

## امکان سنجی کاربرد روش‌های داده‌کاوی در تخمین طبقه کیفی آب رودخانه آجی‌چای

محمدتقی ستاری\*<sup>۱</sup>، علی رضازاده جودی<sup>۲</sup>

تاریخ دریافت: ۹۴/۰۹/۰۴ تاریخ پذیرش: ۹۵/۰۴/۲۲

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

۲- دانش‌آموخته کارشناسی ارشد عمران-آب، باشگاه پژوهشگران جوان و نخبگان، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران

\* مسئول مکاتبات، پست الکترونیکی: mtsattar@gmail.com

### چکیده

دسترسی به منابع آب پاک و با کیفیت یکی از دغدغه‌های اساسی انسان از دیرباز بوده است. از این رو تعیین کیفیت آب برای مصارف مختلف از جمله آبیاری در مناطق مختلف بسیار ضروری می‌باشد. در این تحقیق، ابتدا کیفیت آب آبیاری در رودخانه آجی‌چای در ۴ ایستگاه آخولا، ارزنق، مرکید و ونیاز توسط دیاگرام USSL طبقه‌بندی شد. سپس امکان استفاده از روش‌های طبقه‌بندی‌کننده بردار پشتیبان، K-نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی در تخمین طبقه کیفی آب با استفاده از پارامترهای هیدروشیمیایی مختلف مورد سنجش قرار گرفت. ارزیابی عملکرد روش‌های داده‌کاوی نشان‌دهنده دقت بالا و عملکرد بسیار مناسب این روش‌ها در تعیین طبقه کیفی آب می‌باشد. در این تحقیق بر اساس آماره‌های کاپا و نرخ خطا، روش‌های مورد استفاده از نظر دقت عملکرد رتبه‌بندی گردید. با بررسی دقیق نتایج مشاهده گردید که روش طبقه‌بندی‌کننده بردار پشتیبان که با بهره‌گیری از توابع کرنل توانایی بالایی در حل مسائل مختلف دارد، با میانگین رتبه ۱/۲۵ به‌عنوان کارآمدترین روش داده‌کاوی و پس از آن روش K-نزدیک‌ترین همسایگی با میانگین رتبه ۱/۷۵ و شبکه عصبی مصنوعی با میانگین رتبه ۲ به‌عنوان روش‌هایی مناسب جهت تعیین طبقه کیفی آب می‌باشند.

واژه‌های کلیدی: K-نزدیک‌ترین همسایگی، دیاگرام USSL، شبکه عصبی مصنوعی، طبقه‌بندی کیفیت آب، طبقه‌بندی‌کننده

بردار پشتیبان

## Feasibility Study of Data Mining Methods Application to Estimate Aji Chai River's Water Quality Classification

Mohammad Taghi Sattari<sup>1\*</sup>, Ali Rezazadeh Joudi<sup>2</sup>

Received: 25 November 2015 Accepted: 12 July 2016

1\*- Assist. Prof., Depart. of Water Engineering, Faculty of Agric., Univ. of Tabriz, Iran.

2- M.Sc graduated of Water-Civil Engineering, Young Researchers and Elite Club, Maragheh Branch, Islamic Azad University, Maragheh, Iran

\* Corresponding Author, Email: mtsattar@gmail.com

### Abstract

Accessing to clean and high quality water sources has always been one of the main concerns of humanity. Therefore, determination of water quality is essential for various applications e.g. irrigation. In this study, irrigation water of Aji Chai River in stations (Akhula, Arzanagh, Markid and Vaniar) is initially classified using USSSL diagram. After that, feasibility of support vector classifier, K-nearest neighborhood and artificial neural network classification methods is assessed. Evaluation of data mining methods presents high accuracy and performance of these methods in assessment of water quality levels. In this study, the aforementioned methods are ranked by accuracy using kappa and error rate statistics. Careful examination of the results demonstrates that the support vector classifier which uses kernel functions is highly capable of solving various problems, and with an average ranking of 1.25 is the most efficient mining method followed by K- nearest neighborhood method with an average ranking of 1.75 and artificial neural network with an average rank of 2. These are also suitable methods for determining water quality classification.

**Keywords:** Artificial neural network, K-nearest neighborhood, Support vector classifier, USSSL diagram, Water quality classification

### مقدمه

لزوم صرفه‌جویی در زمان و هزینه‌ها، استفاده از روش‌های جایگزین مانند روش‌های نوین داده‌کاوی می‌تواند راهکار مناسبی جهت پیش‌بینی و طبقه‌بندی کیفیت آب در همه نقاط و در کم‌ترین زمان ممکن باشد. ساهو و همکاران (۲۰۱۱) سیستم استنتاج فازی-عصبی را برای پیش‌بینی شاخص کیفیت آب به کار بردند و توانایی بالای این روش را گزارش کردند. چپود و همکاران (۲۰۱۲) با استفاده از روش‌های سنجش از دور و شبکه عصبی مصنوعی به بررسی کیفیت آب پرداخته

ارزیابی کیفیت آب‌های سطحی از اهمیت بالایی در مدیریت منابع آب برخوردار می‌باشد (دوگان و همکاران ۲۰۰۹). یکی از دیاگرام‌هایی که برای طبقه‌بندی آب آبیاری ارائه شده دیاگرام آزمایشگاه شوری ایالات متحده<sup>۱</sup> با نام اختصاری دیاگرام USSSL می‌باشد (بی‌نام ۱۹۵۴). عدم وجود امکانات لازم در کلیه ایستگاه‌ها و

<sup>1</sup> - United States Salinity Laboratory's Diagram

آب‌های سطحی پرداختند. نتایج نشان داد که روش درخت تصمیم قادر است با استفاده از کمترین تعداد پارامتر ورودی طبقه کیفی آب را پیش‌بینی نماید. میرزاوند و همکاران (۱۳۹۴) مقایسه‌ای بین عملکرد مدل‌های شبکه عصبی مصنوعی و رگرسیون چندمتغیره در تخمین تغییرات کیفی آب‌های زیرزمینی انجام دادند و دقت بالاتر شبکه عصبی مصنوعی پرسپترون چندلایه را در این زمینه نسبت به مدل‌های پایه شعاعی و رگرسیون چند متغیره گزارش کردند. همچنان‌که مشاهده می‌گردد طبق بررسی منابع صورت گرفته در زمینه طبقه‌بندی کیفی آب، بیشتر تحقیقات با استفاده از روش شبکه عصبی مصنوعی و درخت تصمیم صورت پذیرفته و تحقیقات چندان در زمینه کاربرد روش طبقه‌بندی کننده بردار پشتیبان و K- نزدیک‌ترین همسایگی در زمینه طبقه‌بندی کیفی آب آبیاری صورت نپذیرفته است. هدف از این تحقیق ارزیابی عملکرد روش‌های طبقه‌بندی کننده بردار پشتیبان، K- نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی در انجام طبقه‌بندی کیفیت آب رودخانه آجی‌چای واقع در استان آذربایجان شرقی در چهار ایستگاه هیدرومتری آخولا، ارزنق، مرکید و ونیار بر اساس دیاگرام آزمایشگاه شوری ایالات متحده آمریکا می‌باشد.

