

استفاده از الگوریتم CART در پیش‌بینی نوسانات سطح آب زیرزمینی در داخل و خارج از محدوده شبکه

آبیاری (مطالعه موردی: منطقه تحت آبیاری شهرستان قزوین)

سیدحسین میرهاشمی^۱، پرویز حقیقت‌جو^{۲*}، فرهاد میرزایی^۳، مهدی پناهی^۴

۱. دانشجوی دکتری آبیاری و زهکشی، دانشگاه زابل

۲. استادیار گروه آب دانشکده آب و خاک، دانشگاه زابل

۳. دانشیار مهندسی گروه آبیاری و آبادانی پردیس کشاورزی و منابع طبیعی دانشگاه تهران

۴. استادیار گروه مهندسی آب دانشکده کشاورزی، دانشگاه زنجان

(تاریخ دریافت: ۱۳۹۶/۲/۱۷ - تاریخ بازنگری: ۱۳۹۶/۵/۱۰ - تاریخ تصویب: ۱۳۹۶/۷/۱)

چکیده:

با توجه به اهمیت نوسانات سطح آب زیرزمینی به‌عنوان یکی از عوامل مهم و مؤثر بر کشاورزی، در این تحقیق با استفاده از الگوریتم درختی CART از نرم‌افزار داده‌کاوی IBM SPSS Modeler 18.0 به کشف مدل و عوامل مؤثر بر پیش‌بینی نوسانات سطح آب زیرزمینی با توجه به اطلاعات مربوط به آبخوان محدوده تحت آبیاری شهرستان قزوین از سال ۱۳۸۰ تا ۱۳۹۴ پرداخته شد. متغیرهای ورودی به مدل شامل وزن تأثیر هر پیژومتر، نیاز خالص آبی، مصرف در هر پیژومتر، مقدار بارندگی، مقدار آب ورودی و خروجی از سد طالقان، نسبت سطح زیر کشت و مقدار آب ورودی به شبکه آبیاری و همچنین متغیر خروجی به‌عنوان تابع هدف شامل وضعیت نوسانات سطح آب زیرزمینی مورد استفاده قرار گرفت. قابلیت پیش‌بینی مدل به‌وسیله معیارهایی نظیر ضریب همبستگی و متوسط مطلق خطا مورد بررسی قرار گرفت. با توجه به نتایج بدست آمده مشخص شد که الگوریتم CART در پیش‌بینی نوسانات سطح آب زیرزمینی در داخل محدوده شبکه آبیاری نسبت به خارج از محدوده شبکه آبیاری دارای عملکرد بهتری بود. همچنین مشخص شد که مهمترین پارامتر مؤثر در نوسانات سطح آب زیرزمینی در داخل محدوده شبکه آبیاری مقدار آب ورودی به سد می‌باشد.

واژه‌های کلیدی: بارندگی، داده‌کاوی، سطح زیر کشت، سد طالقان

مقدمه

با توسعه کشاورزی و همچنین افزایش جمعیت که نیاز روز افزون به منابع آب را نیز به دنبال داشت، ناپایداری‌هایی در مدیریت سنتی منابع آب ایجاد شد. بخش عمده‌ای از عدم تعادل در منابع آب، ناشی از محدودیت طبیعی منابع آب و چرخه هیدرولوژی بوده و بخش دیگر ناشی از تصمیمات و فعالیت‌های انسانی مؤثر بر منابع آب است که زمینه‌ساز مشکلات زیادی در امر بهره‌برداری از منابع آب زیرزمینی شده است.

نتایج تحقیقات نشان می‌دهد که برداشت بی‌رویه از سفره آب زیرزمینی دشت سیرجان، پیامدهایی همچون تغییر کیفیت آب زیرزمینی، افزایش مصرف انرژی برای استحصال آب زیرزمینی، افزایش آسیب‌پذیری دشت نسبت به خشکسالی، نشست زمین، از بین رفتن اکوسیستم منطقه و خشک شدن باغات و غیره را نیز به دنبال داشته است (AbbasNezhad and

Shahidasht., 2013). سالانه بیش از ۱۹۹ میلیون مترمکعب

اضافه برداشت از منابع آب زیرزمینی این دشت صورت می‌گیرد (Barikani et al., 2011). همچنین با توجه به اینکه منابع آب-های زیرزمینی دشت قزوین یکی از اصلی‌ترین منابع آبی مورد استفاده در بخش کشاورزی استان قزوین است، بنابراین برای نگهداری و برداشت بهینه از آب‌های زیرزمینی مورد استفاده در بخش کشاورزی باید برنامه‌ریزی مناسب‌تری انجام گیرد که برای تحقق این امر علاوه بر پیش‌بینی رفتار آبخوان با توجه تصمیمات عرضه و تقاضا، باید مقدار تأثیر عوامل مختلف را بر نوسانات عمق آبخوان بررسی کرد. ایران از جمله کشورهای است که به دلیل کمبود منابع آب سطحی، بیشترین آب مصرفی در کشاورزی را از آب‌های زیرزمینی تأمین می‌کند؛ بنابراین کمبود منابع آب‌های زیرزمینی و تراز رو به کاهش و منفی آن نیز یکی از بحران‌های حال حاضر کشور محسوب می‌گردد (AbbasNezhad and Shahidasht., 2013).

در سال‌های گذشته نیز تحقیقات مختلفی در ارتباط با

بررسی تغییرات مکانی و تخمین سطح آب‌های زیرزمینی توسط

سنگ‌شناسی، ژئومورفولوژی، شیب، NDVI و NDWI، می‌توان به‌عنوان شاخص کافی برای پتانسیل آبهای زیرزمینی استفاده کرد. ۲- با این پیش دستاورد، الگوریتم درختی داده- کاوی QUEST به‌طور مؤثر در ساختن درخت تصمیم‌گیری برای طبقه‌بندی منطقه مورد مطالعه می‌تواند استفاده شود. ۳- نتیجه طبقه‌بندی دارای دقت قابل قبول و برابر با مقدار درجه صحت $0.92/8$ بوده است (Huajie et al., 2016). (Naghbi et al., 2016) 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، مقدار میزان صحت مدل‌ها به ترتیب 0.81 ، 0.78 و 0.71 بدست آمد؛ بنابراین، مدل BRT بهترین عملکرد را در نقشه‌برداری ذخایر چشمه‌های آب زیرزمینی داشته و بعد از آن به ترتیب مدل‌های CART و RF قرار داشتند. بر اساس نتایج، میزان صحت هر سه مدل بیش از ۷۰ درصد می‌باشد. بنابراین، هر سه مدل برای برنامه‌ریزان و مهندسان در مدیریت منابع آب و زمین و برنامه‌ریزی در منطقه مطالعاتی مناسب است. در مطالعات مدیریت منابع آب معمولاً با حجم بسیار زیادی از داده‌های مکانی-زمانی روبرو هستیم که استفاده از روش‌های تجربی و آماری در تبدیل چنین داده‌هایی به دانش کاربردی، عملاً غیرممکن است. داده‌کاوی یک فناوری توانمند در مدیریت و سازماندهی اطلاعات و همچنین استخراج دانش-های سودمند با حجم بالای داده می‌باشد. در این مقاله فرض بر آن شد که از این روش بتوان برای مدیریت بهتر آبخوان استفاده کرد. داده‌کاوی یک فرآیند شناخت الگوهای معتبر، جدید، ذاتاً مفید و قابل فهم از داده‌ها و همچنین جستجوی خودکار منابع داده‌ای بزرگ، برای یافتن الگوها و وابستگی‌هایی است که تحلیل‌های ساده و معمول آماری قادر به انجام آن نیستند. یکی از زمینه‌های استفاده از این ابزار برای تحلیل داده‌های وسیع و مدل‌سازی پیش‌گویانه با روش‌های محاسباتی جدید، هواشناسی و منابع آب زیرزمینی است. هدف از این تحقیق یافتن عوامل مؤثر بر وضعیت و پیش‌بینی نوسانات سطح آب زیرزمینی منطقه کشاورزی شهرستان قزوین می‌باشد. با شناسایی و کنترل عوامل مؤثر بر وضعیت‌های غیر نرمال در آبخوان، می‌توان از بوجود آمدن این وضعیت‌ها پیش‌گیری نمود.

