Archive of SID

یازدهمین همایش ملی علوم و مهندسی آبخیزداری ایران
توسعه مشارکتی در مدیریت حوزه‌های آبخیز
11th National Conference on Watershed Management Sciences and Engineering of Iran
Participatory Development in Watershed Management
۳۱ فروردین لغایت ۲ اردیبهشت ۱۳۹۵
April 19-21, 2016

کاربرد الگوریتم‌های داده‌کاوی K-Means و CART در تعیین مؤثرترین عوامل کیفیت آب آشامیدنی در دشت نورآباد استان فارس

سید مسعود سلیمان‌پور*[۱]، سید حمید مصباح[۲]، بهرام هدایتی[۳]

۱- عضو هیأت علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس

۲- عضو هیأت علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس

۳- دانش‌آموخته‌ی دوره‌ی کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار، دانشگاه پیام نور تهران

(m.soleimanpour@yahoo.com)*

چکیده

امروزه بحث کیفیت آب در بسیاری از مناطق جهان به عنوان یکی از مباحث کلیدی مطرح است. زیرا این امر ارتباط بسیاری با سلامتی انسان و جامعه‌ی بشری دارد؛ به طوری که بررسی کیفیت آب (به ویژه آب آشامیدنی)، و چگونگی تغییر آن‌ها نقش بسیار مهمی در مدیریت و بهره‌برداری از منابع دارد؛ به علاوه با آگاهی از فاکتورهای مؤثر در تغییر کیفیت آب آشامیدنی می‌توان در جهت هر چه بهتر مدیریت کردن آن‌ها اقدامات مؤثری را به عمل آورد. بدین منظور پژوهش حاضر برای اولین بار نسبت به کاربرد الگوریتم‌های داده‌کاوی K-Means و CART در تعیین مؤثرترین عوامل کیفیت آب آشامیدنی در دشت نورآباد واقع در غرب استان فارس اقدام نموده است. نتایج این تحقیق با دقت تخمین زده شده‌ی ۹۸/۳۳ درصد، با استفاده از مدل‌سازی با بهره‌گیری از تکنیک‌های داده‌کاوی خوشه‌بندی K-Means و درخت تصمیم CART و بر اساس مؤثرترین فاکتورهای کیفیت آب آشامیدنی در نرم‌افزار Clementine (نسخه‌ی ۱۲)، حاصل شده است، نشان داد که مؤثرترین فاکتورهای مطلوب آب آشامیدنی در این منطقه، تابع سختی کل (TH) و هدایت الکتریکی عصاره‌ی اشباع (EC) می‌باشد. بدین ترتیب، در صورتی که سختی کل در این دشت بین ۱۷۵ و ۴۴۰ قسمت در میلیون، و هدایت الکتریکی عصاره‌ی اشباع آن، کمتر یا مساوی ۷۳۱/۶۵ میکروموس بر سانتی‌متر باشد، این آب مناسب آشامیدن می‌باشد. پیشنهاد می‌گردد نسبت به تصفیه و کاهش سختی و هدایت الکتریکی عصاره‌ی اشباع آب توجه جدی شود؛ همچنین بر انجام پایش-های مستمر در قالب نمونه‌برداری‌های دوره‌ای منظم از منابع آب‌های این دشت تأکید می‌گردد.

کلمات کلیدی: آب آشامیدنی، داده‌کاوی، دشت نورآباد، سختی کل، کیفیت آب، هدایت الکتریکی

مقدمه

آب در طبیعت به صورت خالص یافت نمی‌شود؛ بلکه همواره مقادیری املاح، مواد معلق و گازهای محلول را با خود دارد و این موجب می‌شود که آب در مناطق مختلف، ویژگی‌های مختلفی به خود بگیرد. وجود برخی از املاح در آب برای سلامتی انسان ضروری است و این در حالی است که مقدار بیش از حد مجاز آن‌ها سلامتی انسان را به خطر خواهد انداخت. بنابراین وجود آب آشامیدنی سالم، ضامن سلامتی جامعه است و اولین قدم در شناخت آب، بررسی پارامترهای آن است (شریعتی‌شیری‌نسب و همکاران، ۱۳۹۱). بایستی توجه داشت مهمترین مسائلی که در اثر مصرف آب‌های نامناسب ایجاد می‌شود شامل شور شدن ثانویه‌ی خاک‌ها، کاهش

www.SID.ir

Archive of SID

نفوذپذیری و سمیت املاح است؛ که هر کدام به نوعی بر سلامتی و یا رشد و تولید محصولات (کشاورزی، دامدارای، صنعتی و ...) صدمه وارد می‌نماید (سلیمان‌پور و همکاران، ۱۳۹۱).