### مواد و روش‌ها

#### منطقه مورد مطالعه و داده‌های مورد استفاده

رودخانه آجی‌چای در دامنه‌های شمالی کوه سه‌سند واقع است. در این مطالعه، به‌منظور ارزیابی و طبقه‌بندی کیفیت آب آبیاری آب رودخانه آجی‌چای، از داده‌های هیدروشیمیائی ایستگاه هیدرومتری چهار ایستگاه آخولا، ارزنق، مرکید و ونیار استفاده گردید. شکل ۱ نشان‌دهنده موقعیت جغرافیایی منطقه مورد مطالعه به‌همراه رودخانه‌ها و محل‌های ایستگاه‌های هیدرومتری مورد مطالعه می‌باشد. در این مطالعه ۱۰

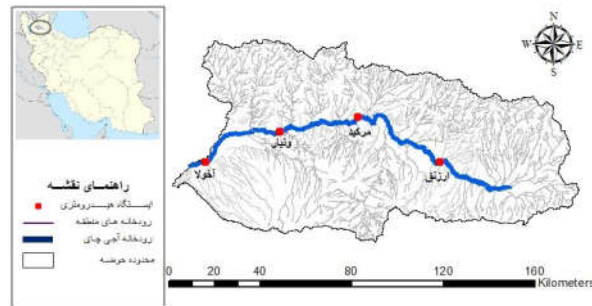
و کارآیی بالای این روش‌ها را در پیش‌بینی دقیق پارامترهای کیفی آب در اکوسیستم‌های پیچیده گزارش کردند. قزاز و همکاران (۲۰۱۲) با استفاده از شبکه عصبی مصنوعی به پیش‌بینی شاخص کیفیت آب در رودخانه‌ای در مالزی پرداختند و اعلام کردند که این روش توانایی بالایی در پیش‌بینی شاخص کیفیت آب دارا می‌باشد. ثاقب‌یان و همکاران (۲۰۱۴) به طبقه‌بندی کیفیت آب‌های زیرزمینی با استفاده از درخت تصمیم بر اساس دیاگرام USSSL پرداختند و اعلام نمودند کاربرد این روش باعث کاهش در هزینه‌های آزمایشگاهی و زمان صرف شده برای طبقه‌بندی کیفیت آب می‌گردد. مدرسی و عراقی‌نژاد (۲۰۱۴) در تحقیقی به ارزیابی روش‌های k- نزدیک‌ترین همسایگی، ماشین بردار پشتیبان و شبکه عصبی احتمالاتی در طبقه‌بندی کیفیت آب بر اساس شاخص کیفی CCME<sup>۲</sup> پرداختند. نتایج نشان داد از بین روش‌های استفاده شده الگوریتم ماشین بردار پشتیبان دارای بهترین کارآیی و روش K- نزدیک‌ترین همسایگی دارای ضعیف‌ترین عملکرد می‌باشد. ستاری و همکاران (۲۰۱۵) با استفاده از روش رگرسیون بردار پشتیبان و K- نزدیک‌ترین همسایگی به تخمین پارامترهای کیفی آب شامل میزان هدایت الکتریکی و کل جامدات محلول پرداخته و کارآیی بالای این روش‌ها را اعلام کردند. سنگرور و همکاران (۲۰۱۵) با استفاده از روش‌های هوشمند نقشه خود سازمان‌ده<sup>۳</sup> و شبکه عصبی مصنوعی به ارزیابی کیفیت آب در رودخانه پرداختند و دقت بالای این روش‌ها را گزارش نمودند. بانژاد و همکاران (۱۳۹۲) با استفاده از مدل ترکیبی شبکه‌های عصبی-موجکی به تخمین پارامترهای کیفی آب پرداختند. نتایج نشان داد می‌توان از این مدل برای حصول اطمینان از نتایج پایش و کاهش هزینه پایش استفاده نمود. همچنین ستاری و همکاران (۱۳۹۳) با استفاده از روش درخت تصمیم به پیش‌بینی کیفیت

<sup>۲</sup> - Canadian Council of Ministers of the Environment

<sup>۳</sup> - Self Organized Map

آجی‌چای در ایستگاه‌های مذکور مورد بررسی قرار گرفتند. مشخصات جغرافیایی ایستگاه‌های مورد مطالعه و طول دوره آماری داده‌های استفاده شده در جدول ۱ ارائه گردیده است. این داده‌ها از شرکت آب منطقه‌ای استان آذربایجان شرقی اخذ شدند.

پارامتر هیدروشیمیایی شامل یون سدیم ( $\text{Na}^+$ )، یون کلسیم ( $\text{Ca}^{2+}$ )، یون منیزیم ( $\text{Mg}^{2+}$ )، یون کلر ( $\text{Cl}^-$ )، بی‌کربنات ( $\text{HCO}_3^-$ )، اسیدیته (pH)، سولفات ( $\text{SO}_4^{2-}$ )، کل نمک محلول (TDS)، نسبت سدیم جذبی (SAR) و هدایت الکتریکی (EC) و پارامتر دبی ماهانه رودخانه (Q)



شکل ۱- موقعیت جغرافیایی محدوده مطالعاتی و محل قرارگیری ایستگاه‌های هیدرومتری.

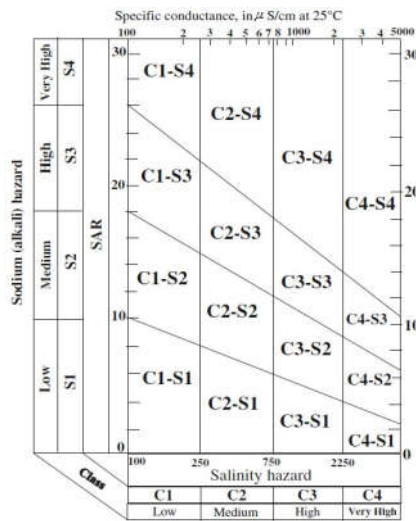
جدول ۱- مشخصات جغرافیایی و طول دوره آماری داده‌های استفاده شده.

