

## استفاده از داده‌کاوی در پیش‌بینی کیفیت آب‌های سطحی (مطالعه موردی: رودخانه‌های دامنه شمالی سهند)

محمدتقی ستاری<sup>۱</sup>، رسول میرعباسی نجف‌آبادی<sup>۲\*</sup>، مهدی عباسقلی نایب‌زاد<sup>۳</sup>

۱. استادیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۲. استادیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شهرکرد

۳. کارشناسی ارشد مهندسی عمران، دانشگاه آزاد مراغه

(تاریخ دریافت ۱۳۹۵/۱۰/۰۵؛ تاریخ تصویب ۱۳۹۵/۱۲/۲۵)

### چکیده

پایش و ارزیابی کیفی آب‌های سطحی فرایندی بسیار پرهزینه و زمان‌بر است. بنابراین، انتخاب روشی که در آن با کمترین پارامترهای هیدروشیمیایی بتوان پیش‌بینی نسبتاً دقیقی از طبقه کیفیت آب داشت، مهم و ضروری است. تصمیم‌گیری درختی به‌عنوان یکی از روش‌های داده‌کاوی با بهره‌گیری از یک ساختار درختی به طبقه‌بندی داده‌ها می‌پردازد. در این مقاله، با استفاده از روش تصمیم‌گیری درختی کیفیت آب برخی از رودخانه‌های واقع در دامنه‌های شمالی کوه سهند در محل ایستگاه‌های هیدرومتری بستان‌آباد، پل سنیخ، ليقوان و ونيار بررسی شد و برای هر یک از رودخانه‌ها طبقه کیفیت آب با استفاده از قوانین اگر-آن‌گاه توسعه داده شد. برای هر یک از رودخانه‌ها دبی و ۱۲ پارامتر هیدروشیمیایی شامل یون‌های کلسیم ( $Ca^{2+}$ )، منیزیم ( $Mg^{2+}$ )، کلر ( $Cl^-$ )، بی‌کربنات ( $HCO_3^-$ )، درصد سدیم ( $Na\%$ )، اسیدیته (pH)، سولفات ( $SO_4^{2-}$ )، مجموع آنیون‌ها (Sum A)، مجموع کاتیون‌ها (Sum C)، کل نمک‌های محلول (TDS)، نسبت جذب سدیم (SAR) و هدایت الکتریکی (EC) برای ایجاد مدل درختی بررسی شد. نتایج نشان داد مدل تصمیم‌گیری درختی بیشتر با استفاده از چهار پارامتر EC، pH، SAR و  $Na^+$  قادر است طبقه کیفیت آب را با دقت بسیار زیادی مشخص کند، به طوری که میزان خطای مدل توسعه داده‌شده در بخش آزمون برای ایستگاه‌های بستان‌آباد، ونيار، پل سنیخ و ليقوان به ترتیب برابر ۳/۴، ۸/۱، ۲۲/۹ و ۱/۶ درصد بود.

**کلیدواژگان:** پارامترهای هیدروشیمیایی، پایش کیفی، داده‌کاوی، درخت تصمیم، کیفیت آب‌های سطحی.

## مقدمه

ارزیابی کیفیت آب‌های سطحی یکی از موارد بسیار مهم در توسعه اراضی کشاورزی، طراحی و بهره‌برداری سیستم‌های آبیاری و انتخاب الگوی کشت در هر منطقه است. برای ارزیابی کیفیت آب‌های سطحی در رودخانه، در بازه‌های زمانی مختلف از آب رودخانه نمونه‌برداری می‌شود. پس از انتقال نمونه‌ها به آزمایشگاه، غلظت یون‌های موجود در آب، هدایت الکتریکی، pH و... اندازه‌گیری و گزارش می‌شود. سپس براساس مقاصد مختلف، کیفیت آب طبقه‌بندی می‌شود. استفاده از دیاگرام‌ها، به دلیل سهولت، یکی از ابزارهای متداول طبقه‌بندی کیفیت آب است. براساس مصارف مختلف آب، دیاگرام‌های مختلفی برای طبقه‌بندی کیفی آب ارائه شده است. یکی از دیاگرام‌هایی که برای طبقه‌بندی آب آبیاری ارائه شده، دیاگرام<sup>۱</sup> USSL است که به‌طور گسترده در همه نقاط دنیا استفاده می‌شود [۱]. محدودیت‌های اساسی در تعیین کیفیت آب‌های سطحی شامل هزینه‌های زیاد آزمایش‌های هیدروشیمیایی در اندازه‌گیری تعداد زیادی پارامتر و نیز فاصله زمانی زیاد بین نمونه‌برداری و گرفتن نتایج از آزمایشگاه است. برای حل این مشکلات می‌توان از مدل‌هایی به‌منظور طبقه‌بندی و پیش‌بینی کیفیت آب سطحی براساس مشاهدات تاریخی و نتایج برگرفته از طبقه‌بندی دیاگرام USSL استفاده کرد. پیشرفت‌های جدید در شبیه‌های مدل‌سازی مبتنی بر روش‌های داده‌کاوی نظیر تصمیم‌گیری درختی جایگزین مناسبی برای طبقه‌بندی کیفیت منابع آب و دیگر موضوعات مرتبط با آب است. روش‌های داده‌کاوی با استفاده از تکنیک‌های متعدد نسبت به آموزش داده‌ها برای استخراج روابط، الگوها، قواعد و نظم پنهان درون داده‌ها اقدام می‌کند. درخت تصمیم به‌عنوان یکی از روش‌های طبقه‌بندی و پیش‌بینی قادر است با استفاده از داده‌های مشاهداتی تاریخی، قواعد اگر - آن‌گاه را طوری تولید کند که بتوان طبقه‌بندی کیفیت آب را پیش‌بینی کرد. استفاده از روش‌های داده‌کاوی سبب کاهش چشمگیر هزینه‌های نمونه‌برداری و آزمایشگاهی می‌شود و این امکان را برای کارشناسان مهیا خواهد کرد تا در مدت خیلی کوتاه و با استفاده از تعداد کمتری پارامتر هیدروشیمیایی، کلاس کیفیت آب را تعیین کنند [۲]. در سال‌های اخیر، با گسترش

پایگاه داده‌ها و امکان اندازه‌گیری‌های کمی و کیفی منابع آب، روش‌های داده‌کاوی نیز مورد توجه متخصصان علوم آب قرار گرفته است.

