مجلهٔ مخاطرات محیط طبیعی، سال چهارم ، شماره ششم، زمستان ۹۴
تاریخ دریافت مقاله: ۹۳/۰۵/۰۱
تاریخ پذیرش مقاله: ۹۴/۱۱/۱۲
صفحات : ۱۹-۱

# ارزیابی کارایی مدل درخت تصمیم رگرسیونی در پیش‌بینی خشکسالی نمونهٔ موردی: ایستگاه سینوپتیک سنندج

***دکتر غلامعلی مظفری [۱]، شهاب شفیعی [۲]، زهرا تقی‌زاده [۳]**

## چکیده

برای مطالعهٔ خشکسالی روش‌های مختلفی وجود دارد. روش تحلیل داده‌های بارندگی، جزو عمومی روش های تحلیل خشکسالی به شمار می رود؛ لذا پیش‌بینی دقیق و پیش از وقوع بارش می تواند شرایط را برای ارزیابی وضعیت خشکسالی فراهم نماید. هدف این پژوهش، بررسی تأثیر پیش‌پردازشِ داده‌های بارش ماهانهٔ ایستگاه سینوپتیک سنندج بر عملکرد مدل درخت تصمیم در پیش‌بینی خشکسالی در ایستگاه سینوپتیک سنندج می‌باشد. در این پژوهش از الگوریتم **CART** به عنوان یکی از انواع درختان تصمیم رگرسیونی جهت پیش‌بینی بارش ۱۲ ماه بعد استفاده شده و جهت ارزیابی درخت‌های ایجاد شده از معیارهای آماری مختلف استفاده شده‌است. داده‌های مورد استفاده در این پژوهش مربوط به آمار ماهانهٔ بارندگی، رطوبت نسبی، دمای حداکثر، دمای متوسط، جهت باد و سرعت باد در دورهٔ آماری (۱۳۸۹- ۱۳۴۹) است. نتایج حاصل از پژوهش نشان می‌دهد که در ایستگاه سینوپتیک سنندج درخت تصمیم‌گیری رگرسیونی، مدلی نسبتاً کارا در پیش‌بینی خشکسالی می‌باشد؛ به‌طوری‌که در شبیه‌سازی‌های صورت گرفته، زمانی که از میانگین متحرّک پنج سالهٔ داده‌ها برای اجرای مدل استفاده گردید، ترکیب بارش قبلی و دمای حداکثر به عنوان مناسب‌ترین حالت با مقدار خطای ۰/۰۶ شناسایی شده و اعمال میانگین متحرک روی داده‌های اصلی در بهبود کارایی مدل مؤثر است. در این شرایط، روش درخت تصمیم رگرسیونی ایستگاه سنندج با ضریب اطمینان بالایی میزان بارش را ۱۲ ماه پیش از وقوع بر آورد نمایند.

**کلید واژه ها:** درخت تصمیم، پیش بینی بارش، الگوریتم **CART**، سنندج

۱- *دانشیار آب و هواشناسی، دانشگاه یزد (نویسنده مسئول) gmozafar@yazd.ac.ir
۲- دانشجوی دکتری آب و هواشناسی، دانشگاه یزد
۳- دانشجوی دکتری آب و هواشناسی، دانشگاه یزد

**مقدمه**

در دهه‌های اخیر تغییر سیستم اقلیمی زمین به ویژه گرمایش سیاره‌ای ناشی از مصرف سوخت‌های فسیلی به طور گسترده‌ای مورد توجه قرار گرفته است. هیأت بین‌الدول تغییر اقلیم (IPCC) در سال ۲۰۰۱ گزارش داد که تغییر و گرمایش جهانی در حال وقوع است. در سال‌های اخیر به علت نتایج اقتصادی، اجتماعی و آسیب‌های مالی، توجه به تغییرات اقلیمی اهمیت پیدا کرده است. یکی از اثرات تغییر سیستم اقلیمی کره زمین، بروز خشکسالی در بسیاری از نقاط و یا تشدید سیل و سیل‌گیری در نقاط دیگر است که برخلاف سایر بلایای طبیعی به آرامی آغاز و به طور خزنده‌ای گسترش می‌یابد و تقریباً در همهٔ رژیم‌های اقلیمی دیده می‌شود *(ویل هایت،۱۹۹۳: ۲۴۰)*. خشکسالی یکی از مخاطرات خزندهٔ محیطی است که در مناطق خشک و نیمه‌خشک جهان نمود بیشتری دارد *( رضیئی و همکاران، ۱۳۸۶: ۲۹۳-۳۱۰)* و یک پدیدهٔ متناوب طبیعی است که همراه با کمبود منابع آبِ در دسترس در یک منطقهٔ جغرافیایی پهناور، در یک دورهٔ زمانی قابل ملاحظه می‌باشد *(پیری و همکاران،۱۳۹۲: ۲۶)*. خشکسالی بدون هیچ یک از علائم هشداردهنده و در محدوده‌ای با مرزهای غیر قابل تشخیص اتفاق می‌افتد؛ طی سال‌های اخیر در ایران، خشکسالی‌های مستمر و شدیدی رخ داده که منجر به خشک یا کم شدن منابع آب‌های سطحی و زیرزمینی شده است *(کشاورز و همکاران، ۱۳۹۱: ۹۸)*. وجود نوسانات شدید در ریزش‌های جوّی یکی از مشکلات عمدهٔ بارش کشور است و با توجه به همین نوسانات شدید، خشکسالی‌ها از جمله بلایای طبیعی هستند که در شرایط نوسان منفی و به عبارت دیگر، کاهش بارش‌های جوّی نسبت به میانگین دراز مدت، رخ می‌دهند. یکی از عمده‌ترین عواملی که در تعریف خشکسالی استفاده می گردد ، بارش است *( فرج زاده، ۱۳۸۴: ۱۶-۱۷)*. با کم شدن بارندگی یا نبود آن برای یک مدت طولانی در سال (خشکسالی)، مراتع، جنگل‌ها، مزارع و باغ‌هایی که منبع آبی آن‌ها وابسته به ریزش-های جوی است و نیز خاک و دیگر منابع طبیعی به طور مستقیم زیان می‌بینند. به این ترتیب با وقوع خشکسالی، فعالیت ها و تاسیسات وابسته به آب، چه شهری و روستایی، چه صنعتی و غیره تحت تأثیر قرار گرفته و خسارت خواهند دید *(کردوانی، ۱۳۸۰: ۱۳۴)*. از این رو پیش بینی خشکسالی می‌تواند در اعمال روش‌های موثر مدیریت منابع آب و تعدیل خسارات ناشی از خشکسالی موثر واقع شود. بنابراین در بسیاری از برنامه های بلند مدت، لازم است که چشم اندازی از وضعیت آینده بارندگی و دوره‌های خشک و تر برای منطقه ترسیم گردد *(حسن زاده و همکاران، ۱۳۹۱: ۴۹)*. بر این اساس روش‌های مختلفی برای مطالعهٔ خشکسالی وجود دارد. روش تحلیل داده‌های بارندگی، جزو روش-های عمومی تحلیل خشکسالی به شمار می‌رود؛ علت این امر دسترسی راحت‌تر و آسان تر به انواع داده‌های بارندگی در بخش‌های مختلف کره زمین است. از طرفی مقادیر بارندگی جزو بی ثبات‌ترین متغیرهای آب و هوایی، مخصوصاً در مناطق خشک محسوب می شود، به همین دلیل نمایه خوبی برای مطالعه خشکسالی می‌باشد *(سلسو آگوستو، ۲۰۰۹: ۳۰۲-۳۱۲)*. 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) اشاره کرد. درختان تصمیم[1] به همراه قوانین تصمیم یکی از روش‌های داده کاوی به شمار می آیند *(کانتاردزیک،۱۳۸۵: ۷۸)* که یکی از ابزارهای قوی و متداول برای دسته بندی و پیش

