تحقیقات منابع آب ایران
Iran-Water Resources Research

سال هشتم، شماره ۳، زمستان ۱۳۹۱
Volume 8, No. 3, Winter 2013 (IR-WRR)
۱۴-۲۷

بررسی کارایی مدل درخت تصمیم در پیش بینی بارش (مطالعه موردی ایستگاه سینوپتیک یزد)

محمدتقی دستورانی[۱]*، اعظم حبیبی‌پور[۲]،
محمدرضا اختصاصی[۳]، علی طالبی[۳] و جواد محجوبی[۴]

چکیده

وقوع خشکسالی اثرات نامطلوبی بر بخش‌های کشاورزی و اقتصادی کشور و به طور خاص بر عرصه‌های طبیعی تحمیل می‌کند. امروزه روش‌ی مختلفی جهت پیش‌بینی مؤلفه‌های اصلی خشکسالی از جمله بارش ارائه شده است. در دهه‌های اخیر، استفاده از مدل‌های جدید کامپیوتری در این زمینه رواج یافته و در اغلب موارد توانایی خود را به خوبی نشان داده است. درخت تصمیم به عنوان یکی از این نوع مدل‌ها، با بررسی پارامترها از جزء به کل، به تولید قانون می‌پردازند و نهایتاً به دانش قابل فهم از بین داده‌های آماری موجود دست می‌یابند. در این تحقیق از مدل درخت تصمیم به عنوان یکی از روش‌های داده کاوی جهت پیش‌بینی بارش و ارزیابی وضعیت خشکسالی درایستگاه سینوپتیک یزد استفاده شد. شبیه‌سازی‌ها در چهار حالت صورت گرفت و در کلیه شبیه‌سازی‌ها از متغیرهای بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت، سرعت باد، جهت باد در مقیاس ماهانه به عنوان متغیرهای مستقل مدل استفاده و میزان بارش ۱۲ ماه قبل از وقوع پیش‌بینی گردید. در نهایت جهت ارزیابی دقت و صحت درخت‌های ایجاد شده در حالات فوق، معیارهای آماری مختلف مورد استفاده قرار گرفت. نتایج نشان داد که در ایستگاه یزد، مدل درخت تصمیم گیری خصوصاً در شرایطی که از میانگین متحرک ۵ ساله داده‌ها استفاده گردد، دارای توانایی مناسبی در پیش‌بینی میزان بارش می‌باشد. پیش‌بینی مقدار بارندگی و به تبع آن ارزیابی وضعیت خشکسالی با دقت مناسب و قبل از وقوع آن، می‌تواند به برنامه‌ریزی جهت کاهش خسارات حاصل از آن کمک قابل توجهی نماید.

کلمات کلیدی: مدل‌سازی، درخت تصمیم، یزد، داده کاوی، بارندگی، خشکسالی.

تاریخ دریافت مقاله: ۱۹ بهمن ۱۳۸۹
تاریخ پذیرش مقاله: ۱۰ آبان ۱۳۹۱

۱- عضو هیئت علمی دانشکده منابع طبیعی دانشگاه یزد، در حال حاضر مامور در دانشکده منابع طبیعی و محیط زیست دانشگاه فردوسی مشهد- یزد- ایران
۲- دانشجوی دکترای منابع طبیعی – آبخیزداری- دانشگاه یزد- یزد- ایران.
۳- عضو هیئت علمی دانشکده منابع طبیعی- دانشگاه یزد- یزد- ایران.
۴- کارشناس ارشد عمران- سازه های هیدرولیکی- دانشگاه علم وصنعت- تهران- ایران
*- نویسنده مسئول

Evaluation of the Decision Tree Model in Precipitation Prediction (Case study: Yazd Synoptic Station)

M. T. Dastourani [1], A. Habibipoor [2], M. R. Ekhtesasi[3], A. Talebi[3] and* J. Mahjoobi [4]

Abstract

Undesirable effects of droughts on the agricultural and economical sectors and especially on the natural resources are intense. Different methods have been presented to predict the main factors of drought such as precipitation and during the recent decades some new computer based models have been developed for drought prediction. In most cases these models have presented quite satisfactory results. Decision tree, as one of these models, produces rules based on evaluation of the parameters from portion (component) to the whole, and finally reaches understandable knowledge from the existing statistical data. In this research, decision tree model has been used as a data mining method to predict precipitation and evaluation of drought in Yazd synoptic meteorological station. Simulations were carried out in four different conditions. Related variables including previous monthly precipitation, mean temperature, maximum temperature, humidity, wind speed, wind direction, and evaporation were used as independent input variables for all these four conditions and the amount of precipitation was predicted 12 months in advance. Finally for evaluation of the model performance in different conditions, statistical criteria were employed. Results indicated that the decision tree model is able to presents suitable prediction of precipitation especially when 5-year moving average of data is used. Precise prediction of precipitation and the accurate evaluation of drought conditions are of great importance for a better management and planning for drought damages reduction.

Keywords: Modelling, Decision tree, Yazd, Data mining, Precipitation, Drought.

Received: February 8, 2011
Accepted: October 31, 2012

1- Lecturer, Faculty of Natural Resources, Yazd University, Yazd, iran, Email: mdastorani @yazduni.ac.ir
2- PhD student of Watershead Management Engineering, Yazd University, Yazd, iran
3-Lecturer, Faculty of Natural Resources, Yazd University, Yazd, iran
4- Msc., Civil Engineering- Water Engineering, Iran University of Science and Technology, Tehran, Iran
*- Corresponding Author

۱- مقدمه

کشور ایران 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 گریزناپذیر تبدیل نموده است. کمبود بارش نسبت به میانگین دراز مدت یک منطقه به عنوان مؤلفه اصلی رخداد خشکسالی محسوب می‌گردد. خشکسالی یک شکل طبیعی از اوضاع آب و هوایی است که تکرار آن در طول زمان اجتناب ناپذیر می‌باشد. در بین سال‌های ۱۹۹۲-۱۹۶۷ حدود ۱/۳ میلیون نفر قربانی پدیده خشکسالی بودند (Obasi, 1994). حدود ۵۰ درصد از نواحی پرجمعیت جهان در برابر خشکسالی آسیب‌پذیرند (USDA, 1994). هرچند تعاریف مختلفی برای خشکسالی ارائه شده است، ولی در هر حال پیش‌بینی مقادیر بارندگی در ماه‌ها و فصول آتی، اصلی ترین ابزار جهت ارزیابی وضعیت خشکسالی بالقوه بوده و اطلاع از خصوصیات خشکسالی می‌تواند کمک شایانی به برنامه ریزی جهت تخفیف خسارات آن نماید. هر چند روش‌های مختلفی برای مطالعه وضعیت خشکسالی وجود دارد، روش تحلیل داده‌های بارندگی جزء عمومی‌ترین روش‌های تحلیل خشکسالی به شمار می‌رود. علت این امر نیز دسترسی راحت‌تر و آسان‌تر به انواع داده‌های بارندگی در بخش‌های مختلف کره زمین است. از طرفی مقادیر بارندگی جزء بی‌ثبات‌ترین متغیرهای آب و هوایی مخصوصاً در مناطق خشک محسوب می‌شود و از این جهت شاخص خوبی برای مطالعه خشکسالی می‌باشد. در حقیقت بارش‌های جوّی مهم‌ترین متغیری است که تغییرات آن به طور مستقیم در رطوبت خاک و جریان‌های سطحی، مخازن زیرزمینی و غیره منعکس شده و بنابراین اولین مولفه‌ای است که می‌تواند در مطالعه هر حالتی از خشکسالی مورد توجه قرار گیرد. امروزه حجم عظیمی از داده‌ها توسط سازمان‌ها و ارگان‌های متولی امر در مقیاس‌های زمانی مختلف ثبت می‌گردد. نظر به گسترش سیستم پایگاه داده و ایجاد ابزارهای متعدد برای ذخیره حجم بالایی از اطلاعات، داده کاوی به عنوان یکی از شاخه‌های علوم بین رشته ای توسعه روزافزونی یافته است. داده‌کاوی به فرایند جستجو و کشف مدل‌های گوناگون، مختصرسازی و اخذ مقادیر از مجموعه‌ای از داده‌های معلوم اطلاق می‌گردد (کانتاردزیک، ۱۳۸۵). دسته‌بندی و پیش بینی دو نوع عملیات داده کاوی برای تحلیل داده ها و استخراج مدل به منظور توصیف دسته‌های مهم داده‌ها، فهم و پیش‌بینی رفتار آن‌ها در آینده است. مدل‌های دسته بندی در تحلیل داده‌های گسسته و طبقه‌ای به کار رفته و مدل‌های پیش‌بینی یا رگرسیون بیشتر بر روی داده‌های پیوسته کار می‌کنند (مشکانی و ناظمی، ۱۳۸۸). درختان تصمیم[1] به همراه قوانین تصمیم یکی از روش های داده‌کاوی به شمار می‌آیند (کانتاردزیک، ۱۳۸۵). درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته بندی و پیش‌بینی 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 داده ها لزوماً به صورت عددی باشند (مشکانی و ناظمی، ۱۳۸۸). تحقیقات متعددی بر روی کارایی درخت تصمیم در زمینه علوم طبیعی صورت گرفته است. Mahesh and Mather, (2003) از مجموعه داده‌های آموزش و آزمون جدا از دو منطقه جغرافیایی و با دو سنجنده مختلف برای ارزیابی توانایی درخت تصمیم تک‌متغیره و چندمتغیره در طبقه‌بندی پوشش زمین استفاده کردند. نتایج نشان داد که درخت تصمیم تک‌متغیره در مقایسه با دیگر طبقه‌بندی‌کننده‌ها، به جز برای داده‌های با ابعاد بالا، به طور قابل قبولی بهتر اجرا شد. اما هیچ کدام از مدل‌های درخت تصمیم تک‌متغیره و چندمتغیره برای داده‌های با ابعاد بالا، به اندازه طبقه‌بندی‌کننده‌های مدل شبکه عصبی مصنوعی خوب اجرا نشدند. Bhattacharya et al., (2007) از دو روش یادگیری ماشینی شامل شبکه عصبی مصنوعی و مدل‌های درخت تصمیم، جهت مدل‌سازی انتقال بار کل و بار کف رودخانه استفاده کردند و نتایج مدل‌های انتقال بار کف را با مدل‌های Bagnold (1941)، Einstein (1942)، Parker et al (1992) و همچنین نتایج مدل‌های انتقال بار کل را با مدل‌های Bagnold (1941)، Engelhund and Hansun (1967) مقایسه کردند. 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 داد. Giessen et al., (2007) فرسایش آبی سطحی و زیرسطحی را با مدل درخت تصمیم در تاباسکو (جنوب‌شرقی مکزیک) مورد ارزیابی قرار دادند. نتایج نشان داد که مدل وقوع فرسایش‌های تونلی و گودالی را با موفقیت پیش‌بینی می‌کند. گیسن و همکاران همچنین بیان کردند که مدل درختان تصمیم استقرایی، ابزاری ارزشمند جهت تجزیه و تحلیل‌های مقدماتی و ایجاد فرضیات در مناطق با کمبود اطلاعات است و می‌تواند نقشه خطر فرسایش و طرح حفاظت خاک را مدیریت کند. Rusjan and Micos (2008) از الگوریتم درخت تصمیم رگرسیونی M5 برای مشاهده نقش متقابل وضعیت‌های فصلی و هیدرولوژیکی که آب‌شویی نیترات را در طول وقایع هیدرولوژیکی در یک حوزه جنگلی واقع در جنوب‌غربی اسلوانی کنترل می‌کند استفاده کردند. آن‌ها بر اساس معیار هرس درخت رگرسیونی از پیش تعریف شده،

