

بررسی و برآورد تاثیر عوامل مختلف بر تغییرات عمق آبخوان با استفاده از الگوریتم درختی، مطالعه موردی: دشت تیروان و کرون

نگار اکبری افوسی^۱، مسعود نصری^{۲*} و سیدحسن میرهاشمی^۳

^۱ دانش‌آموخته کارشناسی ارشد، دانشگاه آزاد اسلامی، واحد اردستان، ^۲ استادیار، دانشگاه آزاد اسلامی، واحد اردستان و

^۳ دانش‌آموخته دکتری، دانشکده آب و خاک، دانشگاه زابل

تاریخ پذیرش: ۱۳۹۹/۰۳/۱۱

تاریخ دریافت: ۱۳۹۸/۰۹/۲۵

چکیده

هدف از این پژوهش، استفاده از الگوریتم درختی داده‌کاوی، به منظور مدیریت مناسب بر آبخوان دشت تیروان و کرون است. در این خصوص، از هفت عامل مختلف انسانی و طبیعی تأثیرگذار بر تغییرات عمق آبخوان استفاده شد. در ابتدا، پیش‌بینی سه الگوریتم درختی CART، CHAID و MP5 در تغییرات آبخوان با استفاده از شاخص‌های آماری مورد ارزیابی قرار گرفتند. الگوریتم CAHID با توجه به ضریب رگرسیون برابر ۰/۸۲ و متوسط مطلق خطا برابر ۰/۱۲ دارای عملکرد بهتری نسبت به دو الگوریتم CART و MP5 است. بیشترین مقدار بالآمدگی آبخوان در ماه‌های آذر، دی، بهمن، اسفند و در زمانی که مقدار حجم بارندگی بین ۰/۸ تا ۰/۷۲ میلیون متر مکعب و درصد رطوبت هوا بیشتر از ۷۲ درصد بوده و همچنین، بیشترین مقدار افت آبخوان در ماه‌های شهریور، مرداد و در زمانی که دمای هوا بیش از ۲۵ درجه سانتی‌گراد و حجم آب برداشتی از چاه‌های کشاورزی بیش از ۱/۳۲ میلیون متر مکعب بوده است، به وسیله نمودار درختی الگوریتم CHAID پیش‌بینی شد. از عوامل طبیعی، دمای هوا و از عوامل انسانی، حجم آب برداشتی از چاه کشاورزی بیشترین تأثیر را در تغییرات عمق آبخوان در دشت مذکور داشته است. دو عامل درصد رطوبت هوا و مقدار حجم بارش، تنها عواملی بودند که رابطه مستقیم با بالآمدگی عمق آبخوان داشته‌اند. تأثیرگذارترین عوامل در پیش‌بینی مقدار تغییرات عمق آبخوان دشت تیروان و کرون به ترتیب دمای هوا، حجم آب برداشتی از چاه کشاورزی و حجم بارندگی و بقیه پارامترها تقریباً تأثیرشان با هم برابر بوده است.

واژه‌های کلیدی: افت آبخوان، الگوریتم CHAID، چاه کشاورزی، دمای هوا، مدیریت آبخوان

مقدمه

هیدرولوژی بوده، بخش دیگر فعالیت‌های انسانی مؤثر بر روی منابع آب است. Ganji Khoramdel و همکاران (۲۰۰۹)، به منظور بهینه‌کردن شبکه پایش، روشی را ارائه دادند که بتوان با دقت مناسب بیلان آب زیرزمینی را تخمین زد. در این روش، با استفاده از زمین‌آمار، درجه اهمیت هر چاه مشاهده‌ای بررسی شده و تأثیر آن در تعیین بیلان آب زیرزمینی به دست

امروزه، با توجه به توسعه و پیشرفت کشاورزی و همچنین، افزایش جمعیت، نیاز روزافزون به منابع آب، بیشتر مورد توجه قرار گرفته است. استفاده از مدیریت سنتی در بیشتر موارد نمی‌تواند جوابگوی این حجم مصرف باشد. بخش عمده‌ای از مدیریت نامناسب، ناشی از محدودیت طبیعی منابع آب و چرخه

آذر سال ۱۳۹۲ تا ماه مرداد سال ۱۳۹۳ جمع‌آوری و سازماندهی کردند. سپس، نقشه‌های خطوط هم‌سطح آب زیرزمینی تهیه شد. نتایج، حاکی از افزایش موضعی سطح آب زیرزمینی در نقاط مختلف شهر مشهد بود. نتایج این بررسی‌ها طی جلساتی با حضور کارشناسان تمامی دستگاه‌های حاکمیتی مرتبط، اساتید دانشگاه و کارشناسان خبره و در قالب کارگاه مهندسی ارزش به نقد گذاشته و تصمیماتی در خصوص چگونگی برخورد با این پدیده اتخاذ شد. به‌منظور کاهش مخاطرات ناشی از بالآمدگی، با بهره‌برداری از چاه‌های سطح شهر، طی زمستان سال ۱۳۹۲ و بهار سال ۱۳۹۳ سطح آب به میزان چشمگیری پایین آمد. طی این مدت، هشدارهای فنی از طریق دستگاه‌های متولی حسب تصمیمات کارگروه برای ساخت و سازها ابلاغ شد و خوشبختانه گزارش خاصی در مورد هرگونه خرابی ساختمان و تاسیسات ناشی از بالآمدگی گزارش نشد. در ادامه، نتایج حاصل از بررسی داده‌های میدانی و همچنین فرصت‌ها، چالش‌ها و مدیریت شرایط، مورد بررسی قرار گرفته‌اند. با توجه به شرایط خشکسالی ایران که به‌صورت طبیعی میزان عرضه آب به مناطق کشاورزی و مقدار بارش کاهش می‌یابد، اگر کاهش قابل توجهی در بخش مصرف آب کشاورزی که با الگوی کشت مناسب رخ می‌دهد، انجام گیرد، می‌توان احتمال وقوع افت زیاد عمق آبخوان دشت قزوین را به مقدار زیادی کاهش داد (Mirhashemi و همکاران، ۲۰۱۶).

همچنین، مدیران در بخش مدیریت منابع آب با روابط پیچیده و خصوصیات بسیار متنوع از حجم بسیار زیادی از داده‌های جمع‌آوری شده مواجه هستند که تجزیه و تحلیل و مدیریت آن‌ها به‌وسیله روش‌های تجربی و آماری، امری دشوار و در بسیاری از حوضه‌ها عملاً ناممکن است. داده‌کاوی یک فرایند شناخت الگوهای معتبر، جدید، ذاتا مفید و قابل فهم از داده‌ها می‌باشد. داده‌کاوی شاخه توسعه‌یافته و پیچیده علم آمار است. یکی از ویژگی‌های مهم داده‌کاوی، یک سامانه پشتیبان تصمیم برای پیش‌بینی آینده و اتفاقات محتمل پیش‌رو است. Mirhashemi و همکاران (۲۰۱۶) با توجه به نتایج الگوریتم داده‌کاوی CART مشخص کردند که مؤثرترین پارامتر در

آمد. استخراج و بهره‌برداری بیش از حد منابع آب زیرزمینی نسبت به تغذیه آن‌ها، منجر به افت و محدودیت این منابع شده است (Ebrahimi و همکاران، ۲۰۱۶). Poormohammadi و همکاران (۲۰۱۱)، به ارزیابی و برآورد اجزای بیلان آبی در حوضه‌های مناطق خشک با به‌کارگیری سنجش از دور و سامانه اطلاعات جغرافیایی حوزه آبخیز منشا یزد پرداختند. نتایج آنان نشان داد که بیشتر هدررفت آب در این حوضه به‌وسیله تبخیر و تعرق واقعی صورت می‌گیرد که مقدار آن در سال آبی ۸۶-۸۵ حدود ۵۴۰ میلی‌متر در سال و هدررفت به‌وسیله رواناب و نفوذ، نزدیک به هم و به‌ترتیب ۱۱۷ و ۱۲۵ میلی‌متر در سال است.

