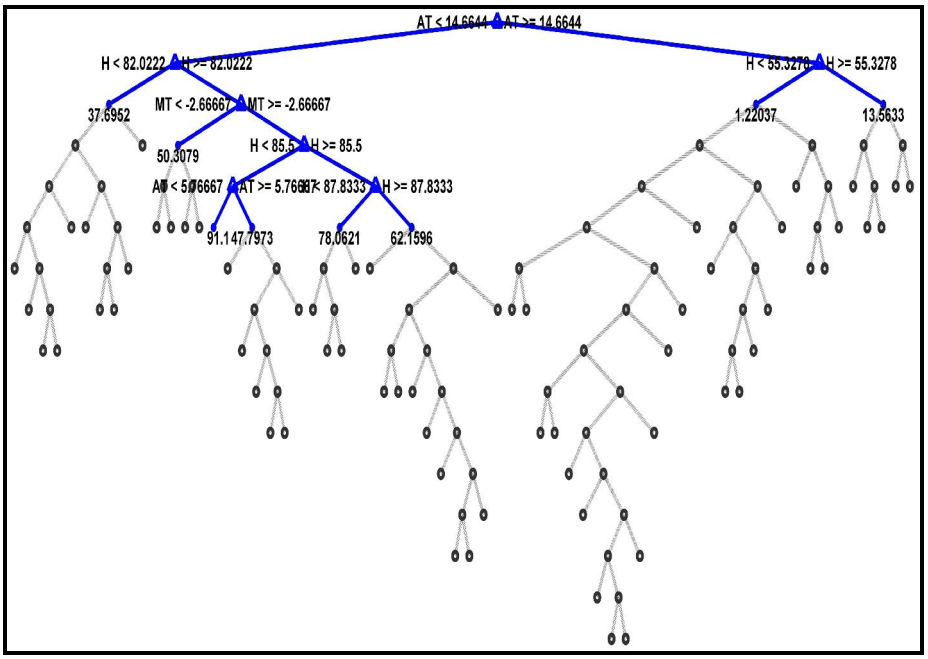
یافته‌های پژوهش

نتایج حاصل از اجرای مدل درخت تصمیم در شرایط استفاده از اصل داده‌ها

در این حالت مدل با ۵ متغیر مستقل شامل بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی، سرعت باد و ۵۱۶ سطر داده اجرا شده است. لازم به ذکر است که از این تعداد

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

۳۶۱ سطر در مرحله آموزش و ۱۵۵ سطر در مرحله آزمون استفاده گردیده است که ساختار درخت ایجاد شده در این حالت در شکل (۳) نمایش داده شده است.



شکل(۳) ساختار درخت تصمیم ایجاد شده در حالت اول(استفاده از اصل داده‌ها)

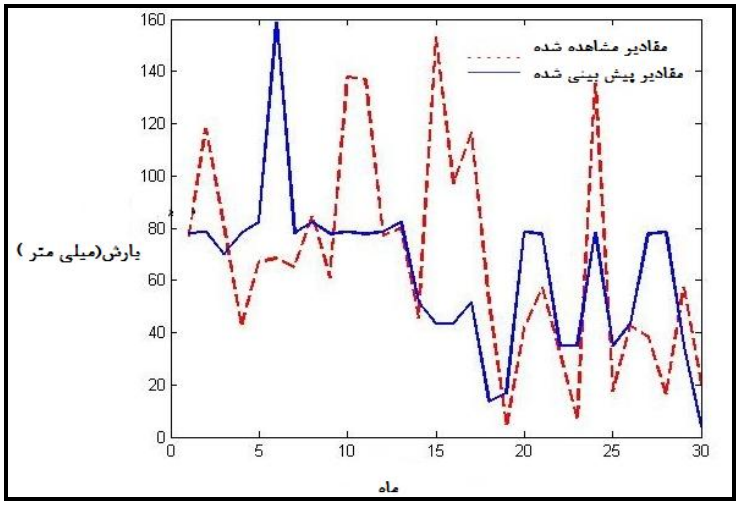
شکل(۳) نشان می‌دهد مدل درخت تصمیم در حالت اول ۵ عنصری را در اختیار داشته است و بارش، دمای متوسط، دمای ماکزیمم و سرعت باد را انتخاب و درخت بر اساس این ۴ عنصر ساخته شده است. در ادامه، کارایی درخت تصمیم ایجاد شده با استفاده از چهار شاخص آماری (R, RMSE, MAE, Bias) بررسی شده که مقادیر این شاخص‌ها در جدول(۳) درج شده است. در جدول زیر متغییر هدف (Pn) بارش می باشد.