مواد و روش‌ها

منطقه مطالعاتی

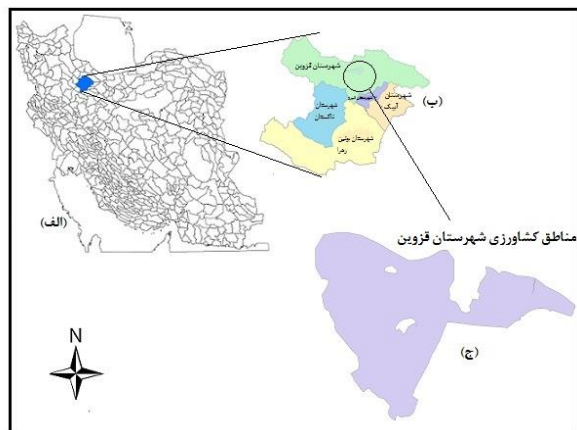
دشت قزوین با مساحتی حدود ۴۵۰۰۰۰ هکتار در محدوده‌ی

محققین صورت گرفته است (Jang et al., 2013). از شاخص استاندارد شده بارش (SPI) برای بررسی اثر خشکسالی و بارندگی روی سطح آب زیرزمینی در سه منطقه تحت آبیاری در حوضه ماری-دارلین استرالیا استفاده شد که نتایج آن‌ها نشان داد همبستگی خوبی بین شاخص SPI و تغییرات سطح آبخوان در منطقه برقرار است و به‌وسیله آن می‌توان الگوی خشکسالی‌های اصلی در استرالیا را تعیین کرد (Khan et al., 2008).

نتایج پیش‌بینی نشان داد که مدل درختی داده‌کاوی CART با همبستگی بین متغیرها و کاهش اطلاعات غیرمفید می‌تواند صحت و دقت پیش‌بینی را افزایش دهد و از مقایسه مدل CART با مدل PSO-SVR مشخص شد که مدل CART با دقت و توانایی پیش‌بینی بهتر می‌تواند برای پیش‌بینی کردن افت سطح آب زیرزمینی استفاده شود (Zhao et al., 2016). مدیران و تصمیم‌گیرندگان آب‌های زیرزمینی با استفاده از اطلاعات کلی و جزئی درخت تصمیم‌گیری داده‌کاوی می‌توانند در اجرای برنامه‌هایی برای حفاظت از منابع آب‌های زیرزمینی حمایت کنند (Stumpp et al., 2016). با استفاده از داده‌کاوی و داده‌های قدیمی آبخوان‌های دره Toluca مشخص شد که داده-کاوی این آبخوان‌ها منجر به تولید دانش جدید شده و با استفاده از اطلاعات الگوریتم داده‌کاوی مشخص می‌شود که مدیریت آب‌های زیرزمینی تحت تأثیر عوامل اجتماعی و اقتصادی از جمله تولید ناخالص ملی و ساختار جمعیتی است (Corona et al., 2016). از روش‌های طبقه‌بندی داده‌کاوی برای دسته‌بندی عوامل اصلاح کیفیت آب استفاده شد. تجزیه و تحلیل خوشه‌ای تمایل به دسته‌بندی بر اساس ویژگی‌های آلودگی کیفیت آب زیرزمینی دارد (Oorkavalan et al., 2016).

Kotsiantis et al. (2008) با استفاده از داده‌های ایستگاه هواشناسی آزمایشگاه انرژی و فیزیک محیط بخش فیزیک دانشگاه پاتریس در یونان و تکنیک‌های داده‌کاوی، دمای بیشینه، کمینه و متوسط را تخمین زده و نتیجه گرفتند که الگوریتم رگرسیونی داده‌کاوی قادر به پیش‌بینی دمای بیشینه، کمینه و متوسط با دقت رضایت‌بخشی است. هم‌چنین آن‌ها یک تکنیک هیبرید داده‌کاوی برای تخمین مقادیر روزانه دمای میانگین ارائه نمودند و به همان نتیجه دست یافتند.

Kolli and Seshadri (2013) با استفاده از قوانین انجمنی حاصل از داده‌کاوی نتیجه گرفتند که الگوریتم‌های داده‌کاوی برای نظارت و مدیریت آلودگی آب‌های زیرزمینی در منطقه تدپلی مجاور قسمت جنوبی رودخانه کریشنا می‌تواند مفید باشد. نتایج حاصل به این صورت خلاصه شد که ۱- بر اساس تحلیل عوامل مؤثر بر نفوذ آب زیرزمینی، از پنج موضوع لایه،



شکل ۱. الف- موقعیت دشت قزوین در بین دشتهای ایران ب- موقعیت شهرستان‌های قزوین در دشت قزوین ج- منطقه کشاورزی شهرستان قزوین

شهرستان قزوین مرکز استان قزوین است. رازمیان، معلم‌کلايه، سیردان، اقبالیه، کوهین و محمودآباد نمونه دیگر از مناطق شهری آن هستند. کل مساحت کشاورزی شهرستان قزوین حدود ۷۹۰۶۹ هکتار می‌باشد.

موقعیت سد طالقان در ۱۳۵ کیلومتری شمال غرب تهران و دارای طول جغرافیایی ۳۷' و ۵۰° الی ۱۰' و ۵۱° عرض جغرافیایی ۵' و ۳۶° الی ۲۵' و ۳۶° است و بر روی رودخانه طالقان در محل روستایی روشناپدر احداث شده است.

شبکه آبیاری دشت قزوین دارای یک کانال به طول ۹۴ کیلومتر بوده که روی خط تراز ۱۳۰۰ متر احداث شده است و تقریباً به موازات اتوبان تهران-نجان قرار دارد. ظرفیت کانال اصلی از ۳۰ تا ۱ مترکعب در ثانیه متغیر است. مشخصات کلی شبکه آبیاری قزوین در جدول (۱) ارائه شده است (Sotoodehnia and Sotoodehnia, 2016).

