



کنگره علوم و مهندسی آب و فاضلاب ایران

دانشگاه تهران، تهران

۲۶ و ۲۷ بهمن ماه ۱۳۹۵

10510-NWWCE

بررسی توانایی رگرسیون مقاوم در تخمین اکسیژن خواهی شیمیایی خروجی زلال ساز تصفیه خانه صنعتی

غلامرضا نبی بیدهندی^۱، پرهام پهلوانی^۲، میلاد ابوذری^۳

۱- استاد گروه مهندسی محیط زیست دانشگاه تهران

۲- استادیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی دانشگاه تهران

۳- دانشجوی کارشناسی ارشد مهندسی محیط زیست دانشگاه تهران

milad.abuzari@ut.ac.ir

خلاصه

در راستای افزایش بهبود کیفیت پساب صنعتی و مدیریت بهتر آن‌ها، باید راهکاری ساده و با دقت مناسب برای تخمین فرآیندها ایجاد نمود و از طرفی با توجه به این که فرایندهای تصفیه به صورت سیستم جعبه سیاه^۱ می‌باشند و به دلیل تأثیرپذیری اکثر عوامل دخیل در سیستم و مشکلات جمع آوری داده در مدل‌های فیزیکی لذا استفاده از آمار و روش‌های رگرسیونی می‌تواند راهگشا باشد. لذا هرچه مدل ساده‌تر و با متغیرهای ورودی کمتری باشد مدل مربوطه اهمیت بیشتری خواهد داشت. ورودی مدل پیشنهادی شامل داده‌های خروجی واحد بیولوژیکی و پارامتر خروجی مدل، میزان اکسیژن خواهی شیمیایی^۲ واحد زلال‌ساز می‌باشد. همچنین برای مقایسه کارایی مدل‌ها از ضریب تبیین^۳، میانگین مجموع خطاها^۴ و ضریب همبستگی^۵ استفاده می‌کنیم. تمام تمرکز این تحقیق روی ارائه یک مدل داده‌منا و بهبود آن و سپس مقایسه آن با روش‌های مشابه می‌باشد. این تخمین‌ها برای داده‌های خروجی واحدهاست و داده‌های ورودی و خروجی به صورت داده‌های کیفی می‌باشند. در نهایت یک رابطه خطی با مدل رگرسیون مقاوم^۶ با شاخص‌های $R=0.784727$ ، $MSE=0.089054$ و $R\text{-Square}=0.6096$ ارائه شده است.

کلمات کلیدی: تصفیه خانه پتروشیمی فجر، مدل‌سازی خطی، رگرسیون مقاوم، اکسیژن خواهی شیمیایی

۱. مقدمه

از مزایای عمده روش‌های داده‌منا، عدم نیاز به درک پیشرفته‌ای از فیزیک و ریاضی در استفاده از آن‌ها می‌باشد که این اهم، احتمال خطای محاسباتی یا مفهومی را کاهش می‌دهد [1]. یکی از ابزارهایی که در دهه اخیر به دلیل وفور جمع‌آوری داده‌ها، بسیار مورد استفاده قرار می‌گیرد روش‌های محاسبات هوشی^۷ می‌باشد. اگر ما بتوانیم مدلی کارآمد و مناسب برای تخمین پارامترهای خروجی هر واحد تصفیه‌خانه بیابیم در نتیجه با توجه به انواع فرایندهای فیزیکی و شیمیایی و بیولوژیکی، بهترین تصمیم‌گیری‌های فنی را برای بهره‌برداری بهتر به کار خواهیم برد. در این تحقیق تنها ابزار ما داده‌های گذشته واحد فرایند مربوطه می‌باشد که به با توجه به شرایط واقعی^۸ توسط ابزارهای مربوطه یا روش‌های آزمایشگاهی و یا با ابزار دقیق به صورت ۸ ساعتی و به مدت ۱۳ ماه اندازه‌گیری شده‌اند. در رابطه با پیشینه‌ی تحقیق حاضر مطالعات مشابهی و با روش‌های متفاوت در سال‌های قبل صورت گرفته

- 1- Black Box.
- 2- Chemical Oxygen Demand (COD).
- 3- Coefficient of Determination (R-Square).
- 4- Mean square error (MSE).
- 5- Correlation of Coefficient (R).
- 6- Robust regression.
- 7- Intelligent methods
- 8- Full-Scale

