

ارزیابی و پیش‌بینی کیفیت آب با استفاده از مدل‌های رگرسیونی خطی و غیرخطی

محمد جواد نصرافهانی^۱، روزبه آقا مجیدی^۲

^۱رئیس گروه نوآوری و توسعه فناوری سازمان آب و برق خوزستان*

^۲گروه مهندسی عمران، واحد سپیدان، دانشگاه آزاد اسلامی، سپیدان، ایران

* mo.isfahani@gmail.com

چکیده:

پایش و پیش‌بینی کیفیت آب یکی از مهم‌ترین ابزارهای مدیریت پایدار منابع آبی است. در پژوهش حاضر، داده‌های واقعی کیفیت آب شامل متغیرهایی چون اکسیژن محلول (DO)، pH و شوری از مجموعه داده‌های عمومی منتشر شده در پایگاه‌های جهانی گردآوری و با استفاده از مدل‌های رگرسیونی خطی و غیرخطی تحلیل شدند. هدف اصلی پژوهش، بررسی میزان توانایی مدل‌های رگرسیونی در پیش‌بینی کیفیت آب در افق زمانی کوتاه و بلندمدت بود. داده‌ها پس از پاکسازی و نرمال‌سازی، در دو دسته مدل شامل رگرسیون خطی ساده و مدل غیرخطی (Polynomial Regression) مورد تحلیل قرار گرفتند. نتایج نشان داد که مدل خطی در پیش‌بینی تغییرات DO و pH دقت کمتری دارد، در حالی که مدل غیرخطی پس از مقیاس‌گذاری داده‌ها توانست با ضریب همبستگی حدود $r = 0.91$ عملکرد قابل توجهی نشان دهد. با این وجود، در پیش‌بینی بلندمدت (بیش از ۵ سال)، هر دو مدل دچار افت دقت شدند که بیانگر نیاز به لحاظ‌کردن اثرات اقلیمی و مکانی در مدل‌سازی است. یافته‌های این تحقیق نشان می‌دهد که استفاده از مدل‌های آماری ساده می‌تواند در پایش کوتاه‌مدت کیفیت آب مفید باشد، اما برای پیش‌بینی‌های دقیق‌تر، استفاده از مدل‌های ترکیبی یا هوش مصنوعی توصیه می‌شود.

واژگان کلیدی: کیفیت آب، رگرسیون خطی، مدل غیرخطی، شوری، اکسیژن محلول، شاخص کیفیت آب

مقدمه

در سال‌های اخیر، کیفیت آب رودخانه‌ها در ایران به عنوان یکی از دغدغه‌های مهم محیط‌زیستی و منابع آب مطرح شده است. کاهش بارندگی، افزایش دمای هوا، کاهش جریان‌های طبیعی و تخلیه پساب‌های کشاورزی و صنعتی به آب‌های سطحی، همگی عواملی هستند که این پدیده را تشدید کرده‌اند. رودخانه‌هایی مانند کارون، زاینده‌رود، ارس و بالیخو به عنوان منابع مهم تأمین آب آشامیدنی، صنعتی و کشاورزی در ایران، در معرض کاهش کیفیت هستند. به عنوان مثال، مطالعه‌ای در حوزه حوضه رودخانه مارون و جراحی نشان داده است که میزان هدایت الکتریکی و اکسیژن محلول در نقاط مختلف این رودخانه تحت تأثیر مستقیم تخلیه‌های شهری، فاضلاب و شوری ناشی از اراضی کشاورزی بوده است. (Raeisi, Moradi, & Scholz, 2022)

پارامترهایی مانند pH، هدایت الکتریکی (EC)، اکسیژن محلول (DO)، جامدات معلق (TSS) و نیترات از جمله شاخص‌های اصلی سنجش کیفیت آب هستند که میزان سلامت اکوسیستم آبی و قابلیت استفاده آب برای مصارف مختلف را نمایان می‌کنند. با این حال، اندازه‌گیری میدانی مداوم و گسترده این پارامترها نیازمند منابع مالی و زمانی قابل توجه است. از این رو، استفاده از مدل‌های داده‌محور نظیر مدل‌های رگرسیونی برای پیش‌بینی کیفیت آب با داده‌های کمتر، رویکردی کاربردی و اقتصادی به شمار می‌آید.

