



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

S. H. Ahmadi¹, M. J. Khosrowjerdi^{2*}

1- PhD. Student of Control Engineering, Tabriz University of Technology

2- Professor of Control Engineering, Tabriz University of Technology

Email: khosrowjerdi@sut.ac.ir

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.

Received: 20 December 2024

Accepted: 20 April 2025

Page Number: 50-59

Keywords:

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

Please Cite this Article Using:

Ahmadi, S. H., & Khosrowjerdi, M. J. (2026). Data-Driven Fault Diagnosis in Gas Refining Processes: A Case Study Using Support Vector Regression and Gaussian Process Regression. *Iranian Chemical Engineering Journal*, 24(143), 50-59, [In Persian].



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، دانشگاه صنعتی تبریز

۲- استاد مهندسی برق، دانشگاه صنعتی تبریز

پیام‌نگار: khosrowjerdi@sut.ac.ir

چکیده

تشخیص سریع عیب در فرایندهای پیچیده پالایشگاه‌های گازی به دلیل اهمیت راه‌بردی این تجهیزات و هزینه‌های بالای نگهداری آن‌ها، یکی از چالش‌های کلیدی در مهندسی و کنترل فرایندها است. با وجود 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 با روش‌های رگرسیون و یا مدل‌های دقیق انجام‌پذیر است. این مقاله، روشی مبتنی بر یادگیری ماشین برای مدل‌سازی رگرسیونی سامانه‌های صنعتی با داده‌های پرت و با نوفه برای کاربردهای تشخیص عیب ارائه می‌کند. برای تشخیص و مکان‌یابی عیب‌ها، دو روش رگرسیونی، شامل رگرسیون ماشین بردار پشتیبان و رگرسیون فرایند گاوسی، به کار گرفته شده است. این مدل‌ها قادر به پیش‌بینی رفتار سامانه‌های صنعتی و شناسایی عیوب بدون نیاز به مدل‌های فیزیکی دقیق هستند. نوآوری اصلی این مقاله، معرفی یک چهارچوب یادگیری ماشین است که می‌تواند در مقایسه با روش‌های سنتی تشخیص عیب به وسیله متخصص باتجربه در شرایط داده‌های ناقص و نوفه‌ای به‌طور دقیق‌تر و سریع‌تری عمل کند.

کلیدواژه‌ها:

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

* تبریز، دانشگاه صنعتی سهند، دانشکده مهندسی برق

استناد به مقاله:

احمدی، سید حسین، و خسروجردی، محمدجواد. (۱۴۰۴). تشخیص عیب داده‌محور در فرایند پالایش گاز: مطالعه موردی با استفاده از رگرسیون ماشین بردار پشتیبان و فرایند گاوسی، نشریه مهندسی شیمی ایران، ۲۴(۱۴۳)، ۵۹-۵۰.

۱. مقدمه

پالایشگاه‌های گازی با استفاده از سامانه‌های کنترل توزیع شده^۱، فرایندهای خود را در اتاق‌های کنترل به دست مهندسان، مدیریت و نظارت می‌کنند. با این حال، تشخیص عیب در این سامانه‌ها کمتر جلب توجه کرده است، در حالی که روش‌های داده‌محور می‌تواند جای‌گزینی سریع‌تر و مؤثرتر برای روش‌های دستی باشد. اهمیت تشخیص عیب در این حوزه با کاهش هدررفت انرژی، آلودگی محیط زیست و هزینه‌های نگهداری مرتبط است. روش‌های سنتی به نیروی انسانی با تجربه نیاز دارد که همیشه در دسترس نیست و دانش آن‌ها غالباً محدود به شرایط خاص است [۱۲-۸]. آشکارسازی و تشخیص عیب به دو روش مدل بیس [۱۰ و ۹ و ۲] و داده‌محور [۶] و یا ترکیبی از این دو [۱۵] انجام می‌گیرد. در روش‌های داده‌محور که با ظهور روش‌های جدید یادگیری ماشین که محل توجه بسیار است، از روش‌های کلاس‌بندی عیب برای آشکارسازی و محل‌یابی عیب و یا تشخیص عیب با استفاده از مدل‌های رگرسیونی [۱۳] برای تعیین دقیق مقدار عیب استفاده می‌شود. با استفاده از روش‌های کلاس‌بندی عیب، می‌توان جدولی از ویژگی‌ها یا مشخصه‌های مؤثر را در ایجاد هر عیب تشکیل داد و هر عیب را به عنوان یک کلاس مجزا تعریف کرد. سپس، سامانه را آموزش داد تا در صورت وقوع مجدد آن عیب، به‌طور خودکار آن را تشخیص دهد. این رویکرد، احتمال خرابی را کاهش و قابلیت اطمینان سامانه را افزایش می‌دهد [۱]. در روش‌های تشخیص عیب، علاوه بر آشکارسازی وجود عیب، می‌توان مقدار دقیق عیب را نیز اندازه‌گیری کرد. این امر امکان محاسبه میزان تغییرات لازم را برای جبران عیب فراهم می‌کند. البته، برای دستیابی به این هدف، به یک مدل دقیق از فرایند واقعی یا سامانه مورد نظر نیاز است.