جدول (۳) کیفیت نتایج در پیش بینی بارش با استفاده از داده‌های خام (اصل داده‌ها)

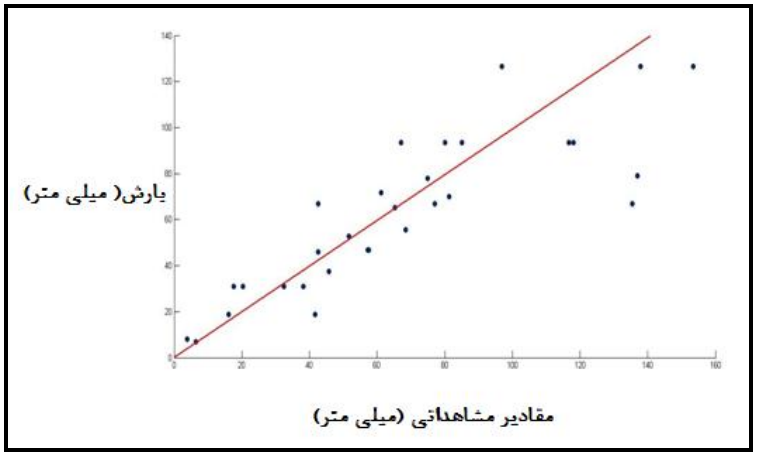
ردیف	متغییر مستقل	متغییر هدف	R	RMSE (mm)	MAE(mm)	Bias
حالت اول	P-Tmean-Tmax-Ws	Pn	۰/۹۱	۵/۷۸	۳/۳	۰/۳۱

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

نمودار نقطه‌ای (الف) و نمودار خطی (ب) مربوط به نتایج مدل درخت تصمیم‌گیری و مقادیر واقعی مربوط به حالت استفاده از اصل داده‌ها نیز در شکل (۴) نمایش داده شده است.



(الف)



(ب)

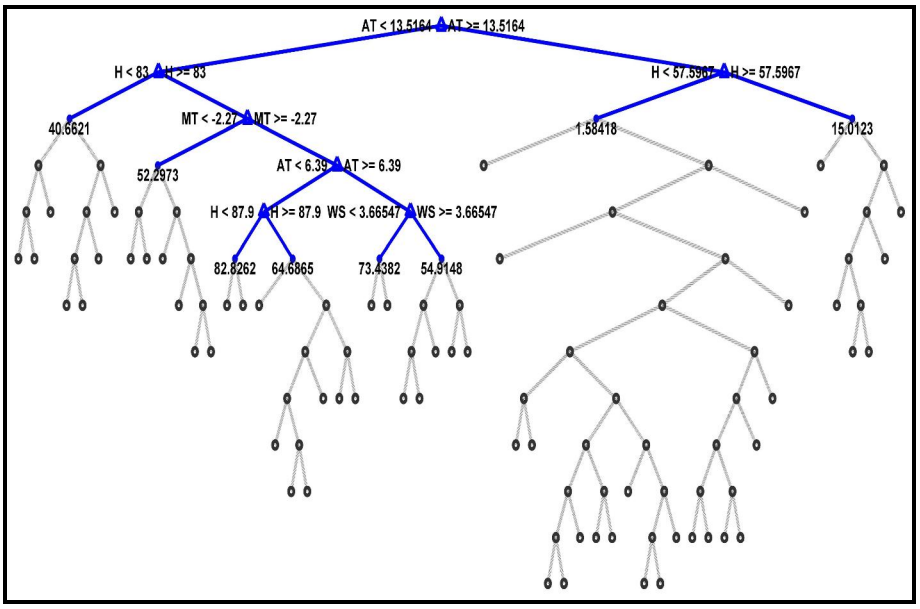
شکل(۴)(الف و ب) تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی درحالت استفاده از اصل داده‌ها

مدل ایجاد شده در شکل (۴) در پیش بینی مقادیر اوج از کارایی بسیار کمی برخوردار است. ضمن اینکه مدل مزبور در غالب بازه‌های زمانی ۳۰ ماهه، پیش بینی‌های خود را نزدیک به دو مقدار عددی حداقل و حداکثر انجام داده است و درخت ایجاد شده با اصل داده‌ها روند پیش بینی نسبتاً ثابتی داشته است

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

نتایج مدل درخت تصمیم با استفاده از میانگین متحرک سه ساله داده‌های اولیه

در این حالت جهت اجرای مدل، ۵ متغیر مستقل شامل میانگین متحرک سه ساله داده‌های بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی و سرعت باد با ۴۹۶ سطر داده در اختیار سیستم قرار گرفته است که از این تعداد ۳۴۸ سطر داده در مرحله آموزش و ۱۴۸ سطر در مرحله آزمون مورد استفاده قرار گرفته است. ساختار درخت ایجاد شده در این حالت در شکل (۵) نمایش داده شده است.



