

## آب، هدایت الکتریکی

## مقدمه

آب در طبیعت به صورت خالص یافت نمی‌گردد، 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 [۲۴]، به طوری که هر ساله ۸۴۲ هزار نفر به علت در دسترس نبودن آب سالم و مصرف آب‌های آلوده جان خود را از دست می‌دهند [۷ و ۱۳]، بنابراین وجود آب آشامیدنی سالم، ضامن سلامتی جامعه است و اولین قدم در شناخت آب، بررسی پارامترهای آن است.

بایستی توجه داشت مهمترین مسائلی که در اثر مصرف آب‌های نامناسب ایجاد می‌شود شامل شور شدن ثانویه خاک‌ها، کاهش نفوذپذیری و سمیت املاح است، که هر کدام به نوعی بر سلامتی و یا رشد و تولید محصولات (کشاورزی، دامداری، صنعتی و ...) صدمه وارد می‌نماید [۲۵]. اخیراً با توسعه‌ی محیط‌زیست و تکنیک‌های جدید، مطالعات و بررسی‌های کیفیت آب شکل نوینی پیدا کرده است [۱۹ و ۳۲]. به طور کلی این باور پذیرفته شده است که خصوصیات فیزیکی و شیمیایی محیط‌های آبی می‌تواند منعکس کننده‌ی واقعی وضعیت سلامت یک اکوسیستم باشد [۳۳]. مدیریت کیفیت آب جهت اخذ استراتژی‌های مدیریتی منابع آب در حوزه‌ی آبخیز ضروری می‌باشد [۲] و دسترسی به آب ناسالم و آلوده باعث بیماری‌ها و مشکلات عدیده‌ای می‌شود که آسیب‌های اجتماعی، اقتصادی و ... را به دنبال خواهد داشت؛ در این راستا آگاهی از کیفیت منابع آب یکی از مهمترین نیازهای اساسی در استفاده از آب در زمینه‌های کشاورزی، صنعت، شرب می‌باشد [۱ و ۱۷]. جهت نیل به این هدف پایش و تحقیقات علمی مناسب، و استفاده از روش‌های نوین و کاربردی، با تأکید بر داشتن اطلاعات جامع، صحیح و قابل اطمینان با دوره‌های زمانی مناسب می‌تواند نقش بسیار مهمی در تصمیم‌گیری و سیاست‌گذاری‌های صحیح در این خصوص ایفا نماید [۲۹]. کشور ایران، کشوری پهناور است که از نظر منابع آبی محدود می‌باشد. به همین جهت از زمان‌های دور، آب در کشور ما از ارزش و اهمیت والایی برخوردار بوده است. شرایط اقلیمی، زمین‌شناسی و آب‌شناسی کشور ایران نیز طوری است که

 تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی با  
 استفاده از روش داده‌کاوی QUEST  
 (مطالعه‌ی موردی: شهرستان پاسارگاد استان فارس)

سید مسعود سلیمان‌پور<sup>۱\*</sup>، سید حمید مصباح<sup>۲</sup> و بهرام هدایتی<sup>۳</sup>  
 تاریخ دریافت: ۱۳۹۶/۰۲/۱۶ تاریخ پذیرش: ۱۳۹۷/۰۱/۲۱

## 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 نقش بسیار مهمی در مدیریت و بهره‌برداری از منابع دارد. به این منظور پژوهش حاضر برای اولین بار نسبت به تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی با استفاده از روش داده‌کاوی در شهرستان پاسارگاد واقع در ۱۰۵ کیلومتری شمال شرقی شیراز اقدام نموده است. نتایج این تحقیق که از مدل‌سازی با بهره‌گیری از درخت تصمیم QUEST در نرم‌افزار Clementine (نسخه ۱۲)، حاصل شده است، نشان داد که مؤثرترین فاکتورهای کیفیت آب آشامیدنی در این منطقه، تابع سختی کل (TH) و هدایت الکتریکی (EC)، می‌باشد. بدین ترتیب، در صورتی که سختی کل (TH) در این شهرستان کمتر از ۲۸۲/۲۳۲ قسمت در میلیون، و هدایت الکتریکی (EC) آن، کمتر از ۸۲۲/۷۸۷ میکروموس بر سانتی‌متر باشد، این آب مناسب آشامیدن می‌باشد. بنابراین توصیه می‌شود به اقدامات تصفیه و کاهش سختی آب جهت مصارف شرب (انسان) توجه گردد و انجام پایش‌های مستمر در قالب نمونه‌برداری‌های دوره‌ای منظم از منابع آب در این شهرستان در دستور کار قرار گیرد.

## واژه‌های کلیدی: آب آشامیدنی، داده‌کاوی، سختی کل، کیفیت

۱- استادیار بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس، سازمان تحقیقات، آموزش و ترویج کشاورزی، شیراز، ایران، نوسنده مسئول،  
 Email: m.soleimanpour@yahoo.com

۲- مربی پژوهشی بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس، سازمان تحقیقات، آموزش و ترویج کشاورزی، شیراز، ایران

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

بهره‌برداری از آب با کیفیت را با مسایل و مشکلات خاص خود همراه می‌کند. هر انسان در روز حداقل به ۲/۵ لیتر آب شیرین سالم و با کیفیت استاندارد برای آشامیدن نیاز دارد تا بتواند احتیاجات خود را بر طرف کند. علاوه بر آن، برای هر نفر در کشور ایران به طور متوسط حداقل روزانه به ۱۱۰ لیتر آب برای مصارف شخصی و خانگی نیاز می‌باشد که در واقع همان سرانه‌ی مصرف آب در شهرهای کشور است [۱۴]. یکی از راه‌های اثر بخش و تأثیرگذار در این حوضه، استفاده‌ی بهینه از منابع آبی کشور با تأکید بر اصول کیفیت آب می‌باشد. 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 [۲۵].

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 [۲۶].