**1- Decision Trees**

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 شبکهٔ عصبی مصنوعی، ضرورتی وجود ندارد که داده‌ها لزوماً به صورت عددی باشند (*مشکانی و ناظمی،۱۳۹۱: ۱۸*). 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 اساساً دارای مسائل غیر خطی هستند به خوبی از عهدهٔ حل مسائل بر آمده‌است؛ لذا کم‌کم پای این ابزار جدید به علوم مهندسی آب و منابع طبیعی که جزو پدیده‌های غیر خطی و پیچیده هستند نیز باز شده‌است *(طالبی و اکبری، ۱۳۹: ۱۱۰)*. امروزه حجم زیادی از داده‌ها توسط سازمان‌های متولی امر در مقیاس‌های زمانی مختلف ثبت می‌گردد. نظر به گسترش سیستم پایگاه داده و ایجاد ابزارهای متعدد برای ذخیرهٔ حجم بالایی از داده‌ها، داده‌کاوی به عنوان شاخهٔ علوم بین رشته‌ای توسعهٔ روز افزونی یافته است. داده‌کاوی به جستجو و کشف مدل‌های گوناگون، چکیده‌سازی و اخذ مقادیر از مجموعه‌ای از داده‌های معلوم اطلاق می‌گردد *(کانتاردزیک، ۱۳۸۵:۱۳۸)*. دسته‌بندی و پیش‌بینی دو نوع عملیات داده‌کاوی برای تحلیل داده‌ها و استخراج مدل به منظور توصیف دسته‌های مهم داده‌ها، فهم و پیش‌بینی رفتار آنها در آینده است. مدل‌های دسته‌ای در تحلیل داده-های گسسته و طبقه‌ای، و مدل‌های پیش‌بینی یا رگرسیون بیشتر روی داده‌های پیوسته کار می‌کنند *(مشکانی و ناظمی :۱۳۸۸. ۱۸۵)*.

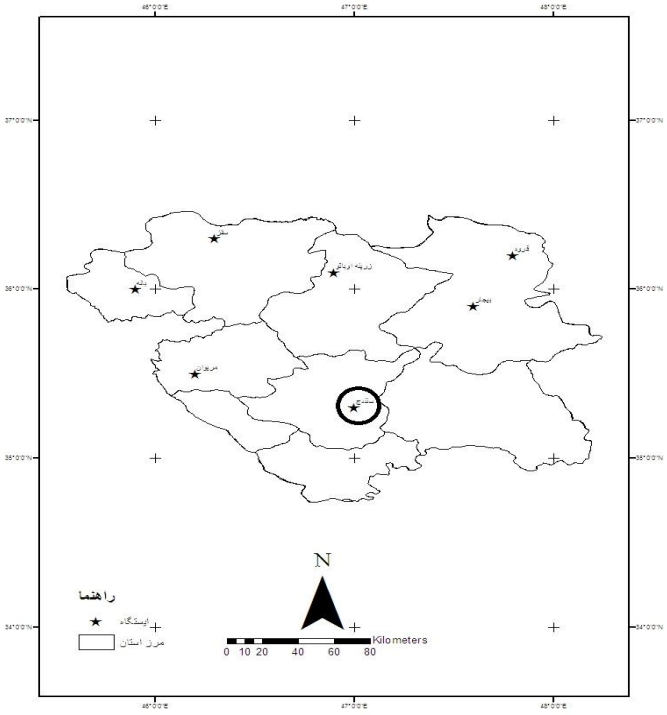
از جمله تحقیقات بین‌المللی و داخلی صورت گرفته بر روی کارایی درخت تصمیم در زمینهٔ علوم طبیعی می‌توان به مطالعات ماهش و مادر (۲۰۰۳)، گیسن و همکاران (۲۰۰۷)، باتاچاریا و همکاران (۲۰۰۷)، روس جان و میکوس (۲۰۰۸)، کوسوو و همکاران (۲۰۰۹)، باتاچاریا و همکاران (۲۰۱۲)، محجوبی و اعتماد شهیدی (۱۳۸۶)، محجوبی و صمغی (۱۳۸۸) ایوب‌لو و همکاران (۱۳۸۹) شایق (۱۳۹۰) طالبی و اکبری (۱۳۹۲)، اشاره نمود.

همچنین از مطالعات صورت گرفته در زمینهٔ استفاده از درخت تصمیم در رابطه با خشکسالی و بارش می‌توان به موارد ذیل اشاره نمود: یوریک لی و همکاران (۲۰۱۲) از تکنیک درخت تصمیم جهت ارزیابی و پیش بینی طبقات خشکسالی در حوزه‌ای در ترکیه استفاده کرده‌اند و نتایج حاصله را با روش طبقه‌بندی و تحلیل خشکسالی بر اساس شاخص بارش مقایسه نموده‌اند. مقایسهٔ نتایج، تفاوتِ قابل ملاحظهٔ (SPI) استاندارد را در طبقه‌بندی خشکسالی بر اساس این دو روش نشان نداده است. افخمی و همکاران (۱۳۸۹) با استفاده از روش شبکهٔ عصبی نقش تعداد و نوع عوامل اقلیمی را در دقت پیش‌بینی خشکسالی استان یزد مورد ارزیابی قرار دادند. از مدل درخت تصمیم رگرسیونی جهت ارزیابی پروژه‌های باروری ابرها در استان فارس استفاده نموده‌اند. دستورانی و همکاران (۱۳۹۱) به بررسی کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه سینوپتیک یزد پرداخته‌اند که در این تحقیق از مدل درخت تصمیم به عنوان یکی از روش‌های داده‌کاوی جهت پیش‌بینی بارش و ارزیابی وضعیت خشکسالی در ایستگاه سینوپتیک یزد استفاده شده‌است. نتایج نشان داده‌اند که در ایستگاه یزد، مدل درخت تصمیم‌گیری خصوصاً در شرایطی که از میانگین متحرک ۵ ساله داده‌ها استفاده گردد، دارای توانایی مناسبی در پیش‌بینی میزان بارش می‌باشد.

به طور کلی نتایج این تحقیقات مدل درخت تصمیم را به عنوان ابزاری دقیق و قابل اعتماد معرفی می‌کند؛ 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 این تحقیق نیز بررسی کارایی مدل درخت تصمیم رگرسیونی در پیش‌بینی بارش در حالت استفاده از داده‌های اصلی و مقایسه آن با کارکرد مدل در شرایط استفاده از میانگین متحرک سه ساله، میانگین متحرک نرمال شدهٔ سه ساله و میانگین متحرک پنج ساله در ایستگاه سینوپتیک سنندج می‌باشد.