شکل(۵) ساختار درخت تصمیم در حالت دوم (استفاده از میانگین متحرک سه ساله)

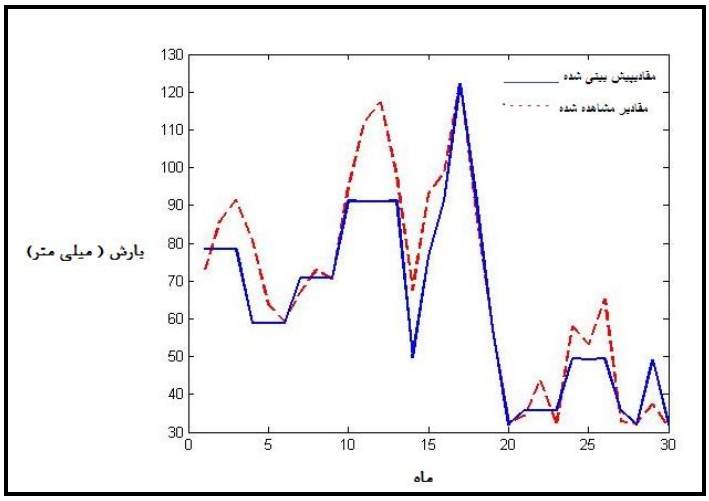
پس از آن کارایی درخت ایجاد شده، با استفاده از پارامترهای آماری ارزیابی گردیده که مقادیر این پارامترها در جدول (۴) درج شده است و میزان شاخص‌های مورد نظر به دست آمده است.

جدول (۴) کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله داده‌ها

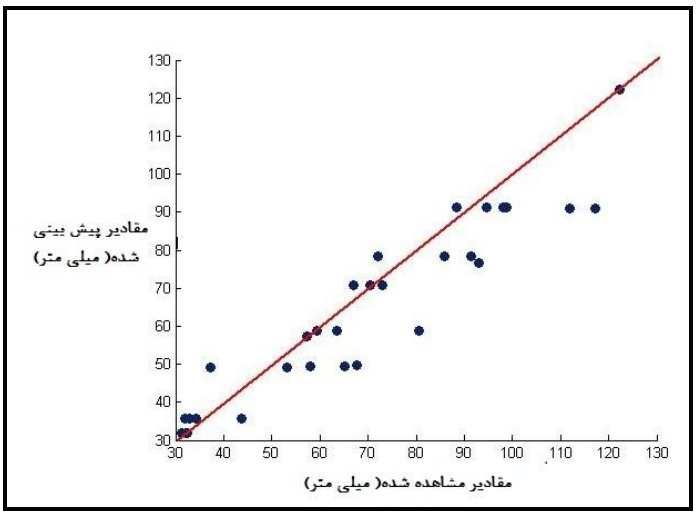
ردیف	متغییر مستقل (میانگین متحرک سه ساله)	متغییر هدف (میانگین متحرک سه ساله)	R	RMSE (mm)	MAE (mm)	Bias
حالت دوم	P -T_{max}-W_s	P_{nm3}	۰/۹۴	۳/۰۲	۱/۹	۰/۱۲

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

نمودار نقطه‌ای (الف) و نمودار خطی (ب) مربوط به نتایج مدل درختان تصمیم و مقادیر واقعی مربوط به حالت استفاده از میانگین متحرک سه ساله داده‌ها در شکل (۶) نمایش داده شده است.



(الف)



(ب)

شکل(۶)(الف و ب) تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه ساله داده‌ها

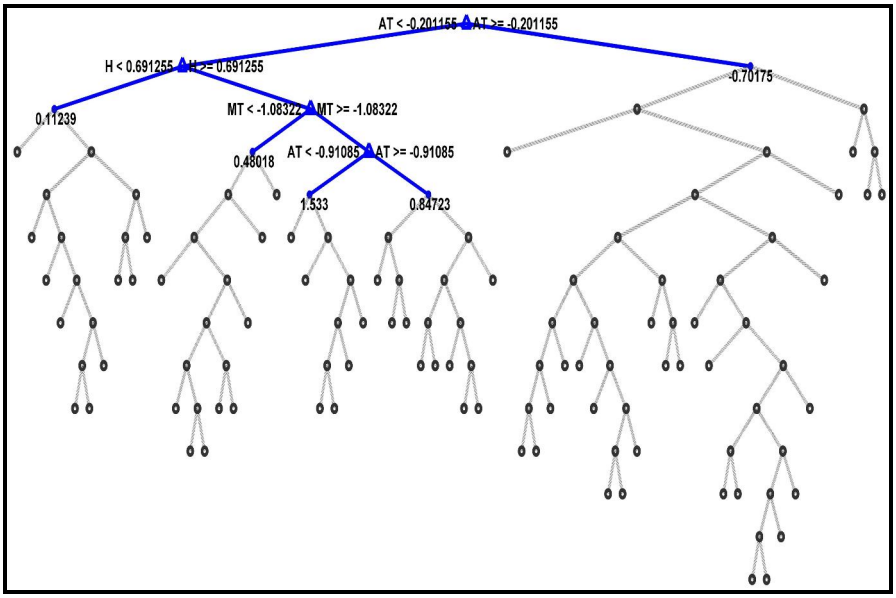
[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

