



Evaluation of Data Mining Algorithms in Studying and Predicting the Qazvin Plain Aquifer Conditions

Seyed Hasan Mirhashemi¹, Parviz Haghghatjou^{2*}, Farhad Mirzaei³, and Mehdi Panahi⁴

Abstract

Due to population growth and agricultural development, predicting aquifer conditions and also determining factors effecting its performance are very important. In this research study, three algorithms of data mining package, i.e. IBM SPSS Modeler were applied for detecting models effective on predicting an aquifer conditions. Furthermore, human and natural factors were considered as input variables effective on predicting Qazvin plain aquifer conditions during 2001 to 2015. The accuracy of CHAID, CART, and CVM algorithms were 0.67, 0.75, and 0.74, respectively. The values obtained for different indices including TPR, TNR, ACC, FPR, FNR, FM, GM, and ER for CART decision tree model showed that the performance of the mentioned algorithm was better comparing to other algorithms. In addition, the results based on CART decision tree model show that the most important factors affecting water table fluctuations were air temperature, evapotranspiration and agricultural water demand. Thus, the developed model lead to predict water table depth in an aquifer and optimum management of the water resources.

Keywords: Agricultural Water Demand, Air Temperature, Evapotranspiration, Decision Tree Model, Management.

Received: 2016/06/25
Accepted: 2017/09/12

ارزیابی الگوریتم‌های داده‌کاوی در بررسی و پیش‌بینی وضعیت آبخوان دشت قزوین

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

چکیده

با توجه به افزایش جمعیت و توسعه کشاورزی پیش‌بینی وضعیت آبخوان و همچنین شناسایی عوامل تأثیرگذار بر آن از اهمیت زیادی برخوردار می‌شود. در این تحقیق با استفاده از سه الگوریتم نرم‌افزار داده‌کاوی SPSS Modeler IMB به کشف مدل‌ها و الگوهای مؤثر بر پیش‌بینی وضعیت آبخوان پرداخته شد. همچنین از عوامل انسانی و طبیعی تأثیرگذار بر پیش‌بینی مقدار تغییرات آبخوان دشت قزوین از سال ۱۳۸۰ تا ۱۳۹۴ به‌عنوان متغیرهای ورودی استفاده شد. مقدار صحت بین الگوریتم‌های اجرا شده CHAID، CART و CVM به ترتیب برابر با مقدار ۰/۶۷، ۰/۷۵ و ۰/۷۴ به دست آمد. مقادیر به دست آمده برای شاخص‌های حساسیت، شفافیت، صحت، نرخ مثبت غلط، نرخ منفی غلط، شاخص F، میانگین هندسی و نرخ خطا برای مدل درخت تصمیم CART نشان دهنده عملکرد بهتر این الگوریتم نسبت به سایر الگوریتم‌ها بود. قوانین استخراج شده توسط درخت تصمیم CART نشان داد که بیشترین تأثیر روی تغییرات عمق آبخوان به ترتیب مقدار دما و تبخیر و تعرق و تقاضای آب کشاورزی می‌باشد. از این رو مدل ارائه شده به پیش‌بینی مقدار نوسانات عمق آبخوان و نیز مدیریت صحیح آبخوان که منجر به کاهش اثرات ناشی از افت می‌شود، کمک می‌کند.

واژه‌های کلیدی: تبخیر و تعرق، تقاضای آب کشاورزی، درخت تصمیم، دما، مدیریت.

تاریخ دریافت مقاله: ۱۳۹۵/۰۴/۰۵

تاریخ پذیرش مقاله: ۱۳۹۶/۰۶/۲۱

1-Ph.D. candidate, Dept. of Water Engineering, University of Zabol.
2-Assistant Professor, Dept. of Water Engineering, University of Zabol.
3-Associate Professor, Dept. of Irrigation and Drainage, University of Tehran.
4-Assistant Professor, Dept. of Water Engineering, University of Zanjan.
*- Corresponding Author: hassan.mirhashemi@yahoo.com

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

* _ نویسنده مسئول

مقدمه

با توسعه کشاورزی و همچنین افزایش جمعیت که نیاز روزافزون به منابع آب را نیز به دنبال داشت، ناپایداری‌هایی را در مدیریت سنتی منابع آب ایجاد نمود. بخش عمده‌ای از عدم تعادل در منابع آب، ناشی از محدودیت طبیعی منابع آب و چرخه هیدرولوژی بوده و بخش دیگر تصمیمات و فعالیت‌های انسانی مؤثر بر روی منابع آب است که زمینه‌ساز مشکلات زیادی در امر بهره‌برداری از منابع آب زیرزمینی گشته است. همچنین مدیران در این بخش 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 می‌باشد. در سال‌های گذشته نیز تحقیقات مختلفی در ارتباط با بررسی تغییرات مکانی و تخمین سطح آب‌های زیرزمینی توسط محققین صورت گرفته است (جانگ و همکاران، ۲۰۱۳). از شاخص استاندارد شده بارش (SPI) برای بررسی اثر خشک‌سالی و بارندگی روی سطح آب زیرزمینی در سه منطقه تحت آبیاری در حوزه ماری-دارلین استرالیا استفاده شد که نتایج آن‌ها نشان داد که همبستگی خوبی بین شاخص SPI و نوسانات سطح آب زیرزمینی در منطقه برقرار است و به‌وسیله آن می‌توان الگوی خشک‌سالی‌های اصلی در استرالیا را تعیین کرد (خان و همکاران، ۲۰۰۸). نتایج پیش‌بینی نشان داد که مدل درختی داده‌کاوی CART با همبستگی بین متغیرها و کاهش اطلاعات غیرمفید می‌تواند صحت دقت پیش‌بینی را افزایش دهد و نتیجه مقایسه مدل CART با مدل PSO-SVR، مشخص شد که مدل CART با دقت بهتر و توانایی پیش‌بینی بهتر، می‌تواند برای پیش‌بینی کردن افت سطح آب زیرزمینی استفاده شود (ژائو و همکاران، ۲۰۱۶). همچنین نتایج 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 (استامپ و همکاران، ۲۰۱۶). با استفاده از داده‌کاوی و داده‌های آبخوان‌های قدیمی دره Toluca مشخص شد که با داده‌کاوی این آبخوان‌ها، قادر به تولید دانش جدید می‌باشند و همچنین با استفاده از اطلاعات الگوریتم داده‌کاوی مشخص می‌شود که مدیریت آب‌های زیرزمینی تحت تأثیر عوامل

