فصلنامه علوم محیطی، دوره سیزدهم، شماره ۴، زمستان ۱۳۹۴

۵۳-۵۸

ارزیابی دو الگوریتم داده‌کاوی CART و CHAID در پیش‌بینی دمای هوای ایستگاه سینوپتیک اراک

مهدی پناهی[۱*] و سید حسن میرهاشمی[۲]

[۱] استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه زنجان
[۲] دانشجوی دکتری آبیاری و زهکشی، دانشکده آب و خاک، دانشگاه زابل

تاریخ دریافت: ۹۴/۴/۳۰ تاریخ پذیرش: ۹۴/۹/۵

چکیده

امروزه با توجه به خشک‌سالی‌های اخیر و گرمایش زمین مدل‌سازی و پیش‌بینی پارامترهای اقلیمی امری اجتناب‌ناپذیر به نظر می‌رسد که در میان روش‌های موجود استفاده از الگوریتم‌های داده‌کاوی در پیش‌بینی عناصر اقلیمی کاربرد وسیعی پیدا کرده است، لذا در این پژوهش دو الگوریتم داده‌کاویCART و CHAID درپیش‌بینی دمای هوای ایستگاه سینوپتیک اراک مورد ارزیابی قرار گرفتند. داده‌های هواشناسی شامل متوسط ساعات آفتابی، دمای نقطه شبنم، درصد رطوبت نسبی، سرعت باد، کمبود فشار بخار اشباع طی دوره چهل‌وشش ساله (۲۰۰۵-۱۹۶۰) مورد استفاده قرار گرفته شد. پس از معرفی داده‌های هواشناسی به صورت متوسط ماهانه به عنوان متغییرهای ورودی و متوسط دمای ماهانه ماه بعد به عنوان متغیرهای خروجی به دو الگوریتم مذکور، الگوریتم‌ها توسط دو شاخص آماری ضریب همبستگی(R^2) و متوسط مطلق خطا (MAE) مورد ارزیابی قرار گرفتند. بر اساس دو شاخص آماری مدل درختی CHAID با مقدار R = 0.915 و MAE = 2.77 عملکرد بهتری در پیش‌بینی دمای متوسط ماه بعد نسبت به الگوریتم CART داشته است.

کلمات کلیدی: پیش‌بینی، مدل درختی،CART،CHAID، دمای متوسط ماهانه، ایستگاه سینوپتیک اراک.

Assessment Among Two Data Mining Algorithms CART and CHAID in Forecast Air Temperature of the Synoptic Station of Arak

Mehdi Panahi[1*] & Seyyed Hassan Mirhashemi[2]

[1] Assistant Professor, Department of Water Engineering, Faculty of Agriculture, University of Zanjan, Zanjan
[2] PhD. Student of Irrigation and Drainage, Faculty of Water and Soil, University of Zabol, Zabol

Abstract

Due to recent droughts and global warming, in the modeling and prediction of climatic parameters it seems inevitable that the approach of using data mining algorithms to predict climatic elements has been widely used .Therefore, in this study the use of the CARD and CHAID data mining algorithms to predict the air temperature at Arak synoptic stations was evaluated. The data provided are the average monthly data from the Arak weather station, including: "sunshine hours", "dew point", "relative humidity", "average wind speed" and "saturation vapor pressure deficit" over a forty-six-year period from 1960 to 2005. The output variable used was "average temperature months later" on a monthly basis. After introducing the weather data as a monthly mean to the algorithm as an input variable of air temperature months later. Then to the CART and CHAID algorithms were evaluated using correlation coefficient (R2) and mean absolute error (MAE). According to the two statistical indices, the CHIAD Tree model has a better function by R2 =0.915 and MAE= 2.77 in forecasting the monthly average temperature for the months later.

Keywords: Assessment, Tree Model, CART, CHAID, Average Temperature Months, Synoptic Station of Arak.