## داده و روش‌شناسی

برای انجام این پژوهش از داده‌های هواشناسی ایستگاه سینوپتیک سنندج استفاده شده‌است. همان‌طور که در شکل (۱) مشاهده می‌شود این ایستگاه در موقعیت جغرافیایی ۴۷ ° طول جغرافیایی، ۲۰ ' ۳۵ ° عرض جغرافیایی و در ارتفاع ۱۳۷۳ متر از سطح دریا قرار دارد. داده‌های مورد استفاده در این تحقیق شامل آمار ماهیانهٔ بارندگی، رطوبت نسبی، دمای حداکثر، دمای متوسط و سرعت باد در دورهٔ آماری (۱۳۸۹- ۱۳۴۹) است. داده‌ها هیچ‌گونه نقص آماری نداشته و جهت کنترل کیفیت آمار و اطلاعات موجود از آزمون توالی استفاده شده‌است. نتایج نشان دادند که داده‌های مزبور از همگنی لازم برخوردار می‌باشند. برای انجام این پژوهش از نرم‌افزارهای مت‌لب[1] و مینی‌تب[2] استفاده شده‌است.




**شکل ۱: موقعیت جغرافیایی ایستگاه سینوپتیک سنندج**

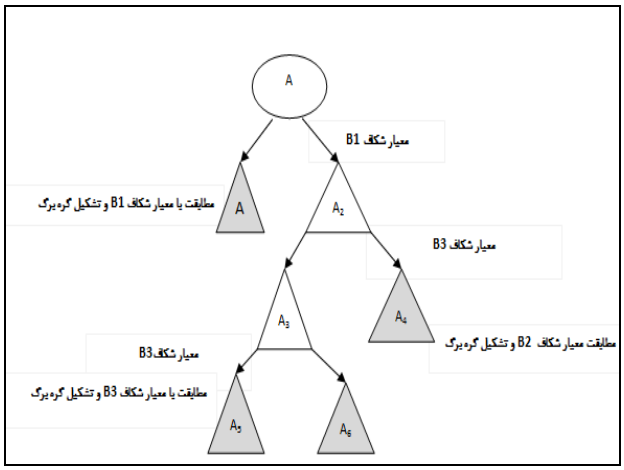
**1- MATLAB**
**2- MINITAB**

جهت کنترل کیفیت آماری مشاهدات از نمودار $\overline{X}$ و $\overline{S}$ استفاده شده‌است. در این نمودار حدود بالایی (UCL)[1]، پایینی (LCL)[2] و مرز میانی نمودار ($\overline{\overline{X}}$ و $\overline{\overline{S}}$) به صورت زیر محاسبه می‌شود (هریدی و وو، ۲۰۰۹).

رابطه (۱) $$UCL,LCL=\overline{\overline{X}}\pm \frac{3}{d2\sqrt{n}}\overline{\overline{R}}$$

**درخت تصمیم**

همان‌طور که گفته شد در این پژوهش از الگوریتم داده‌کاوی درخت تصمیم رگرسیونی CART[3] به عنوان یکی از انواع درختان تصمیم رگرسیونی جهت پیش‌بینی استفاده شده است (شکل ۲).




شکل ۲: نمونه‌ی از یک درخت تصمیم ساده

ساختار این درختان بر سه اصل استوار است:

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

۲) بهترین معیار شاخه‌زدن جهت انتخاب بهترین متغیر مستقل برای ایجاد شاخهٔ جدید.

۳) ایجاد آمار خلاصه برای گرهٔ انتهایی (بریمان و همکاران، ۱۹۸۴: ۲۸۲).

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

1- Lower Control Limit
2- Upper Control Limit
3- Classification and regression tree

رابطه(۲)

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

SS (t): معیاری جهت ایجاد شاخه و تولید درخت تصمیم.

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

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

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

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

رابطه (۳)

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

که در آن ss(t R) و SS(t L) به ترتیب میزان ss(t) در شاخهٔ سمت راست و سمت چپ گرهٔ t می باشد. در این فرایند داده‌ها به دو دسته تقسیم می گردند: داده‌های آموزشی[1] و داده‌های آزمایشی[2]؛ به طوری‌که در این پژوهش ۷۰ ۷۰ ٪ از کل داده‌ها به آموزش مدل تعلق دارد و ۳۰ ٪ باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردیده است. انتخاب داده‌های آموزشی و آزمونی به صورت سیستماتیک و توسط کاربر انجام شده است و بدیهی است که داده‌های آزمون در مرحلهٔ آموزش مورد استفاده قرار نگرفته‌اند.

## پارامترهای مورد استفاده در شبیه‌سازی

در شبیه‌سازی با استفاده از مدل درخت تصمیم‌گیری از عناصر مختلفی استفاده شده‌است. این عناصر به عنوان متغیر مستقل به مدل معرفی شده و شبیه‌سازی‌ها برای پیش‌بینی متغیر هدف صورت گرفته است. متغیرهای مستقل مورد استفاده در این تحقیق و علائم اختصاری آن‌ها در جدول (۱) نمایش داده شده است:

**جدول۱: معرفی پارامترهای ورودی مدل (دوره آماری ۱۳۴۹ تا ۱۳۸۹)**

| میانگین ماهانه | حداقل | حداکثر | عناصر ورودی | علامت اختصاری |
|---|---|---|---|---|
| ۳۵/۹۴ | ۰ | ۲۰۴ | بارش (میلیمتر) | P |
| ۱۳/۵۵ | -۱۳/۳ | ۲۹/۳ | دمای متوسط (درجه سانتیگراد) | Tmax |
| ۲۱/۳۲ | -۶/۳ | ۳۸/۸ | دمای حداکثر (درجه سانتیگراد) | Tmean |
| ۴۰ | ۱۷ | ۸۰ | رطوبت نسبی (درصد) | H |
| ۳/۹۵ | ۰/۴ | ۹/۱ | سرعت باد غالب (متربرثانیه) | WS |

**1-Training data**
**2-Testing data**

**الف)استفاده از اصل داده‌ها بدون نرمال‌سازی:**

در این حالت بر روی داده‌ها هیچ‌گونه عملیات تغییر مقیاس انجام نشد. ورودی و خروجی همان داده‌های خام بودند که به مدل معرفی شدند؛ به عبارت دیگر در این حالت، پارامترهای درج شده در جدول (۱) به عنوان ورودی در اختیار مدل قرار گرفت و بارش ۱ سال بعد پیش‌بینی گردید.

**ب)استفاده از میانگین متحرک سه سالهٔ داده‌های خام:**

بعد از انجام چندین شبیه‌سازی جهت بالا بردن ضریب کارایی در پیش‌بینی بارش، به جای داده‌های ماهیانه از میانگین متحرک سه ساله استفاده شد. به این ترتیب میانگین متحرک سه ساله برای کل پارامترهای ورودی مدل محاسبه شده‌است و با در نظر گرفتن این پارامترها به عنوان متغیر مستقل، میانگین متحرک سه سالهٔ بارش برای یک سال بعد پیش‌بینی می‌گردد.

