

Data-Driven Fault Diagnosis in Gas Refining Processes: A Case Study Using Support Vector Regression and Gaussian Process Regression

Seyed Hossein Ahmadi¹, Mohammad Javad Khosrowjerdi^{*2}

1- Ph. D. Student of Control Engineering, Sahand University of Technology

2- Professor of Control Engineering, Sahand University of Technology

Abstract

Rapid fault diagnosis in complex processes of gas refineries is a key challenge in process engineering and control, due to the strategic importance of this equipment and the high maintenance costs. Despite the vast amount of data generated in these processes, the effective exploitation of this data has been limited so far. In data-driven fault detection methods, two main approaches exist: fault detection and classification, and fault diagnosis and fault measurement, with the latter being achievable through regression techniques or accurate models. This paper presents a machine learning-based approach for the regression modeling of industrial systems with outlier and noisy data, tailored for fault diagnosis applications. For fault diagnosis, two regression methods, including Support Vector Machine Regression and Gaussian Process Regression, have been utilized. These models are capable of predicting the behavior of industrial systems and diagnosis faults without the need for precise physical models. The main innovation of this paper is the introduction of a machine learning framework that can perform more accurately and rapidly compared to traditional expert fault diagnosis methods, particularly in scenarios with outlier and noisy data.

Keywords:

Data Driven Fault Diagnosis, Machine Learning, Support Vector Machine Regression, Gaussian Process Regression, Distributed Control Systems, Gas Refining Processes.

*Corresponding author: Control Engineering Professor, Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. **Tel:** +98 41 3345 9369, **E-mail:** *khosrowjerdi@sut.ac.ir*

تشخیص عیب داده‌محور در فرایند پالایش گاز: مطالعه موردی با استفاده از رگرسیون ماشین بردار پشتیبان و فرایند گاوسی

سید حسین احمدی^۱، محمدجواد خسروجردی^{۲*}

۱- دانشجوی دکتری مهندسی برق گرایش کنترل، دانشکده مهندسی برق، دانشگاه صنعتی سهند sh_ahmadi98@sut.ac.ir

۲- استاد تمام مهندسی برق گرایش کنترل، دانشکده مهندسی برق، دانشگاه صنعتی سهند، تبریز، ایران khosrowjerdi@sut.ac.ir

چکیده

تشخیص سریع عیب در فرایندهای پیچیده پالایشگاه‌های گازی به دلیل اهمیت استراتژیک این تجهیزات و هزینه‌های بالای نگهداری آن‌ها، یکی از چالش‌های کلیدی در مهندسی و کنترل فرایندها است. علی‌رغم تولید حجم وسیعی از داده‌ها در این فرایندها، بهره‌برداری مؤثر از آن‌ها تاکنون محدود بوده است. در روش‌های تشخیص عیب داده‌محور دو رویکرد آشکارسازی و کلاس‌بندی عیب و تشخیص و اندازه‌گیری دقیق عیب وجود دارد که رویکرد دوم توسط روش‌های رگرسیون و یا مدل‌های دقیق قابل‌انجام است. این مقاله، روشی مبتنی بر یادگیری ماشین برای مدل‌سازی رگرسیونی سیستم‌های صنعتی با داده‌های پرت و با نوفه در جهت کاربردهای تشخیص عیب ارائه می‌دهد. برای تشخیص و مکان‌یابی عیب‌ها، دو روش رگرسیونی، شامل رگرسیون ماشین بردار پشتیبان و رگرسیون فرایند گاوسی، به کار گرفته شده‌اند. این مدل‌ها قادر به پیش‌بینی رفتار سیستم‌های صنعتی و شناسایی عیوب بدون نیاز به مدل‌های فیزیکی دقیق هستند. نوآوری اصلی این مقاله، معرفی یک چارچوب یادگیری ماشین است که می‌تواند در مقایسه با روش‌های سنتی تشخیص عیب توسط متخصص با تجربه در شرایط داده‌های ناقص و نوفه‌ای به طور دقیق‌تر و سریع‌تری عمل کند.

واژه‌های کلیدی:

تشخیص عیب داده‌محور، یادگیری ماشین، رگرسیون ماشین بردار پشتیبان، رگرسیون فرایند گاوسی، سیستم‌های کنترل گسترده، فرایندهای پالایش گاز.

* محمد جواد خسروجردی، استاد تمام مهندسی کنترل، دانشکده مهندسی برق، دانشگاه صنعتی سهند، تبریز، ایران khosrowjerdi@sut.ac.ir

پالایشگاه‌های گازی با استفاده از سیستم‌های کنترل توزیع‌شده^۱، فرایندهای خود را در اتاق‌های کنترل توسط مهندسان، مدیریت و نظارت می‌کنند. با این حال، تشخیص عیب در این سیستم‌ها کمتر مورد توجه قرار گرفته است، در حالی که روش‌های داده‌محور می‌توانند جایگزینی سریع‌تر و مؤثرتر برای روش‌های دستی باشند. اهمیت تشخیص عیب در این حوزه به کاهش هدررفت انرژی، آلودگی محیط‌زیست و هزینه‌های نگهداری مرتبط است. روش‌های سنتی به نیروی انسانی با تجربه نیاز دارند که همیشه در دسترس نیست و دانش آن‌ها غالباً محدود به شرایط خاص است (Donald R. Woods, 2021; Norman P. Lieberman - *Troubleshooting Process Plant Control_ A Practical Guide to Avoiding and Correcting Mistakes*-John Wiley & Sons (2024), n.d.) مدل بیس (Akbarpour & Khosrowjerdi, 2024; Habibi et al., 2019; M. J. Khosrowjerdi & Mahdi Aliyari- Shoorehdeli, 2014) و داده‌محور (Chen et al., 2023) و یا ترکیبی از این دو (Tidriri et al., 2016) انجام می‌گیرد. در روش‌های داده‌محور که با ظهور روش‌های جدید یادگیری ماشین بسیار مورد توجه قرار گرفته است از روش‌های کلاس‌بندی عیب برای آشکار سازی و محل یابی عیب و یا تشخیص عیب با استفاده از مدل‌های رگرسیونی (Sarmast Alizadeh & Esfandiari, 2024) جهت تعیین دقیق مقدار عیب استفاده می‌شود. با استفاده از روش‌های کلاس‌بندی عیب، می‌توان جدولی از ویژگی‌ها یا پارامترهای مؤثر در ایجاد هر عیب تشکیل داد و هر عیب را به‌عنوان یک کلاس مجزا تعریف کرد. سپس سیستم را آموزش داد تا در صورت وقوع مجدد آن عیب، به‌طور خودکار آن را تشخیص دهد. این رویکرد احتمال خرابی را کاهش داده و قابلیت اطمینان سیستم را افزایش می‌دهد (Ahmadi & Khosrowjerdi, 2021). در روش‌های تشخیص عیب، علاوه بر آشکار سازی وجود عیب، می‌توان مقدار دقیق عیب را نیز اندازه‌گیری کرد. این امر امکان محاسبه میزان تغییرات لازم برای جبران عیب را فراهم می‌کند. البته، برای دستیابی به این هدف، نیاز به یک مدل دقیق از فرایند واقعی یا سیستم مورد نظر است.