یکی از راه‌های اثربخش و تأثیرگذار برای مبارزه با این معضل ملی، استفاده‌ی بهینه از منابع آبی کشور با تأکید بر اصول کیفیت آب می‌باشد، به طوری که امروزه بررسی‌های کیفی آب، دامنه‌ی گسترده‌ای پیدا کرده و مسائل مربوط به آلودگی آب‌های سطحی و زیرزمینی را نیز شامل می‌شود. این مبحث نه تنها در کشورهای صنعتی، بلکه در کشورهای در حال توسعه نیز مطرح می‌باشد.

با عنایت به مباحث فوق، باید اذعان نمود که امروزه بحث کیفیت آب در بسیاری از مناطق جهان به عنوان یکی از مباحث کلیدی مطرح است. زیرا این امر ارتباط بسیاری با سلامتی انسان و جامعه‌ی بشری دارد؛ به طوری که بررسی کیفیت آب (به ویژه آب آشامیدنی)، و چگونگی تغییر آن‌ها نقش بسیار مهمی در مدیریت و بهره‌برداری از منابع دارد. به علاوه با آگاهی از فاکتورهای مؤثر در تغییر کیفیت آب آشامیدنی می‌توان در جهت هر چه بهتر مدیریت کردن آن‌ها اقدامات مؤثری را به عمل آورد. بدین منظور پژوهش حاضر نسبت به کاربرد الگوریتم‌های داده‌کاوی K-Means و CART در تعیین مؤثرترین عوامل کیفیت آب آشامیدنی در دشت نورآباد واقع در غرب استان فارس اقدام نموده است.

امروزه افزایش سریع حجم پایگاه داده‌ها به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای قدرتمند میسر نمی‌باشد. در این وضعیت، تصمیم‌گیری‌ها به جای تکیه بر اطلاعات، بر درک مدیران و کاربران تکیه دارند. زیرا تصمیم‌گیرندگان ابزار قوی برای استخراج اطلاعات باارزش را در دست ندارند (شهرابی، ۱۳۹۲). در واقع شرایط فعلی توصیف کننده‌ی حالتی است که ما از لحاظ داده غنی، اما از لحاظ اطلاعات ضعیف هستیم. حال با توجه به شدت رقابت‌ها در عرصه‌های مختلف، استفاده‌ی مؤثر از داده‌ها توسط مدیران، یک هدف عمده برای بهبود وضعیت موجود محسوب می‌شود. بدین منظور داده‌کاوی[20]، به عنوان مجموعه‌ای از روش‌ها در فرآیند کشف دانش است که برای تشخیص الگوها و رابطه‌های نامعلوم در داده‌ها مورد استفاده قرار می‌گیرد (ناوی، ۱۳۸۷). از دلایل رشد و توسعه‌ی دانش داده‌کاوی در علوم مختلف، می‌توان به افزایش حجم پایگاه داده‌ها و عدم توانایی انسان برای درک و استخراج از آن‌ها اشاره کرد. بنابراین درک این داده‌ها بدون ابزارهای قدرتمند میسر نمی‌باشد؛ زیرا تصمیم‌گیران ابزار قوی برای استخراج اطلاعات با ارزش را در دست ندارند. در واقع شرایط فعلی توصیف کننده‌ی حالتی است که ما از لحاظ داده غنی، اما از لحاظ اطلاعات ضعیف هستیم. حال با توجه به شدت رقابت‌ها در عرصه‌های مختلف، استفاده‌ی مؤثر از داده‌ها توسط مدیران، یک هدف عمده برای بهبود وضعیت موجود محسوب می‌شود (شهرابی، ۱۳۹۲).

