Research Article

DOI: 10.22034/ijche.2023.408953.1334

This journal is an open access journal licensed under an Attribution-NonCommercial-NoDerivatives 4.0 International license(CC BY-NC-ND 4.0).

Measuring the Quality of Industrial Water Used in a Chemical Plant Using Artificial Intelligence (Logistic Regression, Naive Bayesian, Support Vector Machine, Random Forest, and Decision Tree)

F. Sarmast Alizadeh[1], N. Esfandiari[2*]

1- M. Sc. Student of Chemical Engineering, Department of Chemical Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

2- Associate Professor of Chemical Engineering, Department of Chemical Engineering, Marvdasht Branch,Islamic Azad University, Marvdasht, Iran

Email: nadia.esfandiari@iau.ac.ir

Abstract

Continuous measurement of factory water quality is important. Current methods of measuring water quality are not efficient enough. In this research, a new method using the concepts of artificial intelligence and machine learning has been proposed to solve the mentioned challenges. The proposed research method has been trained and validated using 472 samples of chemical data in MATLAB software. Each data sample has 6 input attributes (pH, conductivity, water hardness, total water-soluble solids, free chlorine, and alkalinity) and one output attribute (target). The parameters of disturbance matrix, precision, accuracy, and readability have been used to evaluate the efficiency of water quality measurement. The highest accuracy is related to the random forest method. The decision tree, simple Bayes, and vector machine methods are the same. The most refreshing rate is related to the decision tree method. The artificial intelligence method of the proposed decision tree with an accuracy equal to 70%, accuracy equal to 98%, and recall equal to 96% compared to logistic regression methods, Naive Bayesian method, support vector machine, and random forest, shows more efficiency and less error.

Received: 27 July 2023
Accepted: 10 October 2023
Page Number: 90-101

Keywords:

Quality,
Industrial Water,
Chemical Plant,
Artificial Intelligence

Please Cite this Article Using:

Sarmast Alizadeh, F., & Esfandiari, N. (2024). Measuring the Quality of Industrial Water Used in a Chemical Plant Using arTificial Intelligence (Logistic Regression, Naive Bayesian, Support Vector Machine, Random forest, and Decision Tree). *Iranian Chemical Engineering Journal*, *23*(135), 90-101, [In Persian].

مقـــالـــــه پـــــژوهشی

DOI: 10.22034/ijche.2023.408953.1334

This journal is an open access journal licensed under an Attribution-NonCommercial-NoDerivatives 4.0 International license(CC BY-NC-ND 4.0).

سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی بااستفاده‌از هوش مصنوعی

(وایازش لجستیک، بیز ساده، ماشین بردار پشتیبان، جنگل تصادفی و درخت تصمیم)

فرید سرمست علیزاده[۱]، نادیا اسفندیاری[۲*]

۱- دانشجوی کارشناسی ارشد مهندسی شیمی، گروه مهندسی شیمی، واحد مرودشت، دانشگاه آزاد اسلامی، مرودشت، ایران

۲- دانشیار مهندسی شیمی، گروه مهندسی شیمی، واحد مرودشت، دانشگاه آزاد اسلامی، مرودشت، ایران

پیام نگار: nadia.esfandiari@iau.ac.ir

چکیده

سنجش مستمر کیفیت آب کارخانه بسیار مهم است. روش‌های کنونی سنجش کیفیت آب از کارایی کافی برخوردار نیست. در این پژوهش برای حل چالش‌های ذکرشده، کاهش خطای انسانی، آنالیز دقیق‌تر و افزایش سرعت تشخیص، روشی نوین بااستفاده‌از مفاهیم هوش مصنوعی و یادگیری ماشین مطرح شده‌است. روش پیشنهادی پژوهش بااستفاده‌از ۴۷۲ نمونه‌داده‌های شیمیایی در نرم‌افزار MATLAB آموزش و اعتبار شده‌است. هر یک از نمونه‌داده‌ها دارای ۶ ویژگی ورودی (pH، هدایت‌سنجی، سختی آب، کل مواد جامد محلول در آب، کلر آزاد و قلیائیت) و یک ویژگی خروجی (هدف) هستند. برای ارزیابی میزان کارایی سنجش کیفیت آب، از معیارهای ماتریس اغتشاش، دقت، صحت و بازخوانی استفاده شده‌است. بیشترین دقت مربوط‌به روش جنگل تصادفی است. صحت روش‌های درخت تصمیم، بیز ساده و ماشین بردار پشتیبان با هم برابر است. بیشترین میزان بازخوانی مربوط‌به روش درخت تصمیم است. روش هوش مصنوعی درخت تصمیم پیشنهادی با دقتی معادل با ۷۰ درصد، صحت معادل با ۹۸ درصد و بازخوانی برابر با ۹۶ درصد نسبت‌به روش‌های وایازش لجستیک، بیز ساده، ماشین بردار پشتیبان و جنگل تصادفی، کارایی بیشتر و خطای کمتری را نشان داد.

تاریخ دریافت: ۱۴۰۲/۰۵/۰۵
تاریخ پذیرش: ۱۴۰۲/۰۷/۱۸
شماره صفحات: ۹۰ تا ۱۰۱

کلیدواژه‌ها:
کیفیت،
آب صنعتی،
کارخانهٔ شیمیایی،
هوش مصنوعی

* مرودشت، دانشگاه آزاد اسلامی، گروه مهندسی شیمی، واحد مرودشت

استناد به مقاله:
سرمست علیزاده، فرید، و اسفندیاری، نادیا. (۱۴۰۳). سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی بااستفاده‌از هوش مصنوعی (وایازش لجستیک، بیز ساده، ماشین بردار پشتیبان، جنگل تصادفی و درخت تصمیم)، *نشریه مهندسی شیمی ایران*، ۲۳ (۱۳۵)، ۹۰-۱۰۱.