**ج) استفاده از میانگین متحرک سه سالهٔ نرمال شده:**

در این حالت میانگین متحرک سه سالهٔ داده‌ها با استفاده از رابطهٔ ۴ نرمال شد و سپس درخت تصمیم برای انجام پیش‌بینی ایجاد و مورد ارزیابی قرار گرفته است.

رابطه ۴:

$$Xn=\frac{X_0 - X_{\min}}{X_{\max} - X_{\min}}$$

X0 و Xn به ترتیب داده‌های نرمال‌شده و داده‌های اولیه را نشان می‌دهد.

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

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

**د) استفاده از میانگین متحرک پنج سالهٔ داده‌های خام:**

در این حالت میانگین متحرک پنج ساله برای کل پارامترهای ورودی مدل محاسبه شد و با در نظر گرفتن این پارامترها به عنوان متغیر مستقل، میانگین متحرک پنج سالهٔ بارش برای یک سال بعد پیش‌بینی گردیده‌است.

**ارزیابی کارایی مدل درخت تصمیم**

جهت ارزیابی کارایی مدل درختان تصمیم در این تحقیق از چهار پارامتر آماری به شرح ذیل استفاده گردیده است:

رابطه ۵:

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

رابطه ۶ :

$$MAE=\frac{\sum|o_i-p_i|}{N}$$

رابطه ۷:

$$Bias=\overline{P}-\overline{O}$$

رابطه ۸:

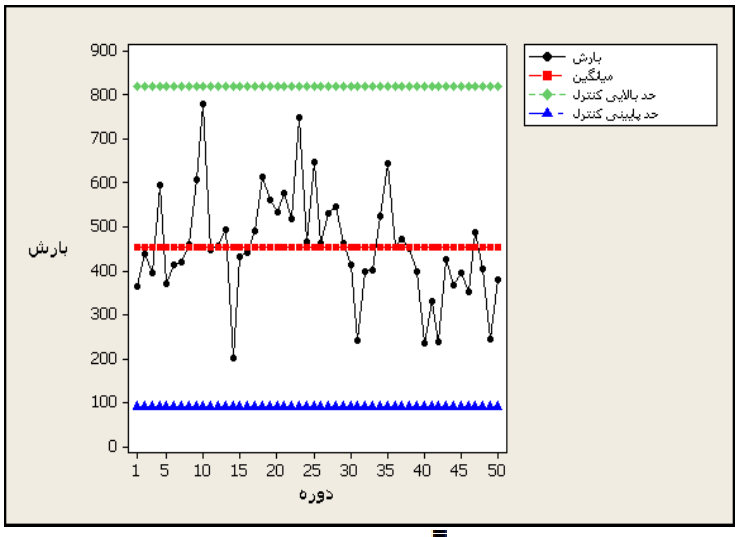
$$R=\sqrt{1-\frac{\sum_{i-1}^{n}(o_{i-}p_i)}{\sum_{i-1}^{n}(o_i-\overline{o})}}$$

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

**نتایج و بحث**

**کنترل کیفیت مشاهدات**

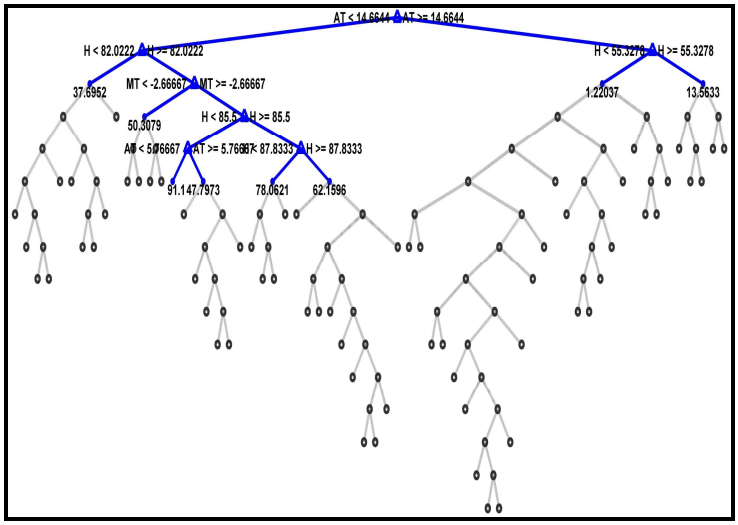
به منظور کنترل کیفیت آماری مشاهدات (داده‌های پرت، مفقوده) از نمودار $\bar{\bar{X}}$ استفاده شده است. با توجه به نمودار مشاهدات بین دو حد، سه برابر انحراف معیار در بالا و پایین میانگین نوسان دارند و با توجه به نمودار در بارش سالانه مشاهده پرت دیده نمی‌شود.




**شکل ۲: نمودار کنترل $\bar{X}$ متوسط بارش سالانه ایستگاه سنندج**

## نتایج مدل درخت تصمیم در شرایط استفاده از اصل داده ها

در این حالت، مدل با ۵ متغیر مستقل شامل بارش، میانگین دما، حداکثر دما، حداقل دما، رطوبت نسبی، سرعت و جهت باد و ۵۱۶ سطر داده اجرا شده‌است. لازم به ذکر است که از این تعداد ۳۶۱ سطر در مرحلهٔ آموزش و ۱۵۵ سطر در مرحلهٔ آزمون استفاده گردیده‌است، ساختار درخت ایجاد شده در این حالت را در شکل (۳) مشاهده می‌شود:



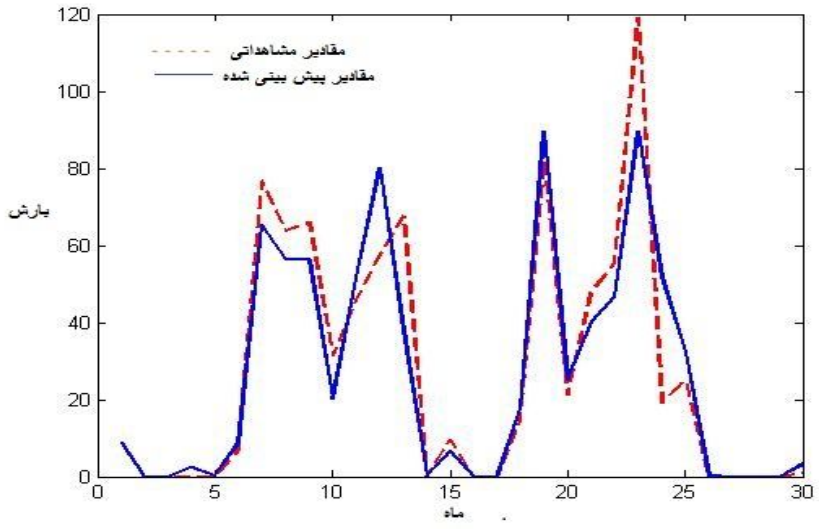

**شکل۳: ساختار درخت تصمیم ایجاد شده در حالت اول(استفاده از اصل داده‌ها)**

شکل (۳) نشان می‌دهد که مدل درخت تصمیم در حالت اول، ۵ پارامتری که در اختیار داشته: بارش، دمای متوسط، رطوبت نسبی، دمای ماکزیمم و سرعت باد، همهٔ پارامترها را اجرا کرده و درخت بر اساس این ۴ پارامتر عمل کرده است که مدل از پارامتر رطوبت نسبی در ساخت درخت تصمیم استفاده نکرده است. در ادامه، کارایی درخت تصمیم ایجاد شده با استفاده از چهار پارامتر آماری (R, RMSE, MAE, Bias) بررسی شده که مقادیر این پارامترها در جدول (۳) درج شده‌است. 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.

