



## Prediction of Groundwater Quality of Khanmirza plain Using Decision tree Method

Zahra Hasani<sup>1</sup>, Rasoul Mirabbasi Najafabadi<sup>\*2</sup>,  
Ahmad Reza Ghasemi<sup>2</sup>

### Abstract

Groundwater is an important resource for agricultural, industrial and drinking uses in arid and semi-arid regions. Therefore, study and management of these valuable resources is necessary. Since the monitoring and investigation of groundwater quality is a time consuming and costly process, so finding a method which enables us to forecast groundwater quality with the least number of parameters lead to save money and time. The aim of this study is prediction of groundwater quality class, based on the USSSL diagram in the Khanmirza plain by decision tree method. The chemical parameters and monthly and cumulative precipitation were used as model inputs. For this purpose, the water quality data of 19 wells which located in the Khanmirza plain, were used for the period 1991-2011. The results showed that decision tree method is able to classify water quality with high accuracy, based on only 4 chemical parameters (i.e. EC, Na, SAR and Cation). In the other words, in order to classify the groundwater quality of Khanmirza plain in the future, measuring these four parameters is enough, which significantly reduces the cost and time of analysis.

**Keywords:** Khanmirza Plain, USSSL Diagram, Decision Tree Method, Water Quality, Agricultural Uses.

Received: 2016/09/25  
Accepted: 2017/02/26

1- MSc student of Water Resources Engineering, Shahrekord University.

2, 3- Assistant Professor, Department of Water Engineering, Shahrekord University

\*- Corresponding Author: mirabbasi\_r@yahoo.com

## پیش بینی کیفیت آب زیرزمینی دشت خانمیرزا با استفاده از روش تصمیم گیری درختی

زهرا حسنی<sup>۱</sup>، رسول میرعباسی نجفآبادی<sup>\*۲</sup>، احمدرضا قاسمی<sup>۲</sup>

### چکیده

آب زیرزمینی یکی از منابع مهم برای کشاورزی، شرب و صنعت در مناطق خشک و نیمه خشک به شمار می رود. بنابراین، مطالعه و مدیریت این منابع ارزشمند امری حیاتی می باشد. به دلیل اینکه پایش و بررسی کیفیت آب های زیرزمینی نیازمند صرف زمان و هزینه زیاد است، بنابراین یافتن روشی که بتوان با استفاده از آن و با تعداد پارامتر هیدروشیمیایی محدود، طبقه کیفیت آب را پیش بینی نمود، موجب صرفه جویی در زمان و هزینه خواهد شد. هدف از این مطالعه، پیش بینی طبقه کیفی آب زیرزمینی دشت خانمیرزا بر اساس دیاگرام USSSL با استفاده از روش تصمیم گیری درختی می باشد. پارامترهای شیمیایی آب و همچنین بارش ماهانه و تجمعی به عنوان ورودی های مدل استفاده شدند. بدین منظور از داده های کیفی ۱۹ حلقه چاه واقع در این دشت، در بازه زمانی ۹۰-۱۳۷۰ استفاده شد. نتایج به دست آمده از مدل نشان داد روش تصمیم گیری درختی قادر است تنها با استفاده از ۴ پارامتر هیدروشیمیایی (EC, Na, SAR و Cation) طبقه کیفیت آب را با دقت بسیار بالایی تعیین نماید. به عبارت دیگر، برای تشخیص کلاس کیفیت آب دشت خانمیرزا در آینده، اندازه گیری این چهار پارامتر کافی می باشد که این مسئله باعث کاهش قابل توجه هزینه ها و زمان تحلیل آزمایش ها می گردد.

واژه های کلیدی: دشت خانمیرزا، دیاگرام USSSL، روش تصمیم گیری درختی، کیفیت آب، مصارف کشاورزی.

تاریخ دریافت مقاله: ۱۳۹۵/۰۷/۰۴

تاریخ پذیرش مقاله: ۱۳۹۶/۰۱/۰۵

۱- کارشناسی ارشد مهندسی منابع آب، دانشگاه شاهرود.

۲ و ۳- استادیار گروه مهندسی آب، دانشکده کشاورزی دانشگاه شاهرود.

\*- نویسنده مسئول

## مقدمه

همکاران، ۲۰۰۲). در این شیوه از یک ساختار درختی برای طبقه‌بندی داده‌ها استفاده می‌شود. در مدل‌سازی پدیده‌ها، وجود شرایط محلی ممکن است باعث شود تا استفاده از یک رابطه کلی نتایج خوبی را به همراه نداشته باشد و تغییرات محلی به خوبی دیده نشوند. در صورت امکان، شناسایی محدوده‌های همگن و ارائه روابط ساده خطی برای هر یک از این محدوده‌ها می‌تواند باعث افزایش دقت مدل شود. بر این اساس معمولاً برای حل یک مسئله پیچیده، آن را به چند مسئله کوچک‌تر و ساده‌تر تقسیم نموده و سپس جواب‌های به دست آمده را با هم ترکیب می‌کنند. همین ایده ساده در مدل‌های درخت تصمیم مورد استفاده قرار می‌گیرد. به این منظور، فضا یا محدوده مقادیر داده‌های ورودی به چند زیر بازه یا ناحیه تقسیم شده و برای هر ناحیه یک معادله یا مدل ساده استخراج می‌شود (ظهیری و قربانی، ۱۳۹۲). نتایج داده‌کاوی می‌تواند در فرم ساده‌ای تحت عنوان درخت تصمیم‌گیری ارائه شود. این شیوه غیر پارامتری بوده و به وجود داده‌های مفقود غیرحساس می‌باشد، لذا از آن به عنوان یک روش پیش‌بینی قوی یاد شده است. درختان تصمیم‌گیری در زمینه‌های مختلف شامل علوم رایانه، پزشکی، محیط‌زیست، سنجش از راه دور و ... کاربرد دارند. درختان تصمیم‌گیر قادر به تولید توصیفات قابل درک برای انسان، از روابط موجود در یک مجموعه داده هستند و می‌توانند برای وظایف دسته‌بندی و پیش‌بینی به کار روند (ثاقبیان و همکاران، ۱۳۹۲).