* Corresponding Author. *E-mail Address*: panahi40@yahoo.com

۱- مقدمه

ایران به‌دلیل قرارگیری در کمربند خشک‌سالی و مجاورت با میدان‌های پرفشار جنب حاره‌ای دارای اقلیمی خشک و نیمه خشک است و همین امر سبب شده تا در بیشتر سال‌ها دچار خشک‌سالی باشد. دو پارامتر اصلی و تأثیرگذار اقلیمی در ایجاد و تداوم خشک‌سالی یک منطقه، بارش و متوسط درجه حرارت هستند[۱]. دما در کنار بارش از مهم‌ترین عناصر اقلیمی محسوب می‌شود که در تعیین نقش و پراکندگی دیگر عناصر اقلیمی نیز مؤثر است. همچنین دما از عوامل اصلی و اساسی در پهنه‌بندی و طبقه-بندی اقلیمی محسوب می‌شود، بر این اساس نوسانات و تغییرپذیری دما دارای اهمیت زیادی است[۲]. لذا مطالعه تغییرات دما حایز اهمیت است. پیرامون تغییرات اقلیم و بررسی عناصر هواشناسی مطالعات متعددی در سطح جهان و ایران صورت گرفته است. طی تحقیقاتی افزایش دما در کانادا[۳]، استرالیا و نیوزلند[۴] گزارش شده است. همچنین نتایج بررسی ۱۹ ایستگاه هواشناسی در اروپای شمالی و مرکزی افزایش دمای متوسط و حداکثر در شمال و مرکز اروپا را نشان داد [۵]. روند زمانی دمای متوسط سالانه در ایران با استفاده از داده‌های ۳۴ ایستگاه سینوپتیک بررسی شد. نتایج بررسی روشن ساخت که برای کل دوره آماری ۱۹۹۸-۱۹۶۸ که در ۵۹ درصد ایستگاه‌ها روند مثبت و در ۴۹ درصد ایستگاه‌ها روند منفی وجود دارد [۶]. شهاب‌فر و همکاران (۲۰۰۴) تغییرات زمانی روزهای یخبندان را در مشهد مورد مطالعه قرار دادند. با استفاده از داده‌های ماهانه روند میانگین دما، حداقل مطلق و تعداد روزهای یخبندان بررسی شد روند افزایشی کلی دما و کاهش در تعداد روزها مشاهده شد[۷]. وانگ و همکاران (۲۰۱۳) روند حداکثر دمای چین را طی سال‌های ۲۰۰۹- ۱۹۶۰ مورد بررسی قرار دادند. در این تحقیق تغییرات قابل توجهی در روند دما از اوایل سال ۱۹۹۰ مشاهده شد[۸].تحقیقات صورت گرفته در زمینه‌ی تغییر اقلیم در جهان، حاکی از این واقعیت است که تغییر هر چند کم در دما موجب تغییر در وقوع پدیده‌های نظیر خشک‌سالی، بارش‌های سنگین و طوفان می‌شود [۹].