در این مقاله، دو روش اصلی یادگیری ماشین، یعنی رگرسیون ماشین بردار پشتیبان^۲ و رگرسیون فرایند گاوسی^۳ برای مدل سازی رگرسیونی جهت تشخیص عیب معرفی شده‌اند. این روش‌ها به‌ویژه در سیستم‌های پیچیده‌ای که نمی‌توان مدل دقیق فیزیکی برای آن‌ها به دست آورد، مفید هستند (Alzghoul et al., 2014; Cheliotis et al., 2020; Yan et al., 2013; Zhao et al., 2013; al., 2023). چالش‌هایی مانند حضور نوفه (Clavijo et al., 2019)، داده‌های پرت^۴ و پراکنده^۵ و انتخاب ویژگی‌های مناسب (Chandrashekar & Sahin, 2014) نیز در این مقاله مورد بررسی قرار گرفته و راهکارهایی برای مقابله با این مشکلات ارائه شده است.

باتوجه به اینکه تاکنون در تشخیص عیب در سیستم‌های کنترل گسترده پالایشگاه‌ها تحقیقات کاربردی زیادی انجام نشده است، این مقاله تلاش دارد با فراهم نمودن زمینه خودکار سازی فرایند تشخیص عیب و ایجاد یک پایگاه داده

¹ Distributed Control System (DCS)

² Support Vector Machine Regression (SVR)

³ Gaussian Process Regression (GPR)

⁴ Outliers

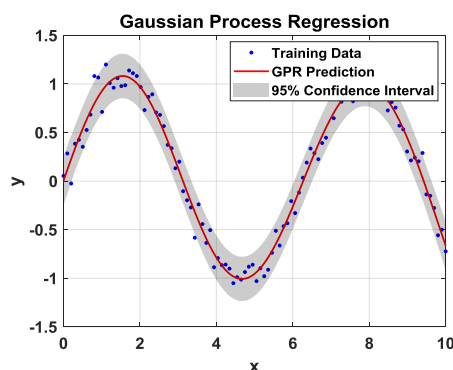
⁵ Sparse Data

هوشمند، این محدودیت‌ها را برطرف کند و با تحلیل داده‌های واقعی پالایشگاه‌ها، راهکارهای عملی ارائه دهد. نوآوری اصلی این پژوهش در توسعه و پیاده‌سازی یک سیستم هوشمند مبتنی بر یادگیری ماشین برای تشخیص عیب است که قادر است با جمع‌آوری داده‌های جامع از پالایشگاه‌ها و تجزیه و تحلیل آن‌ها، عیوب را سریع‌تر و دقیق‌تر شناسایی کند. در این مقاله، ابتدا روش‌های یادگیری ماشین معرفی می‌شوند و سپس مدل‌سازی رگرسیونی سیستم‌های خطی در چهار سطح مختلف بررسی می‌شود. نتایج نشان‌دهنده دقت بالای این مدل‌سازی‌ها است. چالش‌های مختلف مانند نوفه و داده‌های پرت تحلیل شده و راهکارهایی برای مقابله با آن‌ها ارائه می‌شود. در بخش سوم، مطالعات موردی در یک پالایشگاه واقعی بررسی و اثربخشی روش‌های پیشنهادی ارزیابی می‌شود. در نهایت، نتیجه‌گیری و پیشنهاداتی برای پژوهش‌های آینده ارائه شده است.

۲- معرفی روش‌ها

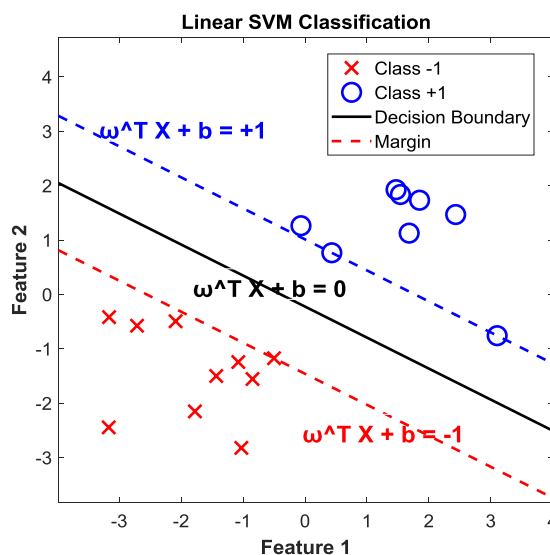
۲-۱- رگرسیون فرایند گاوسی و رگرسیون ماشین بردار پشتیبان

در یک مسئله عمومی رگرسیون، هدف این است که با استفاده از تعدادی نقطه مشاهده‌شده، تابعی پیدا کنیم که بهترین توصیف را از این نقاط ارائه دهد. سپس از این تابع برای پیش‌بینی نقاطی که قبلاً مشاهده نشده‌اند استفاده می‌شود. هرچند برای مجموعه‌ای متناهی از نقاط در یک فضای پیوسته، بی‌شمار تابع وجود دارد که از این نقاط عبور می‌کند. رگرسیون فرایند گاوسی تلاش می‌کند تا توزیع احتمالی روی این توابع بی‌نهایت ممکن را به دست آورد و با استفاده از این توزیع، پیش‌بینی مقادیر نقاط دیده نشده را انجام دهد. توزیع احتمال مذکور روی توابع، در واقع یک فرایند گاوسی است که با یک تابع میانگین و کوواریانس اولیه شروع می‌شود و در هر مرحله با اضافه شدن یک نقطه مشاهده‌شده جدید، به‌روزرسانی می‌شود. از آنجاکه خروجی این مدل توزیعی روی تمام توابع ممکن است، می‌توان میانگین این توابع را به‌عنوان پیش‌بینی مدل و واریانس آن‌ها را به‌عنوان معیاری از اطمینان پیش‌بینی انجام‌شده در نظر گرفت. در شکل ۱، مدل‌سازی یک فرایند با روش GPR نشان داده شده است.



شکل ۱. مدل‌سازی با روش GPR
Figure 1. Modeling with GPR Method.

ماشین بردار پشتیبان^۶ یکی از روش‌های یادگیری با نظارت است که از آن هم برای طبقه‌بندی و هم رگرسیون استفاده می‌شود. مبنای دسته‌بندی‌کننده SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم ابر صفحه‌ای را انتخاب کنیم که حاشیه اطمینان بیشتری برای کلاس‌بندی و جداسازی داده‌ها داشته باشد. پیدا کردن خط بهینه برای جداسازی داده‌ها به وسیله روش‌های برنامه‌سازی غیرخطی صورت می‌گیرد. برای مثال یک مسئله طبقه‌بندی دو کلاسه با خروجی $y = \{-1, 1\}$ در نظر بگیرید که در آن دو کلاس متفاوت از داده‌ها به صورت خطی قابلیت جداسازی دارند. بدیهی است که به تعداد نامتناهی مرز تصمیم‌گیری خطی برای این مسئله وجود دارد. در مسائل با ابعاد و تعداد ویژگی بالاتر مرز تصمیم‌گیری، یک ابر صفحه خواهد بود. مرز تصمیم‌گیری مناسب‌تر است که تا جایی که می‌شود از داده‌های دو کلاس دور باشد. از آنجاکه مرز کلاس‌ها، خطی فرض شده است، تابع آن به صورت $f(\mathbf{X}) = \boldsymbol{\omega}^T \mathbf{X} + b = 0$ تعریف می‌شود، که $\boldsymbol{\omega}$ و b پارامترهای خط و یا صفحه می‌باشند این ابر صفحه یک طبقه‌بندی‌کننده خطی است و $\text{sign}(f(\mathbf{X}))$ برچسب کلاس هر داده جدید \mathbf{X} را در مرحله آزمون تعیین خواهد کرد. در واقع الگوهای با برچسب مثبت یک می‌بایست $f(\mathbf{X})$ مثبت و الگوهای با برچسب کلاس منفی یک باید دارای $f(\mathbf{X})$ منفی باشند. فاصله نزدیکترین داده هر کلاس تا ابر صفحه را می‌توان فاصله داده‌های آن کلاس تا مرز جداسازی در نظر گرفت. فاصله هر نقطه دلخواه تا ابر صفحه مورد نظر برابر با $\frac{|\boldsymbol{\omega}^T \mathbf{X} + b|}{\|\boldsymbol{\omega}\|}$ می‌باشد. واضح است که ابر صفحه می‌تواند طوری قرار گیرد که نزدیک‌ترین داده هر یک از دو کلاس تا آن برابر $\frac{a}{\|\boldsymbol{\omega}\|}$ باشد به طوری که دو ابر صفحه $\boldsymbol{\omega}^T \mathbf{X} + b = \pm a$ در دو سوی ابر صفحه مرزی قرار گیرند. این مفاهیم در شکل ۲، نشان داده شده است.