۱. مقدمه

آب صنعتی مورد نیاز در صنایع و کارخانه‌های شیمیایی ایرانی از گروه آب‌های صنعتی بسیار حساس [۱] است و از روش‌های بسیار پیشرفتهٔ تصفیه برای آن‌ها استفاده می‌شود. کمترین خطای متخصصان و سامانه‌های مانیتورینگ سنجش کیفیت آب صنعتی مورد نیاز صنایع و کارخانه‌های شیمیایی ایرانی، ممکن است موجب به‌وجودآمدن خسارت‌های جبران‌ناپذیر مالی و حیاتی بر کارکنان، جوش‌آورهای پرفشار، برج خنک‌کننده و غیره شود. باتوجه‌به این موارد، ارائهٔ روش‌هایی کارا و دقیق برای سنجش کیفیت آب صنعتی مورد نیاز در صنایع و کارخانه‌های شیمیایی ایرانی بسیار حیاتی و ضروری است که این موضوع اهمیت و ضرورت انجام پژوهش را نمایان می‌کند. مطابق بررسی‌های انجام‌شده، در صنایع و کارخانه‌های شیمیایی ایرانی برای سنجش کیفیت آب صنعتی مورد نیاز، عموماً از روش‌های آزمایشگاهی و غیر دیجیتالی استفاده می‌شود که بسیار زمان‌بر است و به‌دلیل استفاده‌از عامل انسانی و اشتباه در تعریف قوانین جدید، برخط نبودن سامانه‌ها، عدم قابلیت یادگیری و غیره دارای خطای زیاد و کارایی ناکافی است. برای حل این چالش‌ها، حذف دخالت عامل انسانی و افزودن قابلیت یادگیری به‌روش‌های فعلی، برای نخستین بار اندیشهٔ استفاده‌از روش‌های مبتنی‌بر هوش مصنوعی و یادگیری ماشین برای بهبود کارایی سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی داخلی پیشنهاد شده‌است. درخت تصمیم[1] یکی از ابزارها و شگردهایی است که در مهارت‌های داده‌کاوی بسیار پرکاربرد است. زمانی‌که حجم داده‌ها بسیار بالا باشد این روش می‌تواند مفید باشد. درخت تصمیم یک روش رایج برای نشان دادن فرایند تصمیم‌گیری به‌وسیلهٔ ساختاری درخت‌مانند و شاخه‌دار است. مدل درخت تصمیم بصری است و درک فرایند تجزیه‌وتحلیل آن آسان است[۳و۲]. یکی از قدیمی‌ترین و پرکاربردترین مدل‌های یادگیری، وایازش[2] است؛ فرایندی آماری است که برای تخمین روابط موجود بین متغیرها استفاده می‌شود. چنان‌که از نام این الگوریتم پیدا است، الگوریتم جنگل تصادفی[3] یک طبقه‌بندی است که شامل تعدادی درخت تصمیم در زیرمجموعه‌های مختلف مجموعه‌داده قرار دارد و برای بهبود دقت پیش‌بینی آن مجموعه‌داده، میانگین می‌گیرد. جنگل تصادفی به‌جای تکیه‌بر یک درخت تصمیم، پیش‌بینی را از هر درخت و براساس اکثریت آرا پیش‌بینی می‌کند و نتیجهٔ نهایی را به‌عنوان خروجی درنظر می‌گیرد. تعداد بیشتر درختان در جنگل منجربه دقت بالاتری می‌شود و از بروز مشکل جلوگیری می‌کند[۴]. بیز ساده[4] از ساده‌ترین الگوریتم‌های پیش‌بینی است که دقت قابل قبولی هم دارد. دقت آن را می‌توان بااستفاده‌از برآورد چگالی کرنل به‌صورت چشم‌گیری بالا برد. شیوهٔ یادگیری در روش بیز ساده از نوع یادگیری با نظارت است[۵]. یکی از روش‌های یادگیری با نظارت ماشین بردار پشتیبان[5] است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کاری دسته‌بندی‌کننده، دسته‌بندی خطی داده‌ها است. در تقسیم خطی داده‌ها از ابرصفحه‌ای که حاشیهٔ اطمینان بیشتری داشته‌باشد، استفاده می‌شود. خط بهینه برای داده‌ها به‌وسیلهٔ روش‌های برنامه‌سازی غبرخطی که روش‌های شناخته‌شده‌ای در حل مسائل محدودیت‌دار هستند، انجام می‌گیرد[۶].

فرایند انعقاد در تصفیه‌خانهٔ شرکت پتروشیمی فجر بااستفاده‌از شبکه‌های عصبی مصنوعی بررسی شد[۷]. از شبکهٔ عصبی مصنوعی برای پیش‌بینی جریان افقی مایع_مایع استفاده شد[۸]. خان[6] و همکاران[۹]، براساس روش‌های وایازش مؤلفهٔ اصلی[7] و روی‌کرد طبقه‌بندی‌کنندهٔ تقویت گرادیان[8] اقدام به پیش‌بینی و طبقه‌بندی کیفیت آب کرده‌اند؛ نتایج ارزیابی آنان، دقت[9] پیش‌بینی ۹۵ درصدی را برای روش وایازش مؤلفه‌های اصلی و دقت طبقه‌بندی ۱۰۰ درصدی را برای روش طبقه‌بندی‌کنندهٔ تقویت گرادیان نشان داده‌است که عملکرد معتبری را درمقایسه‌با مدل‌های پیشرفته نشان می‌دهد. آزور[10] و همکاران[۱۰]، از الگوریتم‌های یادگیری ماشین برای پیش‌بینی کارامد کیفیت آب استفاده کرده‌اند. مطابق نتایج ارزیابی‌های آنان، اثبات شد که استفاده‌از الگوریتم‌های وایازش چندگانه[11] در پیش‌بینی شاخص کیفیت آب مهم و مؤثر بوده‌است. علاوه‌بر این، استفاده‌از شبکهٔ عصبی مصنوعی[12]