در زمینه بررسی رفتار داده‌های آماری انتخاب، شناسایی و مدل‌سازی از مقادیر زاید داده‌ها برای کشف نظم و یا رابطه‌ای که در ابتدا ناشناخته است می‌توان از الگوریتم-های داده‌کاوی استفاده نمود. داده‌کاوی به صورت تشخیص ساختاری جالب در داده تعریف می‌شود که این ساختار، الگوها و مدل‌های پیش‌بینی و آماری از داده‌ها و روابط بین بخش‌های آن را طراحی می‌نماید [۱۰]. این تعاریف به مقدار زیادی به پیش‌زمینه‌ها و نقطه‌نظرهای افراد بستگی دارد. پس می‌توان گفت داده‌کاوی مجموعه‌ای از روش‌ها در فرآیند کشف دانش می‌باشدکه برای تشخیص الگوها و روابط نامعلوم در داده‌ها مورد استفاده قرار می‌گیرد و یا به عبارت دیگر داده‌کاوی به فرآیند استخراج اطلاعات نهفته، قابل فهم، قابل پیگیری ازپایگاه داده‌های بزرگ و استفاده از آن در تصمیم‌گیری‌های تجاری مهم، اطلاق می‌شود. داده‌کاوی تکنیکی است که آزمون‌های فرض را ترکیب می‌کند و داده‌های مشتقی را کشف می‌کند. در آزمون‌های فرض، محقق ایده را در برابر داده آزمون می‌کند تا اعتبار آن را تأیید یا رد کند. شایان ذکر است که داده‌کاوی از ترکیب چندین رشته نشئت می‌گیرد. آمار، یادگیری ماشین، روش‌های بهینه‌سازی، روش‌های تشخیص و شناخت الگو، بانک اطلاعاتی، تجسم‌سازی، شبکه‌های عصبی، که در کشف مدل‌های ریاضی، بازیابی اطلاعات، الگوریتم ژنتیک و هوش مصنوعی فنونی هستند که داده‌کاوی از آن‌ها بهره می‌برد [۱۱]. یک محقق نتیجه را از داده‌ها ترسیم می‌کند و به داده‌ها اجازه می‌دهد تا نتیجه را بپذیرد. اغلب مسائل داده‌کاوی با استفاده از ترکیب هر دو روش حل می‌شود. به‌طور مثال از نتیجه ممکن است فرضیه جدیدی شود که می‌تواند این آزمون، تایید یا رد شود [۱۲]. در تعریف دیگری، فرآیند انتخاب، کاوش و مدل‌بندی داده‌های حجیم جهت کشف روابط نهفته با هدف به دست آوردن نتایج واضح و مفید برای مالک پایگاه داده‌ها را، داده‌کاوی می‌گویند[۱۳]. تفاوت اصولی که بین داده‌کاوی و آمار وجود دارد این است که داده‌کاوی یک رهیافت بدون پیش فرض است در حالیکه بیشتر تکنیک‌های آماری معمول نیاز به پیش فرض دارند و آماردان‌ها در جستجوی معادلاتی برای مطابقت دادن با پیش فرض‌ها هستند. در مقابل الگوریتم‌های داده‌کاوی می‌توانند این معادلات را به‌طور اتوماتیک از اطلاعات موجود در مجموعه داده‌ها توسعه دهند [۱۴]. به عنوان مثال در تحقیقی، پال و همکاران (۲۰۰۹) از مدل درختی M5P برای مدل‌سازی تبخیر و تعرق در ایستگاه دیویس در کالیفرنیا استفاده کردند و پتانسیل این مدل را با روش‌های پنمن-مونتیث فائو۵۶ و هارگریوز-سامانی کالیبره شده مقایسه کردند، نتایج نشان داد از مدل مذکور به عنوان مدل مرجع در تعیین تبخیر و تعرق می‌توان استفاده نمود[۱۵]. با استفاده از مشاهدات هواشناسی

مؤسسه تحقیقات کشاورزی هند طی دوره ۴ ساله (۲۰۰۷-۲۰۰۴) شامل دمای بیشینه و کمینه، ساعات آفتابی، سرعت باد، رطوبت نسبی I و II (مربوط به ساعات ۷,۲۱ صبح و ۲,۲۱ بعد از ظهر) به مقایسه مقادیر تبخیر روزانه تشت و تخمین‌های شبکه عصبی مصنوعی، رگرسیون و مدل‌های اقلیمی پرداخته شد. همچنین در تحقیقی دیگر عملکرد شبکه‌های عصبی مصنوعی در تخمین تبخیرتعرق پتانسیل با حداقل داده هواشناسی مورد بررسی قرار گرفت [۱۷]. فتاحی و همکاران (۱۳۹۱) از شاخص بارندگی و با استفاده از روش قواعد انجمنی که یکی از روش‌های داده‌کاوی است به پیش‌بینی خشک‌سالی پرداختند. نتایج نشان داد الگوریتم تهیه شده قادر است شرایط خشکی متوسط تا رطوبت شدید همچنین شرایط خنثی با احتمال بسیار زیاد و اطمینان بالا را پیش بینی نماید [۱۸]. ترانکسو و همکاران (۲۰۱۵) با استفاده از الگوریتم RTs اقدام به پیش‌بینی سرعت باد در مناطق بادخیز نمودند[۱۹]. مرور منابع نشان می‌دهد مطالعات در زمینه‌ی مدل‌سازی متغیرهای اقلیمی مانند دمای هوا به کمک تکنیک‌های داده‌کاوی بسیار معدود بوده و هنوز کاربرد آن‌ها گسترش زیادی نیافته است در حالیکه این تکنیک‌ها می‌توانند در مناطقی که دسترسی به داده وجود ندارد و یا اندازه‌گیری آن‌ها مشکل است به کار گرفته شوند. لذا این مطالعه برآن است تا الگوریتم‌هایCART و CHAIDرا در پیش‌بینی دمای ایستگاه سینوپتیک اراک بررسی نموده و بهترین الگوریتم را معرفی نماید.