سانتوس و همکارانش [۳] از یک چارچوب درختی برای انتخاب پارامترهای بیوشیمیایی بررسی شده در سنجش از دور کیفیت آب سیلاب‌ها استفاده کردند. میرعباسی و همکارانش [۲] با استفاده از منطق فازی به ارزیابی کیفیت آب آبیاری در دشت سیرجان اقدام کردند. یحیی و همکارانش [۴] با استفاده از روش تحلیل رگرسیون، گسترش آلودگی و تغییرات فصلی کیفیت آب رودخانه ایندوس در پاکستان را در دوره‌های پیش و پس از بارش‌های موسمی سال‌های ۲۰۰۸ و ۲۰۰۹ بررسی کردند. نتایج نشان داد با استفاده از روش رگرسیون می‌توان گسترش آلودگی و تغییرات فصلی را مدل کرد و این روش سبب صرفه‌جویی در زمان و هزینه‌های آزمایشگاهی می‌شود. رحمانی و همکارانش [۵] کیفیت آب رودخانه‌های جاری در دشت همدان - بهار را برای مقاصد آبیاری بر مبنای دیاگرام ویلکاکس مطالعه کردند. بدین‌منظور پس از شناسایی سرشاخه‌های رودخانه‌ها، یون‌های سدیم، کلسیم، منیزیم، اسیدپته (pH) و هدایت الکتریکی (EC) اندازه‌گیری و نسبت جذب سدیم (SAR) محاسبه شد. سپس با استفاده از دیاگرام ویلکاکس کیفیت آب رودخانه‌ها ارزیابی شد. نتایج نشان داد کیفیت آب با حرکت به سمت پایین‌دست مسیر رودخانه کاهش می‌یابد. گلجان و همکارانش [۶] برای بررسی و طبقه‌بندی کیفیت آب رودخانه‌های شهرستان نور از دیاگرام پایپر استفاده کردند. آنها همچنین تأثیرات زیست‌محیطی را بر کیفیت آب رودخانه‌ها بررسی کردند. علیایی و همکارانش [۷] در تحقیقی از مدل شبکه عصبی مصنوعی پرسپترون چندلایه برای مدل‌سازی شاخص‌های کیفی آب رودخانه مرادبیک در همدان استفاده کردند. شاخص‌های کیفی بررسی‌شده شامل اکسیژن مورد نیاز بیولوژیکی (BOD) و اکسیژن محلول (DO) بود. نتایج، کارایی مدل شبکه عصبی مصنوعی را به‌عنوان تکنیکی برتر برای شبیه‌سازی تغییرات شاخص‌های کیفی آب نشان داد. حاجیان‌نژاد و رهسپار [۸] در مطالعه‌ای تأثیر رواناب‌ها و سبب تصفیه‌خانه فاضلاب را بر کیفیت آب رودخانه زاینده‌رود بررسی کردند. نتایج پژوهش آنها نشان داد کیفیت آب رودخانه زاینده‌رود حتی در شرایطی که بارندگی صورت نمی‌گیرد، از رواناب‌های شهر اصفهان متأثر است؛ اما شدت

هدف از این پژوهش، طبقه‌بندی کیفیت آب‌های سطحی در ایستگاه‌های هیدرومتری استان آباد، پل سنخ، ليقوان و ونيار واقع در دامنه‌های شمالی کوه سهند با استفاده از روش تصمیم‌گیری درختی و توسعه قوانین اگر-آن‌گاه برای انجام طبقه‌بندی کیفیت آب برای مصارف کشاورزی بود.

### مواد و روش‌ها

#### منطقه مطالعه‌شده و داده‌های استفاده‌شده

منطقه مطالعه‌شده دامنه‌های شمالی کوه سهند بود که در حوضه دریاچه ارومیه واقع شده است. موقعیت جغرافیایی این منطقه بین ۳۷ درجه و ۴۵ دقیقه تا ۳۸ درجه و ۴۵ دقیقه عرض شمالی و ۴۶ درجه تا ۴۷ درجه طول شرقی واقع شده است. در شکل ۱ موقعیت جغرافیایی منطقه مطالعه‌شده در حوضه دریاچه ارومیه به همراه رودخانه‌ها و محل ایستگاه‌های هیدرومتری نشان داده شده است. در این مطالعه، به‌منظور ارزیابی و پیش‌بینی کیفیت آب رودخانه‌ها از داده‌های هیدروشمیایی چهار ایستگاه هیدرومتری شامل استان آباد، پل سنخ، ليقوان و ونيار استفاده شد. مشخصات ایستگاه‌های هیدرومتری منتخب و برخی آماره‌های سری جریان سالانه آنها در جدول ۱ ارائه شده است. داده‌های استفاده‌شده پارامترهای هیدروشمیایی و دبی ماهانه رودخانه‌ها در ایستگاه‌های بررسی شده بود. این داده‌ها از شرکت آب منطقه‌ای استان آذربایجان شرقی گرفته و درست و نادرستی آنها بررسی شد. با توجه به اینکه در ایجاد مدل تصمیم‌گیری درختی ترتیب وقوع داده‌ها اهمیت ندارد، در ایجاد مدل فقط از داده‌های ماه‌هایی استفاده شد که کلیه پارامترهای هیدروشمیایی آنها کامل بود و چنانچه داده‌های دبی در این ماه‌ها ناقص بود، با روش ایجاد همبستگی با ایستگاه‌های مجاور بازسازی شد.

#### داده‌کاوی و روش تصمیم‌گیری درختی

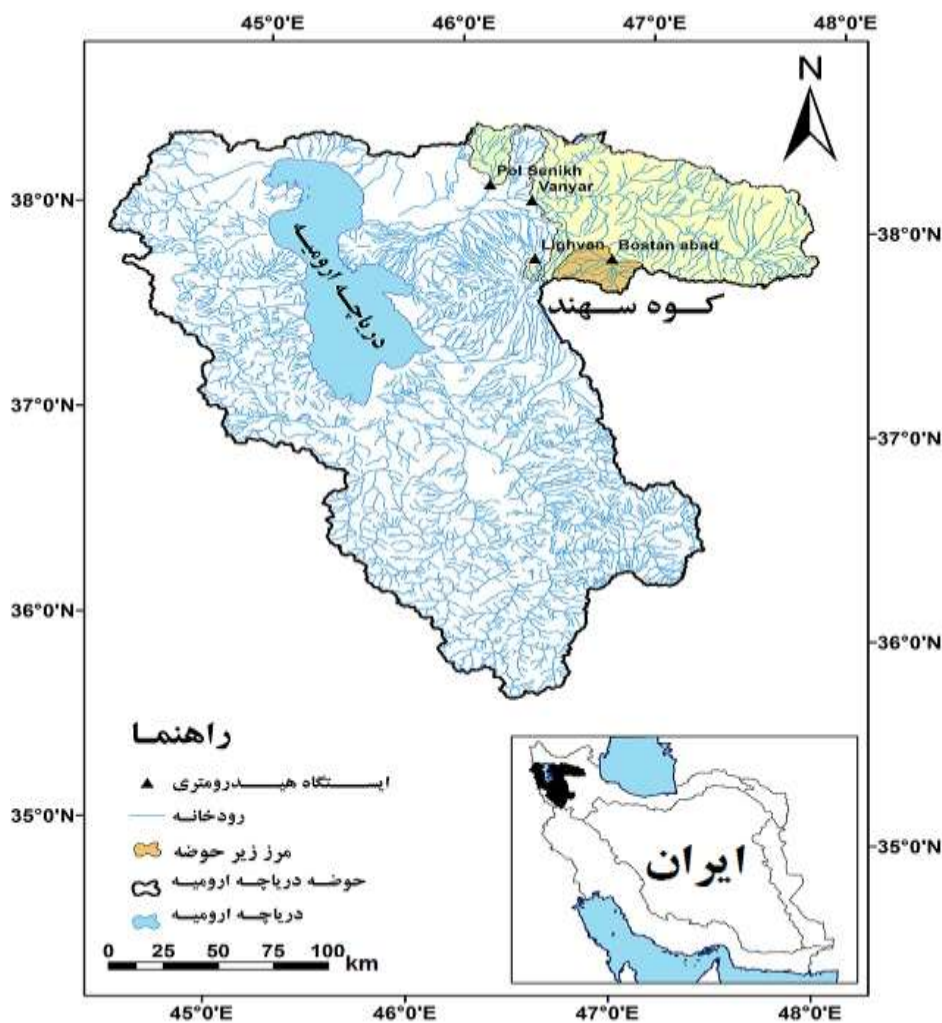
داده‌کاوی به بررسی و تجزیه و تحلیل مجموعه بزرگی از داده‌ها به‌منظور کشف الگوها و قوانین پنهان و معنادار درون داده‌ها گفته می‌شود. داده‌کاوی به دو نوع هدایت‌شده و هدایت‌نشده<sup>۲</sup> تقسیم می‌شود. داده‌کاوی هدایت‌شده، متغیر هدف خاص و از پیش تعیین‌شده دارد