در این مقاله، دو روش اصلی یادگیری ماشین، یعنی رگرسیون ماشین بردار پشتیبان^۲ و رگرسیون فرایند گاوسی^۳ برای مدل‌سازی رگرسیونی در تشخیص عیب معرفی شده است. این روش‌ها به‌ویژه در سامانه‌های پیچیده‌ای که نمی‌توان مدل دقیق فیزیکی برای آن‌ها به دست آورد، مفید هستند [۱۷ و ۱۶ و ۳]. چالش‌هایی مانند حضور نوفه [۷]، داده‌های پرت^۴ و پراکنده^۵ و انتخاب ویژگی‌های مناسب [۴]

1. Distributed Control System (DCS)
2. Support Vector Machine Regression (SVR)
3. Gaussian Process Regression (GPR)
4. Outliers
5. Sparse Data

نیز در این مقاله، بررسی و راه‌کارهایی برای مقابله با این مشکلات ارائه شده است.

با توجه به این که تاکنون در تشخیص عیب در سامانه‌های کنترل گسترده پالایشگاه‌ها تحقیقات کاربردی زیادی انجام نشده است، این مقاله تلاش دارد که با فراهم آوردن زمینه خودکارسازی فرایند تشخیص عیب و ایجاد یک پایگاه داده 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 سامانه 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.

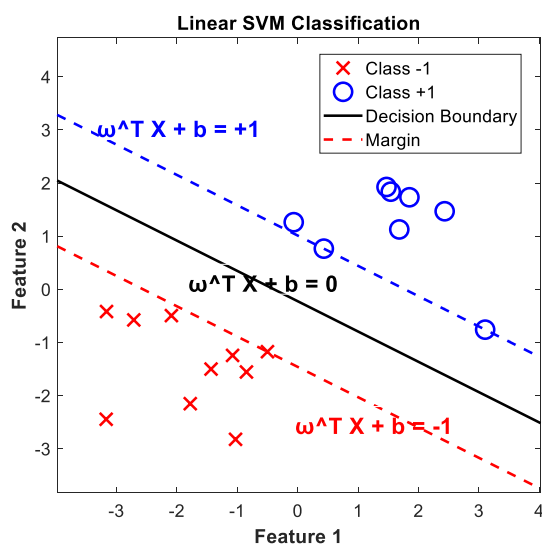
در این مقاله، ابتدا روش‌های یادگیری ماشین، معرفی و سپس، مدل‌سازی رگرسیونی سامانه‌های خطی در چهار سطح مختلف بررسی می‌شود. نتایج نشان‌دهنده دقت بالای این مدل‌سازی‌ها است. چالش‌های مختلف، مانند نوفه و داده‌های پرت، تحلیل و راه‌کارهایی برای مقابله با آن ارائه می‌شود. در بخش سوم، مطالعات موردی در یک پالایشگاه واقعی بررسی و اثربخشی روش‌های پیشنهادی ارزیابی می‌شود. سرانجام، نتیجه‌گیری و پیشنهادهای پژوهش‌های آینده ارائه شده است.

۲. معرفی روش‌ها

۲-۱ رگرسیون فرایند گاوسی و رگرسیون ماشین بردار پشتیبان

در یک مسأله عمومی رگرسیون، هدف این است که با استفاده از تعدادی نقطه مشاهده شده، تابعی پیدا کنیم که بهترین توصیف را از این نقاط ارائه دهد. سپس، از این تابع برای پیش‌بینی نقاطی که قبلاً مشاهده نشده است استفاده می‌شود؛ هرچند که برای مجموعه‌ای متناهی از نقاط در یک فضای پیوسته، بی‌شمار تابع وجود دارد که از این نقاط عبور می‌کند. رگرسیون فرایند گاوسی می‌کوشد تا توزیع احتمالی بر 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 مشاهده شده جدید، به روزرسانی می‌شود.