شکل ۲: طبقه‌بندی‌کننده ماشین‌های بردار پشتیبان با استفاده از ماکزیمم حاشیه بین کلاس‌ها

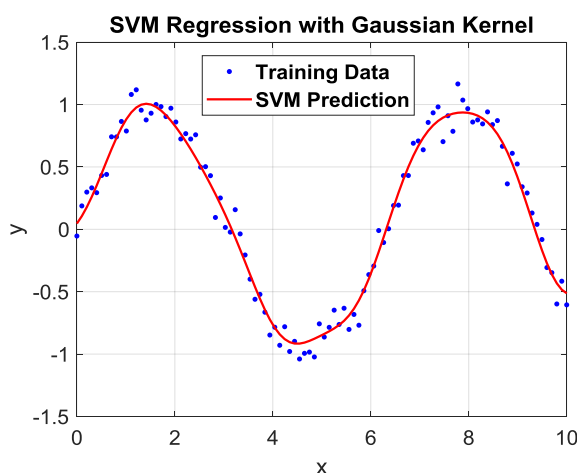
Figure 2. Support Vector Machine Classifier Using Maximum Margin Between Classes

با اعمال ضربی به کل رابطه ابر صفحه‌ها تغییری در آن ایجاد نخواهد شد، لذا می‌توان ابر صفحه را نسبت به a نرمال کرد و یا به عبارتی پارامتر $a=1$ قرار داده شود. دو ابر صفحه کناری محدوده امنی به طول $\frac{2}{\|\boldsymbol{\omega}\|}$ برای ابر صفحه

⁶ Support Vector Machine (SVM)

مرزی $f(X) = 0$ به وجود می‌آورند. به این محدوده حاشیه گفته می‌شود. هر چه این حاشیه بیشتر باشد جداسازی بهتری صورت گرفته است بنابراین باید حاشیه $\frac{2}{\|\omega\|}$ را بیشینه کرد. با حل این مسئله بهینه سازی خط جدا سازی پیدا خواهد شد که فاصله حاشیه ای ماکزیمی نیز نسبت به نزدیکترین داده های هر کلاس داشته باشد.

در حالت رگرسیون ماشین بردار پشتیبان^۷ برعکس حالت کلاس بندی، می‌بایست دنبال یک کانال با حاشیه مناسب بود که تمام نمونه‌ها در آن قرار بگیرند. در شکل ۳، مدل سازی رگرسیونی با روش SVM که به آن SVR نیز گفته می‌شود آورده شده است، هدف جا دادن داده های جدید در یک کانال با کمترین حاشیه ممکن است که توسط داده های گذشته بوجود آمده است. در حالت رگرسیون نیز مانند حالت کلاس بندی می‌توان از انواع کرنل های مختلف مانند کرنل گاوسی^۸ یا مربعی^۹ و مکعبی^{۱۰} برای روش ماشین بردار پشتیبان استفاده نمود.



شکل ۳. مدل سازی رگرسیونی با روش ماشین بردار پشتیبان
Figure 3. Regression Modeling with SVM method (SVR)

۲-۲- مدل سازی رگرسیونی یک سیستم مرتبه اول

یک سیستم ساده مرتبه اول به سیستمی اطلاق می‌شود که با یک معادله دیفرانسیل خطی توصیف می‌شود و در آن تنها خروجی و مشتق مرتبه اول آن (با توان یک) ظاهر می‌شوند. به عنوان مثال، در یک سیستم هیدرولیکی، اگر ارتفاع آب در مخزن به عنوان خروجی و دبی ورودی آب به عنوان ورودی سیستم در نظر گرفته شود، این سیستم به عنوان یک سیستم مرتبه اول مدل سازی می‌شود. یک سیستم مرتبه اول با پاسخ $y = 2e^{-0.5x}$ را در نظر بگیرید که برای x بین صفر تا ۵ و با فواصل نیم، آموزش داده شده و سپس خروجی توسط مدل سازی رگرسیونی برای 0 تا 5 با فواصل یکصدم در شکل ۴ به دست آمده است. در شکل ۴، مشاهده می‌شود که مدل GPR به طور دقیق بر معادله منطبق شده است و روش های SVM, Deep Learning کمی خطا دارند. در مرحله بعد فاصله نمونه‌ها برای آموزش از نیم به یکدهم کاهش یافته است. در شکل ۵، روش GPR و روش های SVM و Deep Learning نشان داده شده است که خطای مدل GPR نسبت

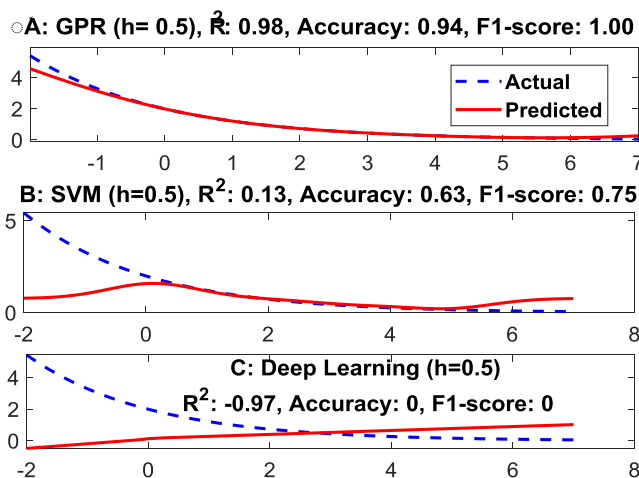
⁷ Support Vector Machine Regression (SVR)