یک مدل رگرسیونی قابل فهم متناسب با اطلاعات حوزه که به اندازه کافی قادر به شرح غلظت‌های مختلف نیترات باشد به دست آوردند. Kheir et al., (2008) از مدل درخت تصمیم لینک شده با GIS برای پیش‌بینی توزیع خاک و سنگ بستر مستعد فرسایش خندقی (گالی) در منطقه‌ای در لبنان استفاده کردند و نتیجه گرفتند که این مدل می‌تواند برای مناطق دیگر مخصوصاً زمانی که اطلاعات خاک و سنگ منطقه محدود باشد استفاده شود. Chen et al., (2008) از مدل درختان تصمیم به همراه شبکه‌های عصبی وصنوعی جهت تحلیل وقوع سیلاب در حوزه‌ای در تایوان استفاده کردند و نتایج حاصله را موفقیت آمیز توصیف نمودند. Cheng et al., (2008) جهت مدیریت مخزن سد در بهینه‌سازی مقدار دبی تخلیه‌ای از مجزن از تکنیک درختان تصمیم استفاده کردند. در مخزن مورد نظر که هم کنترل سیلاب و هم ذخیره آب مورد نیاز اهمیت ویژه‌ای داشته مدیریت مناسب مخزن از نظر میزان تخلیه بسیار مهم بوده و محققین مربوطه کارایی مدل درختان تصمیم را در این رابطه مورد تایید قرار داده‌اند. Kocev et al., (2009) از روش‌های یادگیری ماشینی مختلف (درختان تصمیم تک‌منظوره و چندمنظوره) برای پیش‌بینی وضعیت یا کیفیت گیاهان بومی باقی مانده در منطقه وسیعی در شمال‌شرق استرالیا– ایالت ویکتوریا استفاده کردند. بررسی نتایج این مطالعه نشان داد که مدل‌سازی چندمنظوره بر تک‌منظوره برتری دارد، هر چند تفاوت مهمی از نظر آماری بین این دو مدل در اجرا وجود نداشت. Vega et al., (2009) جذب و نگهداری فلزات سنگین خاک (سرب، مس، کرم، کادمیوم، نیکل و روی) را از طریق یک نمونه معرف از خاک‌های گالیسا (شمال‌غربی اسپانیا) با دقت قابل ملاحظه ای به وسیله مدل درختان تصمیم‌گیری رگرسیونی دودویی، که به وسیله الگوریتم کارت[۲] ساخته شده‌اند، به جای مدل‌های رگرسیونی خطی بازسازی کردند. از بین شش فلز رقابت‌کننده در مناطق جذب، در این آزمایشات سرب، مس و کرم به مقدار بیشتری نسبت به کادمیوم، نیکل و روی جذب و حفظ شدند. مخصوصاً برای کادمیوم، نیکل و روی؛ مدل‌های درختان تصمیم‌گیری رگرسیونی غیرخطی، که به وسیله الگوریتم کارت ساخته شدند، نسبت به مدل‌های خطی تطابق بیشتری با داده‌ها نشان دادند. علاوه بر آن سرب، مس و کرم نسبت به کادمیوم، نیکل و روی؛ با هر دو نوع مدل داده‌ها تطابق بهتری داشتند (اختلاف برای مدل‌های خطی بیشتر نمایان بود).

اکبری (۱۳۸۹)، به بررسی کارایی درختان تصمیم در برآورد رسوب حوزه سد ایلام پرداخت و سپس نتایج این روش با منحنی سنجه رسوب مقایسه گردید. تحقیق مزبور نشان داد که در تمام شبیه‌سازی‌ها دقّت مدل درخت تصمیم‌گیری نسبت به روش منحنی سنجه بالاتر بوده است. Ayoubloo et al., (2010)، از مدل درخت تصمیم و شبکه عصبی مصنوعی برای تخمین عمق آبشستگی پای پل‌ها استفاده کردند. 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 کردند. در تحقیق آن‌ها، شبکه عصبی مصنوعی آموزش دیده با پارامترهای دارای بعد نتایج بهتری در مقایسه با پارامترهای بدون بعد ارائه نمود، ولی مدل درخت تصمیم مبتنی بر ورودی‌های بدون بعد، عمق کنش نرمال شده را بهتر پیش‌بینی کرد. آن‌ها براساس نتایج تحقیق خود یک مدل ترکیبی مبتنی بر مفاهیم ژئومتریک ارائه نمودند. براساس آنالیز حساسیت انجام شده توسط آن‌ها، پارامتر *KC* (نیروی دراگ به نیروی اینرسی) به عنوان مهمترین پارامتر مدل شناسایی گردید. شایق (۱۳۹۰) از مدل درخت تصمیم رگرسیونی جهت ارزیابی پروژه‌های باروری ابرها در استان فارس استفاده کرد.

در این راستا شبیه‌سازی‌ها در چهار حالت صورت گرفت. نتایج نشان داد که مدل درخت تصمیم با داده‌های خام و آموزش تصادفی بهتر آموزش دیده و پیش‌بینی‌های دقیق‌تری را از حجم بارش مناطق هدف ارائه می‌دهد. Sattari et al., (2012) از مدل درخت تصمیم جهت تجزیه و تحلیل تداوم خشکسالی‌ها در آنکارا در ترکیه استفاده نمودند. در این تحقیق از داده‌های ماهانه و سالانه بارش، دما، باد و رطوبت جهت ارزیابی وضعیت خشکسالی استفاده گردید که اهمیت بارش (برای تمام ماهها) و باد (برای ماه ژانویه) به عنوان مهمترین پارامترها در این خصوص معرفی گردیدند. Yurekli et al., (2012) از تکنیک درخت تصمیم جهت ارزیابی و پیش‌بینی طبقات خشکسالی در حوزه‌ای در ترکیه استفاده کردند و نتایج حاصله را با روش طبقه‌بندی و تحلیل خشکسالی براساس شاخص بارش استاندارد (SPI) مقایسه نمودند. مقایسه نتایج تفاوت قابل ملاحظه‌ای را در طبقه بندی خشکسالی بر اساس این دو روش نشان نداد.

