JEHE Alborz University of Medical Sciences Journal of Environmental Health Engineering , 2024; 11(4):380-393 Original Article

Effluent quality prediction of one of the municipal wastewater treatment plants using machine learning algorithms

Received: 17 April 2024, Accepted: 11 June 2024

Amin Bagheri [1] , Mohsen Sadani [2] , Mohammadreza Karimi[3*]

[1]*Department of Health, Safety and Environment, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2] *Department of Environmental Health, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[3]*MPH student School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

***Corresponding Author:**

reza575261@gmail.com

How to Cite This Article:

Bagheri A, Sadani M, Karimi M. Effluent quality prediction of one of the urban wastewater treatment plants using machine learning algorithms. Journal of Environmental Health Engineering. 2024;11(4):380-93.

DOI:
10.61186/jehe.11.4.380

ABSTRACT

Background: Mathematical and statistical simulators can significantly reduce the management costs of wastewater treatment systems. This research aimed to predict the effluent quality of an urban wastewater treatment plant in Tehran using machine learning algorithms from 2017 to 2021.

Materials and Methods: This descriptive-analytical study utilized monitoring data from the influent and effluent of the wastewater treatment plant, which were prepared for analysis. In the second stage, the data were refined, processed, and converted into dummy variables to facilitate entry into data mining algorithms. The Artificial Neural Network (ANN) algorithm and the M5 tree model were then evaluated to identify the best model for predicting the concentration of Chemical Oxygen Demand (COD) in the effluent. In this process, 70% of the data were allocated for training and 30% for validation using Python software. The best model was selected based on regression analysis, comparing the R^2 and RMSE indices.

Results: The findings indicated that the ANN model, with a coefficient of determination (R^2) of 0.72, outperformed the M5 model, which had an R^2 of 0.68, in predicting the output COD concentration—an indicator of the treatment plant's efficiency. Additionally, regression analysis revealed that BOD_5 and TSS exhibited the highest correlation with COD_{out}.

Conclusion: The results of the ANN and M5 models were within an acceptable range based on statistical indicators, demonstrating their potential for effectively estimating data in wastewater treatment plants.

Keywords: Machine learning algorithms, artificial neural network, M5 model tree, regression, chemical oxygen demand

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

Copyright © 2024 The Authors.
Published by Alborz University of Medical Sciences
This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License. (https://creativecommons.org/licenses/by-nc/4.0/).Non-commercial uses of the work are permitted,provided the original work is properly cited.

JEHE
Alborz University of Medical Sciences
مجله مهندسی بهداشت محیط، دوره ۱۱، شماره ۴، تابستان ۱۴۰۳، صفحات ۳۸۰ تا۳۹۳
مقاله پژوهشی

پیش‌بینی کیفیت پساب یکی از تصفیه‌خانه‌های فاضلاب شهری شهر تهران با استفاده از الگوریتم‌های یادگیری ماشین

تاریخ دریافت: ۱۴۰۳/۰۱/۲۹ تاریخ پذیرش: ۱۴۰۳/۰۳/۲۲

امین باقری[۱]، محسن سعدانی[۲]، محمدرضا کریمی[۳*]

[۱] استادیارگروه سلامت، ایمنی و محیط زیست، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران،ایران

[۲] استادیارگروه مهندسی بهداشت محیط، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران،ایران

[۳] دانشجوی گروه MPH، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران،ایران

***پست الکترونیکی نویسنده مسئول:**
reza575261@gmail.com

نحوه استناد به این مقاله:

Bagheri A, Sadani M, Karimi M. Effluent quality prediction of one of the urban wastewater treatment plants using machine learning algorithms. Journal of Environmental Health Engineering. 2024;11(4):380-93.

DOI:
10.61186/jehe.11.4.380

چکیده

زمینه و هدف: برای کاهش هزینه‌های مدیریت سیستم‌های تصفیه فاضلاب، می‌توان از شبیه‌سازهای ریاضی و آماری استفاده نمود. این پژوهش باهدف پیش‌بینی کیفیت پساب یکی از تصفیه‌خانه‌های فاضلاب شهری شهر تهران با استفاده از الگوریتم‌های یادگیری ماشین طی سال‌های ۱۳۹۶ تا ۱۴۰۰ انجام گردید.

مواد و روش ها: این مطالعه یک پژوهش توصیفی – تحلیلی است که در آن اطلاعات سیستم‌های پایش ورودی و خروجی تصفیه‌خانه فاضلاب دریافت و پاک‌سازی داده‌ها انجام گرفت. در مرحله دوم تبدیل داده‌ها به‌منظور آماده‌سازی ورود آن‌ها به الگوریتم‌های داده‌کاوی از طرق پالایش، پردازش و ایجاد متغیر ساختگی (Dummy) انجام شد. سپس، الگوریتم شبکه عصبی مصنوعی (ANN) و مدل درختی M5 به‌منظور یافتن بهترین مدل جهت پیش‌بینی غلظت COD در خروجی تصفیه‌خانه مورد بررسی قرار گرفت؛ در این راستا ۷۰ درصد داده‌ها جهت یادگیری ماشین و ۳۰ درصد به‌منظور اعتبارسنجی در نرم‌افزار پایتون مورداستفاده قرار گرفت. درنهایت با مدل رگرسیونی و مقایسه شاخص‌های R^2 و RMSE به انتخاب بهترین مدل پرداخته شد.

یافته ها: نتایج نشان داد که ANN با ضریب تعیین ۰/۷۲عملکرد بهتری نسبت به مدل M5 با ضریب۰/۶۸در پیش‌بینی غلظت COD خروجی به‌عنوان شاخص کارایی تصفیه‌خانه دارد.

همچنین بر اساس نتایج تحلیل رگرسیون از بین متغیرهای مستقل BOD5e و TSSe بیشترین همبستگی را با CODout داشتند.

نتیجه گیری: در پژوهش حاضر، نتایج مدل ANN و M5 بر اساس شاخص‌های آماری در محدوده قابل‌قبولی قرار گرفتند و می‌توان با موفقیت برای تخمین داده‌ها در تصفیه‌خانه‌های فاضلاب استفاده کرد.