نتایج مدل با استفاده از میانگین متحرک سه ساله نرمال

در این بخش میانگین متحرک سه ساله داده‌ها نرمال شده است و سپس ۵۹۶ سطر داده به نسبت ۷۰ ٪ و ۳۰٪ به ترتیب جهت آموزش و آزمون به مدل معرفی گردیده است و میانگین متحرک سه ساله نرمال شده بارش، ۳۰ ماه قبل از وقوع پیش بینی گردیده است. درخت تصمیم ایجاد شده در این حالت در شکل (۷) نمایش داده شده است.

جدول(۵) کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله نرمال شده

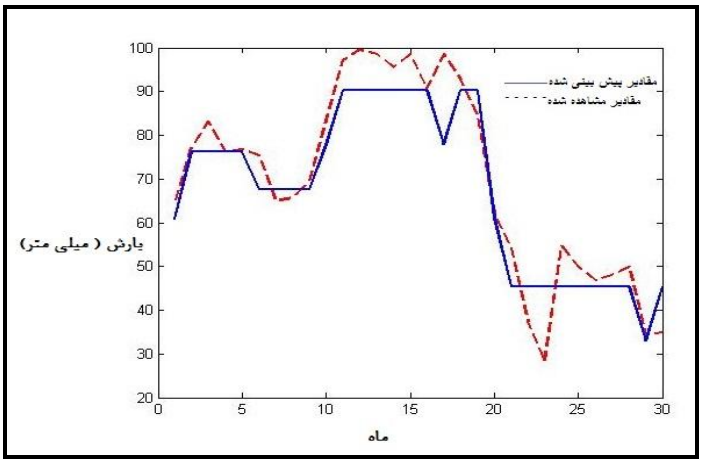
ردیف	متغییر مستقل(میانگین متحرک سه ساله نرمال شده)	متغییر هدف (میانگین متحرک سه ساله نرمال شده)	R	RMSE (mm)	MAE (mm)	Bias
۱	$P - T_{max} - W_s$	P_{nm3}	۰/۹۴	۰/۰۶	۰/۰۵	۰/۰۰۴



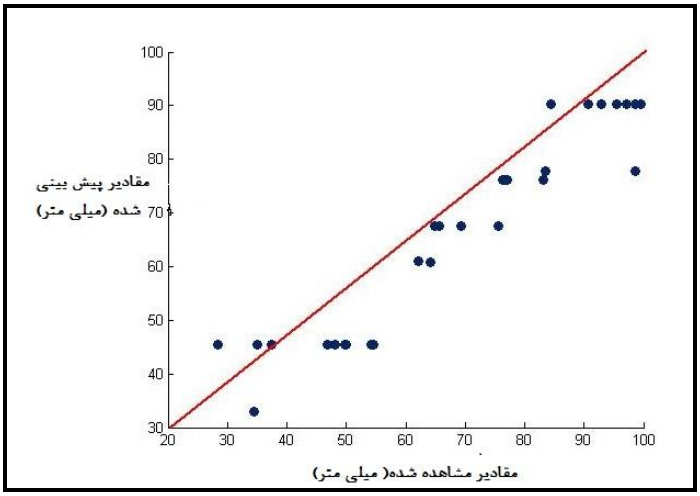
شکل(۷) درخت تصمیم در حالت چهارم(استفاده از میانگین متحرک ۳ ساله داده‌ها)

جهت بررسی کارایی درخت ایجاد شده در شکل (۸) (الف و ب) نیز همانند حالت قبلی از شاخص‌های آماری استفاده گردیده که مقادیر این شاخص‌ها برای این حالت در جدول (۵) آمده است که شاخص R ۰/۹۴، RMSE ۰/۰ ۶ ، MAE۰/۰۵ ، Bias ۰/۰۰۴ می‌باشد.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]



(الف)



(ب)

