ISSN: 2588-2805
http://www.ieaj.ir

مجله علمی پژوهش در شیمی و مهندسی شیمی (سال دهم)
شماره ۲۹ / پاییز و زمستان ۱۴۰۳ / ص ۴۳- ۳۲

کاربرد مدل‌های یادگیری ماشین جهت بررسی و امکان سنجی یا بهینه سازی حذف آلاینده های زیست محیطی با استفاده از جاذب ها

مجتبی عزیزی[۱]، امید عابدی[۲]، سعید علیخانی[۳]

[۱] دانشکده شیمی و مهندسی شیمی دانشگاه صنعتی مالک اشتر تهران، ایران.

[۲] دانشکده مهندسی شیمی دانشگاه صنعتی امیرکبیر.

[۳] گروه مهندسی کامپیوتر، دانشکده هوش مصنوعی دانشگاه آزاد اسلامی واحد تهران شمال تهران، ایران.

نام و نشانی ایمیل نویسنده مسئول:

مجتبی عزیزی

azizi.m_58@yahoo.com

چکیده

امروزه با توجه به افزایش انواع آلاینده‌های صنعتی، یکی از اصلی‌ترین چالش‌های پیش رو علوم زیست محیطی، مدیریت پساب‌های صنعتی است. در این پژوهش ما تکنیک جذب به عنوان یک روش موثر برای تصفیه فاضلاب و کاهش فلزات سنگین، را مورد مطالعه قرار داده‌ایم. مدل‌های هوش مصنوعی و الگوریتم‌های یادگیری ماشین در جهت بهبود فرآیند حذف آلاینده‌ها، به کار رفته‌اند. در این مدل‌ها امکان پیش‌بینی توانایی حذف فلزات سنگینی نظیر آهن، سرب، کادمیوم، مس، نیکل و کروم از طریق آنالیز پارامترهای کلیدی مانند نوع آلاینده، نوع جاذب، دما، غلظت اولیه، میزان جاذب، زمان و pH، فراهم شده است. الگوریتم‌های مختلف یادگیری ماشین از جمله رگرسور تقویت کننده گرادیان[۱]، رگرسور تقویت کننده طبقه ای[۲]، رگرسور تقویت کننده شدید[۳]، رگرسور جنگل تصادفی[۴] و شبکه عصبی، در این مطالعه مورد استفاده قرار گرفته‌اند. بر طبق داده‌های حاصل از این پژوهش، مدل رگرسیون تقویت‌کننده شدید با ضریب همبستگی (R^2) برابر با ۰.۹۳، به عنوان مدل بهینه پیش‌بینی بازده حذف در نظر گرفته شد. در این مطالعه به اهمیت استفاده از تکنیک‌های نوین یادگیری ماشین در مدیریت پساب‌های صنعتی در جهت کمک به توسعه روش‌های کارآمدتر، بسیار پرداخته شده است.

واژگان کلیدی: تصفیه فاضلاب، فلزات سنگین، رویکردهای داده محور، مدل‌های یادگیری ماشین.

[1] Gradient Boosting Regressor
[2] Cat Boost Regressor
[3] XGBRegressor
[4] Random Forest Regressor

ISSN: 2588-2805
http://www.ieaj.ir

مقدمه

یکی از چالش‌های اصلی جهانی، آلودگی آب خصوصاً از طریق فلزات سنگین موجود در فاضلاب‌های صنعتی است. این فلزات سبب القا خطرات اکوسیستمی شده و ریسک به خطر افتادن سلامت انسان و کیفیت کلی منابع آبی را دارند. یکی از مضرترین فلزات این دسته، انواع سنگین مانند سرب، کادمیوم، کروم و جیوه هستند که سمیت و ماندگاری بالایی در محیط دارند که می‌توانند در موجودات آبزی تجمع پیدا کرده و وارد زنجیره غذایی شوند. در نتیجه حذف این آلاینده‌ها از اهمیت به سزایی در حفاظت اکوسیستم و سلامت انسان برخوردار است و ضرورت توسعه تکنیک‌های کارآمد تصفیه را توجیه می‌کند. نظر به حوزه تحقیقاتی ما در رابطه با رعایت اصول شیمی سبز و تصفیه پساب های صنعتی جهت حذف آلاینده های فلزی سمی و سنگین و جلوگیری از ورود انها به اکوسیستم و فاضلابهای شهری[۱، ۲، ۳]، در این پروژه نیز پیش بینی حذف آلاینده ها با استفاده از جاذب ها بوسیله تکنیک یادگیری ماشین مورد بررسی قرار گرفت. یکی از موثرترین تکنیک‌های تصفیه، جذب سطحی است که برای استخراج فلزات سنگین از محلول های آبی، کاربردی است. در این تکنیک از چسبندگی یون‌های فلزی بر روی سطح جاذب‌های جامد استفاده می‌شود که سبب کاهش غلظت آن‌ها در محیط آبی می‌شود. در مطالعات مختلفی، شرایط عملیاتی مختلف و انواع جاذب‌ها به منظور بهینه‌سازی استراتژی کارآمد حذف آلاینده فلزی از فاضلاب، مورد بررسی قرار گرفته است. به منظور استفاده از أنواع مختلف مواد مانند کربن فعال معمولی یا جایگزین های نوآورانه کم هزینه، می‌توان از تطبیق‌پذیری جذب استفاده کرد. به منظور حذف فلزاتی مانند آهن، سرب و نیکل در صنعت، از جاذب‌های مقرون به صرفه مانند پوسته برنج، خاکستر بادی و محصولات جانبی مختلف کشاورزی، استفاده می‌کنند. با استفاده از این مواد هم رویکردی پایدار جهت تصفیه آب ارائه شده و هم از طریق کاهش بقایای کشاورزی، ضایعات نیز کاهش می‌یابند

مطابق مطالعات، راندمان جذب کاملاً وابسته به عملیاتی خاص، مانند دما، pH، زمان تماس و ویژگی‌های آلاینده‌های هدف، می‌باشد. به عنوان مثال، مطابق مطالعات انجام شده، می‌توان از نانوذرات آهن جهت جذب کروم از آب آلوده کمک گرفت[۴].

pH محلول به طور قابل توجهی می‌تواند پروسه جذب را تحت تاثیر قرار دهد بدین صورت که؛ هرچه pH پایین‌تر باشد، سرعت جذب بیشتر شده که به دلیل افزایش حلالیت یون‌های فلزی بوده که عامل تقویت تعامل آن‌ها با جاذب است. به علاوه، افزایش دما سبب تقویت بازده حذف آهن در یک واکنش گرماگیر شده که افزایش دهنده تعامل جاذب و آلاینده است[۵].