⁸ Gaussian Support Vector Machine

⁹ Quadratic Support Vector Machine

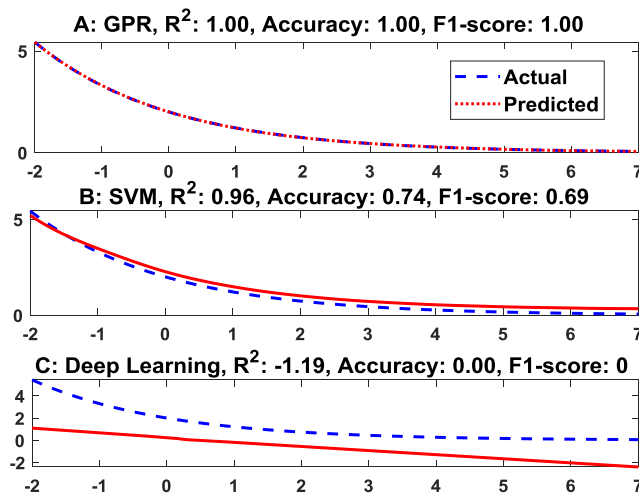
¹⁰ Cubic Support Vector Machine

به دو روش دیگر کمتر می باشد. در هر شکل پارامتر R^2 که نشان دهنده میزان تطابق مدل با واقعیت یا معادله می باشد آورده شده است میزان یک تطابق کامل می باشد. برای قدم بعدی حضور نوفه و داده‌های پرت نیز بر مدل سازی رگرسیونی بررسی می شود. نوفه با دامنه های یکدهم و نیم در دو سطح به صورت اعداد تصادفی به نمونه‌های تحت آزمون اضافه می شود که مدل تقریباً دقیقاً با حضور نوفه به دست می آید شکل ۶ الف، در مرحله بعد چندین داده پرت به داده‌های آموزشی اضافه می شود و مدل به دست می آید. شکل ۶ ب.



شکل ۴. مدل سازی با روش GPR, SVR, Deep Learning^{۱۱} با فواصل بین داده های آموزشی برابر نیم

Figure 4. First-order system model with training data spacing $h=0.5$ GPR, SVR, Deep Learning Method



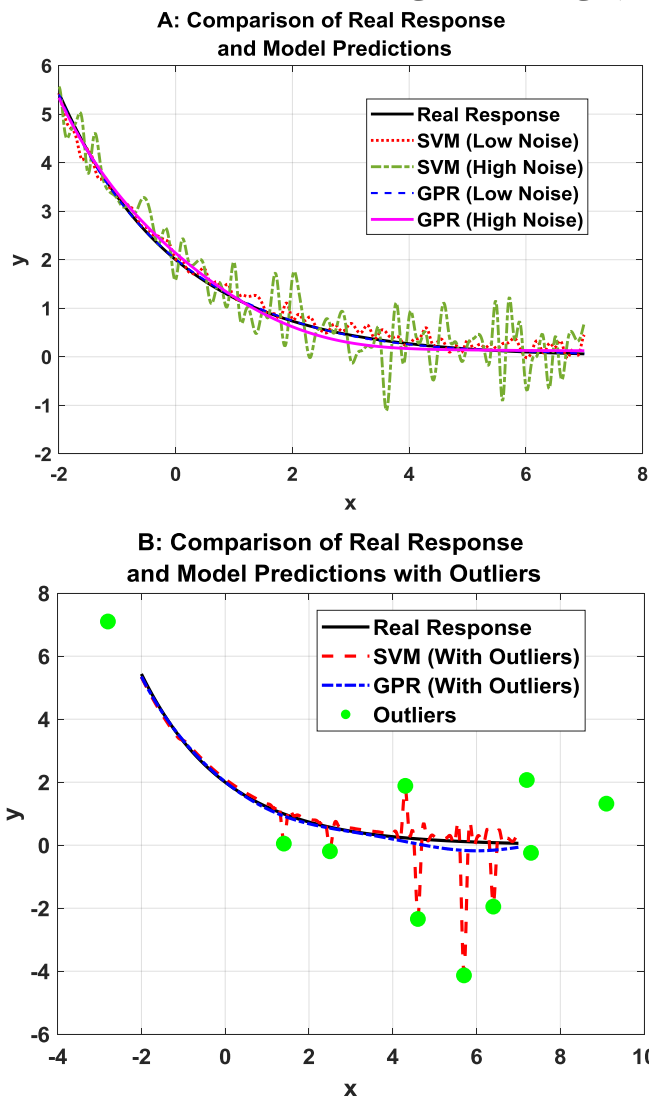
شکل ۵. مدل سازی با روش GPR, SVR, Deep Learning با فواصل بین داده های آموزشی برابر یک دهم

Figure 5. First-order system model with training data spacing $h=0.1$ GPR, SVR, Deep Learning

اگرچه روش‌های متعددی برای حذف داده‌های پرت و کاهش نوفه در مقالات موجود است، نتایج نشان می‌دهند که اثر نوفه بر مدل کم است و این داده‌های پرت هستند که بیشترین تأثیر را بر دقت مدل سازی دارند. با توجه به این که تنها

¹¹Deep Neural Network (Matlab: Fully connected layer = 10)

داده‌های آموزشی پرت که اختلاف زیادی با داده‌های قبلی دارند می‌توانند بر کیفیت مدل تأثیرگذار باشند، با استفاده از یک روش تفاضل‌گیری ساده می‌توان این داده‌ها را شناسایی و حذف کرد. در ادامه، مدل‌سازی رگرسیونی یک سیستم مرتبه دوم نوسانی (زیر میرا) انجام می‌شود و خروجی مدل با حضور نوفه و داده‌های پرت نیز بررسی خواهد شد.



شکل ۶. (الف) مدل سیستم مرتبه اول با روش‌های SVM و GPR در حضور نوفه در دو حالت کم و زیاد (ب) مدل سیستم مرتبه اول در حضور داده‌های پرت

Figure 6. (A) First-order system modeling using GPR, SVM methods in the presence of noise (B) First-order system modeling in the presence of outlier data

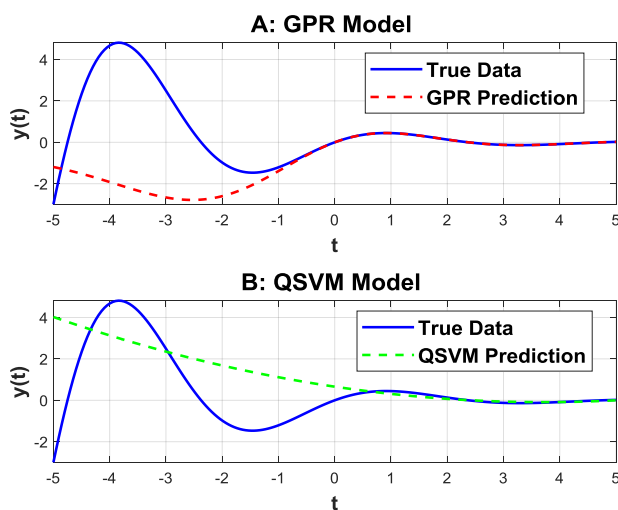
۲-۳- مدل‌سازی رگرسیونی سیستم خطی مرتبه دوم

یک سیستم مرتبه دوم نوسانی (زیر میرا) با معادله (۱) در نظر بگیرید که پاسخ آن نیز در رابطه (۲) آورده شده است.

$$y'' + y' + 2y = 0, \{ y(0) = 0, y'(0) = 1 \} \quad (1)$$

$$y(x) = 2e^{-t/2} \sin(0.5(\sqrt{7}t))/\sqrt{7} \quad (2)$$

مدل‌سازی رگرسیونی این معادلات با متلب در شکل‌های ۷ الف و ۷ ب، با دو روش $QSV\text{M}$ ^{۱۲}، GPR انجام شده است.