واژه های کلیدی: الگوریتم‌های یادگیری ماشین، شبکه عصبی مصنوعی، درخت مدل M5، رگرسیون، اکسیژن‌خواهی شیمیایی

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

مقدمه

با توجه به رشد جمعیت، ارتقاء سطح زندگی و کاهش سرانه آب و از سوی دیگر محدود بودن منابع آب، دسترسی به منابع آبی جدید ضروری است. ازآنجایی‌که تصفیه پساب نسبت به سایر روش‌های تهیه آب شیرین، مقرون‌به‌صرفه می‌باشد بهتر آن است که پساب مورد مصرف مجدد قرار گرفته و به هدر نرود[1]. از طرفی افزایش نگرانی در مورد مسائل زیست محیطی متخصصین را تشویق کرده است که توجه خود را روی عملکرد و کنترل درست تصفیه خانه های فاضلاب متمرکز کنند [2]. با توجه به اینکه تعداد تصفیه خانه های فاضلاب و اهمیت آنها رو به افزایش می باشد، لذا پیش بینی اولیه و سپس تجزیه و تحلیل پارامترهای آلاینده براساس روشهای جدید ضروری است [3]. عملیات و کنترل ایمن یک تصفیه خانه فاضلاب می تواند با توسعه یک ابزار مدلسازی برای پیش بینی عملکرد تصفیه خانه براساس مشاهدات گذشته از پارامترهای کلیدی خاص کیفیت به دست آید. تصفیه خانه فاضلاب شامل چندین فرآیند فیزیکی، بیولوژیکی و شیمیایی پیچیده است. اغلب این فرآیندها رفتار غیرخطی را نشان می دهند که توصیف آن با مدلهای ریاضی خطی دشوار است[4]. برای ارزیابی و شبیه سازی فرآیند می توان از روش های ریاضی، داده کاوی و نرم افزار استفاده کرد. اکسیژن خواهی بیوشیمیایی ۵ روزه (BOD_5) و اکسیژن خواهی شیمیایی (COD) از جمله مهمترین پارامترهای کیفی فاضلاب است. اندازه گیری و پیش بینی این پارامترها جهت پیش بینی میزان آلودگی و ارزیابی عملکرد تصفیه خانه فاضلاب ضروری است[5]. با توجه به محدود بودن منابع آب در جهان، استفاده از آب های تصفیه شده به خصوص در مناطق خشک و نیمه خشکی مانند ایران ضروری به نظر می رسد[6]. ارزیابی های زیست محیطی نیاز به توسعه و تجزیه و تحلیل حجم زیادی از داده ها داشته و تصمیم گیرندگان باید مقادیر قابل توجهی از اطلاعات را فقط برای به دست آوردن مجموعه ای از گزینه های قابل دوام مدیریت کنند [7]. اخیراً الگوریتم هوش مصنوعی و مدل‌های داده محور توجه محققان را به خود جلب کرده و به‌طور گسترده‌ای به‌عنوان جایگزینی برای مدل‌های مفهومی مورداستفاده قرارگرفته است. این مدل‌ها کاربرپسند تر از مدل‌های عددی هستند و می‌توانند داده‌ها را از منابع مختلف با مقیاس‌های مختلف و ساختارهای غیرخطی برای پیش‌بینی فرآیندهای پیچیده بپذیرند[8]. در سال‌های اخیر، رویکردهای هوش مصنوعی (AI) مانند شبکه عصبی مصنوعی(ANN)[1] ، سیستم تداخل عصبی-فازی تطبیقی (ANFIS)، برنامه‌ریزی بیان ژن (GEP)، رگرسیون چند جمله‌ای تکاملی (EPR)، درخت مدل M5 و ماشین بردار پشتیبان (SVM) برای تجزیه و تحلیل جنبه های مختلف زیست محیطی فرآیندهای تصفیه خانه فاضلاب استفاده شده است [9]. ANN یکی از قدیمی‌ترین روش‌های داده‌کاوی است، که می‌توان آن را برای پیش‌بینی بهتر عملکرد فرایند به دلیل دقت بالا، کفایت و کاربردهای کاملاً امیدوارکننده در مهندسی، علوم آب و زمینه‌های زیست‌محیطی مورداستفاده قرار داد[10]. در درخت مدل M5، معیارهای تقسیم داده ها و خروجی مدل قابل استخراج و مشاهده می باشد، در حالی که در شبکه عصبی مصنوعی فقط پیش بینی بیان می شود و نحوه عملکرد آن در سیستم پنهان می ماند[11]. به عبارت دیگر، این ساختار تصمیم‌گیری می‌تواند در قالب تکنیک‌های ریاضی و محاسباتی نیز معرفی شود که به توصیف، طبقه‌بندی و تعمیم مجموعه‌ای از داده‌ها کمک می‌کند[12]. یوان شی و همکاران در پژوهشی در سال ۲۰۲۲ با عنوان " افزایش پیش‌بینی به موقع کیفیت پساب تصفیه‌خانه فاضلاب بر اساس شبکه عصبی بهبودیافته همراه با الگوریتم بهینه‌سازی" مقدار شاخص همبستگی(R) ۰/۷۵ را جهت پیش بینی COD_{out} تصفیه خانه فاضلاب با استفاده از شبکه عصبی مصنوعی به دست آوردند[13]. سین و همکاران در سال ۲۰۲۲ در پژوهش خویش با عنوان"مدل تصفیه فاضلاب هوشمند مبتنی بر IOT با استفاده از هوش

1.Artificial Neural Network

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

مصنوعی" بیان می‌نمایند فاضلابی که توسط شرکت های داروسازی ایجاد می‌شود به یک نگرانی بزرگ برای اکوسیستم تبدیل شده است. مقدار قابل توجهی سموم وجود دارد که به‌طور مداوم از شرکت های داروسازی متعدد دفع می‌شود که به دلیل دارا بودن مواد آلی بالا و همچنین بارهای معدنی، آسیب های جدی به محیط‌زیست و سلامت عمومی وارد می‌کند و بنابراین نیاز به تصفیه مناسب قبل از دفع نهایی به اکوسیستم دارد[۱۴]. Asami و همکاران در مطالعه‌ای تحت عنوان «شبیه‌سازی نیاز اکسیژن بیوشیمیایی و شیمیایی و کل مواد جامد معلق در تصفیه‌خانه‌های فاضلاب: رویکرد داده‌کاوی» در سال ۲۰۲۱ اظهار داشتند که BOD_5، DO (اکسیژن محلول)، COD،TSS(کل مواد معلق)، دما و کدورت در ورودی مهم‌ترین پارامترهای تأثیرگذار بر خروجی تصفیه‌خانه می‌باشد[۱۵]. شرقی و همکاران در پژوهش خویش با عنوان "پایش کیفیت پساب تصفیه‌خانه فاضلاب با روش شبکه عصبی مصنوعی مبتنی بر خوشه" که در سال ۲۰۱۹ انتشار یافته عنوان می‌نمایند شبکه عصبی مصنوعی، به‌عنوان یک رویکرد مبتنی بر داده، ابزار قدرتمندی برای پیشبینی کیفیت پساب تصفیه فاضلاب است[۱۶].طبق یافته ها در اکثر مطالعات از ANN استفاده شده است و استفاده از مدل درخت M5 کمتر در نظر گرفته شده است. بنابراین در تحقیق حاضر از ANN که یکی از قدیمی ترین روش های داده کاوی است و همینطور درخت مدل M5 بعنوان روشی جدید برای شبیه سازی COD استفاده شد و در نهایت کارایی هر دو مدل ارزیابی شد.در این مطالعه با توسعه و ارزیابی مدل های یاد شده ، امکان پیش بینی کیفیت پساب یکی از تصفیه خانه های فاضلاب شهری در شهر تهران طی دوره ۵ ساله ۱۳۹۶ تا ۱۴۰۰ مورد بررسی قرار گرفت.