$f(X)$ مثبت و الگوهای بابرچسب کلاس منفی یک باید دارای $f(X)$ منفی باشند. فاصله نزدیک‌ترین داده هر کلاس تا ابرصفحه می‌توان فاصله داده‌های آن کلاس تا مرز جداساز در نظر گرفت. فاصله هر نقطه دلخواه تا ابرصفحه مورد نظر برابر با $\frac{|\omega^T X + b|}{\|\omega\|}$ است. واضح است که ابرصفحه می‌تواند طوری قرار گیرد که نزدیک‌ترین داده هر یک از دو کلاس تا آن برابر $\frac{a}{\|\omega\|}$ باشد؛ به طوری که دو ابرصفحه $\omega^T X + b = \pm a$ در دو سوی ابرصفحه مرزی قرار گیرند. این مفاهیم در شکل (۲)، نشان داده شده است.



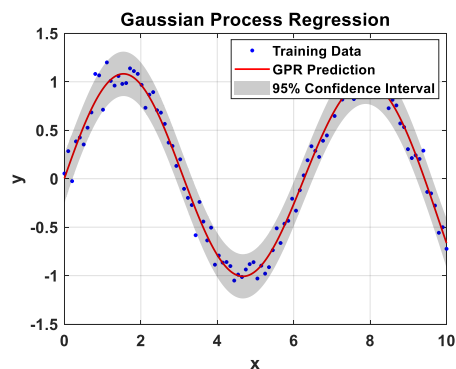
شکل ۲. طبقه‌بندی‌کننده ماشین‌های بردار پشتیبان با استفاده از بیشینه حاشیه بین کلاس‌ها.

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

با اعمال ضریبی به کل رابطه ابرصفحه‌ها تغییری در آن ایجاد نخواهد شد، لذا می‌توان ابرصفحه را نسبت به a نرمال کرد و یا به عبارتی، مشخصه $a=1$ قرار داده شود. دو ابرصفحه کناری محدوده امنی به طول $\frac{2}{\|\omega\|}$ برای ابرصفحه مرزی $f(X) = 0$ به وجود می‌آورند. به این محدوده، حاشیه گفته می‌شود. هرچه این حاشیه بیشتر باشد، جداسازی بهتری انجام گرفته است؛ بنابراین، باید حاشیه $\frac{2}{\|\omega\|}$ را بیشینه کرد. با حل این مسأله، بهینه‌سازی خط جداسازی پیدا خواهد شد که فاصله حاشیه‌ای ماکزیمی نیز نسبت به نزدیک‌ترین داده‌های هر کلاس داشته باشد.

در حالت رگرسیون ماشین بردار پشتیبان^۲ برعکس حالت

از آنجایی که خروجی این مدل توزیعی روی تمام توابع ممکن است، می‌توان میانگین این توابع را به عنوان پیش‌بینی مدل و واریانس آن‌ها را به عنوان معیاری از اطمینان پیش‌بینی انجام شده در نظر گرفت. در شکل (۱)، مدل‌سازی یک فرایند با روش GPR نشان داده شده است.



شکل ۱. مدل‌سازی با روش GPR.

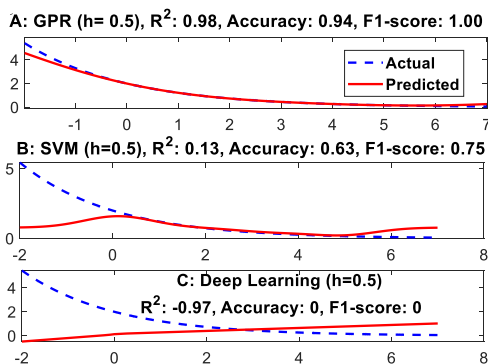
Figure 1. Modeling with GPR Method.