ایستگاه	طول جغرافیایی (°E)	عرض جغرافیایی (°N)	ارتفاع (m)	دوره آماری
آخولا	۴۶° ۰۲' ۵۳"	۳۸° ۰۰' ۴۵"	۱۳۲۶	۱۳۸۰-۱۳۹۰
ارزنق	۴۷° ۱۲' ۳۹"	۳۷° ۵۹' ۰۷"	۱۶۱۰	۱۳۸۲-۱۳۹۰
مرکید	۴۶° ۴۸' ۲۶"	۳۸° ۰۹' ۵۴"	۱۵۰۰	۱۳۶۹-۱۳۹۰
وینار	۴۶° ۲۴' ۱۲"	۳۸° ۰۶' ۵۶"	۱۴۷۰	۱۳۴۵-۱۳۹۰

غلظت منیزیم و کلسیم اشاره کرد (بی‌نام ۱۹۵۴). نمایی از این دیاگرام در شکل ۲ نشان داده شده است. در این نمودار محور افقی نشان‌دهنده میزان شوری آب و محور عمودی نشان‌دهنده میزان نسبت جذبی سدیم می‌باشد. مشخصات مربوط به کیفیت آب در منطقه‌ای از این نمودار با حروف C از نظر شوری و S از نظر سدیم مشخص می‌گردد. مقادیر ۱، ۲، ۳ و ۴ به ترتیب نشان‌دهنده کم، متوسط، زیاد و خیلی زیاد می‌باشد. به‌عنوان مثال طبقه C2S4 بیانگر شوری متوسط و سدیم خیلی زیاد آب می‌باشد.

#### دیاگرام شوری ایالات متحده آمریکا (USSL)

استفاده از دیاگرام شوری ایالات متحده آمریکا (USSL) ۱۹۵۴ جهت ارزیابی کیفیت آب به‌منظور اطمینان از عدم وجود نمک و ترکیبات مضر در آن ضروری می‌باشد. به‌منظور طبقه‌بندی آب آبیاری، برخی مشخصه‌های شیمیایی آن باید تعیین گردد (ثاقبیان و همکاران ۲۰۱۳). از جمله مشخصه‌هایی که بیشترین اهمیت را در تعیین کیفیت آب برای آبیاری دارند می‌توان به میزان تمرکز نمک‌های محلول در آب، سهم نسبی سدیم به عناصر دیگر مانند منیزیم، کلسیم و پتاسیم، غلظت بور و عناصر دیگر که ممکن است سمی باشند و در برخی شرایط غلظت بی‌کربنات نسبت به



شکل ۲- دیاگرام USSSL (بی‌نام ۱۹۵۴).

شعاع محور (رابطه ۴) و تابع کرنل پییرسون (رابطه ۵) اشاره کرد.

$$K(x_i, x_j) = (x_i^T x_j + 1)^p \quad [۲]$$

$$k(x_i, x_j) = \frac{(x_i^T x_j + 1)^p}{\sqrt{(x_i^T x_i)(x_j^T x_j)}} \quad [۳]$$

$$k(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad [۴]$$

$$k(x_i, x_j) = \frac{1}{1 + \left( 2\sqrt{\|x_i, x_j\|^2} \sqrt{2^{1/w} - 1/\sigma} \right)^{2w}} \quad [۵]$$

هر کدام از این توابع کرنل در ساختار خود دارای پارامترهایی هستند که با نام هایپارامتر شناخته می‌شوند. در کاربرد روش‌های مبتنی بر تابع کرنل علاوه بر انتخاب تابع کرنل بهینه، تعیین مقدار بهینه هایپارامترهای مربوط به هر تابع بسیار حائز اهمیت است.

الگوریتم K-نزدیک‌ترین همسایگی<sup>۶</sup> (K-NN)

این الگوریتم یک روش رایج برای طبقه‌بندی است و بر اساس سنجش فاصله می‌باشد. در این روش ابتدا داده‌های آموزشی و طبقه‌های متناظر با آنها در نظر گرفته می‌شود. سپس برای طبقه‌بندی یک نمونه جدید، فاصله آن تا هریک از نمونه‌های آموزشی محاسبه

طبقه‌بندی کننده بردار پشتیبان<sup>۴</sup> (SVC)

ماشین بردار پشتیبان به عنوان یکی از روش‌های یادگیری با نظارت، اولین بار توسط واپنیک (۱۹۹۵) برای طبقه‌بندی و پیش‌بینی مورد استفاده قرار گرفت. هدف طبقه‌بندی‌کننده بردار پشتیبان، یافتن ابرصفحه‌ای با بهترین توانایی تعمیم جهت جداسازی داده‌های متعلق به دو طبقه، با حاشیه بیشینه می‌باشد. از دیدگاه هندسی، حاشیه برابر با فاصله موجود بین ابرصفحه و نزدیک‌ترین داده نمونه آموزشی است. فاصله هندسی مطلوب میان نمونه X و ابر صفحه بهینه عبارتست از:

$$r = \frac{g(x)}{\|w\|} \quad [۱]$$

در رابطه ۱،  $g(x) = W^T x + b$  عبارتست از تابع تشخیص که توسط ابر صفحه تعریف می‌شود که  $W$  و  $b$  به ترتیب عبارتند از بردار وزن نمونه و یک اسکالر (شهرابی و ذوالقدر شجاعی ۱۳۹۰). در طبقه‌بندی کننده بردار پشتیبان برای دستیابی به نتایج دقیق، انتخاب بهینه تابع کرنل<sup>۵</sup> که وظیفه خطی‌سازی خصوصیات غیرخطی را بر عهده دارد، بسیار مهم می‌باشد. از جمله مهمترین توابع کرنل می‌توان به کرنل چندجمله‌ای ساده (رابطه ۲)، چند جمله‌ای نرمال شده (رابطه ۳)، تابع کرنل

<sup>۶</sup> - K-nearest neighbors

<sup>۴</sup> - Support Vector Classifier

<sup>۵</sup> - Kernel function

از عصبی مصنوعی پرسپترون چند لایه استفاده شده است.

معیارهای ارزیابی مدلها  
آماره کاپا کوهن<sup>۹</sup>

آماره‌ی کاپا به‌عنوان معیار اندازه‌گیری توافق مورد استفاده قرار می‌گیرد. اگر  $P_{ij}$  نسبت کل اجزایی از سری زمانی باشد که به طبقه  $i$  از یک سری و طبقه  $j$  از سری دوم اختصاص داشته باشد. این نسبت‌ها می‌توانند داخل یک ماتریس شکل بگیرند که قطر اصلی آن شامل نسبت‌های منحصر به فرد  $P_{ii}$  اجزایی است که در هر دو سری مطابقت دارند.  $P_0$  مجموع مقادیر  $P_{ii}$  است. با در نظر گرفتن  $P_{e,ii}$  به‌عنوان حاصل‌ضرب مجموع نسبت‌های ردیف و ستون مربوطه، مقدار قابل انتظار توافق، برابر است با جمع مقادیر  $P_{e,ii}$ ؛ این جمع  $P_e$  می‌باشد آماره کاپا به‌صورت رابطه ۷ تعریف می‌شود.