است. یورلین و همکاران (۲۰۱۰) در تحقیق خود با استفاده از MLR و روش کمترین مربعات و رگرسیون تکه‌ای^۱ به تخمین میزان COD در تصفیه‌خانه واقع در ایالت زاگرب کرواسی پرداختند و داده‌های آن‌ها به صورت روزانه جمع‌آوری می‌شد و در نهایت مدل رگرسیونی برای قسمت مربوطه ارائه نمودند [2]. حمید زارع (۲۰۱۴) به بررسی تفاوت میزان تخمین COD و BOD^۲ در تصفیه‌خانه شهری اکباتان تهران به دو روش استفاده از رگرسیون‌های خطی^۳ و شبکه‌های عصبی^۴ پرداخت. همچنین میزان قدرت مدل‌ها را نیز با R و RMSE^۵ اندازه‌گیری نمود. نتایج حاکی از آن بود که قدرت ANN به مراتب از MLR بیشتر می‌باشد [3]. موردی که همه محقق‌ها در صدد آن بوده‌اند و در واقع در تحقیق خود در مورد آن کوتاهی نموده‌اند آن است که در بعضی از متغیرهای ورودی و خروجی واحدهای تصفیه‌خانه، شاهد داده‌های اصطلاحاً پرت هستیم که مقادیرشان با مقادیر دیگر مشاهدات متفاوت است که معمولاً قبل از استفاده‌ی برخی روش‌های آزمون آماری آن‌ها را کنار می‌گذارند در حالی که ممکن است زنگ هشدار برای مدلسازی باشند و نباید اینگونه عمل کرد [4]. در برخورد با داده‌های پرت ابتدا باید در صورت امکان تحقیق کرد که وجود آن‌ها ناشی از خطای اندازه‌گیری یا نوشتاری نباشد و در صورتی که موردی مشاهده نشد ترجیح بر این است نقطه پرت را در نظر گرفت و ممکن است داده‌ی معتبری باشد و در صورتی که امکان تکرار با در نظر گرفتن وقت و هزینه و امکان عمل نباشد به صرف اینکه داده غیرعادی است نباید کنار گذاشته شود. در نتیجه در صورتی که با روش‌های مختلف نرمال سازی نتوان نرمال بودن داده‌های بعضی متغیرها را انجام داد به جای فرایند حذف بهتر است از روش‌های مقاوم به داده‌های پرت استفاده نمود هدف اصلی ما در این تحقیق تخمین داده‌های خروجی واحد زلال‌ساز تصفیه‌خانه فجر با ساده‌ترین و بهترین حالت و با استفاده از شناخت داده‌های ورودی تصفیه‌خانه در هر بخش می‌باشد که به عبارتی مدیریت فرایند را علاوه بر تجربی و آزمایشگاهی که نیازمند صرف وقت و هزینه است به صورت ساده‌ترین حالت آماری بیان می‌کنیم [5]. قابل توجه است که با توجه به نام‌گذاری این واحدها توسط بهره‌بردار و جهت جلوگیری از سردرگمی، واحد زلال‌ساز به آدرس ۶۸۱۳ و واحدهای بیولوژیکی A و B به صورت 6806-A و 6806-B نام‌گذاری شده‌اند.