1. Decision Tree
2. Logestic Regression
3. Random Forest
4. Naive Bayesian
5. Support Vector Machine (SVM)
6. Khan
7. Principal Component Regression
8. Gradient Boosting Classifier
9. Accuracy
10. Azrour
11. Multiple Regression Algorithms
12. Artificial Neural Network

کارامدترین راه را برای طبقه‌بندی کیفیت آب فراهم می‌کند. وو[1] و وانگ[2][۱۱]، یک مدل ترکیبی[3] برای پیش‌بینی کیفیت آب براساس روش‌های شبکهٔ عصبی مصنوعی، تبدیل موجک گسسته[4] (DWT) و حافظهٔ کوتاه‌مدت[5] (LSTM) مطرح کرده‌اند. نتایج ارزیابی‌ها نشان داده‌است که مدل ANN-WT-LSTM ارائه‌شده در این مقاله بهتر از مدل‌های قبلی در بسیاری از شاخص‌های ارزیابی عمل کرده‌است. بنابراین، روش‌های پیشنهادی این مطالعه می‌تواند پشتیبانی فنی و مرجع عملی مناسبی برای پایش کیفیت آب و مدیریت رودخانهٔ جین‌جیانگ و سایر حوضه‌ها فراهم کند. پراساد[6] و همکاران[۱۲]، تحلیل و پیش‌بینی کیفیت آب را بااستفاده‌از روش‌های یادگیری عمیق[7] و یادگیری عمیق خودکار[8] (AutoDL) مطرح کرده‌اند مطابق نتایج ارزیابی‌های مقاله، دقت مدل یادگیری عمیق معمولی در تحلیل و پیش‌بینی کیفیت آب ۹۹-۹۸٪ برای تحلیل و پیش‌بینی بوده، درحالی‌که روش یادگیری عمیق خودکار حدود ۹۸٪-۹۶ بوده‌است. با این حال، مدل یادگیری عمیق خودکار وظیفهٔ یافتن مدل یادگیری عمیق مناسب را آسان کرد و بدون دخالت دستی کارایی بهتری را نشان داد.

باتوجه‌به پژوهش‌های این مطالعه برای نخستین بار بااستفاده‌از داده‌های سنجش کیفیت آب صنعتی که از یک کارخانهٔ شیمیایی داخلی استخراج شده‌است، به‌کمک روش‌های هوش مصنوعی و بدون تعریف قوانین و پروتکل‌های کیفیت آب در سامانه‌های فعلی به‌وسیلهٔ عامل انسانی و با دقت و کارایی کافی روشی ارائه شد که قابلیت یادگیری خودکار این قوانین از داده‌های موجود و اعمال آن‌ها را داشته‌باشد و از ایجاد خسارات مالی و زمانی ناشی‌از ناکارامدی روش‌های فعلی پیش‌گیری نماید.

۲. روش کار

مجموعه‌داده‌های مورد نیاز این پژوهش برای آموزش، آزمون و اجرای روش هوش مصنوعی پیشنهادی پژوهش برای سنجش کیفیت آب صنعتی از یک کارخانهٔ شیمیایی داخلی، به‌صورت تصادفی و متعادل گرداوری شده‌است تا دارای توزیع نرمال داده‌ای در دسته‌های مختلف مد نظر کیفیت آب صنعتی باشد. تعداد نمونه‌های گرداوری‌شده در مجموعه‌داده‌های پژوهش برابر با ۴۷۲ نمونه است. هر یک از نمونه‌داده‌های موجود در جامعهٔ داده‌ای پژوهش دارای ۶ ویژگی ورودی (اسیدی و قلیایی بودن آب، هدایت‌سنجی، سختی آب، کل مواد جامد محلول در آب، کلر آزاد، قلیائیت) و یک ویژگی خروجی (هدف) دارای دو کلاس کیفیت کافی و ناکافی آب صنعتی هستند. هر ویژگی ورودی دارای محدودهٔ استاندارد و معین است. این مقادیر با محدوده مقایسه می‌شود و معیار مثبت صحیح، منفی کاذب، منفی صحیح و مثبت کاذب تعریف و مقادیر آن مشخص می‌شود. سپس معیارهای صحت، دقت و بازخوانی حساب می‌شود و از روی آن‌ها در خصوص کارایی هر یک از روش‌ها تصمیم‌گیری می‌شود. جزئیات کامل داده‌های گرداوری‌شدهٔ پژوهش و هم‌چنین محدودهٔ ۶ مشخصهٔ سنجش کیفیت آب صنعتی کارخانهٔ شیمیایی به‌شرح است.

۲-۱ ویژگی‌های ورودی

۲-۱-۱ اسیدی و قلیایی‌بودن آب[9]

در بسیاری از صنایع و کارخانه‌های تولیدی، به‌دلیل شرایط خاص، pH آب در برج خنک‌کننده درحالت خنثی قرار ندارد؛ به‌عنوان مثال، در برج خنک‌کنندهٔ صنایع روی، آب (محلول حاوی یون روی) بسیار اسیدی است و در کارخانه‌های دارای خط تولید اسیدسولفوریک[10] و اسید فسفریک[11] نیز آب برج خنک‌کننده[12] به‌شدت اسیدی است، اما در برخی کاربری‌ها آب برج خنک‌کننده می‌تواند قلیایی باشد. باید توجه داشت که اسیدی و قلیایی بودن و pH غیر معمول آب موجب خوردگی‌های شدید در قطعات برج‌های خنک‌کنندهٔ معمولی می‌شود و به همین دلیل در زمان طراحی کلیهٔ قطعات باید باتوجه‌به شرایط خاص آب در آن پروژه طراحی شوند[۱۳]. در این پژوهش انجام‌گرفته محدودهٔ pH مقدار۷/۸ تا ۸/۳ در نظر گرفته شده‌است.