ماشین بردار پشتیبان^۱ 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 بر آن است تا ابرصفحه‌ای انتخاب شود که حاشیه اطمینان بیشتری برای کلاس‌بندی و جداسازی داده‌ها داشته باشد. پیدا کردن خط بهینه برای جداسازی داده‌ها به وسیله روش‌های برنامه‌سازی غیرخطی انجام می‌گیرد. برای مثال، یک مسأله طبقه‌بندی دو کلاسه با خروجی $y = \{-1, 1\}$ را 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 مسأله وجود دارد. در مسائل با ابعاد و ویژگی بالاتر مرز تصمیم‌گیری، یک ابرصفحه خواهد بود. مناسب‌تر است که مرز تصمیم‌گیری تا جایی که می‌شود از داده‌های دو کلاس دور باشد. از آنجایی که مرز کلاس‌ها، خطی فرض شده است، تابع آن به صورت $f(X) = \omega^T X + b = 0$ تعریف می‌شود، که ω و b مشخصه‌های خط و یا صفحه است، این ابرصفحه یک طبقه‌بندی‌کننده خطی است و $\text{sign}(f(X))$ برچسب کلاس هر داده جدید X را در مرحله آزمون تعیین خواهد کرد. در واقع، الگوهای با برچسب مثبت یک باید

2. Support Vector Machine Regression (SVR)

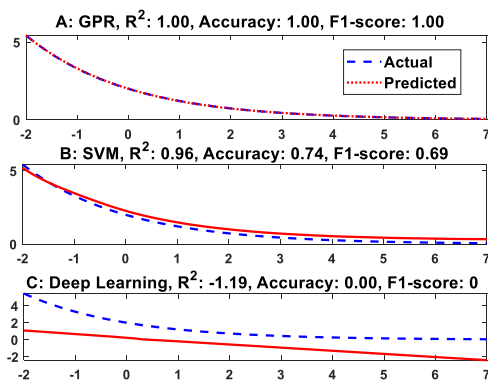
1. Support Vector Machine (SVM)

کاهش یافته‌است. در شکل (۵)، روش GPR و روش‌های SVM و Deep Learning نشان داده شده‌است که خطای مدل GPR نسبت به دو روش دیگر کمتر است. در هر شکل مشخصه R^2 که نشان‌دهنده میزان تطابق مدل با واقعیت یا معادله است، آورده شده و میزان تطابق کامل است. برای قدم بعدی حضور نوفه و داده‌های پرت نیز با مدل‌سازی رگرسیونی بررسی می‌شود. نوفه با دامنه‌های یک‌دهم و نیم در دو سطح به صورت اعداد تصادفی به نمونه‌های تحت آزمون اضافه می‌شود که مدل تقریباً دقیقی با حضور نوفه به دست می‌دهد. شکل (۶-a)، در مرحله بعد چندین داده پرت به داده‌های آموزشی اضافه می‌شود و مدل به دست می‌آید. شکل (۶-b).



شکل ۴. مدل‌سازی با روش SVM، GPR، Deep Learning با فواصل بین داده‌های آموزشی برابر نیم.

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

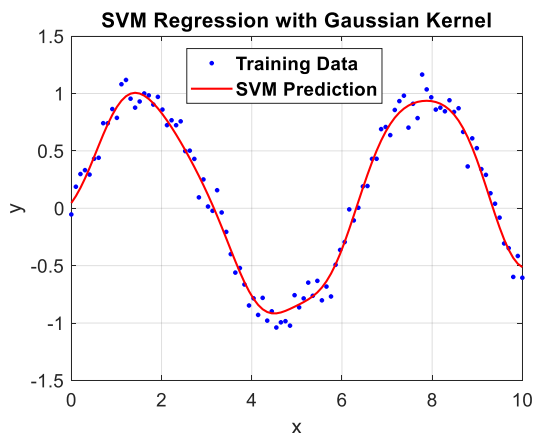


شکل ۵. مدل‌سازی با روش SVM، GPR، Deep Learning با فواصل بین داده‌های آموزشی برابر یک‌دهم.

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

4. Deep Neural Network (Matlab: Fully Connected Layer = 10)

لازم است که دنبال یک کانال با حاشیه مناسب بود که تمام نمونه‌ها در آن قرار بگیرد. در شکل (۳)، مدل‌سازی رگرسیونی با روش SVM - که به آن SVR نیز گفته می‌شود - آورده شده‌است، هدف جادادن داده‌های جدید در یک کانال با کمترین حاشیه ممکن است که با داده‌های گذشته به وجود آمده‌است. در حالت رگرسیون نیز مانند حالت کلاس‌بندی می‌توان از انواع کرنل‌های مختلف، مانند: کرنل گاوسی^۱ یا مربعی^۲ و مکعبی^۳ برای روش ماشین بردار پشتیبان استفاده کرد.