طول جغرافیایی ۴۹ درجه و ۲۵ دقیقه تا ۵۰ درجه و ۳۵ دقیقه شرقی و عرض جغرافیایی ۳۵ درجه و ۲۵ دقیقه تا ۳۶ درجه و ۲۵ دقیقه شمالی قرار دارد. این دشت از جلگه آبرفتی پهناوری شامل رسوبات جریان‌های سطحی کوه‌های اطراف تشکیل شده است (Mohammadi *et al.*, 2011). مجموع تغذیه آبخوان دشت قزوین ۱۲۵۹/۴۶ میلیون مترمکعب است. مجموع عوامل تخلیه آبخوان دشت قزوین ۱۴۵۸/۶۶ میلیون مترمکعب است. بر این اساس سهم تخلیه بخش کشاورزی حدود ۱۳۵۲/۹۲ میلیون مترمکعب است که حدود ۸۵۷/۳ میلیون هکتار آن سهم بخش زراعی است. با توجه به محدودیت منابع آب سطحی و فصلی بودن این منابع، بخش عمده آب آبیاری از منابع زیرزمینی استحصال می‌شود. در وضعیت موجود، برداشت باعث افت سالیانه ۱/۵ متر سطح سفره‌های زیرزمینی و فرو نشست زمین در این منطقه تا حدود ۲۵ سانتیمتر در سال شده است. با توجه به اهمیت دشت قزوین به‌عنوان یک منطقه مستعد کشاورزی از یک‌سو و وجود معضل افت شدید عمق سطح آب زیرزمینی در این منطقه از سوی دیگر، توجه به پایداری منابع آب زیرزمینی در تولید محصولات کشاورزی و انتخاب الگوی کشت این منطقه ضروری به نظر می‌رسد (Barikani *et al.*, 2011). استان قزوین دارای شش شهرستان آبیک، آوج، البرز، بوئین‌زهرا، تاکستان و قزوین می‌باشد. از این شش شهرستان بخش‌هایی از شهرستان‌های آبیک، البرز، بوئین‌زهرا، تاکستان و قزوین داخل دشت قزوین هستند و با توجه به متفاوت بودن رفتار آبخوان دشت قزوین، در بخش‌های مختلف دشت، در این مطالعه بخشی از آبخوان دشت قزوین که در محدوده منطقه کشاورزی شهرستان قزوین می‌باشد، مورد مطالعه قرار گرفت (شکل ۱).

جدول ۱. مشخصات و ظرفیت کانال‌های شبکه آبیاری قزوین

نوع کانال	شکل مقطع	نوع پوشش	ظرفیت طراحی M ³ /S	طول کانال Km	مشخصات مقطع	
					عمق (متر)	عرض کف (متر)
درجه یک	مستطیلی	بتنی	۳۰	۳	۳	۴
درجه دو	دورنقه‌ای	بتنی	۲/۹-۳۰	۹۱	۰/۹-۲/۵	۱/۲-۴
درجه سه	دورنقه‌ای	بتنی	۰/۶-۷/۴	۲۲۰	۰/۶۵-۱/۵	۰/۸-۲
درجه چهار	دورنقه‌ای	بتنی	۰/۱۷-۱	۳۲۰	۰/۶-۰/۸	۰/۶
			۰/۱۷-۰/۳۴	۵۶۰	۰/۴-۰/۵۵	۰/۳-۰/۶

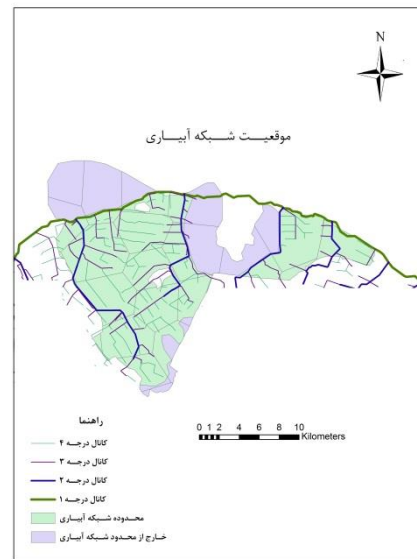
شبکه آبیاری و مناطق سبزرنگ با وسعت حدود ۲۲۰۷۹/۷۰ هکتار، داخل محدوده شبکه آبیاری قرار دارند.

شکل (۲) محل قرارگیری شبکه آبیاری در منطقه کشاورزی شهرستان قزوین را نشان می‌دهد. مناطق آبی‌رنگ در شکل (۲) با وسعت حدود ۹۶۰۴/۲۵ هکتار، خارج از محدوده

باشد. در تجزیه دو زیر گروه تعیین می‌شود و هر کدام در مرحله بعد به دو زیر گروه دیگر تقسیم می‌شوند و این روند ادامه می‌یابد تا زمانی که یکی از معیارهای توقف برآورده شود. درخت CART بازگشتی دو دویی است که گره‌های والدین را دقیقاً به دو گروه فرزند منشعب می‌کند و به‌طور بازگشتی منشعب کردن را تا زمانی که انشعاب دیگری نتواند ساخته شود ادامه می‌یابد. در تقسیم داده‌ها به‌صورت بازگشتی به زیر گروه‌ها، فقط دو زیر گروه را پشتیبانی می‌کند. این الگوریتم سریع است که هرس نمودن آن رو به عقب است (Chattamvelli, 2011).

به‌منظور اعتبارسنجی مدل درختی، داده‌ها به دو بخش داده‌های آموزش و آزمون تقسیم شدند. مدل درختی با استفاده از داده‌های آموزش ساخته شد و مدل ساخته شده با داده‌های آزمون مورد ارزیابی قرار گرفت. درصد نمونه‌هایی از داده‌های آزمون که ویژگی هدف آن‌ها توسط مدل، درست تشخیص داده شده بود دقت مدل را بیان می‌کند (Gupta, 2011). برای الگوریتم درخت تصمیم CART، ۷۰ درصد داده‌ها به‌صورت تصادفی به‌عنوان داده‌های آموزش انتخاب شدند و ۳۰ درصد مابقی به‌عنوان داده‌های آزمون مورد آزمایش قرار گرفتند. بعد از مرتب‌سازی داده‌ها، متغیرهای مستقل و وابسته به نرم‌افزار معرفی شدند. متغیرهای مستقل شامل داده‌های مربوط به مقدار بارندگی مؤثر (میلیون مترمکعب)، مقادیر آب ورودی و خروجی از سد، مقدار کل حجم آب تحویلی به شبکه آبیاری (میلیون مترمکعب)، مصرف آب زیرزمینی در محدوده هر پیزومتر (میلیون مترمکعب) و نام ماه و متغیرهای وابسته شامل داده‌های نوسانات سطح آب زیرزمینی بود که در جدول (۲) محدوده داده‌ها به‌صورت نام مدل گسسته‌سازی شد.

با توجه به اینکه الگوریتم CART یکی از الگوریتم‌های پیش‌بینی کننده و با نظارت از نرم‌افزار داده‌کاوی IBM SPSS Modeler 18 و همچنین یکی از الگوریتم‌های پرکاربرد در زمینه ایجاد درخت تصمیم بازگشتی دو دویی می‌باشد، در این تحقیق برای پیش‌بینی رفتار آبخوان شهرستان قزوین در دو محدوده داخل و خارج از شبکه آبیاری استفاده شد. شهرستان قزوین دارای ۲۴ پیزومتر اندازه‌گیری عمق آبخوان در محدوده دشت قزوین می‌باشد. با توجه به پراکندگی پیزومترهای اندازه‌گیری عمق آب زیرزمینی در قسمت‌های مختلف آبخوان و همچنین برای تشخیص تقریبی میزان بهره‌مندی و تأثیر هر پیزومتر، از پارامترهای مختلف همچون بارندگی، عرضه و مصرف آب و از روش تیسن برای تعیین محدوده مساحت تأثیر هر پیزومتر استفاده شد. برای تحقق این امر از نرم‌افزار ArcGIS 10 استفاده گردید. در این خصوص با استفاده از نرم‌افزار GIS ابتدا لایه‌های