به منظور بهینه سازی فرآیندهای حذف، دانستن گزینش‌پذیری أنواع جاذب‌ها حائز اهمیت است. برای مثال، در مطالعات متعددی از زئولیت طبیعی کلینوپتیلولیت در جهت اثربخشی حذف انتخابی آهن، کروم و مس استفاده شده است. اهمیت این گزینش‌پذیری در طراحی سیستم های تصفیه‌ای هدف گیرنده آلاینده‌های خاص و به حداقل رساندن حذف مواد مغذی ضروری، نشان داده شده است. به علاوه، خواص خود آلاینده مانند pH، قدرت یونی و غلظت نیز اهمیت بالایی در پروسه جذب دارند[۶].

با استفاده از داده‌های تحقیقات ورمیکولیت منبسط شده و غیر منبسط شده، اهمیت انتخاب بهترین جاذب بر طبق هزینه، کارایی و دسترسی را مورد بررسی قرار داد[۷].

با وجود پیشرفت‌های صورت گرفته در زمینه تکنیک‌های جذب، هنوز هم روش‌های آزمایشگاهی ارزیابی حذف آلاینده‌ها، از نظر هزینه و زمان، به صرفه نیستند. در این نوع روش‌ها از جمع‌آوری نمونه‌های پساب و انجام آزمایش‌ها در شرایط مختلف، به منظور ارزیابی میزان جذب فلز، استفاده می‌شود. به دلیل تنوع ذاتی این روش‌ها، نتایج حاصل متناقض بوده و بهینه‌سازی فرآیند درمان را پیچیده می‌کند. به علاوه به دلیل پیچیدگی این روش‌ها، امکان اجرای به موقع راه‌حل‌های موثر برای رسیدگی به چالش‌های آلودگی آب، فراهم نمی‌شود. با ظهور هوش مصنوعی، شرایط مناسبی برای پیشرفت این فرآیندها فراهم شد. می‌توان از طریق ایجاد مدل‌های پیش‌بینی کننده توسط هوش مصنوعی، کارایی حذف فلز را در شرایط عملیاتی مختلف ارزیابی کرد و فرآیند ساده‌سازی انتخاب پارامتر بهینه را انجام داد. هوش مصنوعی می‌تواند با استفاده از داده‌های بسیار زیاد حاصل از مطالعات انجام شده، الگوها و همبستگی‌هایی را شناسایی کند که از طریق متدهای آنالیز مرسوم امکان‌پذیر نیست. در نتیجه شرایط برای تصمیم گیری آگاهانه‌تر در طراحی و بهره برداری از سیستم های تصفیه فاضلاب، فراهم می‌شود. در رشته مهندسی شیمی غالباً مدل‌های پایه مبتنی بر اصول ریاضی حاکم بر پدیده‌هایی مانند انتقال جرم، انتقال حرارت و انتقال تکانه می‌باشند. با این وجود، در سیستم‌های پیچیده، به دلیل تعاملات پیچیده، مدل‌سازی دقیق یک چالش است. اما مدل‌های یادگیری ماشین و هوش مصنوعی

ISSN: 2588-2805
http://www.ieaj.ir

قابلیت شناسایی و آنالیز الگوهای رفتاری را داشته و می‌توانند رویکردهای با سازگاری بیشتر را انتخاب کنند. مزیت اصلی این مدل‌ها، یادگیری مدام آن‌ها از داده‌های جدید است که آن‌ها را تبدیل به گزینه مناسبی برای محیط‌های پویا مانند تأسیسات تصفیه فاضلاب می‌کند[۸].

همزمان با پیشرفت قدرت محاسباتی و با توجه به داده‌های بزرگ در دسترس، کاربرد فنلوری‌های هوش مصنوعی در صنعت افزایش داشته است. در مهندسی شیمی می‌تواند از هوش مصنوعی در جهت مدل سازی راکتور گرفته تا بهینه سازی و کنترل، کمک گرفت. مزایای رویکردهای داده محور شامل؛ حذف نیاز به دانش قبلی از سیستم هدف، توانایی کشف الگوهایی با تشخیص دشوار توسط انسان و افزایش سرعت اجرای محاسبات در مقایسه با روش‌های سنتی، هستند. این قابلیت‌ها شرایط را برای تعدیل سریع در استراتژی‌های تصفیه بر اساس داده‌های بلادرنگ فراهم کرده که سبب حذف کارآمدتر و موثرتر آلاینده‌ها می‌شود[۹].

با این حال، اثربخشی این مدل‌های داده‌محور به کیفیت و کمیت داده‌های ورودی بستگی دارد. جمع‌آوری داده‌های جامع برای ساخت مدل‌های قوی که می‌توانند نتایج را تحت شرایط مختلف به دقت پیش‌بینی کنند، ضروری است. علاوه بر این، عملکرد مدل‌های یادگیری ماشینی می‌تواند تحت تأثیر نویز یا نقاط پرت در داده‌ها قرار گیرد، که می‌تواند منجر به پیش‌بینی‌های اشتباه شود. برای کاهش این چالش‌ها، جمع‌آوری مجموعه‌های داده‌های متعدد در شرایط مختلف بسیار مهم است، تا اطمینان حاصل شود که مدل‌ها بر روی داده‌های دقیق و معرف آموزش داده شده‌اند

باید دقت داشت که کمیت و کیفیت داده‌های ورودی بر اثربخشی این مدل‌های داده محور موثر هستند. به علاوه، باید برای ساخت مدل‌های قدرتمند پیش‌بینی کننده نتایج در شرایط مختلف، داده‌های جمع‌آوری شده کامل باشند. به علاوه، باید دقت داشت که نویزها یا نقاط پرت داده‌ها، بر مدل‌های یادگیری ماشینی اثر گذاشته و سبب پیش‌بینی نادرست آن‌ها می‌شوند. جهت رفع این مشکل باید مجموعه داده‌ها در أنواع موقعیت‌های مختلف انجام شود تا از آموزش مدل‌ها با داده‌های دقیق اطمینان حاصل شود[۱۰].

بر طبق مطالعات هوش مصنوعی در حوزه زیست محیطی مانند توسعه تالاب های ساخته شده با جریان عمودی زیرسطحی برای تصفیه آب طوفان شهری نیز کاربردی هستند. با استفاده از مدل شبکه عصبی در مطالعات، حذف نیتروژن و فسفر تام از آب طوفان موفقیت آمیز بوده است که نشان می‌دهد چگونه هوش مصنوعی می‌تواند فرآیندهای نظارت و کاهش هزینه‌ها را ساده‌سازی کند. محققان با استفاده از تکنیک های تجزیه و تحلیل داده های اکتشافی، مانند تجزیه و تحلیل مؤلفه های اصلی (PCA)، تعداد متغیرهای ورودی را در عین حفظ دقت پیش‌بینی کاهش داده‌اند. بهبود استراتژی‌های عملیاتی سیستم تصفیه و پیشبرد مدیریت آب شهری از طریق ادغام هوش مصنوعی به صورت مدل‌سازی شبکه عصبی مصنوعی، امکان‌پذیر است[۱۱].