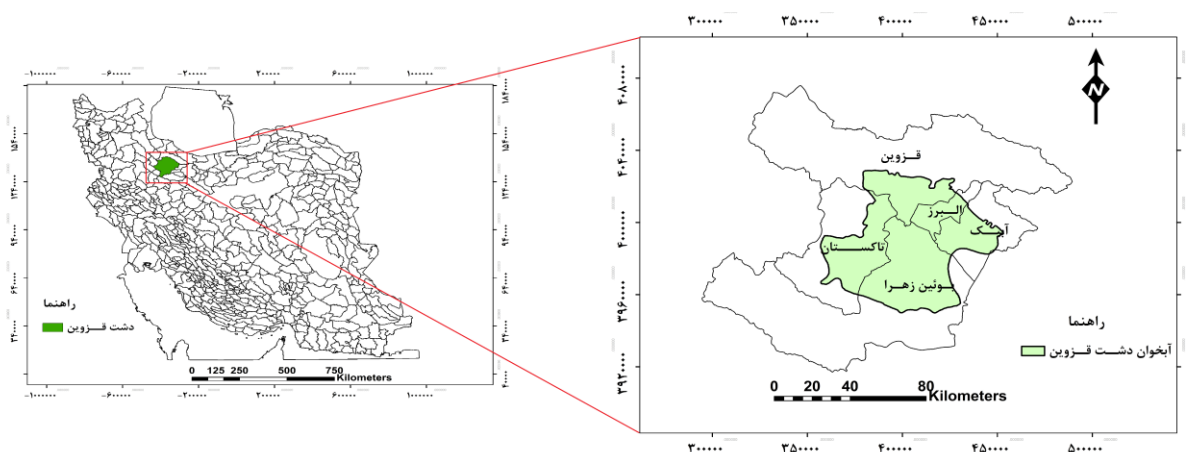
اجتماعی و اقتصادی از جمله تولید ناخالص ملی و ساختار جمعیتی است (کرنا و همکاران، ۲۰۱۶). روش‌های طبقه‌بندی داده‌کاوی برای دسته‌بندی عوامل اصلاح کیفیت آب استفاده کردند. تجزیه و تحلیل خوشه‌ای تمایل به دسته‌بندی بر اساس ویژگی‌های آلودگی کیفیت آب زیرزمینی دارد (ارکولن و همکاران، ۲۰۱۶). نتایج بدست آمده به سه صورت خلاصه شد، ۱- بر اساس تحلیل عوامل نفوذ زیرزمینی، پنج موضوع لایه، سنگ‌شناسی، ژئومورفولوژی، شیب، NDVI و NDWI، می‌تواند به‌عنوان شاخص کافی برای پتانسیل آب‌های زیرزمینی استفاده کرد. ۲- با این پیش‌دستاوردها، الگوریتم درختی داده‌کاوی QUEST به‌طور مؤثر در ساختن درخت تصمیم‌گیری برای طبقه‌بندی منطقه مورد مطالعه می‌تواند استفاده شود. ۳- نتیجه طبقه‌بندی به دقت قابل قبولی با مقدار درجه صحت $92/8\%$ بوده است (هوجیه و همکاران، ۲۰۱۶). ارزیابی و مقایسه چهار الگوریتم معروف با نظارت داده‌کاوی SVM، NB، KNN و CBA به‌منظور صحت در طبقه‌بندی، دقت و مقدار FI در رابطه با مجموعه داده‌های آب‌های زیرزمینی کشور اردن مورد بررسی قرار گرفت. نتایج آزمون نشان داد که الگوریتم SVM از سه الگوریتم مورد استفاده دیگر در رابطه با روش‌های اندازه‌گیری بهتر عمل می‌کند. نتایج همچنین نشان داد که می‌توان حداکثر استفاده را از الگوریتم‌های داده‌کاوی در پیش‌بینی مناطق آب‌های زیرزمینی می‌تواند داشت (نوروزی و همکاران، ۲۰۱۶). چندین مطالعه در زمینه، استفاده از مجتمع بر سیستم یکپارچه اطلاعات جغرافیایی (GIS)، سنجش از دور (RS) و روش‌های زمین‌آمار برای نقشه‌برداری آب‌های زیرزمینی انجام داد (وریشاکیس، ۲۰۱۵). اقدام به شناسایی مناطقی از آب‌های زیرزمینی با پتانسیل بالا به‌وسیله استفاده از 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، نتایج مقدار میزان صحت مدل‌ها به ترتیب $0/81$ ، $0/78$ و $0/71$ بدست آمد. بنابراین، مدل BRT بهترین عملکرد در نقشه‌برداری ذخایر چشمه‌های آب زیرزمینی داشته و بعد از آن به ترتیب مدل‌های CART و RF بودند. بر اساس نتایج، میزان صحت هر سه مدل بیش از ۷۰ درصد می‌باشد. بنابراین، هر سه مدل می‌تواند برای برنامه‌ریزان و مهندسان در مدیریت منابع

آب و زمین و برنامه‌ریزی در منطقه مطالعاتی مورد استفاده قرار گیرد (نقی و همکاران، ۲۰۱۶). با استفاده از داده‌های ایستگاه هواشناسی آزمایشگاه انرژی و فیزیک محیط بخش فیزیک دانشگاه پاتریس در یونان، از تکنیک‌های داده‌کاوی در تخمین دمای بیشینه، کمینه و متوسط بهره گرفتند و نتیجه گرفتند که الگوریتم رگرسیونی داده‌کاوی این امکان را می‌دهند که دمای بیشینه، کمینه و متوسط را با دقت رضایت‌بخشی پیش‌بینی نمایند. همچنین آن‌ها یک تکنیک هیبرید داده‌کاوی برای تخمین مقادیر روزانه دمای میانگین ارائه نمودند و به همان نتیجه دست یافتند (کتسیانتس و همکاران، ۲۰۰۸).

مواد و روش‌ها منطقه مطالعاتی

دشت قزوین با مساحتی حدود ۴۵۰۰۰۰ هکتار در محدوده‌ی 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 (شکل ۱). این دشت از جلگه آبرفتی پهناوری تشکیل شده که از رسوبات جریان‌های سطحی کوه‌ها اطراف تشکیل شده است (محمدی و همکاران، ۱۳۹۰). 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 سالانه ۱/۵ متر سطح سفره‌های زیرزمینی و فرونشست 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 سطح ایستابی 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 (باریکانی و همکاران، ۱۳۹۰).

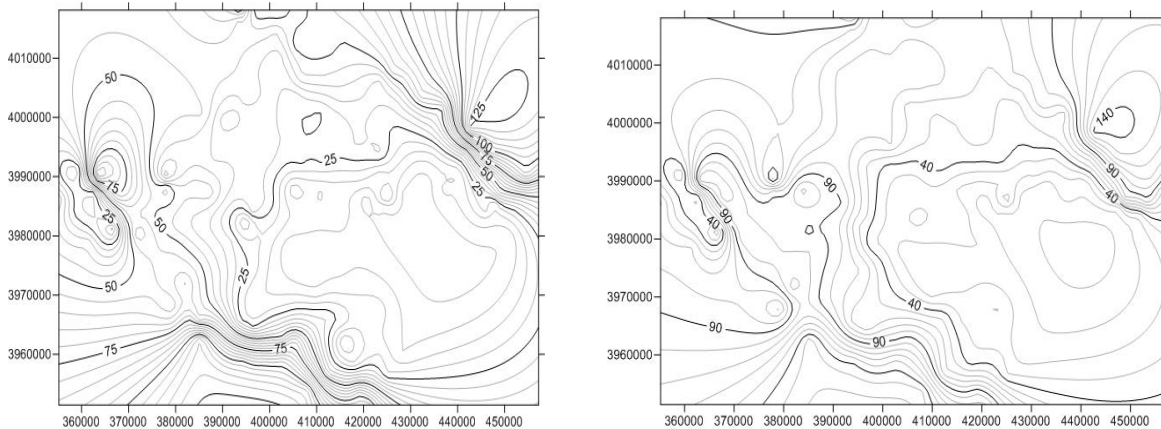
داده‌های مکانی-زمانی در زمینه مدیریت منابع آب بسیار زیاد می‌باشد، که استفاده از روش‌های تجربی و آماری در تبدیل چنین داده‌های به دانش کاربردی، عملاً غیرممکن است. داده‌کاوی یک فناوری توانمند در مدیریت و سازمان‌دهی اطلاعات و همچنین استخراج دانش‌های سودمند با حجم بالای داده می‌باشد. در این تحقیق فرض بر آن است که از این روش بتوان برای مدیریت بهتر آبخوان استفاده کرد. لذا با توجه به نیاز به الگوریتمی قوی و مناسب در این زمینه و توانمندی الگوریتم‌های داده‌کاوی در خصوص مدیریت آبخوان ضرورت به‌کارگیری آن احساس می‌شود. 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