۲- مواد روش‌ها

در این مطالعه از دو الگوریتم داده‌کاوی CHAID و CART و نرم افزار SPSS Clementine استفاده شد. داده‌های هواشناسی متوسط ماهانه ایستگاه سینوپتیک هواشناسی اراک به عنوان داده‌های ورودی که شامل دمای هوا (c)، ساعات آفتابی(ساعت)، دمای نقطه شبنم(c)، رطوبت نسبی (درصد)، متوسط سرعت باد (متر بر ثانیه) و کمبود فشار بخار اشباع (میلی‌بار) طی دوره چهل‌وشش ساله (۱۹۶۰ تا ۲۰۰۵ میلادی) و از متوسط دمای ماهانه ماه بعد به عنوان متغییر خروجی به نرم افزار SPSS Clementine برای ارزیابی دو الگوریتم داده‌کاویCHAID و CART در پیش‌بینی دمای ماه بعد ایستگاه سینوپتیک اراک معرفی شدند. داده‌های هواشناسی از سایت هواشناسی کشور مورد استفاده قرار گرفت و سپس در فایل Excel به‌صورت ۸ ستون و ۵۵۳ ردیف، که هفت ستون معرف یک داده هواشناسی و یک ستون هم معرف تمام ماه‌های سال‌های ۱۹۶۰ بین ۲۰۰۵ میلادی است، مورد استفاده قرار گرفت. یکی از مزیت‌های نرم‌افزار داده‌کاوی SPSS Clementine استفاده از چند روش مختلف در برخورد با داده‌های گمشده یا ثبت نشده است. استفاده از داده تصادفی، میانگین مجموعه داده‌ها و پیش‌بینی داده گمشده از روش‌های انتخابی نرم‌افزار برای برخورد با داده‌های گمشده است. در این تحقیق از روش میانگین مجموعه داده‌ها استفاده شد. از ۷۵درصد داده‌ها به عنوان تولید مدل درختی با استفاده از الگوریتم‌های CHIAD و CART و از ۲۵ درصد آن به عنوان آزمون مدل استفاده شد. آزمون مدل‌ها توسط دو شاخص آماری ضریب رگرسیون (R^2) و متوسط مطلق خطا (MAE) پس از اجرای مدل در نرم افزار محاسبه و نشان داده شد.