تعیین پارامترهای مؤثر در تعیین کلاس کیفیت آب‌های زیرزمینی با استفاده از روش تصمیم‌گیری درختی، باعث خواهد شد که علاوه بر کاهش چشم‌گیر هزینه‌های نمونه‌برداری و آنالیز شیمیایی نمونه‌ها، زمان تعیین کلاس کیفیت نیز کاهش یابد و در عین حال نتایج از دقت قابل قبولی برخوردار باشند. این شیوه خصوصاً در مواردی که طرح مورد اجرا کوچک و هزینه‌های مطالعات محدود می‌باشند (مثل طراحی سیستم‌های آبیاری تحت فشار کوچک)، از ارجحیت برخوردار است. در این شیوه با اندازه‌گیری تعداد محدودی از پارامترها و با استفاده از برخی پارامترهای هواشناسی یا هیدرولوژیکی نظیر بارش یا دبی جریان، می‌توان کلاس کیفیت آب را با دقت خوبی تعیین نمود. در مورد استفاده از روش تصمیم‌گیری درختی در مسائل مرتبط

کشور ایران از جمله کشورهایی است که در منطقه خشک و نیمه‌خشک واقع گردیده و در سال‌های اخیر با خشک‌سالی‌ها و کمبود آب شدید مواجه شده است. در چنین شرایطی کنترل کیفیت آب‌های موجود و برنامه‌ریزی جهت بهره‌برداری از این آب‌ها با توجه به طبقه‌بندی کیفیت آن‌ها ضروری است. این در حالی است که پیش‌بینی کیفیت آب‌های زیرزمینی بسیار پرهزینه و زمان‌بر بوده و انتخاب روشی که در آن با حداقل پارامترهای هیدروشیمیایی بتوان پیش‌بینی نسبتاً دقیقی از طبقه کیفیت آب برای مصارف مختلف داشت، مهم و ضروری است. روش‌های نوین داده‌کاوی که در پی کشف دانش نهفته و استخراج الگوها و قوانین موجود در بین داده‌هاست، در سال‌های اخیر رشد چشم‌گیری داشته و در اغلب موضوعات کاربرد فراوانی پیدا نموده‌اند. این شیوه‌های جدید برای مدل کردن فرآیندهایی که یا داده‌های موجود برای فرمول‌بندی یک معادله ریاضی بسیار محدود بوده و یا داده‌های کمی برای تعیین یک مدل مناسب در دسترس باشند، مناسب هستند. شیوه‌های محاسبه نرم به‌عنوان یک شیوه جایگزین و یا مکمل برای شیوه‌های قدیمی‌تر مبتنی بر مدل‌های فیزیکی مورد استفاده قرار می‌گیرند. این کاربردها نشان می‌دهد که شیوه‌های محاسبه نرم<sup>۱</sup> اطلاعات جدیدی در مورد پدیده تولید نمی‌کند، اما از اطلاعات موجود در مورد پدیده و با استفاده از تکنیک‌های رگرسیون برای برازش بهتر بر داده‌های مشاهداتی در انتخاب متغیرهای ورودی و خروجی استفاده می‌کند.

تصمیم‌گیری درختی یکی از شیوه‌های محاسبه نرم می‌باشد که از داده‌های مشاهده‌ای برای توسعه مکانیسم تصمیم‌گیری برای مدیران استفاده می‌نماید (ثاقبیان و همکاران، ۲۰۱۴). توسعه مدل‌های پشتیبانی تصمیم‌گیری برای کمک به مدیران برای برنامه‌ریزی در مورد استراتژی برای حفاظت از سیستم‌های توزیع آب ضروری می‌باشد (راسخ و براملو، ۲۰۱۴). به دلیل نیاز به کاهش هزینه‌های تجزیه و تحلیل کیفیت آب، به‌ویژه در کشورهای در حال توسعه، روش‌هایی برای افزایش سرعت و کاهش تجزیه و تحلیل کیفیت آب ضروری است (آکیلدیز و

1- Calculation software

داد که طبقه‌بندی کیفیت آب زیرزمینی توسط تصمیم‌گیری درختی دقیق‌تر و کارا تر و کارا تر از روش مؤلفه‌های اصلی می‌باشد. روش پیشنهادی توانست کلاس کیفیت آب زیرزمینی دشت اردبیل را تنها با دو پارامتر هدایت الکتریکی (EC) و بارش تجمعی ۱۱ ماه قبل به‌خوبی تعیین کند. مدل توسعه داده شده قادر به پیش‌بینی کلاس کیفیت آب بر اساس تنها دو متغیر بود و این باعث کاهش تعداد متغیرهای موردنیاز در تعیین کلاس کیفیت آب و در نتیجه کاهش چشم‌گیر هزینه‌های آزمایشگاهی و زمان تلف شده بین نمونه‌برداری و دریافت نتایج از آزمایشگاه می‌شود.

هدف از مطالعه حاضر، پیش‌بینی کلاس کیفیت آب زیرزمینی دشت خانمیرزا برای مصارف کشاورزی، بر اساس دیاگرام USSL، با استفاده از پارامترهای شیمیایی و میزان بارش با روش تصمیم‌گیری درختی می‌باشد. تعیین پارامترهای مؤثر در پیش‌بینی کلاس کیفیت آب زیرزمینی دشت خانمیرزا از اهداف دیگر این مطالعه می‌باشد.

#### مواد و روش‌ها

##### منطقه مورد مطالعه و داده‌های مورد استفاده

حوضه خانمیرزا در استان چهارمحال و بختیاری قرار دارد. استان چهارمحال و بختیاری در بخش مرکزی کوه‌های زاگرس واقع شده است. این استان از شمال و شرق به استان اصفهان، از غرب به استان خوزستان، از جنوب به کهگیلویه و بویراحمد و از شمال غرب به استان لرستان محدود است. محدوده دشت خانمیرزا در شهرستان لردگان و به‌طور دقیق‌تر در بخش خانمیرزا واقع شده است. این شهرستان با وسعت ۳۴۲۳ کیلومتر مربع حدود ۲۰/۷ درصد از کل مساحت استان را تشکیل می‌دهد. محدوده آبخوان خانمیرزا دارای وسعت حدود ۱۲۸ کیلومتر مربع است که در طول جغرافیایی ۵۱ درجه و ۲ دقیقه تا ۵۱ درجه و ۱۱ دقیقه شرقی و عرض جغرافیایی ۳۱ درجه و ۲۵ دقیقه تا ۳۱ درجه و ۳۶ دقیقه شمالی قرار گرفته است. شکل ۱ موقعیت آبخوان خانمیرزا در استان چهارمحال و بختیاری را نشان می‌دهد.