به طور خلاصه، ادغام تکنیک‌های هوش مصنوعی و جذب، تلاشی آگاهانه برای کاهش آلودگی آب ناشی از فلزات سنگین در فاضلاب صنعتی است. هوش مصنوعی به محققان و مهندسان امکان انتخاب و توسعه روش‌های کارآمدتر، مقرون‌به‌صرفه‌تر و سازگارتر را برای مقابله با این چالش حیاتی زیست‌محیطی، فراهم می‌کند. همزمان با کشف روش‌های نوین تصفیه فاضلاب، نقش هوش مصنوعی در این حوزه بیشتر شده و امکان داشتن آینده‌ای پایدارتر و پاک‌تر فراهم می‌شود. در این مقدمه ما به مرور جامع موضوع پرداخته و اهمیت تکنیک‌های جذب و هوش مصنوعی در پرداختن به چالش حذف فلزات سنگین از پساب‌های صنعتی را نشان دادیم. هدف این پروژه، توسعه یک مدل هوش مصنوعی با قابلیت پیش‌بینی بازده حذف فلزات سنگین از پساب‌های صنعتی است. ما عوامل کلیدی موثر بر حذف فلز را شناسایی و داده های مربوطه را جمع آوری کرده‌ایم. طراحی این مدل برای ارزیابی بازده حذف بر اساس شرایط عملیاتی خاص و بر اساس ویژگی‌های فاضلاب مورد نظر، مجزا صورت خواهد گرفت.

روش شناسی

شناسایی ورودی ها و خروجی های مدل

ما در این مقاله با کمک مدل‌های یادگیری ماشینی، امکان پیش‌بینی بازده حذف جاذب‌های خاص تحت شرایط مختلف عملیاتی و ویژگی‌های فاضلاب بررسی را بررسی کردیم. کیفیت داده‌های مورد استفاده بر اعتبار مدل‌های یادگیری اثرگذار هستند. داده‌های این پروژه حاصل از مقالات تجربی است. در هر مقاله، حذف آلاینده‌های خاص از ضایعات فلزی با استفاده از جاذب خاصی

ISSN: 2588-2805
http://www.ieaj.ir

انجام شده بود و در نهایت در همه مقالات میزان حذف، بازده حذف یا غلظت نهایی آلاینده گزارش شده بود. بر طبق مطالعات انجام شده، عوامل زیر بر بازده حذف موثر هستند[۱۲]:

نوع آلاینده: میزان جذب متاثر از ماهیت شیمیایی و خواص فیزیکی آلاینده است. آلاینده‌ها با جاذب‌ها در تعامل بوده که این عمل بر عملکرد کلی تاثیرگذار است.

نوع جاذب: اهمیت نوع جاذب مختلف مانند کربن فعال، زئولیت‌ها یا زیست‌توده با توجه به سطوح ویژه، اندازه منافذ و ترکیبات شیمیایی آن‌ها است که بر جذب آلاینده‌ها تاثیرگذار هستند.

دما: اثر دما بر انرژی جنبشی مولکول های درگیر در فرآیند جذب مشخص بوده و هرچه دما بالاتر باشد سرعت جذب نیز افزایش می‌یابد. البته بسته به ماهیت جاذب و آلاینده ممکن است سرعت جذب را کاهش دهد.

غلظت اولیه آلاینده: در ابتدای واکنش، غلظت آلاینده بر نیروی محرکه اثر گذاشته و هرچه غلظت بالاتر باشد، ظرفیت اشباع جذب نیز افزایش می‌یابد.

دوز جاذب: ارتباط مستقیمی بین مقدار جاذب مصرفی با بازده حذف آلاینده های سیستم وجود دارد. انتخاب دوز بهینه به منظور اثربخشی حداکثری بدون اتلاف منابع، ضروری است.

زمان:یک عامل حیاتی مدت زمان تماس جاذب و آلاینده است. اگر زمان کافی در نظر گرفته شود، تعاملات موثرتر بوده و راندمان جذب نیز افزایش می‌یابد.

PH: اسیدی یا قلیایی بودن محلول بر بار و حلالیت آلاینده و جاذب اثر می‌گذارد و سبب تغییر فرآیند جذب می‌شود. باید در نظر داشت که جذب بهینه برخی آلاینده‌ها در PH خاصی صورت می‌گیرد.

هدف ما در این پروژه، پیش‌بینی بازده حذف انواع جاذب‌ها در شرایط متفاوت است. تمامی نتایج حاصل با معادله زیر به بازده حذف تبدیل شدند[۱۳]:

$$R = \frac{Q \times M_a \times 100}{V \times C_0}$$

ظرفیت جذب

بازده حذف در جاذب‌ها به عنوان نسبت کاهش آلاینده‌های موجود در یک سیستم یا محیط پس از گذر از فرآیند تصفیه یا جذب تعریف می‌شود. این مفهوم نشان می‌دهد که چه میزان از آلاینده‌ها توسط جاذب حذف شده‌اند. بازده حذف تحت تأثیر عوامل مختلفی قرار دارد، مانند مقدار جاذب، غلظت اولیه آلاینده و شرایط عملیاتی مانند pH و زمان تماس.

در تحقیقات انجام شده بر روی برخی جاذبان، مشاهده شده است که با افزایش مقدار جاذب تا نقطه‌ای خاص، بازدهی حذف افزایش می‌یابد؛ اما پس از آن ممکن است به دلیل تشکیل کلوخه‌های بزرگتر و کاهش سطح ویژه، بازدهی کاهش یابد [۱۲].

پیش پردازش داده ها و کاربرد مدل های یادگیری ماشین

پس از بررسی مسائل اصلی تاثیرگذار بر مدیریت پسماند، وارد مرحله اساسی جمع‌آوری داده می‌شویم. ما در این مطالعه داده‌ها را از مقالات آورده شده در بخش رفرنس‌ها، جمع‌آوری کرده‌ایم. غالب روش‌های موجود در این منابع مبتنی بر تکنیک‌های آزمایشگاهی بوده که داده‌های آن‌ها به صورت جدول و نمودار آورده شده است. از نرم افزار GetData Graph Digitizer برای آنالیز موثر داده‌ها استفاده شده است. داده‌های غیرعددی، مانند نوع آلاینده و جاذب، برچسب گذاری شده و برای مقادیر عددی مدل تعریف شد.