۲-۱- منطقه مطالعاتی

به منظور انجام این تحقیق شناسنامه اطلاعاتی آب و هوایی ایستگاه هواشناسی سینوپتیک اراک از سازمان هواشناسی کشور تهیه شد. شهرستان اراک از شمال به تفرش، ازمغرب به همدان و ملایر، از مشرق به شهرستان محلات و از جنوب به شهرستان خمین و شازند محدود شده و ارتفاع متوسط شهرستان حدود ۱۷۰۰متر از سطح دریاست و از لحاظ وسعت دومین شهرستان استان بعد از شهرستان ساوه است. شهرستان اراک به لحاظ برخورداری از عوامل آب و هوایی (مانند مجاورت با کویر میقان، وجود ارتفاعات و...) دارای نوسانات اقلیمی است و بطورکلی تابستان‌های نسبتاً ملایم و زمستان‌های سرد تا نسبتاً سرد ازخصوصیات اقلیمی شهرستان اراک است. استان مرکزی به مرکزیت شهر اراک؛ تقریبا در مرکز ایران بین مدارهای ۳۳ درجه و ۳۰ دقیقه تا ۳۵ درجه و ۳۵ دقیقه‌ی پهنای شمالی و ۴۸ درجه و ۵۷ دقیقه تا ۵۱ درجه‌ی درازای خاوری از نصف النهارگرین‌ویچ قرار دارد.

۲-۲- نرم‌افزار داده‌کاوی SPSS Clementine

نرم‌افزارSPSS Clementine یکی ازمطرح‌ترین نرم‌افزارها در زمینه داده‌کاوی است. این نرم‌افزار از سری نرم‌افزارهای معروف SPSS بوده و مانند نرم‌افزارهای آماری قبلی از امکانات بسیار زیادی در زمینه تحلیل داده‌ها برخورداراست. آخرین نسخه این نرم‌افزار ۱۲ است که پس از انتشار این نسخه، نسخه بعدی با نامPASW Modeler منتشرشد. از مزایای این نرم‌افزار، داشتن روش‌های بسیارمتنوع برای تحلیل داده‌ها، سرعت بسیار بالا در انجام

محاسبات و استفاده ازاطلاعات پایگاه داده‌ها، داشتن محیط گرافیکی به منظور راحتی بیشرکاربر برای انجام کارهای تحلیلی می‌باشد.

۲-۳- الگوریتمCART

همان طور که نام آن نشان می‌دهد، CART ساخت درختان رگرسیونی دودویی را نیز تأمین می‌نماید. درختان رگرسیونی از درختان طبقه‌بندی ساده‌ترند زیرا معیار رشد و هرس نمودن آنها مشابه است. ساختار درخت رگرسیونی مشابه درخت طبقه‌بندی است فقط هر برگ یک مقدار حقیقی را پیش‌بینی می‌کند. الگوریتمCART یک پروسه افزار بازگشتی دودویی است که گره‌های والدین را دقیقا به دوگره فرزند منشعب می‌کند و به‌طور بازگشتی منشعب کردن را تا زمانی‌که به یک معیار همگن برسد یا معیار توقف دیگری وجود داشته باشد (مثل تمام روش‌های رشد درخت) ادامه می‌دهد. توسعه‌ها با پرسیدن یک سوال با جواب بلی یا خیر تعیین می‌شوند. CART داده‌ها را به دو زیرمجموعه افراز می‌کند بطوریکه رکوردهای درون هر زیرمجموعه نسبت به زیر مجموعه قبلی همگن‌ترند. یک فیلدپیش‌بینی‌کننده چندین مرتبه درچندین سطح در درخت تکرارمی‌شود. این الگوریتم ازجداکننده جانشین بهره می‌گیرد تا بهترین استفاده از داده با مقادیرگمشده را داشته باشد.