روش کار

۲-۱ محدوده مطالعه

داده‌های مورد بررسی مربوط به یک تصفیه‌خانه فاضلاب شهری در شهر تهران است که دبی فاضلاب ورودی به آن به طور متوسط ۶۵۰۳۴۳ متر مکعب در روز و دبی فاضلاب تصفیه شده به طور متوسط ۴۳۰۶۵۰ متر مکعب در روز می‌باشد.

غلظت COD طراحی این تصفیه‌خانه ۵۳۳ میلی گرم بر لیتر می‌باشد و روش تصفیه فاضلاب بهداشتی روش لجن فعال می‌باشد. پساب حاصل از تصفیه‌خانه مورد استفاده کشاورزی قرار می‌گیرد. در این مطالعه اطلاعات سیستم‌های پایش ورودی و خروجی تصفیه‌خانه فاضلاب دریافت و پاک‌سازی داده‌ها انجام گرفت. در ادامه تبدیل داده‌ها به‌منظور آماده‌سازی ورود آن‌ها به الگوریتم‌های داده‌کاوی از طرق پالایش، پردازش و ایجاد متغیر ساختگی انجام شد. در نهایت، الگوریتم شبکه عصبی مصنوعی (ANN) و مدل درخت تصمیم M5 به منظور یافتن بهترین مدل جهت پیش‌بینی غلظت COD در خروجی تصفیه‌خانه موردبررسی قرار گرفت؛ در این راستا ۷۰ درصد داده‌ها جهت یادگیری ماشین و ۳۰ درصد به‌منظور اعتبارسنجی در نرم‌افزار پایتون مورداستفاده قرار گرفت. درنهایت با مدل رگرسیونی و مقایسه شاخص‌های R^2،R^2 Adjusted و RMSE به انتخاب بهترین مدل پرداخته شد.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

جدول ۱. مشخصات آماری پارامترهای تصفیه خانه فاضلاب

پارامتر	۱۳۹۶			۱۳۹۷			۱۳۹۸			۱۳۹۹			۱۴۰۰		
	میانگین	حداقل	حداکثر	میانگین	حداقل	حداکثر	میانگین	حداقل	حداکثر	میانگین	حداقل	حداکثر	میانگین	حداقل	حداکثر
دبی فاضلاب ورودی به تصفیه خانه (متر مکعب بر روز)	۳۹۷۰۰۳	۱۱۱۱۱۱	۵۰۲۵۸۴	۵۶۶۱۶۳	۱	۱۵۶۳۷۸۵۶	۶۵۶۳۹۴	۴۰۷۳۵۰	۸۱۲۱۴۱	۶۸۲۶۰۰	۶۵۶	۷۹۳۳۷۹	۶۵۰۳۴۳	۵۳۱۰۲۴	۱۰۸۲۸۲۴
دبی فاضلاب تصفیه شده خروجی از تصفیه خانه (متر مکعب بر روز)	۳۵۹۹۸۰	۶۰۰۰	۴۵۲۴۶۴	۳۷۹۷۳۲	۱	۴۲۲۹۷۶	۳۶۹۱۳۲	۶۰۲۰۶	۴۰۷۳۵۰	۳۷۹۳۷۰	۳۹	۴۰۷۴۸۵	۴۳۰۶۵۰	۳۳۳۹۳۲	۲۴۲۹۳۰۱
BOD۵ (میلی گرم بر لیتر)	۲۴۳	۱۳۰	۲۹۰	۲۳۶	۱۱۰	۳۴۰	۲۴۵	۷۰	۴۵۰	۱۹۴	۷۰	۴۵۰	۲۰۷	۴۶	۴۰۲
BOD۵- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۱۷	۱	۱۹۰	۱۹	۴	۴۵	۱۶	۲	۶۰	۲۴	۲	۵۶	۲۲	۲	۱۲۲
COD (میلی گرم بر لیتر)	۵۰۷	۳۰۰	۸۸۷	۴۵۰	۴۲	۷۵۸	۴۵۹	۱۴۵	۱۳۰۰	۳۵۸	۱۱۰	۱۲۸۴	۳۸۱	۱۰۹	۸۵۸
COD- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۶۸	۱۳	۱۶۸	۷۲	۱۷	۱۵۵	۶۸	۶	۲۸۴	۷۷	۱۵	۱۹۴	۶۸	۱۵	۲۵۹
کل جامدات معلق (TSS) (میلی گرم بر لیتر)	۲۱۳	۳۴	۵۵۰	۲۲۹	۱۰۵	۴۲۸	۲۱۲	۲۶	۱۰۱۰	۱۶۸	۲۴	۴۳۸	۱۶۹	۷۴	۳۱۵
کل جامدات معلق (TSS)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۳۸	۲	۱۲۶	۴۱	۴	۱۱۸	۳۴	۲	۱۸۵	۵۰	۲	۱۳۷	۴۰	۲	۱۳۵
نیترات (NO_3)- بر حسب نیترات (میلی گرم بر لیتر)	۴	۱	۸	۳	۲	۷	۳	۰	۱۰	۸	۰	۱۷	۴	۲	۲۳
نیترات (NO_3)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۴۳	۲	۱۰۵	۱۷	۱	۲۸	۱۴	۰	۵۸	۱۸	۲	۵۲	۲۶	۱	۸۵
نیتروژن کل (TN)- بر حسب N (میلی گرم بر لیتر)	۶۸	۳۹	۲۷۰	۶۱	۴۴	۱۱۹	۷۳	۲۰	۸۶۳	۴۹	۳۳	۱۷۸	۵۵	۲۲	۱۳۶
نیتروژن کل (TN)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۲۱	۵	۷۳	۱۷	۵	۳۹	۱۵	۴	۵۳	۱۵	۶	۲۸	۱۷	۶	۳۴
آمونیاک (NH_3) - بر حسب آمونیاک (میلی گرم بر لیتر)	۴۹	۸	۹۶	۴۸	۵	۱۰۰	۵۰	۲	۱۰۰	۳۵	۱۴	۸۶	۴۷	۲۹	۸۷
آمونیاک (NH_3)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۱۷	۱۷	۱۷				۹	۷	۱۰				۷	۴	۱۲
آمونیوم (NH_4)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۱۱	۰	۲۹	۱۰	۰	۳۱	۱۰	۰	۳۹	۱۰	۲	۴۱	۹	۱	۳۶
کل جامدات محلول (TDS) (میلی گرم بر لیتر)	۵۸۰	۵۲۰	۶۸۰	۶۲۰	۶۰۰	۶۴۰							۵۷۵	۴۵۳	۷۴۷
کل جامدات محلول (TDS)- فاضلاب تصفیه شده (میلی گرم بر لیتر)	۵۸۰	۵۰۰	۶۶۰	۶۱۳	۶۰۰	۶۲۰							۵۷۳	۳۶۸	۵۹۳