۲. منطقه مورد مطالعه

شرکت پتروشیمی فجر در سال ۱۳۷۷ ه.ش با هدف تامین مواد مورد نیاز مجتمع‌های منطقه ویژه اقتصادی پتروشیمی بندر امام خمینی به صورت متمرکز احداث گردید. با توجه به اینکه تصفیه پساب مجتمع‌های منطقه ویژه از اقدامات زیست‌محیطی شرکت پتروشیمی فجر در منطقه محسوب می‌گردد. در این بخش فرایند واحد تصفیه پساب ارائه گردیده است. این منطقه در بین عرض جغرافیایی "29° 00' 30" شمالی و طول جغرافیایی "49° 04' 59.8" شرقی قرار دارد. ظرفیت این مجموعه ۴۶۰ متر مکعب بر ثانیه می‌باشد [6]. پساب تصفیه شده در قسمت بیولوژیکی پس از عبور از حوضچه‌های هوادهی از طریق دیواره تیغه‌ای موجود در حوضچه‌ها سرریز کرده و توسط نیروی ثقل به طرف زلال‌ساز ارسال می‌گردد. پساب از طریق یک لاین به مرکز زلال‌ساز وارد می‌شود. در مرکز زلال‌ساز یک توزیع‌کننده بتنی قرار دارد که پساب را از طریق سوراخ‌های خود به صورت مساوی و در جهات مختلف وارد مخزن دایره‌ای شکل زلال‌ساز می‌کند. در حقیقت پساب به گونه‌ای وارد زلال‌ساز می‌شود که باعث برهم زدن آب نخواهد شد. لجن فعال پس از ورود به زلال‌ساز با زمان ماندی که به آن داده می‌شود به آرامی از پساب تصفیه جدا شده و ته‌نشین می‌شود. پساب تصفیه شده نیز از طریق دیواره‌های کنگره‌ای که دورتا دور زلال‌ساز قرار گرفته سرریز کرده و با پساب زلال‌شده زلال‌ساز دیگر ترکیب شده و به مرحله بعد ارسال می‌گردد. خروجی زلال‌ساز از طریق نیروی ثقل به طرف حوضچه کلرزنی جریان می‌یابد. بر سر راه لاین خروجی زلال‌سازها به حوضچه کلرزنی یک لاین کوچک تزریق آب جهت تامین کلر آزاد وجود دارد. آب خروجی از حوضچه کلرزنی که اینک آب تصفیه شده خوانده می‌شود از طریق تیغه سرریز به خروجی هدایت شده و به خور می‌ریزد. از این آب می‌توان جهت آب باغبانی و یا رقیق‌سازی سایر پساب‌های ورودی به واحد نیز استفاده کرد (شکل ۱) [7].

- 1- Particular least square (PLR)
- 2- Biological Oxygen Demand
- 3- Multiple Linear Regression (MLR)
- 4- Artificial Neural Network (ANN)
- 5- Relative Mean Square Error

۳. روش تحقیق

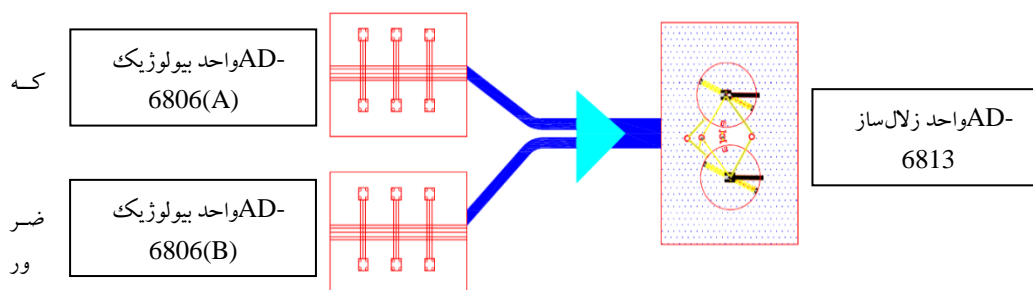
۳.۱. رگرسیون خطی ساده و چندگانه

به طور کل هدف از رگرسیون، پیش بینی یک متغیر مجهول با استفاده از متغیرهای موجود دیگر می باشد. مدل رگرسیونی ساده خطی به صورت زیر است:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

که $i=1,2,\dots,n$ و y_i متغیر وابسته یا متغیر پاسخ و x_i متغیر مستقل می باشد و β_0 عرض از مبدا و β_1 شیب معادله با n مشاهده می باشد. خطای ε_i یک متغیر تصادفی است به عبارتی مقدار آن تحت کنترل تحلیل گر نیست و از تغییرپذیری طبیعی در ذات سیستم نشأت می گیرد. لازم به ذکر است که استفاده از روش های رگرسیونی ساده تنها با استفاده از شرط های زیر می باشد:

- رابطه خطی با X دارد و شکل مدل درست است.
- داده های استفاده شده برای مدل معرف داده های مورد نظر هستند.
- واریانس باقیمانده ها ثابت است و به X یا هر چیز دیگری بستگی ندارد.
- باقی مانده ها مستقل اند.
- باقی مانده ها توزیع مستقل دارند.