شکل ۲. موقعیت شبکه آبیاری در منطقه کشاورزی شهرستان قزوین

نرم‌افزار داده‌کاوی IBM SPSS Modeler

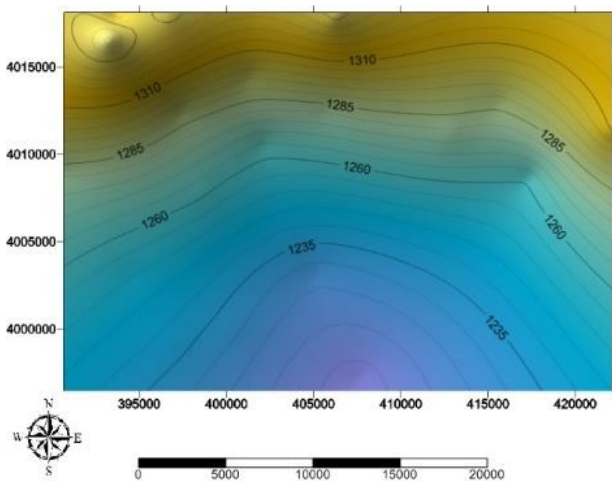
نرم‌افزار رایانه‌ای IBM SPSS برنامه‌ای است که برای تحلیل‌های آماری به کار می‌رود. نام SPSS مخفف عبارت Statistical Package for the Social Sciences به معنی بسته‌ی آماری برای علوم اجتماعی می‌باشد. بسته IBM SPSS Modeler نرم‌افزاری از شرکت SPSS است که در ابتدا با نام کلمنتاین^۱ ارائه می‌شد که در نسخه جدید به SPSS Modeler تغییر نام پیدا کرد. این نرم‌افزار یکی از بهترین ابزارهای داده‌کاوی است و برنامه‌ای حرفه‌ای برای انجام محاسبات پیچیده و آنالیزهای آماری به صورت سرور و کلاینت می‌باشد. نرم‌افزار SPSS IBM Modeler که از بهترین ابزارهای داده‌کاوی است دارای مزیت‌هایی نسبت به نرم‌افزار Clementine 12 است: که از آن جمله می‌توان بهینه کردن تکنیک‌ها برای مجموعه داده‌های بزرگ، افزایش مصورسازی الگوریتم‌های کلیدی مانند شبکه‌های عصبی و درخت تصمیم، محیط مصورسازی جدید برای الگوریتم‌های کلیدی و مدل‌های ترکیبی در ساخت نتایج برای فهم بهتر، به‌روزرسانی الگوریتم‌های شبکه عصبی و خطی، توانایی در خواندن و نوشتن داده از xml و بهینه کردن پردازش پایگاه داده بزرگ برای الگوریتم‌های کلیدی را اشاره کرد.

درخت تصمیم CART^۲

الگوریتم CART متغیرهای ورودی را برای یافتن بهترین تجزیه می‌آزماید تا شاخص ناخالصی حاصل از تجزیه، کمترین مقدار

1 - Clementine
2 - Classification and Regression Tree

به‌منظور مشخص نمودن تراز متوسط سطح آب زیرزمینی در محدوده منطقه کشاورزی شهرستان قزوین از آمار ماهانه مربوط به ۳۴ حلقه پیزومتر مشاهده‌ای موجود استفاده شد. خطوط هم‌تراز سطح ایستابی توسط نرم‌افزار SURFER 13 رسم شد. رسم خطوط هم‌تراز سطح ایستابی در شکل (۴) نشان داده شده است.



شکل ۴. خطوط هم‌تراز بار هیدرولیکی آبخوان شهرستان قزوین از سطح آزاد دریا

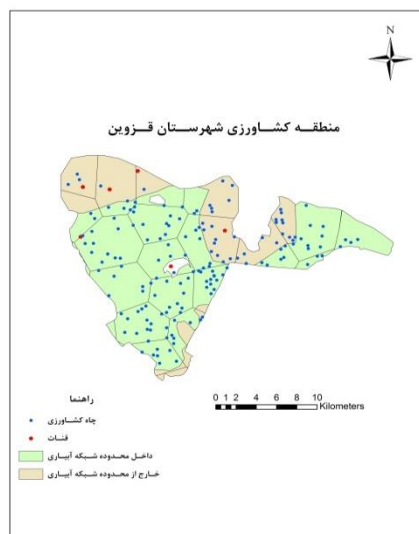
با توجه به اینکه مقادیر خروجی و ورودی به سد طالقان در محدوده زمانی مختلفی قرار دارد و مقادیر آن‌ها در ماه‌های مختلف با هم متفاوت می‌باشد، بدین منظور برای هر ماه از سال متوسط داده‌های همان ماه محاسبه و تغییرات آن نسبت به متوسط مربوطه همان ماه مقایسه گردید. برای انجام بهتر محاسبات و اجرای الگوریتم، مقادیر ورودی و خروجی از سد طالقان نرمال‌سازی شدند. به‌طوری که مقادیر مثبت و منفی شاخص ورودی به سد به ترتیب مشخص‌کننده مقدار بیشتر و کمتر از متوسط بلندمدت ورودی به سد و مقادیر مثبت و منفی شاخص خروجی از سد به ترتیب مشخص‌کننده مقادیر بیشتر و کمتر از متوسط بلندمدت خروجی از سد طالقان می‌باشد.

برای محاسبه مصرف آب زیرزمینی در محدوده هر پیزومتر، ابتدا متوسط سطح زیر کشت در محدوده تأثیر هر پیزومتر محاسبه شد. سپس، با ضرب مقدار مساحت مربوط به وزن تأثیر هر پیزومتر در نیاز خالص آبی تمام محصولات باغی و زراعی مورد کشت در ماه مورد نظر، مقدار مصرف آب کشاورزی برای محدوده تأثیر هر پیزومتر در ماه مورد نظر بر حسب میلیون مترمکعب محاسبه شد. مقادیر باران مؤثر بر حسب میلی‌متر از شرکت آب منطقه‌ای استان قزوین، به‌صورت ماهانه تهیه گردید. با توجه به اینکه میزان اثر باران برای هر پیزومتر

اطلاعاتی پیزومتر و مرزهای شهرستان قزوین در محدوده دشت قزوین تهیه گردید.

مجموع کل وزن تأثیر پیزومترهای اندازه‌گیری عمق آبخوان شهرستان قزوین که در منطقه کشاورزی قرار دارند، حدود ۳۱۶۸۳/۹۵ هکتار می‌باشد. از کل این مقدار، وزن تأثیر برخی از پیزومترها به‌طور کامل در داخل یا در خارج از شبکه آبیاری و برخی هم قسمتی در داخل و یا در خارج از شبکه آبیاری قرار دارند که با استفاده نرم‌افزارهای GIS 10 و Google Earth محدوده وزن تأثیر پیزومترهای که داخل و خارج شبکه آبیاری هستند، مشخص گردید. حدود ۲۲۰۷۹/۷۰ هکتار در داخل شبکه و حدود ۹۶۰۴/۲۵ هکتار در خارج از شبکه آبیاری قرار گرفتند (شکل ۳).