۲-۲. جمع آوری داده‌ها

در این مطالعه پارامترهای C،BOD_{5in}،Cl_2،TN_i،NH_{4e}،TN_e،NO_{3e}،TSS_e،BOD_{5e} NH_3،TSS_{in}،OD_{in} (میانگین روزانه، هفتگی، ماهانه) تصفیه‌خانه فاضلاب از سال ۱۳۹۶ تا۱۴۰۰ که به‌صورت روزانه یا سایر بازه‌های زمانی اندازه گیری و ثبت شده است بعنوان داده ورودی به الگوریتم‌ها مورد استفاده قرار گرفت. در این مطالعه مجموعاً ۱۹۱۸ داده از اطلاعات تصفیه خانه استفاده شده است که ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای تست انتخاب گردیده است.

۲-۳. تکنیک‌های پیش پردازش در داده کاوی

تکنیک‌های پیش‌پردازش برای داده‌کاوی ضروری است که در این پژوهش به علت کمی بودن داده‌ها در مدل M5 نیاز به تبدیل داده‌ها نبوده و فقط پالایش و پردازش اولیه روی داده‌ها صورت گرفت، درحالی‌که در مدل شبکه عصبی مصنوعی، دو داده سال و ماه از نوع کیفی است، لذا برای ورود به مدل ابتدا ساختگی شده و سپس مورد پالایش و پردازش قرار گرفت.

۲-۳-۱. شبکه عصبی مصنوعی ANN

شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به‌دست‌آمده در جهت بیش بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. این سیستم از شمار زیادی عناصر پردازشی فوق‌العاده به‌هم‌پیوسته با نام نورون تشکیل شده که برای حل یک مسئله با هم هماهنگ عمل می‌کنند و توسط سیناپس‌ها (ارتباطات الکترومغناطیسی) اطلاعات را منتقل می‌کنند. در این شبکه‌ها اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آن را جبران کرده، و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری‌اند.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

۲-۳-۲. درخت مدل M5

M5 یا درخت تصمیم، روشی در یادگیری ماشین برای ساختاربندی (شکل‌دهی یا سازمان‌دهی به الگوریتم است. یک الگوریتم درخت تصمیم برای تقسیم ویژگی‌های مجموعه داده (Data Set) از طریق تابع هزینه (Cost Function) مورداستفاده قرار می‌گیرد. این الگوریتم قبل از انجام بهینه‌سازی و حذف شاخه‌های اضافه، به گونه‌ای رشد می‌کند که دارای ویژگی‌های نامرتبط با مسئله است؛ به همین دلیل، عملیات هرس کردن (Pruning)برای حذف این شاخه‌های اضافه در آن انجام می‌شود. در الگوریتم درخت تصمیم، پارامترهایی ازجمله عمق درخت تصمیم را نیز می‌توان تنظیم کرد تا از بیش‌برازش[1] یا پیچیدگی بیش از حد درخت (Overly Complex Tree) تا حد امکان جلوگیری شود. درخت تصمیم معمولاً برای برنامه‌ریزی و ترسیم تصمیم‌های عملیاتی در کسب‌وکارها به‌عنوان «نمودار» یا همان «فلوچارت بصری (Visual Flowchart) نیز استفاده می‌شود.

۲-۳-۳. معیارهای ارزیابی مدل‌ها

پس از انجام مدل‌سازی با دو روش لازم است عملکرد آنها با معیارهای مناسب مورد ارزیابی قرار گیرد، برای این منظور روش‌های متنوعی وجود دارد که عمده‌ترین این روش‌ها مقایسه مقادیر مشاهداتی و محاسباتی مدل‌ها با استفاده از معیارهای ارزیابی است. در این مطالعه از معیارهای پرکاربرد مهندسی شامل ضریب همبستگی[2] (R)، ضریب تبیین DC، ریشه میانگین مربعات خطا (RMSE) استفاده گردیده است که در فرمول شماره ۱ آورده شده است.

$$R=\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}(y_i-\bar{y})^2}}$$

$$DC=1-\frac{\sum_{i=1}^{N}(x_i-y_i)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(x_i-y_i)^2}{N}}$$

فرمول شماره ۱. فرمول محاسبه ضریب همبستگی، ضریب تبیین و ریشه میانگین مربعات خطا

در این روابط xi, yi و y , x به ترتیب داده‌های مربوط به نتایج مدل، داده‌های مشاهداتی، میانگین داده‌های مربوط به نتایج مدل و میانگین داده‌های مشاهداتی است. مدلی بهترین نتیجه را خواهد داشت که برای دو معیار R وDC مقداری نزدیک به یک و برای معیار RMSE مقداری نزدیک به صفر را به همراه داشته باشد.

1.Overfitting
2.Correlation Coefficient

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]
[DOI: 10.61186/jehe.11.4.380]

یافته‌ها

۳-۱. پیش پردازش داده‌ها

در این تحقیق در ابتدا داده‌های موجود وارد نرم‌افزار پایتون گردید و نسبت به ایجاد کتابخانه و فعال‌سازی در نرم‌افزار جهت بهینه‌کردن کدها اقدام گردید. در برنامه‌نویسی رایانه، کتابخانه‌های نرم‌افزاری متعددی از قبیل Pandas-BeautifulSoup-SciPy-Seaborn-TensorFlow-Scikit-Learn-Matplotlib، نوشته شده برای زبان برنامه‌نویسی پایتون برای دست‌کاری و تجزیه‌وتحلیل داده‌ها است. پس از جمع‌آوری داده‌ها و تشکیل دیتاست (شکل شماره ۱)، داده‌های پرت حذف و در نهایت ۱۹۳ رکورد حاوی نتایج پارامترهای اندازه گیری شده شامل BOD_{5e}،TSS_e،NO_{3e}، TN_e، NH_{4e}، TNi، Cl_2، BOD_{5in}، COD_{in}، TSS_{in}،NH_3، COD_{out} در ورودی و خروجی تصفیه خانه جهت مدلسازی مورد استفاده قرار گرفت.