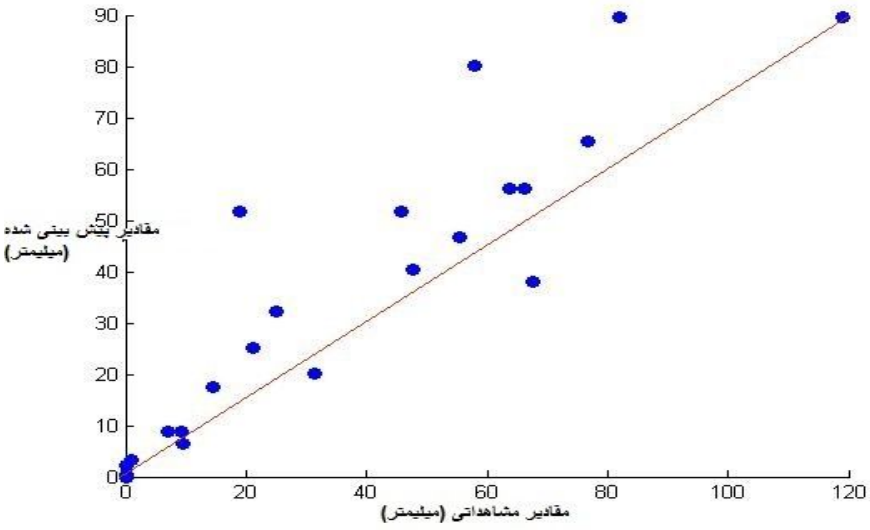
**جدول ۲: کیفیت نتایج در پیش بینی بارش با استفاده از داده های خام (اصل داده‌ها)**

| ردیف | متغیر مستقل | متغیر هدف | R | RMSE (mm) | MAE (mm) | Bias |
|---|---|---|---|---|---|---|
| حالت اول | P-Tmean-Tmax-Ws | Pn | ۰/۹۲ | ۴/۵۹ | ۲/۶۸ | ۰/۳۱ |

مدل ایجاد شدهٔ فوق، در پیش‌بینی مقادیر اوج از کارایی بسیار کمی برخوردار است. مدل مزبور در غالب بازه‌های زمانی ۱۲ ماهه، پیش‌بینی‌های خود را نزدیک به دو مقدار عددی حداقل و حداکثر انجام داده‌است؛ به عبارتی درخت ایجاد شده با استفاده از اصل داده‌ها روند پیش‌بینی نسبتاً ثابتی داشته‌است و نتوانسته تغییرات را به خوبی شبیه‌سازی نماید.



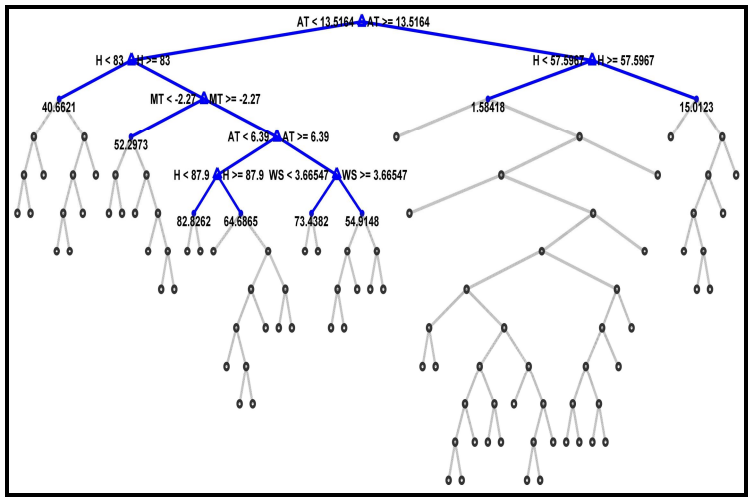

(الف)




(ب)

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

**نتایج مدل درخت تصمیم با استفاده از میانگین متحرک سه سالهٔ داده‌های اولیه:** در این حالت جهت اجرای مدل، ۵ متغیر مستقل شامل میانگین متحرک سه سالهٔ 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 مرحلهٔ آموزش و ۱۴۸ سطر در مرحلهٔ آزمون استفاده گردیده‌است. ساختار درخت ایجاد شده در این حالت در شکل (۵) نمایش داده شده است:



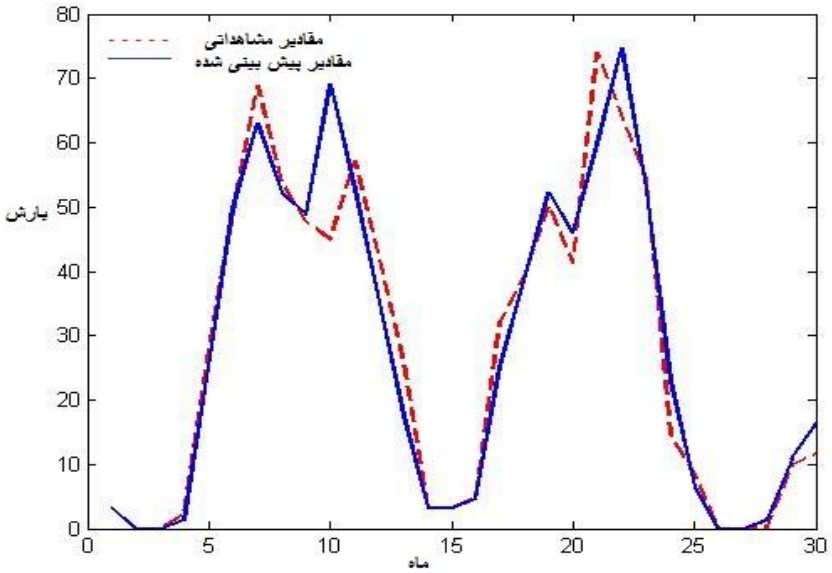

**شکل ۵: ساختار درخت تصمیم در حالت دوم (استفاده از میانگین متحرک سه ساله)**

پس از آن کارایی درخت ایجاد شده با استفاده از پارامترهای آماری مورد استفاده ارزیابی شد که مقادیر این پارامترها در جدول (۴) درج شده‌است. همان‌طور که در جدول (۳) مشاهده شد مقدار همبستگی نتایج مدل با استفاده از داده‌های خام ۰/۹۲ است؛ اما میزان همبستگی نتایج مدل با استفاده از میانگین متحرک سه سالهٔ داده‌ها ۰/۹۴ است بنابراین مدل با استفاده از میانگین متحرک سه ساله داده‌ها نسبت به داده‌های خام نتایج بهتری را نشان داده است.