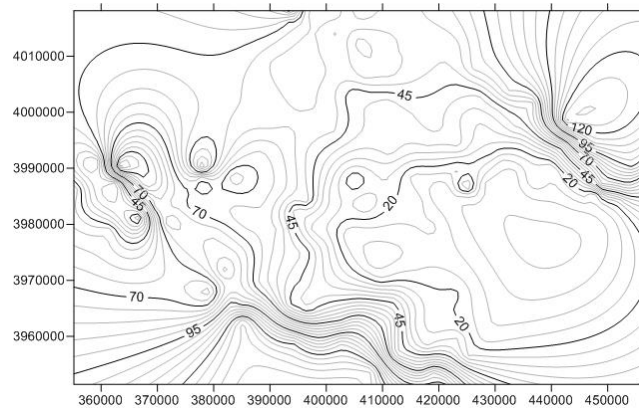
شکل ۱- موقعیت محدوده آبخوان دشت قزوین در بین شهرستان‌های قزوین.

تراز سطح ایستابی دشت قزوین

به‌منظور مشخص نمودن تراز متوسط سطح آب زیرزمینی در محدوده دشت قزوین از آمار ماهانه مربوط به ۱۶۸ پیژومتر مشاهده‌ای موجود که به‌عنوان پیژومترهای معرف منطقه می‌باشند، استفاده شد. خطوط هم‌تراز سطح ایستابی توسط نرم‌افزار SURFER رسم شد. رسم کمترین، بیشترین و میانگین سطح خطوط هم‌تراز آب زیرزمینی از سطح آزاد دریا مربوط به ماه شهریور از سال ۱۳۸۰ تا ۱۳۹۴ به ترتیب در شکل‌های (۲، ۳ و ۴) مشخص شده است. سطح تراز آب زیرزمینی در هر سه حالت نشان می‌دهد که روند کلی جریان، از طرف شمال شرقی به سمت مرکز و جنوب شرقی می‌باشد. سمت غرب و جنوب غرب آبخوان نیز تقریباً خطوط هم‌تراز با یکدیگر برابر و نشان دهنده سطح ایستابی بدون شیب یا همان تخت^۱ می‌باشد.



شکل ۲- کمترین سطح تراز آبخوان دشت قزوین از سطح آزاد دریا شکل ۳- بیشترین سطح تراز آبخوان دشت قزوین از سطح آزاد دریا



شکل ۴- میانگین سطح تراز آبخوان دشت قزوین از سطح آزاد دریا

^۱ - Flat

شبکه آبیاری دشت قزوین

بوده که روی خط تراز ۱۳۰۰ متر احداث شده است و تقریباً به موازات اتوبان تهران-زنجان قرار دارد. ظرفیت کانال اصلی از ۱ تا ۳۰ مترمکعب در ثانیه متغیر است. مشخصات کلی شبکه آبیاری قزوین در جدول (۱) ارائه شده است (ستوده‌نیا و ستوده‌نیا، ۱۳۹۵).

موقعیت شبکه در فاصله ۵۰ کیلومتری غرب شهر تهران بین ۳۶ درجه و ۲۰ دقیقه تا ۳۶ درجه عرض شمالی و ۴۹ درجه و ۴۰ دقیقه الی ۵۰ درجه و ۳۵ دقیقه طول شرقی واقع شده است. شبکه آبیاری دارای یک کانال به طول ۹۴ کیلومتر

جدول ۱ - مشخصات و ظرفیت کانال‌های شبکه آبیاری قزوین

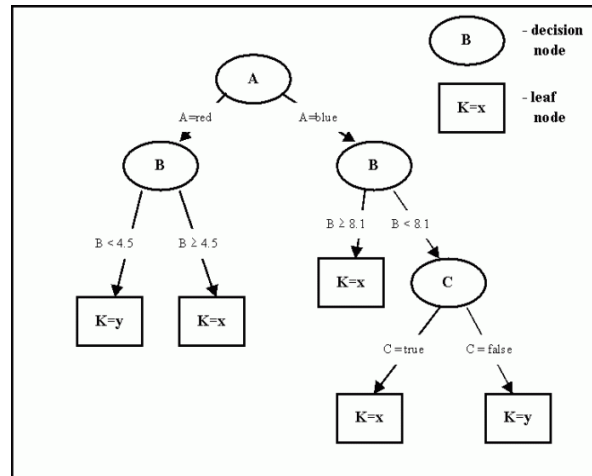
نوع کانال	شکل مقطع	نوع پوشش	ظرفیت طراحی $M^3 S^{-1}$	مشخصات مقطع	
				طول کانال Km	عمق (متر) عرض (متر)
درجه یک	مستطیلی	بتنی	۳۰	۳	۴
درجه دو	دو زنگه‌ای	بتنی	۲/۹-۳۰	۰/۹-۲/۵	۱/۲-۴
درجه سه	دو زنگه‌ای	بتنی	۰/۶-۷/۴	۰/۶۵-۱/۵	۰/۸-۲
درجه چهار	دو زنگه‌ای	بتنی	۰/۱۷-۱	۰/۶-۰/۸	۰/۶
			۰/۱۷-۰/۳۴	۰/۴-۰/۵۵	۰/۳-۰/۶

درخت تصمیم^۳

یکی از الگوریتم‌های دسته‌بندی، درخت تصمیم‌گیری است که مدل خود را بر اساس یک درخت پیاده‌سازی می‌کند. در این الگوریتم با توجه به مجموعه آموزش یک درخت بر اساس ویژگی‌های مختلف آن درست می‌شود که با استفاده از این درخت باید بتوان یک عضو جدید را در دسته خاصی طبقه‌بندی کرد. با جستجو در گره‌های درخت می‌توان به یک برگ رسید که آن برگ مشخص کننده نوع عضو جدید است. ریشه درخت محل آغازین پیمایش درخت است که به بررسی اولین متغیر می‌پردازد. گره‌ها نیز مثل ریشه به بررسی متغیرها می‌پردازد. نتیجه بررسی آن است که با شاخه‌های آن گره به گره مورد نظر دیگری برود و یا به برگ برسد. برگ‌های نهایی پایان درخت هستند. نقاطی که وضعیت داده را مشخص می‌کند (شکل ۵).

نرم‌افزار داده‌کاوی IBM SPSS Modeler

نرم‌افزار IBM SPSS Modeler برنامه‌ای از شرکت SPSS است که در ابتدا با نام کلمنتاین^۲ ارائه شد که در نسخه جدید به SPSS Modeler تغییر نام پیدا کرد. این نرم‌افزار یکی از بهترین ابزارهای داده‌کاوی است و برنامه‌های 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 مزیت‌های نسبت به نرم‌افزار Clementine 12 است. مانند 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.