خشکسالی‌های اخیر و پایین رفتن سطح ایستابی آب‌های زیرزمینی و خشک‌شدن چاه‌های شمالی دشت فریدن، کشاورزی منطقه را با مخاطره و چالش جدی روبه‌رو ساخته است. افت سطح آب‌های زیرزمینی این دشت، ذهن را بدین سو متبادر می‌سازد که تا چه اندازه این موضوع با نوسانات بارش ارتباط می‌یابد (Arvin و همکاران، ۲۰۱۶). Sabuhi و Tavana (۲۰۰۷)، با مطالعه بر روی آثار جانبی و منفی ناشی از بهره‌برداری بیش از حد منابع آب زیرزمینی دشت لارستان مشخص کردند که با کم‌کردن دبی آب چاه‌های آبیاری از طریق ایجاد محدودیت در مجوز احداث چاه، به‌خصوص چاه‌های با دبی زیاد، این آثار جانبی و منفی کاهش می‌یابد. احداث چاه در جایی از مزرعه که کمینه فاصله را نسبت به محدوده زمین زراعی داشته باشد، از مصرف بیش از حد آب و در نتیجه ایجاد آثار جانبی منفی جلوگیری می‌کند.

امروزه، تحقیقات مختلفی روی آب‌های سطحی و زیرزمینی انجام شده است تا شناختی اصولی‌تر در مدیریت منابع آب حاصل شود. Hazarika و Shahid (۲۰۰۹)، تاثیر خشکسالی بر آب زیرزمینی در شمال غربی بنگلادش را مورد بررسی قرار دادند. نتایج آن‌ها نشان داد که کاهش سطح آب زیرزمینی در ۴۲ درصد از این منطقه، هر ساله پدیدار است. به‌منظور بررسی تغییرات مکانی و زمانی سطح آب زیرزمینی، Ghandehary و همکاران (۲۰۱۴)، آمار ۱۱۱ حلقه چاه مشاهده‌ای شهر مشهد را طی شش نوبت از ماه

CART و RF بودند. بر اساس نتایج، میزان صحت هر سه مدل بیش از ۷۰ درصد است. بنابراین، هر سه مدل می‌توانند برای برنامه‌ریزان و مهندسان در مدیریت منابع آب و زمین و برنامه‌ریزی در منطقه مطالعاتی مورد استفاده قرار گیرند.

بخش کشاورزی به‌عنوان بزرگترین مصرف‌کننده آب زیرزمینی محسوب می‌شود. استفاده بیش از اندازه از آب‌های زیرزمینی مشکلات اساسی را در بخش کشاورزی ایجاد کرده است. منابع آب زیرزمینی از اصلی‌ترین منابع آبی مورد استفاده در بخش کشاورزی شهرستان تیروان و کرون است. هدف از این پژوهش، استفاده از الگوریتم درختی داده‌کاوی در بررسی و پیش‌بینی عوامل موثر بر تغییرات عمق آبخوان است. در ابتدا، نتایج الگوریتم‌های درختی در پیش‌بینی دشت مورد مطالعه با استفاده از شاخص‌های آماری بررسی شدند و سپس، با استفاده از الگوریتم منتخب، به بررسی و پیش‌بینی تغییرات عمق آبخوان با توجه به عوامل مختلف انسانی و طبیعی پرداخته شد.

مواد و روش‌ها

مشخصات منطقه مطالعاتی: شهرستان تیروان و کرون (کروند) در $9^{\circ} 51'$ طول جغرافیایی و 42° عرض جغرافیایی و ارتفاع ۱۶۴۰ متری از سطح دریا و در ۴۵ کیلومتری شهر اصفهان واقع شده است. تیروان و کرون از جنوب غربی به استان چهارمحال بختیاری و از شرق و شمال به شهرستان نجف‌آباد، از غرب به شهرستان‌های فریدن و چادگان از جنوب به شهرستان لنجان محدود می‌شود. این شهرستان، دارای هوایی نسبتاً معتدل و زمستان‌های سرد است. از محصولات مهم کشاورزی این شهرستان می‌توان به بادام، گردو، انگور، انار، زردآلو، آلبالو، هلو، زعفران و محصولات گلخانه‌ای اشاره کرد.

داده‌های مورد استفاده در این پژوهش، از مجموعه اطلاعات مربوط به دشت تیروان و کرون از سازمان آب منطقه‌ای اصفهان، جهاد کشاورزی و ایستگاه سینوپتیک استان اصفهان طی سال‌های ۱۳۸۲ تا ۱۳۹۶ به‌صورت ماهانه است. اطلاعات مربوط به چاه‌های مشاهداتی، مقادیر بارندگی، حجم آب ورودی به شبکه آبیاری از سازمان آب منطقه‌ای، مقادیر حجم آب برداشتی از چاه کشاورزی و محصولات زراعی و

تغییرات عمق آبخوان منطقه تحت آبیاری شهرستان قزوین، مقدار دما است. به‌طوری‌که افزایش دما باعث افزایش افت می‌شود و بیشترین احتمال افت در ماه‌های تیر، مرداد، شهریور رخ می‌دهد. لذا، بهتر است که در این سه ماه که معمولاً دارای بیشترین مقدار دمای هوا و افت آبخوان هستند، مدیران و کشاورزان از الگوی کشت مناسب‌تر و کم‌مصرف‌تر استفاده کنند تا مشکلات ناشی از افت کمتر رخ دهد. نتایج پیش‌بینی Zhao و همکاران (۲۰۱۶)، نشان داد که مدل درختی داده‌کاوی CART با همبستگی بین متغیرها و کاهش اطلاعات غیر مفید می‌تواند صحت دقت پیش‌بینی را افزایش دهد و نتیجه مقایسه مدل‌های CART و PSO-SVR، مشخص کرد که مدل CART با دقت بهتر و توانایی پیش‌بینی بهتر، می‌تواند برای پیش‌بینی کردن افت سطح آب زیرزمینی استفاده شود. همچنین، نتایج کار Huajie و همکاران در سال ۲۰۱۶، نشان داد که مدیران و تصمیم‌گیرندگان آب‌های زیرزمینی با استفاده از اطلاعات درخت تصمیم‌گیری داده‌کاوی می‌توانند در اجرای برنامه‌هایی برای حفاظت از منابع آب‌های زیرزمینی، حمایت کنند (Stumpp و همکاران، ۲۰۱۶). نتایج به‌دست آمده به سه صورت خلاصه شد. (۱) بر اساس تحلیل عوامل نفوذ زیرزمینی، پنج موضوع لایه، سنگ شناسی، ژئومورفولوژی، شیب، NDVI و NDWI، می‌تواند به‌عنوان شاخص کافی برای پتانسیل آب‌های زیرزمینی استفاده کرد. (۲) با این پیش‌دست‌آورد، الگوریتم درختی داده‌کاوی QUEST به‌طور موثر در ساختن درخت تصمیم‌گیری برای طبقه‌بندی منطقه مورد مطالعه می‌تواند استفاده شود. (۳) نتیجه طبقه‌بندی به دقت قابل قبولی با مقدار درجه صحت ۹۲/۸ درصد بوده است. Naghibi و همکاران (۲۰۱۶)، اقدام به شناسایی مناطقی از آب‌های زیرزمینی با پتانسیل بالا به‌وسیله یک روش الگوریتم درختی داده‌کاوی CART و دو روش RF و BRT کردند و در نقشه‌برداری ذخایر چشمه‌های حوضه کوه‌رنگ با استفاده از مدل‌های BRT، CART و RF، میزان صحت مدل‌ها به‌ترتیب ۰/۸۱، ۰/۷۸ و ۰/۷۱ به‌دست آمد. بنابراین، مدل BRT بهترین عملکرد را در نقشه‌برداری ذخایر چشمه‌های آب زیرزمینی داشته و بعد از آن به‌ترتیب مدل‌های