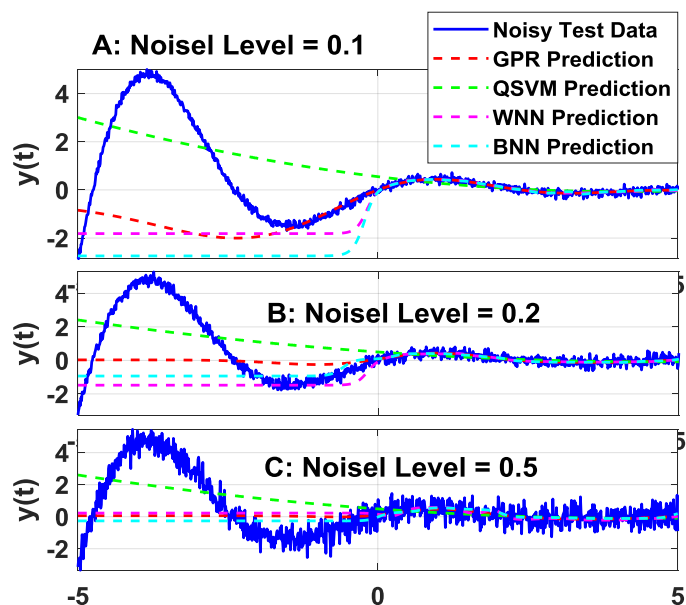
شکل ۷. (الف) مدل سیستم مرتبه دوم با روش GPR (ب) مدل سیستم مرتبه دوم با روش $QSV\text{M}$
Figure 7. (A) Second-order system modeling using the GPR (B) Second-order system modeling using the FGSVM method

برای مشاهده اثر نوفه، اعداد تصادفی با دامنه‌های یکدهم، دودهم و نیم (در سه سطح مختلف از نویز از کم به زیاد) به نمونه‌های آموزشی اضافه می‌شود. در شکل ۸، اثر نوفه بر مدل مشاهده می‌شود که اثر زیادی بر خروجی مدل ندارند. باتوجه به شکل ۸، مشخص است که بازهم روش GPR نسبت به روش‌های WNN ^{۱۴}، BNN ^{۱۳}، $QSV\text{M}$ بهترین پاسخ و دقیق‌ترین مدل را ارائه می‌دهد (زیرا این روش بر روی توزیع احتمالات کار می‌کند و در بیشتر فرایندهای واقعی نیز توزیع احتمال شبیه به توزیع احتمال گاوسی هستند).

¹² Quadratic SVM (Kernel function = Quadratic)

¹³ Bilayered Neural Network (Number of fully connected layer is 2)

¹⁴ Wide Neural Network (MATLAB: 1 layer but size of layer is 100)

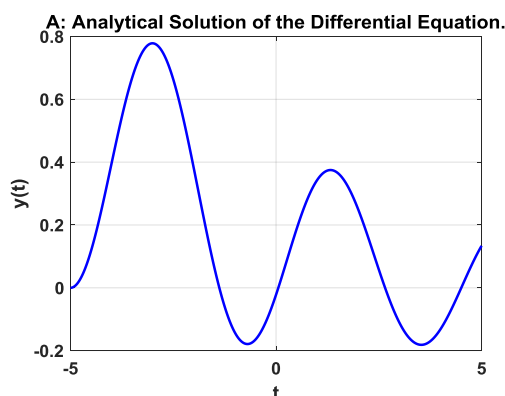


شکل ۸. مدل سازی رگرسیونی سیستم مرتبه دوم با حضور نوفه با سه سطح مختلف از کم به زیاد در داده های آموزشی
Figure 8. Regression modeling of the second-order system with three level of noise in the training data

۲-۴- مدل سازی رگرسیونی سیستم های خطی مرتبه بالاتر از دو

برای سیستم های پیچیده تر از مرتبه دوم، مشابه موارد قبل می توان مدل سازی را انجام داد و رگرسیون توانایی ساختن مدل های دقیق حتی در حضور نوفه و داده های پرت برای سیستم های پیچیده تر را نیز دارد. برای مثال سیستم مرتبه ۴ در معادله (۳) را در نظر بگیرید، پاسخ به صورت فرمول ریاضی بسیار پیچیده ای است که خروجی آن در دامنه زمانی ۵- تا ۵، آورده شده است.

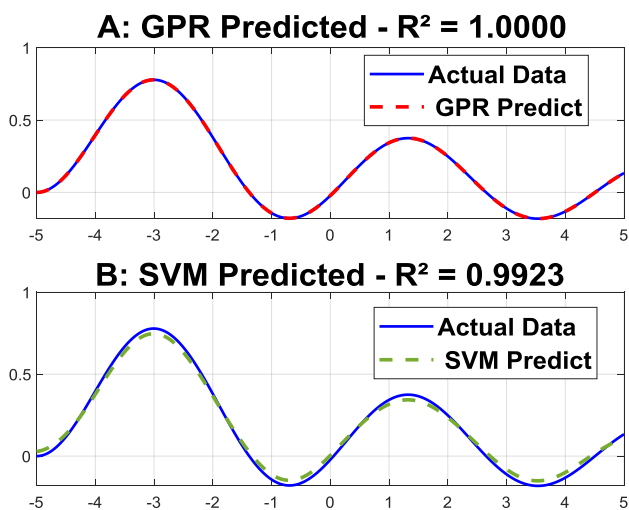
$$y'''' + 2y''' + 3y'' + 4y' + y = 0, y(0) = 0. y'(0) = 0. y''(0) = 1. y'''(0) = 0 \quad (3)$$



شکل ۹. حل معادله نمونه مرتبه چهارم باتوجه به شرایط اولیه معادله ۳
Figure 9. Solution of a sample fourth-order equation (equation 3)

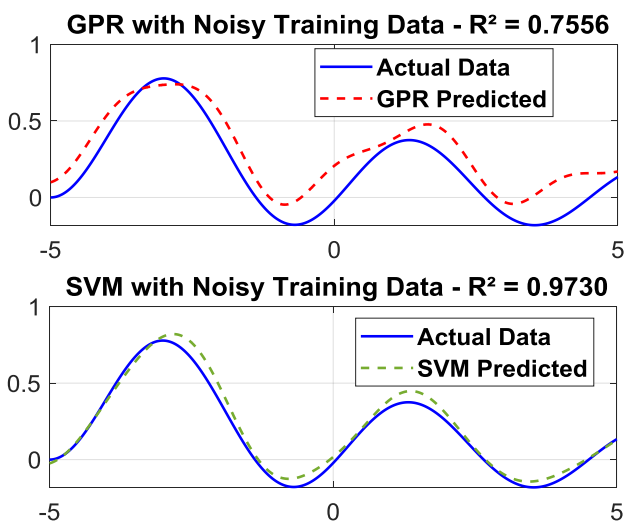
عملیاتی که در قسمت های قبل برای آموزش سیستم مرتبه اول و دوم انجام شد برای این سیستم نیز می تواند صورت پذیرد تنها مشکل زمان انجام محاسبات است که باتوجه به پیچیدگی پاسخ یک سیستم مرتبه چهارم محاسبات

آن برای هر مرحله می‌تواند بسیار بیشتر از قبل باشد. در شکل ۱۰ الف و ب، مدل به ترتیب با روش رگرسیونی GPR و SVM به‌دست‌آمده است. همچنین در شکل ۱۱، مدل سازی در حضور نوفه انجام شده است.