۲-۱-۲ هدایت یا رسانایی[13]

مقدار هدایت الکتریکی آب نشان‌دهندهٔ میزان وجود املاح در آب

1. Wo
2. Wang
3. Hybrid Model
4. Discrete Wavelet Transform
5. Long Short-Term Memory
6. Prasad
7. Deep Learning Techniques
8. Auto Deep Learning Techniques
9. Water pH
10. Sulfuric Acid
11. Phosphoric Acid
12. Cooling Tower
13. Conductivity

است. بهترین روش برای تشخیص نمونهٔ آب و فاضلاب، سنجش میزان هدایت الکتریکی است؛ زیرا آنیون‌ها[1] و کاتیون‌های[2] موجود در فاضلاب بیشتر از آب است. در سنجش هدایت الکتریکی که حساس‌ترین آزمایش است، نمونه باید تا زمانی با الکترود دستگاه در تماس باشد که عدد ثابت و دقیق‌تری را نشان دهد. واحد ارزیابی کیفیت هدایت‌سنجی آن میکروزیمنس بر سانتی‌متر[3] است[۱۴].

۲-۱-۳ سختی آب[4]

سختی آب، یکی از عواملی است که بر عملکرد برج‌های خنک‌کنندهٔ صنعتی و مسکونی تأثیر می‌گذارد. سختی آب بر اثر وجود یون‌های فلزاتی مانند کلسیم، منیزیم و در برخی موارد آلومینیوم یا آهن و غیره ایجاد می‌شود. تقریباً در بیشتر موارد، سختی آب به‌دلیل حل‌شدن یون‌های کلسیم و منیزیم است که البته وجود یون کلسیم تأثیر بیشتری دارد[۱۵]. میزان سختی قابل قبول در این پژوهش بیشینهٔ ۷۰۰ میلی‌گرم در لیتر است.

۲-۱-۴ کل مواد جامد محلول در آب (TDS)[5]

این ویژگی نشان‌دهندهٔ کل مواد جامد محلول در آب است که برابر با مجموع غلظت همهٔ یون‌های موجود در آب است. مواد محلول در آب ممکن است ازنظر ماهیت، آلی یا معدنی باشند. واحد ارزیابی کیفیت TDS، میلی‌گرم در لیتر است[۱۶].

۲-۱-۵ کلر آزاد

باتوجه‌به این‌که تمامی منابع آب حاوی مقادیر بسیار زیادی از ریزاندامگان[6] متفاوت است، برای جلوگیری از ایجاد و رشد جلبک برروی خطوط انتقال، آب قبل از ورودی به سامانهٔ خنک‌کننده با کلر ضدعفونی می‌شود. باتوجه‌به این‌که بازده سامانهٔ کلرزنی تنها بااستفاده‌از ارزیابی کیفیت کلر باقی‌مانده ممکن است، درکنار سامانه‌های تزریق کلر، سامانهٔ پایش برخط کلر باقی‌مانده نیز الزامی است[۱۷]. کلر آزاد در این پژوهش بیشینهٔ ۵ برابر مقدار کلر در آب خام است.

۲-۱-۶ قلیائیت[7]

این ویژگی، ازنظر شیمیایی نشانگر مقدار ظرفیت خنثی‌شدن آب به‌وسیلهٔ اسید است. خنثی‌شدن آب به‌معنی رسیدن pH آب به حدود ۷ است. قلیائیت آب با دو روش قلیائیت ساده و قلیائیت کل ارزیابی می‌شود[۱۸].

۲-۲ ویژگی‌های ورودی

مجموعه‌داده‌های پژوهش دارای دو ویژگی خروجی هدف است که به‌صورت واقعی سنجش شده و شامل کیفیت کافی آب صنعتی مصرفی کارخانهٔ شیمیایی و کیفیت ناکافی آب صنعتی مصرفی کارخانه شیمیایی است. در بخش بعد، معماری مفهومی روش پیشنهادی پژوهش، برای حل مسئلهٔ سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی تشریح شده‌است.

۲-۳ معیارهای ارزیابی پژوهش

برای تجزیه‌وتحلیل اطلاعات و ارزیابی کارایی روش هوش مصنوعی پیشنهادی پژوهش به‌منظور سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی، ابتدا به استخراج ماتریس اغتشاش[8] مطابق با جدول (۱) پرداخته خواهدشد.

در ادامه از معیارهای ارائه‌شده در جدول (۱) برای تجزیه‌وتحلیل اطلاعات و ارزیابی استفاده شد و معیارهای ارزیابی پژوهش شامل دقت (معادلهٔ (۱))، صحت[9] (معادلهٔ (۲))، بازخوانی[10] (معادلهٔ (۳) و F1 (معادلهٔ (۴)) حساب شد. از آن‌ها برای ارزیابی میزان کارایی روش هوش مصنوعی پیشنهادی پژوهش و مقایسهٔ روش‌های سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی استخراج شد. مقادیر بیشتر معیارهای صحت، دقت و بازخوانی ارائه‌شده در جدول (۲)، نشان‌دهندهٔ افزایش دقت و کارایی روش هوش مصنوعی پیشنهادی یا روش‌های مورد مقایسهٔ سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی است. برای برنامه‌نویسی از نرم‌افزار MATLAB استفاده شد.

1. Anions
2. Cations
3. Microsiemens Per Centimeter
4. Total Hardness
5. Total Dissolved Solid
6. Microorganisms
7. Alkality
8. Confusion Matrix
9. Precision
10. Recall

جدول ۱. ماتریس اغتشاش مورد استفاده برای ارزیابی روش هوش مصنوعی پژوهش[۱۹].

Table 1. The confusion matrix used to evaluate the research artificial intelligence method [19].

		Diagnostic quality of industrial water used in a chemical plant	
		Sufficient quality	Insufficient quality
The actual quality of the industrial water used in the chemical plant	Sufficient quality	True positive (TP)	False positive (FP)
	Insufficient quality	False negative (FN)	True negative (TN)

جدول ۲. معیارهای مورد استفاده برای ارزیابی کارایی روش هوش مصنوعی پژوهش[۱۹].