امروزه افزایش سریع حجم پایگاه داده‌ها به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای قدرتمند میسر نیست [۲۲]. در این وضعیت، تصمیم‌گیری‌ها به جای تکیه بر اطلاعات، بر درک مدیران و کاربران تکیه دارند؛ زیرا تصمیم‌گیرندگان ابزار قوی برای استخراج اطلاعات با ارزش را در دست ندارند [۲۳]. در واقع شرایط فعلی توصیف‌کننده‌ی 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 [۲۱]. داده‌کاوی فرآیندی است که در آغاز دهه‌ی ۱۹۹۰ پا به عرصه‌ی ظهور گذاشته و با نگرشی نو، به مسأله‌ی استخراج اطلاعات از پایگاه داده‌ها می‌پردازد [۲۳]. در حال حاضر، داده‌کاوی مهمترین فن‌آوری جهت بهره‌برداری از داده‌های حجیم است و اهمیت آن رو به فزونی است. به بیان دیگر، این فرآیند، مجموعه‌ای از روش‌ها در فرآیند کشف دانش است که برای تشخیص الگوها و رابطه‌های نامعلوم در داده‌ها مورد استفاده قرار می‌گیرد [۲۱]. این دانش، یکی از ده دانش در حال توسعه است که دهه‌ی آینده را با انقلاب تکنولوژیک مواجه خواهد ساخت و بدین رو در سال‌های اخیر در دنیا گسترش فوق‌العاده سریعی داشته است [۲۳]. دانش داده‌کاوی فرآیند کشف دانش پنهان درون داده‌ها است که با برخورداری از دامنه‌ی وسیع زیرزمینه‌های تخصصی با توصیف، تشریح، پیش‌بینی و کنترل پدیده‌های گوناگون پیرامونی، امروزه دارای کاربرد بسیار وسیع در حوزه‌های مختلف از جمله صنعتی، پزشکی، ارتباطات، کشاورزی، انرژی، علوم اجتماعی،

فرهنگی، سیاسی، اقتصادی، بازرگانی، نظامی، آموزشی و ... دارد؛ به گونه‌ای که امروزه مرز و محدودیتی برای کاربرد این دانش در نظر گرفته نشده و زمینه‌های کاری این دانش را از ذرات کف اقیانوس تا اعماق فضا می‌دانند [۲۳]. به این منظور پژوهش حاضر برای اولین بار نسبت به تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی با استفاده از روش‌های داده‌کاوی در شهرستان پاسارگاد استان فارس اقدام نموده است.

سلیمان‌پور و همکاران [۲۵] در پژوهشی به بررسی کیفیت منابع آب زیرزمینی و روند تغییرات عناصر کیفی آن طی مدت ۶ سال در دشت خسویه‌ی ساچون استان فارس پرداختند. نتایج نشان از روند کاهشی pH و روند افزایشی  $SO_4$ ,  $Ca$ ,  $HCO_3$ ,  $Na$ ,  $TH$  و در نتیجه پایین آمدن کیفیت آب منطقه در مدت زمان ۶ ساله‌ی این پژوهش داشت. همچنین این منطقه طبق طبقه‌بندی Wilcox در کلاس  $C_4S_3$  قرار گرفت که حتی از نظر آبیاری نیز نامناسب می‌باشد. ابراهیم‌پور و محمدرزاده [۸] در تحقیقی به ارزیابی و پهنه‌بندی کیفیت آب دریاچه‌ی زریوار با استفاده از شاخص‌های کیفی  $OWQI$ ,  $CWQI$  و  $NSFWQI$  پرداختند. نتایج این پژوهش نشان داد کیفیت آب دریاچه‌ی زریوار از نظر شاخص  $OWQI$  در محدوده‌ی کیفی متوسط و از نظر شاخص  $NSFWQI$  به طور کلی بد، جهت آشامیدن خوب، برای آبیاری بد و به منظور تفریح، آبیاری و استفاده‌ی احشام، عالی می‌باشد. آذرنگ و همکاران [۴]، 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 نتایج اطلاعات کیفی آن از محل ۱۱ ایستگاه آب‌سنجی تهیه 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 [۲۶]، در پژوهشی 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 شد نشان داد مؤثرترین فاکتورهای تأثیرگذار در کیفیت آب آشامیدنی در دشت کازرون عبارتند از:  $TDS$  و  $Ca$  بدین ترتیب، در صورتی که مجموع املاح محلول ( $TDS$ ) در این دشت کمتر یا مساوی ۴۹۵ میلی‌گرم بر لیتر، و کلسیم ( $Ca$ ) آن، کمتر یا مساوی ۶/۱۵۰ میلی‌اکی‌والان بر لیتر باشد، این آب برای آشامیدن مناسب می‌باشد. غفاری و همکاران [۱۰]، در پژوهشی با عنوان تجزیه و تحلیل منطقه‌ای کلر و سدیم کیفیت آب حوزه‌ی آبخیز گرگان‌رود استان گلستان که با استفاده از خوشه‌بندی سلسله‌مراتبی انجام گرفت به این نتیجه رسیدند که همبستگی مناسبی بین پارامترهای کیفیت آب و خصوصیات حوضه وجود دارد. محتشمی و ناصری [۲۰]، نسبت به طبقه‌بندی کیفی آب جهت مصارف شرب، کشاورزی و صنعت در دشت درمیان اسداباد 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. سلیمان‌پور و همکاران [۲۷]، 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 کردند. 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، نشان داد که مؤثرترین فاکتورهای مطلوب آب آشامیدنی در این منطقه، تابع سختی کل و هدایت الکتریکی عصاره‌ی اشباع می‌باشد. به این 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. سلیمان‌پور و همکاران [۲۸]، در پژوهشی دیگر اقدام به کاربرد تکنیک داده‌کاوی درخت تصمیم در تعیین مؤثرترین فاکتورهای نامطلوب آب آشامیدنی در دشت کازرون واقع در غرب استان فارس اقدام نموده است. نتایج این تحقیق با دقت تخمین زده شده‌ی ۹۰ درصد، با استفاده از مدل‌سازی با بهره‌گیری از تکنیک داده‌کاوی درخت تصمیم CART در نرم‌افزار Clementine حاصل شد. نتایج نشان داد دو عامل مجموع املاح محلول و کلسیم، تأثیر بیشتری بر عدم کیفیت آب آشامیدنی در این دشت دارند. به این ترتیب، در صورتی که مجموع املاح محلول در این دشت کمتر یا مساوی ۴۹۵ میلی‌گرم بر لیتر، و کلسیم آن، بیشتر از ۶/۱۵۰ میلی‌اکی‌والان بر لیتر باشد، این آب برای آشامیدن مناسب نمی‌باشد. چو [۶] در پژوهشی به ارزیابی مدل کیفیت آب در حوزه‌های آبخیز با استفاده از داده‌کاوی پرداخت و به این نتیجه رسید که مدل ارزیابی کیفیت آب بر اساس داده‌کاوی، جایگزین مناسبی برای مدل‌های فیزیکی در حوزه‌های آبخیز است. فو چنگ و زو ژائو [۹] از روش‌های داده‌کاوی به منظور طبقه‌بندی و ارزیابی کیفیت آب‌های سطحی روستایی ۳۳ ایستگاه در شهر لیان یانگانگ استفاده و مناطق آلوده به آمونیاک، نیتروژن را تعیین کردند. ایشان کاربرد این روش‌ها را در شناسایی مناطق آلوده و ارزیابی کیفی منابع آب، مطلوب ارزیابی نمودند. آزار و همکاران [۵] هم از منطق فازی و داده‌کاوی به صورت توأم در بررسی استانداردهای کیفیت آب استفاده کردند و اعلام داشتند به منظور نظارت بر کیفیت آب، بهره‌گیری از این روش‌ها مناسب است. به علاوه گو و همکاران [۱۱] در پژوهشی به شناسایی و ارزیابی عوامل تأثیرگذار بر کیفیت آب در ۷۳ مخزن آب آشامیدنی در استان ژجیانگ چین اقدام کردند. نتایج این تحقیق که با بهره‌گیری از روش‌های داده‌کاوی به دست آمد بیانگر ۸۱ درصد دقت در تخمین کیفیت آب بود. با توجه به نتایج مطلوب، ایشان استفاده از این روش را در مطالعه‌ی کیفیت آب آشامیدنی سایر مناطق پیشنهاد و بیان کردند که این روش می‌تواند به طور بالقوه به عنوان ابزار عملیاتی برای برنامه‌ریزان و مدیران عمل نماید.