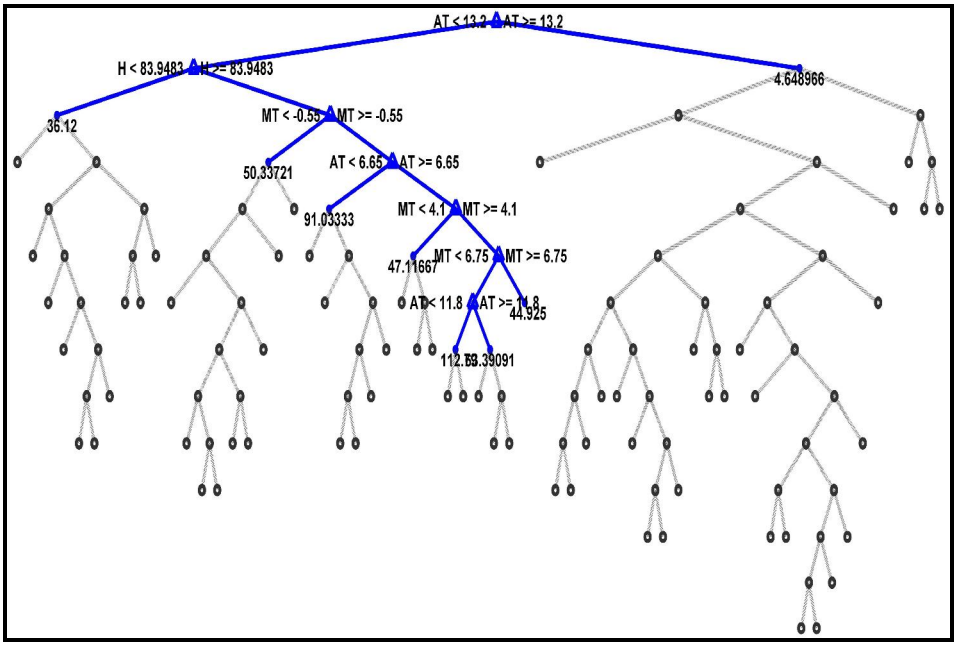
شکل(۸) تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه ساله نرمال شده داده‌ها

نتایج استفاده از میانگین متحرّک سه ساله نرمال شده نسبت به میانگین متحرّک سه ساله نشان می‌دهد که عملیات نرمال سازی کارکرد مدل را تحت الشعاع قرار داده است و کارایی مدل درخت تصمیم را اندکی کاهش داده است.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

نتایج مدل در شرایط استفاده از میانگین متحرک پنج ساله داده‌های اصلی

در ادامه جهت بررسی و بالا بردن کارایی مدل در پیش بینی بارش، از میانگین متحرک پنج ساله داده‌ها نیز استفاده گردیده که ۵ متغیر مستقل شامل میانگین متحرک پنج ساله داده-های؛ بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی و سرعت باد به مدل معرفی و میانگین متحرک پنج ساله بارش برای سال بعد پیش بینی گردیده است. در این حالت ۴۸۴ سطر داده (با توجه به تشکیل میانگین متحرک ۵ ساله تعداد داده‌ها اندکی کمتر شده است) جهت اجرای مدل در اختیار سیستم قرار گرفته است که از این تعداد ۳۴۰ سطر داده در مرحله آموزش و ۱۴۴ سطر داده در مرحله آزمون استفاده گردیده و درخت ایجاد شده در این حالت در شکل (۹) نمایش داده شده است.



شکل(۹) درخت تصمیم در حالت چهارم (استفاده از میانگین متحرک ۵ ساله داده‌ها)

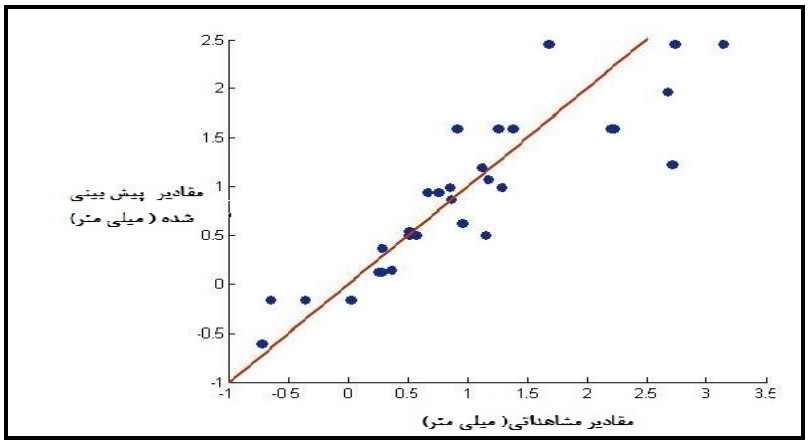
مقادیر شاخص‌های آماری مربوط به ارزیابی نتایج درخت ایجاد شده است در این حالت و نیز متغیرهای مستقل استفاده شده در جدول (۶) آمده است که به ترتیب میزان ۰/۹۷ R، ۱/۷ RMSE، ۰/۱۱ MAE، ۰/۰۴ Bias می‌باشد.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