Table 2. The criteria used to evaluate the effectiveness of the research artificial intelligence method [19].

	Equation	
Accuracy	Accuracy = (TP + TN) / (TP + TN + FP + FN)	(1)
Precision	Precision = (TP) / (TP + FP)	(2)
Recall	Recall = (TP) / (TP + FN)	(3)
F1	F1 = (2 × Precision × Recall) / (Precision + Recall)	(4)

۳. نتایج و بحث

در این بخش نتایج میزان کارایی روش‌های هوش مصنوعی مورد مقایسهٔ پژوهش (درخت تصمیم[۲۰]، وایازش لجستیک[۱][۲۱]، بیز ساده[۲][۲۲]، ماشین بردار پشتیبان[۳][۲۳]، جنگل تصادفی[۴][۲۴]) برای سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی ارائه و تجزیه‌وتحلیل شده‌است.

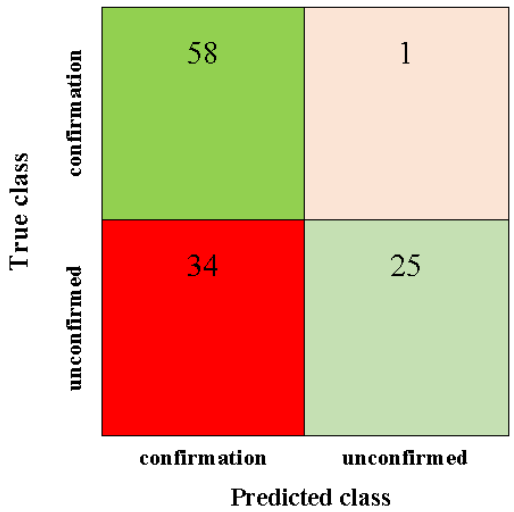
۳-۱ ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم

در این بخش برای ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم ابتدا به بررسی ماتریس اغتشاش این روش پرداخته شد و در ادامه نتایج معیارهای دقت، صحت و بازخوانی این روش استخراج و تحلیل شد. در شکل (۱)، ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم ارائه شده‌است.

همان‌گونه که در شکل (۱) پیدا است، تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت کافی بوده و با روش درخت تصمیم، به‌درستی دارای کیفیت کافی تشخیص داده شده‌است (مثبت صحیح)، برابر با ۵۸ است. تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت ناکافی بوده و با روش درخت تصمیم، به‌اشتباه دارای کیفیت کافی تشخیص داده شده‌است (منفی کاذب)، برابر با ۳۴ است. تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت ناکافی بوده و با روش درخت تصمیم، به‌درستی دارای کیفیت ناکافی تشخیص داده شده‌است (منفی صحیح)، برابر با ۲۵ است و تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت کافی بوده و با روش درخت تصمیم، به‌اشتباه دارای کیفیت ناکافی تشخیص داده شده‌است (مثبت کاذب)، برابر با ۱ است.

در ادامه بااستفاده‌از نتایج ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم مطابق با شکل (۱)، نتایج ارزیابی کارایی این روش مطابق با معیارهای دقت، صحت و بازخوانی در جدول (۳) قابل مشاهده‌است.

1. Logestic Regression
2. Naive Bayesian
3. Support Vector Machine(SVM)
4. Random Forest



شکل ۱. ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم.

Figure 1. Confusion matrix of measuring the quality of industrial water used in a chemical plant by the decision tree method.

جدول ۳. نتایج ارزیابی کارایی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم.

Table 3. The results of evaluating the effectiveness of measuring the quality of industrial water used in a chemical plant by the decision tree method.

Methods	True Negative	False Negative	True Positive	False Positive	Accuracy	Precision	Recall	F1
Decision tree	25	34	58	1	70%	98%	96%	96.98%

معیار دقت برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم، برابر با ۷۰ درصد است. به‌بیانی دیگر، این روش در ۷۰ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، کیفیت کافی و ناکافی آب صنعتی را به‌صورت صحیح سنجش کرده‌است، معیار صحت برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم، برابر با ۹۸ درصد است. به‌بیانی دیگر در این روش ۹۸ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، به‌درستی کیفیت کافی آب صنعتی را سنجش داده و معیار بازخوانی برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش درخت تصمیم برابر با ۹۶ درصد است. به‌بیانی دیگر، این روش در ۹۶ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، به‌طور جامعی سنجش کیفیت آب صنعتی را انجام داده‌است. معیار دقت، تفاوتی بین خطای منفی کاذب و خطای مثبت کاذب درنظر نمی‌گیرد؛ بنابراین برای تصمیم‌گیری، محاسبهٔ معیارهای بیشتری مثل صحت و بازخوانی ضروری است. صحت درصد بلی‌های درست الگوریتم را حساب می‌کند. بازخوانی محاسبهٔ پوشش برروی کل داده‌ها را مد نظر دارد. معیار سومی به‌نام معیار F1 هم وجود دارد که درواقع ترکیبی از معیارهای دقت و صحت است که در این روش برابر ۹۶/۹۸ درصد به‌دست آمده که دقت قابل قبول این روش را برای پیش‌بینی نشان می‌دهد.

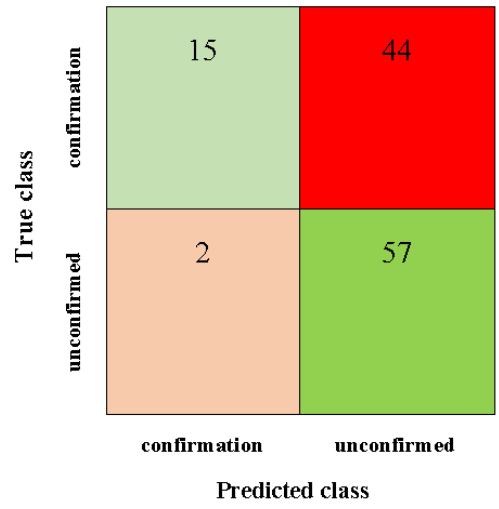
۳-۲ ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با وایازش لجستیک

در این بخش برای ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش وایازش لجستیک، ابتدا به بررسی ماتریس اغتشاش این روش پرداخته و در ادامه نتایج معیارهای دقت، صحت و بازخوانی این روش استخراج و تحلیل شد. در شکل (۲)، ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش وایازش لجستیک ارائه شده‌است.