همچنین لی و لی [۱۵] در تحقیقی پتانسیل بهره‌وری آب زیرزمینی را با استفاده از تکنیک داده‌کاوی و سیستم اطلاعات جغرافیایی در شهرهای بوریونگ و پوهانگ کره، مورد بررسی قرار دادند. نتایج ایشان نیز مؤید آن است که روش‌های داده‌کاوی می‌تواند برای مطالعه و توسعه‌ی منابع آب زیرزمینی مفید باشند. تامسون [۳۱] نیز در پژوهشی با استفاده از تکنیک‌های داده‌کاوی به بررسی کیفیت آب آشامیدنی در کانادا پرداخت و اعلام نمود نتایج حاصل از روش‌های داده‌کاوی می‌تواند در اطلاع‌رسانی، و تصمیم‌گیری‌های آینده مفید واقع گردد.

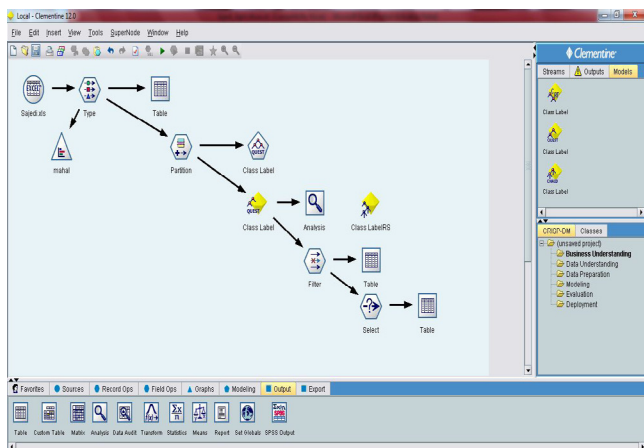
با عنایت به مباحث فوق، باید اذعان نمود که امروزه بحث کیفیت آب در بسیاری از مناطق جهان به عنوان یکی از مباحث کلیدی مطرح است. زیرا این امر ارتباط بسیاری با سلامتی انسان و جامعه‌ی بشری دارد. به طوری که بررسی کیفیت آب (به ویژه آب آشامیدنی)، و چگونگی تغییر آن‌ها نقش بسیار مهمی در مدیریت و بهره‌برداری از منابع 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. به این منظور پژوهش حاضر نسبت به کاربرد تکنیک داده‌کاوی درخت تصمیم QUEST در تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی در شهرستان پاسارگاد استان فارس اقدام نموده است.