درخت تصمیم[21]، یکی از مشهورترین تکنیک‌های دسته‌بندی است که در فرآیند داده‌کاوی کاربرد دارد. شیوه‌ی نمایش درخت تصمیم به این صورت است که با ارائه نمودن یک درخت، روال دسته‌بندی را خلاصه‌سازی می‌کند. از درخت‌های تصمیم‌گیری به منظور پیش‌بینی کردن عضویت اشیاء به دسته‌های مختلف استفاده می‌شود. انعطاف‌پذیری این تکنیک باعث شده تا در میان روش‌های جذاب داده‌کاوی، بیشتر مورد استفاده قرار گیرد. متدولوژی مربوط به درخت تصمیم شامل دو فاز اصلی است:

الف- ساخت درخت اولیه: با استفاده از مجموعه داده‌های آموزشی، ساخت درخت تصمیم تا زمانی که هر برگ، خالص (همگن) شود ادامه می‌یابد.

ب- هرس کردن: در این فاز، به منظور افزایش دقت مدل، درختِ رشد یافته با توجه به مجموعه داده‌های آزمایشی هرس می‌شود (پاوت‌سیچوز، ۲۰۱۵).

آذرنگ و همکاران (۱۳۹۴)، در پژوهشی اقدام به ارزیابی کیفیت آب برای مصارف شرب و کشاورزی در رودخانه‌ی کرخه نمودند. دوره‌ی بررسی آماری در این تحقیق از سال ۱۳۸۱ تا ۱۳۹۰ است که اطلاعات کیفی آن از محل ۱۱ ایستگاه آب‌سنجی تهیه

[20] Data Mining
[21] Decision Tree

www.SID.ir

Archive of SID

شده بود. نتایج نشان داد در قسمت‌های بالایی رودخانه‌ی کرخه، کیفیت آب از نظر مصارف شرب و کشاورزی مناسب‌تر است و مناطق پایین‌دست دارای کیفیت نامناسب می‌باشند.

سلیمان‌پور و همکاران (۱۳۹۴)، در پژوهشی نسبت به پیاده‌سازی الگوریتم CART جهت تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی در دشت کازرون استان فارس اقدام نمودند. نتایج این تحقیق که از مدل‌سازی در نرم‌افزار Clementine، حاصل شده است، نشان داد که مؤثرترین فاکتورهای تأثیرگذار در کیفیت آب آشامیدنی در دشت کازرون عبارتند از: TDS و Ca. بدین ترتیب، در صورتی که مجموع املاح محلول در این دشت کمتر یا مساوی ۴۹۵ میلی‌گرم بر لیتر، و کلسیم آن، کمتر یا مساوی ۶/۱۵۰ میلی‌اکی-والان بر لیتر باشد، این آب برای آشامیدن مناسب می‌باشد.

غفاری‌نیا و همکاران (۱۳۹۴)، در پژوهشی با عنوان تجزیه و تحلیل منطقه‌ای کلر و سدیم کیفیت آب حوزه‌ی آبخیز گرگان-رود استان گلستان که با استفاده از خوشه‌بندی سلسله مراتبی انجام گرفت به این نتیجه رسیدند که همبستگی مناسبی بین پارامترهای کیفیت آب و خصوصیات حوضه وجود دارد.

محتشمی و ناصری (۱۳۹۴)، نسبت به طبقه‌بندی کیفی آب جهت مصارف شرب، کشاورزی و صنعت در دشت درمیان اسدآباد واقع در استان خراسان جنوبی اقدام نمودند. بدین منظور در ۲۱ نقطه از این دشت، نمونه‌گیری انجام و با نمودارهای ویلکوکس و شولر طبقه‌بندی‌های لازم انجام شد. نتایج نشان داد اکثر نمونه‌ها برای کشاورزی و شرب نامناسب می‌باشند.

ظریف و اسلامی (۲۰۱۴)، در پژوهشی به بررسی مقادیر کیفی EC و SAR جهت اهداف کشاورزی در دشت دزفول و اندیمشک پرداختند. ایشان با استفاده از آمار ۱۰۵ حلقه چاه، کلاس کیفی دشت را C3S1 تشخیص دادند و اقدام به تهیه‌ی نقشه‌ی نهایی تغییرات EC و SAR نمودند.