چنان‌که در شکل (۲) قابل مشاهده‌است، تعداد نمونه‌های آب

صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت کافی بوده و با روش وایازش لجستیک، به‌درستی دارای کیفیت کافی تشخیص داده شده (مثبت صحیح)، برابر با ۱۵ است. تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت ناکافی بوده و با روش وایازش لجستیک، به‌اشتباه دارای کیفیت کافی تشخیص داده شده (منفی کاذب) برابر با ۲ است. تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت ناکافی بوده و با روش وایازش لجستیک به‌درستی دارای کیفیت ناکافی تشخیص داده شده (منفی صحیح)، برابر با ۵۷ است و تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت کافی بوده و با روش وایازش لجستیک، به‌اشتباه دارای کیفیت ناکافی تشخیص داده شده (مثبت کاذب) برابر با ۴۴ است. در ادامه، نتایج ارزیابی کارایی این روش در جدول (۴) قابل مشاهده‌است.

معیار دقت در روش وایازش لجستیک برابر با ۶۲ درصد حساب شده‌است. معیار صحت برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش وایازش لجستیک برابر با ۲۵ درصد است. یعنی ۲۵ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، به‌درستی کیفیت آب صنعتی را سنجش کرده و معیار بازخوانی برابر با ۶۵ درصد است. به‌بیانی دیگر، این روش در ۵۶ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، به‌طور جامعی سنجش کیفیت آب صنعتی را انجام داده‌است. مقدار F1 برابر ۳۴/۵۶ درصد است که نشان می‌دهد این روش دقت قابل قبولی برای سنجش در این مطالعه ندارد.



شکل ۲. ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش وایازش لجستیک.

Figure 2. Confusion matrix of measuring the quality of industrial water used by the chemical plant by the logistic regression method.

جدول ۴. نتایج ارزیابی کارایی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش وایازش لجستیک.

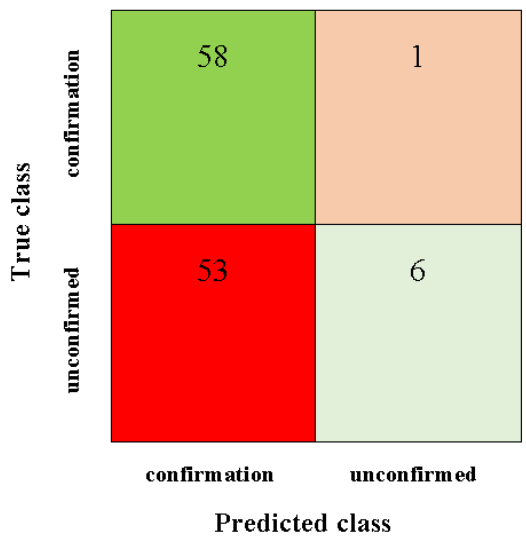
Table 4. The results of evaluating the effectiveness of measuring the quality of industrial water used in a chemical plant by the logistic regression method.

Methods	True Negative	False Negative	True Positive	False Positive	Accuracy	Precision	Recall	F1
Logistic regression	57	2	15	44	61%	25%	56%	34.56

۳-۳ ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده

در این بخش برای ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده، ابتدا به بررسی ماتریس اغتشاش این روش پرداخته و در ادامه نتایج معیارهای دقت، صحت و بازخوانی این روش حساب شد. در شکل (۳)، ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده ارائه شده‌است. تعداد نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی که در واقعیت دارای کیفیت کافی بوده و با روش نیوبیزین، به‌درستی دارای کیفیت کافی تشخیص داده شده (مثبت صحیح)، برابر با ۵۸ است. تعداد نمونه‌هایی که در واقعیت دارای کیفیت ناکافی بوده و با روش نیوبیزین، به‌اشتباه دارای کیفیت کافی تشخیص داده شده (منفی کاذب)، برابر با ۵۳ است. تعداد نمونه‌هایی که در واقعیت دارای کیفیت ناکافی بوده و با روش نیوبیزین، به‌درستی دارای کیفیت ناکافی تشخیص داده شده (منفی صحیح)، برابر با ۶ است و نمونه‌هایی که در واقعیت دارای کیفیت کافی بوده و با روش نیوبیزین، به‌اشتباه دارای کیفیت ناکافی تشخیص داده شده (مثبت کاذب) برابر با ۱ است.

در ادامه بااستفاده‌از نتایج ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده مطابق با شکل (۳)، نتایج ارزیابی کارایی این روش حساب شده و در جدول (۵) آمده‌است. معیار دقت برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده برابر با ۵۴ درصد است. یعنی ۵۴ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، کیفیت کافی و ناکافی آب صنعتی را به‌صورت صحیح سنجش کرده‌است. در این روش مقدار F1 برابر ۹۱/۶ درصد به‌دست آمده که دقت قابل قبول این روش را برای پیش‌بینی نشان می‌دهد.



شکل ۳. ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده.

Figure 3. Confusion matrix for measuring the quality of industrial water used in a chemical plant by the Naive Bayesian method.

جدول ۵. نتایج ارزیابی کارایی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش بیز ساده.

Table 4. The results of evaluating the effectiveness of measuring the quality of industrial water used in a chemical plant by the Naive Bayesian method.