## مواد و روش‌ها

### معرفی منطقه‌ی مورد مطالعه

شهرستان پاسارگاد در ۱۰۵ کیلومتری شمال‌شرقی شیراز، و در موقعیت جغرافیایی ۵۳ درجه و ۱۳ دقیقه و ۳۳ ثانیه‌ی طول شرقی و ۳۰ درجه و ۰۶ دقیقه و ۶۰ ثانیه‌ی عرض شمالی قرار دارد (شکل ۱). ارتفاع شهرستان از سطح آب‌های آزاد حدود ۱۷۰۰ متر است. این شهرستان از شمال به شهرستان خرم‌بید و آباد، از غرب به شهرستان اقلید و مرودشت، از جنوب به شهرستان مرودشت، و از شرق به شهرستان ارسنجان مشرف می‌باشد. طبق سرشماری سال ۱۳۸۵، جمعیت این شهر ۱۶۹۳۲ نفر اعلام شده است. از نظر اقلیمی دارای آب و هوای معتدل مدیترانه‌ای با زمستان‌های نسبتاً سرد و تابستان‌های معتدل می‌باشد. میانگین بارش منطقه حدود ۴۰۰ میلی‌متر در سال است که عمدتاً به صورت برف در ارتفاعات، و باران در دشت‌ها می‌باشد. بر اساس نقشه‌های زمین‌شناسی، سازندهای رخنمون شده در منطقه‌ی مورد مطالعه از قدیم به جدید عبارتند از: فهلیان- سورمه، گدون، داریان، کژدمی، سروک، و رسوبات کواترنری. سازند سورمه و فهلیان، تحت عنوان یک واحد به نام فهلیان- سورمه مشخص شده‌اند و از نظر سنگ‌شناسی شامل آهک، آهک‌های مارنی و رسی به رنگ‌های خاکستری تا قهوه‌ای، دولومیت، و آهک‌های دولومیتی به رنگ خاکستری تیره بوده و ضخامت لایه‌های آن از حد متوسط تا بسیار ضخیم متغیر است. سازند گدون، شامل آهک‌های رسی متوسط تا ضخیم لایه، و شیل و مارن‌های نازک تا متوسط لایه،

و مانند یک دستگاه عمل می‌کند؛ یعنی مجموعه‌ای از ورودی‌ها را می‌پذیرد و یک یا چند خروجی تولید می‌نماید.



شکل ۲: 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 تحقیق، از بین مشخصات ۱۵۰ حلقه چاه شرب، در نهایت تعداد ۴۲ حلقه چاه (شکل ۱) بر اساس فاکتورهای کیفی آب آشامیدنی موجود در جدول ۱، مناسب برای آشامیدن در نظر گرفته شدند و به عنوان ورودی الگوریتم QUEST تعیین گردیدند. جدول ۱، فاکتورهای کیفی گروه‌بندی شده مناسب جهت شرب انسان بر مبنای طبقه‌بندی ویلکوکس، شورلر، و اوکین را نشان می‌دهد.

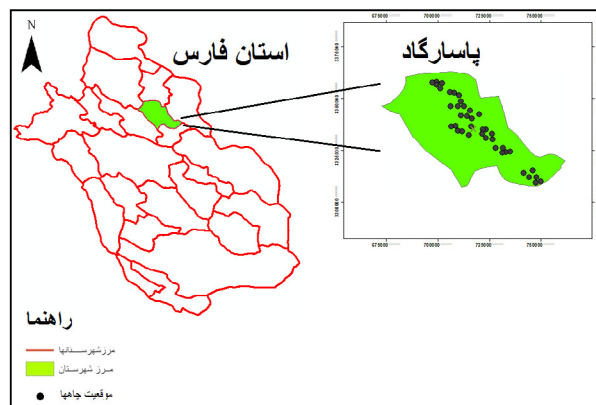
جدول ۱: فاکتورهای کیفی آب آشامیدنی

واحد	فاکتورهای کیفی و محدوده‌های مجاز
میلی‌اکی‌والان بر لیتر	$0 < Na < 230$
میلی‌اکی‌والان بر لیتر	$0 < Cl < 350$
میلی‌اکی‌والان بر لیتر	$0 < SO_4 < 280$
قسمت در میلیون	$0 < TH < 500$
میلی‌گرم بر لیتر	$0 < TDS < 1000$
میکروموس بر سانتی‌متر	$0 < EC < 750$
-	$0 < SAR < 18$
-	$5/6 < pH < 9$

در الگوریتم‌های دسته‌بندی<sup>۶</sup>، کل مجموعه‌ی داده‌ها به دو قسمت مجموعه داده‌های آموزشی، و مجموعه داده‌های آزمایشی تقسیم‌بندی می‌شوند [۳۰]. همان‌طور که در شکل ۳ نشان داده می‌شود الگوریتم‌های دسته‌بندی شامل دو مرحله‌ی آموزش و آزمایش هستند. در مرحله‌ی آموزش، الگوریتم یادگیرنده بر اساس

6- Classification

و سازند داریان شامل سنگ‌های آهکی متوسط تا ضخیم لایه به رنگ‌های خاکستری روشن تا تیره می‌باشد. سازند کژدمی نیز شامل شیل و مارن‌های نازک تا متوسط لایه به رنگ سبز، زرد تا خاکستری و آهک‌های رسی متوسط تا ضخیم لایه به رنگ خاکستری تا سیاه می‌باشد؛ و سازند سروک، شامل آهک رسی تیره با میان لایه‌های مارنی، آهک‌های کرم رنگ و خاکستری است. لایه‌های مارنی نازک لایه و آهک‌های متوسط تا بسیار ضخیم لایه می‌باشد [۱۲].



شکل ۱: موقعیت جغرافیایی شهرستان پاسارگاد و چاه‌های مورد مطالعه

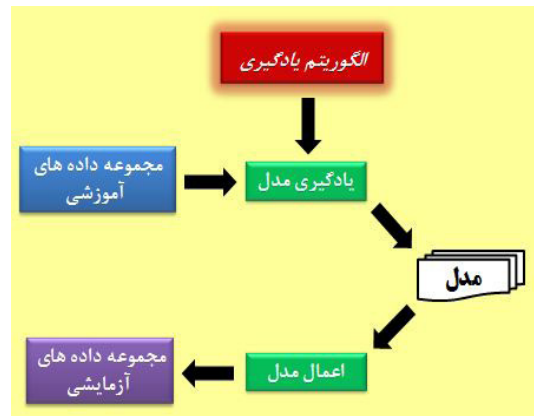
#### روش تحقیق

جهت انجام این پژوهش، با مراجعه به سازمان آب منطقه‌ای استان فارس (سال ۱۳۹۶)، نسبت به تهیه‌ی آمار فاکتورهای کیفی ۱۵۰ حلقه چاه شرب واقع در شهرستان پاسارگاد اقدام شد. آمارهای اخذ شده برای هر حلقه چاه، مشتمل بر فاکتورهای کیفی  $Na$ ،  $SAR$ ،  $TH$ ،  $SO_4$ ،  $Cl$ ،  $EC$  و  $pH$  بود.