تأثیر پساب تخلیه‌شده از تصفیه‌خانه فاضلاب جنوب اصفهان بسیار زیادتر است. سلاجقه و همکارانش [۹] اثر تغییرات کاربری اراضی و آثار آن بر کیفیت آب رودخانه‌های حوضه آبخیز کرخه را مطالعه کردند. نتایج پژوهش آنها نشان داد تغییرات کاربری اراضی در این حوضه سبب کاهش شدید کیفیت آب رودخانه‌ها شده است. ثاقبیان و همکارانش [۱۰] یک شیوه جدید مبتنی بر تصمیم‌گیری درختی برای طبقه‌بندی و پیش‌بینی کیفیت آب زیرزمینی مبتنی بر دیاگرام USSL با استفاده از داده‌های آبخوان اردبیل پیشنهاد کردند. ارزیابی عملکرد روش پیشنهادی بیانگر مناسب بودن شیوه طبقه‌بندی مبتنی بر تصمیم‌گیری درختی برای مجموعه داده‌های استفاده‌شده بود. نتایج تحقیق آنها نشان داد طبقه‌بندی کیفیت آب زیرزمینی توسط تصمیم‌گیری درختی دقیق‌تر و کارتر از روش مؤلفه‌های اصلی است. روش پیشنهادی توانست کلاس کیفیت آب زیرزمینی دشت اردبیل را فقط با دو پارامتر هدایت الکتریکی (EC) و بارش تجمعی ۱۱ ماه قبل (CP11) به‌خوبی تعیین کند. مدل توسعه داده‌شده قادر به پیش‌بینی کلاس کیفیت آب فقط براساس دو متغیر بود و همین سبب کاهش تعداد متغیرهای مورد نیاز در تعیین کلاس کیفیت آب و در نتیجه کاهش چشم‌گیر هزینه‌های آزمایشگاهی و زمان تلف‌شده بین نمونه‌برداری و دریافت نتایج از آزمایشگاه می‌شود.

حسینی و همکارانش [۱۱] طبقه کیفیت آب زیرزمینی دشت خان‌میرزا در استان چهارمحال و بختیاری را براساس دیاگرام USSL با استفاده از روش تصمیم‌گیری درختی پیش‌بینی کردند. نتایج تحقیق آنها نشان داد روش تصمیم‌گیری درختی قادر است فقط با استفاده از چهار پارامتر هیدروشمیایی مجموع کاتیون‌ها، SAR، Na<sup>+</sup> و EC طبقه کیفیت آب این دشت را با دقت بسیار زیادی تعیین کند. نوروزی و همکارانش [۱۲] به‌منظور بررسی آلودگی آب زیرزمینی دشت ملکان به آرسنیک از روش جنگل تصادفی<sup>۱</sup> (RF) استفاده کردند که یک روش یادگیری مبتنی بر دسته‌ای از درختان تصمیم است. براساس پیش‌بینی مدل RF، مقدار قابلیت انتقال، نیترات، هدایت هیدرولیکی و وجود شهرها به‌عنوان تأثیرگذارترین پارامترها در وجود آنومالی آرسنیک شناخته شدند.

موارد داده‌ها از منابع متفاوتی جمع‌آوری می‌شوند که باید به‌صورت پایگاهی از داده‌های مناسب درآیند تا عملیات داده‌کاوی بهتر انجام شود؛ ۴. انتخاب مجموعه متغیرهای هدف؛ ۵. یافتن ویژگی‌های استفاده‌شده و تعیین ویژگی‌های جدید؛ ۶. نمایش داده‌ها به‌صورتی که بتوان در داده‌کاوی استفاده کرد؛ ۷. انتخاب عملیات داده‌کاوی (طبقه‌بندی، خوشه‌بندی، پیش‌بینی و غیره)؛ ۸. انتخاب روش داده‌کاوی (شبکه‌های عصبی، درخت تصمیم و نظایر آن)؛ ۹. انجام داده‌کاوی و جست‌وجو برای یافتن الگوی مناسب؛ ۱۰. ارزیابی و تحلیل الگوی به‌دست‌آمده و حذف الگوهای نامناسب؛ ۱۱. تفسیر نتایج داده‌ها و استنتاج اطلاعات با ارزش [۱۳].

که به دنبال الگویی خاص می‌گردد، در حالی که هدف داده‌کاوی هدایت‌نشده، یافتن الگوها یا تشابهات بین گروه‌هایی از اطلاعات، بدون داشتن متغیر هدف خاص و یا مجموعه‌ای از دسته‌ها و الگوهای از پیش تعیین‌شده است [۱۳]. داده‌کاوی بیشتر 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 می‌کند. یک مدل تحت شرایط درست می‌تواند به بینش درست منجر شود. داده‌کاوی شامل مراحل مختلفی است که عبارت‌اند 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



شکل ۱. موقعیت جغرافیایی حوضه دریاچه ارومیه و زیرحوضه‌های ایستگاه‌های هیدرومتری منتخب

جدول ۱. مشخصات ایستگاه‌های هیدرومتری مطالعه‌شده و برخی آماره‌های سری جریان سالانه آنها (دوره آماری ۱۳۴۸-۱۳۹۰)

نام رودخانه	نام ایستگاه	عرض جغرافیایی دقیقه-درجه	طول جغرافیایی دقیقه-درجه	ارتفاع (متر)	میانگین دبی سالانه (m <sup>3</sup> /s)	انحراف معیار (m <sup>3</sup> /s)	ضریب تغییرات	ضریب چولگی
اوجان‌چای	بستان‌آباد	۳۷-۵۰	۴۶-۵۰	۱۷۰۰	۱/۷۴	۰/۸۰	۰/۴۶	۰/۳۲
لیقوان‌چای	لیقوان	۳۷-۵۰	۴۶-۲۶	۲۲۰۰	۰/۷۸	۰/۲۲	۰/۳۸	۰/۶۴
آجی‌چای	ونیار	۳۸-۰۷	۴۶-۲۶	۱۴۵۰	۱۲/۷۸	۶/۷۷	۰/۵۳	۰/۸۷
سنیخ‌چای	پل سنیخ	۳۸-۱۱	۴۶-۰۹	۱۳۸۰	۱/۰۹	۱/۴۰	۱/۲۸	۲/۴۳

هدف C مقدار مختلف داشته باشد، آنتروپی S نسبت به این دسته‌بندی C گانه به صورت زیر تعریف می‌شود [۱۴]:

$$\text{Entropy}(S) = \sum_{i=1}^C -p_i \log p_i \quad (1)$$

که در آن،  $p_i$  نسبتی از S است که به دسته i تعلق دارد. توجه شود که لگاریتم بر مبنای ۲ در نظر گرفته می‌شود. در این حالت بیشترین 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 بیان دیگر، بهره اطلاعات Gain(S,A) برای یک ویژگی نظیر A نسبت به مجموعه مثال‌های S به صورت زیر تعریف می‌شود [۱۴]:

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

که در آن، Values(A) مجموعه‌ای از مقادیر ممکن برای ویژگی‌های A بوده و  $S_v$  زیرمجموعه‌ای از S است که در آن ویژگی A دارای مقدار v است. در این پژوهش، برای توسعه مدل درختی کیفیت آب سطحی و استخراج قواعد اگر-آن‌گاه از نرم‌افزار See5 استفاده شد که مبتنی بر الگوریتم C5.0 بوده و توسط کوپنلن [۱۵] توسعه داده شده است.