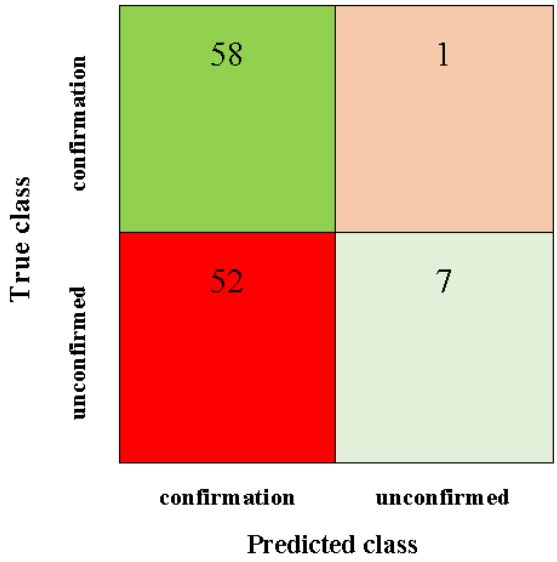
Methods	True Negative	False Negative	True Positive	False Positive	Accuracy	Precision	Recall	F1
Naive Bayesian	6	53	58	1	54%	98%	86%	91.6%

۳-۴ ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش ماشین بردار پشتیبان

همانند روش‌های قبلی ماتریس اغتشاش در شکل (۴) و مقادیر دقت، صحت و بازخوانی در جدول (۶) حساب شد. همان‌گونه‌که در شکل (۴) قابل مشاهده‌است، مثبت صحیح برابر با ۵۸، منفی کاذب برابر با ۵۲، منفی صحیح برابر با ۷ و تعداد مثبت کاذب برابر با ۱ است. معیار دقت برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش ماشین بردار پشتیبان برابر با ۵۵ درصد است. به‌بیانی دیگر، این روش در ۵۵ درصد از کل نمونه‌های آب صنعتی مصرفی کارخانهٔ شیمیایی موجود در جامعهٔ داده‌ای پژوهش، کیفیت کافی و ناکافی آب صنعتی را به‌صورت صحیح سنجش کرده‌است. مقدار F1 برابر ۹۲/۷۳ درصد است که نشان می‌دهد این روش دقت قابل قبولی برای سنجش در این مطالعه دارد.

۳-۵ ارزیابی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش جنگل تصادفی

ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش جنگل تصادفی ارائه شد. مثبت صحیح، منفی کاذب، منفی صحیح و مثبت کاذب برابر با ۵۰، ۱۶، ۴۳ و ۹ به‌دست آمد. دقت مقدار ۷۹ درصد، صحت ۸۵ درصد و بازخوانی ۸۳ درصد حساب شد. معیار F1 عدد ۸۳/۹۸ درصد به‌دست آمد که بالاتر از روش وایازش لجستیک و کمتر از درخت تصمیم، بیز ساده و ماشین بردار پشتیبان است.



شکل ۴. ماتریس اغتشاش سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش ماشین بردار پشتیبان.

Figure 4. Confusion matrix of measuring the quality of industrial water used in a chemical plant by support vector machine(SVM) method.

جدول ۶. نتایج ارزیابی کارایی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش ماشین بردار پشتیبان.

Table 6. The results of evaluating the effectiveness of measuring the quality of industrial water used in a chemical plant by support vector machine(SVM) method.

Methods	True Negative	False Negative	True Positive	False Positive	Accuracy	Precision	Recall	F1
support vector machine	7	52	58	1	55%	98%	88%	92.73%

۳-۶ مقایسهٔ روش‌ها

در جدول (۷) به ارزیابی کارایی روش‌های مورد مقایسهٔ پژوهش پرداخته شده‌است. دقت روش جنگل تصادفی برابر ۷۹ درصد و بالاتر از سایر روش‌ها است؛ اما مقادیر صحت و بازخوانی آن از روش درخت تصمیم کمتر است. درواقع، هنگامی‌که دقت و صحت در کنار هم قرار بگیرند، نتایج اندازه‌گیری، تکرارپذیر و نزدیک به مقدار واقعی خواهد بود. بنابراین درخت تصمیم به‌عنوان روش برتر شناخته شد. معیار دقت یک نقطهٔ ضعف دارد. این معیار، نمی‌تواند تفاوتی بین خطای منفی کاذب و خطای مثبت کاذب داشته‌باشد. درواقع این معیار، تمامی خطاها را یکسان درنظر می‌گیرد. برای حل این نقطهٔ ضعف، دو معیار صحت و بازخوانی تعریف شدند. صحت، تعیین می‌کند که الگوریتم چند درصدِ «بلی»هایش درست بوده‌است؛ اما معیار کاربردی بازخوانی به‌دنبال محاسبهٔ پوشش برروی کل داده‌ها است. هدف اصلی معیار بازخوانی برخلاف معیار صحت برروی داده‌هایی است که واقعاً «بلی» بوده‌اند. معیار دیگری به‌نام معیار F1 وجود دارد که درواقع ترکیبی از معیارهای دقت و صحت است. همان‌طورکه در جدول (۷) نشان داده شده، درخت تصمیم بالاترین مقدار F1 را دارد و می‌تواند به‌عنوان بهترین روش در این مطالعه عمل کند.