آزمایشات نشان می‌دهد که هموارسازی به صورت متعاقب صحت پیش‌بینی‌ها را افزایش می‌دهد.

دقیقا همان فرایند هموارسازی می‌تواند پس از ساخت درخت با شرکت‌دادن مدل‌های داخلی در هر برگ، مدل انجام شود. در طی فرایند طبقه‌بندی تنها مدل‌های بزرگ استفاده می‌شوند. عیب آن این است که مدل‌های برگ، بزرگ‌تر و فهم آن پیچیده‌تر می‌شود. زیرا بسیاری از ضرابی که قبلا صفر بودند، به هنگام شرکت‌دادن مدل‌های داخلی غیر صفر می‌شوند.

مدل CART: Chattamvelli (۲۰۱۱)، نتیجه گرفت که الگوریتم CART متغیرهای ورودی را برای یافتن بهترین تجزیه، می‌آزماید، تا شاخص ناخالصی حاصل از تجزیه کمترین مقدار باشد. در تجزیه دو زیر گروه تعیین می‌شود و هر کدام در مرحله بعد به دو زیر گروه دیگر تقسیم خواهند شد و این روند ادامه می‌یابد تا زمانی که یکی از معیارهای توقف برآورده شود. درخت CART بازگشتی دودویی است، که گره‌های والدین را دقیقا به دو گروه فرزند منشعب می‌کند و به‌طور بازگشتی منشعب‌کردن را تا زمانی که انشعاب دیگری نتواند ساخته شود، ادامه می‌دهد. در تقسیم داده‌ها به‌صورت بازگشتی به زیرگروه‌ها فقط دو زیر گروه را پشتیبانی می‌کند و یک الگوریتم سریع است که هرس نمودن آن رو به عقب است.

مدل CHAID: Chattamvelli (۲۰۱۱)، بیان کرد که به‌طور کلی روش CHAID برای ساخت یک درخت تصمیم‌گیری، داده‌ها را به‌نوبت به زیرمجموعه‌های مشابه افراز می‌کند تا آنجا که هر زیر مجموعه شامل تعداد مشخصی نمونه شود. این الگوریتم می‌تواند درختی تولید کند که در برخی از مواقع به‌صورت غیر دودویی عمل کند. در واقع، از روش جداکردن چندتایی به‌جای جداکردن دودویی استفاده می‌کند. به این‌صورت که می‌توان گره پدر را به تعداد بیش از دو بار تقسیم کند. این الگوریتم از آزمون کای دو برای تصمیم‌گیری در هر تقسیم برای مشخص کردن نودهای فرزند استفاده می‌کند. سپس، شاخه‌های درخت ساخته شده تا تحقق معیار توقف یا رسیدن به سطح پیچیدگی خواسته‌شده، هرس می‌شوند. به بیان دیگر، CHAID ابتدا تفاوت‌های هر نمونه را با سایر نمونه‌ها می‌یابد و درخت مورد نظر را تولید می‌کند.

باغی از جهاد کشاورزی و همچنین، مقادیر دمای هوا، سرعت باد، درصد رطوبت هوا و ساعات خورشیدی از ایستگاه سینوپتیک استان اصفهان اخذ شدند. متغیرهای ورودی مورد استفاده، شامل مقدار حجم آب برداشتی از چاه‌های کشاورزی (میلیون متر مکعب)، حجم آب ورودی به شبکه آبیاری (میلیون متر مکعب)، مقدار حجم تقاضای آب کشاورزی (میلیون متر مکعب)، مقدار حجم بارندگی (میلیون متر مکعب)، دمای هوا (سانتی‌گراد)، رطوبت هوا (درصد) و تبخیر و تعرق پتانسیل (میلی‌متر در روز) هستند.

مدل‌های درختی: وقتی یک مدل درختی برای پیش‌بینی مقدار برای یک نمونه آزمایشی استفاده می‌شود، درخت به‌صورت نرمال به سمت پائین تا یک برگ پیش می‌رود و از مقادیر صفات نمونه برای تصمیم‌گیری در هر گره استفاده می‌کند. برگ شامل یک مدل خطی بر مبنای برخی صفات خواهد بود و برای هر نمونه آزمایشی به‌منظور بازده یک مقدار خام پیش‌بینی شده ارزیابی می‌شود. به جای استفاده از این مقدار خام به‌صورت مستقیم، بهتر است از فرایند هموارسازی برای جبران گسستگی‌های شدیدی که ناگزیر بین مدل‌های خطی مجاور در برگ‌های درخت هرس شده رخ می‌دهد، استفاده کرد. این مسئله خاصی برای مدل‌هایی است که از تعداد کمی نمونه‌های آموزشی استفاده می‌کنند. هموارسازی می‌تواند با تولید مدل‌های خطی هر گره داخلی و نیز برای برگ‌ها در زمانی که درخت ساخته می‌شود، انجام شود. سپس، وقتی مدل برگ برای به‌دست آوردن یک مقدار خام پیش‌بینی شده برای یک نمونه آزمایشی استفاده می‌شود، این مقدار در طول مسیر به طرف عقب تا ریشه فیلتر می‌شود و در هر گره با ترکیب آن با مقدار پیش‌بینی شده با مدل خطی برای آن گره هموار می‌شود.

یک رهیاف محاسبه هموارسازی عبارت از:

$$P' = \frac{np+kq}{n+k} \quad (1)$$

که در آن، P' مقدار پیش‌بینی رسیده به گره مجاور بالایی، p مقدار پیش‌بینی که از زیر به این گره رسیده است و q مقدار پیش‌بینی شده به‌وسیله مدل در این گره و n تعداد نمونه‌های آموزشی است که به گره زیری می‌رسد و k ثابت هموارسازی است.