در این پژوهش به منظور تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی از تکنیک داده‌کاوی درخت تصمیم<sup>۱</sup> QUEST یا درخت آماری سریع، بی‌طرف و کارآمد<sup>۲</sup> 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<sup>۳</sup> ارائه می‌دهد. شکل ۲ نمایی از مدل‌سازی انجام شده در پژوهش حاضر را با استفاده از این نرم‌افزار نشان می‌دهد. در این نرم‌افزار، کلیه‌ی فرآیندها به صورت جریان<sup>۴</sup> طراحی می‌شوند. یک جریان شامل مجموعه‌ای از گره‌ها<sup>۵</sup> است که به ترتیب به یکدیگر متصل می‌شوند؛ به گونه‌ای که خروجی هر گره، ورودی گره بعدی باشد تا در نهایت هدفی را برآورده سازند. هر گره در جریان، وظیفه‌ای بر عهده دارد

- 1- Decision Tree
- 2- Quick Unbiased and Efficient Statistical Tree
- 3- Machine Learning
- 4- Stream
- 5- Nodes

مجموعه داده‌های آموزشی، یک مدل را تولید می‌کند [۲۲]. مدل ساخته شده به الگوریتم یادگیرنده مورد استفاده بستگی دارد که در این پژوهش از الگوریتم درخت تصمیم QUEST استفاده شده است، بنابراین مدل ساخته شده، یک درخت تصمیم خواهد بود.



شکل ۳: مراحل مختلف فرآیند دسته‌بندی

مزیت اصلی الگوریتم QUEST نسبت به سایر روش‌های درخت تصمیم، بی‌طرف بودن در انتخاب متغیر تفکیک‌کننده است. این در حالی است که الگوریتمی مثل درخت تصمیم CART تمایل به انتخاب متغیرهایی دارد که باعث تفکیک‌پذیری بیشتری می‌شوند. بی‌طرف بودن در انتخاب متغیر جداکننده، سبب می‌شود که متغیرهای تعیین‌کننده مرتبط با هدف که در ابتدا جداکنندگی کمتری دارند، از قرار گرفتن در درخت تصمیم و در نهایت از حضور در قواعد پایانی باز نمانند. به علاوه، الگوریتم QUEST ضمن حفظ دقت نتایج از سرعت بالاتری در ساخت درخت تصمیم نسبت به سایر روش‌های ساخت درخت تصمیم دودویی برخوردار می‌باشد.

فرآیند رشد درخت تصمیم QUEST شامل انتخاب یک متغیر پیشگوی<sup>۱</sup> جدا کننده، انتخاب یک نقطه تفکیک<sup>۲</sup> جهت متغیر پیشگوی انتخاب شده، و توقف می‌باشد. در این الگوریتم فقط تقسیم‌های تک متغیره در نظر گرفته می‌شوند [۱۶].

به طور کلی درخت تصمیم روشی برای نمایش یک سری از قوانین است که منتهی به یک رده یا مقدار می‌شوند. در درخت تصمیم‌گیری تعدادی پرسش وجود دارد و با مشخص شدن پاسخ هر پرسش، سؤالی دیگر طرح می‌شود. اگر سؤال‌ها درست و سنجیده پرسیده شوند، تعدادی کمی از پرسش‌ها برای پیش‌بینی دسته رکورد جدید کافی خواهد بود. عملکرد درخت تصمیم به این صورت است که یک گره ریشه در بالای آن کشیده شده و برگ‌های آن در پایین قرار دارند. یک رکورد در گره ریشه وارد می‌شود و در این گره یک تست (آزمون) صورت می‌گیرد تا روشن شود که این رکورد به کدام یک از گره‌های فرزند (شاخه‌های پایین‌تر) خواهد رفت.

معمولاً روش‌های متفاوتی برای انتخاب این آزمون اولیه وجود دارند ولی هدف همه آن‌ها گزینش روشی است که بهترین جداسازی را در کلاس‌های هدف انجام دهد. این فرآیند آن قدر ادامه پیدا می‌کند تا رکورد به گره برگ برسد [۳].

پس از ساخت درخت، باید اثربخشی درخت ایجاد شده اندازه‌گیری شود. برای این کار، از مجموعه‌ای از رکوردها به نام داده‌های آزمایشی استفاده می‌شود. این مجموعه با داده‌های اولیه‌ای که درخت را ایجاد کرده‌اند، متفاوت است. معیاری که در این بخش استفاده می‌شود عبارت است از: میزان داده‌هایی که به طور صحیح دسته‌بندی شده‌اند به کل داده‌ها. هر مسیر ایجاد شده از ریشه به برگ، معادل یک قانون یا قاعده می‌باشد. فرآیند رشد درخت تصمیم QUEST شامل مراحل زیر است:

۱- انتخاب یک متغیر پیشگوی<sup>۳</sup> جدا کننده: متغیر پیشگو متغیری است که در مورد آن از رکوردهای ورودی به الگوریتم سؤال پرسیده می‌شود. در واقع هر متغیر پیشگو یکی از گره‌های درخت تصمیم و یکی از متغیرهایی است که در قواعد نهایی ظاهر خواهد شد. سپس بر اساس پاسخی که در مورد هر رکورد به این متغیر داده می‌شود، رکوردهای موجود در آن گره از طریق شاخه‌ها تفکیک و در گره‌ها یا برگ‌های پایین‌تر قرار خواهند گرفت.

۲- انتخاب یک نقطه تفکیک<sup>۴</sup> جهت متغیر پیشگوی انتخاب شده: نقطه تفکیک یک مقدار آستانه است که بر اساس مقایسه مقادیر رکوردها با مقدار آستانه، تفکیک رکوردهای موجود در گره صورت می‌پذیرد.