در این مطالعه، به‌منظور پیش‌بینی کلاس کیفی آب زیرزمینی دشت خانمیرزا با تعداد پارامتر شیمیایی کمتر و بارش تجمعی، به‌طوری که تا حد امکان از صرف زمان و هزینه‌های زیاد

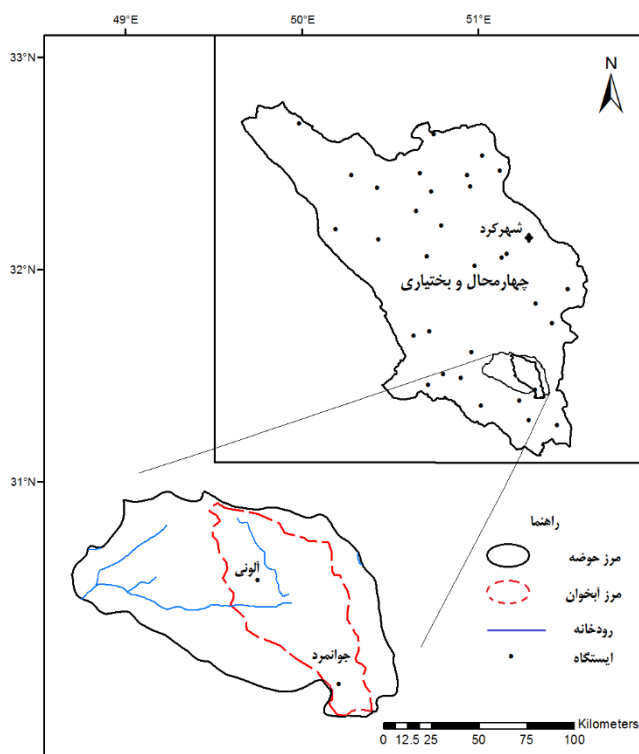
با مهندسی آب و به‌ویژه در ارزیابی و پیش‌بینی کیفیت آب می‌توان به موارد زیر اشاره نمود:

ایسنبرگ و مک کئون (۱۹۹۸) روشی مبتنی بر تصمیم‌گیری درختی برای برآورد اثرات دخالت‌های انسانی بر کیفیت آب توسعه دادند. اسپرویل و همکاران (۲۰۰۲) از مدل تصمیم‌گیری درختی برای طبقه‌بندی منابع نیترات در آب‌های زیرزمینی استفاده نمودند. آن‌ها دو مدل طبقه‌بندی درختی آماری از ۴۸ نمونه آب محتوی نیترات از ۵ دسته منبع توسعه دادند. نتایج نشان داد که هر دو مدل قادر به تشخیص هر پنج دسته منبع با موفقیت کلی بهتر از ۸۰ درصد و با موفقیت ۷۱ تا ۱۰۰ درصد در دسته‌های انفرادی با استفاده از نمونه‌های یادگیری بودند. لیتائور و همکاران (۲۰۱۰) یک مدل تصمیم‌گیری درختی دوتایی برای ارزیابی شاخص‌های هیدروشیمیایی آبخوان به عنوان ورودی به کار بردند. نتایج آن‌ها نشان داد که مدل درختی به‌طور واضح شاخص‌های هیدروشیمیایی را که می‌توان از تحلیل‌های آزمایشگاهی حذف نمود، مشخص می‌کند. بنابراین، این روش منجر به کاهش چشمگیری در تعداد پارامترهای اندازه‌گیری و زمان آزمایش‌ها می‌گردد.

محبوبی و همکاران (۱۳۸۹) مقایسه عملکرد دو الگوریتم شبکه عصبی مصنوعی پرسپترون چندلایه و درخت تصمیم‌گیری رگرسیونی در پیش‌بینی تغییرات شوری آب رودخانه کارون را مورد ارزیابی قرار دادند. داده‌های استفاده شده مربوط به ایستگاه ملاثانی در بازه سال‌های ۱۳۴۷ تا ۱۳۸۴، شامل غلظت‌های سدیم، کلسیم، منیزیم، کلر، سولفات، بی‌کربنات، pH و دبی به‌عنوان پارامترهای ورودی مدل‌ها و کل مواد محلول (TDS) به‌عنوان پارامتر خروجی استفاده شده است. مقایسه عملکرد مدل‌ها نشان داد که با وجود دقت بالاتر شبکه عصبی مصنوعی پرسپترون چند لایه در پیش‌بینی غلظت مواد محلول، خروجی درخت تصمیم‌گیری رگرسیونی که در قالب قوانینی ارائه می‌گردد، قابل تفسیرتر می‌باشد. ثاقبیان و همکاران (۲۰۱۴) یک شیوه جدید مبتنی بر تصمیم‌گیری درختی برای طبقه‌بندی و پیش‌بینی کیفیت آب زیرزمینی مبتنی بر دیاگرام آزمایشگاه شوری ایالات متحده (USSL) با استفاده از داده‌های آبخوان اردبیل پیشنهاد کردند. ارزیابی عملکرد روش پیشنهادی حاکی از مناسب بودن شیوه طبقه‌بندی مبتنی بر تصمیم‌گیری درختی برای مجموعه داده‌های مورد استفاده بود. نتایج نشان

جلوگیری شود، از روش تصمیم‌گیری درختی استفاده گردید. بدین منظور از داده‌های کیفی ۱۹ حلقه چاه در بازه زمانی ۹۰-۱۳۷۰ استفاده شد که شامل، یون کلسیم (Ca)، یون منیزیم (Mg)، یون کلر (Cl)، بی‌کربنات (HCO<sub>3</sub>)، درصد سدیم (Na%)، اسیدیته (pH)، سولفات (SO<sub>4</sub>)، مجموع آنیون‌ها، مجموع کاتیون‌ها، کل نمک محلول (TDS)، نسبت سدیم جذبی (SAR) و هدایت الکتریکی (Ec) می‌باشند. همچنین از داده‌های بارش ماهانه ایستگاه باران‌سنجی آلونی در دوره آماری ۹۰-۱۳۷۰ استفاده شد. علت استفاده از بارش در این مطالعه و لحاظ کردن آن به عنوان یک ورودی به مدل تصمیم‌گیری

درختی این است که مشخص شود آیا میزان بارش در طول زمان تأثیری بر کیفیت آب زیرزمینی دشت خانمیرزا داشته است؟ چون بارش با نفوذ و حرکت در لایه‌های مختلف خاک می‌تواند مواد محلول را در اعماق مختلف خاک حین نفوذ با خود حمل کند و بر کیفیت آب زیرزمینی اثرگذار باشد. در این تحقیق، بارش به صورت تجمعی برای بازه‌های یک تا ۱۲ ماه قبل در مدل استفاده می‌شود. میانگین پارامترهای هیدروشیمیایی در ۱۹ حلقه چاه انتخابی در دوره آماری ۹۰-۱۳۷۰ در جدول ۱ ارائه شده است.