هدف از پژوهش حاضر بررسی کارایی مدل درخت‌های تصمیم‌گیری در پیش‌بینی مقدار بارش می‌باشد و برابر بررسی‌های انجام شده، تاکنون هیچ‌گونه تحقیقی در زمینه پیش‌بینی بارش با استفاده از مدل درخت‌های تصمیم‌گیری صورت نگرفته است. منطقه تحقیق یزد می‌باشد که از آمار و اطلاعات ایستگاه سینوپتیک یزد استفاده شده است. با توجه به واقع شدن ایستگاه در یک منطقه خشک و بیابانی و تغییرات قابل توجه داده‌های بارش، پیش بینی مقادیر آن البته کار ساده‌ای نبوده و اغلب مدل‌ها در انجام آن توانایی محدودی دارند.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه و آماده‌سازی داده‌ها

ایستگاه سینوپتیک یزد با بالغ بر ۵۰ سال آمار، دارای بلندترین طول دوره آماری در بین ایستگاه‌های استان یزد می‌باشد. داده‌های مورد استفاده در این تحقیق مربوط به آمار ماهیانه بارندگی، تبخیر، رطوبت نسبی، دمای ماکزیمم، دمای متوسط، سرعت باد و جهت باد غالب در ایستگاه یزد می‌باشد. جدول ۱ مشخصات اصلی ایستگاه سینوپتیک یزد را نشان می‌دهد.

قبل از استفاده از داده‌های مورد نظر، اقداماتی جهت آماده‌سازی داده‌ها صورت گرفت تا مشکلی از نظر همگنی، صحت و نقص داده‌ها وجود نداشته باشد. در ابتدا جهت کنترل کیفیت آمار و اطلاعات موجود از آزمون توالی استفاده شد. نتایج نشان داد که داده‌های مزبور از همگنی لازم برخوردار می‌باشند. در قدم بعدی جهت برطرف کردن نواقص آماری داده ها از روش نسبت نرمال استفاده گردید. برای شناسایی ایستگاه‌هایی که دارای بیشترین ضریب همبستگی با ایستگاه یزد هستند، ماتریس همبستگی بین ایستگاه یزد و ۱۳ ایستگاه دیگر استان یزد که دارای طول دوره آماری نسبتا مناسبی بودند تشکیل و سپس معنی‌دار بودن نتایج در سطح اعتماد ۱ درصد و ۵ درصد آزمایش شد. جدول ۲ ماتریس همبستگی بین ایستگاه‌ها را نمایش می‌دهد. ردیف‌های ۱ تا ۱۴ به ترتیب ایستگاه‌های ابرکوه، اردکان، مزرعه‌آقا خرانق، مزرعه‌نو عقدا، ساغند، حاج‌آباد زرّین، قطروم، باجگان، پاچنار تفت، دهشیر، گاریز، نصرآباد، یزد و خویدک را نشان می‌دهد. در مجموع با توجه به سطح معنی داری؛ برای آماده‌سازی داده‌های ایستگاه یزد از اطلاعات و آمار ایستگاه‌های ابرکوه، اردکان، مزرعه‌آقا خرانق، مزرعه‌نو عقدا، ساغند، حاج‌آباد زرّین، قطروم، باجگان، پاچنار تفت، نصرآباد، یزد و خویدک استفاده گردید.

همانگونه که قبلاً ذکر شد در این تحقیق، نواقص آماری با استفاده از روش نسبت نرمال بازسازی شده است. در این روش ایستگاه ناقص با استفاده از آمار ایستگاه‌هایی که همبستگی آنها در سطح اعتماد ۱٪ معنی‌دار است بازسازی می‌شود. به این منظور از رابطه (۱) استفاده گردید.

جدول ۱- مشخصات ایستگاه هواشناسی یزد

نام ایستگاه	نوع ایستگاه	سال تاسیس	موقعیت			
			شهرستان	ارتفاع (m)	طول جغرافیایی	عرض جغرافیایی
یزد	سینوپتیک	۱۹۶۱	یزد	۱۲۳۰	۵۴° ۱۷ ´	۳۱°۵۳´ ۵۶ ˝

جدول ۲- ماتریس همبستگی بین ایستگاه‌های مورد استفاده در این تحقیق جهت بازسازی نواقص آماری ایستگاه یزد

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
۱	۱	#٪۴۱	#٪۴۳	*٪۴۹	&٪۳۴	&٪۴۲	#٪۴۲	&٪۲۳	&٪۲۷	*٪۶۰	&٪۳۷	*٪۵۹	*٪۵۹	#٪۴۵
۲		۱	*٪۶۳	*٪۸۴	*٪۷۷	*٪۶۷	*٪۴۸	#٪۳۹	*٪۶۵	#٪۴۵	*٪۵۹	*٪۵۹	*٪۷۲	*٪۷۹
۳			۱	*٪۶۱	#٪۵۳	&٪۲۹	*٪۵۵	*٪۵۳	*٪۴۶	*٪۵۱	*٪۶۳	*٪۵۱	*٪۶۳	*٪۶۳
۴				۱	*٪۷۱	*٪۶۳	#٪۴۶	&٪۲۱	*٪۶۸	#٪۴۴	#٪۴۵	*٪۵۹	*٪۸۱	*٪۷۳
۵					۱	*٪۶۹	&٪۳۰	&٪۴۰	*٪۶۹	&٪۴۷	#٪۴۹	*٪۷۰	*٪۶۳	*٪۶۰
۶						۱	&٪۱۹	&٪۱۵	*٪۵۲	#٪۴۸	&٪۳۷	#٪۳۸	*٪۴۷	*٪۵۹
۷							۱	&٪۳۷	*٪۵۱	&٪۳۸	&٪۳۶	*٪۶۹	*٪۵۸	*٪۶۳
۸								۱	*٪۵۳	#٪۴۷	*٪۶۱	&٪۳۸	*٪۴۸	*٪۵۴
۹									۱	#٪۴۳	*٪۶۴	*٪۵۴	*٪۵۶	*٪۶۴
۱۰										۱	*٪۶۶	#٪۴۵	#٪۴۵	*٪۵۰
۱۱											۱	*٪۵۸	#٪۴۵	*٪۵۰
۱۲												۱	*٪۶۵	*٪۶۲
۱۳													۱	*٪۸۳
۱۴														۱

توضیح: * معنی‌دار در سطح ۱ درصد، # معنی‌دار در سطح اعتماد ۵ درصد، & در سطح اعتماد ۱ درصد و ۵ درصد معنی‌دار نیست

$$P_x = \frac{1}{n}\left[\left(\frac{\overline{P}_x}{\overline{P}a} * P_a\right) + \left(\frac{\overline{P}_x}{\overline{P}b} * P_b\right) + ...\right] \quad (۱)$$

P_x : بارندگی ایستگاه ناقص در ماه مورد نظر.

n : تعداد ایستگاه‌های شاهد.

$\overline{\mathbf{P}}_x$: بارندگی متوسط در ایستگاه ناقص برای آمارهای موجود.

$\overline{\mathrm{P}}b$ $\overline{\mathrm{P}}a$, : بارندگی متوسط در ایستگاه‌های شاهد و همزمان با آمار ایستگاه ناقص.

P_b و P_a : بارندگی در ایستگاه شاهد a و b در ماه مورد نظر برای تکمیل آمار ایستگاه ناقص.

۲-۲- معرفی روش مورد استفاده

درخت تصمیم یکی از روش های داده کاوی و از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد که برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد. یعنی درخت تصمیم پیش‌بینی خود را در قالب یکسری قوانین توضیح می‌دهد (شکل ۱). در حالی که در شبکه‌های عصبی تنها پیش‌بینی بیان می‌شود و چگونگی آن در خود شبکه پنهان باقی می‌ماند. علاوه بر آن در درخت تصمیم‌گیری بر خلاف شبکه‌های عصبی، می‌توان از داده‌های غیرعددی نیز استفاده نمود.

در این تحقیق از الگوریتم CART (Classification and regression tree) به عنوان یکی از انواع درختان تصمیم رگرسیونی جهت «پیش‌بینی بارش۱۲ ماه قبل از وقوع» استفاده شد. ساخت این درختان بر سه اصل استوار است:

۱- مجموعه‌ای از سؤالات به شکل ?x≤d که در آن x یک متغیر مستقل و d یک مقدار ثابت است و جواب هر سؤال بله/خیر است.

۱- بهترین معیار شاخه زدن جهت انتخاب بهترین متغیرمستقل برای ایجاد شاخه.

۲- ایجاد آمار خلاصه برای گره انتهایی (Breiman et al., 1984).