شکل ۱: سیستم جداگانه مربوط به ارتباط واحد زلزال ساز و قسمت تصفیه بیولوژیکی

ت برآورد کردن آن ها با هدفی که از به کارگیری مدل رگرسیونی انتظار می رود دنبال می شود. همچنین انتظار است که خط رگرسیونی، به صورت میانگین شرطی، کمابیش مانند میانگین یک نمونه به داده های پرت حساس باشد. یکی از مدل های رگرسیونی مدل حداقل مربعات معمولی^۱ می باشد که بر پایه مینیم کردن اختلاف مشاهدات و برآوردها در هر مرحله می باشد. حال ممکن است در داده ها شرایط به گونه ای باشد که فرض واریانس ثابت و نرمال بودن باقیمانده ها ایفا نشود لذا در این شرایط به دنبال روش های تعمیمی و جایگزین برای برازش تابع به داده ها می رویم که این روش ها شامل روش های غیر پارامتریک رتبه ای یا خطوطی که به جز مربع باقیمانده ها را مینیمم کند می باشد. رگرسیون خطی چندگانه (MLR) بسط رگرسیون خطی ساده^۲ برای چندین متغیر توصیفی می باشد. مدل MLR به صورت زیر نمایش داده می شود:

1- Ordinary Least Square (OLS).
2- Simple Linear Regression (SLR).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

که در آن y متغیر پاسخ، β_0 عرض از مبدا و β_i شیب هر متغیر و ε باقیمانده یا به عبارتی خطای داده‌ها می‌باشد.

۳.۲. رگرسیون مقاوم

یکی دیگر از مدل‌های رگرسیونی موفق رگرسیون مقاوم یا رگرسیون روست می‌باشد که به روش‌های رگرسیون گفته می‌شود که رفتار باثبات و مقاومی در برابر وجود داده‌ی غیر معمول دارند. بعضی از روش‌های معمول رگرسیون مانند کمترین مربعات در صورت صدق فرض‌های آنان به خوبی کار می‌کنند اما در مورد داده‌هایی که از فرض‌های آنان تخلف می‌کنند شاید به خوبی عمل نکنند. به ویژه، روش کمترین مربعات نسبت به داده‌ی پرت حساس است. مسئله‌ی دیگر وجود ناهم‌وارianسی در داده است. روش‌های پارامتری و ناپارامتری مختلفی برای رگرسیون باثبات پیشنهاد شده‌است. به عنوان روش کمترین قدرمطلق معادل باثبات‌تری برای کمترین مربعات می‌باشد [8].

$$s = \sum_{i=1}^n |y_i - f(x_i)| \quad (3)$$

همان‌گونه که در معادله (۳) مشخص می‌باشد به جای توان دوم خطای رگرسیون از قدر مطلق خطا استفاده می‌شود که وجود داده‌ی پرت تاثیر کمتری بر قدر مطلق خطا نسبت به مربع خطا دارد.

۳.۳. رگرسیون ریج^۱

یکی دیگر از روش‌های رگرسیونی رگرسیون ریج می‌باشد. این رگرسیون یک روش حل برای آنالیز داده‌هایی است که از چند بعد دارای شرایط غیرخطی باشند. وقتی که داده‌های چندبعدی^۲ باشند در نتیجه روش حداقل مربعات معمولی نااریب می‌شود و واریانس آن‌ها از مقدار واقعی فاصله می‌گیرد. با افزودن درجه‌ای از بایاس در تقریب رگرسیون، این روش از خطای استاندارد داده‌ها می‌کاهد. این کار باعث ایجاد رابطه بهتر جهت تخمین داده‌ها می‌گردد. همچنین روش‌های مشابه فراوانی وجود دارند که در بین آن‌ها این روش از محبوبیت بیشتری برخوردار است. رابطه رگرسیونی به صورت ماتریسی به صورت معادله زیر می‌باشد [9].

$$Y = XB + e \quad (0)$$

که Y متغیر وابسته و X متغیر مستقل می‌باشد. همچنین B ضریب رگرسیون و e خطای باقیمانده‌هاست. در رگرسیون ریج استاندارد کردن متغیرهای مستقل و وابسته اولین قدم می‌باشد که بعد از آنکه رگرسیون محاسبه شد دوباره مقادیر به حالت اولیه خود تبدیل می‌شوند. در این رگرسیون ابتدا مانند کمترین مربعات به محاسبه B می‌پردازیم:

$$\hat{B} = (X'X)^{-1}X'Y \quad (2)$$

همچنین زمانی که متغیرها استاندارد باشند رابطه زیر برقرار است:

$$X'X=R \quad (3)$$

که R ماتریس آزادی می‌باشد و روابط ماتریس واریانس و کواریانس و امید ریاضی به صورت زیر است:

$$E(\hat{B}) = B \quad (4)$$

$$V(\hat{B}) = \sigma^2 R^{-1} \quad (5)$$

که فرض ما بر این اساس است که $\sigma = 1$. از معادلات بالا به رابطه زیر می‌رسیم:

$$V(\hat{b}_j) = r^{jj} = \frac{1}{1 - R_j^2} \quad (6)$$

1- Ridge Regression Estimates

2- Multicollinearity

که R_j^2 همان میزان R به توان دو یا R-Square رگرسیون می‌باشد. حال روش رگرسیون با اضافه کردن یک مقدار K به ماتریس کرولیشن عمل می‌کند:

$$\bar{B} = (R + KI)^{-1}X'Y \quad (6)$$

که K یک مقدار مثبت کمتر از یک می‌باشد و در حالت معمول مقدار آن ۰.۳ می‌باشد. همچنین مقدار بایاس از رابطه زیر محاسبه می‌شود:

$$E(\bar{B} - B) = [(X'X + KI)^{-1}X'X - I]B \quad (7)$$

همچنین ماتریس کواریانس طبق زیر محاسبه می‌شود:

$$V(\bar{B}) = (X'X + KI)^{-1}X'X(X'X + KI)^{-1} \quad (8)$$

می‌توان اثبات کرد یک مقداری برای k وجود دارد که خطای رگرسیون از خطای روش معمول حداقل مربع کمتر شود. همچنین برای محاسبه K روابط زیادی وجود دارد که یکی از معروف‌ترین آن رابطه‌ی هیر و کینارد^۱ (۱۹۷۶) می‌باشد که به صورت زیر محاسبه می‌شود:

$$K = \frac{\rho s^2}{\bar{B}'\bar{B}} \quad (9)$$

۳.۴. رگرسیون وایت با رویکرد کمترین مربعات^۲

روش رگرسیونی وایت با رویکردی مشابه ریح عمل می‌کند به این صورت که اگر معادله زیر یک معادله رگرسیونی باشد:

$$y = XB + \varepsilon \quad (10)$$

که $E(\varepsilon\varepsilon') = \Phi$ و $E(\varepsilon) = 0$ که یک مقدار مثبت می‌باشد. لذا به جای محاسبه OLS به صورت رابطه $\hat{B} = (X'X)^{-1}X'y$ می‌توان از رابطه زیر استفاده کرد:

$$\text{Var}(\hat{B}) = (X'X)^{-1}X'\Phi X(X'X)^{-1} \quad (11)$$

که اگر خطاها دارای واریانس همسانی باشند رابطه $\Phi = \sigma^2 I$ برقرار می‌شود و معادله بالا ساده می‌گردد. حال با توجه به $e_i = y_i - x_i\hat{B}$ و با توجه به اینکه X_i ردیف i ام از X می‌باشد لذا ماتریس کواریانس حداقل مربعات را طبق رابطه زیر می‌توان تقریب زد:

$$\text{OLSCM} = \frac{\sum e_i^2}{N - K} (X'X)^{-1} \quad (12)$$

که N تعداد نمونه و K تعداد متغیرها می‌باشد. رابطه OLSCM برای بررسی اطمینان از همگنی پراکنش یا همسانی واریانس داده‌های رگرسیون به کار می‌رود که در صورت صحیح بودن طبق روابط بالا و ساده شدن محاسبه $\hat{\Phi}$ به محاسبه رگرسیون می‌پردازیم. در اغلب مواقع این اتفاق نمی‌افتد و باید از ماتریس HCCM^۳ استفاده کنیم. ایده‌ای که پشت این رابطه نهفته است، استفاده از e_i^2 برای تخمین Φ می‌باشد که برابر رابطه $\hat{\Phi} = \text{diag}[e_i^2]$ می‌باشد که در نهایت معادله زیر بدست می‌آید:

$$\text{HCO} = (X'X)^{-1}X'\hat{\Phi}X(X'X)^{-1} = (X'X)^{-1}X'\text{diag}[e_i^2]X(X'X)^{-1} \quad (13)$$

که این معادله شایع‌ترین معادله HCCM می‌باشد و همانطور که وایت و همکاران در سال ۱۹۸۰ بیان کردند، HCO برای مواقعی که همسانی واریانس \hat{B} نامشخص باشد کاربرد دارد. همچنین یکسری اصلاحات در زمان‌های بعد به این معادله تحت عنوان HC1, HC2 و HC3 توسط دیگر دانشمندان اعمال شد که در این تحقیق نمی‌گنجد [10].