۳- توقف: فرآیند تفکیک رکوردها از طریق متغیرها و نقاط تفکیک سرانجام باید متوقف شود. با توجه به کاربرد و میزان داده‌ها، شرط توقف می‌تواند متفاوت باشد. به عنوان مثال می‌تواند بر اساس تعداد تفکیک‌های صورت گرفته، میزان ارتفاع درخت ساخته شده، درصد رکوردهای باقیمانده جهت تفکیک بیشتر و ... باشد [۳].

در یک گره<sup>۵</sup> فرض نمایید که یک متغیر پیش‌بینی  $X$  برای تقسیم انتخاب شده است. مرحله‌ی بعدی برای تعیین نقطه تقسیم است. اگر  $X$  متغیر پیش‌بینی پیوسته باشد یک نقطه تقسیم  $d$  در تقسیم  $X$   $d \leq$  تعیین می‌شود. اگر  $X$  متغیر پیش‌بینی مطلقاً صوری باشد، زیر مجموعه‌ی  $K$  از مجموعه مقادیر گرفته شده توسط  $X$  در تقسیم  $X$   $K \in$  تعیین می‌شود. گام توقف با توجه به موارد زیر، شرایط توقف الگوریتم را بررسی می‌نماید [۱۸]:

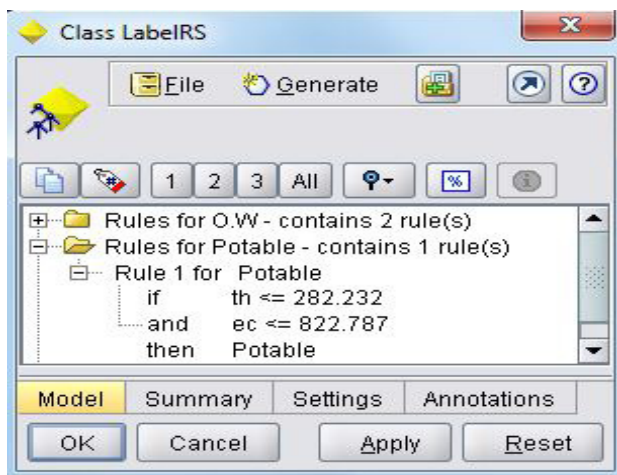
اگر یک گره، خالص<sup>۶</sup> شود؛ یعنی همه‌ی موارد موجود در آن گره، متعلق به یک کلاس باشند. در این صورت گره، قابل تقسیم شدن نیست.

اگر همه‌ی موارد موجود در یک گره، مقادیر یکسانی برای هر متغیر پیشگو داشته باشند که در این صورت گره، نمی‌تواند تقسیم شود.

3- Predictor  
4- Split point  
5- Node  
6- Pure

1- Predictor  
2- Split point





شکل ۵: قانون استخراج شده جهت تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی

### نتیجه گیری

با تحلیل مجموعه داده‌ها و پارامترهای انتخاب شده و با بهره‌گیری از تکنیک داده‌کاوی درخت تصمیم و الگوریتم QUEST مؤثرترین فاکتورهای کیفیت آب آشامیدنی در شهرستان پاسارگاد شناسایی گردیدند. با توجه به دو عامل سختی کل (TH) و هدایت الکتریکی (EC)، به نظر می‌رسد این دو فاکتور تأثیر بیشتری بر کیفیت آب آشامیدنی در این شهرستان دارند.

دلیل اثرگذاری بیشتر این دو عامل را می‌توان به ساختار سازندهای زمین‌شناسی منطقه و وجود آهک در ترکیب آن‌ها مرتبط دانست؛ زیرا میزان شدت سختی آب، به بستر جریان آب در سطح و زیر زمین بستگی دارد و آب‌های نواحی آهکی سختی زیادتری نسبت به آب‌های سایر نواحی (گرانیتی و یا شنی) دارند. به علاوه هدایت الکتریکی آب، نشان‌دهنده‌ی یون‌های موجود در آب می‌باشد؛ زیرا یون‌ها جریان الکتریسته را هدایت می‌کنند و از این نظر، قابلیت هدایت الکتریکی نسبت معین و مستقیمی با سختی آب دارد.

نتایج این پژوهش با یافته‌های آزار و همکاران [۵]، چو [۶]، فو چنگ و زو ژائو [۹]، گو و همکاران [۱۱]، لی و لی [۱۵]، سلیمان‌پور و همکاران [۲۶ و ۲۷ و ۲۸]، تامسون [۳۱] همخوانی دارد و مؤید کاربرد تکنیک‌های داده‌کاوی، و تأیید کننده‌ی اثربخشی این روش در تحقیقات مرتبط با منابع آب و تعیین مؤثرترین فاکتورهای کیفیت آن است. در پایان پیشنهاد می‌گردد نسبت به تصفیه و کاهش سختی آب توجه جدی شود؛ همچنین بر انجام پایش‌های مستمر در قالب نمونه‌برداری‌های دوره‌ای منظم از منابع آب چاه‌های این شهرستان تأکید می‌گردد.

عمق درخت فعلی به مقدار آستانه‌ی حداکثری تعیین شده توسط کاربر برسد که در این وضعیت، فرآیند رشد درخت متوقف می‌گردد. اندازه‌ی یک گره، کمتر از مقدار آستانه‌ی حداقلی تعیین شده توسط کاربر باشد که در این صورت گره، قابلیت تقسیم شدن را ندارد.

اگر با تقسیم یک گره، گره فرزندى تولید می‌شود که اندازه‌ی آن کمتر از مقدار آستانه‌ی حداقلی تعیین شده توسط کاربر باشد، آن‌گاه گره نمی‌تواند تقسیم شود.