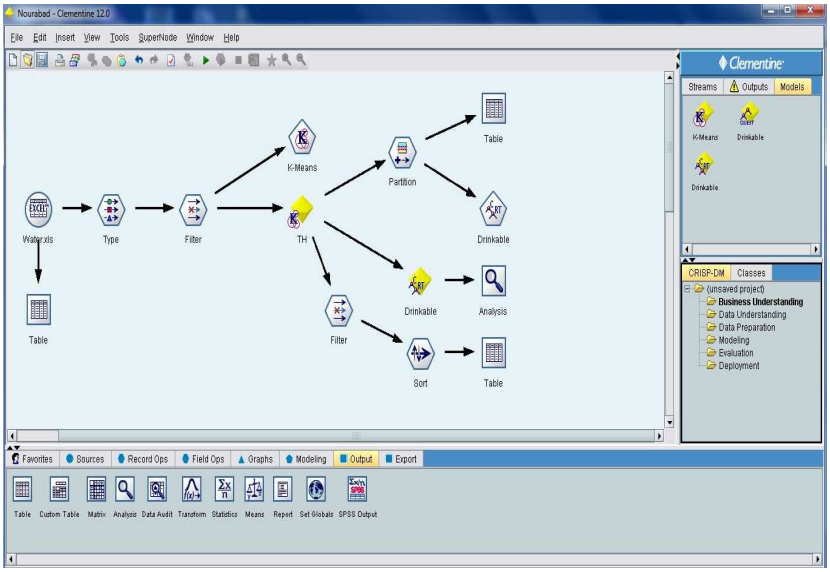
سلیمان‌پور و همکاران (۲۰۱۵)، در پژوهشی اقدام به تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی با استفاده از تکنیک-های داده‌کاوی QUEST در شهرستان سعادت‌شهر استان فارس نمودند. نتایج نشان داد مؤثرترین فاکتورهای کیفیت آب آشامیدنی در این منطقه تابع سختی کل و هدایت الکتریکی می‌باشد. ایشان کاهش سختی آب، و توجه به اقدامات تصفیه را به منظور بهبود وضعیت، پیشنهاد کردند.

مواد و روش‌ها

به منظور انجام این پژوهش، با منابعه به سازمان آب منطقه‌ای استان فارس، نسبت به تهیه‌ی آمار فاکتورهای کیفی ۶۰ حلقه چاه آب واقع در دشت نورآباد اقدام شد. آمارهای اخذ شده برای هر حلقه چاه، مشتمل بر فاکتورهای کیفی SAR، Na، Cl، SO4، TH، TDS، pH، NO3، CaCo3، HCO3، Ca، Mg، K، و EC، بود. به منظور تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی از تکنیک‌های داده‌کاوی خوشه‌بندی K-Means و درخت تصمیم CART استفاده شده است. مدل‌سازی در نرم‌افزار Clementine 12.0 انجام شده است. نرم‌افزار یاد شده ساخت شرکت SPSS است و امکان ایجاد مدل‌های متعددی را بر اساس تئوری‌های آماری، هوش مصنوعی و یادگیری ماشین ارائه می‌دهد. شکل ۱ نمایی از مدل‌سازی انجام شده در پژوهش حاضر را با استفاده از این نرم‌افزار نشان می‌دهد.

www.SID.ir

Archive of SID



شکل ۱- مدل‌سازی انجام شده با استفاده از نرم‌افزار SPSS Clementine 12.0

در مبحث داده‌کاوی، مهمترین موضوع دستیابی به داده‌هایی است که بتوان بر اساس آن‌ها به نتایج مفیدی دست یافت. در این پژوهش، مشخصات ۶۰ حلقه چاه آب موجود در منطقه به عنوان ورودی الگوریتم CART تعیین گردیدند. جدول ۱، فاکتورهای کیفی گروه‌بندی شده مناسب جهت شرب انسان بر مبنای طبقه‌بندی‌های ویلکوکس، شولر، و اوکین، و استاندارد کیفیت آب آشامیدنی ایران را نشان می‌دهد.

جدول ۱- فاکتورهای کیفی آب آشامیدنی

فاکتورهای کیفی و محدوده‌های مجاز	واحد
۰< CaCo3 <۵۰۰	میلی‌اکی‌والان بر لیتر
۰< HCO3 <۵۰۰	میلی‌اکی‌والان بر لیتر
۰< NO3 <۵۰	میلی‌اکی‌والان بر لیتر
۰< SO4 <۴۰۰	میلی‌اکی‌والان بر لیتر
۰< Ca <۳۰۰	میلی‌اکی‌والان بر لیتر
۰< Cl <۴۰۰	میلی‌اکی‌والان بر لیتر
۰< K <۱۲	میلی‌اکی‌والان بر لیتر
۰< Mg <۱۵۰	میلی‌اکی‌والان بر لیتر
۰< Na <۲۰۰	میلی‌اکی‌والان بر لیتر
۰< EC <۷۵۰	میکروموس بر سانتی‌متر
۰< TDS <۱۵۰۰	میلی‌گرم بر لیتر
۰< TH <۵۰۰	قسمت در میلیون
۶/۵< pH <۹	-
۰< SAR <۱۸	-