معیارهای مختلفی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، ولی از آنجا که تحقیق حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل که انحراف حداقل مربعات (LSD)[۳] نام دارد تشریح می‌گردد. این معیار به صورت زیر تعریف می‌شود:

$$SS(t) = \sum_{i=1}^{N_t} \left(y_i(t) - \bar{y}(t)\right)^2 \quad (۲)$$

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

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

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

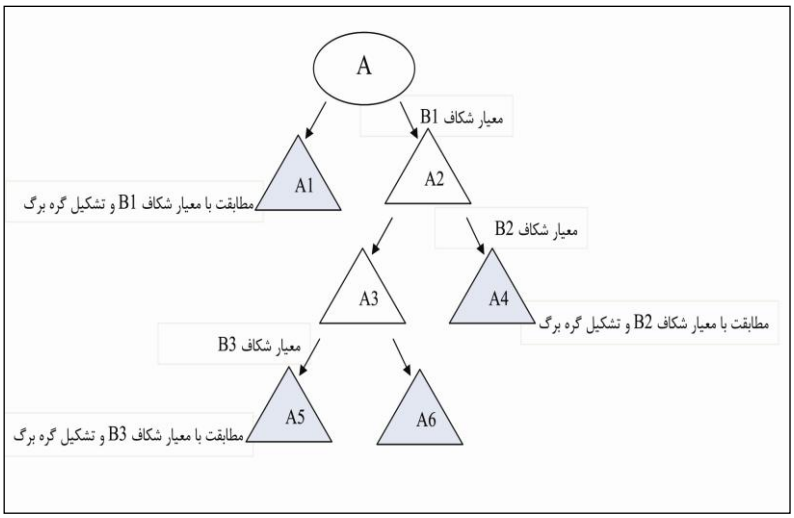
حال متغیر ورودی S زمانی بهترین متغیر برای ایجاد شاخه در گره t می‌باشد که مقدار Q(s,t) را بیشینه نماید.

$$Q(s,t) = SS(t) - SS(t_R) - SS(t_L) \quad (۳)$$

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.

۲-۳- تقسیم‌بندی داده‌ها برای آموزش مدل در ایستگاه یزد

جهت اجرای مدل درختان تصمیم از آمار و اطلاعات مربوط به برخی



شکل۱- نمونه‌ای از یک درخت تصمیم ساده

پارامترهای اقلیمی که در مقیاس ماهیانه در طول سال‌های ۱۳۳۲ لغایت ۱۳۸۴ در ایستگاه سینوپتیک یزد ثبت شده، استفاده گردید.

در این فرایند داده‌ها به دو دسته تقسیم شدند: داده‌های آموزشی[۴] و داده‌های آزمایشی[۵]. در این تحقیق ۷۰٪ از کل داده‌ها به آموزش مدل تعلق گرفت و ۳۰ درصد باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردید. انتخاب داده‌های آموزش و آزمون به صورت سیستماتیک و توسط کاربر انجام شد و بدیهی است که داده‌های آزمون در مرحله آموزش مورد استفاده قرار نگرفته اند.

۲-۴- پارامترهای مورد استفاده در شبیه‌سازی

در شبیه‌سازی با استفاده از مدل درخت تصمیم از پارامترهای مختلفی استفاده شد. این پارامترها به عنوان متغیر مستقل به مدل معرفی و شبیه‌سازی‌ها برای پیش‌بینی متغیر هدف صورت گرفت. متغیرهای مستقل مورد استفاده در این تحقیق و علائم اختصاری آن در جدول ۳ نمایش داده شده است.

۲-۵- حالات مختلف اجرای مدل

الف) استفاده از اصل داده‌ها بدون نرمال‌سازی: در این حالت بر روی داده‌ها هیچ گونه عملیات تغییرمقیاس 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 ۳ به عنوان ورودی در اختیار مدل قرار گرفت و بارش ماهیانه یک سال بعد پیش‌بینی گردید.

ب) استفاده از میانگین متحرک سه ساله داده‌های خام: استان یزد یک استان خشک می‌باشد که از خصوصیات بارز آن نوسانات شدید بارش از یک سال به سال دیگر می‌باشد به گونه‌ای که داده‌های بارش از هیچ گونه روند منظمی برخوردار نمی‌باشند. از این رو بعد از انجام چندین شبیه‌سازی جهت بالا بردن ضریب کارایی در پیش‌بینی بارش، به جای داده‌های ماهیانه از میانگین متحرک سه ساله استفاده گردید. به این ترتیب میانگین متحرک سه ساله برای کل پارامترهای ورودی مدل محاسبه شد و با در نظر گرفتن این پارامترها به عنوان متغیر مستقل، میانگین متحرک سه ساله بارش برای یک سال بعد پیش‌بینی گردید.

ج) استفاده از میانگین متحرک سه ساله نرمال شده: در این حالت میانگین متحرک سه ساله داده‌ها، با استفاده از رابطه ۴ نرمال شد و سپس درخت تصمیم برای انجام پیش‌بینی ایجاد و مورد ارزیابی قرار گرفت.

$$X_n = \frac{X_o - X_{\min}}{X_{\max} - X_{\min}} \qquad (۴)$$

X_n و X_O : به ترتیب داده‌های نرمال شده و داده‌های اولیه را نشان می‌دهد.

X_{min} , X_{max} : به ترتیب مینیمم و ماکزیمم داده‌های اولیه را نشان می‌دهد.

در این حالت میانگین متحرک سه‌ساله نرمال شده داده‌ها، به عنوان ورودی به مدل معرفی شده و میانگین متحرک سه ساله نرمال شده بارش برای یک سال بعد پیش‌بینی گردید.

د) استفاده از میانگین متحرک پنج ساله 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 ساله بارش برای یک سال بعد پیش‌بینی گردید.

۲-۶- ارزیابی کارایی مدل

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

$$RMSE = \left(\frac{\sum_{i=1}^{n} (p_i - o_i)^2}{n} \right)^{1/2} \qquad (۵)$$

$$MAE = \frac{\sum |O_i - P_i|}{N} \qquad (۶)$$

$$Bias = \overline{P} - \overline{O} \qquad (۷)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} (Oi - Pi)}{\sum_{i=1}^{n} (Oi - \overline{O})}} \qquad (۸)$$

در این روابط O_i و P_i به ترتیب مقادیر مشاهده شده و شبیه‌سازی شده، $\overline{O}$ و $\overline{P}$ میانگین مقادیر مشاهده شده و شبیه‌سازی شده و N تعداد نمونه‌ها می‌باشد.

جدول ۳- معرفی پارامترهای ورودی مدل (دوره آماری ۱۳۳۲ تا ۱۳۸۴)

علامت اختصاری	پارامتر ورودی	حداکثر	حداقل	میانگین
P	بارش (میلیمتر)	۷۰	۰	۴/۴۷
T_{mean}	دمای متوسط (درجه سانتیگراد)	۳۵/۵	-۰/۸	۲۰/۱۸
T_{max}	دمای ماکزیمم (درجه سانتیگراد)	۴۲/۶	۵/۸	۲۷/۲۵
H	رطوبت نسبی (درصد)	۷۴	۸	۲۲/۸۴
W_s	سرعت شدید ترین باد(متر بر ثانیه)	۱۱/۱	۱/۶	۵/۷
W_d	جهت باد غالب (درجه)	۳۶۰	۱۰	۲۵۶/۶

جدول۴- کیفیت نتایج در پیش بینی بارش با استفاده از داده‌های خام (اصل داده‌ها)

ردیف	متغیرهای مستقل	متغیر هدف	R	RMSE (mm)	MAE (mm)	Bias
حالت اول	$P\text{-}T_{mean}\text{-}T_{max}\text{-}W_d$	p_n	۰/۴۲	۸/۲	۴/۶۱	۰/۳۷