شکل ۵- نمای از درخت تصمیم حاصل از اجرای الگوریتم درختی

ادامه می‌دهد. در تقسیم داده‌ها به صورت بازگشتی به زیرگروه‌ها فقط دو زیرگروه را پشتیبانی می‌کند. یک الگوریتم سریع است که هرس نمودن آن رو به عقب است (چتامویلی، ۲۰۱۱). به‌طور کلی روش، CHAID برای ساخت یک درخت تصمیم‌گیری، داده‌ها را متناوباً به زیر مجموعه‌های مشابه افزایش می‌کند تا آنجا که هر زیر مجموعه دارای تعداد مشخصی نمونه شود. این الگوریتم می‌تواند درختی تولید کند که در برخی از مواقع به صورت غیر دو دویی عمل کند، در واقع از روش جدا کردن چندتایی به جای جدا کردن دو دویی استفاده می‌کند. به این صورت که می‌توان گره والد را به بیش از دو تقسیم نماید. این الگوریتم از آزمون «کای دو» برای تصمیم‌گیری در هر تقسیم برای مشخص کردن گره‌های فرزند استفاده می‌کند. سپس شاخه‌های درخت ساخته شده تا تحقق معیار توقف یا رسیدن به سطح پیچیدگی خواسته شده، هرس می‌شوند. به بیان دیگر، CHAID ابتدا تفاوت‌های هر نمونه را با سایر نمونه‌ها می‌یابد و درخت مورد نظر را تولید می‌کند. هرس کردن درخت از طریق یافت تفاوت‌های مشابه انجام می‌شود (چتامویلی، ۲۰۱۱).

ماشین بردار پشتیبان یا CVM به‌عنوان یکی از قوی‌ترین و دقیق‌ترین روش‌ها در میان الگوریتم‌ها معروف شناخته می‌شود. ماشین بردار پشتیبان یکی از روش‌های یادگیری با ناظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای

به‌منظور پیش‌بینی ویژگی هدف وضعیت نوسانات عمق آبخوان، از مدل درخت تصمیم استفاده شده است. الگوریتم درخت تصمیم قادر است علاوه بر متغیرهای کمی، متغیرهای کیفی را نیز پیش‌بینی کند (بریمان و همکاران، ۱۹۸۴). الگوریتم درخت تصمیم به گونه‌ای عمل می‌کند که سعی دارد گوناگونی و یا تنوع (از نظر ویژگی هدف) را در گره‌ها به حداقل ممکن برساند. این عدم یکنواختی در گره‌ها با استفاده از معیارهای عدم خلوص قابل اندازه‌گیری است که مهم‌ترین و پرکاربردترین آن شاخص جینی می‌باشد (یونی‌یاما و همکاران ۲۰۰۲). اغلب تفاوت انواع درخت‌های تصمیم در همین معیار اندازه‌گیری عدم خلوص، شیوه شاخه‌بندی و هرس کردن گره‌های درخت می‌باشد. در این تحقیق از سه نوع الگوریتم درخت تصمیم CART^۴، CHAID^۵ و CVM^۶ استفاده شد. الگوریتم CART متغیرهای ورودی را برای یافتن بهترین تجزیه می‌آزماید تا شاخص ناخالصی حاصل از تجزیه کمترین مقدار باشد. در تجزیه دو زیرگروه تعیین می‌شود و هر کدام در مرحله بعد به دو زیرگروه دیگر تقسیم خواهند شد و این روند ادامه می‌یابد تا زمانی که یکی از معیارهای توقف برآورده شود.

درخت CART بازگشتی دوتایی 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

4 - Classification and Regression Tree

5- Chi-square Automatic Interaction Detector

6 - Support Vector Machines

این معیارها که توجه بیشتری به دسته‌بندی مثبت نشان می‌دهند، توانایی دسته‌بندی را در تشخیص دسته مثبت تبیین می‌کنند. معیار حساسیت نشان می‌دهد که دقت تشخیص دسته مثبت چه مقدار است و معیار نرخ مثبت غلط نرخ هشدار غلط را با توجه به دسته منفی بیان می‌کند (سلاویا و خوش‌گفتار، ۲۰۱۱). شاخص‌های مورد نظر طبق روابط (۱) تا (۹) محاسبه می‌گردند (هان، ۲۰۰۰).

TP = تعداد داده‌های برچسب مثبتی که درست دسته‌بندی شده‌اند، FP = تعداد داده‌های برچسب منفی که به نادرست مثبت دسته‌بندی شده‌اند، FN = تعداد داده‌های برچسب مثبتی که به نادرست منفی دسته‌بندی شده‌اند، TN = تعداد داده‌های برچسب منفی که درست دسته‌بندی شده‌اند.

$$TPR = \frac{TP}{TP + FN} \quad \text{رابطه (۱)}$$

$$TNR = \frac{TN}{FP + TN} \quad \text{رابطه (۲)}$$

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{رابطه (۳)}$$

$$PPV = \frac{TP}{TP + FP} \quad \text{رابطه (۴)}$$

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \quad \text{رابطه (۵)}$$

$$FNR = \frac{FN}{TP + FN} = 1 - TPR \quad \text{رابطه (۶)}$$

$$FM = \frac{2 \times TP}{TP + FP + FN} \quad \text{رابطه (۷)}$$

$$GM = \sqrt{TPR \times TNR} \quad \text{رابطه (۸)}$$

$$ER = \frac{FP + FN}{TP + FN + FP + TN} = 1 - Accuracy \quad \text{رابطه (۹)}$$

طبقه‌بندی از جمله شبکه‌های عصبی پرسپترون نشان داده است. مبنای کاری دسته‌بندی کننده‌ی SVM دسته‌بندی خطی داده‌هاست. به منظور اعتبارسنجی مدل درختی داده‌ها به دو بخش داده‌های آموزش و آزمون تقسیم شدند. مدل درختی با استفاده از داده‌های آموزش و آزمون تقسیم شدند. مدل درختی با استفاده از داده‌های آموزش ساخته شد و مدل ساخته شده بر روی داده‌های آزمون مورد تست قرار گرفت. درصد نمونه‌هایی از داده‌های آزمون که ویژگی هدف آن‌ها توسط مدل، درست تشخیص داده شده بود دقت مدل را بیان می‌کند (گوپا، ۲۰۱۱). برای هر سه الگوریتم درخت تصمیم ۷۵ درصد داده‌ها به صورت تصادفی به عنوان داده‌های آموزش انتخاب شدند و ۲۵ درصد مابقی به عنوان داده‌های آزمون مورد آزمایش قرار گرفتند.

برای ارزیابی مدل‌ها و انتخاب بهترین مدل از شاخص‌های حساسیت (TPR^۷)، شفافیت (TNR^۸)، صحت (ACC^۹)، دقت (PPV^{۱۰})، نرخ مثبت غلط (FPR^{۱۱})، نرخ منفی غلط (FNR^{۱۲})، شاخص F (FM^{۱۳})، میانگین هندسی (GM^{۱۴}) و نرخ خطا (ER^{۱۵}) استفاده شد. شاخص‌ها مذکور به ترتیب از رابطه (۱) تا (۹) مشخص شدند. در مسائل واقعی، معیار صحت دسته‌بندی، به هیچ عنوان معیار مناسبی برای ارزیابی کارایی الگوریتم‌های دسته‌بندی نمی‌باشد، به دلیل اینکه در رابطه صحت دسته‌بندی، ارزش رکوردهای دسته‌های مختلف یکسان در نظر گرفته می‌شوند. بنابراین، مسائلی که با دسته‌های نامتعادل سروکار دارند و یا در مسائلی که ارزش دسته‌ای در مقایسه با دسته دیگر متفاوت است، از معیارهای دیگری استفاده می‌شود. در مسائل واقعی معیارهای دیگری نظیر حساسیت و نرخ مثبت غلط اهمیت ویژه‌ای دارند.