جدول ۱: انواع جاذب ها وآلاینده های بررسی شده

آلاینده	جاذب	منبع
آهن، سرب،کادمیوم، نیکل، مس	پوسته برنج	[۴]
آهن، سرب،کادمیوم، نیکل، مس	خاکستر بادی	[۴]
کروم	ذرات نانوآهن	[۵]
مس، کروم، آهن	زئولیت کلینوپتیلولیت	[۶]

ISSN: 2588-2805
http://www.ieaj.ir

مس، کادمیوم، سرب	ورموکولیت منبسط شده	[۷]
مس، کادمیوم، سرب	ورموکولیت منبسط نشده	[۷]
نیکل	نانو لوله کربنی تک جداره	[۸]
نیکل	نانو لوله کربنی چند جداره	[۸]
روی، نیکل، مس	زغال سنگ نارس	[۹]

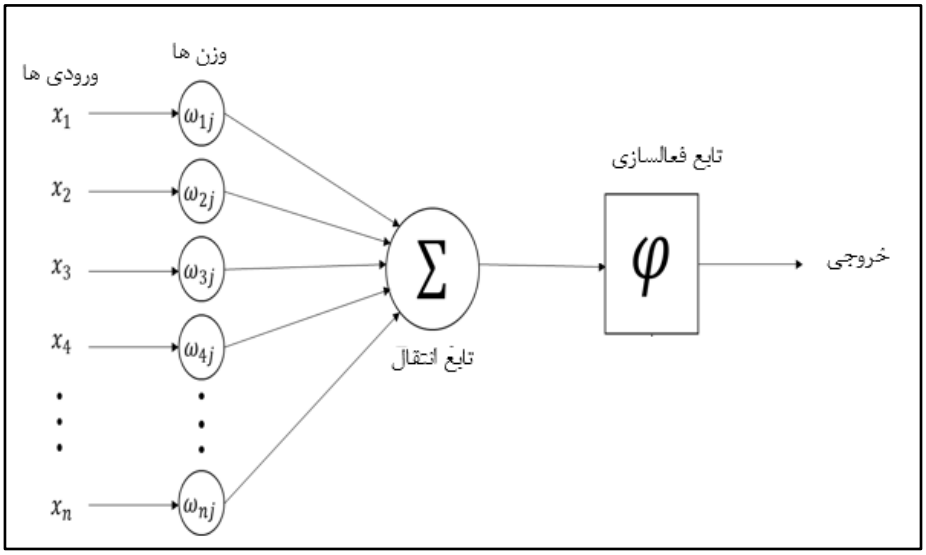
جزئیات بیشتر در مورد هر داده در مقالات مربوطه قابل بررسی است.

ابزارها و متدولوژی های قدرتمند ارائه شده با یادگیری ماشینی، درک ما از فرآیند جذب را بهبود داده و به انتخاب جاذب بهینه کمک می‌کند. با این تکنیک‌ها امکان آنالیز داده‌های پیچیده، شناسایی الگوها و پیش‌بینی نتایج بر اساس داده‌های موجود، وجو دارد[۱۴].

یکی از اولین کاربردهای یادگیری ماشینی در این مطالعه، انتخاب بهینه جاذب‌ها است. آموزش مدل‌های داده‌های مربوط به عملکرد جاذب‌های مختلف امکان پیش‌بینی نوع ماده با بهترین عملکرد در شرایط خاص تعریف شده، فراهم می‌شود. به علاوه، با استفاده از الگوریتم‌های یادگیری ماشین ما قادر به پیش‌بینی کارایی جذب آلاینده‌های مختلف بر اساس پارامترهای کلیدی مانند غلظت، دما و سطوح Ph خواهیم بود. قابلیت پیش‌بینی عامل موثری در توسعه راهبرهای درمانی می‌باشد[۱۵].

یکی دیگر از مزایای یادگری ماشینی، ساده سازی آنالیز داده‌های بزرگ است که سبب تسهیل نتیجه‌گیری معنادار شده و عامل بهبود بخشیدی به فرآیندهای تصمیم‌گیری است. در این پروژه ما از چندین مدل یادگیری ماشین، مانند مدل‌های رگرسیون به منظور درک روابط بین پارامترهای ورودی و کارایی جذب، مدل‌های طبقه‌بندی برای دسته‌بندی جاذب‌ها بر اساس اثربخشی آن‌ها، و تکنیک‌های خوشه‌بندی برای گروه‌بندی جاذب‌های مشابه بر اساس معیارهای عملکردی، استفاده کرده‌ایم[۱۶] .

در مورد الگوهای داده با پیچیدگی بیشتر، به بررسی شبکه‌های عصبی خواهیم پرداخت که قادر به مدل‌سازی روابط غیر خطی بین متغیرها هستند. هدف ما در این مقاله ادغام یادگیری ماشینی در انجام کار به منظور افزایش کارایی و اثربخشی فرآیند جذب در تصفیه پساب های صنعت پوشش فلزی می‌باشد. این روش خلاقانه سبب آنالیز ساده‌تر داده‌ها شده و در توسعه راه حل های تصفیه پایدارتر و موثرتر برای پساب واحدهای پوشش فلزی نیز کارآمد است[۱۵].



تصویر ۱: چارچوب شماتیک شبکه عصبی مصنوعی در مدل‌سازی هوش مصنوعی با استفاده از یادگیری عمیق.

ارزیابی مدل

ارزیابی معیارهای تعیین شده بعد از اعمال مدل‌های ماشینی باید انجام شود. ضریب تعیین یکی از مهم‌ترین معیارهای ارزیابی است. ضریب تعیین (R2)، معیاری آماری است که اثربخشی مدل رگرسیونی را توضیح می‌دهد. به علاوه این ضریب می‌تواند نسبت واریانس در متغیر وابسته را با توجه به متغیرهای مستقل، تبدیل به کمیت کند. به عبارتی، R2 قادر است مشخص کند ویژگی‌های

ISSN: 2588-2805
http://www.ieaj.ir

انتخابی چگونه نتیجه مورد نظر ما را پیش‌بینی کنند. در نتیجه ابزار حیاتی برای ارزیابی کیفیت مدل‌های رگرسیون در یادگیری ماشین است. R2 مقادیری از ۰ تا ۱ را می‌پذیرد. مقدار ۱ نشان دهنده این است که مدل می‌تواند متغیر وابسته را کاملاً پیش‌بینی کند که به معنی محاسبه تمام تغییرات توسط مدل است. مقدار ۰ نیز نشان دهنده این است که مدل قادر به توضیح هیچ یک از واریانس‌های متغیر وابسته نیست یا به عبارتی پیش‌بینی‌های مدل به میانگین متغیر وابسته نیشت. با استفاده از دامنه‌های تعریف شده می‌توان عملکرد مدل را مستقیماً تعریف کرد.