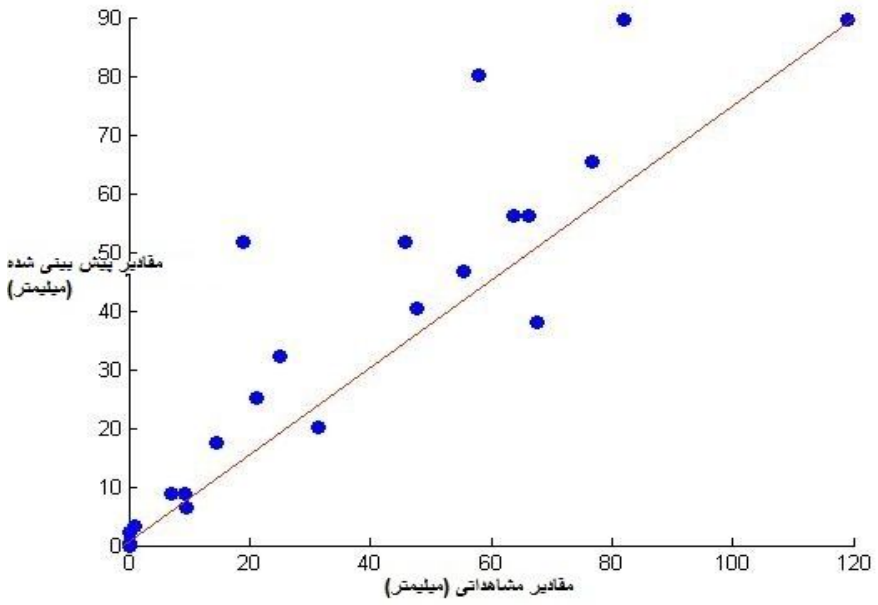
**جدول ۳: کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله داده ها**

| ردیف | متغییر مستقل (میانگین متحرک سه ساله) | متغییر هدف (میانگین متحرک سه ساله) | R | RMSE (mm) | MAE (mm) | Bias |
|---|---|---|---|---|---|---|
| حالت اول | P -Tmax-Ws | Pnm3 | ۰/۹۴ | ۲/۹۹ | ۱/۸۹ | ۰/۲۱ |

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 است که در شکل (۶) نیز نمایش داده شده است.




(الف)




(ب)

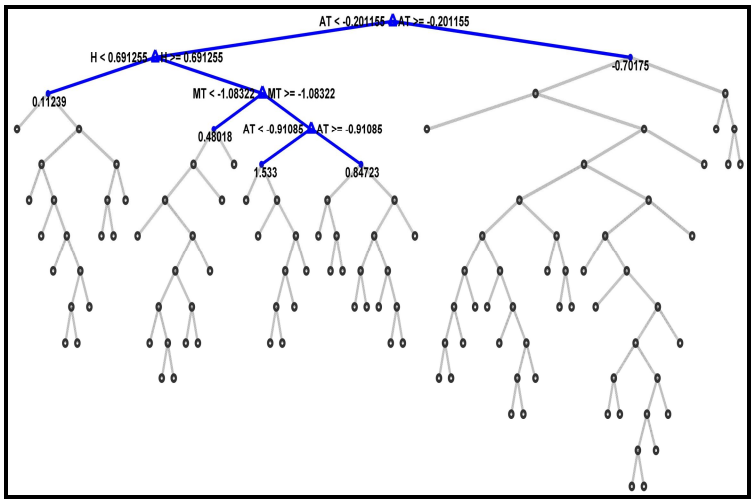
**شکل ۶: (الف و ب) تغییرات مقادیر پیش‌بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه سالهٔ داده‌ها**

## نتایج مدل با استفاده از میانگین متحرک سه سالهٔ نرمال

در این بخش میانگین متحرک سه سالهٔ داده‌ها نرمال شده 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 سالهٔ نرمال شدهٔ بارش، ۳۰ ماه قبل از وقوع پیش‌بینی گردیده است. درخت تصمیم ایجاد شده در این حالت در شکل (۷) نمایش داده شده است. همان‌طور که در جدول (۴) مشاهده می‌شود میزان همبستگی نتایج مدل با استفاده از میانگین متحرک سه سالهٔ نرمال شده ۰/۹۴ و همچنین مقدار خطای آن ۰/۰۶ است که نتایج بهتری را نشان داده است.

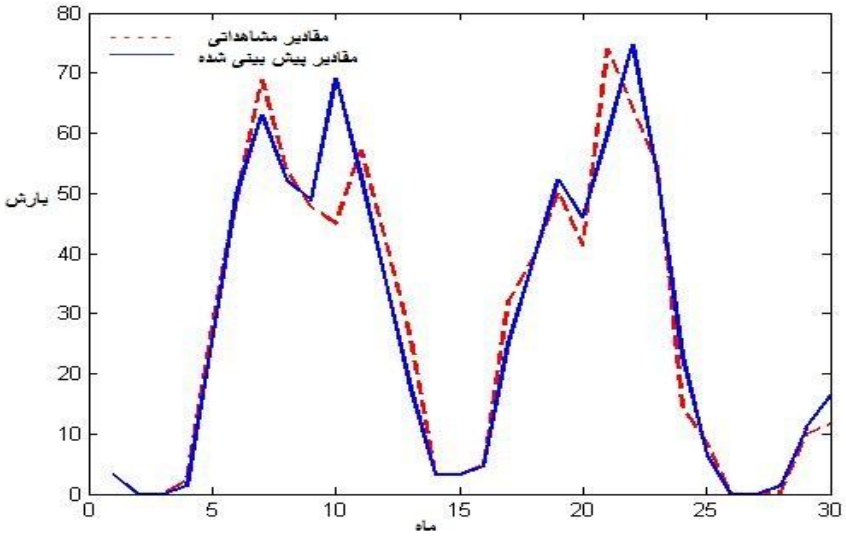
**جدول۴: کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله نرمال شده**

| 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 -Tmax-Ws | Pnm3 | ۰/۹۴ | ۰/۰۶ | ۰/۰۴ | ۰/۰۰۶ |

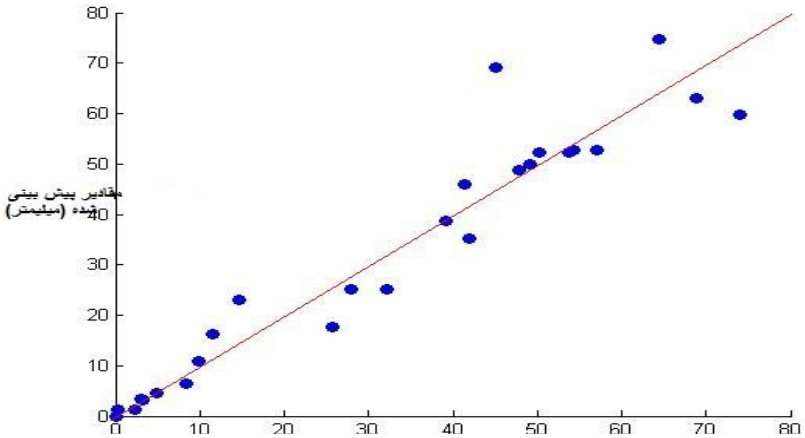



**شکل۷: درخت تصمیم در حالت چهارم(استفاده از میانگین متحرک ۵ ساله داده ها)**

جهت بررسی کارایی درخت ایجاد شده در شکل (۸) ( الف و ب) نیز همانند حالت قبلی از پارامترهای آماری استفاده گردید؛ همان‌طور که در نمودارها مشاهده می‌شود مقادیر مشاهداتی و مقادیر پیش‌بینی شده از همبستگی بالایی برخوردار هستند. مقادیر این پارامترها برای این حالت در جدول( ۵) آمده است.