داده‌های آزمون که ویژگی هدف آن‌ها به‌وسیله مدل، درست تشخیص داده شده بود، دقت مدل را بیان می‌کند (Gupta, ۲۰۱۱). برای هر سه الگوریتم درخت تصمیم، ۷۵ درصد داده‌ها به‌صورت تصادفی به‌عنوان داده‌های آموزش انتخاب خواهند شد و ۲۵ درصد مابقی به‌عنوان داده‌های آزمون مورد آزمایش قرار خواهند گرفت.

از داده‌های چندین ساله مربوط به منابع آب زیرزمینی برای پیش‌بینی الگوریتم‌ها استفاده خواهد شد و با چند آزمون آماری می‌توان مقدار پیش‌بینی صحیح الگوریتم‌ها را ارزیابی کرد و پس از برآورد مقادیر هدف، به‌منظور بررسی عملکرد مدل‌ها از شاخص‌های آماری استاندارد بهره گرفته شد و از شاخص‌های ضریب تبیین (R) و متوسط مطلق خطا (MAE) به‌منظور ارزیابی الگوریتم‌ها استفاده شد.

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_{obs} - y_{pred})^2}{\sum_{i=1}^n (y_{obs} - y_{mean})^2}} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^n |y_{obs} - y_{pred}|}{n} \quad (4)$$

که در آن‌ها، n تعداد داده‌های مشاهده‌شده، y_{obs} داده‌های برداشت‌شده، y_{pred} عمق سطح آب‌های زیرزمینی پیش‌بینی شده، y_{mean} میانگین عمق سطح آب‌های زیرزمینی و y_{obs} مقادیر مشاهده‌شده است.

نتایج و بحث

در ابتدا، نتایج سه الگوریتم درختی به‌وسیله شاخص‌های آماری مورد ارزیابی قرار گرفت. با توجه به جدول ۱، بهترین عملکرد مربوط به الگوریتم CHAID با مقدار ضریب رگرسیون برابر ۰/۸۲ و متوسط مطلق خطا برابر ۰/۱۲ است.

هرس کردن درخت از طریق یافتن تفاوت‌های مشابه انجام می‌شود

مدل MSP: مدل‌های درخت رگرسیونی، مفاهیم رگرسیون را با مقادیر ثابت در برگ‌هایش به‌صورت کلی بیان می‌کنند (Frank و Witten, ۲۰۰۵). این مدل‌ها را می‌توان شبیه به توابع رگرسیون piece-wise و غیر خطی معرفی کرد. مدل MSP یک درخت رگرسیونی دودویی است که در گره‌های نهایی خود (برگ‌ها) توابع رگرسیونی خطی دارند که می‌تواند صفات عددی پیوسته تولید نمایند. مدل‌های بر مبنای درخت به روش تقسیم و نتیجه‌گیری ساخته می‌شوند. تولید مدل درختی به دو مرحله نیاز دارد. اولین مرحله شامل استفاده از یک معیار انشعاب برای تولید یک تصمیم درختی است. معیار انشعاب برای الگوریتم مدل درختی MSP بر اساس رفتار انحراف معیار، مقادیر طبقه‌ای است که به‌عنوان کمیتی از خطا به یک گره می‌رسد و کاهش مورد انتظار در این خطا را به‌عنوان نتیجه آزمون هر صفت در آن گره محاسبه می‌کند. فرمول محاسبه کاهش انحراف معیار (SDR) به‌صورت رابطه (۲) است.

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (2)$$

که در آن، T بیانگر یکسری نمونه‌هایی است که به گره می‌رسد، T_i بیانگر یکسری نمونه‌هایی است که i -امین برآمد تست پتانسیلی را دارند، sd بیانگر انحراف معیار است.

به‌منظور اعتبارسنجی مدل درختی نیز داده‌ها به دو بخش داده‌های آموزش و آزمون تقسیم می‌شوند. مدل درختی با استفاده از داده‌های آموزش ساخته می‌شود و مدل ساخته‌شده بر روی داده‌های آزمون مورد تست قرار خواهد گرفت. درصد نمونه‌هایی از

جدول ۱- ارزیابی نتایج الگوریتم‌های CART، CHAID و MP5 به‌وسیله شاخص‌های آماری

MAE	R	الگوریتم درختی
۰/۱۹	۰/۷۸	CART
۰/۱۲	۰/۸۲	CHAID
۰/۲۱	۰/۷۱	MP5

پرت تا حد زیادی از بین می‌رود. در این پژوهش، داده‌های مربوط به نوسانات عمق آبخوان به صورت شش مدل تقسیم‌بندی انجام شد و همچنین، سعی شد مرزهای بین مقادیر گسسته طوری تعیین شوند که هر مدل به اندازه کافی معرف نمونه داده‌های مربوطه باشد. بعد از انجام گسسته‌سازی برای کل داده‌ها، برای هر محدوده داده به وسیله نام مدل از A تا F نام‌گذاری شد (جدول ۲). به طوری که مقدار افت سطح آب زیرزمینی به صورت نزولی از مدل A تا C کاهش و مقدار بالآمدگی به صورت صعودی از مدل D تا F افزایش می‌یابد. علامت منفی محدوده داده‌ها، مربوط به افت آبخوان است.

با توجه به این‌که الگوریتم CHAID با در نظر گرفتن نتایج شاخص‌های آماری نسبت به دو الگوریتم MP5 و CART عملکرد بهتری داشته است، در ادامه از نتایج الگوریتم CHAID به منظور پیش‌بینی و بررسی وضعیت آبخوان دشت تیروان و کرون استفاده شد.

تحلیل ویژگی‌های درخت تصمیم‌گیری با استفاده از داده‌های گسسته تغییرات عمق آبخوان: برای کاهش هر چه بیشتر اثرات خطاها و همچنین افزایش قابلیت استفاده نتایج برای مدیران، مقادیر تغییرات عمق آب زیرزمینی، دسته‌بندی شد. با استفاده از دسته‌بندی، اثر خطاهای اندازه‌گیری و مقادیر داده‌های

جدول ۲- تعیین محدوده داده‌های تغییرات عمق آبخوان (متر)

نام مدل مربوط به محدوده داده‌ها	محدوده داده‌ها	میانگین داده‌ها	درصد داده‌ها
A	$-1 \geq \dots$	-۱/۶۶	۱۲/۷۷
B	$-0.5 \geq \dots > -1$	-۰/۵۶	۲۳/۹۲
C	$0 > \dots > -0.5$	-۰/۱۵	۲۱/۴
D	$0.5 \geq \dots > 0$	۰/۱۵	۱۹/۷۶
E	$1 \geq \dots > 0.5$	۰/۵۵	۱۴/۶۸
F	$\dots > 1$	۱/۷۳	۷/۵۰

بوده است، به وسیله نمودار درختی الگوریتم CHAID پیش‌بینی شده است. همان‌طور که از نمودار درختی مشخص است، عوامل حجم بارش و درصد رطوبت هوا از عوامل مهم در مقدار بالآمدگی آبخوان و عوامل حجم آب برداشتی از چاه کشاورزی، دمای هوا و تبخیر و تعرق پتانسیل از عوامل مهم در مقدار افت آبخوان می‌باشند.