www.SID.ir

Archive of SID

خوشه‌بندی

هدف از خوشه‌بندی یافتن گروه‌هایی از داده‌ها است که به هم شباهت دارند، لذا با کشف این شباهت‌ها می‌توان رفتارها را بهتر شناسایی کرد و بر مبنای آن‌ها طوری عمل نمود که نتیجه‌ی بهتری حاصل شود؛ بدین منظور از الگوریتم‌های خوشه‌بندی K-Means که یکی از معروف‌ترین دستورالعمل‌های خوشه‌بندی است استفاده گردید. این الگوریتم تلاش می‌کند که بر اساس یک معیار فاصله، داده‌ها را به K خوشه تقسیم نماید. مراحل الگوریتم K-Means به شرح زیر است (تَن و همکاران، ۲۰۰۵):

الف- K مرکز خوشه‌ی اولیه، توسط کاربر تعیین می‌گردد.

ب- هر نمونه به نزدیک‌ترین مرکز خوشه اختصاص می‌یابد به طوری که مجموعه‌ای از نقاط اختصاص یافته به یک مرکز، یک خوشه محسوب می‌شود.

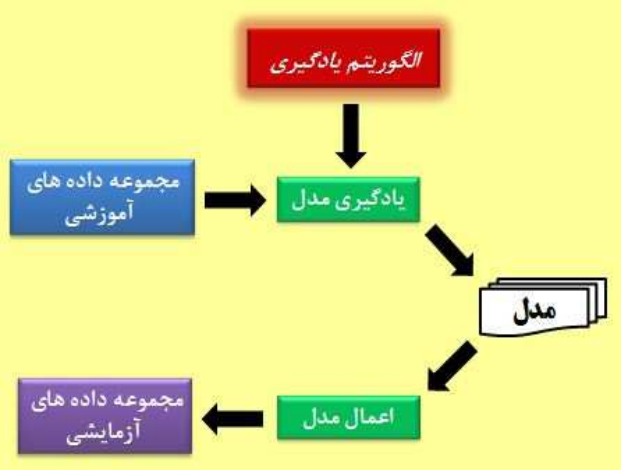
پ- مرکز هر خوشه بر اساس نمونه‌های اختصاص یافته به آن خوشه، به‌روزرسانی می‌شود.

تا زمانی که مراکز خوشه‌ها یکسان باقی بمانند، یا هیچ نمونه‌ای از خوشه‌ها تغییر نکند؛ گام‌های اختصاص و به‌روزرسانی (مراحل ب و پ) تکرار می‌شوند.

به منظور اختصاص یک نمونه به نزدیک‌ترین مرکز خوشه، به یک معیار نزدیکی نیاز است تا بتوان مفهوم نزدیک‌ترین را برای داده‌های مورد نظر تعیین کرد. این معیار نزدیکی برای الگوریتم K-Means، عموما فاصله‌ی اقلیدسی[22] است (غضنفری و همکاران، ۱۳۹۳).

دسته‌بندی

در الگوریتم‌های دسته‌بندی[23]، کل مجموعه‌ی داده‌ها به دو قسمت مجموعه داده‌های آموزشی، و مجموعه داده‌های آزمایشی تقسیم‌بندی می‌شوند (تَن و همکاران، ۲۰۰۵). همان‌طور که در شکل ۲ نشان داده می‌شود الگوریتم‌های دسته‌بندی شامل دو مرحله‌ی آموزش و آزمایش هستند. در مرحله‌ی آموزش، الگوریتم یادگیرنده بر اساس مجموعه داده‌های آموزشی، یک مدل را تولید می‌کند (صنیعی‌آباده و همکاران، ۱۳۹۱). مدل ساخته شده به الگوریتم یادگیرنده‌ی مورد استفاده بستگی دارد که در این پژوهش از الگوریتم درخت تصمیم CART استفاده شده است، بنابراین مدل ساخته شده، یک درخت تصمیم خواهد بود.