شکل ۱۰. آموزش پاسخ سیستم مرتبه ۴ معادله (۳) با دو روش SVM, GPR بافاصله نمونه‌گیری $h=0.5$

Figure 10. Response of the fourth-order system in Equation (1) with a sampling interval of $h=0.5$ using GPR, SVM



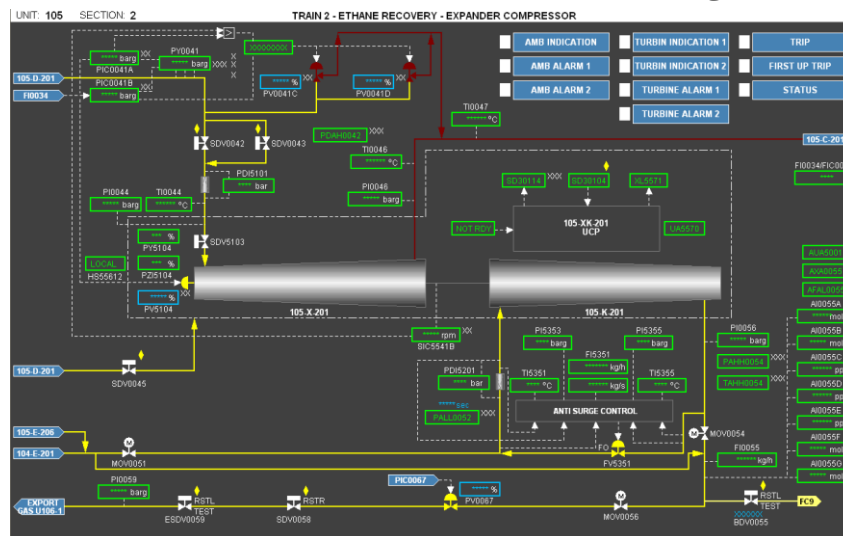
شکل ۱۱. مدل رگرسیونی سیستم مرتبه ۴ در حضور نویز در داده‌های آموزشی

Figure 11. Regression Model of the fourth-order system with Noise in training data

۳- مطالعه موردی

در واحد جداسازی اتان با یک توربوآکسپندر می‌توان دمای گاز را با افت فشار از ۶۰ بار به ۳۰ بار، کاهش داد و اتان را به صورت مایع از گاز جدا کرد (Sayadmosleh et al., 2022). برای مدل‌سازی این قسمت از روش رگرسیونی استفاده شده است و با داشتن تعداد محدودی (تنها ۶۰ عدد داده در دامنه کاری) از مقادیر فشار ورودی PIC0041، فشار

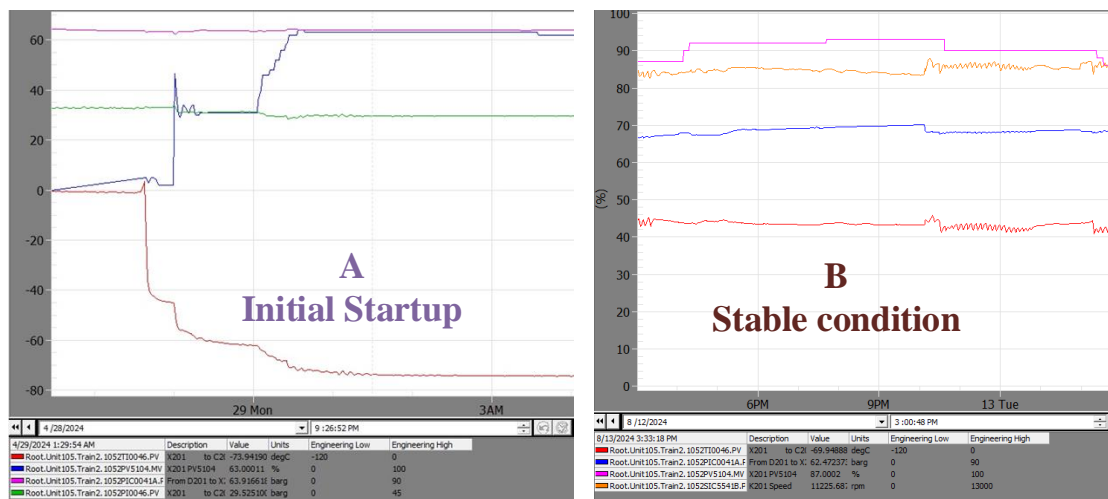
خروجی PI0046، و میزان باز بودن دریچه ورودی کمپرسور PV5104، سرعت کمپرسور SIC5541B به عنوان متغیرهای ورودی و دمای خروجی گاز TI0046 به عنوان خروجی مدل، می توان مدل داده محور را برای دمای خروجی به دست آورد. در شکل ۱۲ نمای این واحد دیده می شود.



شکل ۱۲. شماتیک توربو اکسپندر واحد فراوری اتان

Figure 12. Schematic of the turboexpander in the ethane recovery unit.

در شکل ۱۳ نمودار روندهای^{۱۵} دما و فشار در این واحد را می توان مشاهده کرد که در دو حالت راه اندازی اولیه و مقادیر نهایی تثبیت شده دیده می شوند، از این داده ها برای مدل سازی استفاده شده است. در شکل ۱۴ مدل رگرسیونی به دست آمده با نرم افزار متلب دیده می شود. یکی از قواعد کلی برای انتخاب تعداد داده های آموزش، این است که تعداد نمونه ها باید حداقل ۱۰ برابر تعداد ویژگی ها باشد. در مطالعه موردی توربو اکسپندر تعداد متغیرهای ورودی یا ویژگی ها چهار عدد می باشد بنابراین حداقل به ۴۰ داده برای آموزش نیاز است که در اینجا از ۶۰ داده استفاده شده است. برای مدل سازی ابتدا از روش رگرسیون چند جمله ای درجه چهار استفاده شده است که دقت ۱/۴ درجه سانتی گراد معادل ۹۸ درصد را دارد (شکل ۱۴) و سپس از روش های مدل سازی رگرسیونی GPR و SVM استفاده شده است که دقت ماکزیمم ۹۸/۷ درصد برای روش GPR به دست آمده است. از این مدل ها می توان برای کاربردهای تشخیص عیب استفاده نمود در شکل ۱۶ نحوه اجرای تشخیص عیب با روش مانده ها توسط مدل نشان داده شده است. وقوع عیب باعث می شود که داده های واقعی در لحظاتی با داده های مدل یکسان نباشند و این اختلاف نشان دهنده یک عیب در سیستم می باشد.

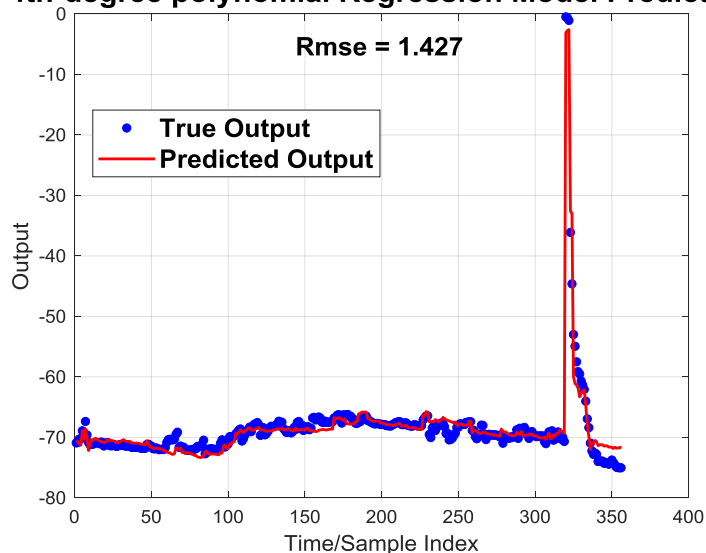


شکل ۱۳. (الف) روندهای توربو اکسپندر واحد فراوری اتان شرایط اولیه (ب) شرایط تثبیت شده

Figure 13. (A) Trends of the turboexpander in the ethane processing unit under initial conditions (B) Stabilized condition.