از ۱۷۴ حلقه چاه مجاز مورد استفاده در منطقه کشاورزی شهرستان قزوین، ۱۳۸ حلقه چاه از نوع بهره‌برداری کشاورزی، تعداد ۲۳ حلقه چاه از نوع بهره‌برداری تلفیقی و تعداد ۱۳ حلقه چاه از نوع بهره‌برداری چندمنظوره هستند. از ۱۷۴ حلقه چاه موجود، تعداد ۴۰ حلقه چاه در خارج از محدوده شبکه آبیاری و تعداد ۱۳۴ حلقه چاه در محدوده شبکه آبیاری قرار دارند. میانگین عمق چاه‌ها ۱۲۳ متر و بیشترین عمق چاه‌ها ۲۰۰ متر و همچنین متوسط آبدهی چاه‌ها ۳۶ لیتر در ثانیه و ماکزیمم آبدهی چاه‌ها ۹۰ لیتر در ثانیه می‌باشد. از ۶ مظهر قناتی که در منطقه کشاورزی شهرستان قزوین قرار دارد، سه مظهر قنات در داخل و سه مظهر قنات در خارج از محدوده شبکه آبیاری و با متوسط دبی ۳۸ لیتر در ثانیه و با حداکثر دبی ۶۰ لیتر در ثانیه می‌باشد (شکل ۳).



شکل ۳. موقعیت چاه‌ها و مظهر قنات‌ها در مناطق کشاورزی شهرستان قزوین

می‌شود که d_0 و dm به ترتیب مقادیر کمینه و بیشینه A با شرایط رابطه (۲) هستند. بنابراین در درخت تصمیم، برای متغیرهای پیوسته، گسسته سازی اتفاق می‌افتد.

$$\{[d_0, d_1], (d_1, d_2], \dots, (dm-1, dm]\} \quad (\text{رابطه ۱})$$

$$\forall i = 0, \dots, m-1: d_i < d_{i+1} \quad (\text{رابطه ۲})$$

بعد از انجام گسسته‌سازی برای کل داده‌ها، هر محدوده داده‌ها با اسم مدل از A تا J نام‌گذاری شد (جدول ۲). به طوری که مقدار افت سطح آب زیرزمینی به صورت نزولی از مدل A تا E کاهش و مقدار بالآمدگی به صورت صعودی از مدل F تا J افزایش می‌یابد.

جدول (۳) شامل خلاصه‌ای از مشخصات آماری به صورت ماهانه مربوط به داده‌های نوسانات سطح آب زیرزمینی منطقه کشاورزی شهرستان قزوین در داخل و خارج از محدوده شبکه آبیاری برای دوره ۱۵ ساله از ۱۳۸۰ تا ۱۳۹۴ می‌باشد.

متفاوت است، لذا میزان مقدار حجم باران مربوط به وزن تأثیر هر پیژومتر، بر حسب میلیون مترمکعب محاسبه شد. برای افزایش قابلیت استفاده نتایج برای مدیران و همچنین کاهش هر چه بیشتر اثرات این خطاها، گسسته‌سازی مقادیر افت آب زیرزمینی انجام شد. اگر خطای قرائت‌ها در حد کمتر از ۱ متر باشد، با استفاده از گسسته‌سازی، اثر این خطاها تا حد زیادی از بین می‌رود. همچنین برای مقادیر بسیار بزرگ و غیرمنطقی‌ای که در میزان افت وجود دارد و راهی برای اثبات یا رد آن وجود ندارد، با انجام فرآیند گسسته‌سازی، این خطا می‌تواند تا حد زیادی رفع گردد. در این روش، داده‌ها بر حسب قواعدی دسته‌بندی شده و دسته‌ای که تعداد داده‌های موجود در آن بسیار کم باشد، کنار گذاشته می‌شود. علت آن این است که این داده‌ها با دیگر داده‌ها تفاوت داشته و بنا به اشتباهاتی به وجود آمده‌اند؛ به عبارت دیگر گسسته‌سازی فرایندی است که صفات پیوسته را به گسسته تبدیل می‌کند به این معنی که مقادیر یک صفت پیوسته مثل A به زیر بازه‌های رابطه (۱) افراز

جدول ۲. تعیین مدل‌های مختلف برای محدوده داده‌ها

خارج از محدوده شبکه آبیاری		داخل محدوده شبکه آبیاری		MODEL		محدوده داده‌ها
درصد	میانگین (m)	تعداد	درصد	میانگین (m)	تعداد	
۰/۶۰	-۳/۱۹	۱۴	۰/۵۸	-۲/۹۱	۲۲	A $-2/2 > n$
۲/۷۴	-۱/۵۶	۶۴	۳/۱۰	-۱/۵۳	۱۱۷	B $-2/2 < n < -1/2$
۳/۹۳	-۰/۹۱	۹۲	۵/۲۶	-۰/۹۲	۱۹۹	C $-1/2 < n < -0/7$
۷/۸۲	-۰/۴۵	۱۸۳	۹/۳۷	-۰/۴۵	۳۵۴	D $-0/7 < n < -0/3$
۱۳/۲۵	-۰/۱۵	۳۱۰	۱۱/۱۱	-۰/۱۶	۴۲۰	E $-0/3 < n < 0$
۶/۸۴	۰/۱۳	۱۶۰	۵/۹۸	۰/۱۳	۲۲۶	F $0 < n < 0/3$
۲/۹۱	۰/۴۶	۶۸	۲/۹۶	۰/۴۴	۱۱۲	G $0/3 < n < 0/7$
۱/۴۱	۰/۸۹	۳۳	۱/۳۰	۰/۹	۴۹	H $0/7 < n < 1/2$
۰/۳۴	۱/۸۵	۸	۰/۱۹	۱/۸۹	۷	I $1/2 < n < 2/2$
۰/۱۷	۲/۸۷	۴	۰/۱۶	۲/۹۷	۶	J $2/2 < n$

جدول ۳. مشخصات آماری مربوط به نوسانات سطح آب زیرزمینی در داخل و خارج محدوده شبکه آبیاری در منطقه کشاورزی شهرستان قزوین

محدوده مطالعاتی	میانگین کل (m)	میانگین مقادیر افت (m)	میانگین مقادیر حداکثر افت حداکثر بالآمدگی (m)
داخل شبکه	-۰/۳۳	-۰/۵۹	۰/۴۰
تعداد داده‌ها	۱۵۱۲	۱۱۱۲	۴۰۰
خارج از شبکه	-۰/۲۶	-۰/۵۴	۰/۴۱
تعداد داده‌ها	۹۳۶	۶۶۳	۲۷۳

از خارج محدوده شبکه آبیاری می‌باشد. با توجه به جدول (۵) هنگامی که داده‌ها به صورت دسته‌بندی مورد استفاده قرار گرفتند مشخص شد که الگوریتم CART در پیش‌بینی نوسانات سطح آب زیرزمینی در داخل محدوده با درصد مقدار پیش‌بینی صحیح بیشتر و درصد پیش‌بینی غلط کمتر دارای عملکردی بهتری نسبت به خارج از محدوده شبکه آبیاری می‌باشد. مقادیر درصد پیش‌بینی درست هرچه بیشتر باشد نشان‌دهنده این است که طبقه‌بندی مورد استفاده، نمونه‌های بیشتری را در جای درست خود طبقه‌بندی کرده است. مقادیر کمتر درصد پیش‌بینی غلط تأیید کننده وقوع خطای کمتر در طبقه‌بندی نمونه‌ها است.