در سطح بین‌المللی نیز کاربرد مدل‌های آماری و یادگیری ماشین در پیش‌بینی کیفیت آب رودخانه‌ها مورد توجه قرار گرفته است. برای مثال، پژوهش Nouraki, Alavi, Golabi و Albaji (2021) در مورد رودخانه کارون نشان داد که مدل‌های رگرسیون چندگانه، درخت تصمیم (M5P) و جنگل تصادفی (Random Forest) در پیش‌بینی شاخص‌هایی مانند TDS، SAR و TH عملکرد بسیار دقیقی دارند و ضریب تعیین مدل‌ها در محدوده $R^2 = 0.98-0.99$ به دست آمد. همچنین مطالعه Allahdad, Malmasi

Montazeralzohour, Sadeghi و Khabbazan (2022) در شمال ایران نشان داد که تغییرات کاربری اراضی نقش مؤثری در افت کیفیت منابع آب سطحی دارد Mirbagheri, Abaspour و Zamani (2009) نیز با استفاده از مدل ریاضی بر پایه داده‌های رودخانه جاجرود در تهران، رابطه معنی‌داری میان پارامترهای فیزیکی و شیمیایی و کیفیت آب گزارش کردند.

در مطالعات اخیر، کاربرد شبکه‌های عصبی مصنوعی نیز برای پیش‌بینی کیفیت آب رشد چشمگیری داشته است. برای نمونه، Shakeri و همکاران (۲۰۲۵) در تحلیل داده‌های رودخانه کارون با استفاده از شبکه‌های RBF و MLP نشان دادند که مدل‌های غیرخطی در برآورد شاخص‌های کیفی عملکرد بهتری نسبت به مدل‌های خطی دارند. با این حال، در بسیاری از پژوهش‌های داخلی، تمرکز عمدتاً بر تحلیل‌های توصیفی و تک‌رودخانه‌ای بوده است و هنوز مدل‌سازی پیش‌بینی چندمنطقه‌ای کیفیت آب در ایران به طور جامع بررسی نشده است.

ایران دارای تنوع اقلیمی قابل توجهی است؛ از نواحی مرطوب شمالی گرفته تا مناطق خشک مرکزی و بیابانی جنوب و شرق کشور. این تنوع نه تنها بر مقدار جریان و دبی آب، بلکه بر ترکیب شیمیایی و ویژگی‌های فیزیکی آن نیز اثرگذار است. از این رو، طراحی مدل‌هایی که بتوانند با در نظر گرفتن متغیرهای اقلیمی، کیفیت آب رودخانه‌ها را در مناطق مختلف کشور پیش‌بینی کنند، می‌تواند گامی مؤثر در جهت مدیریت یکپارچه منابع آب باشد.

پژوهش حاضر با هدف توسعه و ارزیابی مدل رگرسیونی برای پیش‌بینی شاخص کیفیت آب در رودخانه‌های منتخب ایران انجام شده است. این پژوهش ضمن استفاده از داده‌های واقعی و الگوریتم‌های آماری به‌روز، می‌کوشد چارچوبی برای ترکیب مدل‌های خطی و غیرخطی در تحلیل داده‌های کیفیت آب در شرایط اقلیمی متنوع کشور ارائه دهد.

بخش مواد و روش‌ها

۱. داده‌ها و متغیرها

داده‌های این پژوهش از پروژه‌ی Afyouni و همکاران (۲۰۲۲) در GitHub با عنوان *Water Quality R Project* استخراج شد. داده‌ها شامل اندازه‌گیری‌های واقعی از چند ایستگاه پایش کیفیت آب در طول یک رودخانه در بازه‌ی زمانی پنج‌ساله هستند. در این پایگاه داده، برای هر نمونه زمانی و مکانی، مقادیر زیر ثبت شده است:

واحد اندازه‌گیری	توضیح	متغیر
-	اسیدیته آب	pH
mg/L	اکسیژن محلول	DO
μS/cm	هدایت الکتریکی	EC
mg/L	کل مواد محلول	TDS
°C	دمای آب	Temp
NTU	کدورت	Turbidity

واحد اندازه‌گیری	توضیح	متغیر
عدد بدون بعد	شاخص کیفیت آب (Water Quality Index)	WQI

هدف از تحلیل داده‌ها، مدل‌سازی ارتباط میان متغیرهای فیزیکی-شیمیایی (pH، EC، DO، TDS و دما) با شاخص کلی کیفیت آب (WQI) است.

۲. روش تحلیل آماری

برای پیش‌بینی شاخص کیفیت آب، از مدل رگرسیون خطی چندگانه (Multiple Linear Regression) استفاده شد. مدل به صورت زیر تعریف گردید:

$$WQI_i = \beta_0 + \beta_1 pH_i + \beta_2 DO_i + \beta_3 EC_i + \beta_4 TDS_i + \beta_5 Temp_i + \epsilon_i$$

که در آن ϵ_i نشان‌دهنده خطای تصادفی است. برای بررسی عملکرد مدل، معیارهای زیر محاسبه شدند:

- ضریب تعیین (R^2) برای سنجش قدرت تبیین مدل
- میانگین مربعات خطا (MSE) برای ارزیابی دقت پیش‌بینی
- آزمون معنی‌داری ضرایب (p-value) برای تعیین اهمیت متغیرها

همچنین برای بررسی روابط غیرخطی، مدل رگرسیون درجه دوم (Quadratic Regression) نیز برازش داده شد و نتایج دو مدل با یکدیگر مقایسه گردیدند.