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \quad [7]$$

برای این آماره، مقادیر کمتر از ۰/۴ نشان‌دهنده توافق ضعیف تا خیلی ضعیف، مقادیر ۰/۴ تا ۰/۵۵ نشان‌دهنده توافق متوسط، مقادیر ۰/۵۵ تا ۰/۷ نشان‌دهنده توافق خوب و مقادیر بیش از ۰/۷ نشان‌دهنده توافق خیلی خوب تا عالی می‌باشند (کومار و همکاران ۲۰۰۹، بذرافشان و همکاران ۱۳۹۳).

نرخ خطا

نرخ خطا<sup>۱۰</sup> آماره‌ای ساده جهت بیان دقت طبقه‌بندی روش‌های مختلف می‌باشد که از طریق رابطه ۸ قابل محاسبه است. در این آماره  $N$  برابر با تعداد نمونه‌ای است که در طبقه اشتباه قرار گرفته است و  $T$  برابر با تعداد کل نمونه‌ها می‌باشد.

$$ER \% = \frac{N}{T} * 100 \quad [8]$$

می‌گردد و  $K$  همسایه از بین نزدیک‌ترین همسایه‌ها (نمونه‌های آموزشی) انتخاب می‌شوند. در نهایت نمونه جدید در طبقه‌ای قرار می‌گیرد که بیشتر نمونه‌های موجود در  $K$  همسایه‌اش در آن طبقه هستند. برای تعیین فاصله هر نمونه جدید از نمونه‌های مشاهده‌ای از توابع فاصله مانند همینگ، اقلیدسی و چیبیشف استفاده می‌شود که در این تحقیق از تابع فاصله اقلیدسی استفاده گردید. در این روش، برای دو نمونه با  $n$  ویژگی یعنی  $X = (x_1, x_2, \dots, x_n)$  و  $Y = (y_1, y_2, \dots, y_n)$  فاصله اقلیدسی از رابطه ۶ محاسبه می‌شود (یانگ سو ۲۰۱۱).

$$dist(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad [6]$$

در این الگوریتم تعیین مقدار بهینه پارامتر  $K$  برای دستیابی به بهترین نتیجه بسیار مهم می‌باشد که در تحقیق حاضر برای این منظور از روش سعی و خطا استفاده گردیده است.

شبکه عصبی مصنوعی<sup>۷</sup> (ANN)

یک شبکه عصبی مصنوعی در حالت کلی از سه لایه ورودی، لایه پنهان و لایه خروجی تشکیل می‌شود. تعداد نرون‌های موجود در لایه ورودی و خروجی با توجه به نوع مسأله مورد نظر مشخص می‌شوند. تعداد نرون‌های موجود در لایه پنهان و تعداد این لایه‌ها با آزمون و خطا انتخاب می‌گردند. در حالت کلی پارامترهایی وجود دارند که برای طراحی یک شبکه عصبی مصنوعی قابل تغییر و تنظیم به‌صورت دلخواه می‌باشند تا به نتیجه دلخواه رسیده و خطای شبکه به کمینه برسد. این پارامترها شامل تعداد لایه پنهان، تعداد نرون‌های لایه پنهان، نرخ یادگیری در لایه‌های پنهان و خروجی، مومنتم، اپاک<sup>۸</sup> یا تکرار، تابع تحریک و نحوه ترکیب‌بندی داده‌های ورودی می‌باشند. در این مطالعه

<sup>9</sup> - Cohen's kappa coefficient

<sup>10</sup> - Error rate

<sup>7</sup> - Artificial Neural Network

<sup>8</sup> - Epoch

## نتایج و بحث

در این تحقیق ابتدا طبقه کیفیت آب آبیاری رودخانه آجی‌چای در چهار ایستگاه مورد مطالعه به تفکیک با استفاده از دیاگرام USSL تعیین گردید. سپس در این ۴ ایستگاه در دبی ماهانه و ۱۰ پارامتر هیدروشیمیایی مختلف به‌عنوان ورودی مدل‌های داده‌کاوی و طبقه کیفی آب به‌عنوان خروجی و ویژگی هدف در نظر گرفته شد و سعی گردید امکان کاربرد روش‌های نوین داده‌کاوی در تخمین طبقه کیفی آب مورد ارزیابی قرار گیرد. برای تخمین طبقه کیفی توسط روش‌های داده‌کاوی در حالت کلی تعداد ۸۰ درصد از داده‌ها برای قسمت آموزش مدل و ۲۰ درصد از آن برای قسمت آزمون مدل انتخاب گردید. هنگام تخمین طبقه کیفی توسط طبقه‌بندی‌کننده بردار پشتیبان پس از انجام آزمون و خطا و استفاده از توابع کرنل مختلف با مقادیر هایپر پارامترهای گوناگون مشخص شد که تابع کرنل پیرسون بهترین نتیجه را ارائه می‌دهد و مقادیر  $(\sigma=1, \omega=0/8, C=0/8)$  به‌عنوان مقادیر بهینه این تابع کرنل با آزمون و خطا تعیین گردید. به این ترتیب که میزان

دقت و خطای مدل به ازای مقادیر مختلف  $\sigma$  و  $\omega$  بررسی گردید و مقادیر این پارامترها به‌ازای مدلی که دارای بیش‌ترین دقت و کم‌ترین خطا است بهینه‌سازی شد. همچنین هنگام تخمین طبقه کیفی آب، توسط روش K-نزدیک‌ترین همسایگی پس از انجام آزمون و خطا مشخص گردید بهترین نتیجه هنگامی حاصل می‌گردد که از تعداد ۱ همسایه نزدیک استفاده شود. با بررسی انواع مختلف ساختارهای شبکه‌های عصبی مصنوعی پرسپترون چندلایه مشاهده گردید، بهترین نتایج زمانی حاصل می‌گردد، که از تعداد ۲ لایه پنهان استفاده گردد. همچنین تابع سیگموئید به‌عنوان تابع محرک و مقدار نرخ یادگیری برابر با ۰/۲، ضریب مومنتوم برابر با ۰/۴ و تعداد اپاک برابر با ۱۰۰۰ در نظر گرفته شوند.

## نتایج طبقه‌بندی کیفیت آب آبیاری

چنان‌که قبلاً ذکر گردید تعداد ۴ ایستگاه واقع بر روی رودخانه آجی‌چای مورد بررسی قرار گرفتند که نتایج به‌دست آمده از دیاگرام USSL برای این ایستگاه‌ها طبق جدول ۲ می‌باشد.

جدول ۲- طبقه‌بندی کیفیت آب آبیاری رودخانه آجی‌چای در ایستگاه‌های مورد مطالعه توسط دیاگرام USSL.