جدول ۵. نتایج پیش‌بینی الگوریتم CART در نوسانات سطح آب زیرزمینی با استفاده از مقادیر گسسته

محدوده آبخوان	نوع داده‌ها	درصد پیش‌بینی صحیح
داخل شبکه	داده‌های آموزشی	۹۴/۱۲
	داده‌های آزمایش	۸۱/۷۸
خارج از شبکه	داده‌های آموزشی	۷۲/۷۳
	داده‌های آزمایش	۵۲/۸۴

در نهایت با توجه به جدول‌های (۴) و (۵) مشخص شد که الگوریتم CART در پیش‌بینی تغییرات عمق آبخوان در داخل محدوده شبکه آبیاری با توجه به مقدار ضریب همبستگی و درصد پیش‌بینی صحیح بیشتر و همچنین با مقدار متوسط مطلق خطا و درصد پیش‌بینی غلط کمتر نسبت به پیش‌بینی نوسانات سطح آب زیرزمینی در خارج از محدوده شبکه آبیاری دارای عملکرد بهتری می‌باشد. لذا نتایج حاصل از اجرای الگوریتم CART در پیش‌بینی نوسانات سطح آب زیرزمینی در محدوده شبکه آبیاری مورد بررسی قرار گرفت.

شکل‌های (۶) تا (۱۰) شامل پنج نمودار می‌باشد که در محور عمودی تمامی نمودارها، مقدار مدل‌ها یا همان مقادیر گسسته شده عمق آبخوان و در محور افقی به ترتیب مقدار مصرف در پیزومتر (میلیون مترمکعب)، مقدار بارندگی (میلیون مترمکعب)، مقدار دبی کل شبکه (میلیون مترمکعب)، شاخص ورودی و خروجی از سد طالقان را شامل می‌شود. با توجه به شکل‌های زیر، مدل E که با رنگ زرد مشخص شده و با توجه به جدول (۲) دارای مقدار افت سطح آب زیرزمینی از مقدار ۰ تا ۰/۳ متر را شامل می‌شود، دارای بیشترین احتمال رخ دادن، از بین ده مدل می‌باشد. بعد از مدل E، مدل D با رنگ سبز بیشترین مقدار را بین دیگر مدل‌ها در تمامی نمودارها (به غیر از

به‌منظور بررسی عملکرد مدل CART از شاخص‌های آماری استاندارد بهره گرفته شد. در نرم‌افزار IMB SPSS Modeler از شاخص‌های آماری ضریب تبیین (R) و متوسط مطلق خطا (MAE) برای ارزیابی الگوریتم‌ها استفاده شد (روابط ۳ و ۴).

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_{obs} - y_{pred})^2}{\sum_{i=1}^n (y_{obs} - y_{mean})^2}} \quad (\text{رابطه ۳})$$

$$MAE = \frac{\sum_{i=1}^n |y_{obs} - y_{pred}|}{n} \quad (\text{رابطه ۴})$$

که در این روابط n: تعداد داده‌های مشاهده‌شده، y_{obs} : داده‌های مشاهده‌شده، y_{pred} : داده‌های پیش‌بینی‌شده، y_{mean} : میانگین داده‌ها می‌باشند.

نتایج و بحث

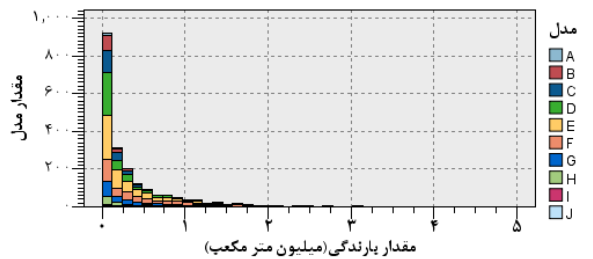
بعد از پیش‌بینی نوسانات سطح آب زیرزمینی در دو محدوده داخل و خارج از شبکه آبیاری و ساخت نمودار درختی و خطی توسط الگوریتم CART، برای بررسی و مقایسه آن در نوسانات سطح آب زیرزمینی در دو محدوده از دو شاخص آماری ضریب همبستگی و متوسط مطلق خطا استفاده شد. با توجه به اینکه تابع هدف الگوریتم CART برای ارزیابی به‌وسیله دو آزمون آماری باید پیوسته باشد، داده‌های واقعی نوسانات سطح آب زیرزمینی به‌عنوان تابع هدف به الگوریتم CART معرفی شد. جدول (۴) نشان می‌دهد که پیش‌بینی مدل CART در نوسانات سطح آب زیرزمینی در محدوده شبکه آبیاری با توجه به میزان همبستگی بیشتر و متوسط خطای کمتر نسبت به خارج از محدود شبکه آبیاری دارای عملکرد بهتری است. علت زیاد شدن مقدار خطا به علت رفتارهای متفاوت آبخوان در قسمت‌های مختلف منطقه مورد مطالعه، مانند اختلاف تعداد چاه‌های بهره‌برداری کشاورزی در نقاط مختلف و همچنین شرایط مختلف توپوگرافی می‌باشد.

جدول ۴. نتایج پیش‌بینی الگوریتم CART در نوسانات سطح آب زیرزمینی با استفاده از مقادیر واقعی

محدوده آبخوان	ضریب همبستگی (R)	متوسط مطلق خطا (MAE)	تعداد داده
داخل محدوده	٪۷۷	۰/۲۶	۴۵۴
خارج از محدوده	٪۶۷	۰/۲۹	۲۸۱

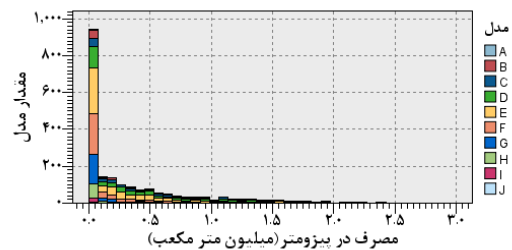
جدول (۵) نشان می‌دهد که میزان درصد پیش‌بینی صحیح الگوریتم CART در داخل محدوده شبکه آبیاری بیشتر

مقدار ورودی و خروجی از سد طالقان از مقدار متوسط بلند مدت آن کمتر می‌شود، احتمال رخ دادن مدل D (افت بین ۰/۳ تا ۰/۷ متر) بیشتر از مدل F (بالآمدگی بین ۰ تا ۰/۳ متر) و همچنین اگر مقدار ورودی و خروجی از سد طالقان از متوسط بلند مدت آن بیشتر شود، احتمال رخ دادن مدل F بیشتر از مدل D می‌شود.

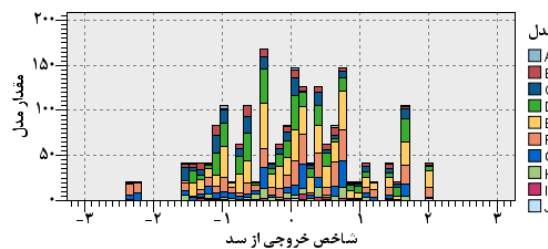


شکل ۶. رابطه بین مقدار مدل‌ها و مقدار بارندگی توسط الگوریتم CART

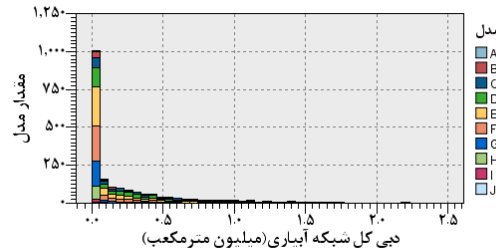
نمودار شکل (۶) را شامل می‌شود. لذا با توجه به احتمال بیشتر رخ دادن مدل‌های E و D در پیش‌بینی تغییرات سطح آب زیرزمینی نسبت به سایر مدل‌های دیگر مشخص می‌شود که احتمال رخ دادن افت سطح آب زیرزمینی بیشتر از احتمال بالآمدگی سطح آب زیرزمینی با توجه به شرایط موجود می‌باشد. در شکل‌های (۹) و (۱۰) مشخص می‌شود هنگامی که