شکل ۱- موقعیت آبخوان خانمیرزا در استان چهارمحال و بختیاری.

جدول ۱- مقادیر متوسط پارامترهای شیمیایی ۱۹ چاه انتخابی در آبخوان خانمیرزا (۹۰-۱۳۷۰)

مقدار	پارامتر	مقدار	پارامتر
۰/۵۹	So <sub>4</sub> (mg/l)	۲/۷۶	Ca <sup>2+</sup> (mg/l)
۷/۹۲	Total Anions (mg/l)	۲/۴۲	mg <sup>2+</sup> (mg/l)
۷/۹۳	Total Kations (mg/l)	۳/۲۳	Cl <sup>-</sup> (mg/l)
۵۱۳/۰۱	TDS (mg/l)	۴/۰۴	Hco <sub>3</sub> (mg/l)
۱/۳۹	SAR	۱۷/۱۷	Na%
۷۸۹/۳۰	Ec (μ mho/cm)	۸/۰۶	pH

## روش‌های مورد استفاده

استفاده نمود. اشیائی که باید دسته‌بندی شوند، معمولاً به وسیله اطلاعاتی در جدول پایگاه داده‌ها یا یک فایل ارائه می‌شوند و عمل دسته‌بندی شامل افزودن ستون جدیدی با کد دسته‌بندی خاصی است. درختان تصمیم‌گیری به منظور پیشگویی یا کلاسه‌بندی داده‌ها بر اساس مجموعه قوانین تصمیم ایجاد شده‌اند. کلاسه‌بندی داده‌ها با درختان تصمیم‌گیری یک فرایند دو مرحله‌ای می‌باشد: در مرحله اول که به آن مرحله آموزش گفته می‌شود، مدلی بر اساس یک الگوریتم کلاسه‌بندی منطبق با داده‌های مربوط به مجموعه آموزشی ساخته می‌شود. مجموعه آموزشی به صورت تصادفی از پایگاه داده انتخاب می‌شود. در مرحله دوم، یادگیری از طریق یک تابع  $y=f(X)$  انجام می‌شود که می‌تواند برچسب کلاس هر رکورد  $X$  از پایگاه داده را پیش‌بینی کند. مرحله یادگیری خود طی دو گام اساسی رشد و هرس انجام می‌شود. در طول فرآیند آموزش الگوریتم درخت تصمیم می‌بایست به صورت مکرر، مؤثرترین روش جهت تقسیم کردن مجموعه رکوردها به فرزندان را بیابد. مرحله هرس<sup>۱</sup> برای جلوگیری از پردازش بیش از حد و بزرگ شدن درخت تصمیم که باعث پیچیدگی و افزایش تعداد قوانین اگر-آنگاه می‌شود، صورت می‌گیرد. در روش درخت تصمیم معیار مورد استفاده برای ایجاد شاخه‌ها و جداسازی، بی‌نظمی یا آنتروپی می‌باشد. اگر ویژگی هدف دارای  $C$  مقدار مختلف باشد، آنتروپی  $S$  نسبت به این دسته‌بندی  $C$  گانه به صورت زیر تعریف می‌شود (کوینلن، ۱۹۹۳):

$$Entropy(S) = \sum_{i=1}^c -p_i \log p_i \quad [1]$$

در رابطه (۱)،  $p_i$  نسبتی از  $S$  است که به دسته  $i$  تعلق دارد. توجه شود که لگاریتم در مبنای ۲ در نظر گرفته می‌شود. در این حالت حداکثر آنتروپی می‌تواند  $\log_2 c$  باشد. بهره اطلاعات یک ویژگی عبارت است از مقدار کاهش آنتروپی که به واسطه جداسازی مثال‌ها از طریق این ویژگی حاصل می‌شود. به عبارت دیگر، بهره اطلاعات  $Gain(S, A)$  برای یک ویژگی نظیر  $A$  نسبت به مجموعه مثال‌های  $S$  به صورت زیر تعریف می‌شود (کوینلن، ۱۹۹۳):

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad [2]$$

داده‌کاوی به بررسی و تجزیه و تحلیل مجموعه بزرگی از داده‌ها به منظور کشف الگوها و قوانین پنهان و معنی‌دار درون داده‌ها اطلاق می‌شود. داده‌کاوی به دو نوع هدایت شده و غیرهدایت شده تقسیم‌بندی می‌شود (ویتن و فرانک، ۲۰۰۵). داده‌کاوی هدایت شده، دارای متغیر هدف خاص و از پیش تعیین شده است که به دنبال الگویی خاص می‌گردد، در حالی که هدف داده‌کاوی غیر هدایت شده، یافتن الگوها یا تشابهات بین گروه‌هایی از اطلاعات، بدون داشتن متغیر هدف خاص و یا مجموعه‌ای از دسته‌ها و الگوهای از پیش تعیین شده می‌باشد. داده‌کاوی عمدتاً با ساختن مدل‌ها مرتبط است. یک مدل اساساً به الگوریتم یا مجموعه‌ای از قوانینی گفته می‌شود که مجموعه‌ای از ورودی‌ها را با هدف یا مقصد خاصی مرتبط می‌نماید. یک مدل تحت شرایط درست می‌تواند منجر به بینش درست شود. داده‌کاوی شامل مراحل مختلفی می‌باشد که عبارتند از: ۱- تعیین اطلاعات گذشته، ۲- پیرایش داده‌ها و پردازش اولیه. در این مرحله خطاهای داده‌ها تصحیح و داده‌های اشتباه جایگزین می‌شوند، ۳- یکپارچه‌سازی داده‌ها. در بیشتر موارد داده‌ها از منابع متفاوتی جمع‌آوری می‌شوند که باید به صورت پایگاهی از داده‌های مناسب درآیند تا عملیات داده‌کاوی بهتر انجام شود، ۴- انتخاب مجموعه متغیرهای هدف، ۵- یافتن ویژگی‌های مورد استفاده و تعیین ویژگی‌های جدید، ۶- نمایش داده‌ها به صورتی که بتوان در داده‌کاوی استفاده نمود، ۷- انتخاب عملیات داده‌کاوی (طبقه‌بندی، خوشه‌بندی، پیش‌بینی و غیره)، ۸- انتخاب روش داده‌کاوی (شبکه‌های عصبی، درخت تصمیم و نظایر آن)، ۹- انجام داده‌کاوی و جستجو برای یافتن الگوی مناسب، ۱۰- ارزیابی و تحلیل الگوی به دست آمده و حذف الگوهای نامناسب، ۱۱- تفسیر نتایج داده‌ها و استنتاج اطلاعات با ارزش. طبقه‌بندی که یکی از معمول‌ترین کارکردهای داده‌کاوی است، شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین شده می‌باشد (ویتن و فرانک، ۲۰۰۵). عمل طبقه‌بندی با تعریف درستی از دسته‌ها و مجموعه‌ای از ویژگی‌ها که حاوی موارد از پیش دسته‌بندی شده هستند، انجام می‌شود. این عمل شامل ساختن مدلی است که بتوان از آن برای طبقه‌بندی داده‌های دسته‌بندی نشده،