7 - True Positive Rate
8 - True Negative Rate
9- Accuracy
10- Positive Predictive Value or Precision
11- False Positive Rate
12- False Negative Rate
13-F - Measure
14-Geometric Mean
15-Error Rate

داده‌ها این مسئله می‌تواند تا حد زیادی رفع گردد. هدف از این روش آن است که داده‌ها را برحسب قواعدی در دسته‌بندی‌هایی قرار گرفته شوند و دسته‌ای را که تعداد داده‌های موجود در آن بسیار کم باشد، کنار گذاشته شود. گسسته‌سازی فرایندی است که صفات پیوسته را به گسسته تبدیل می‌کند، در این تحقیق 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 داده‌ها مربوطه باشد. بعد از انجام گسسته‌سازی برای کل داده‌ها، برای هر محدوده داده‌ها 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 می‌باشد.

مجموعه داده‌های مورد استفاده در این تحقیق از اطلاعات مربوط به دشت قزوین طی دوره سال‌های ۱۳۸۰ تا ۱۳۹۴ به صورت ماهانه می‌باشد. متغیرهای ورودی مورد استفاده از عوامل انسانی و عوامل طبیعی انتخاب شدند که شامل مقدار حجم آب خروجی چاه‌های کشاورزی (میلیون مترمکعب)، حجم آب شبکه آبیاری (میلیون مترمکعب)، مقدار تقاضای آب کشاورزی (میلیون مترمکعب)، مقدار بارندگی (میلیون مترمکعب)، دما (سانتی‌گراد)، رطوبت هوا (درصد) و تبخیر و تعرق (میلی‌متر در روز) می‌باشند. همچنین از داده‌های مربوط به مقدار نوسانات عمق آبخوان به عنوان داده‌های هدف به مدل معرفی شدند. برای افزایش قابلیت استفاده نتایج برای مدیران و همچنین کاهش هر چه بیشتر اثرات این خطاها، گسسته‌سازی (دسته‌بندی) مقادیر نوسانات عمق آب زیرزمینی انجام شد. اگر خطای قرائت‌ها در حد کمتر از یک متر باشد، با استفاده از گسسته‌سازی، اثر این خطاها تا حد زیادی از بین می‌رود. همچنین برای مقادیر داده‌های پرت، با انجام فرآیند دسته‌بندی

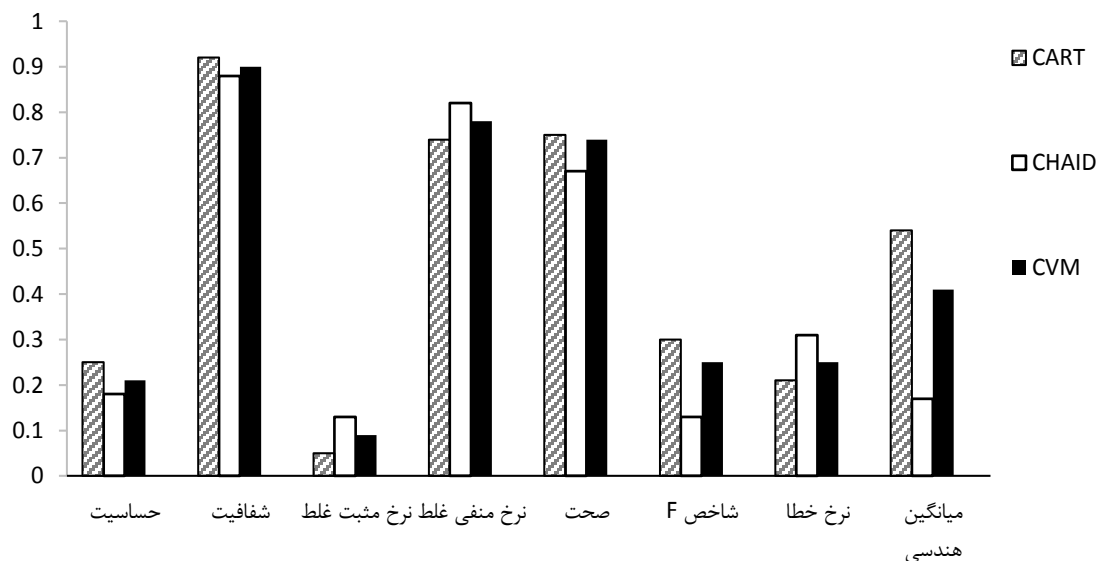
جدول ۲- تعیین مدل برای محدوده داده‌های نوسانات عمق آبخوان

نام مدل	محدوده داده‌ها (m)	میانگین داده‌ها (m)	درصد داده‌ها
A	$-1 \geq \dots$	-۱/۶۶	۱۰/۲۳
B	$-0/1 > \dots \geq -0/3$	-۰/۵۶	۲۵/۷۹
C	$0/3 > \dots \geq 0$	-۰/۱۵	۲۴/۱۴
D	$0 > \dots \geq 0/3$	۰/۱۵	۱۷/۶۶
E	$0/3 > \dots \geq 1$	۰/۵۵	۱۶/۷۸
F	$1 \geq \dots$	۱/۷۳	۵/۴

بیشتری را در جای درست خود طبقه‌بندی کرده است. همچنین شاخص‌های نرخ‌های مثبت غلط، منفی غلط و خطا برای مدل CART نسبت به دو الگوریتم دیگر، کمترین مقدار را دارند. مقادیر کم این شاخص‌ها تأیید کننده وقوع خطای کمتر در طبقه‌بندی نمونه‌ها است. شکل ۶ نمودار نتایج شاخص‌ها جهت مقایسه بین الگوریتم‌ها جهت پیش‌بینی نوسانات عمق آبخوان را نشان می‌دهد.

بحث و نتایج

مقادیر نتایج شاخص‌های بدست آمده توسط سه الگوریتم CART، CHAID و CVM روی مجموعه داده‌ها نشان داد که الگوریتم CART بهترین مدل را تولید نموده است. شاخص‌های حساسیت، شفافیت، صحت، شاخص F و میانگین هندسی برای مدل CART نسبت به دو الگوریتم CHAID و CVM دارای بیشترین مقدار است. مقادیر این شاخص‌ها هر چه بیشتر باشد نشان‌دهنده این است که طبقه‌بندی مورد استفاده نمونه‌های



شکل ۶- نمودار نتایج شاخص‌ها برای مدل‌های تولید شده

۳- اگر در ماه‌های اردیبهشت، خرداد و مهر مقدار حجم آب شبکه آبیاری کمتر مساوی $0/025$ میلیون مترمکعب و مقدار حجم آب خروجی از چاه‌های کشاورزی بیشتر از $0/08$ میلیون مترمکعب باشد، با احتمال ۳۳ درصد مدل B یعنی افتی بین $0/3$ تا ۱ متر در آبخوان اتفاق می‌افتد.

۴- اگر ماه‌های آذر، دی، بهمن و اسفند مقدار تبخیر و تعرق پتانسیل کمتر مساوی $1/7$ میلی‌متر در روز و حجم آب خروجی از چاه‌های کشاورزی کمتر مساوی $0/01$ میلیون مترمکعب باشد، با احتمال ۴۳ درصد مدل E یعنی بالآمدگی بین $0/3$ تا ۱ متر در آبخوان اتفاق می‌افتد.