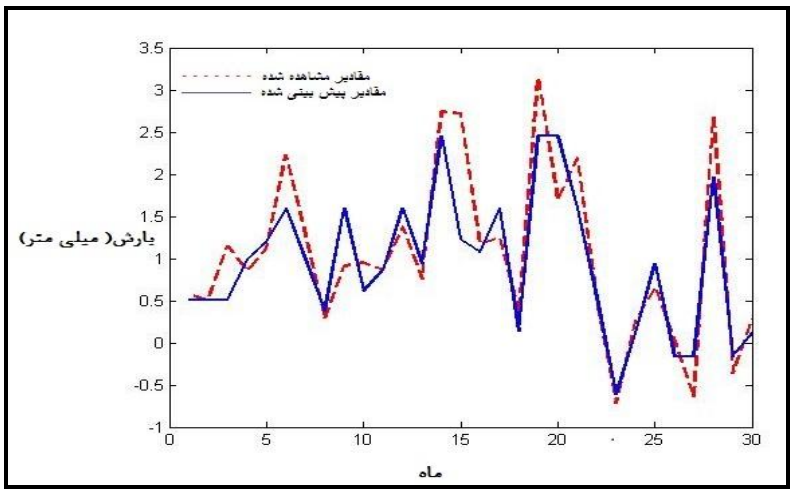
جدول (۶) کیفیت نتایج مدل با استفاده از میانگین متحرک پنج ساله

ردیف	متغییر مستقل (میانگین متحرک پنج ساله)	متغییر هدف (میانگین متحرک پنج ساله)	R	RMSE (mm)	MAE (mm)	Bias
۱	P -Tmax-Ws	Pnm5	۰/۹۷	۱/۷	۰/۱۱	۰/۰۴

نمودارهای نقطه‌ای (الف) و خطی (ب) مربوط به نتایج مدل درختان تصمیم و مقادیر واقعی مربوط به میانگین متحرک پنج ساله داده‌ها در شکل (۱۰) نمایش داده شده است.



(الف)



(ب)

شکل(۱۰) (الف و ب) تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی با استفاده از داده‌های میانگین متحرک پنج ساله

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

در شبیه سازی‌های صورت گرفته زمانی که از میانگین متحرّک پنج ساله داده‌ها برای اجرای مدل استفاده شده است، ترکیب بارش قبلی، دمای ماکزیمم به عنوان مناسب‌ترین حالت شناسایی شده است. همچنین درخت ایجاد شده در حالت میانگین متحرک پنج ساله داده‌ها شکل(۱۰) (الف و ب) قادر است بارش ایستگاه سینوپتیک کرمانشاه را ۳۰ ماه قبل از وقوع پیش بینی نماید.

نتیجه گیری

پیش بینی وضع هوا در محدوده وسیعی از زندگی روزمره دارای کاربرد است. برای نمونه-هایی از این کاربرد می‌توان به پیش بینی وضع هوا در زندگی عادی، مسائل اقتصادی، ورزشی، پزشکی و نظایر آن اشاره نمود. بارش نیز به عنوان یکی از مهمترین عناصر اقلیمی نقش مهمی را در زندگی تمام موجودات دارد. در کشور ایران همچنین به سبب وابستگی بخش عمده‌ای از فعالیت‌های اقتصادی به مقدار و توزیع زمانی بارش آگاهی از میزان و زمان وقوع بارش دارای اهمیت بسیار زیادی است. بنابراین در راستای اهمیت موضوع و با هدف ارزیابی کارایی مدل درخت تصمیم گیری در بر آورد بارش در ایستگاه سینوپتیک کرمانشاه از آمار ماهانه بارندگی، تبخیر، رطوبت نسبی، دمای ماکزیمم، دمای متوسط و سرعت باد در دوره آماری (۱۳۸۹-۱۳۴۹) استفاده شده است. در نتایج حاصل از این پژوهش و تمام شبیه سازی‌های صورت گرفته، پارامترهای بارش قبلی و دمای ماکزیمم دخیل بوده است که این موضوع نشان دهنده اهمیت این پارامترها در پیش بینی فرایند بارش در آینده می‌باشد. بر اساس نتایج به دست آمده در این پژوهش، می‌توان گفت درخت تصمیم گیری رگرسیونی در ایستگاه سینوپتیک کرمانشاه، مدلی نسبتاً کارا در پیش بینی بارش با استفاده از داده‌های پردازش شده می‌باشد که استفاده از میانگین متحرّک نسبت به سایر حالات منجر به افزایش چشمگیر کارایی مدل درخت تصمیم می‌شود و در صورت تعدیل دامنه تغییرات داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان بارش را ۳۰ ماه قبل از وقوع برآورد نماید که در شبیه سازی‌های صورت گرفته در این پژوهش، زمانی که از میانگین متحرّک پنج ساله داده‌ها برای اجرای مدل استفاده شده است، ترکیب بارش قبلی، دمای ماکزیمم به عنوان مناسب‌ترین حالت شناسایی شده است. همچنین درخت ایجاد شده در حالت میانگین متحرک پنج ساله داده‌ها قادر است بارش ایستگاه سینوپتیک کرمانشاه را ۳۰ ماه قبل از وقوع پیش بینی نماید. در ایستگاه کرمانشاه، به نظر می‌رسد دامنه تغییرات داده‌ها برای استفاده در مدل درخت تصمیم زیاد می-باشد و همین موضوع منجر به کاهش کارایی مدل در پیش بینی می‌گردد به همین دلیل می-