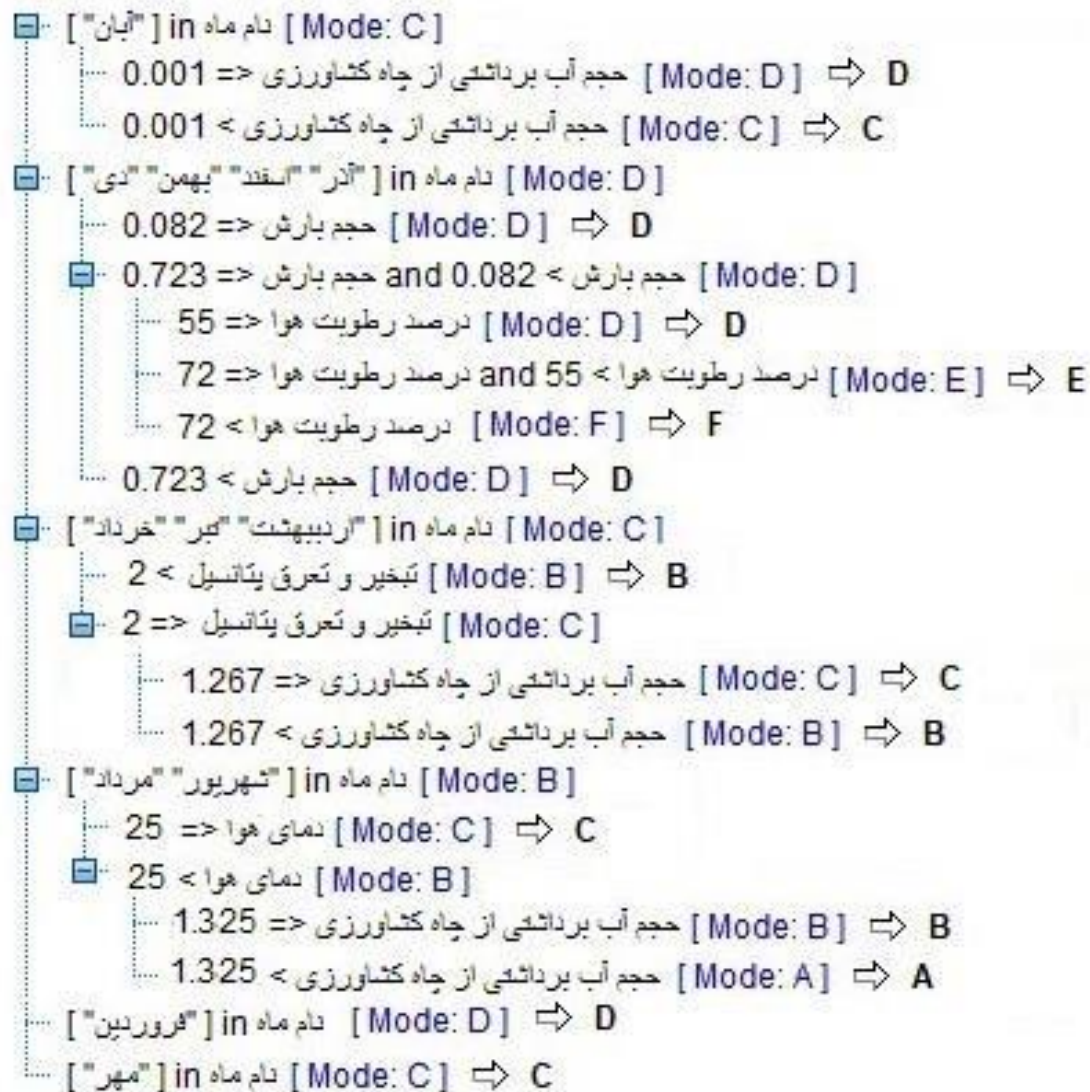
به منظور بررسی تأثیر مستقل هر یک از عوامل بر تغییرات عمق آبخوان به صورت جداگانه، نمودار تأثیر هر یک از عوامل، با تغییرات عمق آبخوان بررسی شد. شکل ۲، شامل هشت نمودار می‌باشد که در محور عمودی، هفت نمودار اول، شامل تعداد مدل‌ها و در محور افقی آن‌ها به ترتیب مقدار حجم آب ورودی به شبکه آبیاری (میلیون متر مکعب)، مقدار حجم بارندگی (میلیون متر مکعب)، مقدار حجم تقاضای آب کشاورزی (میلیون متر مکعب)، مقدار دمای هوا (سانتی‌گراد)، مقدار تبخیر و تعرق پتانسیل (میلی‌متر

بعد از این‌که شش مدل جدول ۲، به عنوان متغیر هدف و عوامل تأثیرگذار به عنوان متغیر ورودی به الگوریتم CHAID معرفی و درصد داده‌های آموزشی و آزمون مشخص شد، الگوریتم CHAID اجرا شد. شکل ۱، نمودار درختی حاصل از اجرای الگوریتم CHAID است.

با توجه به شکل ۱، نمودار درختی الگوریتم CHAID تقریباً از ماه آبان تا فروردین بالآمدگی آبخوان و از ماه اردیبهشت تا مهر افت آبخوان را پیش‌بینی کرده است. بیشترین مقدار بالآمدگی آبخوان (مدل F) در ماه‌های آذر، دی، بهمن و اسفند و در زمانی که مقدار حجم بارش بین ۰/۰۸ تا ۰/۷۲ میلیون متر مکعب و درصد رطوبت هوا بیشتر از ۷۲ درصد بوده، همچنین، بیشترین مقدار افت (مدل A) در ماه‌های شهریور و مرداد، در زمانی که دمای هوا بیشتر از ۲۵ درجه سانتی‌گراد و حجم آب برداشتی از چاه‌های کشاورزی بیشتر از ۱/۳۲ میلیون متر مکعب

نمودار، نام ماه (شمسی) و در محور افقی آن تعداد مدل‌ها را شامل می‌شود.