جدول شماره ۲. نمای اجمالی از دیتاست

ردیف	BOD_5	BOD_e	COD_i	COD_e	TSS_i	TSS_e	NO_{3e}	TN_i	TN_e	NH_{3i}	NH_{4e}	CL_2
۰	۲۹۰	۱۹/۳۲	۴۹۹	۶۹/۳۲	۲۶۸	۳۹/۶۵	۱۳/۲۸	۶۹	۱۵/۲۰	۵۴/۴۰	۲/۷۰	۰/۱۰
۱	۲۴۰	۱۹/۳۲	۴۲۹	۶۹/۳۲	۱۸۰	۳۹/۶۵	۱۳/۷۳	۶۵	۱۴/۳۰	۴۹/۷۸	۲/۷۰	۰/۱۰
۲	۲۷۰	۳۹/۰۰	۴۹۷	۱۱۱/۰۰	۲۱۵	۵۶/۰۰	۱۵/۵۰	۶۷	۱۴/۲۰	۴۷/۴۸	۳/۶۰	۰/۲۲
۳	۲۴۰	۷/۰۰	۴۰۵	۹۲/۰۰	۲۸۰	۴۸/۰۰	۱۲/۴۰	۵۵	۱۱/۱۰	۴۴/۲۰	۲/۵۷	۰/۲۲
۴	۲۵۰	۲۲/۰۰	۴۸۴	۹۲/۰۰	۲۴۵	۵۶/۰۰	۱۲/۸۴	۷۲	۱۴/۹۰	۴۹/۷۸	۳/۹۸	۰/۲۲
------	------	------	------	------	------	------	------	----	------	------	------	-----
۱۸۷	۲۶۰	۱۵/۰۰	۴۸۷	۶۹/۰۰	۱۹۳	۴۱/۰۰	۱۷/۵۸	۵۷	۲۶/۲۰	۴۷/۶۰	۲۰/۶۰	۰/۲۰
۱۸۸	۲۵۰	۸/۰۰	۴۸۷	۵۱/۴۰	۱۷۳	۲۸/۰۰	۱۷/۵۸	۵۷	۲۲/۱۰	۴۸/۵۷	۱۳/۳۷	۰/۳۰
۱۸۹	۲۶۰	۳/۰۰	۵۵۲	۲۲/۰۰	۲۵۰	۱۰/۰۰	۱۶/۳۸	۶۲	۱۹/۷۰	۴۴/۲۰	۱۸/۲۶	۰/۲۰
۱۹۰	۲۸۰	۹/۰۰	۵۱۹	۴۱/۰۰	۲۹۲	۸/۰۰	۲۰/۸۱	۶۴	۱۸/۴۰	۴۳/۹۶	۱۳/۸۸	۰/۲۲
۱۹۱	۲۸۰	۱/۰۰	۵۶۷	۱۹/۰۰	۲۸۰	۲/۰۰	۳۱/۸۸	۶۹	۱۴/۷۰	۵۱/۰۰	۸/۷۴	۰/۲۲
۱۹۲ ردیف و ۱۲ ستون												

همچنان که در جدول شماره ۲ مشاهده می‌گردد در این تحقیق پارامترهایی درنظر گرفته شد که با کمک آنها خروجی مطلوب که همان COD خروجی است را پیش‌بینی نمودیم. بدین منظور پارامترهای موجود در سرستون‌ها به‌عنوان متغیرهای مستقل یا ورودی و پارامتر COD خروجی به‌عنوان متغیر هدف یا خروجی در نظر گرفته شد.

برای تعیین اینکه بدانیم کدام پارامترها یا متغیرهای مستقل بیشترین مشارکت را در پیش‌بینی COD خروجی داشته‌اند از تکنیک‌های انتخاب ویژگی استفاده گردید. در این تحقیق تکنیک انتخاب ویژگی (متغیرهای مستقل) مورداستفاده، روش اطلاعات متقابل یا MI[1] است.

1.Mutual Information

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right)$$

فرمول شماره ۲. تکنیک انتخاب ویژگی

جهت اجرای روش MI و کارکرد مناسب الگوریتم‌ها در داخل نرم‌افزار تمام متغیرهای مستقل را بین ۰ و ۱ مقیاس بندی نمودیم. با اجرای روش اطلاعات متقابل ویژگی‌هایی را که بیشترین همبستگی را با خروجی داشتند تعیین شد. شایان ذکر است هرچقدر اعداد به یک نزدیک‌تر باشند میزان همبستگی بهتر ارزیابی می‌گردد.

جدول شماره ۳. انتخاب ویژگی‌هایی (متغیرهای مستقل) با بیشترین همبستگی

پارامتر	ضریب همبستگی
TSS_e	۰/۷۰۶۳۸۶
$BOD5_e$	۰/۴۵۱۸۷۸
$NO3_e$	۰/۰۹۸۴۷۴
TN_e	۰/۰۹۶۶۸۳
$NH4_e$	۰/۰۹۰۲۸۶
TN_i	۰/۰۲۶۳۶۵
CL_2	۰/۰۰۳۶۵۱
BOD_{5i}	۰/۰۰۰۰۰۰
COD_i	۰/۰۰۰۰۰۰
TSS_i	۰/۰۰۰۰۰۰
NH_{3i}	۰/۰۰۰۰۰۰

با اجرای روش اطلاعات متقابل و مقیاس‌بندی مشخص گردید که دو متغیر مستقل BOD_e وTSS_e بیشترین همبستگی را با COD_e دارند که در ادامه مدل‌سازی صرفاً از این دو متغیر جهت پیش‌بینی استفاده گردید.

در این مطالعه x متغیرهای ورودی مدل جهت پیش‌بینی و yمتغیر پاسخ یا آن چیزی است که قرار است پیش‌بینی شود. در این تحقیق x هاBOD خروجی و کل مواد معلق خروجی از تصفیه‌خانه و متغیر y، COD_e در نظر گرفته شد. سپس مقادیر جدید یکبار دیگر بین صفر تا یک مقیاس‌بندی گردید تا برای ساخت مدل آماده‌سازی شوند؛ بنابراین در این تحقیق دومرتبه عملیات مقیاس‌بندی یکبار جهت انتخاب ویژگی و یکبار هم جهت ساخت مدل صورت گرفت. در ادامه دیتاست به دو بخش ۳۰درصد و ۷۰ درصد تقسیم گردید. ۳۰درصد یا ۵۸ سطر از داده‌هایی که در دیتاست موجود است جهت تست و اعتباربخشی مدل و ۷۰ درصد یا ۱۳۴ رکورد الباقی جهت آموزش الگوریتم استفاده گردید. در عملیات تست ابتدا xtest را به مدل دادیم تا مدل میزان y یا خروجی را پیش بینی کند و در ادامه xtest را با y ای که پیش بینی شد مقایسه نمودیم. در این

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

حالت هرچه قدر اختلاف بیشتر باشد به این معنی است که مدل از دقت کمتری برخوردار بوده است و بالعکس.

۲-۳. تحلیل استنباطی داده‌ها

در این تحقیق دو الگوریتم درخت M5 و شبکه عصبی مصنوعی به‌عنوان دو الگوریتم انتخابی به‌منظور یافتن بهترین مدل جهت پیش‌بینی غلظت COD_{out} در خروجی یک تصفیه‌خانه فاضلاب شهری مورد بررسی و مقایسه قرار گرفت.