بعد از اعمال یادگیری ماشینی باید این مدل‌ها را با معیارهای تعیین شده ارزیابی کرد. ضریب تعیین (R2)، معیاری آماری است که اثربخشی مدل رگرسیونی را توضیح می‌دهد. با این ضریب می‌توان نسبت واریانس در متغیر وابسته را با توضیح متغیر وابسته مدل، کمی‌سازی کند. به عبارت دیگر R2، نشان می‌دهد که ویژگی‌های انتخابی ما بر پیش‌بینی نتایج مدنظر موثر است. بنابراین می‌توان گفت که این ضریب ابزاری کلیدی در ارزیابی کیفیت مدل‌های رگرسیون در یادگیری ماشین است. R2 مقادیری از ۰ تا ۱ را می‌پذیرد و مقادیر آن مدل متغیر کاملاً وابسته را پیش‌بینی می‌کند.

محاسبه R^2 مستلزم درک دو جزء کلیدی است: مجموع مجذورات (SS_{tot}) و مجموع باقیمانده مربعات .(SS_{res}) مجموع مجذورات واریانس کل در متغیر وابسته را اندازه گیری می کند، در حالی که مجموع مجذورات باقیمانده واریانسی را اندازه می گیرد که توسط مدل توضیح داده نشده است. فرمول R^2 به صورت زیر ارائه می شود:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

با این فرمول می‌توان R2 را برای نمایش عددی واضح از توان توضیحی مدل ارائه کرد[۱۷].

نتایج و بحث ها

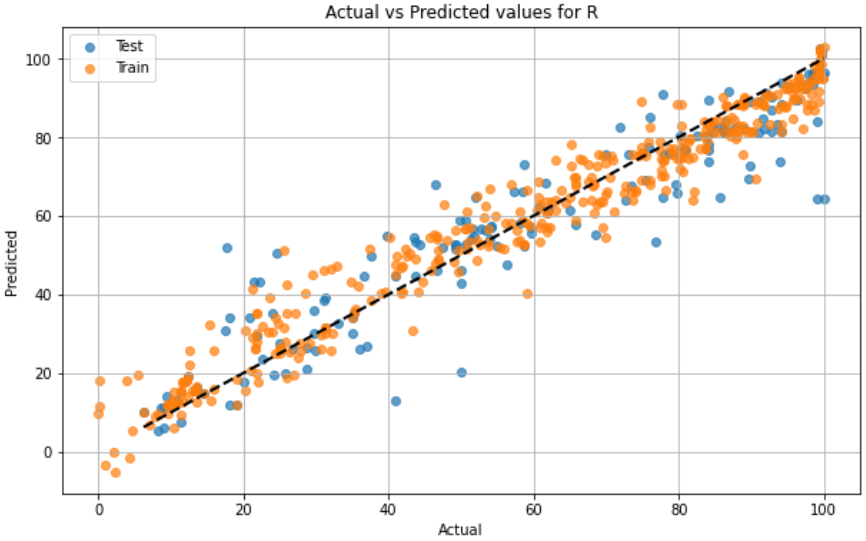
بعد از اینکه ندل‌های یادگیری ماشین اعمال شد باید با استفاده از معیارهای مشخص آن‌ها را ارزیابی کرد تا بتوان مدل بهینه را انتخاب کرد. هر مدل بسته به محدوده آموزش داده شده دارای نقاط ضعف و قدرتی است. بعد از بررسی علمکرد مدل در بین تمامی داده‌ها، محدوده‌ای که مدل بهترین پیش‌بینی را دارد مشخص می‌شود. با این آنالیز می‌توان نقاط قوت و ضعف را شناسایی کرد. به علاوه، با توجه به معیارهای مختلف ارزیابی، امکان بررسی جامع‌تر مدل فراهم می‌شود. این ارزیابی‌ها هم به انتخاب مدل و هم به فراهم کردن زمینه پیشرفت‌های آینده در مدل‌سازی و پیش‌بینی، کمک می‌کند. بنابراین می‌توان نتیجه گرفت که این فرآیند به ما امکان استفاده موثرتر از مدل‌های یادگیری ماشینی در برنامه‌های کاربردی دنیای واقعی را داده و به ارائه پیش‌بینی‌های دقیق‌تر کمک می‌کند.

پیش بینی های رگرسیور تقویت کننده گرادیان

یکی از تکنیک‌های یادگیری گروهی، رگرسیور تقویت کننده گرادیان است که با ترکیب کردن چندین یادگیرنده ضعیف، درخت‌های تصمیم، به صورت متوالی یک مدل را پیش‌بینی می‌کند. در هر سری تکرار، تمرکز الگوریتم بر اصلاح خطاهای قبلی با برازش یک مدل جدید به باقی‌مانده‌های پیش‌بینی‌های ترکیبی. با این روش می‌توان یک تابع از دست دادن متمایز را بهینه‌سازی کرد و از آن برای کارهای رگرسیون مختلف، استفاده کرد. نقطه قوت رگرسیور تقویت کننده گرادیان، مدیریت روابط پیچیده توسط آن است که اثربخشی آن وابسته به کاهش بیش از حد برازش از طریق تکنیک های منظم سازی، می‌باشد[۱۸].

نمودار زیر مقایسه مقدار واقعی و مقدار پیش بینی شده با این مدل را در جهت عملکرد پیش‌بینی این مدل، به نمایش می‌گذارد.