تعداد نمونه	تعداد نمونه صادق در هر طبقه کیفی برای هر ایستگاه											
	ایستگاه آخولا			ایستگاه ارزنق			ایستگاه مرکید			ایستگاه ونیار		
تعداد نمونه	آغاز دوره	میان دوره	انتهای دوره	آغاز دوره	میان دوره	انتهای دوره	آغاز دوره	میان دوره	انتهای دوره	آغاز دوره	میان دوره	انتهای دوره
C2S1	-	-	-	۱۰	۶	۱۲	-	-	-	-	-	-
C3S1	۴	-	-	۱۳	۱۲	۹	۵	۲	۲	-	۶	۳
C3S2	-	۲	۳	۶	۵	۴	۱۱	-	۶	۱۵	۶	۵
C3S3	-	-	-	-	-	-	-	-	۱	۱	-	۱
C4S1	۱۳	-	-	۲	-	-	-	۷	-	-	-	۶
C4S2	۳	۲	۶	۴	۷	۱۲	۲	۶	۱	۹	۳	۶
C4S3	۷	۱۰	۶	۳	۱۰	۳	۸	۹	۲	۱۸	۱۶	۹
C4S4	۱۶	۳۰	۲۹	۶	۳	۱	۳۶	۳۷	۴۹	۹۵	۱۰۶	۱۰۷
تعداد نمونه‌ها	۱۳۱	۱۳۰	۱۸۴	۴۱۲								

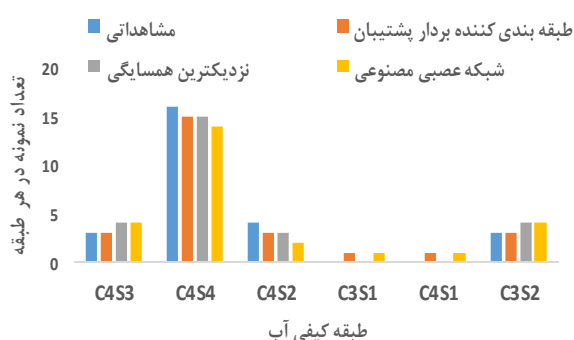
در میان این ۴ ایستگاه، ایستگاه ونیار با قرارگیری بیش از ۸۵ درصد از نمونه‌ها در طبقه کیفی C4S3 و C4S4 دارای بدترین کیفیت و ایستگاه ارزنق با قرارگیری

با بررسی نتایج ارائه شده در جدول ۲ مشخص می‌گردد در مجموع، تمامی ۴ ایستگاه مورد مطالعه در محدوده کیفی بسیار نامناسبی جهت آبیاری قرار دارند.

**جدول ۳- نتایج آماری به دست آمده توسط روش‌های مختلف در تخمین طبقه کیفی آب در ایستگاه آخولا.**

روش داده‌کاوی	آماره کاپا	خطا %	نرخ	تعداد نمونه در طبقه صحیح	رتبه
SVC	۰/۸۷	۷/۶۹	۲۴	۱	۱
K-NN	۰/۸۷	۷/۶۹	۲۴	۱	۱
ANN	۰/۶۷	۱۹/۲۳	۲۱	۲	۲

با بررسی جدول ۳ و شکل ۳ مشاهده می‌گردد هر سه روش توانسته‌اند با دقت بالایی طبقه کیفی آب را مدل‌سازی نمایند اما، روش K-نزدیک‌ترین همسایگی و طبقه‌بندی‌کننده بردار پشتیبان نسبت به شبکه عصبی مصنوعی نتایج بهتری ارائه کرده‌اند زیرا دارای آماره کاپای بیشتر و نرخ خطای کمتری هستند و تعداد نمونه‌ی بیشتری را به صورت صحیح طبقه‌بندی کرده‌اند. با بررسی جداول ۴ و ۵ مشاهده می‌شود که از کل ۲۶ نمونه قسمت آزمون، از میان نمونه‌های طبقه‌بندی شده با روش K-نزدیک‌ترین همسایگی، تنها ۲ نمونه در طبقه اشتباه قرار گرفته‌اند که یکی از آن‌ها به جای قرارگیری در طبقه کیفی C4S4 در طبقه C4S3 قرار گرفته و دیگری به جای قرارگیری در طبقه کیفی C4S2 در طبقه C3S2 قرار گرفته است.



**شکل ۳- مقایسه طبقه‌بندی مشاهداتی و محاسباتی کیفیت آب در ایستگاه آخولا.**

نمونه‌ها در طبقه‌های کیفی به نسبت بهتر دارای بهترین کیفیت می‌باشد. همچنین مشاهده می‌گردد کیفیت آب در ایستگاه‌های مرکید، ونیار و آخولا با گذشت زمان کاهش پیدا کرده است در صورتی‌که کیفیت آب در ایستگاه ارزنق تقریباً دارای ثبات نسبی می‌باشد. به نظر می‌رسد، قرارگیری ایستگاه ارزنق در بالادست سایر ایستگاه‌ها باعث مصون ماندن نسبی آب از آلودگی‌های محیطی بوده است. این درحالی‌ست که احتمالاً تخلیه فضولات انسانی، حیوانی و پساب به آب رودخانه و همچنین کاهش نسبی بارش در این منطقه از عوامل کاهش کیفیت در ایستگاه‌های مرکید، ونیار و آخولا باشد.

برای تخمین طبقه کیفی آب در ایستگاه آخولا در هر سه روش داده‌کاوی مذکور پس از انجام مدل‌سازی طبقه کیفی آب توسط پارامترهای هیدروشیمیایی مختلف همان‌گونه که انتظار می‌رفت مشاهده گردید که، طبقه کیفی آب آبیاری بیشترین تاثیرپذیری را از پارامتر کیفی نسبت جذبی سدیم و سپس هدایت الکتریکی دارا بوده و استفاده از پارامتر هیدروشیمیایی یون منیزیم باعث بهبود دقت نتایج می‌گردد. در جدول ۳ آماره‌های به دست آمده توسط روش‌های مختلف داده‌کاوی برای تخمین طبقه کیفی آب در ایستگاه آخولا و رتبه مربوط به هر کدام از روش‌های داده‌کاوی ارائه شده است. همچنین تعداد نمونه مشاهداتی موجود در هر طبقه کیفی و نمونه‌های محاسباتی به دست آمده از روش‌های مختلف در شکل ۳ برای داده‌های آزمون ایستگاه آخولا نشان داده شده است. همچنین ماتریس اغتشاش به دست آمده از روش‌های طبقه‌بندی کننده بردار پشتیبان و K-نزدیک‌ترین همسایگی که بیانگر نحوه پراکنش نمونه‌ها بین طبقه‌های مختلف می‌باشد در جداول ۴ و ۵ ارائه گردیده است. در این جداول اگر نمونه‌ای خارج از قطر اصلی ماتریس قرار گیرد، بدین معنی است که در طبقه و کلاس اشتباه قرار گرفته است و هرچه تعداد نمونه‌های موجود در قطر اصلی بیشتر باشد نشان‌دهنده دقت و کارایی بیشتر مدل در تخمین طبقه مورد نظر می‌باشد.



برای داده‌های آزمون ایستگاه ارزنق نشان داده شده است.

جدول ۶- نتایج آماری به‌دست آمده توسط روش‌های مختلف در تخمین طبقه کیفی آب در ایستگاه ارزنق.