۴. نتایج و تحلیل آن‌ها

قبل از نرمال سازی داده‌ها به آنالیز اولیه داده‌های خام می‌پردازیم که با احتساب ۱۲ متغیر ورودی و خروجی شامل نتایج جدول است:

- 1- Hoerl and Kinnard
- 2- White's adjusted heteroscedastic consistent Least-squares Regression
- 3- Heteroscedasticity Consistent Covariance Matrix



کنگره علوم و مهندسی آب و فاضلاب ایران

دانشگاه تهران، تهران

۲۶ و ۲۷ بهمن ماه ۱۳۹۵

جدول ۱: توزیع آماری متغیرهای موجود در خروجی واحد بیولوژیکی و زلال ساز

متغیر	میانگین	میانه	مد	مینیم	ماکزیم	واریانس	انحراف از معیار
COD (6813)	۱۰۰.۹۵	۹۰	۵۰	۳۸	۲۹۷	۲۰۹۴.۱۲	۴۵.۷۶۱۵
COD (6806 A)	۱۳۵.۶۳	۱۲۱.۵	۵۰	۴۹	۸۲۵	۵۵۷۴.۷۳	۷۴.۶۶
N (6806 A)	۷.۴۳	۷.۲	-	۰.۴	۲۳.۲	۱۶.۳۲	۴.۰۴
P (6806 A)	۱.۶۶۵۹	۱.۲	۱.۱	۰.۱	۱۵۰	۲۹.۱۹	۵.۴
PH (6806 A)	۷.۲	۷.۳	۷.۴	۳.۹	۸.۸	۰.۳۳	۰.۵۷۷
SST (6806 A)	۲۷۱.۲۱	۲۵۰	-	۵۰	۸۵۰	۱۶۷۸۷	۱۲۹.۵۶۷
COD (6806 B)	۱۴۲.۸۶	۱۲۶.۵	۵۰	۴۳	۸۱۴	۶۹۹۷.۶۶	۸۳.۶۵
N (6806 B)	۸.۱۶۷	۷.۲	۸.۳	۰.۴	۴۶۰	۲۸۸	۱۶.۹۷
P (6806 B)	۱.۵۹۹۴	۱.۲	-	۰.۱	۱۴۳	۲۶.۵۲	۵.۱۴۹۸
PH (6806 B)	۷.۱۹	۷.۳	۷.۴	۳.۹	۸.۵	۰.۳۴	۰.۵۸۳۳
SST (6806 B)	۲۸۷.۰۸	۲۶۰	۲۰۰	۵۰	۷۶۰	۲۰۳۰۵	۱۴۲.۴۹۶

قبل از استاندارد سازی، داده‌های خروجی را یک گام یعنی ۸ ساعت گام می‌دهیم و بدین صورت روش ارائه شده بدین صورت است که با توجه به ویژگی سیال و عدم یک سیستم صفر و یکی برای ورودی و خروجی، داده‌های چند ساعت قبل ورودی هم لزوماً در خروجی مدل تاثیر گذار می‌باشد. لذا بعد از اضافه کردن متغیرهای جدید که همان گام‌های متغیرهای قدیمی می‌باشند و با استفاده از همان متغیرهای قبلی، مجموعاً ۲۰ متغیر خواهیم داشت که بعد از محاسبه ماتریس کرولیشن آن‌ها و با در نظر گرفتن روابط متغیرها و خواص هر یک از آن‌ها و ویژگی‌هایی که دارند ۱۱ متغیر از ۲۰ متغیر خود را که وجود آن‌ها به عنوان متغیر ورودی را حذف کردیم و با کاهش متغیرها در مرحله اول، نتایج را بررسی می‌کنیم. در نهایت متغیرهایی که میزان کرولیشن بالایی با هم هستند را از ورودی مدل آماری کاهش می‌دهیم تا مدل، ساده‌تر گردد (کمک در جهت جلوگیری از اندازه‌گیری داده‌های غیر ضروری و کاهش هزینه). در نهایت ما دارای ۹ متغیر ورودی و ۲ متغیر خروجی مطابق جدول ۲ هستیم که باید استاندارد شوند.