ISSN: 2588-2805
http://www.ieaj.ir

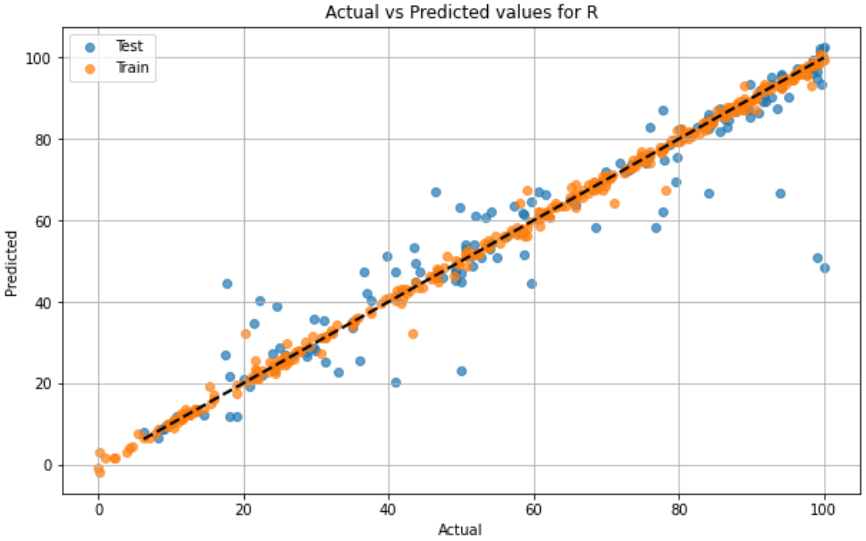


تصویر ۲: مقدار واقعی و مقدار پیش بینی شده برای رگرسیور تقویت کننده گرادیان

پیش بینی های رگرسور تقویت کننده طبقه ای

یکی ار الگوریتم‌های تقویت کننده گرادیان، رگرسور تقویت کننده طبقه ای است که به وسیله Yandex توسعه پیدا کرد تا در دسته‌بندی ویژگی‌ها موثرتر عمل کند. این روش برخلاف سایر روش‌های هم نوع خود، به طور خودکار متغیرهای طبقه‌بندی را مستقل از پیش‌پردازش گسترده مانند رمزگذاری یک‌طرفه، تحت کنترل در می‌آورد. با این قابلیت، یکپارچگی داده‌ها حفظ شده و عملکرد مدل بهبود می‌یابد. این رگسور از یک ساختار درختی متقارن استفاده کرده و برای کمک به کاهش بیش از حد برازش و افزایش تعمیم از تقویت منظم بهره می‌برد. به سبب کارایی و سهولت در استفاده، این روش یک گزینه مناسب برای مجموعه داده‌هایی است که تعداد متغیر قابل توجهی دارند [۱۹].

با استفاده از نمودار مقایسه کننده مقادیر واقعی با مقادیر پیش‌بینی کننده، عملکرد پیشبینی موثر مدل را می‌توان نشان داد. بر طبق شکل ۳، این مدل در سطوح بالای بازده حذف، دارای بهترین عملکرد است.



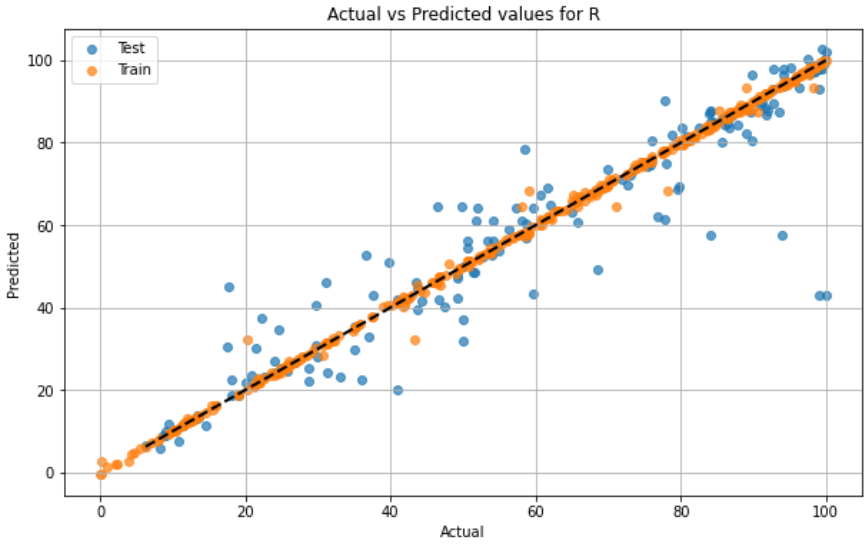
تصویر ۳: مقدار واقعی و مقدار پیش بینی شده برای رگرسور تقویت کننده طبقه ای

ISSN: 2588-2805
http://www.ieaj.ir

پیش بینی های رگرسور تقویت کننده شدید

رگرسور تقویت کننده شدید به عنوان یک الگوریتم تقویت گرادیان، کارایی بالا و دقت بسیاری در مسائل مربوط به یادگیری ماشینی خصوصاً در رگرسیون و طبقه‌بندی دارد. این الگوریتم از تکنیک‌های پیشرفته‌ای مانند پردازش موازی و بهینه‌سازی درختان تصمیم استفاده کرده که سبب افزایش سرعت و کاهش زمان آموزش شده است. رگرسور تقویت کننده شدید می‌تواند روابط غیرخطی پیچیده بین ویژگی‌ها و متغیرهای هدف را مدیریت کرده و اهمیت هر خصیصه را نشان دهد که این قضیه سبب می‌شود در جهت انتخاب بهتر ویژگی‌ها و بهبود عملکرد مدل بسیار کارآمد عمل کند. یکی از دلایلی که این ابزار بین محققان و مهندسان بسیار محبوبیت دارد، امکان جلوگیری از بیش‌برازش و تنظیم دقیق پارامترها توسط این رگرسور است[۲۰].

با استفاده از نمودار مقایسه مقادیر واقعی در مقابل مقادیر پیش بینی شده می‌توان نشان داد که این مدل قابلیت پیش‌بینی قوی دارد. بر طبق شکل ۴، این مدل در مقادیر بازده حذف بیشتر، پیش‌بینی خوبی ارائه می‌کند.



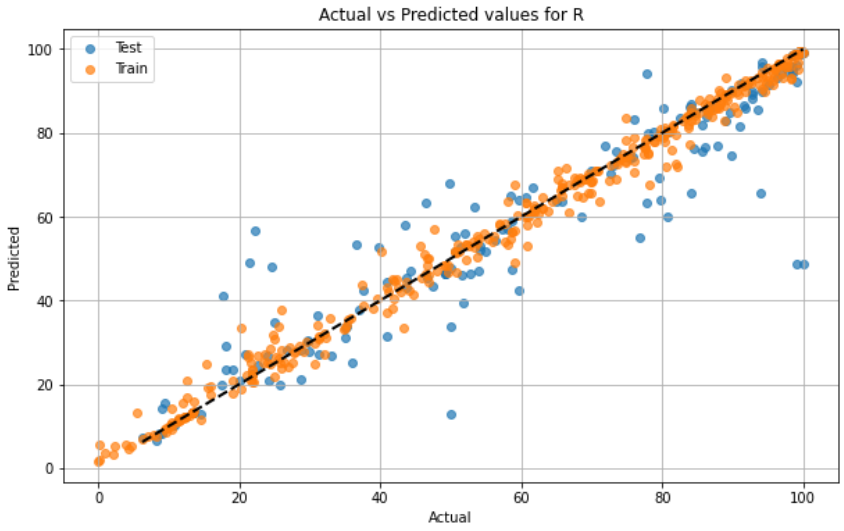
تصویر ۴: مقدار واقعی و مقدار پیش بینی شده برای رگرسور تقویت کننده شدید