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

۲-۲ مدل‌سازی رگرسیونی یک سامانه مرتبه اول

یک سامانه ساده مرتبه اول به سامانه‌ای گفته می‌شود که با یک معادله دیفرانسیل خطی، توصیف و در آن تنها خروجی و مشتق مرتبه اول آن (با توان یک) ظاهر می‌شود. به عنوان مثال، در یک سامانه هیدرولیکی، اگر ارتفاع آب در مخزن به عنوان خروجی و دبی ورودی آب به عنوان ورودی سامانه در نظر گرفته شود، این سامانه به عنوان یک سامانه مرتبه اول مدل‌سازی می‌شود. یک سامانه مرتبه اول را با پاسخ $y = 2e^{-0.5x}$ در نظر بگیرید که برای x بین صفر تا ۵ و با فواصل نیم، آموزش داده شده‌است و سپس، خروجی با مدل‌سازی رگرسیونی برای ۰ تا ۵ با فواصل یک‌صدم در شکل (۴) به دست آمده‌است. در شکل (۴)، مشاهده می‌شود که مدل GPR به طور دقیق بر معادله منطبق شده‌است و روش‌های SVM، Deep Learning کمی خطا دارند. در مرحله بعد، فاصله نمونه‌ها برای آموزش از نیم به یک‌دهم

1. Gaussian Support Vector Machine
2. Quadratic Support Vector Machine
3. Cubic Support Vector Machine

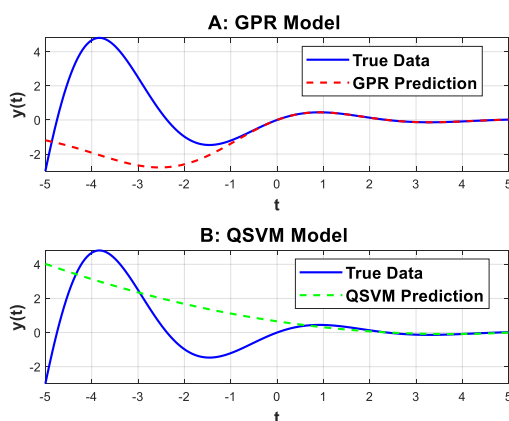
۲-۳ مدل‌سازی رگرسیونی سامانه خطی مرتبه دوم

یک سامانه مرتبه دوم نوسانی (زیر میرا) را 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 (a) و (b) انجام شده است.

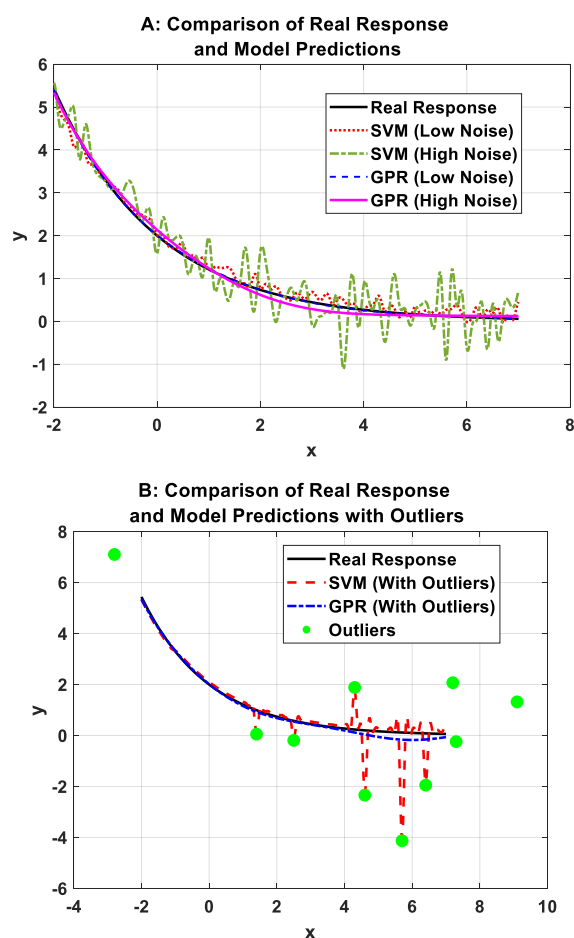


شکل ۷. (a) مدل سامانه مرتبه دوم با روش GPR (b) مدل سامانه مرتبه دوم با روش QSVM.