۴. نتیجه‌گیری

بیشترین دقت مربوط‌به روش جنگل تصادفی است؛ اما مقادیر صحت و بازخوانی آن از روش درخت تصمیم کمتر است. صحت روش‌های درخت تصمیم، بیز ساده و ماشین بردار پشتیبان با هم برابر است. بیشترین میزان بازخوانی مربوط‌به روش درخت تصمیم است. معیار دقت یک مشکل دارد و تفاوتی بین خطای منفی کاذب و خطای مثبت کاذب قائل نمی‌شود. برای حل این مشکل، دو معیار صحت و بازخوانی تعریف شدند. بنابراین، این سه معیار در کنار هم قابل بررسی است. ازطرفی، ترکیب متعادلی بین معیارهای دقت و صحت به‌نام F1 تعریف شده‌است. باتوجه‌به نتایج حاصل‌شده، روش هوش مصنوعی درخت تصمیم با دقتی معادل با ۷۰ درصد، صحت برابر با ۹۸ درصد، بازخوانی برابر با ۹۶ درصد و F1 برابر ۹۶/۹۸ درصد کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی را سنجش کرده و در جمع‌بندی نهایی بهترین نتایج را حاصل کرده‌است. باتوجه‌به نتایج، روش هوش مصنوعی درخت تصمیم به‌عنوان روش پیشنهادی پژوهش گزینش شده و توانسته‌است تا دقت و کارایی بیشتری را برای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی نسبت‌به روش‌های هوش مصنوعی مورد مقایسهٔ وایازش لجستیک، بیز ساده، ماشین بردار پشتیبان و جنگل تصادفی کسب کند. درنهایت، بهبود کارایی و دقت سنجش کیفیت آب صنعتی مصرفی یک کارخانهٔ شیمیایی با روش درخت تصمیم پیشنهادی موجب کاهش خطای سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی می‌شود.

جدول ۷. ارزیابی کارایی سنجش کیفیت آب صنعتی مصرفی کارخانهٔ شیمیایی با روش‌های مورد مقایسهٔ پژوهش.

Table 7. Evaluating the effectiveness of measuring the quality of industrial water used in a chemical plant by the methods compared in the research.

Methods	True Negative	False Negative	True Positive	False Positive	Accuracy	Precision	Recall	F1
Decision tree	25	34	58	1	70%	98%	96%	96.98%
logistic regression	57	2	15	44	61%	25%	56%	34.56%
Naive Bayesian	6	53	58	1	54%	98%	86%	91.60%
support vector machine	7	52	58	1	55%	98%	88%	92.73%
random forest	43	16	50	9	79%	85%	83%	83.98%

مراجع

[1] Torabian, A., Hasani, A. H., Rahmanipour, A., & Mahjouri, M. (2009). Guidance on quality classification of raw water, wastewater and return water for industrial and recreational use. *Planning and Budgeting of the Country, 462*. In Persian.

[2] Rokach, L., & Oded, M. (2014). Data Mining with Decision Trees: Theory and Applications. *World Scientific Publishing Co., Inc.1060 Main Street Suite 1B River Edge, NJ United States.*

[3] Lee, J. H., Lee, J. Y., Lee, M. H., Lee, M. Y., Kim, Y. W., Hyung, J. S, Kim, K. B., Cha, Y. K., & Koo, J. Y. (2022). Development of a short-term water quality prediction model for urban rivers using real-time water quality data. *Water Supply, 22* (4), 4082–4097.

[4] Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial Intelligence risk measurement. *Expert Systems with Applications, 235*, 121220.

[5] Wang, S., Ren, J., & Bai, R. (2023). A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes. *Expert Systems with Applications, 225*, 120094.

[6] Lou, C., & Xie, X. (2023). Multi-view intuitionistic fuzzy support vector machines with insensitive pinball loss for classification of noisy data. *Neurocomputing, 549*, 126458.

[7] Yousefi, M., & Yasin, M. (2015). Optimization of Industrial Water Pretreatment Operational Processes Using Artificial Neural Network and Genetic Algorithm. *Iranian Chemical Engineering Journal, 14* (79), 13-22. In Persian.

[8] Azizi, S., & Karimi, H. (2016). Prediction of horizontal liquid-liquid two-phase flow patterns using artificial neural networkIanian. *Chemical Engineering Journal, 14* (82), 65-74. In Persian.

[9] Khan, M. S., Islam, I. N., Uddin, J., Islam, S., & Nasir, M. K. (2022). Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *Journal of King Saud University-Computer and Information Sciences, 34*, 4781-4773.

[10] Azrour, M., Mabrouki, J., Fattah, G., Guezzaz, A., & Aziz, F. (2022). Machine learning algorithms for efficient water quality prediction. *Modeling Earth Systems and Environment, 8*, 2801-2793.

[11] Wu, J., & Wang, Z. (2022). A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water, 14* (4), 610.

[12] Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Harshana, S., Srividya, S. J., Harrinei, K., & Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of The Total Environment, 821*, 153311.

[13] Galster, H. (1991). pH measurement. *VCH (Verlagsgesellschaft), New York*.

[14] Light, T. S., Licht, S., Bevilacqua, A. C., & Morash, K. R. (2004). The fundamental conductivity and resistivity of water. *Electrochemical and solid-state letters, 8* (1), 16.

[15] Lerga, T. M., & O'Sullivan, C. K. (2008). Rapid determination of total hardness in water using fluorescent molecular aptamer beacon. *analytica chimica acta, 610*, 105-111.

[16] Wang, B. B. (2021). Research on drinking water purification technologies for household use by reducing total dissolved solids (TDS). *Plos one, 16*, e0257865.

[17] Aoki, T., & Munemori, M. (1983). Continuous flow determination of free chlorine in water. *Analytical Chemistry, 55*, 212-209.

[18] Cox, D. (1995). Water Quality: pH and Alkalinity. University of Massachusetts Extension, Department of Plant and Soil Science, Massa.

[19] Pal, O. K. (2022). The Quality of Drinkable Water using Machine Learning Techniques. *International Journal of Advanced Engineering Research and Science, 9* (6), 16-23.

[20] Song, Y. Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry, 27*(2), 130-135.

[21] Nick, T. G., & Campbell, K. M. (2007). Logistic regression, *Topics in biostatistics*. 273-301.

[22] Leung K. M. (2007). Naive Bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, 123-156.

[23] Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification Thinking with Examples for Effective Learning. *Springer New York, NY*.

[24] Devetyarov, D., & Nouretdinov, I. (2010). Prediction with Confidence Based on a Random Forest Classifier. *Artificial Intelligence Applications and Innovations*, 44-37.