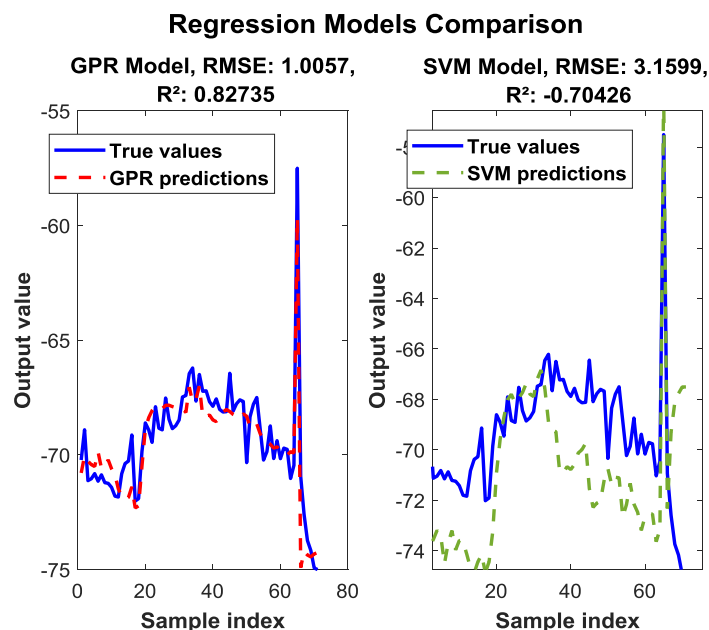
همچنین می توان نمونه دیگری از کاربرد مدل سازی با روش های رگرسیونی جهت تشخیص عیب و بهبود کارکرد کوره های پالایشگاه را ارائه نمود. مصرف انرژی در کوره های موجود در پالایشگاه یکی از موارد مهم برای بهینه سازی مصرف سوخت و افزایش بازده یک پالایشگاه است (Mohsen Maleki Douz-Zadeh & Mohammad Hamid, 2013). در شکل ۱۷، کوره ای مشاهده می شود که در پالایشگاه گاز جهت گرم کردن و تجدید کاتالیست های جذب آب در واحدهای نم گیر پالایش گاز و یا موارد مشابه استفاده می شود. کنترل این کوره ها توسط یک شیر کنترل سوخت انجام می شود که بسته به میزان دمای مورد نظر، میزان سوخت ورودی را کم یا زیاد می کند.

4th-degree polynomial Regression Model Prediction



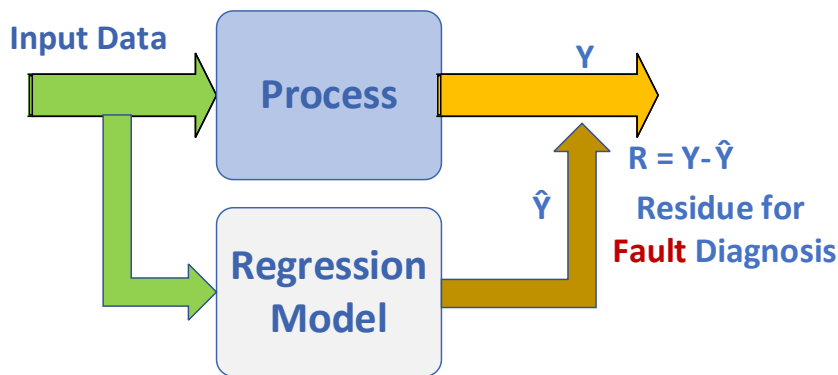
شکل ۱۴. مدل توربو اکسپندر با رگرسیون خطی چند جمله ای درجه چهار (دقت مدل ۱/۴۲۷ درجه سانتی گراد دقت ۹۸٪)

Figure 14. The turbo expander model with a fourth-degree polynomial regression model (Rmse = 1.427 degrees Celsius, 98% Accuracy).



شکل ۱۵. مدل توربو اکسپندر با روش های رگرسیون ماشین بردار پشتیبان و فرایند گاوسی (دقت ماکزیمم با مدل فرایند گاوسی یک درجه سانتی گراد دقت ۹۸/۷٪)

Figure 15. Turboexpander model, GPR and SVM method (maximum model accuracy of 1°C with GPR, 98.7%).



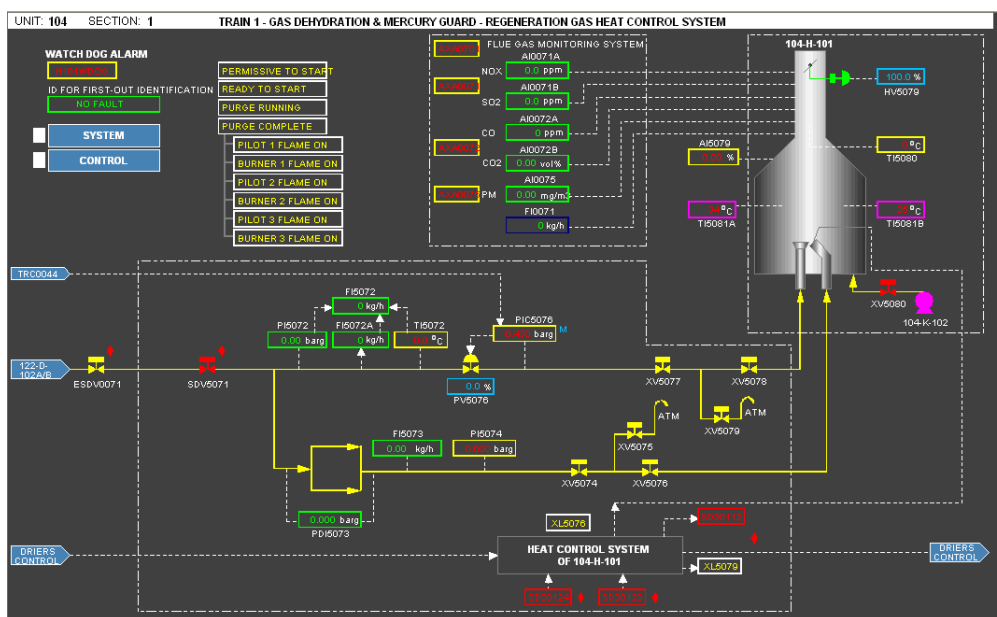
شکل ۱۶. نحوه استفاده از مدل رگرسیونی در کاربردهای تشخیص عیب

Figure 16. Regression model in fault diagnosis applications.

در بالای دودکش این کوره یک دریچه قابل کنترل HV5079 برای جلوگیری از هدررفت انرژی تعبیه شده است. کنترل میزان باز یا بسته بودن این دریچه چالشی مهندسی است، زیرا اگر این دریچه کاملاً باز باشد، باعث هدررفت گرما و انرژی، کاهش بازده و افزایش مصرف سوخت می شود. در حالی که بسته شدن بیش از حد آن می تواند باعث خفگی و خاموش شدن ناگهانی کوره شود. در اینجا می توان با مدل سازی رگرسیونی کوره، با استفاده از داده های آموزشی، بهترین میزان باز یا بسته بودن این شیر (باتوجه به مصرف سوخت و بازده کوره) که توسط یک موقعیت یاب^{۱۶} کنترل می شود، به

¹⁶ Positioner

دست آورد. روش به دست آوردن مدل نیز همانند مطالعه موردی اول، با استفاده از داده‌های آموزشی محدود در نرم‌افزار متلب و با به کارگیری روش‌های SVM و GPR انجام می‌شود.