<sup>1</sup>- Step pruning

استفاده از دیاگرام USSL و بر اساس پارامترهای هیدروشیمیایی مؤثر در ارزیابی کیفیت آب، طبقه و کلاس کیفیت آب مشخص شدند. بر اساس نمونه‌برداری‌های انجام گرفته در ۱۹ حلقه چاه و داده‌های بارش ایستگاه باران‌سنجی آلونی واقع در این دشت، در مجموع برای آب زیرزمینی این آبخوان بارش ماهانه و ۱۲ پارامتر هیدروشیمیایی شامل، یون کلسیم (Ca)، یون منیزیم (Mg)، یون کلر (Cl)، بی‌کربنات (HCO<sub>3</sub>)، درصد سدیم (Na%)، اسیدیته (pH)، سولفات (SO<sub>4</sub>)، مجموع آنیون‌ها، مجموع کاتیون‌ها، کل نمک محلول (TDS)، نسبت سدیم جذبی (SAR) و هدایت الکتریکی (Ec) مورد توجه قرار گرفت. ۱۲ صفت و ویژگی هیدروشیمیایی ذکر شده و بارش تجمعی به عنوان ورودی مدل درخت تصمیم C5.0 و کلاس کیفیت آب به‌عنوان خروجی و ویژگی هدف در نظر گرفته شد. کلاس کیفیت آب که بر اساس دیاگرام USSL تعیین می‌شود و در اینجا به‌عنوان ویژگی هدف مورد استفاده قرار گرفته، دارای ماهیتی گسسته می‌باشد. در زیر نتایج به-دست آمده از تحلیل مدل جهت تعیین کلاس کیفیت آب زیرزمینی آبخوان خانمیرزا ارائه گردیده است.

در این دشت جمعاً به تعداد ۵۴۷ مورد نمونه‌برداری از آب در فواصل زمانی مختلف صورت گرفته و طی آن با توجه به پراکنش پارامترهای هیدروشیمیایی، کلاس کیفیت آب بر اساس دیاگرام USSL تعیین گردید. از بین ۵۴۷ مورد به تفکیک تعداد ۴۰۰ نمونه برای آموزش مدل و مابقی (یعنی ۱۴۷ نمونه) برای آزمون در نظر گرفته شد (ستاری و همکاران، ۲۰۱۳؛ تاپک و همکاران، ۲۰۱۴). خلاصه نتایج مدل تصمیم درختی C5.0 برای آب زیرزمینی دشت خانمیرزا در جدول ۲ ارائه گردیده است. نتایج مدل تصمیم درختی C5.0 شامل چند قانون است که به‌صورت اگر-آنگاه ارائه می‌گردد و هر قانون را می‌توان به‌صورت جداگانه بررسی کرد. به‌طور مثال، در جدول ۲ قانون شماره یک بیان می‌کند که اگر  $Ec \leq 750$  باشد، آنگاه کیفیت آب زیرزمینی دشت خانمیرزا در طبقه C2-S1 قرار دارد. در قوانین ارائه شده از نتیجه مدل C5.0 دقت بر اساس آماره لیفت می‌باشد. به‌طوری که هر قانون دارای یک آماره لیفت و یک آماره دقت لاپلاس است. هرچه مقدار آماره لیفت یک قانون کمتر باشد، دقت آن قانون بیشتر خواهد بود.

در رابطه (۲)، Values (A) مجموعه‌ای از مقادیر ممکن برای ویژگی‌های A بوده و S<sub>v</sub> زیرمجموعه‌ای از S است که در آن ویژگی A دارای مقدار V است. در این مطالعه، جهت توسعه مدل درختی کیفیت آب زیرزمینی و استخراج قواعد اگر-آنگاه از نرم‌افزار See 5 که مبتنی بر الگوریتم C5.0 بوده و توسط (کوینلن، ۲۰۰۰) توسعه داده شده، استفاده گردید. هر قانون در برنامه See 5 شامل شماره قانون و آماره (N, Lift X) یا (N/M) که کارایی قانون را خلاصه می‌کند، می‌باشد. N نشان دهنده تعدادی از مکان‌هایی که همزمان در طول آموزش با قانون بیان شده همخوانی و درست پیش‌بینی شده است، می‌باشد (برآورد درست و صحیح). M تعداد اطلاعات یا داده‌هایی که با قانون همخوانی نداشته و غلط پیش‌بینی شده است، می‌باشد (برآورد غلط یا نادرست). دقت قوانین تخمین زده شده با نرخ یا دقت لاپلاس و نیز LIFT X به‌صورت روابط زیر بیان می‌شوند:

$$[3] \quad \text{دقت لاپلاس} = (N - M + 1) / (N + 2)$$

$$[4] \quad \text{Lift X} = ((N - M + 1) / (N + 2)) \quad (\text{تعداد داده‌های همخوان با قانون} / \text{تعداد کل داده‌های این کلاس یا طبقه})$$

نرخ یا دقت لاپلاس قابل توجه‌ترین و مفیدترین آماره در ارزیابی قوانین می‌باشد (کوینلن ۲۰۰۰). نرم‌افزار See 5 همچنین می‌تواند اطلاعات از دست‌رفته را زمانی که یک مقدار در یک گره از درخت تصمیم‌گیری ناشناخته باشد، دسته‌بندی و طبقه‌بندی کند که این الگوریتم تمامی ترکیبات و طبقه‌بندی‌های به‌دست آمده را بررسی می‌کند و کلاسی را که دارای بیشترین احتمال باشد، به‌عنوان کلاس پیش‌فرض انتخاب می‌کند (کوینلن، ۱۹۹۳).