۵- اگر ماه‌های آذر، دی، بهمن و اسفند مقدار تبخیر و تعرق پتانسیل کمتر مساوی $1/7$ میلی‌متر در روز و حجم آب خروجی از چاه‌های کشاورزی بیشتر از $0/01$ میلیون مترمکعب و حجم بارندگی بیشتر از $0/87$ میلیون مترمکعب باشد، با احتمال ۴۳ درصد مدل E یعنی بالآمدگی بین $0/3$ تا ۱ متر در آبخوان اتفاق می‌افتد.

۶- اگر ماه‌های آذر، دی، بهمن و اسفند مقدار تبخیر و تعرق پتانسیل کمتر مساوی $1/7$ میلی‌متر در روز و حجم آب خروجی از چاه‌های کشاورزی بیشتر از $0/01$ میلیون مترمکعب و حجم بارندگی کمتر مساوی $0/87$ میلیون مترمکعب باشد، با احتمال

از بین الگوریتم‌های مورد استفاده بهترین نتایج مربوط به الگوریتم CART با بیشترین مقدار صحت $0/75$ ، شفافیت $0/92$ ، حساسیت $0/25$ و میانگین هندسی $0/54$ و همچنین کمترین مقدار نرخ‌های مثبت غلط $0/05$ ، منفی غلط $0/74$ و خطا $0/21$ می‌باشد.

با توجه به عملکرد بهتر الگوریتم CART نسبت به دو الگوریتم CAHID و CVM از اطلاعات خروجی این الگوریتم برای پیش‌بینی نوسانات عمق آبخوان دشت قزوین استفاده شد. برخی از قوانین تولید شده در خصوص تغییرات عمق آبخوان حاصل از اجرای الگوریتم CART به شرح زیر می‌باشد:

۱- اگر در ماه‌های فروردین و آبان مقدار حجم آب خروجی از چاه‌های کشاورزی کمتر مساوی $0/04$ میلیون مترمکعب و تبخیر تعرق پتانسیل کمتر مساوی $1/2$ میلی‌متر در روز باشد، با احتمال ۲۸ درصد مدل D یعنی بالآمدگی بین ۰ تا $0/3$ متر در آبخوان اتفاق می‌افتد.

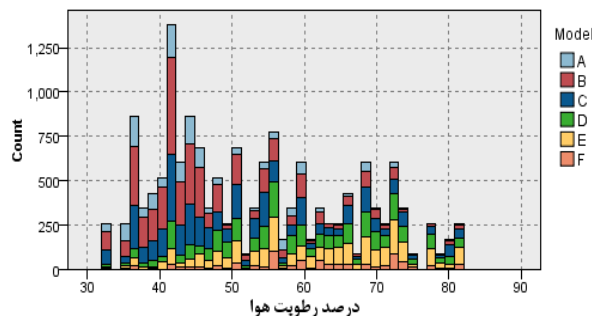
۲- اگر در ماه‌های اردیبهشت، خرداد و مهر مقدار حجم آب شبکه آبیاری کمتر مساوی $0/025$ میلیون مترمکعب و مقدار حجم آب خروجی از چاه‌های کشاورزی کمتر مساوی $0/08$ میلیون مترمکعب باشد، با احتمال ۴۳ درصد مدل C یعنی افتی بین ۰ تا $0/3$ متر در آبخوان اتفاق می‌افتد.

همان‌طور که از رابطه بین درصد رطوبت هوا و مدل‌ها مشخص شد رابطه عکس بین مقدار درصد رطوبت هوا و مقدار افت آبخوان وجود دارد به طوری که با افزایش مقدار درصد رطوبت هوا مقدار افت کمتر می‌شود. در شکل (۹) مشخص می‌شود با افزایش مقدار حجم تقاضای مصرف آب کشاورزی مقدار افت نیز بیشتر می‌شود به طوری که مقدار مدل B (افتی بین ۰/۳ تا ۱ متر) نسبت به سایر مدل‌ها با افزایش مقدار تقاضای آب کشاورزی بیشتر می‌شود. همان‌طور که از شکل (۱۰) مشخص می‌شود دمای هوا تا حدود ۱۵ درجه سانتی‌گراد باعث افزایش مدل E (بالا آمدگی بین ۰ تا ۰/۳ متر) و با افزایش مقدار دما از ۱۵ درجه تا حدود ۳۰ درجه سانتی‌گراد باعث افزایش مدل B (افتی بین ۰/۳ تا ۱ متر) نسبت به سایر مدل‌ها می‌شود. طبق شکل (۱۱) مشخص می‌شود که مقدار تبخیر و تعرق پتانسیل تا مقدار دو میلی‌متر در روز باعث افزایش مدل C (افتی بین ۰ تا ۰/۳ متر) و از دو تا حدود پنج و نیم میلی‌متر باعث افزایش مدل B (افتی بین ۰/۳ تا ۱ متر) نسبت به سایر مدل‌ها می‌شود. طبق شکل (۱۲) مشخص می‌شود که با افزایش مقدار حجم برداشتی از چاه‌های کشاورزی مقدار افت آبخوان نیز افزایش پیدا می‌کند و مقدار مدل B نسبت به سایر مدل‌های دیگر بیشتر می‌شود. از شکل ۱۴ مشخص می‌شود با افزایش مقدار بارندگی مقدار مدل‌های C و D در نوسان است و به‌طور متوسط عمق آبخوان تغییر چندانی نکرده است.

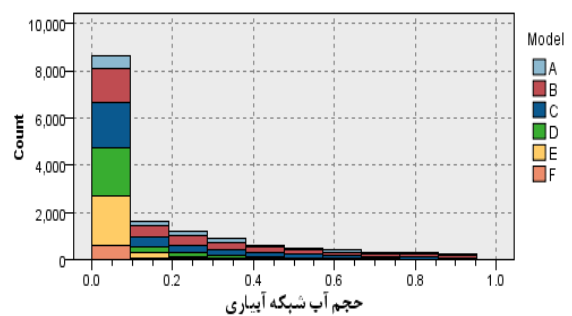
۴۰ درصد مدل D یعنی بالا آمدگی بین ۰ تا ۰/۳ متر در آبخوان اتفاق می‌افتد.

۷- اگر ماه‌های آذر، دی، بهمن و اسفند مقدار تبخیر و تعرق پتانسیل بیشتر از ۱/۷ میلی‌متر در روز و حجم آب خروجی از چاه‌های کشاورزی بیشتر از ۰/۱ میلیون مترمکعب باشد، با احتمال ۴۳ درصد مدل E یعنی بالا آمدگی بین ۰/۳ تا ۱ متر در آبخوان اتفاق می‌افتد.