۱-۲-۳. ANN

در این تحقیق ابتدا الگوریتم شبکه عصبی مصنوعی در نرم‌افزار پایتون که MLPregressor است را اجرا نمودیم. در اینجا با یک‌سری پارامترهایی در شبکه عصبی مصنوعی سرکار داشتیم که با تنظیم آنها می‌توان دقت مدل را افزایش داد که در این تحقیق تعداد لایه‌های پنهان الگوریتم پس از آزمون‌وخطا عدد ۱۰ لحاظ گردید که بهترین خروجی را به ما نشان داد. در ادامه توابع فعال ساز مختلف را تست نمودیم که بهترین تابع فعالساز که بهترین خروجی را حاصل نمود تابع فعال ساز تانژانت هایپربولیک می‌باشد. بهترین Solver انتخابی در این تحقیق نیز lbfgs است.

جدول شماره ۴. نتایج شبکه عصبی مصنوعی برای شبیه سازی پارامترها

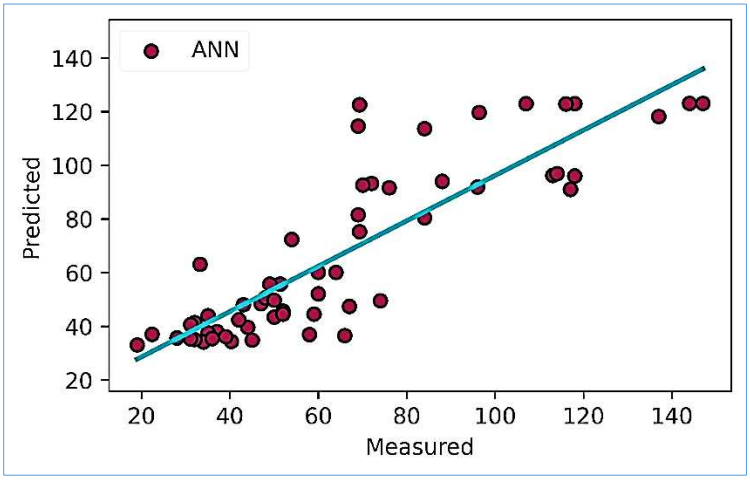
آموزش				تست			
R^2	RMSE	MAE	MSE	R^2	RMSE	MAE	MSE
۰/۷۸	۱۴/۸۲	۱۰/۳۲	۲۱۹/۸۵	۰/۷۲	۱۶/۷	۱۲/۵۱	۲۷۹/۱۸

همان‌طور که در جدول شماره ۴ مشاهده می‌گردد در این تحقیق R^2 در مرحله تست معادل۰/۷۲ به دست آمده است که عدد قابل قبولی می‌باشد. همچنین میزان RMSE ۱۶/۷می‌باشد. به این معنی که مقدار پیش بینی شده COD_e حدوداً با ۱۶ میلی گرم خطا گزارش شده است. همچنین میزان MAE(میانگین قدرمطلق خطا) و میانگین جذر مربعات خطا به ترتیب۱۲/۵۱ و۲۷۹/۱۸به دست آمده است. در واقع مقادیر به دست آمده کارکرد مدل برای زیرمجموعه تست می‌باشد. در ادامه سنجش کارایی مدل برای زیرمجموعه آموزش نیز محاسبه گردید که بتوان با استفاده از آنها متوجه شویم که الگوریتم به چه میزان توانسته است آموزش ببیند و یا به نوعی train را درک نماید.

همچنین در این جدول نتایج کارایی آموزش الگوریتم نشان‌داده‌شده است. در این تحقیق R^2 معادل ۰/۷۸به دست آمده است و میزان RMSE ۱۴/۸۲می‌باشد. همچنین میزان MAE و MSE (میانگین توان دوم خطاها) به ترتیب ۱۰/۳۲ و ۲۱۹/۸۵ به دست آمده است. در ادامه جهت مصور سازی روند پیش بینی شکل شماره۱ ترسیم گردید.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]



شکل شماره ۱. نتایج رگرسیون مدل شبکه عصبی (مرحله تست)

۲-۲-۳. الگوریتم درخت M5

در این نوع الگوریتم هر چقدر تعداد شاخه‌ها بالاتر رود احتمال بیش برازش بالاتر می‌رود و هرچه قدر تعداد شاخه‌ها کمتر گرفته شود احتمال کم برازش بیشتر می‌گردد. با تست در نرم‌افزار پایتون بیشتر از ۳ شاخه مدل دچار بیش برازش و کمتر از ۳ دچارکم برازش می‌گردید. در نتیجه ۳ شاخه به عنوان بهینه‌ترین عدد برای تعداد شاخه‌های این الگوریتم انتخاب گردید.

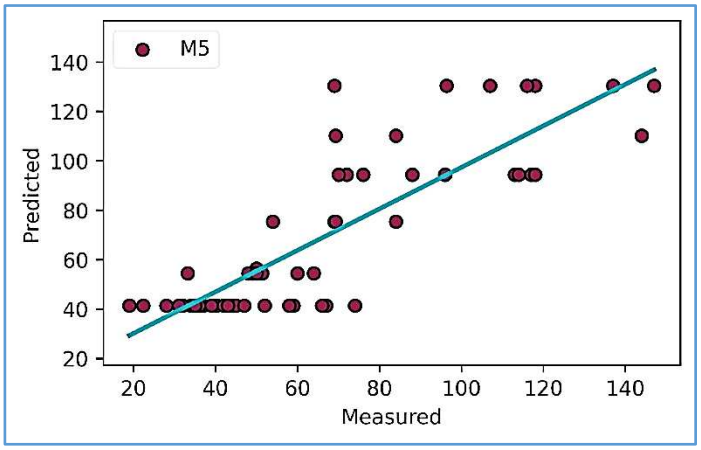
جدول شماره ۵. نتایج درخت M5 برای شبیه سازی پارامترها

آموزش				تست			
R^2	RMSE	MAE	MSE	R^2	RMSE	MAE	MSE
۰/۸۲	۱۳/۲۱	۹/۷۱	۱۷۴/۵۸	۰/۶۸	۱۸/۱۴	۱۴/۰۹	۳۲۹/۱۷

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

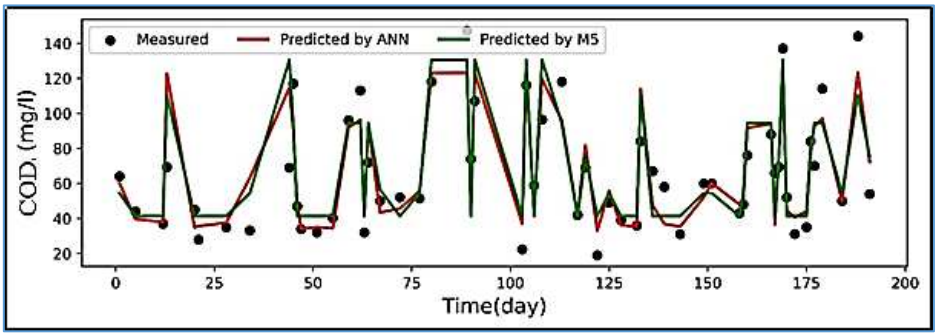
[DOI: 10.61186/jehe.11.4.380]

همان‌طور که در جدول شماره ۵ مشاهده می‌گردد مقدار R^2 ۰/۶۸ درصد برای زیرمجموعه تست و ۰/۸۲٪ جهت زیرمجموعه آموزش به دست آمده است.