۲-۴- الگوریتم CHAIDCHAID

مخفف تخصیص متقابل خودکار کای- اسکوئر است. این رهیافتی است بسیارمؤثر برای رشد درخت است که به وسیله کاس (۱۹۸۰) توسعه یافت [۲۰]. با استفاده از معیار معنی‌دار یک آزمون آماری، CHAID همه مقادیر صفات متغیر هدف را ارزیابی می‌نماید. این الگوریتم مقادیری که از لحاظ آماری همگن (مشابه) قضاوت شده‌اند را با توجه به متغیر هدف متصل و تمامی مقادیر ناهمگن (غیرمشابه) را حفظ می‌کند. سپس بهترین صفت را برای تشکیل اولین شاخه در درخت تصمیم انتخاب می‌کند به‌طوری که هرگره از گروهی از مقادیر همگن از متغیر هدف انتخاب‌شده، ساخته می‌شود. این روند بازگشتی تا زمانی‌که درخت به‌طورکامل رشد کرده باشد ادامه می‌یابد. آزمون آماری استفاده شده بستگی به سطح اندازه‌گیری شده متغیر هدف دارد. اگر متغیر هدف پیوسته باشد آزمون F و اگر متغیر هدف گسسته باشد آزمون آماری کای- اسکوئر استفاده می‌شود.CHAID یک روش درختی دودویی نیست زیرا می‌تواند بیش ازدو طبقه در هر سطح ویژه در درخت تولید کند. بنابراین تمایل دارد تا درخت وسیع‌تری نسبت به روش‌های رونده دودویی تولیدکند. این الگوریتم با انواع از متغیرها کار می‌کند و هم وزن و هم فراوانی متغیرها را می‌پذیرد.

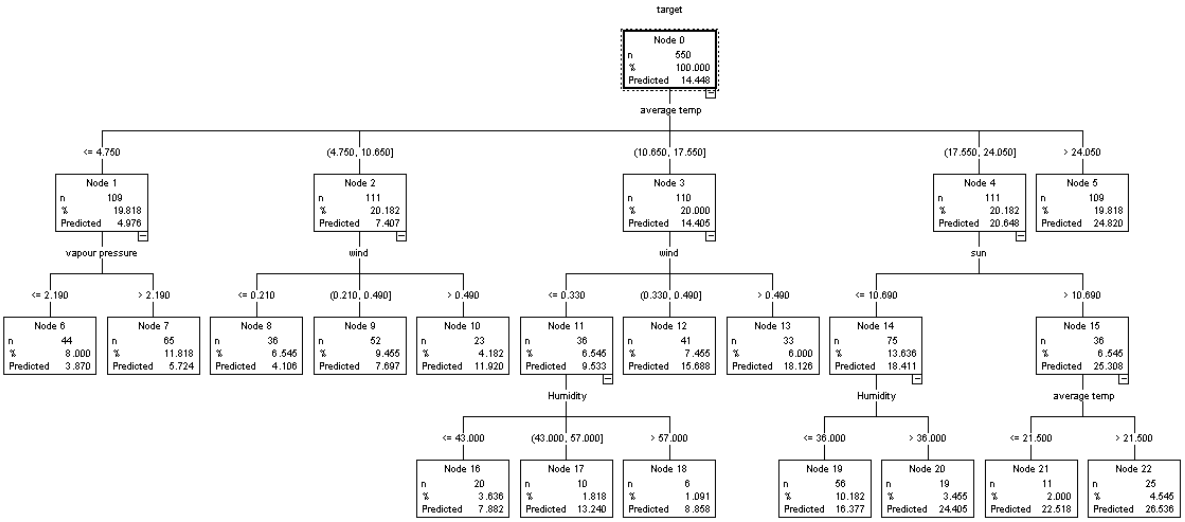
۳- نتایج

از دو الگوریتم نرم افزار داده‌کاوی کلمنتاین برای پیش‌بینی دمای متوسط ماه بعد ایستگاه هواشناسی سینوپتیک اراک استفاده شد. لذا برای پیش‌بینی متوسط دمای ماه بعد از سری داده‌های هواشناسی متوسط ماهانه ایستگاه هواشناسی سینوپتیک اراک استفاده شد. همانطور که در جدول۱ مشاهده می‌شود مدل درختی CHAID با مقدار ضریب همبستگی ۰/۹۱۵ و MAE برابر ۲/۷۷ و الگوریتم CART با مقدار ضریب همبستگی۰/۸۷۲ و MAE برابر ۳/۴۰۵ برای پیش‌بینی متوسط دمای ماهانه برای ماه بعد محاسبه شد. از این محاسبات نتیجه می‌گیریم که مدل درختی CHAID برای پیش‌بینی متوسط دمای ماهانه ماه بعد عملکرد بهتری داشته است. مدل درختی CHAID با تخمین مناسبی از دمای متوسط ماهانه برای ماه بعد نشان داد که دارای توانایی بالایی در تخمین پارامترهای هواشناسی است.