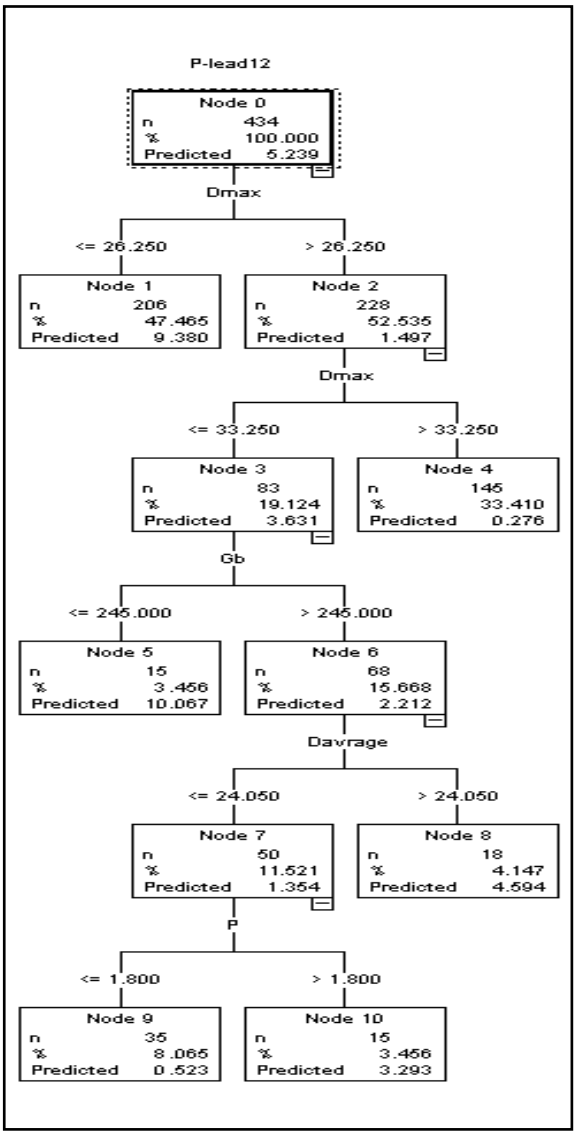
۳- نتایج و تحلیل نتایج

۳-۱- نتایج حاصل از اجرای مدل در شرایط استفاده از اصل داده‌ها

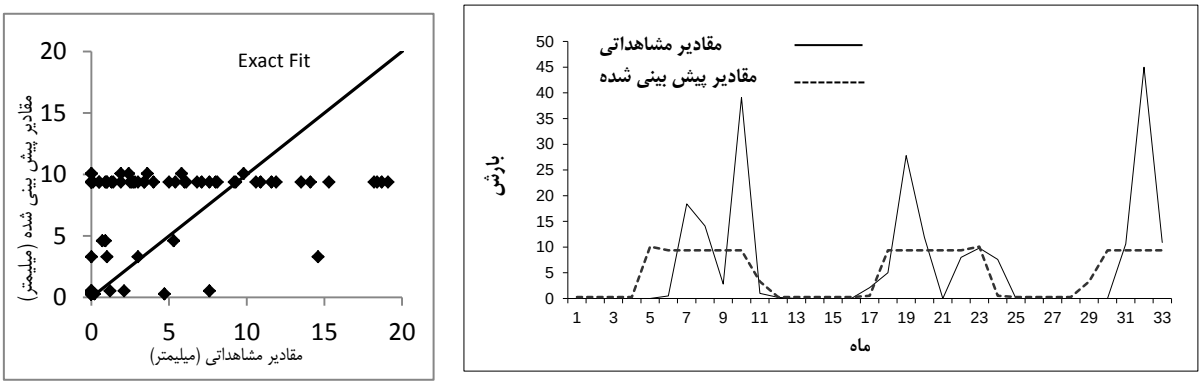
در این حالت مدل با ۶ متغیر مستقل شامل بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی، سرعت باد و جهت باد و ۶۲۰ سطر داده اجرا شد. لازم به ذکر است که از این تعداد ۴۳۴ سطر در مرحله آموزش و ۱۸۶ سطر در مرحله آزمون استفاده گردید. ساختار درخت ایجاد شده در این حالت در شکل ۲ نمایش داده شده است.

شکل ۲ نشان می‌دهد که مدل در حالت اول از بین ۶ پارامتری که در اختیار داشته، بارش، دمای متوسط، دمای ماکزیمم و جهت باد را انتخاب و درخت بر اساس این ۴ پارامتر ساخته شده است. در ادامه، کارایی درخت ایجاد شده با استفاده از چهار پارامتر آماری (R , RMSE MAE , Bias) بررسی شد که مقادیر این پارامترها در جدول ۴ درج شده است. نمودار نقطه‌ای و نمودار خطی مربوط به نتایج مدل درخت تصمیم و مقادیر واقعی مربوط به حالت استفاده از اصل داده‌ها نیز در شکل ۳ نمایش داده شده است.

مدل ایجاد شده فوق، در پیش بینی مقادیر اوج از کارایی بسیار کمی برخوردار است. ضمن اینکه مدل مزبور در غالب بازه‌های زمانی ۱۲ ماهه، پیش‌بینی‌های خود را نزدیک به دو مقدار عددی حداقل و حداکثر انجام داده است. به عبارت بهتر درخت موصوف روند پیش‌بینی نسبتاً ثابتی داشته است.



شکل ۲- ساختار درخت تصمیم ایجاد شده درحالت اول (استفاده از اصل داده‌ها)



شکل۳- نمودار تغییرات مقادیرپیش بینی شده در مقابل مقادیر واقعی در حالت استفاده از اصل داده‌ها

۳-۲- نتایج مدل با استفاده از میانگین متحرک سه ساله داده‌های اولیه:

در این حالت جهت اجرای مدل، ۶ متغیر مستقل شامل میانگین متحرک سه ساله داده های بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی، سرعت باد و جهت باد با ۵۹۶ سطر داده در اختیار سیستم قرار گرفت که از این تعداد۴۱۷ سطر داده در مرحله آموزش و ۱۷۹ سطر در مرحله آزمون استفاده گردید. ساختار درخت ایجاد شده در این حالت در شکل ۴ نمایش داده شده است.

پس از آن کارایی درخت ایجاد شده با استفاده از پارامترهای آماری مورد استفاده ارزیابی شد که مقادیر این پارامترها در جدول ۵ درج شده است. نمودار نقطه ای و نمودار خطی مربوط به نتایج مدل درختان تصمیم و مقادیر واقعی مربوط به حالت استفاده از میانگین متحرک سه ساله داده‌ها نیز در شکل ۵ نمایش داده شده است.

۳-۳- نتایج مدل با استفاده از میانگین متحرک سه‌ساله نرمال

در این بخش میانگین متحرک سه ساله داده‌ها نرمال شد و سپس ۵۹۶ سطر داده به نسبت ۷۰٪ و ۳۰٪ (۴۱۷ و ۱۷۹سطر داده) به ترتیب جهت آموزش و آزمون به مدل معرفی گردید و میانگین متحرک سه ساله نرمال شده بارش، ۱۲ ماه قبل از وقوع پیش‌بینی گردید. درخت تصمیم ایجاد شده در این حالت در شکل ۶ نمایش داده شده است. جهت بررسی کارایی درخت ایجاد شده در شکل ۶ نیز همانند حالت قبلی از پارامترهای آماری استفاده گردید. که مقادیر این پارامترها برای این حالت در جدول ۶ آمده است.

به منظور امکان مقایسه دقیق نتایج به دست آمده در حالات مختلف، مقادیر پیش بینی شده و واقعی مربوط به این مرحله مجدداً آنرمال شدند (به حالت غیر نرمال برگردانده شدند) و پارامترهای خطا برای آن ها محاسبه گردید. جدول ۷ مقادیر خطای پیش بینی داده‌های آنرمال را در این مرحله نمایش می‌دهد.

جدول ۵- کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله داده‌ها

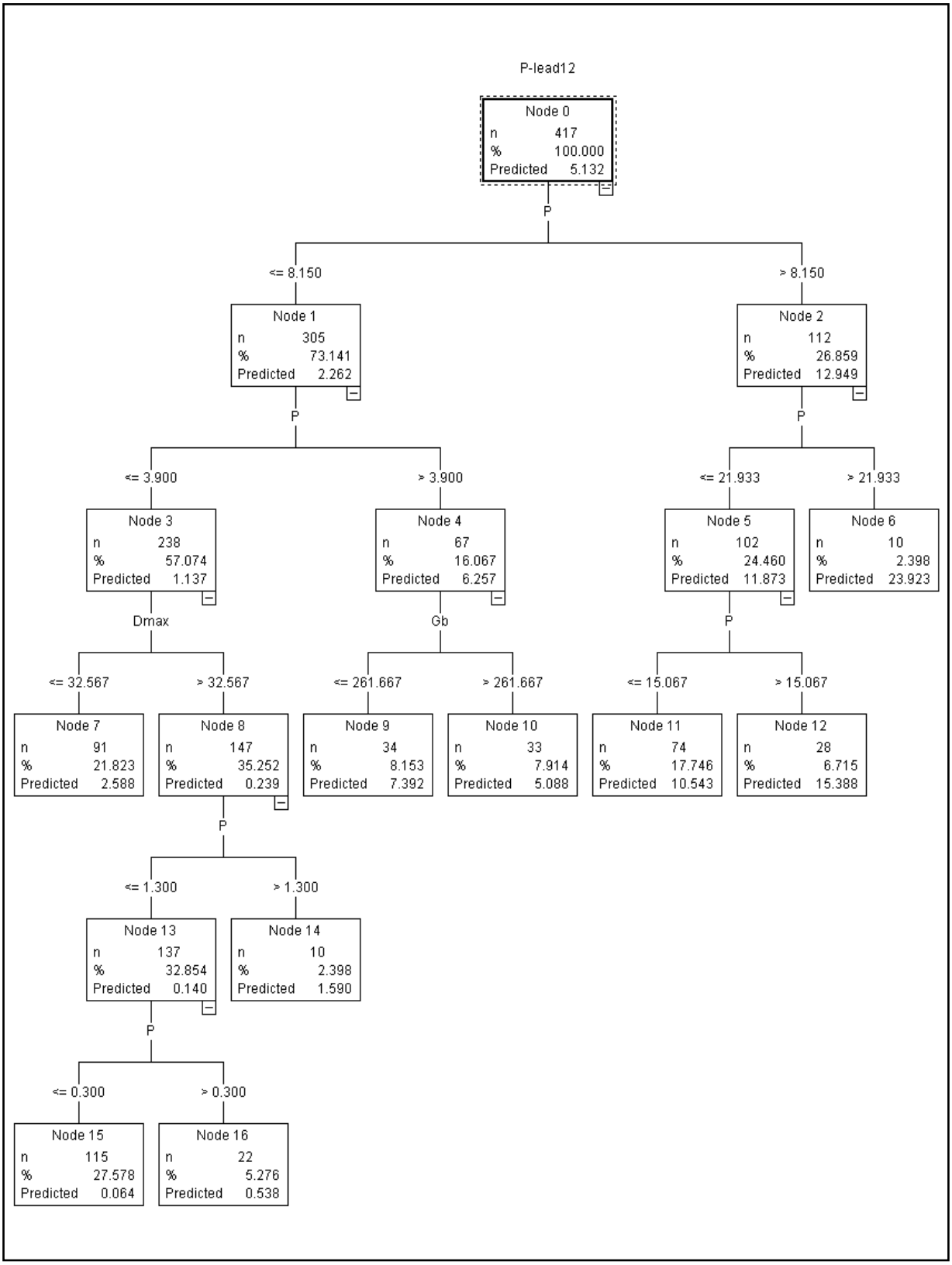
ردیف	متغیرهای مستقل (میانگین متحرک سه ساله)	متغیر هدف (میانگین متحرک سه ساله)	R	RMSE (mm)	MAE (mm)	Bias
حالت اول	$P\text{-}T_{max}\text{-}W_d$	P_{nm3}	۰/۸۳	۳/۴۴	۱/۹۹	۰/۱۷