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

توان گفت که مدل درخت تصمیم گیری در پیش بینی بارش در شرایط استفاده از اصل داده-ها ناموفق عمل کرده و جهت بهتر شدن کارایی مدل پیش بینی از داده‌های میانگین متحرک استفاده شده است. بنابراین شکل (۴) نشان دهنده این موضوع است که مدل در حالت استفاده از اصل داده‌ها در دوره‌های زمانی ۳۰ ماهه پیش بینی‌های خود را در دو حالت حداقل و حداکثر انجام داده است و درخت ایجاد شده با اصل داده‌ها روند پیش بینی نسبتاً ثابتی داشته است و روند مناسبی را دنبال نکرده است. بنابراین استفاده از میانگین متحرّک سه ساله نرمال شده نسبت به میانگین متحرّک سه ساله نشان می‌دهد که عملیات نرمال سازی کارکرد مدل را تحت الشعاع قرار داده است و کارایی مدل درخت تصمیم را اندکی کاهش داده است. به نظر می‌رسد که این موضوع ناشی از نحوه کارکرد مدل باشد و عملیات نرمال سازی که هم داده‌های با دامنه تغییرات گسترده و هم داده‌های با دامنه تغییرات کم را در یک دامنه عددی ۱- ۰ محدود می‌سازد، قدرت درخت را در انتخاب متغیرهای مهمتر در ایجاد شاخه زنی و تفکیک داده‌ها کاهش می‌دهد.

بنابراین در این پژوهش بر اساس نتایج به دست آمده، درخت تصمیم گیری رگرسیونی، مدلی نسبتاً کارا در پیش بینی بارش با استفاده از داده‌های پردازش شده می‌باشد که استفاده از میانگین متحرّک نسبت به سایر حالات منجر به افزایش چشمگیر کارایی مدل درخت تصمیم می‌شود. همچنین درخت ایجاد شده در حالت میانگین متحرک پنج ساله داده‌ها قادر است بارش ایستگاه کرمانشاه را ۳۰ ماه قبل از وقوع پیش بینی نماید.

منابع و ماخذ

۱. امیدوار، کمال، اژدرپور، مهران (۱۳۹۱) **مقایسه شبکه عصبی مصنوعی و مدل – HEC HMS در برآورد بارش – رواناب در حوضه آبریز رودخانه اعظم هرات**، فصلنامه تحقیقات جغرافیایی، سال۲۷، شماره چهارم، شماره پیاپی ۱۰۷، صص ۱۳۹- ۱۵۹.

۲. پوراصغری، فرناز، جهانبخش، سعید، ساری صراف، بهروز، قائمی، هوشنگ، تدینی، معصومه (۱۳۹۲) **پهنه بندی رژیم بارش در نیمه جنوبی ایران**، جغرافیا و برنامه ریزی، سال۱۷، شماره۴۴، صص ۲۷-۴۶.

۳. حلبیان، امیر حسین (۱۳۸۸) **پیش آگاهی و برآورد بارش یزد با استفاده از شبکه-های عصبی مصنوعی**، نشریه تحقیقات کاربردی علوم جغرافیایی، ج۱۱، شماره ۱۴، صص ۷-۲۸.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

۴. حلبیان، امیر حسین، دارند، محمد (۱۳۹۱) **پیش بینی بارش اصفهان با استفاده از شبکه‌های عصبی مصنوعی**، نشریه تحقیقات کاربردی علوم جغرافیایی، شماره ۲۶، صص ۴۷- ۶۳.

۵. دستورانی، محمد تقی، حبیبی پور، اعظم، اختصاصی، محمد رضا، طالبی، علی، محجوبی، جواد (۱۳۹۱) **بررسی کارایی مدل درخت تصمیم در پیش بینی بارش(مطالعه موردی ایستگاه سینوپتیک یزد)**، فصلنامه تحقیقات منابع آب ایران، سال هشتم، شماره ۳، صص ۱۴-۲۷.