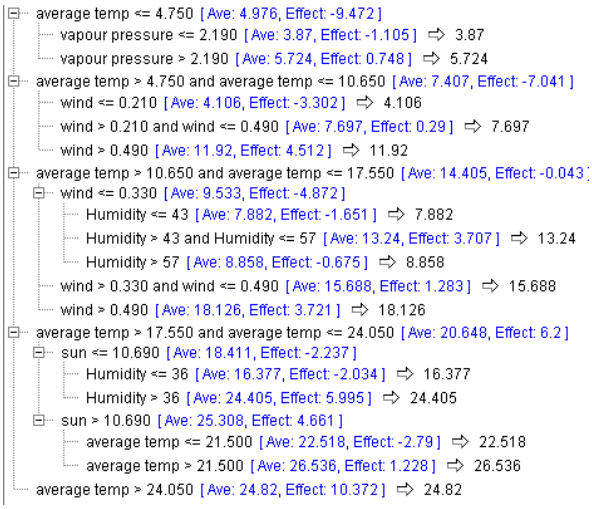
جدول ۱- ضریب همبستگی و متوسط مطلق خطا دو مدل درختی CART و CHAID

آزمون آماری الگوریتم‌ها	R2	MAE
CHAID	۲/۷۷	۰/۹۱۵
CART	۳/۴۰۵	۰/۸۷۲

با توجه به شکل۱ نمودار درختی الگوریتم CHAID از چند شاخه و برگ تشکیل شده است. الگوریتم CHAID در اولین شاخه در نمودار درختی (شکل۱) متوسط دما را به کار برده و نشان‌دهنده تأثیر بیشتر دمای ماه قبل در پیش‌بینی دمای ماه بعد است. در ادامه ترسیم نمودار درختی توسط مدلCHAID مقادیر دمای هوا به چهار برگ تقسیم می‌شود که هر کدام از این برگ‌ها به چند شاخه با مقادیر داده هواشناسی مختلف وصل می‌شود. در نهایت نمودار درختی به ۲۳ برگ و ۲۶ گره تقسیم می‌شود.



شکل۱– نمودار درختی مدل CHAID ایستگاه هواشناسی سینوپتیک اراک



شکل۲–رابطه خطی مدل CHAID توسط نرم افزار SPSS Clementine ایستگاه هواشناسی سینوپتیک اراک

۴- نتیجه‌گیری

با توجه به تحقیقات صورت گرفته تغییرات دما در نقاط گوناگون دنیا گزارش شده است. تغییر هر چند کم در دما موجب تغییر در وقوع پدیده‌هایی نظیر خشک‌سالی، بارش‌های سنگین و طوفان می‌شود لذا پیش‌بینی دما در جهت مدیریت صحیح در وضعیت بحرانی جوی امری ضروری به نظر می‌رسد. استفاده از تکنیک داده‌کاوی از جمله مدل درختی CART می‌تواند جایگزینی در برآورد دمای متوسط ماهانه بر روابطی که تاکنون ارائه شده است باشد. مدل‌های درخت رگرسیونی با توانایی بالایی قادر به پیش‌بینی متوسط دمای ماه بعد هستند. نتایج تحقیق نشان داد مدل درختی CHAID در مقایسه با مدل CART عملکرد بهتری در پیش‌بینی متوسط دمای ماهانه ماه بعد دارد. لذا با بکارگیری این مدل می‌توان داده‌های هواشناسی، ثبت نشده در ایستگاه‌هایی که به علل مختلف دچار نقصان شده‌اند را تعیین نمود.