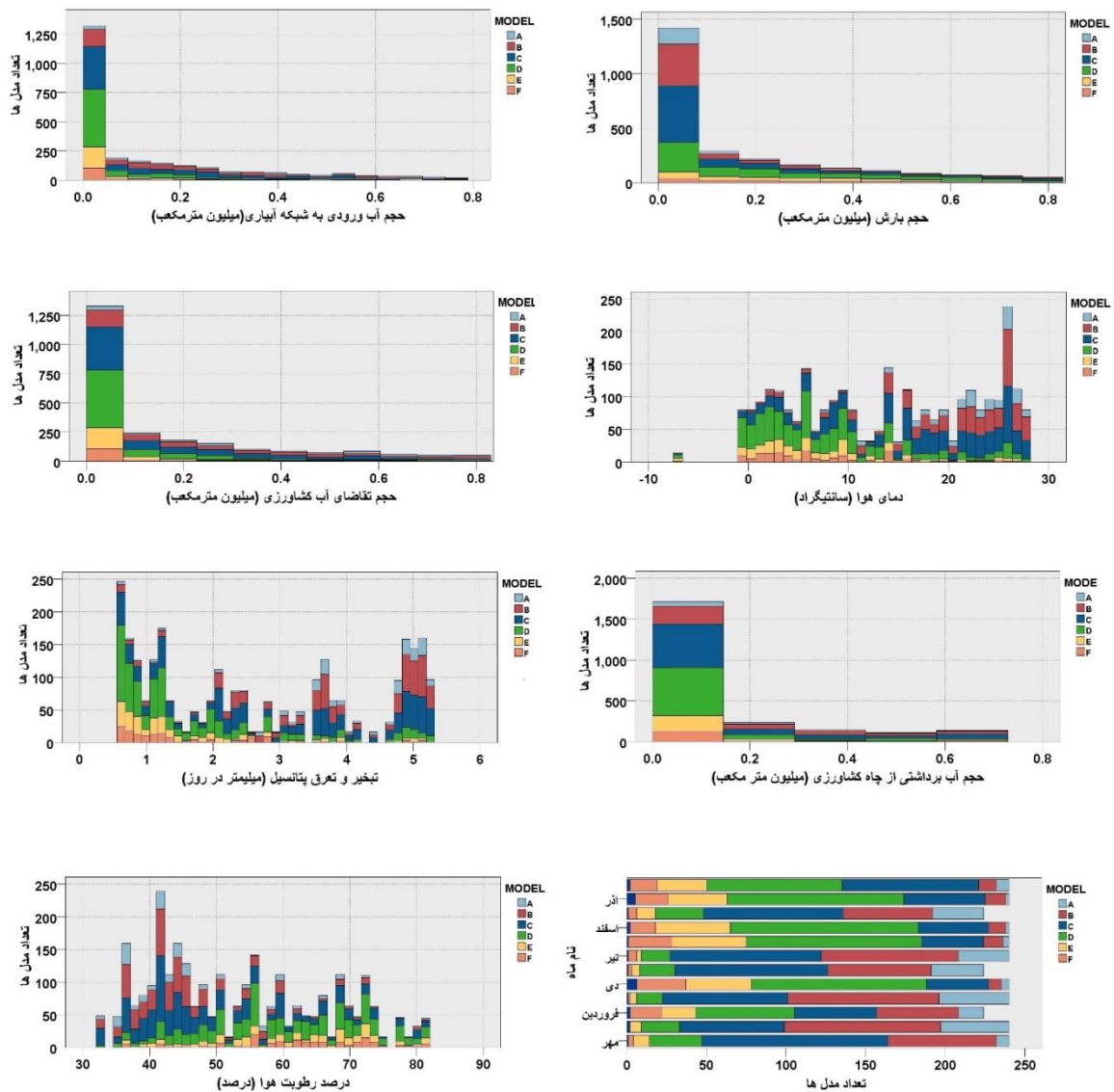
در روز، مقدار حجم آب برداشتی از چاه کشاورزی (متر مکعب)، مقدار رطوبت هوا (درصد) و در آخرین نمودار (دمای هوا) است. در محور عمودی آخرین



شکل ۱- نمودار درختی مربوط پیش‌بینی تغییرات آبخوان به‌وسیله الگوریتم CHAID

تقاضای آب کشاورزی باعث افزایش تعداد مدل‌های B و C (مدل‌های مربوط به افت) نسبت به سایر مدل‌ها شده است. لذا، حجم تقاضای آب کشاورزی رابطه مستقیمی با افت عمق آبخوان دارد. دمای هوا تا حدود ۱۰ درجه سانتی‌گراد باعث افزایش تعداد مدل‌های E و D (مدل‌های مربوط به بالآمدگی) و با افزایش مقدار دمای هوا از حدود ۱۰ درجه سانتی‌گراد باعث افزایش تعداد مدل‌های B و C (مدل‌های مربوط به افت) نسبت به سایر مدل‌ها شده است.

با توجه به شکل ۲، با افزایش مقدار حجم آب ورودی به شبکه آبیاری تعداد مدل‌های B و C (مدل‌های مربوط به افت) نسبت به سایر مدل‌ها افزایش پیدا کرده است. لذا، حجم آب ورودی به شبکه آبیاری رابطه مستقیم با مقدار افت عمق آبخوان دارد. با افزایش مقدار حجم بارش، تعداد مدل D (مدل مربوط به بالآمدگی) نسبت به سایر مدل‌ها افزایش پیدا کرده است. لذا، حجم بارش، رابطه مستقیم با مقدار بالآمدگی عمق آبخوان دارد. افزایش مقدار حجم



شکل ۲- بررسی تأثیر عوامل مختلف بر تغییرات افت آبخوان با استفاده از الگوریتم CHAID

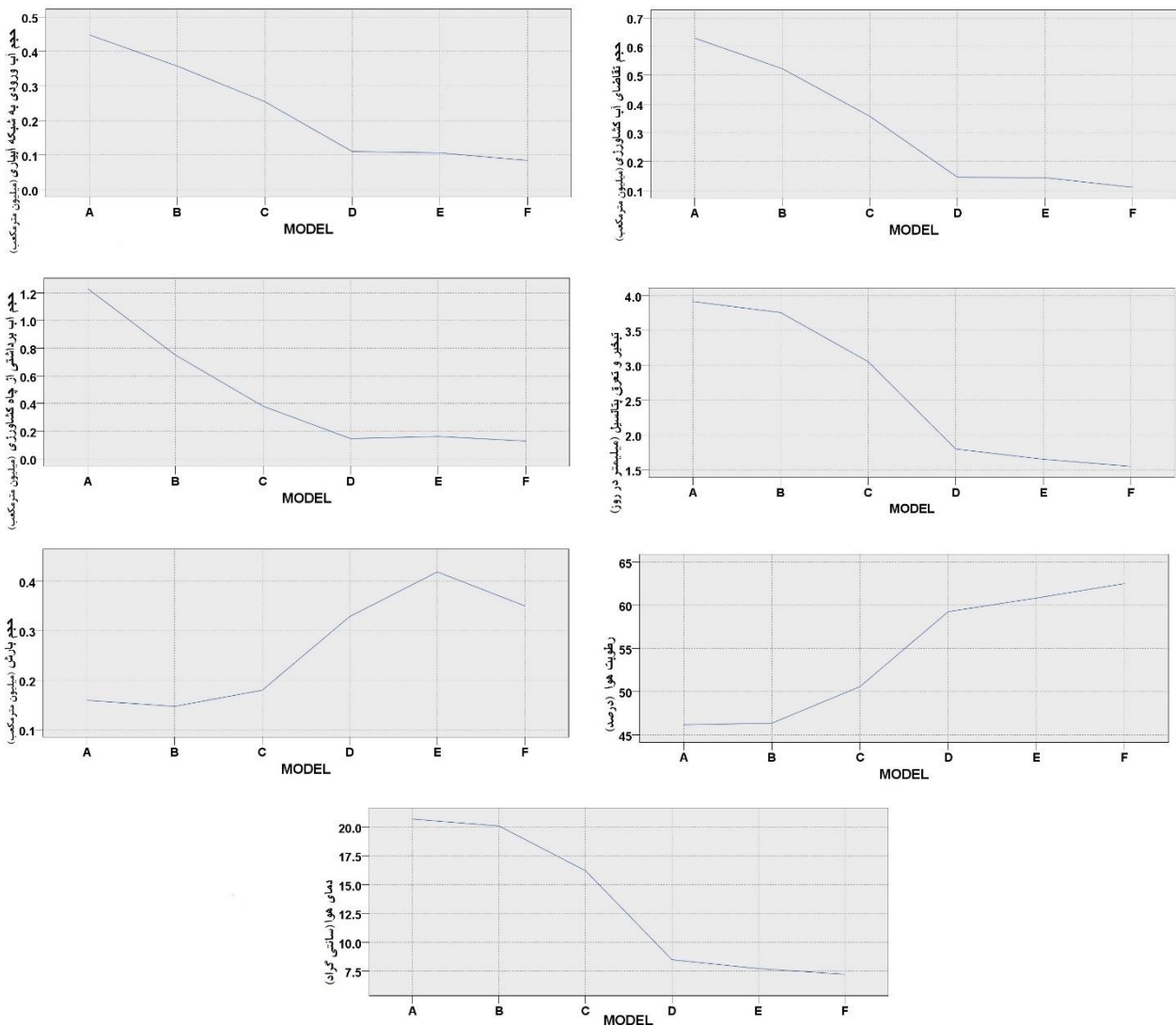
از ۵۵ درصد باعث افزایش تعداد D و E (مدل‌های مربوط به بالآمدگی عمق آبخوان) نسبت به سایر مدل‌ها شده است. همان‌طور که از رابطه بین درصد رطوبت هوا و مدل‌ها مشخص شد، رابطه عکس بین مقدار درصد رطوبت هوا و مقدار افت آبخوان وجود داشته، با افزایش مقدار درصد رطوبت هوا مقدار افت کمتر شده است. تعداد مدل A (افتی بیش از یک متر) بیشتر در شهریورماه و بعد از آن به ترتیب در ماه‌های مرداد و تیرماه بوده است. بیشترین تعداد مدل‌های بالآمدگی عمق آبخوان در ماه‌های دی و بهمن‌ماه مشاهده شده است. به‌طور کلی، بالآمدگی‌های عمق آبخوان بیشتر در ماه‌های آبان، آذر، دی، بهمن و