شکل شماره ۲. نتایج رگرسیون مدل درخت تصمیم (مرحله تست)

شکل شماره ۳ به مقایسه مقادیری COD_{out} پیش بینی شده توسط دو مدل با مقادیر داده‌های واقعی آن می‌پردازد. با توجه به پراکندگی داده‌ها الگوریتم M5 نسبت به الگوریتم ANN، عملکرد ضعیف‌تری داشته است.



شکل شماره ۳. روند پیش بینی دو الگوریتم ANN و M5

بحث

پیش‌پردازش داده‌ها

بر اساس نتایج تحلیل رگرسیون انجام‌شده در طول مدل‌سازی COD_{out} مشخص شد که پارامترهای BODout، COD_{in}، TSS_{out}، YEAR، MONTH به‌صورت مجزا و COD_{in}*TSSout، BOD_{out}*YEAR، BOD_{out}*MONTH، COD_{in}*MONTH، TSS_{in}*MONTH، TSS_{out}*YEAR، TSS_{out}*MONTH، YEAR*MONTH به‌صورت ترکیبی بر COD_{out} مؤثر بوده‌اند. نتایج ترکیب ورودی‌های مختلف در پژوهش‌های مختلف نشان داد که تنها درنظرگرفتن متغیرهایی با بالاترین ضریب همبستگی برای مدل‌سازی دقیق کافی نیست؛ بنابراین، برای افزایش عملکرد مدل، لازم است متغیرهایی با ضریب همبستگی پایین نیز در نظر گرفته شوند. همچنین، پیش‌بینی با استفاده از الگوریتم‌های هوش مصنوعی مختلف و مجموعه‌داده‌های مختلف ممکن است به نتایج متفاوتی منجر شود.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

ANN

دراین پژوهش R=۰/۷۲ برای پیش‌بینی COD_{out} ارائه گردید. ضریب تعیین بالا و مقدار خطای پایین در مراحل آموزش، اعتبارسنجی و تست در مدل با لایه پنهان ۱۰۰ این امکان را به ما می‌دهد که از توابع بیشتری در لایه پنهان استفاده کرده تا به مقدار بهینه دست یابیم. ضریب تعیین بالا در گام آموزش مبین این است که چند درصد تغییرات متغیر وابسته در یک مدل رگرسیونی با متغیرهای مستقل تبیین می‌شود که این خود نشان دهنده انتخاب صحیح پارامترهاست. در شکل شماره ۱ رگرسیون در مراحل مختلف با ضریب تعیین بالا منطبق بودن داده‌ها با رگرسیون حاصل را نشان می‌دهد، که داده‌های انتخابی حتی در صورتیکه دارای اثر ناچیز باشند می‌توانند با ترکیب با پارامترهای دیگر اثر افزایشی خود را بر متغیر وابسته وارد سازند. مقایسه ضریب تعیین رگرسیون ساده با ضریب تعیین مدل شبکه عصبی مصنوعی به ما نشان می‌دهد که در دنیای پیشرفته امروز با کثرت زیادی داده‌ها و تنگی زمان استفاده از یادگیری ماشین در حوزه‌های مختلف امکان صرفه جویی در زمان و هزینه را به ما می‌دهد. در مطالعه‌ای که توسط آقای محمدتقی اعلمی و همکاران در سال ۱۴۰۰ در خصوص بررسی کارآیی روش‌های هوش مصنوعی در پیش بینی عملکرد تصفیه خانه فاضلاب شهر تبریز صورت گرفت میزانR= ۰/۷۸۸ برای پیش بینی COD خروجی به دست آمد که نتایج تقریباً متشابهی را نشان می‌داد (۱۷). آهنرت و همکاران که امکان پیش‌بینی COD بر اساس نرخ جریان را با تجزیه و تحلیل تأسیسات در آلمان و بلژیک نشان داد نیز نتایج متشابهی را نشان داد [۱۸].

درخت تصمیم M5

در شکل شماره ۲ دیاگرام درخت بهینه M5 با دادن تعداد گره‌های مختلف به نرم‌افزار و گرفتن ضریب تعیین، تعیین می‌شود که با توجه به تعداد گره، ضریب تعیین بهینه حاصل می‌شود که این خود یکی از برتری‌های درخت تصمیم می‌باشد. ضریب تعیین بهینه در این مدل ۰/۶۷به‌دست‌آمده است. که عدد به نسبت کمی است. یکی از معایب این مدل تکرر در بررسی ویژگی‌ها و اعتبارسنجی مدل می‌باشد که این خود باعث می‌شود ضریب تعیین عدد پایین‌تری به دست آید. این مدل نسبت به رگرسیون ساده ضریب تعیین پایین‌تری دارد. اما این به معنای برتری شبکه عصبی مصنوعی نسبت به درخت تصمیم نیست. در درخت تصمیم با بررسی RMSE میزان خطای مدل محاسبه شده بسیار کوچک است که این دقت مدل درخت تصمیم را به نظر می‌رساند. در مطالعه‌ای که توسط آقای مهدی قلی زاده و همکاران در سال ۲۰۲۴ در خصوص پیش‌بینی مبتنی بر یادگیری ماشینی کل مواد جامد معلق پساب در یک تصفیه خانه فاضلاب با استفاده از رویکردهای مختلف انتخاب ویژگی صورت گرفت میزان۰/۷۸ =R برای پیش بینی به دست آمد که نتایج تقریباً متشابهی را نشان می‌داد [۱۹]. در پژوهش صورت گرفته توسط Hamada و همکارانش در سال ۲۰۱۸ مشخص شد که ANN نسبت به M5 در مدل‌سازی و پیش‌بینی عملکرد تصفیه‌خانه فاضلاب نتایج بهتری ارائه می‌دهد [۱۲] . مقایسه نتایج نشان داد دقت مدل‌سازی در تحقیق حاضر بیشتر از پژوهش Hamada و همکارانش می‌باشد؛ که می‌توان به نوع معماری کاربردی و تعداد داده‌های آموزش و اعتبار سنجی نسبت داد.