در این مطالعه، به منظور ارزیابی دقت قواعد از دو آماره لاپلاس و لیفت استفاده شد. اگر N تعداد موارد صحیح طبقه‌بندی شده و M تعداد موارد اشتباه طبقه‌بندی شده در یک قانون خاص باشند، دو آماره لاپلاس و لیفت (Lift X) به صورت زیر قابل محاسبه خواهند بود [۱۳]:

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

$$\text{دقت لاپلاس} = (N-M+1)/(N+2) \quad (4)$$

که با قانون همخوانی دارند) Lift X=((N-M+1)/(N+2))

طبقه‌بندی که یکی از کارکردهای معمول داده کاوی است، شامل بررسی ویژگی‌های یک شیء جدید و تخصیص آن به یکی از مجموعه‌های از قبل تعیین شده است. عمل طبقه‌بندی با تعریف درستی از دسته‌ها و مجموعه‌ای از ویژگی‌ها انجام می‌شود که حاوی موارد از پیش دسته‌بندی شده هستند. این عمل شامل ساختن مدلی است که بتواند از آن برای طبقه‌بندی داده‌های دسته‌بندی نشده، استفاده کرد. 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 خاص است. درختان تصمیم‌گیری به منظور پیشگویی یا کلاسه‌بندی داده‌ها براساس مجموعه قوانین تصمیم ایجاد شده‌اند. کلاسه‌بندی داده‌ها با درختان تصمیم‌گیری یک فرایند دومارحله‌ای است. در مرحله اول که به آن آموزش گفته می‌شود، مدلی براساس یک الگوریتم کلاسه‌بندی منطبق با داده کاوی 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<sup>۱</sup> انجام می‌شود. در طول فرایند آموزش الگوریتم درخت تصمیم باید به صورت مکرر مؤثرترین روش برای تقسیم کردن مجموعه رکوردها به فرزندان را بیابد. مرحله هرس برای جلوگیری از پردازش بیش از حد و بزرگ شدن درخت تصمیم که سبب پیچیدگی و افزایش تعداد قوانین اگر-آن‌گاه می‌شود، صورت می‌گیرد [۱۳]. در روش درخت تصمیم معیار استفاده شده برای ایجاد شاخه‌ها و جداسازی، بی‌نظمی یا آنتروپی است. اگر ویژگی

سولفات ( $SO_4^{2-}$ )، مجموع آنیون‌ها (Sum A)، مجموع کاتیون‌ها (Sum C)، کل نمک محلول (TDS)، نسبت سدیم جذبی (SAR) و هدایت الکتریکی (EC) توجه شد. ۱۳ صفت و ویژگی هیدروشیمیایی یادشده به‌عنوان ورودی مدل درخت تصمیم و کلاس کیفیت آب به‌منزله خروجی و ویژگی هدف در نظر گرفته شد. کلاس کیفیت آب که براساس دیاگرام USSL تعیین می‌شود و در اینجا به‌عنوان ویژگی هدف استفاده شد، ماهیتی گسسته دارد. در زیر نتایج به‌دست‌آمده از تحلیل مدل برای تعیین کلاس کیفیت آب برای چهار رودخانه مطالعه شده ارائه شده است. برای هر ایستگاه از بین نمونه‌های بررسی شده ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد برای تست در نظر گرفته شد.

### رودخانه اوجان‌چای

در این رودخانه در مجموع تعداد ۲۳۹ مورد نمونه‌برداری از آب در فواصل زمانی مختلف در محل ایستگاه بستان‌آباد انجام گرفته و طی آن با توجه به پراکنش پارامترهای هیدروشیمیایی کلاس کیفیت آب با استفاده از نمودار USSL تعیین شده است. از بین ۲۳۹ مورد به تفکیک تعداد ۱۸۰ نمونه (۷۵ درصد از کل داده‌ها) برای آموزش مدل و باقی یعنی ۵۹ نمونه (۲۵ درصد از کل داده‌ها) برای آزمون در نظر گرفته شد. خلاصه نتایج مدل تصمیم درختی C5.0 برای ایستگاه بستان‌آباد در جدول ۲ ارائه شده است.

هر چقدر مقدار آماره لاپلاس یک قانون بیشتر و مقدار آماره لیفت آن کمتر باشد، دقت آن قانون بیشتر است. نرخ یا دقت لاپلاس متداول‌ترین و مفیدترین آماره در ارزیابی قوانین مستخرج از درخت تصمیم است [۱۵].

### نتایج و بحث

در این پژوهش، از روش تصمیم‌گیری درختی برای پیش‌بینی کیفیت آب‌های سطحی رودخانه‌های منتخب دامنه شمالی سهند در محل ایستگاه‌های هیدرومتری (بستان‌آباد، پل سنیخ، ليقوان و ونيار) استفاده شد. در این رودخانه‌ها ابتدا با استفاده از دیاگرام USSL و براساس پارامترهای هیدروشیمیایی مؤثر در ارزیابی کیفیت آب، طبقه و کلاس کیفیت آب مشخص شد. در بین چهار ایستگاه مطالعه شده، ایستگاه ليقوان کیفیت آب بهتر و ایستگاه پل سنیخ کیفیت نامطلوب‌تری نسبت به بقیه داشتند. همچنین بهترین کیفیت آب در ایستگاه‌های بررسی‌شده در ماه‌های فروردین و اردیبهشت مشاهده شده و در ماه‌های مرداد و شهریور نامطلوب‌ترین کیفیت آب در این ایستگاه‌ها ثبت شده است که احتمالاً دلیل آن افزایش دبی رودخانه‌های بررسی شده در ماه‌های فروردین و اردیبهشت و کاهش آن در ماه‌های مرداد و شهریور است. بنابراین، در این چهار رودخانه در مجموع برای هر یک از رودخانه‌ها دبی و ۱۲ پارامتر هیدروشیمیایی شامل یون کلسیم ( $Ca^{2+}$ )، یون منیزیم ( $Mg^{2+}$ )، یون کلر ( $Cl^-$ )، بی‌کربنات ( $HCO_3^-$ )، درصد سدیم (Na%)، اسیدیته (pH)،

جدول ۲. خلاصه نتایج مدل تصمیم درختی برای ایستگاه هیدرومتری بستان‌آباد

شماره قانون	اگر		آن‌گاه (طبقه کیفیت آب)	تعداد نمونه صادق غیرصادق /	دقت لاپلاس (درصد) /LIFT
	۱	۲			
۱	$EC \leq 750$	-	C2-S1	۱۱۹/۰	۹۹/۲-۱/۵
۲	$EC > 750$	$Na^+ \leq 8/7$	C3-S1	۵۰/۰	۹۸/۱-۳/۵
۳	$Na^+ > 8/7$	$Na^+ \leq 11$	C3-S2	۶/۱	۷۵/۰-۲۷/۰
۴	$Na^+ > 11$	-	C4-S2	۵/۰	۸۵/۷-۳۰/۹

با ۹۹/۲ درصد و مقدار عددی آماره لیفت برابر با ۱/۵ است. کم‌بودن مقدار آماره لیفت نشان‌دهنده دقت زیاد قانون ۱ است.

قانون ۲ بیان می‌کند که «اگر  $Na^+ \leq 8/7$  و  $EC > 750$  (باشد، آن‌گاه کلاس کیفیت آب C3-S1 خواهد بود». از کل نمونه‌های آموزشی ۵۰ نمونه با این قانون همخوانی

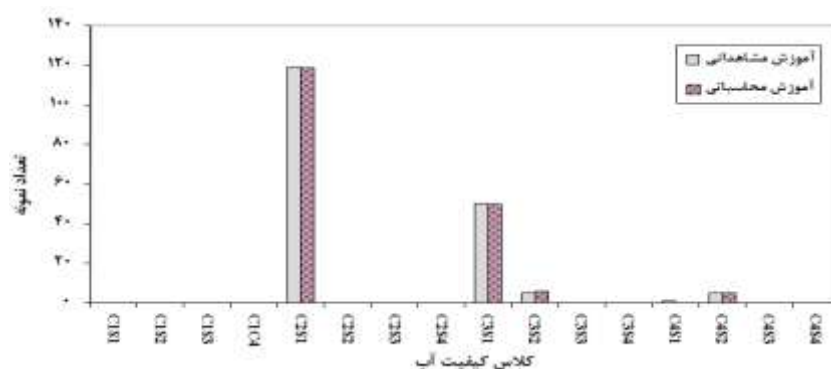
نتایج ارزیابی نمونه‌های آموزشی در ایستگاه بستان‌آباد نشان می‌دهد در این مدل تصمیم درختی چهار قانون برای طبقه‌بندی استفاده شده است. قانون ۱ بیان می‌کند که «اگر  $EC \leq 750$  باشد، آن‌گاه کلاس کیفیت آب C2-S1 خواهد بود». از کل نمونه‌های آموزشی ۱۱۹ نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۱ برابر