شکل ۱۷ کوره با دریچه دودکش قابل تنظیم برای بهترین بازده
Figure 17. Furnace with an adjustable chimney valve for optimal efficiency.

۴- بحث و نتیجه گیری

در پالایشگاه‌های گاز، فرایندهای متنوع و پیچیده‌ای وجود دارند که اغلب فاقد مدل‌های دقیق فیزیکی هستند. تشخیص سریع عیب در این سیستم‌ها همواره چالشی بزرگ بوده که تاکنون به روش‌های سنتی و باتکیه بر تحلیل کارشناسان باتجربه و صرف زمان قابل توجه انجام شده است که این روش‌ها به دلیل زمان بر بودن، باعث هدررفت انرژی و یا مواد باارزش می‌شود. برای تشخیص عیب دقیق هر سیستم نیاز به داشتن مدل دقیقی از آن سیستم است، این مقاله امکان مدل‌سازی رگرسیونی سیستم‌های پیچیده را بدون نیاز به معادلات ریاضی و تنها با بهره‌گیری از داده‌های ورودی و خروجی بررسی کرده است. مدل‌های پیشنهادی قابلیت کاربرد در زمینه‌های متعددی، از جمله تشخیص عیب دستگاه‌ها، را دارند. در این پژوهش، مدل‌سازی رگرسیونی سیستم‌های مرتبه اول و دوم در شرایط حضور نوفه و داده‌های ناقص یا پرت مورد بررسی قرار گرفت و خطاهای مرتبط تحلیل شدند. همچنین نشان داده شد که برای سیستم‌های پیچیده‌تر، می‌توان این روش‌ها را توسعه داد. در مطالعه موردی، توربو اکسپندر واحد تولید اتان در یک پالایشگاه گازی با استفاده از روش‌های پیشنهادی مدل‌سازی شد و خروجی این مدل داده‌محور دقت ۹۸/۷ درصدی را نشان داد که برای کاربرد تشخیص دقیق مقدار عیب بسیار مناسب است. در این پژوهش، مشخص شد که روش‌های رگرسیون فرایند گاوسی و رگرسیون ماشین بردار پشتیبان نسبت به سایر روش‌های یادگیری ماشین، از جمله شبکه‌های عصبی، عملکرد بهتری در حضور نوفه و داده‌های پرت دارند. این روش‌ها می‌توانند مدل‌های رگرسیونی دقیق و قابل اعتماد برای فرایندهای پیچیده

ارائه دهند و از این مدل‌ها به راحتی می‌توان برای کاربردهای تشخیص عیب در فرایندهای پیچیده استفاده نمود. به عنوان کار آینده، پیشنهاد می‌شود که برنامه اپلیکیشنی نوشته شود که عیب‌های تشخیص داده شده در سیستم کنترل گسترده را از طریق یک شبکه امن پیام‌رسان به اطلاع بهره‌بردار و یا رئیس واحد برساند تا عملیات تشخیص عیب با سرعت و دقت بیشتری صورت پذیرد.

۵- تقدیر و تشکر

این مقاله با حمایت مالی و معنوی شرکت مجتمع گاز پارس جنوبی SPGC تهیه شده است.

۶-مراجع

- Ahmadi, S. H., & Khosrowjerdi, M. J. (2021). Fault detection Automation in Distributed Control Systems using Data-driven methods : SVM and KNN. *TechRxiv. Preprint*.
- Akbarpour, S., & Khosrowjerdi, M. J. [In persian] (2024). Estimation of Combustor and Compressor Faults in Industrial Gas Turbines by Using Multiple Model Approach. *Iranian Chemical Engineering Journal*, 23(134).
- Alzghoul, A., Backe, B., Löfstrand, M., Byström, A., & Liljedahl, B. (2014). Comparing a knowledge-based and a data-driven method in querying data streams for system fault detection: A hydraulic drive system application. *Computers in Industry*, 65(8), 1126–1135.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16–28.
- Cheliotis, M., Lazakis, I., & Theotokatos, G. (2020). Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering*, 216.
- Chen, Z., O'Neill, Z., Wen, J., Pradhan, O., Yang, T., Lu, X., Lin, G., Miyata, S., Lee, S., Shen, C., Chiosa, R., Piscitelli, M. S., Capozzoli, A., Hengel, F., Kühner, A., Pritoni, M., Liu, W., Clauß, J., Chen, Y., & Herr, T. (2023). A review of data-driven fault detection and diagnostics for building HVAC systems. *Applied Energy*, 339.
- Clavijo, N., Melo, A., Câmara, M. M., Feital, T., Anzai, T. K., Diehl, F. C., Thompson, P. H., & Pinto, J. C. (2019). Development and application of a data-driven system for sensor fault diagnosis in an oil processing plant. *Processes*, 7(7).
- Donald R. Woods. (2021). *Troubleshooting Guide for Process Engineers: A Complete Course with Applied Examples and Glossary*.
- Habibi, H., Howard, I., & Simani, S. (2019). Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review. In *Renewable Energy* (Vol. 135, pp. 877–896). Elsevier Ltd.
- M. J. Khosrowjerdi, & Mahdi Aliyari-Shoorehdeli. [In Persian] (2014). *Fault Diagnosis and Compensation in Control Systems*. Sahand University of Technology (SUT).
- Mohsen Maleki Douz-Zadeh, & Mohammad Hamid. [In persian] (2013). Investigation of the Effect of Changing the Location of Heat Exchangers in the Abadan Refinery Distillation Unit on the Performance of the Furnace and Flash Drum. *Iranian Chemical Engineering Journal*.
- Norman P. Lieberman - *Troubleshooting Process Plant Control_ A Practical Guide to Avoiding and Correcting Mistakes*-John Wiley & Sons (2024).

- Sarmast Alizadeh, F., & Esfandiari, N. [In persian] (2024). Measuring the Quality of Industrial Water Used in a Chemical Plant Using arTificial Intelligence (Logistic Regression, Naive Bayesian, Support Vector Machine, Random forest, and Decision Tree). *Iranian Chemical Engineering Journal*, 23(135), 90–101.
- Sayadmosleh, E., Shirazi, L., Soleimani, M., & Sarmad, M. [In persian] (2022). Comparison of the Thermodynamic Effect of Ethane and Ethylene in the Mixed Refrigerant Cycle of the Natural Gas Liquefaction Process. *Iranian Chemical Engineering Journal*, 21(123), 7–21.
- Tidiri, K., Chatti, N., Verron, S., & Tiplica, T. (2016). Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. In *Annual Reviews in Control* (Vol. 42, pp. 63–81). Elsevier Ltd.
- Yan, F., Zhang, X., Yang, C., Hu, B., Qian, W., & Song, Z. (2023). Data-driven modelling methods in sintering process: Current research status and perspectives. In *Canadian Journal of Chemical Engineering* (Vol. 101, Issue 8, pp. 4506–4522). John Wiley and Sons Inc.
- Zhao, Y., Wang, S., & Xiao, F. (2013). A system-level incipient fault-detection method for HVAC systems. *HVAC and R Research*, 19(5), 593–601.