مقدار تبخیر و تعرق پتانسیل از صفر تا مقدار دو میلی‌متر در روز باعث افزایش تعداد مدل‌های D و E (مدل‌های مربوط به بالآمدگی عمق آبخوان) و از مقدار دو میلی‌متر بیشتر باعث افزایش تعداد مدل‌های B و C (مدل‌های مربوط به افت عمق آبخوان) نسبت به سایر مدل‌ها شده است. با افزایش مقدار حجم برداشتی از چاه‌های کشاورزی از مقدار ۰/۲ میلیون متر مکعب به بعد، باعث افزایش تعداد مدل‌های C و B (مدل‌های مربوط به افت) نسبت به مدل‌های دیگر شده است. مقدار درصد رطوبت هوا کمتر از ۵۵ درصد باعث افزایش تعداد مدل‌های B و C (مدل‌های مربوط به افت عمق آبخوان) و مقدار درصد رطوبت هوا بیشتر

۰/۲ میلیون متر مکعب بیشتر می‌شود، نمودار، مرتبط با مدل‌های مربوط به افت می‌شود. زمانی که دمای هوا تا مقدار حدود بیشتر از ۱۰ درجه سانتی‌گراد و نمودار تبخیر و تعرق پتانسیل از مقدار از دو میلی‌متر در روز بیشتر می‌شود، نمودار، مرتبط با مدل‌های مربوط به افت می‌شود. همچنین، زمانی که حجم بارش از مقدار بیشتر از ۰/۲ میلیون متر مکعب و درصد رطوبت هوا بیشتر از ۵۵ درصد می‌شود، نمودار، مرتبط با مدل‌های مربوط به بالآمدگی آبخوان می‌شود.

اسفند و در سایر ماه‌ها بیشتر افت عمق آبخوان مشاهده شده است.

با توجه به شکل ۳، از بین هفت عامل تأثیرگذار نمودار، پنج عامل تقریباً به صورت نزولی (رابطه مستقیم با مقدار افت آبخوان) و دو عامل به صورت صعودی (رابطه مستقیم با مقدار بالآمدگی آبخوان) هستند. زمانی که مقدار نمودار سه عامل حجم آب ورودی به شبکه آبیاری، حجم تقاضای آب کشاورزی و حجم آب برداشتی از چاه کشاورزی تقریباً از مقدار



شکل ۳- نمودارهای خطی مربوط به رابطه بین تغییرات عمق آبخوان و عوامل تأثیرگذار بر آن

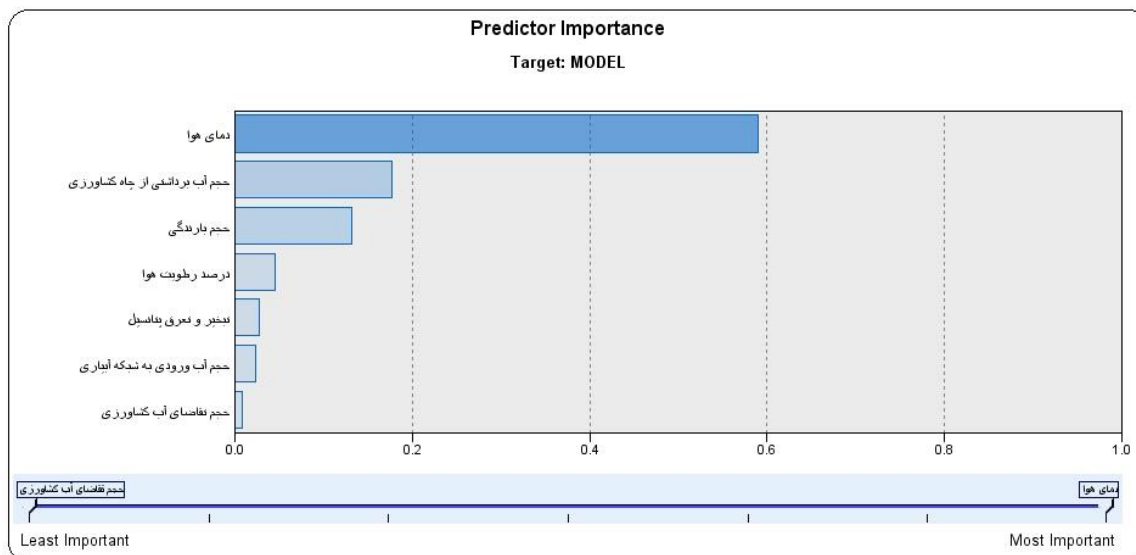
کننده^۱ الگوریتم CHAID استفاده شد. تأثیرگذارترین و با اهمیت‌ترین پارامتر در تعیین پیش‌بینی مقدار تغییرات عمق آبخوان مقدار دمای هوا می‌باشد. عوامل

تعیین مقدار اهمیت عوامل تأثیرگذار در برآورد تغییرات عمق آبخوان: برای تعیین مقدار اهمیت و تأثیر هر یک از پارامترهای ورودی در پیش‌بینی تغییرات عمق آبخوان از قسمت اهمیت پیش‌بینی

¹ Predictor importance

۴). بنابراین، برای برنامه‌ریزی و پیش‌بینی تغییرات عمق آبخوان مهمترین پارامتر دمای هوا است که لزوم توجه به دمای هوای برای پیش‌بینی تغییرات عمق آبخوان امری ضروری به‌نظر می‌رسد.