همان گونه که قبلاً اشاره شد، در این مطالعه در هر ایستگاه ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد برای تست در نظر گرفته شد. بعد از اینکه قوانین استخراج شدند، هم نمونه‌های قسمت آموزش و هم نمونه‌های قسمت تست با قوانین تطبیق داده و موارد درست و نادرست مشخص شد. شایان یادآوری است که در مدل‌سازی با روش‌های هوش مصنوعی معمولاً قسمت آموزش را ارزیابی نمی‌کنند، ولی در این روش برای ارزیابی عملکرد مدل و قوانین به‌دست‌آمده از درخت تصمیم تطبیق هر دو بخش داده‌های آموزش و تست با قوانین ارزیابی می‌شود که این کار توسط نرم‌افزار See5 انجام می‌شود [۱۵].

همچنین ماتریس اغتشاش که بیان‌کننده چگونگی پراکنش نمونه‌ها بین کلاس‌های مختلف است، برای ایستگاه بستان‌آباد در جدول ۳ ارائه شده است. در ماتریس اغتشاش نمونه‌هایی که روی قطر اصلی ماتریس واقع شده‌اند، نمونه‌هایی هستند که درست و صحیح پیش‌بینی شده‌اند. درباره ایستگاه بستان‌آباد فقط یک نمونه آموزشی در خارج از قطر اصلی واقع شده است و به‌جای اینکه نمونه در کلاس C3-S2 قرار بگیرد به اشتباه در کلاس C4-S1 قرار گرفته و سبب شده خطای این بخش ۰/۶ درصد باشد. این امر بیان‌کننده دقت بسیار زیاد مدل در بخش آموزش است.

دارد. دقت آماره لاپلاس برای قانون ۲ برابر با ۹۸/۱ درصد و مقدار عددی آماره لیفت برابر با ۳/۵ است. کم‌بودن نسبی مقدار آماره لیفت نشان‌دهنده دقت زیاد قانون ۲ است. قانون ۳ بیان می‌کند که «اگر  $Na^+ > 8/7$  و  $Na^+ \leq 11$  باشد، آن‌گاه کلاس کیفیت آب C3-S2 خواهد بود». از کل نمونه‌های آموزشی شش نمونه با شرایط این قانون همخوانی دارد. یکی از نمونه‌ها در این قانون صادق نیست و بنابراین مدل در تخمین کلاس کیفیت آب دچار خطا شده است. دقت آماره لاپلاس برای قانون ۳ برابر با ۷۵ درصد و مقدار عددی آماره لیفت برابر با ۲۷ است. زیادبودن نسبی مقدار آماره لیفت نشان‌دهنده دقت نسبتاً کمتر قانون ۳ است. قانون ۴ بیان می‌کند که «اگر  $Na^+ > 11$  باشد، آن‌گاه کلاس کیفیت آب C4-S2 خواهد بود». از کل نمونه‌های آموزشی پنج نمونه با این قانون همخوانی دارد. دقت آماره لاپلاس برای قانون ۴ برابر با ۸۵/۷ درصد و مقدار عددی آماره لیفت برابر با ۳۰ است.

همان گونه که از جدول ۲ استنباط می‌شود، در مجموعه این قوانین مستخرج از نمونه‌های آموزشی فقط یک نمونه از قانون ۳ تبعیت نکرده، این در حالی است که ۱۷۹ مورد درست طبقه‌بندی شده است. نمودار پراکنش نمونه‌های آموزشی در ایستگاه بستان‌آباد در شکل ۲ ارائه شده است.



شکل ۲. نمودار پراکنش نمونه‌های آموزشی در ایستگاه بستان‌آباد

جدول ۳. ماتریس اغتشاش برای نمونه‌های آموزشی ایستگاه هیدرومتری بستان‌آباد

کلاس	C4S2	C3S2	C3S1	C2S1
C2S1	0	0	0	۱۱۹
C3S1	0	0	۵۰	0
C3S2	0	۵	0	0
C4S1	0	۱	0	0
C4S2	۵	0	0	0

۳/۴ درصد خواهد بود. نمودار پراکنش نمونه‌های آزمون در ایستگاه بستان‌آباد در شکل ۳ ارائه شده است.

با توجه به اینکه در این رودخانه بیشترین پراکنش کلاس کیفیت آب مربوط به کلاس C2-S1 است، بنابراین این کلاس به‌عنوان پیش‌فرض در نظر گرفته شده است. همچنین نتایج به‌دست‌آمده نشان می‌دهد مدل تصمیم درختی C5.0 برای پیش‌بینی کیفیت آب فقط از پارامترهای EC و  $Na^+$  استفاده کرده است. به‌بیان دیگر، از بقیه پارامترهای هیدروشیمیایی برای توسعه مدل بهره‌ن گرفته است. بنابراین، برای تخمین کلاس کیفیت آب در ایستگاه بستان‌آباد اندازه‌گیری دو پارامتر EC و  $Na^+$  کافی است.

#### رودخانه آجی‌چای

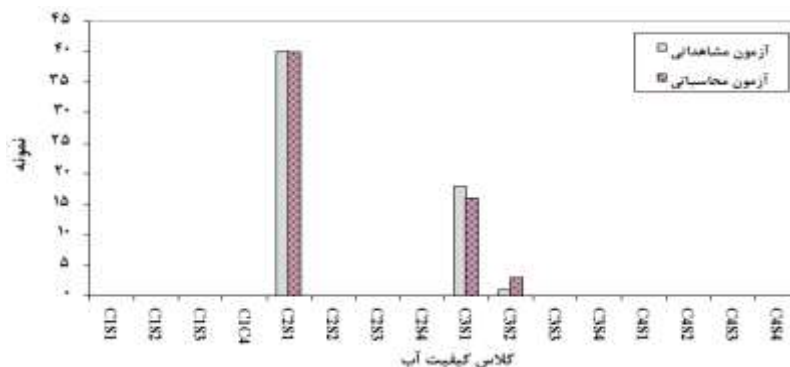
در این رودخانه در مجموع به تعداد ۲۴۵ مورد نمونه‌برداری از آب در فواصل زمانی مختلف در محل ایستگاه و نیار انجام گرفته است. خلاصه نتایج مدل تصمیم درختی C5.0 برای ایستگاه و نیار در جدول ۵ ارائه شده است.

پس از آموزش مدل و استخراج قوانین مربوطه، برای بررسی صحت نتایج، مدل با ۵۹ مورد نمونه آزمون شد. ماتریس اغتشاش مربوط به نمونه‌های آزمون در جدول ۴ ارائه شده است.