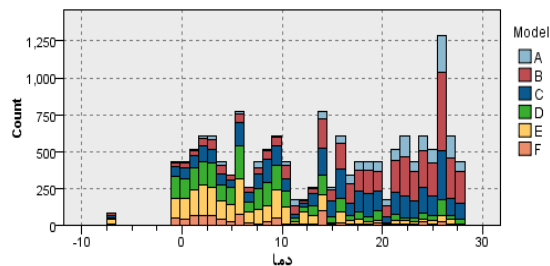
شکل‌های شامل هفت نمودار می‌باشد که در محور عمودی تمامی نمودارها، مقدار مدل‌ها (با توجه به جدول ۲) و در محور افقی به ترتیب مقدار حجم آب شبکه آبیاری (میلیون مترمکعب)، مقدار رطوبت هوا (درصد)، مقدار حجم تقاضای آب کشاورزی (میلیون مترمکعب)، مقدار دما (سانتی‌گراد)، مقدار تبخیر و تعرق پتانسیل (میلی‌متر در روز) و مقدار حجم آب خروجی از چاه کشاورزی (مترمکعب)، مقدار حجم بارندگی (میلیون مترمکعب) را شامل می‌شود. همان‌طور که از شکل ۷ مشخص می‌شود با افزایش مقدار حجم آب شبکه آبیاری از مقدار ۰/۱ میلیون مترمکعب به بعد، مقدار مدل B (افتی بین ۰/۳ تا ۱ متر) نسبت به سایر مدل‌ها افزایش پیدا می‌کند. طبق شکل (۹) نیز مشخص می‌شود درصد رطوبت هوا تا ۵۰ درصد باعث افزایش مدل B (افتی بین ۰/۳ تا ۱ متر) و از ۵۰ درصد تا ۶۰ درصد باعث افزایش مدل C (افتی بین ۰ تا ۰/۳ متر) و از ۶۰ به بعد باعث افزایش مدل E (بالا آمدگی بین ۰ تا ۰/۳ متر) نسبت به سایر مدل‌های دیگر می‌شود.



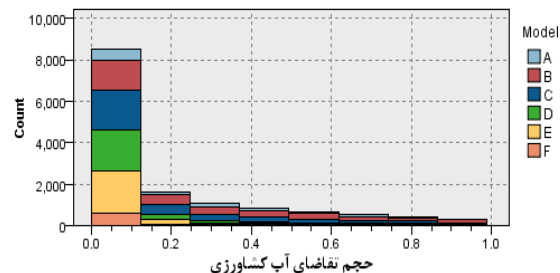
شکل ۸- رابطه بین مقدار درصد رطوبت و مقدار مدل‌ها



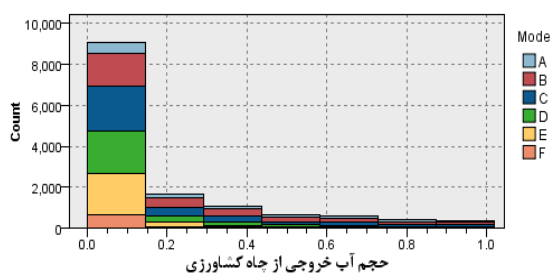
شکل ۷- رابطه بین مقدار حجم شبکه آبیاری و مقدار مدل‌ها



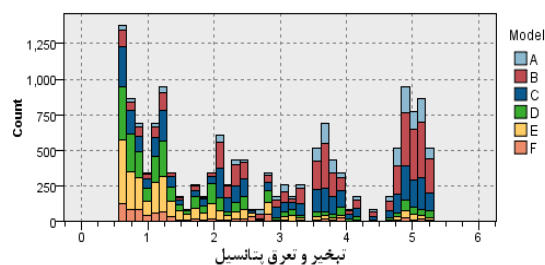
شکل ۱۰- رابطه بین مقدار دما و مقدار مدل‌ها



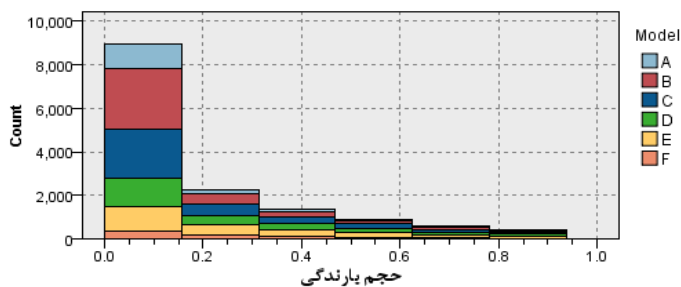
شکل ۹- رابطه بین مقدار حجم تقاضای آب کشاورزی و مقدار مدل‌ها



شکل ۱۲- رابطه حجم آب خروجی از چاه کشاورزی و مقدار مدل‌ها

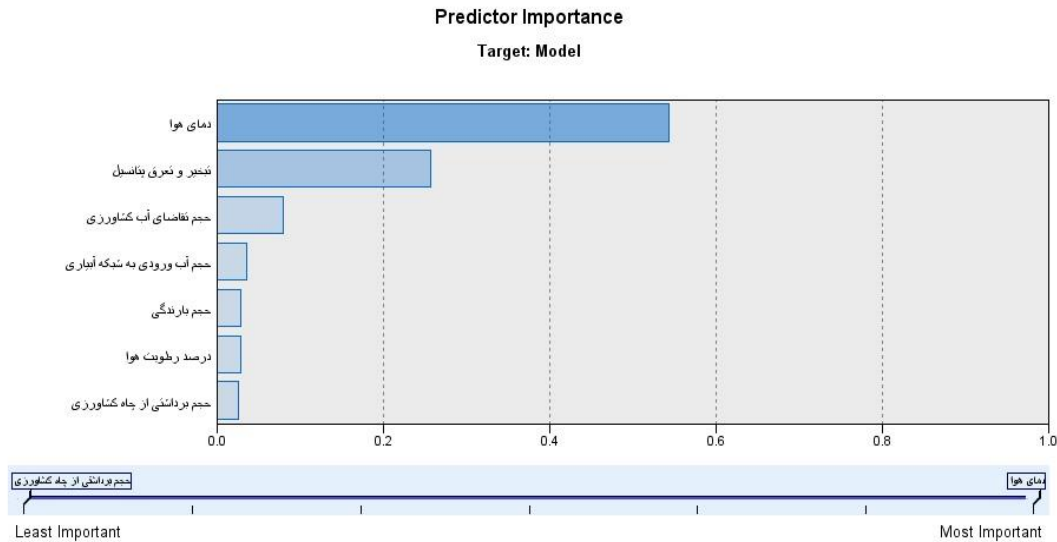


شکل ۱۱- رابطه مقدار تبخیر و تعرق پتانسیل و مقدار مدل‌ها



شکل ۱۳- رابطه بین مقدار حجم بارندگی و مقدار مدل‌ها

با توجه به نتایج خروجی مدل CART در تعیین اهمیت متغیرهای ورودی مشخص شد که تأثیرگذارترین متغیرهای ورودی در پیش‌بینی تغییرات عمق آبخوان به ترتیب دما و تبخیر و تعرق و مقدار تقاضای آب کشاورزی می‌باشد و بقیه متغیرهای ورودی تقریباً دارای اهمیت یکسان می‌باشند (شکل ۱۱).



شکل ۱۱- تعیین اهمیت متغیرهای ورودی در پیش‌بینی تغییرات عمق آبخوان

جدول (۳) نتایج حاصل شده از اجرای الگوریتم CART، برای وضعیت محتمل عمق آبخوان در ماه‌های مختلف سال می‌باشد.