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 ندارد. باتوجه به شکل (۸)، مشخص است که بازهم روش GPR نسبت به روش‌های WNN^2 ، BNN^2 ، QSVM، بهترین پاسخ و دقیق‌ترین مدل را ارائه می‌کند (زیرا این روش بر روی توزیع احتمالات کار می‌کند و در بیشتر فرایندهای واقعی نیز توزیع احتمال شبیه به توزیع احتمال گاوسی است).

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 مدل‌سازی دارند. باتوجه به این که تنها 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 سامانه مرتبه دوم نوسانی (زیر میرا) انجام و خروجی مدل با حضور نوفه و داده‌های پرت نیز بررسی می‌شود.

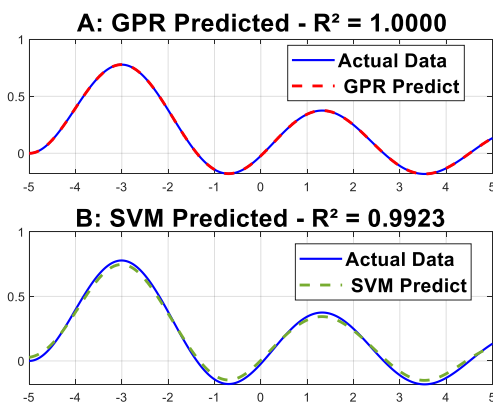


شکل ۶. (a) مدل سامانه مرتبه اول با روش‌های SVM و GPR در حضور نوفه در دو حالت کم و زیاد (b) مدل سامانه مرتبه اول در حضور داده‌های پرت.

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.

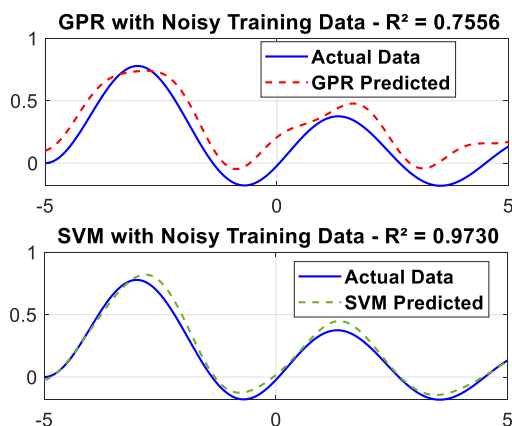
1. Quadratic SVM (Kernel Function = Quadratic)
2. Bilayered Neural Network (Number of Fully Connected Layer is 2)
3. Wide Neural Network (MATLAB: 1 Layer but Size of Layer is 100)

عملیاتی که در قسمت‌های قبل برای آموزش سامانه مرتبه اول و دوم انجام شد برای این سامانه نیز می‌تواند انجام پذیرد، تنها مشکل زمان انجام محاسبات است که باتوجه به پیچیدگی پاسخ یک سامانه مرتبه چهارم، محاسبات آن برای هر مرحله می‌تواند بسیار بیشتر از قبل باشد. در شکل (۱۰) و (۱۱) مدل به ترتیب با روش رگرسیونی GPR و SVM به دست آمده است. هم‌چنین، در شکل (۱۱)، مدل سازی در حضور نوفه انجام شده است.



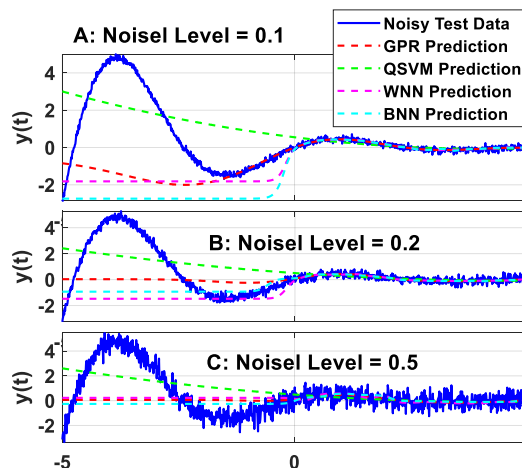
شکل ۱۰. آموزش پاسخ سامانه مرتبه ۴ معادله (۳) با دو روش GPR, SVM با فاصله نمونه‌گیری $h=0.5$.

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 سامانه مرتبه ۴ در حضور نویز در داده‌های آموزشی.