جدول ۲: متغیرها و پارامترهای مورد استفاده در گزینش نهایی برای مدل

COD (6806A)	متغیرهای ورودی مدل
N(6806A)	
P(6806A)	
PH(6806A)	
SST(6806A)	
COD with Lag (6806A)	
N with Lag (6806A)	
N(6806B)	
N with Lag (6806B)	
COD(6813)	متغیرهای خروجی مدل

بعد از انجام مراحل بالا با استفاده از روش نرمال سازی مطابق فرمول ۱۸ عمل می‌کنیم.

$$Z = 2 \times \frac{x_i - \min(X)}{\text{Max}(X) - \text{Min}(X)} - 1 \quad (18)$$

در معادله ۱۸ X_i بیانگر هر کدام از داده‌ها، X بیانگر ماتریس داده‌ها و Z معادل ماتریس داده‌های نرمال شده می‌باشد. بعد از انجام پیش‌پردازش‌های مذکور برای ارزیابی مدل‌سازی‌های انجام شده، بعد از آنکه داده‌ها به بخش‌های آموزشی و اعتبارسنجی و تست یا آزمایشی تقسیم شد، از داده‌های تست جهت ارزیابی مدل‌های پیاده شده با استفاده از شاخص‌های ذکر شده در روابط زیر اقدام می‌کنیم

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (y_i - y'_i)^2}{\sum_{i=1}^N y_i^2} \right) \quad (19)$$

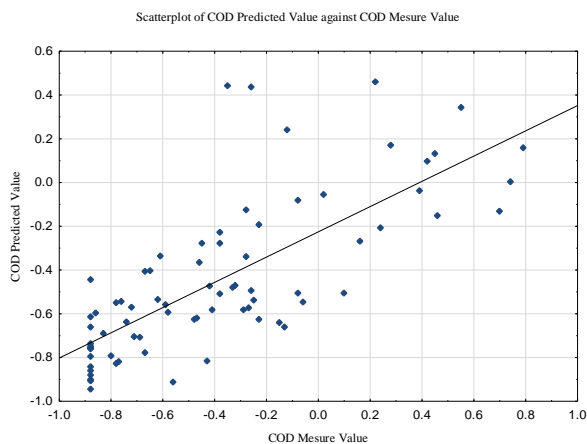
$$MSE = \frac{1}{N} \sum_{i=1}^N (y - y')^2 \quad (20)$$

$$R = \frac{S_{XY}}{S_X S_Y} \quad (21)$$

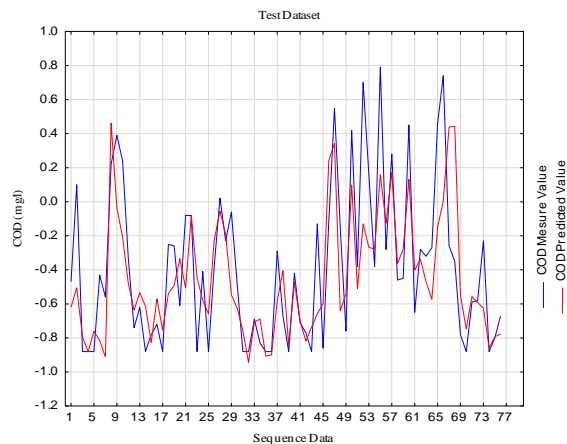
که در معادلات بالا y' نشان‌گر خروجی مدل و y نشان‌گر مشاهدات اندازه‌گیری می‌باشد. در نهایت همه محاسبات در برنامه MATLAB 2016-a و با شرایط مشابه پیاده شده‌اند. حال خروجی مدل ما شامل جدول ۳ می‌باشد.

جدول ۳: مقادیر شاخص‌های مدل ایجاد شده در تخمین COD واحد زلال ساز

ردیف	نوع مدل	R^2	MSE	R
۱	رگرسیون مقاوم	۰.۶۰۹۶	۰.۰۸۹۰۵۴	۰.۷۸۴۷۲۷
۲	رگرسیون ریج	۰.۵۷۳۴	۰.۰۹۰۰۵۰	۰.۷۵
۳	رگرسیون وایت با رویکرد کمترین مربعات	۰.۵۶۸۲	۰.۰۹۰۱۴	۰.۷۵۴۰۱۳
۴	روش کمترین مربعات معمولی	۰.۵۶۲۶	۰.۰۹۰۲۰۴	۰.۷۵۱۲۳۳



(ب)



(الف)

شکل ۲۰: شکل مربوط به قسمت (الف): مقادیر COD تخمینی و مشاهداتی برای داده‌های تست و (ب): نمودار خطی رابطه بین COD مشاهداتی و تخمینی مربوط به مدل رگرسیونی مقاوم

با توجه به اینکه مدل رگرسیون مقاوم با میزان کارایی $R=0.784727$ بهترین تقریب را در خصوص میزان COD واحد زلال‌ساز تصفیه‌خانه پتروشیمی فجر ارائه نمود (شکل ۲) لذا رابطه پیشنهادی این تحقیق با توجه به پارامترهای کیفی خروجی قسمت بیولوژیکی و با توجه به روابط ۴ الی ۱۳ به صورت زیر می‌باشد.