روش داده‌کاوی	آماره کاپا	نرخ خطا %	تعداد نمونه در طبقه صحیح	رتبه
SVC	۱	۰	۲۶	۱
K-NN	۰/۹۴	۳/۸۴	۲۵	۲
ANN	۱	۰	۲۶	۱



شکل ۴- مقایسه طبقه‌بندی مشاهداتی و محاسباتی کیفیت آب در ایستگاه ارزنق.

با توجه به نتایج به‌دست آمده در این ایستگاه روش‌های شبکه عصبی مصنوعی و طبقه‌بندی‌کننده بردار پشتیبان با ارائه مقادیر بالای ضریب کاپا و صفر خطا تمامی نمونه‌ها را به‌درستی طبقه‌بندی نموده‌اند. ماتریس اغتشاش به‌دست آمده از روش‌های شبکه عصبی مصنوعی و طبقه‌بندی‌کننده بردار پشتیبان در جدول ۷ ارائه گردیده است.

با بررسی جدول ۷ مشاهده می‌شود که از کل ۲۶ نمونه قسمت آزمون، روش‌های طبقه‌بندی‌کننده بردار پشتیبان و شبکه عصبی مصنوعی موفق شده‌اند تمامی نمونه‌ها را به‌درستی طبقه‌بندی کنند و قرارگیری تمامی نمونه‌ها در قطر اصلی نشان‌دهنده توافق ۱۰۰ درصدی مقادیر مشاهداتی و تخمین‌زده شده طبقه کیفی آب می‌باشد.

جدول ۴- ماتریس اغتشاش به‌دست آمده از روش SVC برای نمونه‌های آزمون ایستگاه آخولا.

طبقه	C4S3	C4S4	C4S2	C3S1	C4S1	C3S2
C4S3	۳					
C4S4		۱۵			۱	
C4S2			۳			۱
C3S1				۰		
C4S1					۰	
C3S2						۳

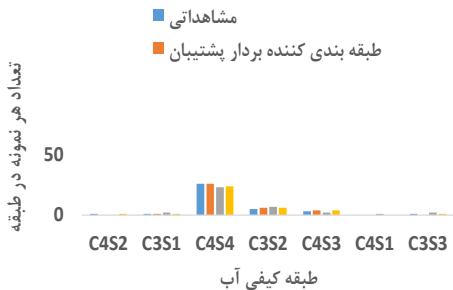
همچنین با بررسی ماتریس اغتشاش به‌دست آمده از طبقه‌بندی‌کننده بردار پشتیبان مشاهده می‌شود، تنها ۲ نمونه در طبقه اشتباه قرار گرفته‌اند که یکی از آن‌ها به‌جای قرارگیری در طبقه کیفی C4S4 در طبقه C4S1 قرار گرفته و دیگری به‌جای قرارگیری در طبقه کیفی C4S2 در طبقه C3S2 قرار گرفته است. در ایستگاه ارزنق در هر سه روش داده‌کاوی مذکور پس از انجام مدل‌سازی طبقه کیفی آب توسط پارامترهای هیدروشیمیایی مختلف مشاهده گردید که، بهترین نتیجه زمانی حاصل می‌گردد که تنها از پارامتر نسبت جذبی سدیم استفاده شود.

جدول ۵- ماتریس اغتشاش به‌دست آمده از روش K-NN برای نمونه‌های آزمون ایستگاه آخولا.

طبقه	C4S3	C4S4	C4S2	C3S1	C4S1	C3S2
C4S3	۳					
C4S4		۱۵			۱	
C4S2			۳			۱
C3S1				۰		
C4S1					۰	
C3S2						۳

در جدول ۶ آماره‌های به‌دست آمده توسط روش‌های مختلف داده‌کاوی برای تخمین طبقه کیفی آب در ایستگاه ارزنق و رتبه مربوط به هر کدام از روش‌های داده‌کاوی ارائه شده است. همچنین تعداد نمونه مشاهداتی موجود در هر طبقه کیفی و نمونه‌های محاسباتی به‌دست آمده از روش‌های مختلف در شکل ۴

روش طبقه‌بندی‌کننده بردار پشتیبان در جدول ۹ ارائه گردیده است.



شکل ۵- مقایسه طبقه‌بندی مشاهداتی و محاسباتی کیفیت آب در ایستگاه مرکید.

جدول ۹- ماتریس اغتشاش به‌دست آمده از روش SVC برای نمونه‌های آزمون ایستگاه مرکید.

طبقه	C4S2	C3S1	C4S4	C3S2	C4S3	C4S1	C3S3
C4S2	۰			۱			
C3S1		۱					
C4S4			۲۶				
C3S2				۵			
C4S3					۳		
C4S1						۰	
C3S3							۱

با بررسی ماتریس اغتشاش ارائه شده در جدول ۹ مشاهده می‌شود که از کل ۳۷ نمونه قسمت آزمون، از میان نمونه‌های طبقه‌بندی‌شده با روش طبقه‌بندی‌کننده بردار پشتیبان تنها ۲ نمونه در طبقه اشتباه قرار گرفته‌اند که یکی از آن‌ها به‌جای قرارگیری در طبقه کیفی C4S2 در طبقه C4S3 قرار گرفته و دیگری به‌جای قرارگیری در طبقه کیفی C4S2 در طبقه C3S2 قرار گرفته است. در صورتی‌که تعداد نمونه قرار گرفته در طبقه اشتباه برای روش شبکه عصبی مصنوعی برابر ۴ نمونه و برای روش K-نزدیک‌ترین همسایگی برابر ۶ نمونه می‌باشد. برای تخمین طبقه کیفی آب در ایستگاه

نتایج به‌دست آمده نشان داد، در ایستگاه مرکید استفاده از پارامتر هیدروشیمیایی یون منیزیم باعث بهبود دقت نتایج می‌گردد.

جدول ۷- ماتریس اغتشاش به‌دست آمده از روش‌های بهینه برای نمونه‌های آزمون ایستگاه ارزنق.

طبقه	C3S1	C2S1	C3S2	C4S3	C4S4	C4S1	C4S2
C4S2	۵						
C4S1		۶					
C4S4			۲				
C4S3				۱			
C3S2					۰		
C2S1						۰	
C3S1							۱۲

در جدول ۸ آماره‌های به‌دست‌آمده توسط روش‌های مختلف داده‌کاوی برای تخمین طبقه کیفی آب در ایستگاه مرکید و رتبه مربوط به هر کدام از روش‌های داده‌کاوی ارائه شده است. همچنین تعداد نمونه مشاهداتی موجود در هر طبقه کیفی و نمونه‌های محاسباتی به‌دست‌آمده از روش‌های مختلف در شکل ۵ برای داده‌های آزمون ایستگاه مرکید نشان داده شده است.

جدول ۸- نتایج آماری به‌دست آمده توسط روش‌های مختلف در تخمین طبقه کیفی آب در ایستگاه مرکید.