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

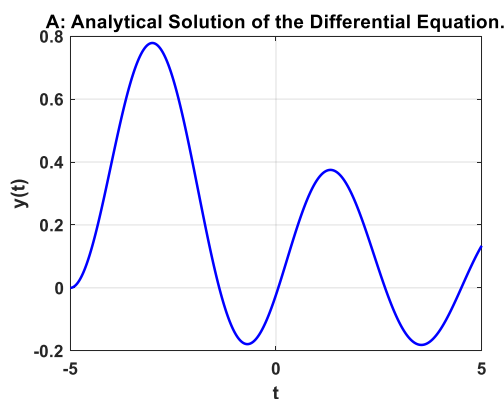


شکل ۸. مدل سازی رگرسیونی سامانه مرتبه دوم با حضور نوفه با سه سطح مختلف از کم به زیاد در داده‌های آموزشی.

Figure 8. Regression modeling of the second-order system with three level of noise in the training data.

۲-۴ مدل سازی رگرسیونی سامانه‌های خطی مرتبه بالاتر از دو برای سامانه‌های پیچیده‌تر از مرتبه دوم، مشابه موارد قبل می‌توان مدل سازی را انجام داد و رگرسیون توانایی ساختن مدل‌های دقیق را حتی در حضور نوفه و داده‌های پرت برای سامانه‌های پیچیده‌تر نیز دارد. برای مثال، سامانه مرتبه ۴ را در معادله (۳) در نظر بگیرید؛ پاسخ به صورت فرمول ریاضی بسیار پیچیده‌ای است که خروجی آن در دامنه زمانی ۵- تا ۵، در شکل (۹)، آورده شده است.

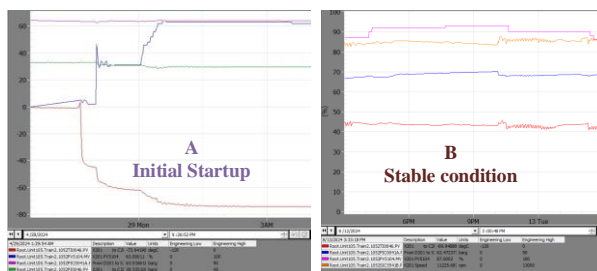
$$y'''' + 2y''' + 3y'' + 4y' + y = 0, y(0) = 0, y'(0) = 0, y''(0) = 1, y'''(0) = 0 \quad (3)$$



شکل ۹. حل معادله نمونه مرتبه چهارم باتوجه به شرایط اولیه معادله (۳).

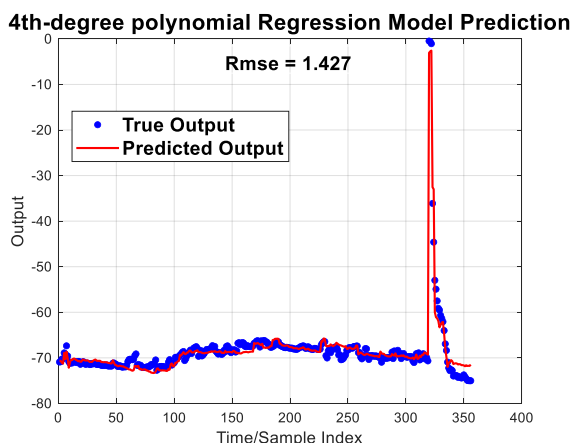
Figure 9. Solution of a sample fourth-order equation (equation 3).

روش‌های مدل‌سازی رگرسیونی GPR و SVM استفاده شده‌است که دقت بیشینه ۹۸/۷ درصد برای روش GPR به دست آمده‌است. از این مدل‌ها می‌توان برای کاربردهای تشخیص عیب استفاده کرد؛ در شکل (۱۶) نحوه اجرای تشخیص عیب با روش مانده‌ها به وسیله مدل نشان داده شده‌است. وقوع عیب در سامانه باعث می‌شود که داده‌های واقعی در لحظاتی با داده‌های مدل یکسان نباشند.



شکل ۱۳. (a) روندهای توربو اکسپندر واحد فراوری اتان شرایط اولیه (b) شرایط تثبیت شده.