پیش بینی های رگرسیون تصادفی جنگل

در رگرسیون تصادفی جنگل به عنوان یک روش یادگیری مجموعه‌ای، چندین درخت تصمیم در جهت بهبود عملکرد کلی و استحکام مدل ترکیب می‌شوند که هر یک از این درخت‌ها، بر روی یک زیر مجموعه تصادفی از ویژگی ها آموزش داده شده و پیش‌بینی نهایی حاصل جمع کل این پیش‌بینی‌ها خواهد بود. رگرسیون تصادفی جنگل قادر به مدیریت داده‌های با حجم زیاد را داشته و می‌تواند تعامل بین ویژگی‌ها را تشخیص داده و ویژگی‌ها را بر اساس اهمیت آن‌ها طبقه‌بندی کند. این مدل در مقایسه با درخت تصمیم‌گیری فردی، کمتر مستعد برازش بیش از حد بوده و قابلیت مدیریت متغیرهای عددی و طبقه‌ای را نیز دارد[۲۱].

با استفاده از نمودار نشانگر رابطه بین مقادیر واقعی و پیش بینی شده، می‌توان عملکرد مدل را در پیش بینی‌ها را نشان داد. بر طبق شکل ۵، در مقادیر بازده حذف بیشتر، مدل دقت افزایش یافته‌ای را نشان می‌دهد.

ISSN: 2588-2805
http://www.ieaj.ir

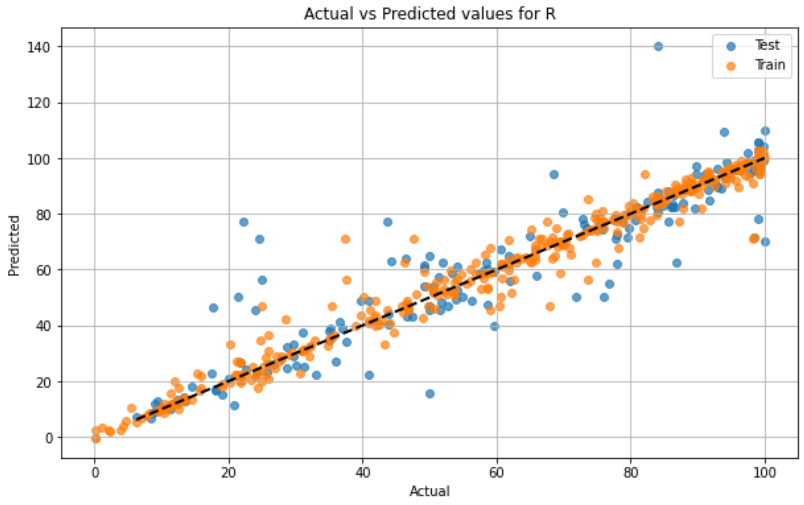


تصویر ۵: مقدار واقعی و مقدار پیش بینی شده برای رگرسیون تصادفی جنگل

پیش بینی شبکه های عصبی

یک دسته از مدل‌های یادگیری ماشینی، شبکه‌های عصبی دسته‌ای هستند که با الهام از ساختار و عملکرد مغز انسان ساخته شده‌اند. این مدل‌ها تشکیل شده از لایه های به هم پیوسته گره ها (نورون ها) بوده که داده های ورودی را برای یادگیری الگوها و روابط پیچیده پردازش می‌کنند. مهارت اصلی این شبکه‌های عصبی در گرفتن روابط غیرخطی است خصوصاً در مورد تشخیص تصویر، پردازش زبان طبیعی و پیش‌بینی سری‌های زمانی بسیار کارآمد هستند. این مدل‌ها به مجموعه داده های بزرگتر و قدرت محاسباتی بیشتری در مقایسه با الگوریتم های سنتی نیاز داشته و علاوه بر این، به دلیل انعطاف پذیری و قابلیت مدل سازی الگوهای پیچیده، تبدیل به ابزاری قدرتمند برای یادگیری ماشینی می‌شود[۱۵].

با استفاده از نموداری که مقادیر واقعی و پیش‌بینی‌شده را در این مدل کنار هم قرار می‌دهد، می‌توان عملکرد پیش‌بینی ان را نشان داد. بر طبق شکل ۶، در مقادیر بازده حذف بیشتر، این مدل موثرتر عمل می‌کند.



تصویر ۶: مقدار واقعی و مقدار پیش بینی شده برای شبکه عصبی

با آنالیز نمودارهای بالا می‌توان نشان داده که در مقادیر جذب کمتر، تمامی مدل‌ها عملکرد بهینه‌تری دارند. در نهایت باید برای انتخاب بهترین مدل، مقادیر R^2 هر مدل را با سایر مدل‌ها مقایسه کرد تا بتوان نزدیک‌ترین مدل به ۱ را انتخاب کرد.

ISSN: 2588-2805
http://www.ieaj.ir

جدول ۲: مقایسه بین مدل های یادگیری ماشینی

مدل های یادگیری ماشین	R^2
شبکه عصبی	۰/۸۳
رگرسور تقویت کننده گرادیان	۰/۸۸
رگرسور تقویت کننده طبقه ای	۰/۹۱
رگرسور تقویت کننده شدید	۰/۹۳
رگرسیون تصادفی جنگل	۰/۸۷

بر طبق جدول ۲، بهترین مدل پیش‌بینی بازده حذف، رگرسیور تقویت کننده شدید با مقدار R^2 = ۰/۹۳ است.