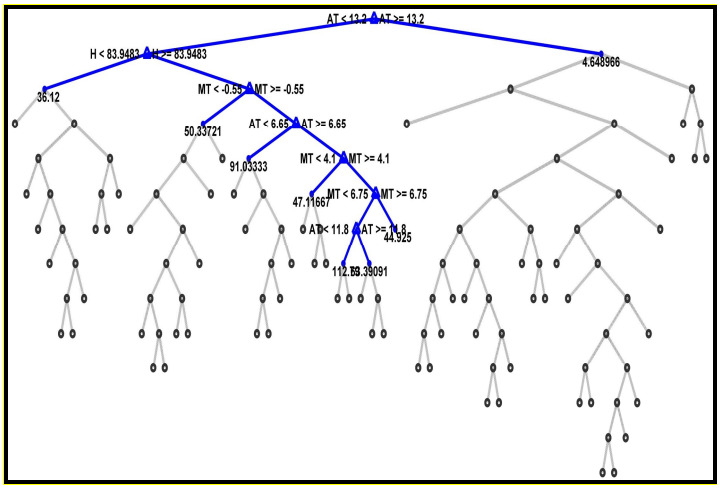

(الف)




(ب)

**شکل۸: تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه ساله نرمال شده داده‌ها**

**نتایج مدل در شرایط استفاده از میانگین متحرک پنج سالهٔ داده‌های اصلی**: در ادامه جهت بررسی و بالا بردن کارایی مدل در پیش‌بینی بارش، از میانگین متحرک پنج سالهٔ داده‌ها نیز استفاده گردیده است که متغیر مستقل شامل میانگین متحرک پنج سالهٔ داده‌های بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی و سرعت باد به مدل معرفی و میانگین متحرک پنج سالهٔ بارش برای سال بعد پیش‌بینی شده است. در این حالت ۴۸۴ سطر داده (با توجه به تشکیل میانگین متحرک ۵ سالهٔ تعداد داده‌ها اندکی کمتر شده است) و 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 است. درخت ایجاد شده در این حالت در شکل (۹) نمایش داده شده است.



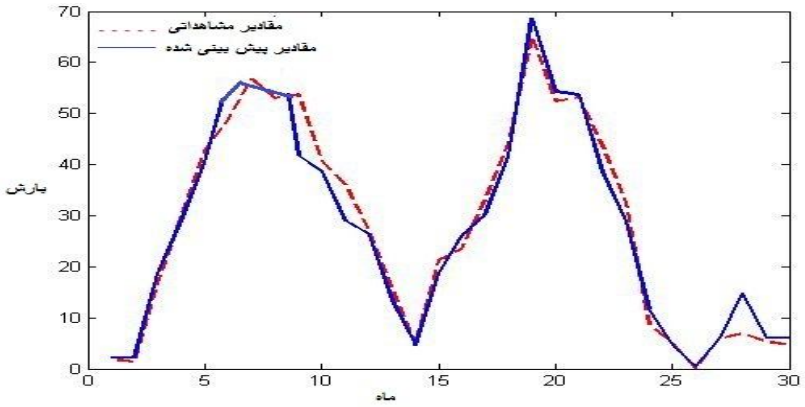

**شکل۹: درخت تصمیم در حالت چهارم (استفاده از میانگین متحرک ۵ ساله داده ها)**

مقادیر پارامترهای آماری مربوط به ارزیابی نتایج درخت ایجاد شده در این حالت و نیز متغیرهای مستقل استفاده شده در جدول (۶ ) آمده است.

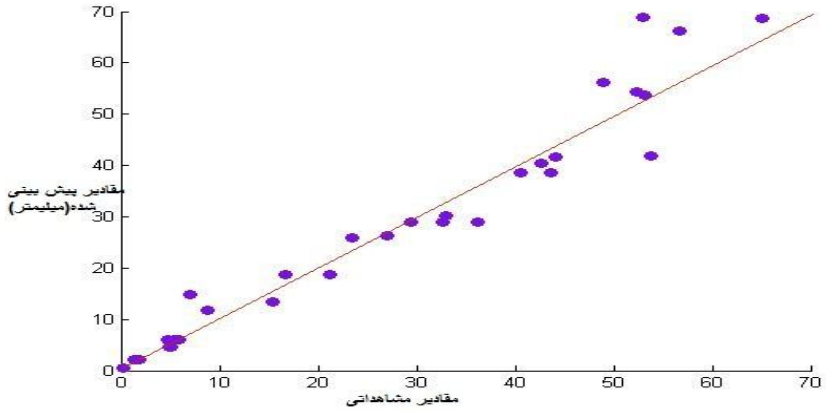
**جدول ۵: کیفیت نتایج مدل با استفاده از میانگین متحرک پنج ساله**

| ردیف | متغییر مستقل (میانگین متحرک پنج ساله) | متغییر هدف (میانگین متحرک پنج ساله) | R | RMSE (mm) | MAE (mm) | Bias |
|---|---|---|---|---|---|---|
| ۱ | P -Tmax-Ws | Pnm5 | ۰/۹۶ | ۱/۶۳ | ۰/۱۰ | ۰/۰۵ |

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 شده‌است:




(الف)




(ب)

**شکل ۱۰: (الف و ب) تغییرات مقادیرپیش بینی شده در مقابل مقادیر واقعی با استفاده از داده های میانگین متحرک پنج ساله داده ها**