### نتایج و بحث

به منظور ایجاد مدل، داده‌ها به طور تصادفی به دو بخش مجموعه داده‌های آموزشی و آزمایشی با تناسب به ترتیب ۵۰ درصد و ۵۰ درصد تقسیم شدند. پس از اجرای مدل با استفاده از درخت تصمیم‌گیری QUEST، مجموعه‌ای از قوانین توسط نرم‌افزار شناسایی و نمایش داده خواهد شد که مؤثرترین قاعده‌ی آن در ادامه ارائه شده است. همان‌طور که در شکل ۴ نشان داده شده است، دقت تخمین زده شده بر اساس مؤثرترین فاکتورهای کیفیت آب آشامیدنی در مدل نهایی، برای مجموعه داده‌های آموزشی برابر با ۹۱/۶۷ درصد و برای مجموعه داده‌های آزمایشی ۸۷/۰۱ درصد می‌باشد.

The screenshot shows the 'Analysis of (Class Label) #1' window. It has a menu bar with 'File' and 'Edit'. Below the menu is a toolbar with icons for file operations and a percentage sign. The main area displays a table titled 'Results for output field Class Label' with a sub-header 'Comparing \$R-Class Label with Class Label'. The table has columns for 'Partition', '1\_Training', and '2\_Testing', and rows for 'Correct', 'Wrong', and 'Total'. The data is as follows:

Partition	1_Training		2_Testing	
Correct	66	91.67%	67	87.01%
Wrong	6	8.33%	10	12.99%
Total	72		77	

At the bottom, there are tabs for 'Analysis' and 'Annotations', and an 'OK' button.

شکل ۴: ارزیابی انجام شده از الگوریتم QUEST

مؤثرترین قانون استخراج شده از درخت تصمیم QUEST در شکل ۵، نشان داده شده است. از این قانون می‌توان به منظور تعیین مؤثرترین فاکتورهای کیفیت آب آشامیدنی استفاده نمود. تفسیر این قانون، به شرح زیر می‌باشد:

مهمترین فاکتورهای تأثیرگذار در کیفیت آب آشامیدنی در شهرستان پاسارگاد (مطابق شکل ۵) عبارتند از: TH و EC بدین ترتیب، در صورتی که سختی کل (TH) در این شهرستان کمتر از ۲۸۲/۲۳۲ قسمت در میلیون، و هدایت الکتریکی (EC) آن، کمتر از ۸۲۲/۷۸۷ میکروموس بر سانتی‌متر باشد، این آب مناسب آشامیدن می‌باشد.

surface Water Quality in Lianyungang City. Fifth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA); 16-17 Jan, Hong Kong, China.

10. Ghafari, A.M. Najafenezhad, A. and Sadeddin, A. 2015. The regional analysis of chlorine and sodium water quality in Gorganroud watershed, Golestan province. Proceedings of 1<sup>st</sup> national conference of water resources quality and sustainable development, Arak, I.R.Iran. (In Persian)

11. Gu, Q. Deng, J. Wang, K. Lin, Y. Li, J. Gan, M. Ma, L. and Hong, Y. 2014. Identification and assessment of potential water quality impact factors for drinking-water reservoirs. International Journal of Environmental Research and Public Health. 11(6): 6069-6084.

12. Haseb Karaji Consultant Engineering Company. 1995. Comprehensive plan for controlling and directing a flood in Fars province (Marvdasht city). Vol 3, 167pp. (In Persian)

13. Kayser, G.L. Urooj, A. Fernanda, D. Jamie, B. and Margaret, E.B. 2015. Drinking Water Quality Governance: A Comparative Case Study of Brazil, Ecuador, and Malawi. Environmental Science and Policy. 48: 186-195.

14. Keyaghaderi, R. Ashrafepour, A. Abdoulzadeh, R. and Nouri, A. 2012. Comparison of water quality standards at the international level and the development of water quality criteria for sweet water resources. Proceedings of the 6<sup>th</sup> National Conference of Environmental Engineering, Tehran, I.R.Iran. (In Persian)

15. Lee, S. and Lee, C.W. 2015. Application of decision-tree model to groundwater productivity-potential mapping. Sustainability. 7(10): 13416-13432.

16. Lim, T.S. Loh, W.Y. and Shih, Y.S. 2000. A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms. Machine Learning publication. 40pp.

17. Lobato, T.C. Hauser-Davis, R.A. Oliveira, T.F. Silveira, A.M. Silva, H.A.N. Tavares, M.R.M. and Saraiva, A.C.F. 2015. Construction of a novel water

1. Abtahi, M. Golchinpour, N. Yaghmaeian, K. and Saeedi, R. 2015. A modified drinking water quality index (DWQI) for assessing drinking source water quality in rural communities of Khuzestan Province, Iran. Ecological Indicators. 53: 283-291.

2. Ashton, P.J. Van Zyl, F.C. and Heath, R.G. 1995. Water quality management in the Crocodile river catchment, Eastern Transvaal, South Africa. Water Science and Technology. 32(5-6): 201-208.

3. Azar, A. Ahmadi, P. and Sebt, M.V. 2010. Determining the Indicators Effective on Selection of Human Resources with the Data Mining Approach. Quarterly Journal of Education. 100(4): 87-111. (In Persian)

4. Azarangh, F. Telvari, A. and Shafaiebestani, M. 2015. Assessment of water quality for drinking and agriculture (Case Study: Karkheh river, downstream of dam). Proceedings of 1<sup>st</sup> national conference of water resources quality and sustainable development, Arak, I.R.Iran. (In Persian)

5. Azhar, S.A.S. Johar, H. Baki, S.M.S. and Tahir, N.M. 2013. Optimization of water quality monitoring based on fuzzy algorithms. Proceeding of IEEE Conference on Systems, Process & Control; 13-15 Dec, Kuala Lumpur, Malaysia.

6. Cho, Y. 2016. A watershed water quality evaluation model using data mining as an alternative to physical watershed models. Water Science and Technology: Water Supply. 16(3): 703-714.