روش داده‌کاوی	آماره کاپا	نرخ خطا %	تعداد نمونه در طبقه صحیح	رتبه
SVC	۰/۸۸	۵/۴۰	۳۵	۱
K-NN	۰/۶۹	۱۶/۲۱	۳۱	۲
ANN	۰/۷۷	۱۰/۸۱	۳۳	۲

با بررسی جدول ۸ و شکل ۵ مشاهده می‌گردد هر سه روش بررسی شده توانسته‌اند با دقت بالایی طبقه کیفی آب را مدلسازی نمایند اما روش طبقه‌بندی‌کننده بردار پشتیبان با ارائه مقادیر بیشتر آماره کاپا و کمتر نرخ خطا نسبت به سایر روش‌ها نتایج بهتری دارد. ماتریس اغتشاش به‌دست آمده از

طبقه کیفی آب را مدل‌سازی نمایند اما از بین روش‌های بررسی شده روش K-نزدیک‌ترین همسایگی با ارائه مقادیر بیشتر آماره کاپا و کمتر نرخ خطا نسبت به طبقه‌بندی‌کننده بردار پشتیبان و شبکه عصبی مصنوعی نتایج بهتری را در این ایستگاه ارائه کرده است. ماتریس اغتشاش به‌دست‌آمده از روش K-نزدیک‌ترین همسایگی در جدول ۱۱ ارائه گردیده است.

با بررسی ماتریس اغتشاش ارائه شده در جدول ۱۱ مشاهده می‌شود از میان ۸۲ نمونه طبقه‌بندی شده با روش K-نزدیک‌ترین همسایگی تنها ۱ نمونه در طبقه اشتباه قرار گرفته‌اند که به‌جای قرارگیری در طبقه کیفی C3S2 در طبقه C4S3 قرار گرفته است. در صورتی‌که تعداد نمونه قرار گرفته در طبقه اشتباه برای روش طبقه‌بندی‌کننده بردار پشتیبان برابر ۳ نمونه و برای روش شبکه عصبی مصنوعی برابر ۴ نمونه می‌باشد.

جدول ۱۱- ماتریس اغتشاش به‌دست آمده از روش K-NN برای نمونه‌های آزمون ایستگاه ونیار.

طبقه	C4S1	C3S1	C4S2	C3S2	C4S3	C4S4
C4S1	۰					
C3S1		۱				
C4S2			۱			
C3S2				۵		
C4S3					۵	
C4S4						۶۹

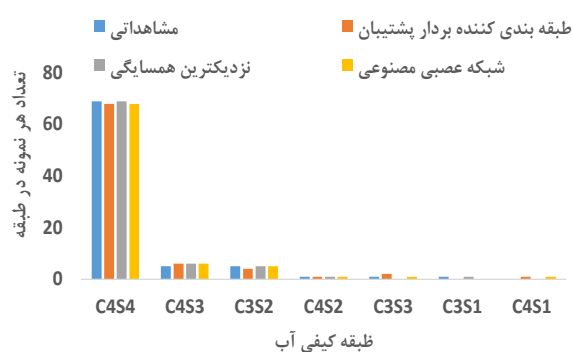
### نتیجه‌گیری کلی

در این تحقیق کیفیت آب آبیاری در رودخانه آجی‌چای در ۴ ایستگاه آخولا، ارزنق، مرکید و ونیار مورد مطالعه قرار گرفت و توسط دیگرام USSL طبقه‌بندی شد. بررسی نتایج نشان داد که در حالت کلی، آب در این ایستگاه‌ها دارای مقادیر بالای پارامترهای نسبت جذبی سدیم و هدایت الکتریکی می‌باشد که به‌علت عبور رودخانه از مسیرهای دارای بستر نمکی می‌باشد. ارزیابی عملکرد روش‌های نوین داده‌کاوی شامل

ونیار توسط روش‌های داده‌کاوی در هر سه روش داده‌کاوی پس از آزمون پارامترهای هیدروشیمیایی مختلف مشاهده گردید بهترین نتیجه زمانی حاصل می‌شود که تنها از پارامترهای هدایت الکتریکی و نسبت جذبی سدیم استفاده گردد. در جدول ۱۰ آماره‌های به‌دست آمده توسط روش‌های مختلف داده‌کاوی برای تخمین طبقه کیفی آب در ایستگاه ونیار و رتبه مربوط به هر کدام از روش‌های داده‌کاوی ارائه شده است. همچنین تعداد نمونه مشاهداتی موجود در هر طبقه کیفی و نمونه‌های محاسباتی به‌دست‌آمده از روش‌های مختلف در شکل ۶ برای داده‌های آزمون ایستگاه ونیار نشان داده شده است.

جدول ۱۰- نتایج آماری به‌دست آمده توسط روش‌های مختلف در تخمین طبقه کیفی آب در ایستگاه ونیار.

روش داده‌کاوی	آماره کاپا	نرخ خطا %	تعداد نمونه در طبقه صحیح	رتبه
SVC	۰/۸۷	۶/۶۵	۷۹	۲
K-NN	۰/۹۵	۱/۲۲	۸۱	۱
ANN	۰/۸۲	۴/۸۸	۷۸	۳



شکل ۶- مقایسه طبقه‌بندی مشاهداتی و محاسباتی کیفیت آب در ایستگاه ونیار.

با بررسی جدول ۱۰ و شکل ۶ مشاهده می‌گردد هر سه روش بررسی شده توانسته‌اند با دقت بالایی

سطحی بر اساس دیاگرام شوری و با استفاده از روش درخت تصمیم پرداختند. آنها برای این منظور در مجموع از پارامتر دبی ماهانه و ۱۰ پارامتر هیدروشیمیایی مختلف که در ایستگاه‌های سنجش کیفیت آب اندازه‌گیری می‌شود، استفاده نمودند. نتایج نشان داد، پارامترهای هدایت الکتریکی، نسبت جذبی سدیم، میزان کل جامدات نامحلول در آب و مجموع کاتیون‌ها بیشترین تأثیر را در تخمین طبقه کیفی آب دارا می‌باشند و روش درخت تصمیم به خوبی می‌تواند برای طبقه‌بندی کیفیت آب به کار گرفته شود. نتایج تحقیق حاضر حاکی از قابل رقابت بودن این روش با روش درخت تصمیم می‌باشد. همچنین در تحقیق دیگری مدرسی و عراقی نژاد (۲۰۱۴) با استفاده از روش‌های شبکه عصبی احتمالاتی، ماشین بردار پشتیبان و نزدیکترین همسایگی به طبقه‌بندی کیفیت آب بر اساس استاندارد CCME پرداختند. آنها برای این منظور از پارامترهای هیدروشیمیایی نیترات و کلرید استفاده نمودند. نتایج نشان داد، ماشین بردار پشتیبان با استفاده از تابع کرنل چند جمله‌ای دارای بیشترین دقت از بین روش‌های استفاده شده می‌باشد. نتایج تحقیق حاضر از نظر دقت در موقعیت نسبتاً خوبی نسبت به تحقیق مدرسی و عراقی نژاد (۲۰۱۴) قرار دارد. در حالت کلی تمامی روش‌های بررسی شده و در رأس آنها روش طبقه‌بندی‌کننده بردار پشتیبان به عنوان روشی کارآمد در این زمینه برای کاربرد در موارد مشابه توصیه می‌گردد. همچنین استفاده از روش‌های یادگیری ماشینی برای طبقه‌بندی کیفیت آب باعث صرفه‌جویی در زمان و هزینه می‌گردد.