حجم آب برداشتی از چاه کشاورزی و تبخیر و تعرق پتانسیل، به‌ترتیب در رتبه‌های دوم و سوم در تعیین مقدار اهمیت پارامتر ورودی در پیش‌بینی تغییرات عمق آبخوان دشت تیروان و کرون قرار دارند (شکل



شکل ۴- تعیین تأثیر و اهمیت پارامترهای ورودی در پیش‌بینی مقدار تغییرات عمق آبخوان

درصد بوده، همچنین، بیشترین مقدار افت در ماه‌های شهریور و مرداد، در زمانی که دمای هوا بیش از ۲۵ درجه سانتی‌گراد و حجم آب برداشتی از چاه‌های کشاورزی بیشتر از ۱/۳۲ میلیون متر مکعب بوده است، به‌وسیله نمودار درختی الگوریتم CHAID پیش‌بینی شد. مهمترین عامل طبیعی تاثیرگذار بر تغییرات عمق آبخوان، دمای هوا است. به‌طوری‌که در زمان‌هایی که دمای هوا تقریباً از ۱۵ درجه سانتی‌گراد بیشتر می‌شود، بیشتر در آبخوان افت مشاهده می‌شود. به‌طورکلی، بالآمدگی‌های عمق آبخوان بیشتر در ماه‌های آبان، آذر، دی، بهمن و اسفند و در سایر ماه‌ها بیشتر افت عمق آبخوان مشاهده شد. مهمترین عامل انسانی تاثیرگذار بر تغییرات آبخوان، حجم آب برداشتی از چاه کشاورزی است. به‌طوری‌که با افزایش مقدار حجم برداشتی از چاه‌های کشاورزی از مقدار ۰/۲ میلیون متر مکعب، بیشتر در آبخوان افت مشاهده شده است.

نتیجه‌گیری

در سال‌های اخیر، به‌علت بارش‌های کمتر و در بعضی مواقع، استفاده از الگوی کشت نامناسب و پرمصرف باعث افزایش افت در آبخوان شده است. در این پژوهش، سعی شد تا با استفاده از الگوریتم درختی داده‌کاوی و همچنین عوامل مختلف طبیعی و انسانی تاثیرگذار بر تغییرات عمق آبخوان به پیش‌بینی و بررسی تاثیر عوامل مختلف بر تغییرات عمق آبخوان پرداخته شود.

در ابتدا، نتایج سه الگوریتم درختی CHAID، CART و MP5 به‌وسیله شاخص‌های آماری مورد ارزیابی قرار گرفتند. با توجه به نتایج شاخص‌های آماری، الگوریتم CHAID دارای عملکرد بهتری در رابطه با پیش‌بینی تغییرات عمق آبخوان دشت تیروان و کرون داشته است. با توجه به نتایج، بیشترین مقدار بالآمدگی آبخوان در ماه‌های آذر، دی، بهمن و اسفند و در زمانی که مقدار حجم بارش بین ۰/۰۸ تا ۰/۷۲ میلیون متر مکعب و درصد رطوبت هوا بیش از ۷۲

منابع مورد استفاده

1. Arvin, A.A., A.H. Hosseinian and M. Baharlou. 2016. The effect of climate fluctuations and water harvesting on changes in groundwater level in the Domain Plain. *Journal of Risks of the Natural Environment*, 5(7): 47-67 (in Persian).
2. Chattamvelli, R. 2011. *Data mining algorithms*. Alpha Science International, 274 pages.
3. Ebrahimi, M., H. Kazemi, M. Ehtashemi and T.D. Rockaway. 2016. Assessment of groundwater quantity and quality and saltwater intrusion in the Damghan Basin, Iran. *Chemie der Erde-Geochemistry*, 76(2): 227-241.
4. Ganji Khoramdel, N., M. Mohammadi and M.J. Mo'am. 2009. Optimization of observational wells network for estimating balance using Groundwater Double Fluctuation Method. *Water and Soil*, 22(2): 358-370 (in Persian).
5. Ghandehary, A., A. Gord Noshahri, A. Barati and K. Hasani. 2014. Local groundwater rise under metropolitans; opportunities and challenges. *Journal of Water and Sustainable Development*, 1(2): 75-82 (in Persian).
6. Gupta, G.K. 2011. *Introduction to data mining with case studies*. Prentice Hall of India, 514 pages.
7. Huajie, D., D. Zhengdong and D. Feifan. 2016. Classification of groundwater potential in Chaoyang area based on Quest algorithm. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 15 pages.
8. Kandahari, A., A. Gurd Noushari, R. Barati and Kh. Hosni. 2014. Groundwater level localization in metropolis; opportunities and challenges, case study: Mashhad. *Water and Sustainable Development*, 1(2): 12-25 (in Persian).
9. Mirhashimi, S.H., P. Haghghat, F. Mirzaei and M. Panahi. 2016. Evaluation of data mining algorithms for investigation and prediction of Qazvin Plain Aquifer situation. *Hydrogeology*, 2(2): 53-66 (in Persian).
10. Mirhashimi, S.H., P. Haghghat, F. Mirzaei and M. Panahi. 2016. Using CART algorithm in predicting groundwater table fluctuations inside and outside of an irrigation system, case study: irrigating area of Qazvin. *Iranian Journal of Soil and Water Research*, 94(2): 385-395 (in Persian).
11. Naghibi, S.A., H.R. Pourghasemi and B. Dixon. 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree and random forest machine learning models in Iran. *Environmental Monitoring and Assessment*, 188(1): 44-61.
12. Poormohammadi, S., M.T. Dastorani, S.A.M. Cheraghi, M.H. Mokhtari and M.H. Rahimian. 2011. Evaluation and estimation of water balance components in arid zone catchments using RS and GIS, case study: Manshad Catchment, Yazd Province. *Journal of Water and Wastewater*, 22(79): 99-108 (in Persian).
13. Sabuhi, M. and H. Tavana. 2007. Negative side effects of groundwater resources abuse, case study of Larestan City. *Agricultural Sciences and Industries*, 21(2): 67-77 (in Persian).
14. Shahid, Sh. and M.K. Hazarika. 2009. Groundwater drought in the northwestern district of Bangladesh. *Water Resources Management*, 24(10): 1989-2006.
15. Stumpp, C.J., A.J. Zurek, P. Wachniew Gargini, A. Gemitzi Filippini and M.S. Witzak. 2016. A decision tree tool supporting the assessment of groundwater vulnerability. *Environmental Earth Sciences*, 75(1057): 1-7.
16. Witten, I.H. and E. Frank. 2005. *Data Mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann, San Francisco, 558 pages.
17. Zhao, Y., Y. Li, L. Zhang and Q. Wang. 2016. Groundwater level prediction of landslide based on classification and regression tree. *Geodesy and Geodynamics*, 7(5): 348-355.