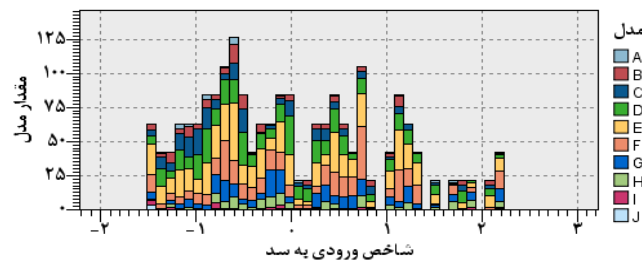
شکل ۵. رابطه بین مدل‌ها و مقدار مصرف در پیژومتر توسط الگوریتم CART



شکل ۸. رابطه بین مدل‌ها و شاخص ورودی به سد توسط الگوریتم CART



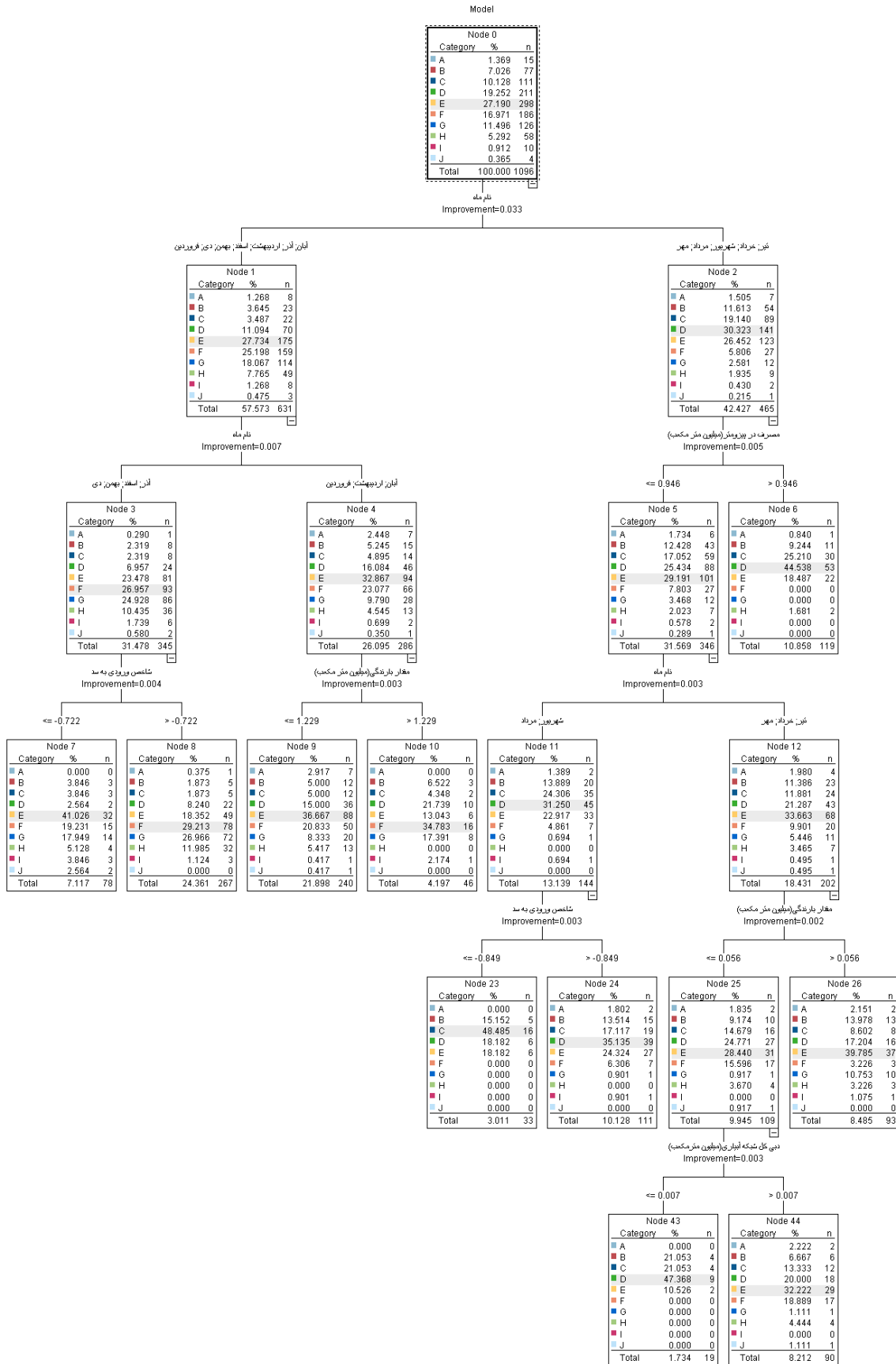
شکل ۷. رابطه بین مقدار مدل‌ها و دبی کل شبکه آبیاری توسط الگوریتم CART



شکل ۹. رابطه بین مقدار مدل‌ها و شاخص خروجی از سد توسط الگوریتم CART

ها از A تا J مربوط به وضعیت نوسانات سطح آب زیرزمینی می‌باشد. نام مدل‌ها از A تا E مشخص‌کننده میزان افت سطح آب زیرزمینی از مقدار زیاد به مقدار کم و از F تا J نشان‌دهنده میزان بالآمدگی سطح آب زیرزمینی از مقدار کم به مقدار زیاد می‌باشد. اولین برگ به دو شاخه تفکیک شده و هر شاخه با توجه به ماه‌های مختلف از هم تفکیک شده و به برگ دیگری متصل می‌شود و به همین ترتیب تقسیم شاخه و تولید برگ ادامه پیدا می‌کند تا به برگ انتهایی برسد که همان پیش‌بینی نهایی (نام مدل) است.

با توجه به این‌که نمودار درختی شامل شاخه‌ها و گره‌ها است. شاخه نشان‌دهنده نتیجه یک آزمون برای طبقه‌بندی یک الگو (بر اساس یک آزمون) با استفاده از یکی از ویژگی‌ها است. گره یا برگ در پایان نشان‌دهنده نهایی انتخاب کلاس برای یک الگو است. شکل (۱۰) نمودار درختی به صورت شاخه و برگ حاصل از اجرای الگوریتم CART برای پیش‌بینی نوسانات سطح آب زیرزمینی در محدوده شبکه آبیاری را نشان می‌دهد. به غیر از برگ اول که شرایط کل داده‌ها را نشان می‌دهد، بقیه برگ‌ها درصد احتمال و تعداد وقوع هر مدل را با توجه به شرایط شاخه قبل از خودش مشخص می‌کند. با توجه به جدول (۲) نام مدل-



شکل ۱۰. نمودار درختی CART در پیش‌بینی نوسانات سطح آب زیرزمینی در محدوده شبکه آبیاری

زیرزمینی در محدوده شبکه آبیاری منطقه کشاورزی شهرستان قزوین با استفاده از الگوریتم CART بدست آمد.