شکل ۲۳- مراحل مختلف فرآیند دسته‌بندی

[22] - Euclidean Distance
[23]- Classification

www.SID.ir

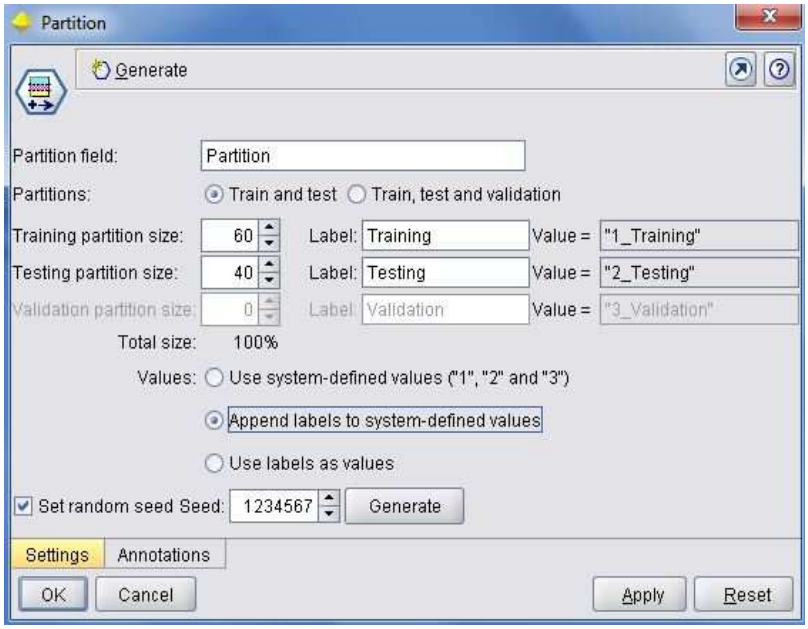
Archive of SID

درخت تصمیم و الگوریتم CART

الگوریتم درخت تصمیم CART یک رویه‌ی تقسیم دودویی بازگشتی است که قادر به پردازش متغیرهای پیشگو و هدف عددی و دسته‌ای می‌باشد. به منظور استفاده از الگوریتم CART، بایستی ابتدا یک فیلد هدف را در مجموعه‌ی داده‌ها مشخص کنیم. این الگوریتم کار خود را از گِرِه ریشه آغاز می‌کند، داده‌ها به دو گِرِه فرزند و سپس هر گِرِه فرزند به ترتیب به دو گِرِه نوه تقسیم می‌شوند. فرآیند رشد درخت تا هنگام رسیدن به درختی با اندازه‌ی ماکزیمم و تا زمانی که عملیات تقسیم به دلیل کمبود داده‌ها متوقف نشود، ادامه خواهد داشت (ناوی، ۱۳۸۷).

پس از ساخت درخت، عملیات هرس کردن درخت با اندازه‌ی ماکزیمم توسط یکی از روش‌های هرس کردن با شروع از برگ‌ها به سمت ریشه انجام می‌شود. مکانیسم CART قصد تولید تنها یک درخت را ندارد، بلکه تلاش می‌کند تا یک توالی از درخت‌های هرس شده‌ی تو در تو را ایجاد نماید به طوری که هر یک از آن‌ها کاندیدایی هستند که می‌توانند در نهایت به عنوان درخت بهینه انتخاب شوند. یک درخت خوب به وسیله‌ی ارزیابی کارایی آن بر روی داده‌های آزمایشی مستقل شناسایی می‌گردد.

همان‌طور که در شکل ۳ نشان داده شده است، به منظور ایجاد مدل، داده‌های این تحقیق به طور تصادفی به دو بخش مجموعه داده‌های آموزشی و آزمایشی با تناسب به ترتیب ۶۰ و ۴۰ درصد تقسیم شدند؛ و پس از اجرای مدل، با استفاده از درخت تصمیم‌گیری CART، مجموعه‌ای از مؤثرترین قوانین کشف شده از نظر نرم‌افزار ارائه گردید.



شکل ۳- تقسیم مجموعه داده‌ها به دو مجموعه داده‌های آموزشی و آزمایشی

نتایج

نتایج خوشه‌بندی داده‌ها بر اساس پارامترهای کیفی با استفاده از الگوریتم K-Means در جدول۲ نشان داده شده است.