7. Clasen, T. Pruss Ustun, A. Mathers, C.D. Cumming, O. Cairncross, S. and Colford, J.M. 2014. Estimating the impact of unsafe water, sanitation and hygiene on the global burden of disease: evolving and alternative methods. Tropical Medicine and International Health. 19(8): 884-893.

8. Ebrahimpour, S. and Mohammadzadeh, H. 2013. Assessment and Zoning of Water Quality Zarivar Lake using CWQI, OWQI and NSFQI quality indexes. Journal of environmental research. 7: 137-146. (In Persian)

9. Fu-Cheng, L. and Xue-Zhao, H. 2013. Application of Fuzzy c-means Clustering for assessing rural

Persian)

27. Soleimanpour, S.M. Mesbah, S.H. and Hedayati, B. 2016. The application of data mining algorithm K-Means and CART most influential factors in determining the quality of drinking water in Nurabad plain of Fars Province. Proceedings of Eleventh National Congress of Watershed Management Science and Engineering, Yasouj, I.R.Iran. (In Persian)
28. Soleimanpour, S.M. Mesbah, S.H. and Hedayati, B. 2016. The application of decision tree data mining techniques to determine the most effective factors on adverse drinking water in Kazerun plain, Fars province. Proceedings of the 6<sup>th</sup> conference on water resources management, Sanandaj, I.R.Iran. (In Persian)
29. Summiya, N. Zaffar Hashmi, M. and Malik, R. 2014. Heavy metals distribution, risk assessment and water quality characterization by water quality index of the River Soan, Pakistan. *Ecological Indicators*. 43: 262-270.
30. Tan, P.N. Steinbach, M. and Kumar, V. 2005. *Introduction to Data Mining*. Addison-Wesley. 134pp.
31. Thompson, E. 2016. Investigating drinking water advisories in first nations communities through data mining. A Thesis for Master of Applied Science in Engineering, University of Guelph, Ontario. Canada.
32. Tu, J. 2013. Spatial variations in the relationships between land use and water quality across an urbanization gradient in the watersheds of northern Georgia, USA. *Environmental management*. 51(1): 1-17.
33. Zheng, B. Lei, K. Liu, R. Song, S. and An, L. 2014. Integrated biomarkers in wild crucian carp for early warning of water quality in Hun River, North China. *Journal of Environmental Sciences*. 26(4): 909-916.
- quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. *Journal of Hydrology*. 522: 674-683.
18. Loh, W.Y. and Shih, Y.S. 1997. Split selection methods for classification trees. *Statistica Sinica*. 7: 815-840.
19. Marinoni, O. Higgins, A. Coad, P. and Garcia, J.N. 2013. Directing urban development to the right places: Assessing the impact of urban development on water quality in an estuarine environment. *Landscape and Urban Planning*. 113: 62-77.
20. Mohtashami, A. and Naseri, A. 2015. Classification of water quality for drinking, agriculture and industry (case study: Darmeyan plain, south Khorasan province). Proceedings of 1<sup>st</sup> national conference of water resources quality and sustainable development, Arak, I.R.Iran. (In Persian)
21. Navi, M. 2007. Identify load components with the use of data mining techniques. M.Sc. Thesis, Tarbiyatmodares University, 116pp. (In Persian)
22. Saneieabadeh, M. Mahmoudi, S. and Taherpour, M. 2012. *Applied Data mining*. Neyazedanesh Publication, 520pp. (In Persian)
23. Shahrabi, J. 2013. *Data mining 2*. Jahadedaneshghahi Publication, 299pp. (In Persian)
24. Shareateshirenasab, A. Soleimanpour, S.M. and Jowkar, L. 2012. Relationship between water quality factors using stepwise in the Zarendasht region, Fars province. Proceedings of the First National Congress of desert, Karaj, I.R.Iran. (In Persian)
25. Soleimanpour, S.M. Shareateshirenasab, A. and Jowkar, L. 2012. Investigation of the quality of groundwater resources and the changes of qualitative elements (Case Study: Plain Khosouyeh Sachun plain, Zarendasht region, Fars province). Proceedings of the First National Congress of desert, Karaj, I.R.Iran. (In Persian)
26. Soleimanpour, S.M. Mesbah, S.H. and Hedayati, B. 2016. Implementation of CART algorithm to determine the most effective drinking water quality factors (case study: Kazerun plain, Fars province). 2<sup>nd</sup> national conference on conservation of natural resources and environment, Ardabil, I.R.Iran. (In





## Abstract

## Determination of the most Effective Drinking Water Quality Factors using QUEST Data Mining Technique (Case Study: Pasargad City, Fars Province)

S. M. Soleimanpour\*<sup>1</sup>, S. H. Mesbah<sup>2</sup> and B. Hedayati<sup>3</sup>

Received: 2017/05/06 Accepted: 2018/04/10

Nowadays, water quality is one of the key issues in most regions all over the world. Because of its high relation with human health and also have important role in management and exploitation of resources. This research has done first time in Pasargad city located in 105 km of eastern north of Shiraz. In order to determine drinking water quality factors using data mining technique, results of this research have obtained using Quest decision tree in Clementine software(12 version) and showed most effective factors of drinking water in this region is function of total hardness and electrical conductivity. In this way, if total hardness in this city is less than 282.232 parts in million and electrical conduction is less than 822.787 micro mohs in centimeter, this water will be suitable for drinking. Therefore recommended more noting to refine and decrease the hardness of water for human drinking water, moreover continuous refining in periodically regular sampling of water resources in this city should be noted.

**Keywords: Drinking water, Data mining, Total hardness, Water quality, Electrical conductivity**

1- Assistant Professor, Soil Conservation and Watershed Management Research Department, Fars Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shiraz, Iran, Corresponding Author, Email: m.soleimanpour@yahoo.com

2- Instructor, Soil Conservation and Watershed Management Research Department, Fars Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shiraz, Iran.

3- Graduate of M.Sc., Software Computer Engineering, Payam-e Noor University, Tehran, Iran.