## نتایج و بحث

در مطالعه حاضر، هدف پیش‌بینی کلاس کیفی آب زیرزمینی دشت خانمیرزا با تعداد پارامتر شیمیایی کمتر و بارش تجمعی است، به‌طوری که تا حد امکان از صرف زمان و هزینه‌های زیاد جلوگیری شود. برای این منظور در این تحقیق، از روش تصمیم‌گیری درختی برای پیش‌بینی کیفیت آب زیرزمینی دشت خانمیرزا استفاده گردید. در این دشت، ابتدا با

جدول ۲- خلاصه نتایج مدل تصمیم درختی C5.0 برای آب زیرزمینی دشت خانمیرزا.

شماره قانون	اگر			آنگاه (طبقه کیفیت آب)	تعداد نمونه غیرصادق	تعداد نمونه صادق	دقت لاپلاس (درصد)	LIFT
	۱	۲	۳					
۱	$Ec \leq 750$	-	-	C2-S1	۰	۲۲۲	۹۹/۱	۱/۸
۲	$Ec > 750$	-	-	C3-S1	۱۰۳	۱۷۸	۵۵/۵	۲/۳
۳	$Na > 9/13$	$Cation \leq 22/63$	$SAR \leq 10/589$	C3-S2	۰	۲۸	۹۶/۶	۱۱/۸
۴	$Na > 7/05$	$Cation \leq 22/63$	$SAR > 4/903$	C3-S2	۴	۳۳	۸۵/۷۱	۱۰/۴
۵	$Cation > 22/63$	$SAR \leq 7/888$	-	C4-S2	۱	۲۶	۹۲/۸۵	۱۴/۳
۶	$Cation > 22/63$	$SAR > 7/888$	$SAR \leq 12/313$	C4-S3	۱	۳۵	۹۴/۵۹	۱۰/۲
۷	$Cation > 22/63$	$SAR > 12/313$	-	C4-S4	۲	۵	۵۷/۱۴	۷۶/۲

جدول ۲ نتایج ارزیابی نمونه‌های آموزشی در دشت خانمیرزا را نشان می‌دهد که در این مدل تصمیم درختی ۷ قانون جهت طبقه‌بندی استفاده شده است. قانون ۱ بیان می‌کند که "اگر  $EC \leq 750$  باشد، آنگاه کلاس کیفیت آب C2-S1 خواهد بود." از کل نمونه‌های آموزشی ۲۲۲ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۱ برابر با ۹۹/۱٪ و مقدار عددی آماره لیفت برابر با ۱/۸ می‌باشد. پایین بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً زیاد قانون ۱ می‌باشد.

قانون ۲ بیان می‌کند که "اگر  $EC > 750$  باشد، آنگاه کلاس کیفیت آب C3-S1 خواهد بود." از کل نمونه‌های آموزشی ۱۷۸ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۲ برابر با ۵۵/۵٪ و مقدار عددی آماره لیفت برابر با ۲/۳ می‌باشد. پایین بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً زیاد قانون ۲ می‌باشد. قانون ۳ بیان می‌کند که "اگر  $Na > 9/13$  و  $Cation \leq 22/63$  و  $SAR \leq 10/589$  باشد، آنگاه کلاس کیفیت آب C3-S2 خواهد بود." از کل نمونه‌های آموزشی ۲۸ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۳ برابر با ۹۶/۶٪ و مقدار عددی آماره لیفت برابر با ۱۱/۸ می‌باشد. پایین بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً کم قانون ۳ می‌باشد. قانون ۴ بیان می‌کند که "اگر  $Na > 7/05$  و  $Cation \leq 22/63$  و  $SAR > 4/903$  باشد، آنگاه کلاس کیفیت آب C3-S2 خواهد بود." از کل نمونه‌های آموزشی ۳۳ نمونه با شرایط این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۴ برابر با ۸۵/۷۱٪ و مقدار عددی آماره لیفت برابر با ۱۰/۴ می‌باشد. بالا بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً کم قانون ۴ می‌باشد. قانون ۵ بیان می‌کند که "اگر  $Cation > 22/63$  و  $SAR \leq 7/888$  باشد، آنگاه کلاس کیفیت آب C4-S2 خواهد بود." از کل نمونه‌های آموزشی ۲۶ نمونه با شرایط این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۵ برابر با ۹۲/۸۵ درصد و مقدار عددی آماره لیفت برابر با ۱۴/۳ می‌باشد. بالا بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً کم قانون ۵ می‌باشد. قانون ۶ بیان می‌کند که "اگر  $Cation > 22/63$  و  $SAR > 7/888$  و  $SAR \leq 12/313$  باشد، آنگاه کلاس کیفیت آب C4-S3 خواهد بود." از کل نمونه‌های آموزشی ۳۵ نمونه با

جدول ۲ نتایج ارزیابی نمونه‌های آموزشی در دشت خانمیرزا را نشان می‌دهد که در این مدل تصمیم درختی ۷ قانون جهت طبقه‌بندی استفاده شده است. قانون ۱ بیان می‌کند که "اگر  $EC \leq 750$  باشد، آنگاه کلاس کیفیت آب C2-S1 خواهد بود." از کل نمونه‌های آموزشی ۲۲۲ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۱ برابر با ۹۹/۱٪ و مقدار عددی آماره لیفت برابر با ۱/۸ می‌باشد. پایین بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً زیاد قانون ۱ می‌باشد.

قانون ۲ بیان می‌کند که "اگر  $EC > 750$  باشد، آنگاه کلاس کیفیت آب C3-S1 خواهد بود." از کل نمونه‌های آموزشی ۱۷۸ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۲ برابر با ۵۵/۵٪ و مقدار عددی آماره لیفت برابر با ۲/۳ می‌باشد. پایین بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً زیاد قانون ۲ می‌باشد. قانون ۳ بیان می‌کند که "اگر  $Na > 9/13$  و  $Cation \leq 22/63$  و  $SAR \leq 10/589$  باشد، آنگاه کلاس کیفیت آب C3-S2 خواهد بود." از کل نمونه‌های آموزشی ۲۸ نمونه با شرایط این قانون همخوانی دارد. دقت

از جدول ۴ ملاحظه می‌شود که از کل مجموع ۱۴۷ مورد نمونه آزمایشی ۶ مورد از نمونه‌ها خارج از قطر اصلی ماتریس اغتشاش می‌باشند. از این ۶ مورد، ۵ نمونه به‌جای اینکه در کلاس C1-S1 قرار گیرد، به‌طور اشتباه در کلاس C2-S1 و ۱ نمونه به‌جای اینکه در کلاس C3-S2 قرار گیرد، در کلاس C3-S1 واقع شده‌اند. بنابراین خطای ناشی از داده‌های بخش آزمایشی مدل برابر با ۴/۱ درصد خواهد بود.