## نتیجه گیری

نوسانات بارش، جزو روش‌های عمومی تحلیل خشکسالی به شمار می‌رود؛ لذا پیش‌بینی دقیق و پیش از وقوع بارش می‌تواند شرایط را برای ارزیابی وضعیت خشکسالی فراهم نماید. هدف این پژوهش، بررسی تأثیر پیش‌پردازش داده‌ها بر عملکرد مدل درخت تصمیم در پیش‌بینی خشکسالی در ایستگاه سینوپتیک سنندج است. در 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 برای انجام این پژوهش در راستای اهمیت موضوع و با هدف ارزیابی کارایی مدل درخت تصمیم‌گیری در برآورد خشکسالی ایستگاه سینوپتیک سنندج از آمار ماهیانهٔ بارندگی، رطوبت نسبی، دمای ماکزیمم، دمای متوسط، جهت و سرعت باد در دورهٔ آماری (۱۳۸۹- ۱۳۴۹) استفاده شده‌است. نتایج حاصل از این پژوهش در تمام شبیه‌سازی‌های صورت گرفته، در پارامترهای بارش پیشین و دمای ماکزیمم دخالت داشته‌است که این موضوع نشان دهندهٔ اهمیت این پارامترها در پیش‌بینی خشکسالی در آینده است. بر اساس نتایج به دست آمده، می‌توان گفت درخت تصمیم‌گیری رگرسیونی در ایستگاه سینوپتیک سنندج، مدلی نسبتاً کارا در پیش‌بینی خشکسالی با استفاده از داده‌های پردازش شده می‌باشد که مطالعات صورت گرفته توسط گیسن و همکاران (۲۰۰۷)، روس جان و میکوس (۲۰۰۸)، باتاچاریا و همکاران (۲۰۱۲)، محجوبی و اعتماد شهیدی (۱۳۸۶)، دستورانی و همکاران (۱۳۹۱)، طالبی و اکبری (۱۳۹۲) نیز این نتیجه را تأیید نموده است. در این پژوهش نیز استفاده از میانگین متحرّک نسبت به سایر حالات منجر به افزایش چشمگیر کارایی مدل درخت تصمیم می‌شود و در صورت تعدیل دامنهٔ تغییرات، داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان بارش را ۱۲ ماه قبل از وقوع برآورد نماید. در شبیه‌سازی‌های صورت گرفته در این پژوهش، زمانی که از میانگین متحرّک پنج سالهٔ داده‌ها برای اجرای مدل استفاده گردیده است، ترکیب بارش قبلی و دمای ماکزیمم به عنوان مناسب‌ترین حالت شناسایی شده است که مقدار خطای آن ۰/۰۶ است. همچنین درخت ایجاد شده در حالت میانگین متحرک پنج سالهٔ داده‌ها قادر است بارش ایستگاه سنندج را ۱۲ ماه قبل از وقوع پیش‌بینی نماید. در ایستگاه سینوپتیک سنندج، به نظر می‌رسد دامنهٔ تغییرات داده‌ها برای استفاده در مدل درخت تصمیم زیاد می‌باشد و همین موضوع منجر به کاهش کارایی مدل در پیش‌بینی می‌گردد؛ به همین دلیل می‌توان گفت که مدل درخت تصمیم‌گیری برای پیش‌بینی خشکسالی در شرایط استفاده از اصل داده‌ها ناموفق عمل کرده و جهت بهتر شدن کارایی مدل پیش‌بینی، از داده‌های میانگین متحرک استفاده شده است.

## منابع


- افخمی، حمیده، محمدتقی دستورانی، حسین ملکی نژاد، محمدحسین مبین (۱۳۸۹). بررسی کارایی شبکهٔ عصبی مصنوعی و پارامترهای اقلیمی در پیش‌بینی خشکسالی در منطقهٔ یزد، علوم آب و خاک - علوم و فنون کشاورزی و منابع طبیعی، دوره (۱۴)، شماره ۵۱، صص: ۱۵۷-۱۶۹.
- پیری، حلیمه، عباس‌زاده، محبوبه، راهداری، وحید، ملکی، سعیده (۱۳۹۲). ارزیابی تطبیقی ۴ نمایه خشکسالی هواشناسی با استفاده از روش تحلیل خوشه‌ای مطالعه موردی استان سیستان و بلوچستان، مهندسی منابع آب، سال۶، شماره ۱۷، صص: ۲۵-۳۶.
- حسن زاده، یوسف، عبدی کردانی، امین، فاخری فرد، احمد (۱۳۹۱). پیش‌بینی خشکسالی با استفاده از الگوریتم ژنتیک و مدل ترکیبی شبکه عصبی – موجکی، فصلنامه آب و فاضلاب، دوره ۲۳، شماره ۳، صص: ۴۸-۵۹.
- دستورانی، محمد تقی، حبیبی پور، اعظم، اختصاصی، محمد رضا، طالبی، علی، محجوبی، جواد (۱۳۹۱). بررسی کارایی مدل درخت تصمیم در پیش بینی بارش (مطالعهٔ موردی ایستگاه سینوپتیک یزد)، فصلنامهٔ تحقیقات منابع آب ایران، سال هشتم، دوره ۸، شماره ۳، صص: ۱۴-۲۳.
- رضیئی، طیب، دانش‌کار آراسته، پیمان، اختری، روح‌انگیز، ثقفیان، بهرام (۱۳۸۶). بررسی خشکسالی هواشناسی در استان سیستان و بلوچستان با استفاده از نمایه SPI و مدل زنجیره مارکف، مجله تحقیقات منابع آب ایران، دوره۳، شماره ۱، صص: ۲۵- ۳۵.
- شایق، محمد علی (۱۳۹۰). ارزیابی پروژه‌های باروری ابرها با استفاده از مدل درختان تصمیم‌گیری رگرسیونی (مطالعه موردی: ایران مرکزی- استان فارس)، پایان‌نامه کارشناسی ارشد مهندسی عمران آب، دانشگاه صنعت آب و برق، تهران.
- طالبی، علی، اکبری، زینب (۱۳۹۲). بررسی کارایی مدل درختان تصمیم‌گیری 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.
- فرجزاده اصل، منوچهر (۱۳۸۴). خشکسالی، از مفهوم تا راهکار ، انتشارات سازمان جغرافیایی نیروهای مسلح.
- کانتاردزیک، م (۱۳۸۵). داده‌کاوی، ترجمه: امین علیخان زاده، نشر علوم رایانه، بابل، چاپ اول.
- کردوانی، پرویز (۱۳۸۰). خشکسالی و راهکارهای مقابله با آن در ایران، انتشارات دانشگاه تهران، چاپ دوم.
- کشاورز، محمدرضا، وظیفه‌دوست، مجید، فتاحی، ابراهیم، بهیار، محمدباقر (۱۳۹۱). الگوی توزیعی جهت و شدت روند تغییرات خشکسالی ایران به کمک نمایه توزیعی شدت خشکسالی پالمر، نشریه تحقیقات کاربردی علوم جغرافیایی، سال۱۲، شماره ۲۷، صص: ۹۷- ۱۱۰.
- محجوبی، جواد، اردلان صمغی، حسین (۱۳۸۸). پیش‌بینی 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: ۶۵- ۷۱.
- مشکانی، علی، ناظمی، عبدالرضا (۱۳۸۸). مقدمه‌ای بر داده‌کاوی، مشهد موسسه چاپ و انتشارات دانشگاه فردوسی، چاپ اول.
- Bhattacharya, B., R.K. Price and D. P. Solomatine. (2007), Machine learning approach to modeling sediment transport. J. Hydraul. Eng. 133(4), Pp 440-450.
- B reiman, L., Friedman J., Olshen R., and 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 .

- Santos, C. A. G., Morais, B. S., & Silva, G. B. (2009), Drought forecast using an artificial neural network for three hydrological zones in San Francisco River basin, Brazil. IAHS publication, 333, 302.
- Mahesh P. and Mather P. M., (2003), an assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86, Pp. 554–565.
- Geissen, V., Kampichler, C., López-de Llergo-Juárez, J. J., & Galindo-Acántara, A. (2007), Superficial and subterranean soil erosion in Tabasco, tropical Mexico: development of a decision tree modeling approach. Geoderma, 139(3), Pp.277-287.
- Kocev,D., Saso D., White M. D., Newell G. R. and 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.
- Rusjan, S. and M. Micos. (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: 645-656.
- Yurekli, K., Taghi Sattari, M.T., Anli, A.S. and Hinis, M.A. (2012), Seasonal and annual regional drought Prediction by using data-mining approach, Atmosfera, 25(1) Pp. 85-105.
- Wilhite, D.A. ed., (2012), Drought Assessment, Management, and Planning: Theory and Case Studies: Theory and Case Studies (Vol. 2). Springer Science & Business Media.