نتیجه گیری

هدف این پروژه تحقیقاتی، توسعه مدل‌های هوش مصنوعی خصوصاً الگوریتم‌های یادگیری ماشینی، به منظور پیش‌بینی بازده حذف جاذب‌های مختلف برای حذف فلزات سنگین از پساب‌های صنعتی است. با انالیز عوامل اصلی از جمله نوع آلاینده، جاذب، دما، غلظت اولیه، دوز، زمان تماس و pH، این مدل‌ها را برای پیش‌بینی کارایی فرآیند جذب می‌توان اموزش داد. بر طبق داده‌های به دست آمده، مدل رگرسیون تقویت کننده شدید، با ضریب همبستگی R^2 نزدیک به یک (۰/۹۳)، بهترین عملکرد را در پیش‌بینی بازده حذف دارد. تاکید مطالعه حاضر بر استفاده از تکنیک‌های پیشرفته یادگیری ماشین برای بهینه‌سازی مدیریت پساب‌های صنعتی و افزایش حذف فلزات سنگین بوده است. می‌توان با پیش‌بینی دقیق بازده حذف تحت شرایط عملیاتی مختلف، انتخاب پارامتر بهینه توسط این مدل‌ها را تسهیل کرده و منجر به کارآمدی و مقرون به صرفه شدن فرآیندهای تصفیه فاضلاب شد. یکی از راه‌های توسعه روش‌های پایدارتر و سازگارتر برای رسیدگی به موضوع مبرم آلودگی آب ناشی از آلودگی فلزات سنگین، ادغام آن‌ها با هوش مصنوعی است. با توجه به افزایش داده‌های در دسترس و تقویت قدرت محاسباتی، هوش مصنوعی می‌تواند پتانسیل خود در برنامه های زیست محیطی را توسعه داده و در جهت داشتن آینده‌ای پتاک‌تر و سالم‌تر برای اکوسیستم‌ها و جوامع بشری کمک کننده باشد. به علاوه، این پروژه فرصتی برای گسترش مطالعات و جمع‌آوری داده‌های اضافی فراهم کرده که سبب می‌شود طیف وسیعی از جاذب‌ها و شرایط عملیاتی متفاوت را شامل شود. این مدل ابتکاری، اعتبار مدل ما را نیز افزایش می‌دهد زیرا هرچه کیفیت داده‌ها ورودی بیشتر باشد، دقت یادگیری ماشینی نیز بیشتر خواهد بود. در نتیجه، به منظور شناسایی و بهینه‌سازی شرایط ایداه‌آل جذب می‌توان کدهای مربوطه را پالایش کرد.

ISSN: 2588-2805
http://www.ieaj.ir

منابع و مراجع

[1] Maleki, A., Mohammad, M., Emdadi, Z., Asim, N., Azizi, M., & Safaei, J. (2020). Adsorbent materials based on a geopolymer paste for dye removal from aqueous solutions. Arabian Journal of Chemistry, 13(1), 3017-3025.

[2] Emdadi, Z., Asim, N., Amin, M. H., Ambar Yarmo, M., Maleki, A., Azizi, M., & Sopian, K. (2017). Development of green geopolymer using agricultural and industrial waste materials with high water absorbency. Applied Sciences, 7(5), 514.

[3] Salehi, M. M., Hassanzadeh-Afruzi, F., Esmailzadeh, F., Choopani, L., Rajabi, K., Kuzekanan, H. N., ... & Maleki, A. (2024). Chlorpyrifos and diazinon elimination through pAAm-g-XG/HKUST-1@ Fe3O4 biopolymer nanoadsorbent hydrogel from wastewater: Preparation, characterization, kinetics and isotherm. Separation and Purification Technology, 334, 126097.

[4] Hegazi, H. A. (2013). Removal of heavy metals from wastewater using agricultural and industrial wastes as adsorbents. HBRC journal, 9(3), 276-282.

[5] Sharma, Y. C., Srivastava, V., Weng, C. H., & Upadhyay, S. N. (2009). Removal of Cr (VI) from wastewater by adsorption on iron nanoparticles. The Canadian Journal of Chemical Engineering, 87(6), 921-929.

[6] Zanin, E., Scapinello, J., de Oliveira, M., Rambo, C. L., Franscescon, F., Freitas, L., ... & Dal Magro, J. (2017). Adsorption of heavy metals from wastewater graphic industry using clinoptilolite zeolite as adsorbent. Process Safety and Environmental Protection, 105, 194-200.

[7] dos Santos, A. C. V., & Masini, J. C. (2007). Evaluating the removal of Cd (II), Pb (II) and Cu (II) from a wastewater sample of a coating industry by adsorption onto vermiculite. Applied Clay Science, 37(1-2), 167-174.

[8] Noh, W., Park, S., Kim, S., & Lee, I. A Hybrid Framework of Knowledge-Based Digital Twin and Data-Driven Machine Learning Models for Optimizing Control Parameters of Chemical Processes. Available at SSRN 4220989.

[9] Chakraborty, A., Serneels, S., Claussen, H., & Venkatasubramanian, V. (2022). Hybrid ai models in chemical engineering–a purpose-driven perspective. Computer Aided Chemical Engineering, 51, 1507-1512.

[10] Nentwich, C., Gebus, P., Brächer, A., & Markovic, A. (2021). Hybrid Process Modeling of an Industrial Process. Chemie Ingenieur Technik, 93(12), 2092-2096.

[11] Kiiza, C., Pan, S. Q., Bockelmann-Evans, B., & Babatunde, A. (2020). Predicting pollutant removal in constructed wetlands using artificial neural networks (ANNs). Water Science and Engineering, 13(1), 14-23.

[12] Li, F., Yuasa, A., Ebie, K., Azuma, Y., Hagishita, T., & Matsui, Y. (2002). Factors affecting the adsorption capacity of dissolved organic matter onto activated carbon: modified isotherm analysis. Water Research, 36(18), 4592-4604.

[13] Onwordi, C. T., Uche, C. C., Ameh, A. E., & Petrik, L. F. (2019). Comparative study of the adsorption capacity of lead (II) ions onto bean husk and fish scale from aqueous solution. Journal of water reuse and desalination, 9(3), 249-262.

[14] Iftekhar, S., Ramasamy, D. L., Srivastava, V., Asif, M. B., & Sillanpää, M. (2018). Understanding the factors affecting the adsorption of Lanthanum using different adsorbents: a critical review. Chemosphere, 204, 413-430.

[15] Zhang, W., Huang, W., Tan, J., Huang, D., Ma, J., & Wu, B. (2023). Modeling, optimization and understanding of adsorption process for pollutant removal via machine learning: Recent progress and future perspectives. Chemosphere, 311, 137044.

[16] Ullah, H., Khan, S., Chen, B., Shahab, A., Riaz, L., Lun, L., & Wu, N. (2023). Machine learning approach to predict adsorption capacity of Fe-modified biochar for selenium. Carbon Research, 2(1), 29.

[17] Matheri, A. N., Ntuli, F., Ngila, J. C., Seodigeng, T., & Zvinowanda, C. (2021). Performance prediction of trace metals and cod in wastewater treatment using artificial neural network. Computers & Chemical Engineering, 149, 107308.

ISSN: 2588-2805
http://www.ieaj.ir

[18] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.
[19] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.
[20] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[21] Forests, R., & Breiman, L. (1999). Statistics department university of california berkeley.