با توجه به اطلاعات شکل‌های از (۵) تا (۱۰)، اطلاعات جدول (۶) برای وضعیت محتمل در نوسانات سطح آب زیرزمینی در ماه‌های مختلف برای پیش‌بینی نوسانات سطح آب

جدول ۶. وضعیت محتمل نوسانات سطح آب زیرزمینی در ماه‌های مختلف سال در محدوده شبکه آبیاری

ماه	وضعیت محتمل نوسانات سطح آب زیرزمینی	توضیحات
فروردین، اردیبهشت، آبان	افت (با احتمال ۶۱/۵۱ درصد)	افت بیشتر از ۲ متر با احتمال ۲/۴۴ درصد افت در حدود ۱ متر با احتمال ۴/۸۹ درصد افت کمتر از ۰/۳ متر با احتمال ۳۱/۵ درصد
اسفند، آذر، بهمن، دی	بالا آمدگی (با احتمال ۶۴/۶۱ درصد)	بالا آمدگی بیشتر از ۲ متر با احتمال ۰/۵ درصد بالا آمدگی در حدود ۱ متر با احتمال ۱۰/۴۳ درصد بالا آمدگی کمتر از ۰/۳ متر با احتمال ۲۶/۹۵ درصد
تیر، خرداد، مهر	افت (با احتمال ۸۰/۱۸ درصد)	افت بیشتر از ۲ متر با احتمال ۱/۹۸ درصد افت در حدود ۱ متر با احتمال ۱۱/۸۸ درصد افت کمتر از ۰/۳ متر با احتمال ۳۳/۶۶ درصد
شهریور، مرداد	افت (با احتمال ۹۳/۷۲ درصد)	افت بیشتر از ۲/۲ متر با احتمال ۱/۳۸ درصد افت در حدود ۱ متر با احتمال ۲۴/۳ درصد افت کمتر از ۰/۳ متر با احتمال ۳۱/۲۵ درصد

نتیجه‌گیری

مهمترین و ساده‌ترین راه‌های مدیریت مصرف در بخش کشاورزی، مدیریت عرضه مربوط به مصرف بهینه و پایدار می‌باشد. از نتایج این تحقیق همچنین مشخص شد که احتمال افت بین ۰/۷ و ۱/۲ متر بیشتر در ماه‌های شهریور و مرداد و در صورتی که مقدار شاخص ورودی به سد کمتر از ۰/۸۴- باشد رخ می‌دهد. لذا در شرایط خشکسالی که هم میزان عرضه آب و هم بارندگی کاهش می‌یابد، اگر کاهش قابل توجهی در بخش مصرف که با الگوی کشت محصولات کم‌مصرف رخ می‌دهد، انجام گیرد می‌توان احتمال وقوع افت زیاد سطح آب زیرزمینی را تا مقدار قابل توجهی کاهش داد. همچنین با توجه به نتایج نمودار درختی مشخص شد که مهم‌ترین پارامتر مؤثر در نوسانات سطح آب زیرزمینی در داخل محدوده شبکه آبیاری مقدار حجم آب ورودی به سد می‌باشد.

از نتایج بدست آمده توسط الگوریتم CART در پیش‌بینی نوسانات سطح آب زیرزمینی در محدوده شبکه آبیاری مشخص شد زمانی که خشکسالی کم تا متوسط داشته باشیم و همچنین محدودیت و کمبود منابع آب سطحی در تخصیص به شبکه‌های آبیاری نیز داشته باشیم، لازم است که توجهی جدی به الگوی کشت و مدیریت مصرف آب کشاورزی و آگاه‌رسانی مصرف‌کنندگان صورت گیرد. اگر در شرایطی که عرضه به صورت محدود و ناکافی انجام گیرد و کشت و مصرف آب در بخش کشاورزی بسیار بیشتر از میزان عرضه پایدار اتفاق بیفتد، این اختلاف، افت نسبتاً قابل توجهی در سطح آب زیرزمینی ایجاد می‌نماید. همچنین با افزایش و کاهش مصرف آب کشاورزی اثرات آن در مقدار دبی شبکه آبیاری مشاهده می‌شود. لذا یکی از

REFERENCES

- AbbasNezhad, A and Shahidasht, A. (2013). Vulnerability of Sirjan Plain Due to Aquifer Over Abstraction. *Journal of Geography and Urban Planning – Regional* 3(7), 85-96. (In Farsi)
- Barikani, A., Ahmadian, M. and Khalilian, S. (2011). Optimal Sustainable Use of groundwater resources in agriculture: Case Study Subsector in Qazvin Plain. *Journal of Agricultural Economics and Development* 25(2), 253-262. (In Farsi)
- Chattamvelli, R. (2011). *Data Mining Algorithms*, (1th ed.). Oxford: Alpha Science, PP: 274-290.
- Corona, O. L., Fuentes, O. E., Casique, E.M., Longoria, P.P. and Moran, T. G. (2016). Data Mining of Historic Hydrogeological and Socioeconomic Data Bases of the Toluca Valley, Mexico. *Journal of Water Resource and*

- Protection, 8(4), 522-533.
- Gupta, G.K. (2011) Introduction to Data Mining with Case Studies, (2th ed.), Prentice hall of India.
- Huajie, D., Zhengdong, D. and Feifan, Deng. (2016). Classification of groundwater potential in chaoyang area based on quest algorithm. College of Defense Engineering, advancing the understanding of our living planet, Beijing, China, 19: 890-893.
- Jang, C.S., Chen., S.K. and Ku, Y.M. (2013). Applying indicator-based geostatistical approaches to determine potential zones of groundwater recharge based on borehole data. *Catena*, 101, 178-187.
- Khan, S., Gabriel, H. F. and Rana, T. (2008). Standard precipitation index to track drought and assess impact of rainfall on watertables in Irrigation and Drainage Systems, 22, 159-177.
- Kolli, K. and R, Seshadri. (2013). Ground Water Quality Assessment using Data Mining Techniques. *International Journal of Computer Applications*, 76(15): 39-45.
- Kotsiantis, S., Kostoulas, A., Lykoudis S., Argirio, A. and Menagias, K. (2008). Using data mining techniques for estimating minimum, maximum and average daily temperature value. *IJMPES*. 1(2): 117-121.
- Mohammadi, M., Mohammadi Ghaleney, M. , Ebrahimi, K. (2011). Spatial and Temporal Variations of Groundwater Quality of Qazvin plain, *Water Research Iran*, 5(8), 41-51. (In Farsi)
- Naghbi, S. A., Pourghasemi, H. R. and Dixon, B. (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
- Oorkavalan, G., Chidambaram, S. M., Mariappan, V., Kandaswamy, G. and Natarajan. S. (2016). Cluster Analysis to Assess Groundwater Quality in Erode District, Tamil Nadu, India. *Circuits and Systems, Computer Science & Communications*, 7(6): 877-890
- Sotoodehnia, A. Sotoodehnia, S. (2016). The Assessment of the role of Taleghan reservoir dam construction on sedimentation in Ghazvin plain irrigation network, 14-15 Dec., Faculty of Engineering and Technology, Imam Khomeini International University, Ghazvin, Iran, pp.1-9. (In Farsi)
- Stumpp, C., J. Zurek, A.J., Wachniew, P., Gargini, Gemitzi, A., Filippini, M. and Witczak, S. (2016). A decision tree tool supporting the assessment of groundwater vulnerability. *Environmental Earth Sciences*, 75(1057), 1-7.
- Zhao, Y., Li, Lifen., Y and Zhang, L. and Wang, Q. (2016). Groundwater level prediction of landslide based on classification and regression tree. *geodesy and geodynamics*, 17(5), 348-355.