جدول ۳- وضعیت محتمل در آبخوان دشت قزوین در ماه‌های مختلف سال

ماه	وضعیت محتمل آبخوان	توضیحات
اردیبهشت، خرداد، مهر	افت (با احتمال ۷۶/۹۳ درصد)	افت بیشتر از ۲/۲ متر با احتمال ۱/۱۳ درصد افت در حدود ۱ متر با احتمال ۱۱/۲۶ درصد افت کمتر از ۰/۳ متر با احتمال ۳۴/۸۴ درصد
تیر، شهریور، مرداد	افت (با احتمال ۸۵/۷ درصد)	افت بیشتر از ۲/۲ متر با احتمال ۱/۱۸ درصد افت در حدود ۱ متر با احتمال ۱۹/۲۱ درصد افت کمتر از ۰/۳ متر با احتمال ۲۹ درصد
فروردین، آبان	افت (با احتمال ۴۳/۸۳ درصد)	افت بیشتر از ۲/۲ متر با احتمال ۱/۵۳ درصد افت در حدود ۱ متر با احتمال ۴/۶۷ درصد افت کمتر از ۰/۳ متر با احتمال ۲۵/۶۳ درصد
آذر، دی، بهمن، اسفند	بالآمدگی (با احتمال ۷۲/۰۳ درصد)	بالآمدگی بیشتر از ۲/۲ متر با احتمال ۱/۵۱ درصد بالآمدگی در حدود ۱ متر با احتمال ۱۲/۲۱ درصد بالآمدگی کمتر از ۰/۳ متر با احتمال ۲۸/۵۳ درصد

نتیجه‌گیری

چندین ماه متوالی اثرات آن در تقاضای آب مصرفی مشخص می‌شود. لذا یکی از بهترین و ساده‌ترین راه‌های مدیریت مصرف آب در بخش کشاورزی، مدیریت عرضه مربوط به مصرف بهینه و پایدار می‌باشد. در ماه‌های مهر تا بهمن، علاوه بر میزان بیشتر عرضه آب، کاهش محسوس مصرف آب نیز مشاهده می‌شود. لذا در این دوره افت در آبخوان کاهش و به‌طور متوسط در آبخوان بالآمدگی مشاهده می‌شود. بنابراین اولویت اصلی در مدیریت

بررسی مطالعات خارجی و داخلی نشان می‌دهد که تاکنون از روش‌های مختلف داده‌کاوی برای پیش‌بینی نوسانات آبخوان استفاده شده است. ناپارامتری بودن روش طبقه‌بندی درختی، تفسیر آسان نتایج، عدم نیاز به پیش‌فرض‌هایی مشابه سایر مدل‌های پیش‌بینی کننده از مزایای الگوریتم‌های درخت تصمیم است. با افزایش و کاهش عرضه آب به آبخوان در

- Breiman, L., Friedman, J., Olshen R, Stone Ch., 1984. Classification and regression trees. Boca Raton: Chapman & Hall/CRC.
- Chattamvelli, R. 2011. Data Mining Algorithms, First Edition. Alpha Science International. 8: 274-290.
- Corona O L, Fuentes O E, Casique E M, Longoria P P, 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.
- Han, J., Kamber, M., 2006. Data Mining: Concepts and Techniques. 2nd ed. Morgan Kaufman.
- Huajie, D., Zhengdong, D., 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, 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, Rana, T., 2008. Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. Irrig Drainage Syst. 22: 159-177.
- Kotsiantis, S., Kostoulas, A., Lykoudis S., Argirio, A, Menagias, K., 2008. Using data mining techniques for estimating minimum, maximum and average daily temperature value. IJMPES. 1(2): 117-121.
- Naghibi, S. A., Pourghasemi, H, R, 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. 88:44-60.
- Norouzi, H., Moghaddam, A., Nadiri, A., 2016. Detemining the vulnerable areas of malekan plain aquifer for nitrate using random. ournal of Environmental Studies. 41(4): 923-942.
- Oorkavalan, G., Chidambaram, S,M., Mariappan, V., Kandaswamy, G., 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.
- Varouchakis, E., 2015. Integrated water resources analysis at Basin scale: a case study in Greece. Journal of Irrigation and Drainage Engineering. 142(3): 1-12.
- Yoneyama, Y., Suzuki S, Sawa R, Yoneyama K, Power GG, Araki T., 2011. Increased plasma adenosine

منابع آب در این دوره مربوط به عرضه و جبران اثرات مصرف آب در دوره مصرف یا همان فصل‌های زراعی می‌باشد. با توجه به اطلاعات خروجی الگوریتم CART نتیجه گرفته می‌شود که ۱- بیشترین احتمال مقدار افت‌ها در ماه‌های تیر، مرداد و شهریور رخ می‌دهد. ۲- مقدار بالآمدگی کمتر از ۰/۳ متر در ماه‌های آذر، دی و بهمن با احتمال بیشتر از مقدار بالآمدگی حدود یک متر در این ماه‌ها اتفاق افتاده می‌افتد. ۳- مقدار افت حدود یک متر در ماه‌های اردیبهشت، خرداد و مهر با احتمال بیشتر از افت بیشتر از ۲/۲ متر در این ماه‌ها اتفاق می‌افتد. ۴- مقدار بالآمدگی در ماه‌های آذر، دی، بهمن و اسفند با احتمال بیشتر از ماه‌های دیگر اتفاق افتاده است. همچنین با توجه به نتایج الگوریتم CART مشخص شد که مؤثرترین پارامتر در تغییرات عمق آبخوان مقدار دما می‌باشد به طوری که 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 دما و افت می‌باشند، مدیران و کشاورزان از الگوی کشت مناسب‌تر و کم‌مصرف‌تر استفاده کنند تا مشکلات ناشی از افت کمتر رخ دهد. با توجه به اهمیت وضعیت آبخوان برای زارعان، مطالعه و پژوهش روی این مسئله می‌تواند بسیار بارز باشد. طراحی و تدوین مدل‌های پیش‌بینی مناسب و با اعتبار بالا می‌تواند کمک زیادی به این مسئله باشد.

منابع

- باریکانی، ا، احمدیان، م، خلیلیان، ص، ۱۳۹۰. بهره‌برداری بهینه پایدار از منابع آب زیرزمینی در بخش کشاورزی: مطالعه موردی زیربخش زراعت دشت قزوین. نشریه اقتصادی و توسعه کشاورزی (علوم و صنایع کشاورزی)، ۲۵: ۲۶۲-۲۵۳.
- ستوده‌نیا، ع، ستوده‌نیا، س، ۱۳۹۵. بررسی نقش احداث سد مخزنی طالقان بر رسوب‌گذاری شبکه آبیاری دشت قزوین، ص: ۹-۱، پانزدهمین کنفرانس هیدرولیک ایران، دانشکده فنی و مهندسی، دانشگاه بین‌المللی امام خمینی، قزوین، ایران.
- محمدی، م، محمدی‌قلعه‌نی، م، ابراهیمی، ک، ۱۳۹۰. تغییرات زمانی و مکانی کیفیت آب زیرزمینی دشت قزوین. مجله پژوهش آب ایران، ۸: ۵۲-۴۱.

concentrations and theseverity of preeclampsia. *Obstet Gynecol.* 100(6): 1266-70.

Zhao, Y., Li, Lifen ., Y, Zhang, L., Wang, Q., 2016. Groundwater level prediction of landslide based on classification and regression tree. *Geodesy and geodynamics.* 17(5): 348-355.

Seliya, N., Khoshgoftaar, T., 2011. The use of decision trees for cost-sensitive classification: an empirical study in software quality prediction. *WIREs Data Mining and Knowledge Discovery.* 1(5): 448-459.

Stumpp C J, Zurek A J, Wachniew P, Gargini, Gemitzi A, Filippini M, Witczak S., 2016. A decision tree tool supporting the assessment of groundwater vulnerability. *Environmental Earth Sciences.* 75: 1-7.