شرکت مهندسی آب و فاضلاب کشور

کنگره علوم و مهندسی آب و فاضلاب ایران

دانشگاه تهران، تهران

۲۶ و ۲۷ بهمن ماه ۱۳۹۵



$$\begin{aligned} \text{COD}_{6813(t)} = & (0.20439 \times \text{COD}_{6806A(t-8hr)}) + (0.036955 \times N_{6806A(t-8hr)}) \\ & + (0.012338 \times P_{6806A(t-8hr)}) - (0.13179 \times \text{PH}_{6806A(t-8hr)}) \\ & + (0.060349 \times \text{SST}_{6806A(t-8hr)}) \\ & - (0.080425 \times \text{COD}_{6806A(t)}) + (0.665653 \times N_{6806A(t)}) \\ & - (0.060715 \times N_{6806B(t-8hr)}) + (0.117426 \times N_{6806B(t)}) \end{aligned} \quad (22)$$

در معادله (۲۲)، t نشان گر زمان حاضر و t-8hr بیان گر زمان مربوط به 8 ساعت قبل می باشد

۵. نتیجه گیری

در این تحقیق با استفاده از چهار مدل رگرسیونی خطی معروف، به بررسی واحد زلال ساز تصفیه خانه صنعتی پرداختیم که در نهایت مدل رگرسیونی مقاوم در مقایسه با دیگر روش های رگرسیونی موجود، COD را تا میزان $R=0.748$ تخمین نمود که یکی از دلایل موفقیت این مدل، ساختار ریاضی آن در مواجهه با داده های شوک بود که از آن می توان به عنوان یک رابطه خطی چند گانه در تخمین میزان COD واحد زلال ساز استفاده نمود. پس در نهایت به این نتیجه رسیدیم که معادله هایی همچون معادله (۲۲) در تصفیه خانه ها می تواند جایگزین روش های مدل فیزیکی و ریاضی باشند و از هزینه زمان و اندازه گیری و آزمایش بکاهند و زیربنایی برای مکانیزه کردن واحدهای تصفیه پساب در آینده باشند که استفاده از آن ها می تواند در جهت بهینه نمودن فرایندهای موجود و حتی مصرف کمترین میزان برق و تولید حداکثر گاز متان کمک کند.

۶. مراجع

1. Xiupeng, Wei., Andrew, Kusiak., (2015). Short-term prediction of influent flow in wastewater treatment plant. DOI 10.1007/s00477-014-0889-0. Stoch Environ Res Risk Assess (2015) 29:241–249.
2. M. ÈURLIN. (2008). Modeliranje procesa biološke obrade otpadne vode, Kem. Ind. 57 (2) 59–67.
3. Hamid Zare Abyaneh., (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J Environ Health Sci Eng. 2014 Jan 23. doi: 10.1186/2052-336X-12-40.
4. Dennis, R., Helsel, & Ronert, M., Hirsch., (1992). Statistical methods in water resource. (U.S.G.S.).
5. حسنلو، ن، مهرداد، ح، نایب، ف، گلپایایی. (۱۳۹۱). استفاده از روش تحلیل عاملی در مدل سازی عصبی واحد تصفیه پساب با نمک پایین تصفیه خانه فجر، هفتمین همایش ملی و نمایشگاه تخصصی مهندسی محیط زیست، تهران، دانشکده محیط زیست.
6. شریعت زاده، م. ۱۳۸۸، سیمای زیست محیطی پتروشیمی فجر.
7. مردانی، ن. ۱۳۸۸، معرفی فرایند تصفیه پساب تصفیه خانه فجر، نشریه داخلی پتروشیمی فجر.
8. James, P., LeSage., (1998). Spatial Econometrics. Department of Economics University of Toledo.
9. Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. Annals of Statistics, 32 (2): 409–499.
10. Badi, H., Baltagi., (2002). Econometrics. (3rd Edition). Springer.