۶. شایق، محمد علی (۱۳۹۰) **ارزیابی پروژه‌های باروری ابرها با استفاده از مدل درختان تصمیم گیری رگرسیونی(مطالعه موردی: ایران مرکزی- استان فارس)**، پایان نامه کارشناسی ارشد مهندسی عمران آب، دانشگاه صنعت آب و برق، تهران.

۷. طالبی، علی، اکبری، زینب (۱۳۹۲) **بررسی کارایی مدل درختان تصمیم گیری در برآورد رسوبات معلق رودخانه‌ای(مطالعه موردی: حوضه سد ایلام)**، علوم و فنون کشاورزی و منابع طبیعی، علوم آب و خاک، سال ۱۷، شماره ۶۳، صص ۱۰۹- ۱۲۱.

۸. عرفانیان، مریم، انصاری، حسین و علیزاده، امین (۱۳۹۲) **پیش بینی بارش و دمای متوسط ماهانه با استفاده از الگوهای پیوند از دور به کمک شبکه‌های عصبی مصنوعی (مطالعه موردی: ایستگاه سینوپتیک مشهد)**، فصلنامه مطالعات جغرافیایی مناطق خشک، شماره ۱۱، صص ۵۳-۷۳.

۹. عساکره، حسین (۱۳۸۶) **کاربرد رگرسیون خطی و تحلیل روند بارش سالانه تبریز**، فصلنامه تحقیقات جغرافیایی، انتشارات دانشگاه اصفهان، شماره۲۲، پیاپی ۸۷، صص ۳- ۲۶.

۱۰.فیروزی، فاطمه، نگارش، حسین، خسروی، محمود (۱۳۹۱) **مدل سازی، پیش بینی و بررسی روند بارش در ایستگاه‌های منتخب استان فارس**، فصلنامه برنامه ریزی منطقه‌ای، سال دوم، شماره ۷، صص ۷۷-۹۱.

۱۱.کانتاردزیک، مهمد(۱۳۸۵) **داده کاوی**، امیر علیخان زاده، بابل، چاپ اول، نشر علوم رایانه، ص ۳۴۴.

۱۲. محجوبی، جواد، صمغی، حسین اردلان (۱۳۸۸) **پیش بینی پارامترهای امواج ناشی از باد در دریای خزر با استفاده از روش درختان تصمیم رگرسیونی و شبکه‌های عصبی مصنوعی**، نشریه مهندسی دریا، سال پنجم، شماره ۹، صص ۶۴- ۷۱.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]

۱۳. مشکانی، علی، ناظمی، عبدالرضا (۱۳۸۸) **مقدمه‌ای بر داده کاوی**، مشهد، چاپ اول، موسسه چاپ و نشر دانشگاه فردوسی، ص ۴۵۶.

Ayoubloo, M.K., Etemad-Shahidi, A. Mahjoobi, J, (2010) ***Evaluation of regular wave scour around a circular pile using data mining approaches***. Applied Ocean Research. 32, pp. 34-39.

Bhattacharya B, Price R K, Solomatine D P, (2007) ***Machine learning approach to modeling sediment transport***. ***J.*** Hydraul. Eng. 133(4), Pp. 440 - 450.

Breiman L, Friedman J, Olshen R, Stone C, (1984) ***Classification and Regression Trees, Chapman & Hall/CRC Press***, Boca Raton, FL. Development of a decision tree modeling approach .Geoderma 139, Pp. 277-287.

Geissen V, Kampichler C, (2007) ***Superficial and subterranean soil erosion in Tabasco, tropical mexico***

Hand D, Heikki M, Padhraic S, (2001) ***Principles of Data Mining***, The MIT press.

Kazeminezhad M H, Etemad Shahidi A, Mousavi, S J, (2005) ***Application of fuzzy inference system in the prediction wave parameters***, Ocean Eng; 32: Pp. 1709-1725.

Kocev D, Saso D M D, Newell G R, Griffioen P, (2009) ***Using single-and multi-target regression trees and ensembles to model a compound index of vegetation condition.*** Ecological Modeling 220, Pp. 1159 –1168.

Mahesh P, Mather P M, (2003) ***an assessment of the effectiveness of decision tree methods for land cover classification***. Remote Sensing of Environment, No, 86, Pp. 554–565.

Rusjan S, Micos M, (2008) ***Assessment of hydrolpgical and seasonal controls over the nitrate flushing from a forested watershed using a data mining technique***. Hydrol .Earth Sys. Sci. 12. Pp. 645-656.

Yurekli K, Taghi Sattari M T, Anli AS, Hinis M A, (2012)***Seasonal and annual regional drought Prediction by using data-mining approach, Atmosfera***, 25(1) Pp. 85-105.

[Downloaded from ndea10.khu.ac.ir on 2024-09-20]