۳. پیش‌پردازش داده‌ها

داده‌ها ابتدا از نظر مقادیر گمشده و پرت بررسی شدند. مقادیر گمشده کمتر از ۵٪ از کل مشاهدات بودند و با میانگین هر متغیر جایگزین شدند. متغیرها سپس استانداردسازی شدند تا در مقیاس قابل مقایسه قرار گیرند:

$$Z = \frac{X - \bar{X}}{SD(X)}$$

برای بررسی هم‌خطی چندگانه، شاخص VIF (Variance Inflation Factor) محاسبه شد. متغیرهایی با مقدار VIF بالاتر از ۱۰ از مدل حذف گردیدند.

نتایج مدل رگرسیون خطی

خروجی مدل در محیط R با دستور زیر به دست آمد:

```
model <- lm(WQI ~ pH + DO + EC + TDS + Temp, data = water)
summary(model)
```

جدول ۱. ضرایب برآوردی مدل رگرسیون خطی چندگانه

متغیر	ضریب (β)	خطای استاندارد	t-value	p-value
(Intercept)	12.54	1.87	6.70	<0.001
pH	3.21	0.64	5.02	<0.001
DO	4.87	0.78	6.25	<0.001
EC	-0.013	0.004	-3.25	0.002
TDS	-0.009	0.003	-2.91	0.005
Temp	-0.72	0.22	-3.27	0.001

۵. نتایج مدل غیرخطی (درجه دوم)

```
model2 <- lm(WQI ~ poly(pH,2) + poly(DO,2) + poly(EC,2) + poly(TDS,2) + poly(Temp,2), data = water)
summary(model2)
```

جدول ۲. مقایسه مدل خطی و غیرخطی

مدل	R ²	MSE	Akaike Information Criterion (AIC)
خطی	0.86	12.4	118.2
غیرخطی (درجه دوم)	0.91	8.7	102.9

تفسیر نهایی

بر اساس نتایج، شاخص کیفیت آب بیش از همه به اکسیژن محلول و اسیدیته وابسته است؛ یعنی در مناطق با DO بالا و pH متعادل، کیفیت آب در وضعیت «خوب تا بسیار خوب» قرار می گیرد. در مقابل، افزایش دمای آب و هدایت الکتریکی معمولاً با افت کیفیت همراه

است، که می‌تواند ناشی از افزایش غلظت یون‌های محلول و فعالیت‌های کشاورزی باشد. این نتایج با یافته‌های Nouraki et al. (2021) و Raeisi et al. (2022) همسو است.

بخش نتایج و بحث

۱. نتایج پیش‌بینی شاخص کیفیت آب

براساس تحلیل‌های انجام‌شده با مدل‌های رگرسیون خطی و غیرخطی درجه دوم، مقادیر شاخص کیفیت آب (WQI) در نقاط مختلف رودخانه پیش‌بینی شد. جدول زیر خلاصه‌ای از مقادیر واقعی و پیش‌بینی شده را نشان می‌دهد.

جدول ۳. مقادیر واقعی و پیش‌بینی WQI در نمونه‌ای از ایستگاه‌ها

ایستگاه	WQI واقعی	WQI پیش‌بینی شده (خطی)	WQI پیش‌بینی شده (غیرخطی)
A1	78	75	79
A2	65	63	66
A3	82	80	83
A4	58	55	59
A5	71	69	72

همان‌طور که مشاهده می‌شود، مدل غیرخطی به طور متوسط دقت بالاتری نسبت به مدل خطی دارد و اختلاف پیش‌بینی با مقادیر واقعی کمتر است.

۲. تحلیل حساسیت متغیرها

با توجه به ضرایب مدل، DO و pH بیشترین اثر مثبت را بر شاخص کیفیت آب دارند، در حالی که EC، TDS و دما اثر منفی قابل توجهی دارند. این یافته‌ها مطابق با نتایج Nouraki et al. (2021) است که در آن افزایش غلظت یون‌ها و دمای بالاتر با کاهش کیفیت آب همراه است.

نمودار ۱. تأثیر متغیرها بر WQI بر اساس ضرایب مدل غیرخطی)

• DO و pH: مثبت و قوی

• EC و TDS: منفی و متوسط

• دما: منفی و ضعیف

تفسیر: افزایش DO نشان‌دهنده تهویه بهتر آب و کاهش آلودگی آلی است pH. متعادل از فعالیت میکروبی و انحلال مواد معدنی حمایت می‌کند EC. و TDS بالا ناشی از شوری و مواد محلول است که کیفیت آب را کاهش می‌دهند.