با توجه به این‌که در آب زیرزمینی این دشت بیش‌ترین پراکنش کلاس کیفیت مربوط به کلاس C2-S1 می‌باشد، بنابراین این کلاس به‌عنوان کلاس پیش‌فرض در نظر گرفته شده است. همچنین با توجه به نتایج به‌دست آمده از مدل تصمیم‌درختی C5.0 برای آب زیرزمینی آبخوان خانمیرزا در جدول ۲ مشاهده می‌شود که این مدل برای پیش‌بینی کیفیت آب زیرزمینی تنها از پارامترهای SAR، EC، NA و Cation استفاده کرده است. به عبارت دیگر، از بقیه پارامترهای هیدروشیمیایی جهت توسعه مدل بهره‌نگرفته است. لذا برای تخمین کلاس کیفیت آب زیرزمینی در دشت خانمیرزا برای مصارف کشاورزی، اندازه‌گیری چهار پارامتر مذکور کافی می‌باشد. همچنین از بین ۷ قانون مورد بررسی، قانون یک بهترین و مناسب‌ترین قانون است، زیرا برای طبقه‌بندی کیفیت آب زیرزمینی تنها به پارامتر EC نیاز دارد و مقدار دقت لاپلاس در این قانون بالا و مقدار آماره لیفت پایین می‌باشد. بنابراین با استفاده از این مدل نه تنها در صرف زمان و هزینه صرفه‌جویی می‌شود، بلکه می‌توان با داشتن تعداد پارامتر کیفی کم در این دشت در کمترین زمان به پیش‌بینی کلاس کیفی آب زیرزمینی پرداخت و به نتایج قابل قبولی دست یافت.

با توجه به جدول ۵ که خلاصه‌ای از نتایج مدل C5.0 را نشان می‌دهد، مشاهده می‌شود که مدل تصمیم‌گیری درختی در آبخوان خانمیرزا تنها با استفاده از ۴ پارامتر هیدروشیمیایی (EC، Na، SAR و Cation) و ۷ قانون و با دقت ۹۵/۵ درصد می‌تواند طبقه کیفیت آب را با دقت بسیار بالایی تعیین نماید.

این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۶ برابر با ۹۴/۵۹٪ و مقدار عددی آماره لیفت برابر با ۱۰/۲ می‌باشد. بالا بودن نسبی مقدار آماره لیفت نشانگر دقت نسبتاً کم قانون ۶ می‌باشد. قانون ۷ بیان می‌کند که "اگر  $Cation > 22/63$  و  $SAR > 12/313$  باشد، آنگاه کلاس کیفیت آب C4-S4 خواهد بود". از کل نمونه‌های آموزشی ۵ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۷ برابر با ۵۷/۱۴٪ و مقدار عددی آماره لیفت برابر با ۷۶/۲ می‌باشد. بالا بودن مقدار آماره لیفت نشانگر دقت کم قانون ۷ می‌باشد. همچنین ماتریس اغتشاش که بیانگر نحوه پراکنش نمونه‌ها بین کلاس‌های مختلف می‌باشد و دقت هر دسته را بیان می‌کند، برای آب زیرزمینی دشت خانمیرزا در جدول ۳ ارائه گردیده است. در ماتریس اغتشاش نمونه‌هایی که روی قطر اصلی ماتریس واقع شده‌اند، نمونه‌هایی هستند که درست و صحیح پیش‌بینی شده‌اند. در مورد آب زیرزمینی دشت خانمیرزا ۸ نمونه آموزشی در خارج از قطر اصلی واقع شده است. از این ۸ مورد ۱ نمونه به‌جای اینکه در کلاس C1-S1 قرار گیرد، به اشتباه در کلاس C3-S2 و ۳ نمونه به‌جای اینکه در کلاس C3-S3 قرار گیرند، در کلاس C3-S2 و ۱ نمونه به‌جای اینکه در کلاس C4-S2 قرار گیرد، در کلاس C4-S3 و ۱ نمونه به‌جای اینکه در کلاس C4-S3 قرار گیرد، در کلاس C4-S1 و ۲ نمونه به‌جای اینکه در کلاس C4-S3 قرار گیرند، به اشتباه در کلاس C4-S4 قرار گرفته‌اند و باعث شده که خطای این بخش ۲/۰ درصد باشد. پس از آموزش مدل و استخراج قوانین مربوطه برای بررسی صحت نتایج، مدل با ۱۴۷ مورد نمونه مورد آزمون قرار گرفت. نتایج پراکنش نمونه‌های آزمایشی در بین کلاس‌های مختلف و ماتریس اغتشاش مربوط به نمونه‌های آزمایشی در جدول ۴ ارائه گردیده است. همان‌طور که قبلاً نیز ذکر گردید نمونه‌هایی که در روی قطر اصلی ماتریس اغتشاش واقع شده‌اند، بیانگر نمونه‌های هستند که درست پیش‌بینی شده‌اند و آن تعداد از نمونه‌هایی که در خارج از قطر اصلی واقع شده‌اند به‌صورت نادرست و غلط پیش‌بینی شده‌اند.



جدول ۳- ماتریس اغتشاش برای نمونه‌های آموزش آب زیرزمینی دشت خانمیرزا.

C1-S1	C1-S2	C1-S3	C1-S4	C2-S1	C2-S2	C2-S3	C2-S4	C3-S1	C3-S2	C3-S3	C3-S4	C4-S1	C4-S2	C4-S3	C4-S4	کلاس
									۱							C1-S1
																C1-S2
																C1-S3
																C1-S4
			۲۲۲													C2-S1
																C2-S2
																C2S3
																C2S4
								۷۵								C3-S1
									۳۳							C3-S2
										۳						C3-S3
																C3-S4
																C4S1
													۲۵	۱		C4S2
													۱	۳۴	۲	C4S3
															۳	C4S4

جدول ۴- ماتریس اغتشاش برای نمونه‌های تست آب زیرزمینی دشت خانمیرزا.

C1-S1	C1-S2	C1-S3	C1-S4	C2-S1	C2-S2	C2-S3	C2-S4	C3-S1	C3-S2	C3-S3	C3-S4	C4-S1	C4-S2	C4-S3	C4-S4	کلاس
				۵												C1-S1
																C1-S2
																C1-S3
																C1-S4
				۱۴۰												C2-S1
																C2-S2
																C2S3
																C2S4
								۱								C3-S1
								۱								C3-S2
																C3-S3
																C3-S4
																C4S1
																C4S2
																C4S3
																C4S4

جدول ۵- خلاصه نتایج مدل تصمیم درختی C5.0 برای طبقه‌بندی کیفیت آب زیرزمینی آبخوان خانمیرزا.