طبقه‌بندی‌کننده بردار پشتیبان، K-نزدیکترین همسایگی و شبکه عصبی مصنوعی نشان‌دهنده دقت نسبتاً بالا و عملکرد مناسب این روش‌ها در تعیین طبقه کیفی آب می‌باشد. بررسی نتایج نشان می‌دهد، هر کدام از روش‌های مورد استفاده در این پژوهش در ایستگاه‌های مختلف دارای عملکرد متفاوتی بوده‌اند. مبنای روش‌های یادگیری ماشینی استفاده از داده‌های بازه آموزش جهت ساخت مدل می‌باشد، از این روی با توجه به تغییرات فیزیکی و شیمیایی موجود در محدوده داده‌های اختصاص یافته برای آموزش در هر یک از ایستگاه‌ها، طبیعتاً روش‌های موجود نیز نتایج و کارایی متفاوتی برای ایستگاه‌های مختلف ارائه خواهند نمود. به همین دلیل برای تعیین کارآمدترین روش، از میانگین رتبه هر روش در ایستگاه‌های مورد مطالعه استفاده شده است. چنان‌که انتظار می‌رفت، در طبقه‌بندی کیفیت آب در ایستگاه‌های مختلف مشاهده گردید که پارامترهای هدایت الکتریکی و نسبت جذب سدیم بیشترین تأثیر را بر روی طبقه‌بندی کیفیت آب دارا هستند ولی در برخی موارد استفاده از یون منیزیم باعث بهبود نتایج می‌گردد. نتایج به‌دست‌آمده نشان داد که روش طبقه‌بندی‌کننده بردار پشتیبان که با بهره‌گیری از توابع کرنل، توانایی بالایی در حل مسائل مختلف دارد، با میانگین رتبه ۱/۲۵ به‌عنوان کارآمدترین روش داده‌کاوی و پس از آن روش K-نزدیکترین همسایگی با میانگین رتبه ۱/۷۵ و شبکه عصبی مصنوعی با میانگین رتبه ۲ به‌عنوان روش‌هایی مناسب جهت تعیین طبقه کیفی آب معرفی می‌گردند. در تحقیق مشابهی ستاری و همکاران (۱۳۹۳) به طبقه‌بندی کیفیت آب‌های

#### منابع مورد استفاده

بانزاد ح، کمالی م، امیر مرادی ک و علیائی ا، ۱۳۹۲. تخمین برخی پارامترهای کیفی رودخانه‌ها با استفاده از مدل هیبرید شبکه‌های عصبی-موجکی (منطقه مطالعاتی: رودخانه جاجرود تهران و قره سو کرمانشاه). مجله سلامت و محیط، جلد ۶، شماره ۳، صفحه‌های ۲۷۷ تا ۲۹۴.

بذرافشان ا، سلاجقه ع، مهدوی م، بذرافشان ج و فاتحی مرج ا، ۱۳۹۳. بررسی کارآیی مدل‌های هیبرید شبکه عصبی مصنوعی- استوکاستیک در پیش‌بینی خشک‌سالی‌های هیدرولوژیکی با استفاده از آماره کاپا (مطالعه موردی: حوزه آبخیز رودخانه گاماسیاب). علوم و مهندسی آبخیزداری ایران، جلد ۸، شماره ۲۷، صفحه‌های ۳۵ تا ۴۸.

ستاری م، عباسقلی نائب‌زاد م و میرعباسی نجف آبادی ر، ۱۳۹۳. پیش‌بینی کیفیت آب‌های سطحی با استفاده از روش درخت تصمیم. مهندسی آبیاری و آب، جلد ۴، شماره ۱۵، صفحه‌های ۷۶ تا ۸۸.

شهرابی ج و ذوالقدر شجاعی ع، ۱۳۹۰. داده‌کاوی پیشرفته، انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر.

میرزاوند م، قاسمیه ه، ساداتی‌نژاد س ج و اکبری م، ۱۳۹۴. مقایسه مدل‌های شبکه عصبی مصنوعی و رگرسیون چندمتغیره در تخمین تغییرات کیفی آب زیرزمینی (مطالعه موردی: آبخوان کاشان). دانش آب و خاک، جلد ۲۵، شماره ۲، صفحه‌های ۲۰۷ تا ۲۲۰.

- Anonymous, 1954. Diagnosis and Improvement of Saline and Alkali Soils: U.S. Dept. Agric. Handbook No.60, 160 p.
- Chebud Y, Naja GM, Rivero RG and Mellese AM, 2012. Water quality monitoring using remote sensing and artificial neural network. *Water, Air and Soil Pollution* 223: 4875-4887.
- Dogan E, Sengorur B and Koklu R, 2009. Modeling biological oxygen demand of the Melen river in Turkey using an artificial neural network technique. *Journal of Environmental Management* 90: 1229-1235.
- Kumar MN, Murthy CS, Sesha Sai MVR and Roy PS, 2009. On The use of standardized precipitation index (SPI) for drought intensity assessment. *Meteorological Applications* 16: 381- 389 .
- Modaresi F, Araghinejad S, 2014. A comparative assessment of support vector machines, probabilistic neural networks and K-nearest neighbors algorithms for water quality classification. *Water Resources management* 28: 4095-4111.
- Saghebian SM, Sattari MT, Mirabbasi R and Pal M, 2014. Ground water quality classification by decision tree method in Ardebil region, Iran. *Arabian Journal of Geosciences* 7(11): 4767-4777.
- Sahu M, Mahapatra SS, Sahu HB and Patel RK, 2011. Prediction of water quality index using neuro fuzzy inference system. *Water Quality, Exposure and Health* 3: 175-191.
- Sattari MT, Rezazadeh Joudi A and Kusiak A, 2015. Estimation of water quality parameters with data-driven models. *American Water Works Association* 108(4): 232-239.
- Sengorur B, Koklu R and Ates A, 2015. Water quality assessment using artificial intelligence techniques: SOM and ANN- A case study of Melen River. *Water Quality, Exposure and Health* 7(4): 469-490.
- Vapnik VN, 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Yang Su M, 2011. Real-Time anomaly detection systems for denial-of-service attacks by weighted K-nearest neighbor classifiers. *Expert Systems with Applications* 38: 3492-3498.