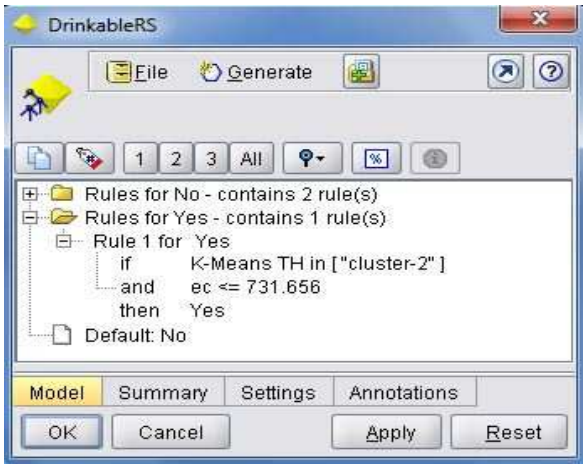
www.SID.ir

Archive of SID

جدول ۲- نتایج خوشه‌بندی الگوریتم K-Means در دشت نورآباد استان فارس

شماره‌ی خوشه	تعداد	کمینه	میانگین	بیشینه
خوشه‌ی ۱	۴۳	۴۵۰	۵۸۱/۰۲	۶۰۰
خوشه‌ی ۲	۱۷	۱۷۵	۳۰۰/۰۸	۴۴۰

با اختصاص نتایج فوق و سایر پارامترها به ورودی الگوریتم درخت تصمیم CART مجموعه‌ای از قوانین توسط نرم‌افزار شناسایی و نمایش داده خواهد شد که مؤثرترین قاعده‌ی آن در شکل ۴ نشان داده شده است.

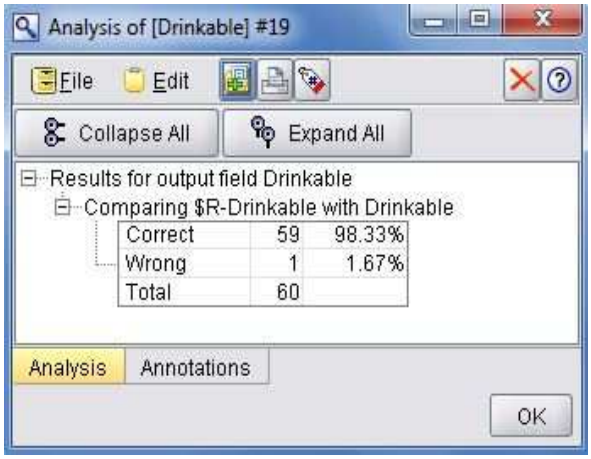


شکل ۴- قانون استخراج شده جهت تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی

مهمترین فاکتورهای تأثیرگذار در کیفیت آب آشامیدنی در دشت نورآباد (مطابق شکل ۴) عبارتند از: TH و EC. بنابراین در صورتی که سختی کل (TH) در این دشت در خوشه‌ی شماره ۲ واقع شده باشد (میانگین خوشه‌ی دوم ۳۰۰/۰۸ قسمت در میلیون است) و با توجه به نتایج مندرج در جدول ۲ و شکل ۴، در شرایطی که سختی کل (TH) در این دشت بین ۱۷۵ و ۴۴۰ قسمت در میلیون، و هدایت الکتریکی عصاره‌ی اشباع (EC) آن، کمتر یا مساوی ۷۳۱/۶۵ میکروموس بر سانتی‌متر باشد، این آب مناسب آشامیدن می‌باشد.

www.SID.ir

Archive of SID



شکل ۵- نتیجه‌ی ارزیابی انجام شده از الگوریتم QUEST

با توجه به شکل ۵ مشخص گردید دقت تخمین زده شده بر اساس مؤثرترین فاکتورهای کیفیت آب آشامیدنی در مدل نهایی، برابر با ۹۸/۳۳ درصد می‌باشد. با توجه به دو عامل سختی کل (TH) و هدایت الکتریکی عصاره‌ی اشباع (EC)، به نظر می‌رسد این دو فاکتور تأثیر بیشتری بر کیفیت آب آشامیدنی در این دشت دارند؛ بدین منظور، پیشنهاد می‌گردد نسبت به تصفیه و کاهش سختی و هدایت الکتریکی عصاره‌ی اشباع آب توجه جدی شود؛ همچنین بر انجام پایش‌های مستمر در قالب نمونه‌برداری‌های دوره‌ای منظم از منابع آب‌های این دشت تأکید می‌گردد.