تعداد قانون	تعداد پارامترهای هیدروشیمیایی	آموزش		آزمون		Default Class (کلاس پیش‌فرض)
		خطا برحسب درصد	دقت مدل بر حسب درصد	خطا برحسب درصد	دقت مدل بر حسب درصد	
۷	۳	۲/۰	۹۸	۴/۱	۹۵/۵	C2-S1

## نتیجه‌گیری

در این مطالعه، به منظور پیش‌بینی کلاس کیفی آب زیرزمینی دشت خانمیرزا بر اساس دیاگرام USSL، با تعداد پارامتر شیمیایی کمتر و بارش ماهانه و بارش تجمعی از روش تصمیم‌گیری درختی استفاده گردید. نتایج طبقه‌بندی نمونه‌های کیفی آب زیرزمینی این دشت نشان داد که بیش‌ترین فراوانی کلاس کیفیت نمونه‌های آب مورد بررسی، مربوط به کلاس C2-S1 می‌باشد. همچنین نتایج به‌دست آمده از مدل تصمیم‌گیری درختی نشان داد که این مدل در آبخوان خانمیرزا تنها با استفاده از ۴ پارامتر هیدروشیمیایی (Na, EC, SAR و Cation) و ۷ قانون و با دقت ۹۵/۵ درصد می‌تواند طبقه کیفیت آب را با دقت بسیار بالایی تعیین نماید. بررسی نشان داد که بارش در طول دوره مورد مطالعه، تأثیر قابل توجهی بر تغییرات کیفیت آب زیرزمینی این دشت نداشته است. بنابراین، مشخص شد که تنها با ۴ پارامتر کیفی و با استفاده از روش تصمیم‌گیری می‌توان از کیفیت آب زیرزمینی این دشت برای مصارف کشاورزی آگاهی یافت، به‌طوری که در آینده اندازه‌گیری این ۴ پارامتر برای تشخیص کلاس کیفیت کافی می‌باشد که این خود باعث کاهش هزینه‌ها و زمان تحلیل آزمایش‌ها می‌گردد. نتایج محققان دیگر که از تصمیم‌گیری درختی استفاده کرده‌اند نیز حاکی از مفید و مناسب بودن این روش می‌باشد. به‌طور مثال ثاقبیان و همکاران (۲۰۱۴) کیفیت آب زیرزمینی دشت اردبیل را با روش تصمیم‌گیری درختی طبقه‌بندی کردند و نتایج آن‌ها حاکی از این بود که این روش توانست کلاس کیفیت را تنها با دو پارامتر هدایت الکتریکی (EC) و بارش تجمعی ۱۱ ماه قبل به‌خوبی تعیین کند. پیشنهاد می‌شود که در مطالعات آتی کارایی روش‌های دیگر مبتنی بر داده همچون روش ماشین بردار پشتیبان در پیش‌بینی کیفیت

آب زیرزمینی این دشت مورد بررسی قرار گرفته و با نتایج مطالعه حاضر مقایسه شود.

## منابع

- ثاقبیان، م. ستاری، م. ت. و میرعباسی نجف‌آبادی، ر. ۱۳۹۲. طبقه‌بندی کیفیت آب زیرزمینی آبخوان دشت اردبیل با استفاده از روش تصمیم‌گیری درختی، گزارش طرح تحقیقاتی، دانشگاه آزاد اسلامی واحد اهر.
- ظهیری، ع. و قربانی، خ. ۱۳۹۲. شبیه‌سازی دبی جریان در مقاطع مرکب به کمک مدل درخت تصمیم M5 1. نشریه پژوهش‌های حفاظت آب و خاک، ۲۰(۳): ۱۱۳-۱۳۲.
- محبوبی، ع. و تجریشی، م. ۱۳۸۹. مقایسه عملکرد الگوریتم‌های شبکه عصبی مصنوعی و درختان تصمیم‌گیری در پیش‌بینی تغییرات شوری آب رودخانه‌ها (مطالعه موردی: رودخانه کارون)، چهارمین همایش تخصصی مهندسی محیط زیست، ۱۰-۱۱ آبان، دانشکده محیط‌زیست دانشگاه تهران، تهران.
- Akyildiz, I.F. Su, W. Sankarasubramaniam, Y. Cayirci, E. 2002. Wireless sensor networks: a survey, *Computer Networks*, 38: 393-422.
- Eisenberg, J.N.S. and Mc Kone, T.E. 1998. Decision tree method for the classification of chemical pollutants: incorporation of across-chemical variability and within-chemical uncertainty, *Environmental Science and Technology*, 32: 3396-3404.
- Litaor, M.I. Briemann, H. Reichmann, O. and Shenker, M. 2010. Hydrochemical analysis of groundwater using a tree-based model. *Journal of Hydrology* 387: 273-282.
- Quinlan, J.R. 1993. C4.5 Programs for machine learning, Morgan, Kaufmann, San Mateo, California.
- Quinlan, J.R. 2000. Data mining tools See5 and C5.0 [cited Feb 2012]. Available from <http://www.rulequest.com/see5-info.html>.
- Rasekh, A. and Brumbelow, K. 2014. Drinking water distribution systems contamination management to reduce public health impacts and system service interruptions. *Journal Environmental Modelling & Software*, 51: 12-25.
- Saghebain, S.M. Sattari, M.T. Mirabbasi, R. Pal, M. 2014. Ground water quality classification by decision tree method in Ardebil region, Iran, *Arabian Journal of Geosciences*, 7: 4746-4777.

- Tapak, L. Rahmani, A.R. and Moghimbeigi, A. 2014. Prediction the groundwater level of Hamadan-Bahar Plain, west of Iran using Support Vector Machines, *Journal of Research in Health Sciences*, 14(1): 81-86.
- Witten, I.H. Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, Second edition, Elsevier: San Francisco, ISBN 0-12-088407-0.
- Sattari, M.T., Mahesh P. and Apaydin, H. 2013. M5 model tree application in Daily River flow forecasting in Sohu Stream, Turkey, *Water Resources*, 4(3): 233-242.
- Spruill, T.B. Showers, W.J. and Howe, S.S. 2002. Application of classification tree methods to identify nitrate sources in ground water. *Environmental Quality*, 31: 1538-1549.