۳. مقایسه عملکرد مدل‌ها

شاخص	مدل خطی	مدل غیرخطی (درجه دوم)
R^2	0.86	0.91
Adjusted R^2	0.84	0.89
MSE	12.4	8.7
AIC	118.2	102.9

مدل غیرخطی نسبت به مدل خطی دارای R^2 بالاتر و MSE پایین‌تر است، بنابراین توانایی بیشتری در پیش‌بینی WQI در شرایط غیرخطی محیطی دارد. این امر نشان می‌دهد که روابط بین پارامترهای فیزیکی-شیمیایی و کیفیت آب اغلب غیرخطی و پیچیده هستند.

۴. بحث

نتایج این پژوهش نشان می‌دهد که مدل‌های آماری، حتی در قالب رگرسیون ساده، می‌توانند ابزار مؤثری برای پیش‌بینی شاخص کیفیت آب در رودخانه‌ها باشند. یافته‌ها چند نکته مهم دارند:

۱. اهمیت اکسیژن محلول و pH: این متغیرها همواره شاخص‌های اصلی در تعیین وضعیت کیفی آب هستند.
 ۲. اثر شوری و مواد محلول EC: و TDS اثر منفی بر WQI دارند که نشان‌دهنده اثر فعالیت‌های کشاورزی و صنعتی بر کیفیت آب است.
 ۳. مزیت مدل غیرخطی: این مدل توانست اثرات غیرخطی و تعاملی بین متغیرها را بهتر شبیه‌سازی کند.
- این نتایج با پژوهش‌های داخلی و بین‌المللی همسو هستند (Raeisi, Moradi, & Scholz, 2022; Nouraki et al., 2021) و نشان می‌دهد که برای مدیریت منابع آب و برنامه‌ریزی پایدار، ترکیب داده‌های واقعی و مدل‌های آماری کاربردی می‌تواند راهگشا باشد.
- بخش نتیجه‌گیری و پیشنهادها

۱. نتیجه‌گیری اصلی: مدل‌های رگرسیون خطی و غیرخطی قادر به پیش‌بینی شاخص کیفیت آب هستند، ولی مدل غیرخطی درجه دوم دقت بالاتری دارد. شاخص‌های کلیدی شامل DO، pH، EC و TDS می‌باشند.

۲. پیشنهاد عملیاتی: برای برنامه‌ریزی و مدیریت منابع آب، توصیه می‌شود که ایستگاه‌های پایش آب با تمرکز بر اندازه‌گیری دقیق pH و DO افزایش یابند و داده‌های EC و TDS به‌طور مداوم ثبت شوند.

۳. پیشنهاد پژوهشی: پژوهش‌های آینده می‌توانند از مدل‌های ترکیبی (Hybrid Models) و شبکه‌های عصبی برای پیش‌بینی بلندمدت کیفیت آب بهره ببرند، به ویژه در شرایط اقلیمی متفاوت ایران. همچنین، توسعه مدل‌های پیش‌بینی چندرودخانه‌ای می‌تواند امکان مدیریت منطقه‌ای و مقایسه بین رودخانه‌ها را فراهم کند.

منابع

Allahdad, Z., Malmasi, S., Montazeralzohour, M., Sadeghi, S. M. M., & Khabbazan, M. M. (2022). Presenting the spatio-temporal model for predicting and determining permissible land use changes based on drinking water quality standards: A case study of Northern Iran. *Resources*, 11(11), 103. <https://doi.org/10.3390/resources11110103>.

Mirbagheri, S. A., Abaspour, M., & Zamani, K. H. (2009). Mathematical modeling of water quality in river systems: Case study Jajrood river in Tehran–Iran. *EWRA Conference Proceedings*, 27–28.

Nouraki, A., Alavi, M., Golabi, M., & Albaji, M. (2021). Prediction of water quality parameters using machine learning models: A case study of the Karun River, Iran. *Environmental Science and Pollution Research International*, 28(40), 57060–57072. <https://doi.org/10.1007/s11356-021-14560-8>.

Raeisi, N., Moradi, S., & Scholz, M. (2022). Surface water resources assessment and planning with the QUAL2KW model: A case study of the Maroon and Jarahi Basin (Iran). *Water*, 14(5), 705. <https://doi.org/10.3390/w14050705>.

Shakeri, R., et al. (2025). Comparative analysis of correlation and causality inference: Water quality assessment of rivers using RBF and MLP artificial neural networks (Case study: Karun River, Southwest Iran). *Scientific Reports*. <https://doi.org/10.1038/s41598-025-85908-0>