جدول ۴. ماتریس اغتشاش برای نمونه‌های آزمون ایستگاه هیدرومتری بستان‌آباد

کلاس	C3S2	C3S1	C2S1
C2S1			۴۰
C3S1	۲	۱۶	
C3S2	۱		

در جدول ۴ ملاحظه می‌شود که از کل مجموع ۵۹ مورد نمونه بخش آزمون فقط دو مورد نادرست طبقه‌بندی شده و خارج از قطر اصلی ماتریس اغتشاش است. این دو مورد به جای اینکه در کلاس C3-S2 پیش‌بینی شوند به اشتباه در کلاس C3-S1 واقع شده‌اند. بنابراین، خطای ناشی از داده‌های بخش آزمون مدل برابر با



شکل ۳. نمودار پراکنش نمونه‌های آزمون در ایستگاه هیدرومتری بستان‌آباد

جدول ۵. خلاصه نتایج مدل تصمیم درختی C5.0 برای ایستگاه هیدرومتری و نیار

شماره قانون	اگر			آن‌گاه (طبقه کیفیت آب)	تعداد نمونه صادق / غیرصادق	دقت لاپلاس (درصد) /LIFT
	۱	۲	۳			
۱	$EC \leq 2170$	$SAR \leq 4/696$	-	C3-S2	۶/۰	۸۷/۵ - ۲۶/۷
۲	$EC \leq 2170$	$SAR > 4/696$	-	C3-S2	۱۴/۱	۸۷/۵ - ۱۲/۳
۳	$HCO_3^- > 8$	-	-	C4-S1	۲/۱	۵۰/۰ - ۱۸/۳
۴	$pH > 7/99$	$\%Na > 81/63$	$Q > 1/35$	C4-S2	۳/۰	۸۰/۰ - ۱۶/۳
۵	$EC > 2170$	$Na^+ \leq 18$	-	C4-S2	۷/۱	۷۷/۸ - ۱۵/۸
۶	$Na^+ > 18$	$Na^+ \leq 28/3$	-	C4-S3	۱۸/۲	۸۵/۰ - ۹/۱
۷	$pH \leq 7/99$	$HCO_3^- \leq 8$	$Na^+ > 28/3$	C4-S4	۱۱۸/۰	۹۹/۲ - ۱/۴
۸	$HCO_3^- \leq 8$	$Na^+ > 28/3$	$Q \leq 1/35$	C4-S4	۷۶/۰	۹۸/۷ - ۱/۴
۹	$HCO_3^- \leq 8$	$Na^+ > 28/3$	$\%Na \leq 81/63$	C4-S4	۶۳/۱	۶۹/۹ - ۱/۳



بخش آزمون مدل برابر با ۸/۱ درصد خواهد بود. با توجه به اینکه در این رودخانه بیشترین پراکنش کلاس کیفیت مربوط به کلاس C4-S4 است، بنابراین این کلاس به‌عنوان کلاس پیش‌فرض در رودخانه و نیار در نظر گرفته شده است. همچنین نتایج به‌دست‌آمده نشان می‌دهد، مدل تصمیم درختی C5.0 برای پیش‌بینی کیفیت آب از پارامترهای SAR، EC،  $\text{Na}^+$ ، pH،  $\text{Na}\% \text{HCO}_3^-$  و دبی (Q) استفاده کرده است. به بیان دیگر، از بقیه پارامترهای هیدروشمیایی برای توسعه مدل بهره نگرفته است. بنابراین، برای تخمین کلاس کیفیت آب در ایستگاه و نیار اندازه‌گیری پارامترهای یادشده کافی است.

#### رودخانه سنخ‌چای

در این رودخانه تعداد ۱۹۵ مورد نمونه‌برداری از آب در فواصل زمانی مختلف در محل ایستگاه هیدرومتری پل سنخ انجام گرفته است. خلاصه نتایج مدل تصمیم درختی C5.0 برای رودخانه پل سنخ در جدول ۸ ارائه شده است.

نتایج ارزیابی نمونه‌های آموزشی در ایستگاه پل سنخ نشان می‌دهد در این مدل تصمیم درختی ۸ قانون برای طبقه‌بندی استفاده شده است. همچنان که از جدول ۹ ملاحظه می‌شود از مجموع ۱۴۷ مورد، ۱۳۶ مورد درست طبقه‌بندی شده است. از این‌رو، در مجموع خطا برابر با ۷/۵ درصد خواهد بود. این امر بیانگر دقت بسیار زیاد مدل در بخش آموزش است. در این بخش از بین ۱۴۷ نمونه آموزشی شش نمونه در کلاس C3-S2، نه نمونه در کلاس C4-S1، ۱۵ نمونه در کلاس C4-S2، ۴۶ نمونه در کلاس C4-S3، ۷۱ نمونه در کلاس C4-S4 توزیع شده است.

نتایج ارزیابی نمونه‌های آموزشی در ایستگاه و نیار نشان می‌دهد در این مدل تصمیم درختی نه قانون برای طبقه‌بندی استفاده شده است.

همچنان که در جدول ۵ ملاحظه می‌شود از بین ۱۸۳ نمونه آموزشی، ۱۷۷ مورد درست طبقه‌بندی شده است. از کل نمونه‌ها، ۶ نمونه در کلاس C3-S1، ۱۴ نمونه در کلاس C3-S2، ۲ نمونه در کلاس C4-S1، ۱۰ نمونه در کلاس C4-S2، ۱۸ نمونه در کلاس C4-S3، ۱۳۳ نمونه در کلاس C4-S4 توزیع شده است. در جدول ۶ ماتریس اغتشاش ایستگاه و نیار ارائه شده است. درباره ایستگاه و نیار شش نمونه آموزشی در خارج از قطر اصلی واقع شده است و از این شش مورد، یک نمونه به‌جای اینکه در کلاس C3-S3 قرار بگیرد، در کلاس C3-S2 و ۱ نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرد، در کلاس C4-S4 و ۲ نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرد، در کلاس C4-S3 و یک نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرد، در کلاس C4-S2 و یک نمونه به‌جای اینکه در کلاس C4-S3 قرار بگیرد، به اشتباه در کلاس C4-S1 قرار گرفته‌اند و سبب شده که خطای این بخش ۳/۳ درصد باشد. این امر بیان‌کننده دقت بسیار زیاد مدل در بخش آموزش است.

پس از آموزش مدل و استخراج قوانین مربوطه برای بررسی صحت نتایج مدل با ۶۲ مورد نمونه آزموده شد. نتایج پراکنش نمونه‌های آزمون در بین کلاس‌های مختلف و ماتریس اغتشاش مربوط به نمونه‌های آزمون در جدول ۷ ارائه شده است.

در جدول ۷ ملاحظه می‌شود که از کل مجموع ۶۲ مورد نمونه آزمون، پنج مورد از نمونه‌ها خارج از قطر اصلی ماتریس اغتشاش است. بنابراین، خطای ناشی از داده‌های

جدول ۶. ماتریس اغتشاش برای نمونه‌های آموزش ایستگاه هیدرومتری و نیار

کلاس	C3S1	C3S2	C4S1	C4S2	C4S3	C4S4
C3S1	۶					
C3S2		۱۳				
C3S3			۱			
C4S1			۱	۱	۲	۱
C4S2				۹		
C4S3					۱۶	
C4S4						۱۳۲

جدول ۷. ماتریس اغتشاش برای نمونه‌های آزمون ایستگاه هیدرومتری و نیار

کلاس	C3S2	C4S1	C4S2	C4S3	C4S4
C3S1	۱				
C3S2	۱				
C4S1				۱	
C4S2			۲		
C4S3				۸	
C4S4					۴۶

اینکه در کلاس C3-S1 قرار بگیرد، در کلاس C3-S2 و دو نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرند، در کلاس C4-S3 و یک نمونه به‌جای اینکه در کلاس C4-S3 قرار بگیرد، در کلاس C4-S4 و پنج نمونه به‌جای اینکه در کلاس C4-S3 قرار بگیرد، در کلاس C4-S2 و یک نمونه به‌جای اینکه در کلاس C4-S4 قرار بگیرد در کلاس C4-S3 و یک نمونه به‌جای اینکه در کلاس C4-S4 قرار بگیرد، در کلاس C4-S1 واقع شده‌اند. بنابراین، خطای ناشی از داده‌های بخش آزمایشی مدل برابر با ۲۲/۹ درصد خواهد بود. با توجه به اینکه در این رودخانه بیشترین پراکنش کلاس کیفیت مربوط به کلاس C4-S4 است، بنابراین این کلاس به‌عنوان کلاس پیش‌فرض در نظر گرفته شده است. همچنین نتایج به‌دست‌آمده نشان می‌دهد مدل تصمیم درختی C5.0 برای پیش‌بینی کیفیت آب فقط از پارامترهای SAR، EC،  $Mg^{2+}$ ، pH،  $Na^+$ ،  $HCO_3^-$ ،  $SO_4^{2-}$  استفاده کرده است. به‌بیان دیگر، برای تخمین کلاس کیفیت آب در رودخانه سنخ‌چای اندازه‌گیری پارامترهای یادشده کافی است.