جدول۶-کیفیت نتایج مدل با استفاده از میانگین متحرک سه ساله نرمال شده

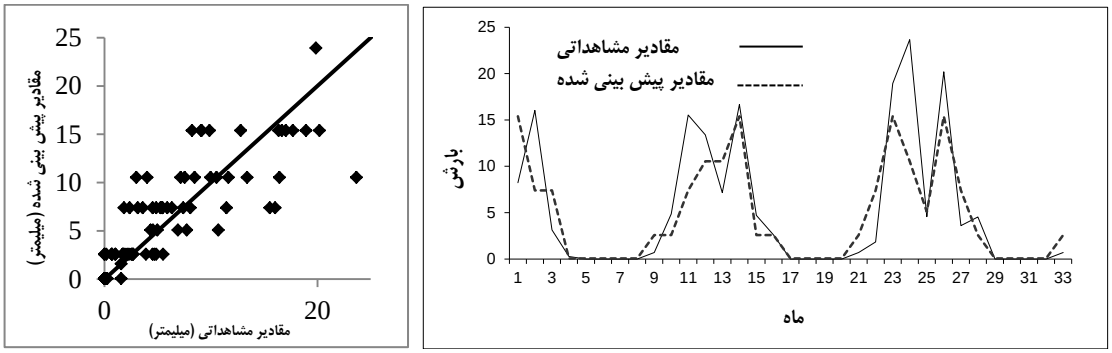
ردیف	متغیر مستقل (میانگین متحرک سه ساله نرمال شده)	متغیر هدف (میانگین متحرک سه ساله نرمال شده)	R	RMSE (mm)	MAE (mm)	Bias
۱	$P\text{-}T_{max}\text{-}W_d$	P_{nm3n}	۰/۸۳	۰/۰۹	۰/۰۶	۰/۰۰۶

جدول۷- خطای مقادیر آنرمال پیش بینی در شرایط استفاده از میانگین متحرک سه ساله نرمال شده

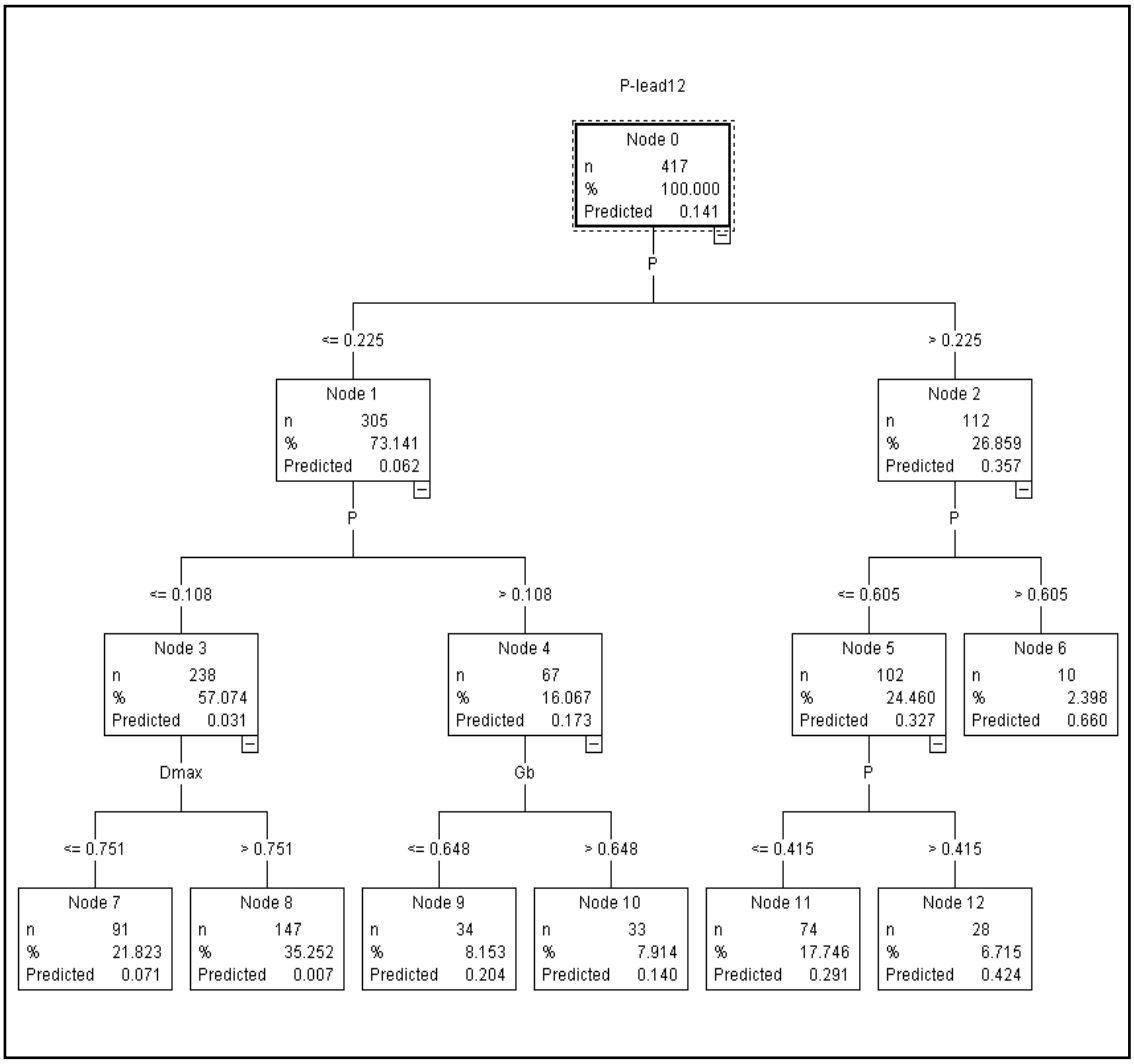
ردیف	متغیر مستقل (میانگین متحرک سه ساله نرمال شده)	متغیر هدف (آنرمال)	R	RMSE (mm)	MAE (mm)	Bias
حالت اول	$P\text{-}T_{max}\text{-}W_d$	P_{nm3a}	۰/۸۳	۳/۵	۲/۱۱	۰/۲۳



شکل ۴- ساختار درخت تصمیم در حالت دوم (استفاده از میانگین متحرک سه ساله)



شکل ۵- نمودار تغییرات مقادیرپیش بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه ساله داده‌ها



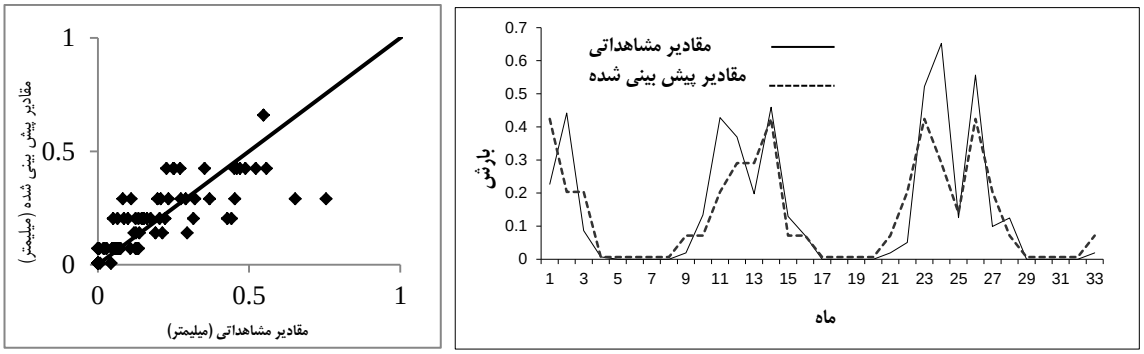
شکل۶ - ساختار درخت تصمیم در حالت سوم (استفاده از میانگین متحرک سه ساله نرمال شده)

نمودار نقطه‌ای و نمودار خطی مربوط به نتایج مدل درخت تصمیم و مقادیر واقعی مربوط به حالت استفاده از میانگین متحرک سه ساله نرمال شده در شکل ۷ نمایش داده شده است.