منابع

آذرنگ، ف.، تلوری، ع.، صدقی، ح.، و شفاعی‌بجستان، م.، (۱۳۹۴)، ارزیابی کیفیت آب برای مصارف شرب و کشاورزی (مطالعه‌ی موردی: رودخانه‌ی کرخه پایین‌دست سد مخزنی)، مجموعه مقالات اولین همایش ملی کیفیت منابع آب و توسعه‌ی پایدار، اراک.

سلیمان‌پور، س. م.، شریعتی‌شیری‌نسب، آ.، و جوکار، ل.، (۱۳۹۱)، بررسی کیفیت منابع آب زیرزمینی و روند تغییرات عناصر کیفی آن (مطالعه‌ی موردی: دشت خسویه ساچون در منطقه‌ی زرین‌دشت استان فارس)، مجموعه مقالات اولین همایش ملی بیابان، کرج.

سلیمان‌پور، س.م.، مصباح، س.ح.، و هدایتی، ب.، (۱۳۹۴)، پیاده‌سازی الگوریتم CART جهت تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی (مطالعه‌ی موردی: دشت کازرون استان فارس)، مجموعه مقالات دومین همایش ملی صیانت از منابع طبیعی و محیط زیست، اردبیل.

شریعتی‌شیری‌نسب، آ.، سلیمان‌پور، س. م.، و جوکار، ل.، (۱۳۹۱)، تعیین روابط بین فاکتورهای کیفی آب با استفاده از روش گام به گام در منطقه‌ی زرین‌دشت استان فارس، مجموعه مقالات اولین همایش ملی بیابان، کرج.

شهرابی، ج.، (۱۳۹۲)، داده‌کاوی ۲، انتشارات جهاد دانشگاهی تهران، واحد صنعتی امیرکبیر، ۲۹۹ صفحه.

صنیعی‌آباده، م.، محمودی، س.، و طاهرپور، م.، (۱۳۹۱)، داده‌کاوی کاربردی، انتشارات نیاز دانش، ۵۲۰ صفحه.

غضنفری، م.، علیزاده، س.، و تیمورپور، ب.، (۱۳۹۳)، داده‌کاوی و کشف دانش، انتشارات دانشگاه علم و صنعت ایران، ۴۰۳ صفحه.

غفاری‌نیا، م.، نجفی‌نژاد، ع.، و سعدالدین، ا.، (۱۳۹۴)، تجزیه و تحلیل منطقه‌ای کلر و سدیم کیفیت آب حوزه‌ی آبخیز گرگان‌رود استان گلستان، مجموعه مقالات اولین همایش ملی کیفیت منابع آب و توسعه‌ی پایدار، اراک.

محتشمی، ع.، و ناصری، م.، (۱۳۹۴). طبقه‌بندی کیفی آب جهت مصارف شرب و کشاورزی و صنعت (مطالعه‌ی موردی: دشت درمیان اسدآباد استان خراسان جنوبی) ، مجموعه مقالات اولین همایش ملی کیفیت منابع آب و توسعه‌ی پایدار، اراک.

www.SID.ir

Archive of SID

ناوی. م.، (۱۳۸۷)، شناسایی مؤلفه‌های تشکیل‌دهنده بار با استفاده از تکنیک‌های داده‌کاوی، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس، ۱۱۶ صفحه.

Pawet Cichosz, R., (2015), Data mining algorithms explained using, Wiley press, 684 pp.

Soleimanpour, S.M., Hedayati, B., Zolfaghari, M. and Zare, M., (2015), Determination of Effective Factors of Drinking Water Quality by using the QUEST Data Mining Technique in Saadatsharh-Fars Province, 3rd International Conference on Rain Catchment System, Birjand, I.R.Iran.

Tan, P.N., Steinbach, M. and Kumar, V., (2005), Introduction to Data Mining, Addison-Wesley, 134pp.

Zarif, F. and Eslami, H., (2014), Spatial assessment of SAR and EC of groundwater for agricultural purposes (case study: Dezfoul-Andimeshk plain), Bulletin of Environment Pharmacology and Life Sciences, 3(11): 141-145.

www.SID.ir