ماتریس اغتشاش نمونه‌های آموزشی ایستگاه پل سنخ در جدول ۹ ارائه شده است. درباره رودخانه پل سنخ ۱۱ نمونه آموزشی در خارج از قطر اصلی واقع شده است و از این ۱۱ مورد، یک نمونه به‌جای اینکه در کلاس C3-S1 قرار بگیرد، در کلاس C3-S2 و دو نمونه به‌جای اینکه در کلاس C3-S3 قرار بگیرد، در کلاس C3-S2 و دو نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرد، در کلاس C4-S1 و یک نمونه به‌جای اینکه در کلاس C4-S1 قرار بگیرد، در کلاس C4-S2 و چهار نمونه به‌جای اینکه در کلاس C4-S3 قرار بگیرد، در کلاس C4-S4 و یک نمونه به‌جای اینکه در کلاس C4-S4 قرار بگیرد، به اشتباه در کلاس C4-S1 قرار گرفته‌اند. پس از آموزش مدل و استخراج قوانین مربوطه برای بررسی صحت نتایج، مدل با ۴۸ نمونه آزموده شد. ماتریس اغتشاش مربوط به نمونه‌های آزمون در جدول ۱۰ ارائه شده است.

در جدول ۱۰ ملاحظه می‌شود که از کل مجموع ۴۸ نمونه آزمون، ۱۱ مورد از نمونه‌ها خارج از قطر اصلی ماتریس اغتشاش است. از این ۱۱ مورد، یک نمونه به‌جای

جدول ۸. خلاصه نتایج مدل تصمیم درختی برای ایستگاه هیدرومتری پل سنخ

شماره قانون	اگر					آن‌گاه (طبقه کیفیت آب)	تعداد نمونه صادق غیرصادق /	دقت لاپلاس /LIFT (درصد)
	۱	۲	۳	۴	۵			
۱	$EC \leq 2240$	-	-	-	-	C3-S2	۶/۱	۵۰-۲۴/۵
۲	$pH \geq 7.69$	$HCO_3^- \geq 3$	$Mg^{2+} \leq 4.62$	$Na^+ \geq 27.1$	$SAR \geq 12.758$	C4-S1	۹/۱	۸۱/۸-۱۰/۹
۳	$EC \geq 2240$	$SO_4^{2-} \geq 12.17$	$SAR \leq 9.234$	-	-	C4-S2	۱۵/۱	۸۸/۲-۹/۳
۴	$EC \geq 2240$	$SAR \geq 9.234$	$SAR \leq 12.758$	-	-	C4-S3	۴۲/۲	۹۳/۳-۲/۹
۵	$EC \geq 2240$	$SO_4^{2-} \geq 12.17$	$SAR \leq 12.758$	-	-	C4-S3	۵/۰	۸۵/۷-۲/۶
۶	$pH \leq 7.69$	$Na^+ \geq 27.1$	$SAR \geq 12.758$	-	-	C4-S4	۳۳/۰	۹۷/۱-۲/۱
۷	$Mg^{2+} \geq 4.62$	$SAR \geq 12.758$	-	-	-	C4-S4	۵۲/۱	۹۶/۴-۱/۲
۸	$HCO_3^- \leq 3$	$SAR \geq 12.758$	-	-	-	C4-S4	۱۲/۰	۹۳/۳-۲/۰

جدول ۹. ماتریس اغتشاش برای نمونه‌های آموزش ایستگاه هیدرومتری پل سنیخ

کلاس	C3S2	C4S1	C4S2	C4S3	C4S4
C3S1	۱				
C3S2	۳				
C3S3	۲				
C4S1		۸	۱	۲	
C4S2			۱۴		
C4S3				۴۴	۴
C4S4					۶۷

جدول ۱۰. ماتریس اغتشاش برای نمونه‌های آزمون ایستگاه هیدرومتری پل سنیخ

کلاس	C3S2	C4S1	C4S2	C4S3	C4S4
C3S1	۱				
C3S2	۵				
C4S1		۰		۲	
C4S2			۳		
C4S3				۷	۱
C4S4					۲۲

### رودخانه ليقوان چای

در این رودخانه در مجموع به تعداد ۲۴۵ مورد نمونه‌برداری از آب در فواصل زمانی مختلف در محل ایستگاه ليقوان انجام گرفته است. خلاصه نتایج مدل تصمیم درختی C5.0 برای رودخانه ليقوان چای در جدول ۱۱ ارائه شده است.

نتایج ارزیابی نمونه‌های آموزشی در رودخانه ليقوان چای نشان می‌دهد در این مدل تصمیم درختی ۲ قانون برای طبقه‌بندی استفاده شده است. طبق جدول ۱۱ ملاحظه می‌شود در مجموعه این قوانین مستخرج از نمونه‌های آموزشی یک نمونه از قانون ۲ تبعیت نکرده، این در حالی است که ۱۸۲ مورد درست طبقه‌بندی شده است. در این بخش از بین ۱۸۳ نمونه آموزشی ۵۴ نمونه در کلاس C1-S1، ۱۲۹ نمونه در کلاس C2-S1، توزیع شده است. ماتریس اغتشاش نمونه‌های آموزشی رودخانه ليقوان در جدول ۱۲ ارائه شده است. درباره ایستگاه ليقوان فقط یک نمونه آموزشی در خارج از قطر اصلی واقع شده است و به‌جای اینکه نمونه در کلاس C2-S1 قرار بگیرد، به‌اشتباه در کلاس C3-S1 قرار گرفته و سبب شده که خطای این بخش ۰/۵ درصد باشد.

پس از آموزش مدل و استخراج قوانین مربوطه برای بررسی صحت نتایج مدل با ۵۹ نمونه آزموده شد. ماتریس اغتشاش مربوط به نمونه‌های آزمایشی در جدول ۱۳ ارائه شده است. در جدول ۱۳ ملاحظه می‌شود که از کل ۶۲ نمونه آزمایشی فقط یک مورد نادرست طبقه‌بندی شده و خارج از قطر اصلی ماتریس اغتشاش است. این یک مورد به‌جای اینکه در کلاس C1-S1 پیش‌بینی شود، به‌اشتباه در کلاس C2-S1 واقع شده است. بنابراین، خطای ناشی از داده‌های بخش آموزش مدل برابر با ۱/۶ درصد خواهد بود. با توجه به اینکه در این رودخانه بیشترین پراکنش کلاس کیفیت مربوط به کلاس C2-S1 است، بنابراین این کلاس به‌عنوان کلاس پیش‌فرض در نظر گرفته شده است. همچنین نتایج به‌دست‌آمده نشان می‌دهد مدل تصمیم درختی C5.0 برای پیش‌بینی کیفیت آب فقط از پارامتر EC استفاده کرده است. به بیان دیگر، برای تخمین کلاس کیفیت آب در رودخانه ليقوان فقط اندازه‌گیری پارامتر EC کافی است.