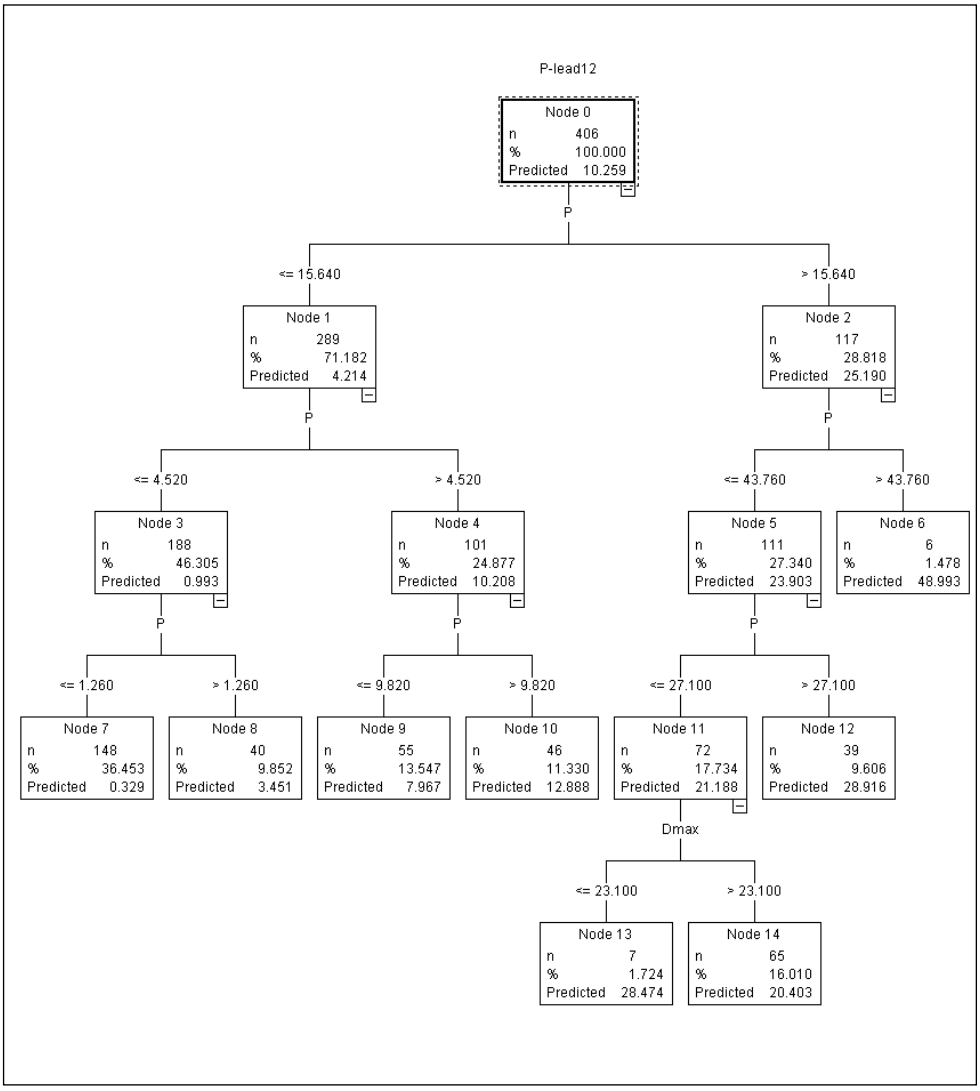
۳-۴- نتایج مدل در شرایط استفاده از میانگین متحرک پنج ساله داده‌های اصلی

در ادامه جهت بررسی و بالا بردن کارایی مدل در پیش بینی بارش، از میانگین متحرک پنج ساله داده ها نیز استفاده گردید. ۶ متغیر

مستقل شامل میانگین متحرک پنج ساله داده‌های؛ بارش قبلی، دمای متوسط، دمای ماکزیمم، رطوبت نسبی، سرعت باد و جهت باد به مدل معرفی و میانگین متحرک پنج ساله بارش برای سال بعد پیش‌بینی گردید. در این حالت ۵۸۲ سطر داده (با توجه به تشکیل میانگین متحرک ۵ ساله تعداد داده‌ها اندکی کمتر شد) جهت اجرای مدل در اختیار سیستم قرار گرفت که از این تعداد ۴۰۸ سطر داده در مرحله آموزش و ۱۷۴ سطر داده در مرحله آزمون استفاده گردید. درخت ایجاد شده در این حالت در شکل ۸ نمایش داده شده است.



شکل ۷- نمودار تغییرات مقادیر پیش بینی شده در مقابل مقادیر واقعی با استفاده از میانگین متحرک سه ساله نرمال شده داده‌ها



شکل ۸- درخت تصمیم در حالت چهارم (استفاده از میانگین متحرک ۵ ساله داده‌ها)

جدول ۸- کیفیت نتایج مدل با استفاده از میانگین متحرک پنج ساله داده‌ها

ردیف	متغیر مستقل (میانگین متحرک پنج ساله)	متغیر هدف (میانگین متحرک پنج ساله)	R	RMSE (mm)	MAE (mm)	Bias
۱	$P\text{-}T_{max}$	P_{nm5}	۰/۹۱	۴/۵۹	۲/۶۷	۰/۰۷

مقادیر پارامترهای آماری مربوط به ارزیابی نتایج درخت ایجاد شده در این حالت و نیز متغیرهای مستقل استفاده شده در جدول ۸ آمده است.

نمودارهای نقطه‌ای و خطی مربوط به نتایج مدل درختان تصمیم و مقادیر واقعی مربوط به میانگین متحرک پنج ساله داده‌ها در شکل ۹ نمایش داده شده است.

نتایج نهایی حاصل از این تحقیق را به صورت خلاصه بشرح زیر بیان نمود:

– در تمام شبیه سازی‌های صورت گرفته، پارامتر بارش قبلی و دمای ماکزیمم دخیل بوده است. این موضوع نشان دهنده اهمیت این پارامترها در پیش‌بینی فرایند بارش در آینده می‌باشد.

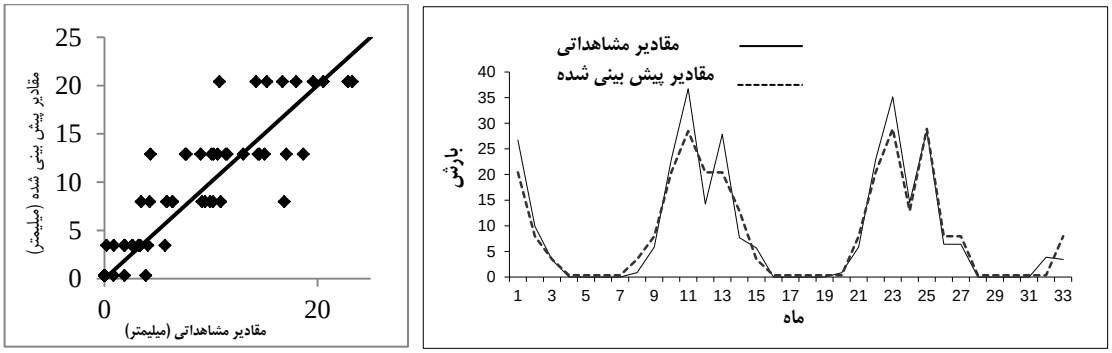
– بر اساس نتایج به دست آمده از این تحقیق، می‌توان گفت درخت تصمیم رگرسیونی در ایستگاه یزد، مدلی نسبتاً کارا در پیش‌بینی بارش با استفاده از داده‌های پردازش شده می‌باشد، که در صورت تعدیل دامنه تغییرات داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان بارش را ۱۲ ماه قبل از وقوع برآورد نماید. به طور مشخص استفاده از میانگین متحرّک نسبت به سایر حالات منجر به افزایش چشمگیر کارایی مدل درخت تصمیم می‌شود.