مشکل پیش‌بینی کیفیت فاضلاب ورودی از تصفیه‌خانه نیز توسط دوگان و همکاران مطرح شد که تأثیر قابل‌توجهی از COD، TSS، TN و TP بر روی مقدار BOD_5 نشان دادند. مقدار داده‌های ورودی لازم برای تعیین BOD_5 در مقایسه با آنچه در تحقیق پیشنهاد شده است، هزینه‌های بالاتری را

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

برای تعیین شـــاخص‌های کیفیت فاضـــلاب ایجاد می‌کند و زمان انجام تعیین آنها را افزایش می‌دهد. [۲۰].

نتیجه گیری

در پژوهش حاضر، از مقایسه دو مدل فوق نتیجه می‌گیریم ANN نسبت به مدل M5 به دلیل ضریب تعیین بالاتر و میزان RMSE و MSE ، مدلی با عملکرد بهتر برای پیش‌بینی کیفیت پساب تصفیه‌خانه فاضلاب است. همچنین نتایج تحقیق نشان‌دهنده دقت مناسب‌تر ANN نسبت به M5 به علت دارابودن ضریب تعیین بالاتر است. باتوجه‌به نتایج مشخص شد که روش‌های ذکر شده بر اساس معیارهای آماری توانایی قابل‌قبولی برای شبیه‌سازی COD خروجی دارند. برای استفاده از روش‌های داده‌کاوی مانند ANN و درخت تصمیم، تعداد داده‌های مورداستفاده و پردازش اولیه حائز اهمیت است. ازآنجائی‌که مدل‌های M5 و ANN نتایج قابل‌قبولی را در تفسیر روابط بین متغیرها و پیش‌بینی کیفیت پساب تصفیه‌خانه ارائه داده‌اند، می‌توانند برای ارائه تخمینی از مقادیری که برای اندازه‌گیری واحدهای تصفیه در نظر گرفته می‌شود در غیاب اقدامات مستقیم توصیه شوند. این روش‌ها می‌توانند برای تصمیم‌گیری‌های زیست‌محیطی و تخمین داده‌های گمشده در تصفیه‌خانه‌های فاضلاب برای افزایش کارایی سیستم در آینده مفید باشند. در کنار ساختار مدل، دقت داده‌ها، طول و ساختار ممکن است تأثیر مؤثری بر عملکرد مدل داشته باشد.

سپاسگزاری

بدین‌وسیله از گروه محترم MPH دانشکده بهداشت و ایمنی دانشگاه علوم پزشکی شهید بهشتی به‌ویژه استاد گران‌قدر جناب آقای دکتر حاتمی مدیرگروه محترم MPH که ما را در انجام این تحقیق یاری نمودند و سرکار خانم مهندس حمیده سمیعی کارشناس شرکت مهندسی آب و فاضلاب کشور که در گردآوری اطلاعات و داده‌ها و انجام این تحقیق کمک‌حال بنده بودند کمال تشکر و قدردانی را داریم.

مقاله فوق با کد مصوبه اخلاق IR.SBMU.PHNS.REC.1401.042 در دانشکده بهداشت و ایمنی دانشگاه علوم پزشکی شهید بهشتی مصوب گردیده است و به اجرا در آمده است.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]

References

1. Torabian A, Motalebi M. Management plan for reuses of treated wastewater (case study: ekbatan treatment plant). 2004.
2. Yel E, Yalpir S. Prediction of primary treatment effluent parameters by Fuzzy Inference System (FIS) approach. procedia computer science. 2011;3:659-65.
3. Fresner J. Cleaner production as a means for effective environmental management. Journal of cleaner production. 1998;6(3-4):171-9.
4. Türkmenler H, Pala M. Performance assessment of advanced biological wastewater treatment plants using artificial neural networks. International Journal of Engineering Technologies IJET. 2017;3(3):151-6.
5. Shirvani H, Ganjidoust H, Hemmati M, Zarasvand Asadi R. Investigation of oil refinery wastewater treatment using a submerged membrane bioreactor. Journal of Petroleum Research. 2013;22(70):43-55.
6. Mjalli FS, Al-Asheh S, Alfadala H. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. Journal of environmental management. 2007;83(3):32938.
7. Ward J, Ward P, Saint C, Mantzioris E. The urban agriculture revolution. Water: Journal of the Australian Water Association. 2014;41(1):69-74.
8. Mehrdadi N, Ghasemi M. Modeling of Tehran South Water Treatment Plant Using Neural Network and Fuzzy Logic Considering Effluent and Sludge Parameters. Numerical Methods in Civil Engineering. 2021;6(1):63-76.
9. Ramírez Y, Kraslawski A, Cisternas LA. Decision-support framework for the environmental assessment of water treatment systems. Journal of Cleaner Production. 2019;225:599-609.
10. Khosravi K, Mao L, Kisi O, Yaseen ZM, Shahid S. Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. Journal of Hydrology. 2018;567:165-79.
11. Najafzadeh M, Zeinolabedini M. Prognostication of waste water treatment plant performance using efficient soft computing models: an environmental evaluation. Measurement. 2019;138:690-701.
12. Hamada M, Adel Zaqoot H, Abu Jreiban A. Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip. Journal of Applied Research in Water and Wastewater. 2018;5(1):399-406.
13. Wong J. Pollution prevention/waste minimization in california petroleum refineries. OCEESA J. 2002;19(1):306.
14. Singh DN, Murugamani C, Kshirsagar PR, Tirth V, Islam S, Qaiyum S, et al. IOT based smart wastewater treatment model for industry 4.0 using artificial intelligence. Scientific Programming. 2022;2022:1-15.
15. Asami H, Golabi M, Albaji M. Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: data-mining approach. Journal of Cleaner Production. 2021;296:126533.
16. Sharghi E, Nourani V, AliAshrafi A, Gökçekuş H. Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method. Desalination and Water Treatment. 2019;164:86-97.
17. Aalami MT, Hejabi N, Nourani V, SAGHEBIAN S. Investigation of artificial intelligence approaches capability in predicting the wastewater treatment plant performance (Case study: Tabriz wastewater treatment plant). Amirkabir Journal of Civil Engineering. 2021;53(3):1033-48.
18. Ahnert M, Marx C, Krebs P, Kuehn V. A black-box model for generation of site-specific WWTP influent quality data based on plant routine data. Water Science and Technology. 2016;74(12):2978-86.
19. Gholizadeh M, Saeedi R, Bagheri A, Paeezi M. Machine learning-based prediction of effluent total suspended solids in a wastewater treatment plant using different feature selection approaches: A comparative study. Environmental Research. 2024;246:118146.
20. BAKİ1a OT, Egemen A. Estimation of BOD in wastewater treatment plant by using different ANN algorithms. 2018.

[Downloaded from jehe.abzums.ac.ir on 2026-11-18]

[DOI: 10.61186/jehe.11.4.380]