منابع

[1] Babazadeh H, Shamsnia SA, Bostani F, Norozi eghdam A, Khodakaramidehkordi D. Evaluation of drought, wet and prediction of Shiraz climatic parameters precipitation and temperature by using stochastic methods. Journal of Geography and Urban Planning; **2012**;**16** (41):23-47. **[In Persian]**

[2] Sfandiari F, Hosseini SA, Azadimobaraki M, Hejazizadeh Z. Predict the average monthly temperature in Sanandaj station using the model (MLP) MLP neural network.Iran Geographic; **2010**;**8**(27):45- 65. **[In Persian]**

[3] Bootsma A. Long-term (100 years) climate trends for agriculture at selected locations in Canada. Climatic Change, **1994; 26**: 65–88.

[4] Plummer N, Salinger M J, Nicholls N, Suppiah R, Hennessy K J, Leighton R M, Trewin B, Page C M, Lough J M. Changes in climate extremes over the Australian region and New Zealand during the twentieth century. Climatic Change; **1999**; **42**: 183–202.

[5] Heino R, Brázdil R, Forland E, Tuomenvirta H, Alexandersson H, Beniston M, Pfister C, Rebetez M, Rosenhagen G, Rösner S, Wibig J. Progress in the study of climatic extremes in northern and central Europe. Climatic Change; **1999; 42**: 151–181.

[6] Khalili A. Report of the National Water Master Plan, the synthesis report, the weather, the possible effect of climate change on water resources. Ministry of Energy; **2001.[In Persian]**

[7] Shahabfar A, Mohammadnia gharaii S, Javedani Khalifa N, The time variations of frost days in Mashhad. Proceedings of the Third Regional Conference and the National Conference on Climate Change; **2004:** 74-81.

[8] Wang B, Zhang M, Wei J, Wang S, Li S, Ma Q, Li X, Pan S. Changes in extreme events of temperature and precipitation over Xinjiang, northwest China, during 1960–2009. Quaternary International; **2013**;17(298):141-151.

[9] Balling JRRC, Idso SB. Effects of greenhouse warming on maximum summer temperatures. Agric for meteoro; **1990; l53**:143–147.

[10] Giudici P. Applied data Mining: statistical methods for business and industry. Wily, London; **2003:** 364.

[11] Mahmoudi A, Rostami H, Canopy M, Moradi A. A review of the science of data mining and its applications in the offshore industry. Fifth National Conference on offshore industries OIC. Sharif University; **2014. [In Persian]**

[12] Vanderberg H, Sogard P, Motoroni S. Mine Set TM 3.0 Enterprise Edition Tutorial for Windows , Silicon Graphics Inc. **1999;** Doc. No. 007-4006-001.

[13] Meshkani A, Nazmi A. Introduction to Data Mining, Islamic Azad University. Neyshabur; **2010:**456.**[In Persian]**

[14] Cabena PH, Stadler R, Verhees J, Zanasi. Discovering data mining: From concept to implementation, IMB, New Jersey; **1998:**195.

[15] Pal M, Deswal S. M5 model tree based modeling of reference evapotranspiration. Hydrol Process; **2009**; **23**:1437-1443.

[16] Shirsath PB, Singh AK. A comparative study of daily pan evaporation estimation using ANN, Regression and climate based models. J. Water resource management; **2010; 24**:1571-1581.

[17] Diamantopoulou MJ, Georgiou PE, Papamichial DM. Perforeance evaluation of artificial neural network in estimating reference evapotranspiration with minimal meteorological data. Global nest Journal; **2010;**13 (1):18-27.

[18] Fatahi MH, Bamdad A, Rahimikhob A. Application of association rules to monitor rainfall and drought events using sea surface temperature (Case Study: Khozestan). Journal of Water Resources Engineering; **2012**: 109- 118. **[In Persian]**

[19] Troncoso A, Salcedo-Sanz S, Casanova-Mateo C, Riquelme JC, Prieto L. Local models-based regression trees for very short-term wind speed prediction. Renewable Energy; **2015**;**81**:589-98.

[20] Kass G. An Exploratory technique for investigating large quantities of categorical data. Applid Statistics; **1980; 1**:119-127.