– در شبیه سازی های صورت گرفته در این تحقیق، زمانی که از میانگین متحرّک پنج ساله داده‌ها برای اجرای مدل استفاده گردید، ترکیب "بارش قبلی، دمای ماکزیمم" به عنوان مناسب ترین حالت شناسایی شد. طبق نتایج این تحقیق، درخت ایجاد شده در حالت استفاده از میانگین متحرک پنج ساله داده‌ها قادر است با ۸ قانون، میانگین متحرک پنج ساله بارش را در ایستگاه یزد، ۱۲ ماه قبل از وقوع پیش بینی نماید. بدیهی است با کسر بارش ۴ سال آخر از عدد مزبور می‌توان به میزان بارش سال بعد، ۱۲ ماه قبل از وقوع دست یافت. بر این اساس دستیابی به این پارامتر (میزان بارش ۱۲ ماه قبل از وقوع) با محاسبات ساده ریاضی امکان پذیر است.

– در ایستگاه یزد، مدل درخت تصمیم در پیش‌بینی بارش در شرایط استفاده از اصل داده‌ها ناموفق بود. به نظر می‌رسد دامنه تغییرات این داده‌ها برای استفاده در مدل درخت تصمیم زیاد می‌باشد و همین موضوع منجر به کاهش کارایی مدل در پیش‌بینی می‌گردد. بهتر شدن کارایی مدل جهت پیش بینی در حالت استفاده از داده های میانگین متحرک مؤید این موضوع می‌باشد.

– با توجه به نمودارهای ترسیم شده مبتنی بر نتایج این تحقیق، در تمام شبیه سازی‌های صورت گرفته، مدل مورد استفاده در پیش‌بینی مقادیر اوج با چالش روبروست.

– شکل ۳ نشان می‌دهد که مدل در حالت استفاده از اصل داده‌ها در غالب دوره‌های زمانی ۱۲ ماهه پیش بینی‌های خود را در قالب دو مقدار عددی حداقل و حداکثر انجام داده است. به عبارتی درخت ایجاد شده با اصل داده‌ها روند پیش‌بینی نسبتاً ثابتی داشته و نتوانسته است روند مناسبی را دنبال کند.



شکل ۹- نمودار تغییرات مقادیرپیش بینی شده در مقابل مقادیر واقعی با استفاده از داده‌های میانگین متحرک پنج ساله داده‌ها

– استفاده از میانگین متحرّک سه ساله نرمال شده نسبت به میانگین متحرّک سه ساله و مقایسه نتایج نهایی مدل (داده‌های آنرمال) نشان می‌دهد که عملیات نرمال‌سازی کارکرد مدل را تحت‌الشعاع قرار داده و کارایی مدل درخت تصمیم را اندکی کاهش می‌دهد. به نظر می‌رسد این موضوع ناشی از نحوه کارکرد مدل بوده 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 شکاف (شاخه زنی) و تفکیک داده‌ها کاهش می‌دهد.

– هرس، یکی از مراحل اصلی اجرای مدل درخت تصمیم است که در غالب موارد منجر به افزایش کارایی درخت در پیش‌بینی مقادیر بارش ۱۲ ماه قبل از وقوع گردید.

– به طور کلی، مدل درخت تصمیم به لحاظ سادگی و ایجاد قوانین پیش‌بینی بر مدل های جعبه سیاه ارجحیت دارد. از طرفی، مدل درخت تصمیم قادر است بدون دخالت کاربر، ورودی‌های مهم تر را برای ایجاد قوانین پیش‌بینی استفاده و ورودی‌های ضعیف‌تر را حذف نماید.

۴ – نتیجه گیری

در مجموع وقوع خشکسالی ها خصوصاً در دهه های اخیر خسارات جبران‌ناپذیری را به منابع طبیعی و انسانی بخش‌های مختلف کشور وارد ساخته است. یکی از مؤلفه های مؤثر در مدیریت خشکسالی‌ها و سازماندهی امکانات و نیز تصمیم‌گیری در خصوص نحوه استفاده از منابع آبی در این مواقع اطلاع نسبی از وضعیت خشکسالی قبل از وقوع آن است. میزان بارندگی اصلی‌ترین شاخص در ارزیابی شدت و مدت خشکسالی است و لذا لازمه ارزیابی خشکسالی قبل از وقوع آن پیش بینی مقدار بارندگی با دقت و اطمینان کافی است. از این رو هر مدلی که بتواند مقدار بارش را خصوصاً در مناطق خشک که تغییرات داده های اقلیمی قابل توجه است، با دقت کافی پیش بینی نماید اهمیت فوق العاده ای در مدیریت خشکسالی ها و کاهش خسارات ناشی از آنها دارد. علاوه بر آن پارامتر های هواشناسی و به ویژه پارامتر بارش از پایگاه داده مناسبی در ارگان‌های زیربط برخوردار می‌باشد، لذا می توان از طریق داده کاوی به پردازش، شبیه سازی و تجزیه و تحلیل داده‌های مذکور و کشف روابط حاکم بر داده‌های اقلیمی موثر بر رخداد خشکسالی پرداخت. این در حالی است که نوسانات شدید بارش به خصوص در مناطق خشک، توانایی مدل‌های داده کاوی را در یادگیری فرایند حاکم بر وقوع خشکسالی تحت الشعاع قرار می دهد. از این رو پردازش داده ها به منظور افزایش کارایی این مدل ها می‌تواند در تحقق اهداف تحقیق مثمر ثمر قرار گیرد. نتایج این تحقیق نشان داد که اعمال پروسه میانگین متحرک به عنوان یکی از عملیات موثر در تعدیل دامنه تغییرات داده ها، به نحو چشمگیری بر افزایش کارایی مدل درخت تصمیم رگرسیونی تاثیرگذار بوده است و مدل درخت تصمیم رگرسیونی مدلی کارا در پیش‌بینی خشکسالی با استفاده از داده‌های پردازش شده می‌باشد، که در صورت تعدیل دامنه تغییرات داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان بارش را ۱۲ ماه قبل از وقوع برآورد نمایند.

پی نوشت‌ها

1- Decision trees
2- CART
3- Least Square Deviation
4- Training data
5- Testing data

۵- مراجع

اکبری، ز.(۱۳۸۹). بررسی کارایی مدل درخت تصمیم‌گیری رگرسیونی در برآورد میزان رسوب حوزه سد ایلام، پایان نامه کارشناسی ارشد، دانشکده منابع طبیعی و کویر شناسی، دانشگاه یزد.

شایق، م. ع. (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.

کانتاردزیک، م. (۱۳۸۵).داده کاوی، مترجم: امیر علیخانزاده؛ بابل، نشر علوم رایانه.

مشکانی، ع. و ناظمی، ع. (۱۳۸۸). مقدمه ای بر داده کاوی، مشهد، موسسه چاپ و انتشارات دانشگاه فردوسی.

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, and Solomatine D.P., (2007). Machine learning approach to modeling sediment transport. *Journal of Hydraulic Engineering*. 133(4), pp. 440-450.

Breiman, L., Friedman J., Olshen R., and Stone C.,1984, *Classification and Regression Trees*, Chapman & Hall/CRC Press, Boca Raton, FL.

Obasi,G.O.P., (1994), WMO,s role in the international decade for natural disaster reduction, Bull,*Am, Meteorole, Soc*.75(1), 665-16661.

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

Taghi Sattari, M., Anli, A.S., Apaydin, H. and Kodal, S., (2012), Decision tree to determine the possible drought periods in Ankara, *Atmosfera* 25 (1), pp. 65-83.

USDA., (1994), Major world crop areas and climatic profile, World Agricultural out look Board, US Department of Agriculture, Agricultural Hand book No, 664, pp. 157-170.

Vega, F.A., Matías J. M., Andrade M. L., Reigosa M. J. and Covelo E.F. (2009). Classification and regression trees (CARTs)for modeling the sorption and retention of heavy metals by soil. *Journal of Hazardous Materials*, 167, pp. 615 –624.

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.

Chen, J.C., Ning, S.K., Chen, H.W. and Shu, C.S. (2008), Flooding probability of urban area estimated by decision tree and artificial neural networks, *Journal of Hydroinformatics*, Vol.10, No.1, pp. 57-67.

Cheng, C.C., Hsu, N.S. and Wei C.C. (2008), Decision-tree analysis on optimal release of reservoir storage under typhoon warnings, *Natural Hazards*, 44, pp. 65-84.

Geissen, V., Kampichler C., Lopez-de Llergo J.J. and Galindo-Acantara A. (2007).superficial and subterranean soil erosion in Tabasco, tropical Mexico: development of a decision tree modeling approach .*Geoderma* 139, pp. 277-287.

Kheir,B.F., Chorowicz J., Chadi A. and Damien D., (2008). Soil and bedrock distribution estimated from gully form and frequency: A GIS-based decision-tree model for Lebanon. *Geomorphology* 93, pp. 482-.492.

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.

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.