جدول ۱۱. خلاصه نتایج مدل تصمیم درختی C5.0 برای ایستگاه هیدرومتری ليقوان

شماره قانون	اگر		آن گاه (طبقه کیفیت آب)	تعداد نمونه صادق غیرصادق /	دقت لایپلاس (درصد) /LIFT
	۱	۲			
۱	EC ≤ ۲۴۵	-	C1-S1	۵۴/۰	۹۸/۲ - ۳/۳
۲	EC > ۲۴۵	-	C2-S1	۱۲۹/۱	۹۸/۵ - ۱/۴

جدول ۱۲. ماتریس اغتشاش برای نمونه‌های آموزش ایستگاه هیدرومتری ليقوان

کلاس	C3S1	C2S1	C1S1
C1S1			۵۴
C2S1		۱۲۸	
C3S1		۱	

جدول ۱۳. ماتریس اغتشاش برای نمونه‌های تست ایستگاه هیدرومتری ليقوان

کلاس	C2S1	C1S1
C1S1	۱	۱۰
C2S1	۵۱	

جدول ۱۴. خلاصه نتایج مدل تصمیم درختی C5.0 برای رودخانه‌های دامنه شمالی سهند

ایستگاه	تعداد قانون	آموزش		آزمون	
		خطا (درصد)	دقت مدل (درصد)	خطا (درصد)	دقت مدل (درصد)
بستان‌آباد	۴	۰/۶	۹۹/۴	۳/۴	۹۶/۶
لیقوان	۲	۰/۵	۹۹/۵	۱/۶	۹۸/۴
پل سنیخ	۸	۷/۵	۹۲/۵	۲۲/۹	۷۷/۱
ونیار	۹	۳/۳	۹۶/۷	۸/۱	۹۱/۹

استفاده از چهار پارامتر EC، pH، SAR و  $Na^+$  قادر به طبقه‌بندی کلاس کیفیت آب سطحی رودخانه‌های منتخب دامنه شمالی سهند است.

### نتیجه‌گیری

در این پژوهش، از روش تصمیم‌گیری درختی برای پیش‌بینی کیفیت آب‌های سطحی در ایستگاه‌های هیدرومتری رودخانه‌های منتخب دامنه شمالی سهند (بستان‌آباد، پل سنیخ، ونیار و ليقوان) استفاده شد. در این رودخانه‌ها ابتدا با استفاده از دیاگرام USSL و براساس پارامترهای هیدروشیمیایی مؤثر در ارزیابی کیفیت آب، طبقه و کلاس کیفیت آب مشخص شد. براساس نمونه‌برداری‌های انجام‌یافته در این چهار رودخانه در مجموع برای هر یک از رودخانه‌ها دبی و ۱۲ پارامتر هیدروشیمیایی

خلاصه نتایج به‌دست‌آمده از روش تصمیم درختی برای رودخانه‌های منتخب دامنه شمالی سهند در جدول ۱۴ آورده شده است. نتایج به‌دست‌آمده در مطالعه حاضر با نتایج ثاقب‌یان و همکارانش [۱۰] قابل مقایسه است. مدل درختی که آنها ارائه کردند، فقط با استفاده از دو پارامتر هدایت الکتریکی (EC) و بارش تجمعی ۱۱ ماه قبل (CP11) قادر به پیش‌بینی طبقه کیفیت آب زیرزمینی دشت اردبیل بر مبنای دیاگرام USSL بود. همچنین در مطالعه حسنی و همکارانش [۱۱] مدل درختی توسعه داده شده قادر بود با استفاده از چهار پارامتر مجموع کاتیون‌ها، SAR،  $Na^+$  و EC طبقه کیفیت آب زیرزمینی دشت خان‌میرزا را بر مبنای دیاگرام USSL پیش‌بینی کند. اما در مطالعه حاضر مدل درختی توسعه داده شده بیشتر با

- Journal of Agricultural Research, 2007; 8(1b): 27-35. [Persian]
- [6]. Goljan F, Karbasi AR, Hajizadeh Zaler N, Nabi Bidhendi GR. Water quality of Nour City rivers, Journal of Water Sciences Research, 2009; 1(1): 35-48. [Persian]
- [7]. Olyaie E, Banejad H, Samadi MT, Rahmani AR, Saghi MH, Performance Evaluation of Artificial Neural Networks for Predicting Rivers Water Quality Indices (BOD and DO) in Hamadan Morad Beik River, Water and Soil Science, 2010; 20(3): 199-210. [Persian]
- [8]. Hajian Nejad M, Rahsepar AR, Measurement and Simulation of Dissolved Oxygen in Zayande Rood River, Journal of Health System Research, 2010; 6(2): 821-828. [Persian]
- [9]. Salajegheh A, Razavizadeh S, Khorasani N, Hamidifar M, Salajegheh S, Land use Changes and its Effects on Water Quality (Case study: Karkheh Watershed), 2011; 58:81-86. [Persian]
- [10]. Saghebiam SM, Sattari MT, Mirabbasi R, Pal M. Ground water quality classification by decision tree method in Ardebil region, Iran. Arabian Journal of Geosciences. 2013; 7(11): 4767-4777.
- [11]. Hasani Z, Mirabbasi Najafabadi R, Ghasemi AR. Prediction of groundwater quality in Khanmirza plain using decision tree method, Hydrogeology. 2016; 1(3): 15-30. [Persian]
- [12]. Norouzi H, Nadiri A, Asghari Moghaddam A. Investigation of Malikan Plain Groundwater's Pollution to Arsenic, Ecohydrology. 2016; 3(2): 151-166. [Persian]
- [13]. Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques, second edition, Elsevier: San Francisco. ISBN 0-12-088407-0; 2005.
- [14]. Quinlan JR. C4.5 Programs for machine learning, Morgan, Kaufmann, 1993; San Mateo, California
- [15]. Quinlan JR. Data mining tools See5 and C5.0 [cited Feb 2012]. Available from <http://www.rulequest.com/see5-info.html>. 2000.
- مورد توجه قرار گرفت. ۱۲ صفت و ویژگی هیدروشیمیایی یادشده به‌عنوان ورودی مدل درخت تصمیم و کلاس کیفیت آب به‌منزله خروجی و ویژگی هدف در نظر گرفته شد. نتایج به‌دست‌آمده نشان داد تصمیم‌گیری درختی بیشتر با استفاده از چهار پارامتر EC, pH, SAR و  $\text{Na}^+$  قادر به طبقه‌بندی کلاس کیفیت آب با دقت بسیار زیادی است. به‌تبع تعیین پارامترهای مؤثر در تعیین کلاس کیفیت آب‌های سطحی با استفاده از روش تصمیم‌گیری درختی، سبب خواهد شد که علاوه بر کاهش چشمگیر هزینه‌های نمونه‌برداری و آنالیز شیمیایی نمونه‌ها، زمان اختصاص‌یافته برای تعیین کلاس کیفیت آب نیز کاهش یابد و در عین حال نتایج دقت قابل قبولی داشته باشند.

#### منابع

- [1]. U.S. Salinity Laboratory Staff, Diagnosis and improvement of saline and alkali soils: U.S. Dept. Agric. Handbook; 1954. No.60, 160 p.
- [2]. Mirabbasi R, Mazlounzadeh SM, Rahnama MB. Evaluation of irrigation water quality using fuzzy logic, Research Journal of Environmental Sciences, 2008; 2(5): 340-352.
- [3]. Santos MF, Cortez P, Quintela H, Neves J, Vicente H & Arteiro J. Ecological Mining - A Case Study on Dam Water Quality. In A. Zanasi, C. Brebbia and N. Ebecken (Eds.), Data Mining VI - Data Mining, Text Mining and their Business Applications, WIT Transactions of Information and Communication Technologies 35, 523-531, WIT Press, ISBN:1-84564-017-9, ISSN:1746-4463; 2005.
- [4]. Yahya SM, Rahman AU, Abbasi HN. Assessment of seasonal and polluting effects on the quality of river water by using regression analysis: A case study of River Indus in province of Sindh, Pakistan. International Journal of Environmental Protection. 2012; 2(2): 10-16.
- [5]. Rahmani AR, Samadi MT, Heydari M. Water quality assessment of Hamadan-Bahar Plain rivers using Wilcox diagram for irrigation,