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



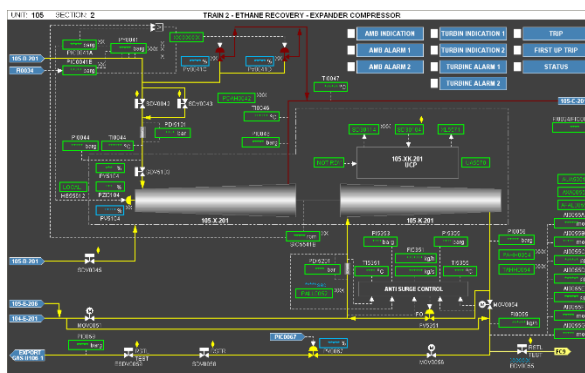
شکل ۱۴. مدل توربو اکسپندر با رگرسیون خطی چندجمله‌ای درجه چهار (دقت مدل ۱/۴۲۷ درجه سلسیوس دقت ۹۸٪).

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

هم‌چنین، می‌توان نمونه دیگری از کاربرد مدل‌سازی را با روش‌های رگرسیونی برای تشخیص عیب و بهبود کارکرد کوره‌های پالایشگاه ارائه کرد. مصرف انرژی در کوره‌های موجود در پالایشگاه یکی از موارد مهم برای بهینه‌سازی مصرف سوخت و افزایش بازده یک پالایشگاه است [۱۱]. در شکل (۱۷)، کوره‌ای مشاهده می‌شود که در پالایشگاه گاز برای گرم کردن و تجدید کاتالیست‌های جذب آب در واحدهای

۳. مطالعه موردی

در واحد جداسازی اتان با یک توربو اکسپندر 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 [۱۴]. برای مدل‌سازی این قسمت از روش رگرسیونی استفاده شده‌است و با داشتن تعداد محدودی (تنها ۶۰ عدد داده در دامنه کاری) از مقادیر فشار ورودی PIC0041، فشار خروجی PI0046، و میزان بازبودن دریچه ورودی کمپرسور PV5104، سرعت کمپرسور SIC5541B به عنوان متغیرهای ورودی و دمای خروجی گاز TI0046 به عنوان خروجی مدل، می‌توان مدل داده‌محور را برای دمای خروجی به دست آورد. در شکل (۱۲) نمای این واحد دیده می‌شود.



شکل ۱۲. طرحواره توربو اکسپندر واحد فراوری اتان.

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

در شکل (۱۳) نمودار روندهای دما و فشار را در این واحد می‌توان مشاهده کرد که در دو حالت راه‌اندازی اولیه و مقادیر نهایی تثبیت شده دیده می‌شود؛ از این داده‌ها برای مدل‌سازی استفاده شده‌است. در شکل (۱۴) مدل رگرسیونی به دست آمده با نرم‌افزار متلب دقت دیده می‌شود. یکی از قواعد کلی برای انتخاب تعداد داده‌های آموزش، این است که تعداد نمونه‌ها باید حد اقل ۱۰ برابر تعداد ویژگی‌ها باشد. در مطالعه موردی توربو اکسپندر تعداد متغیرهای ورودی یا ویژگی‌ها چهار عدد است؛ بنابراین، حد اقل به ۴۰ داده برای آموزش نیاز است که در اینجا از ۶۰ داده استفاده شده‌است. برای مدل‌سازی، ابتدا از روش رگرسیون چندجمله‌ای درجه چهار استفاده شده‌است که دقت ۱/۴ درجه سلسیوس، معادل ۹۸ درصد را دارد (شکل (۱۴)) و سپس، از

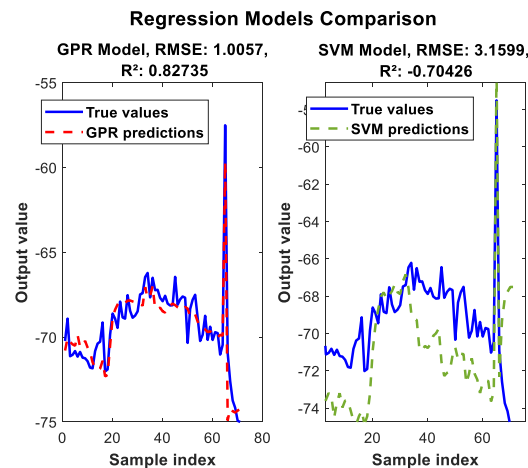
1. Trends

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 می‌شود؛ در حالی که بسته شدن بیش از حد آن می‌تواند باعث خفگی و خاموشی ناگهانی کوره شود. در اینجا می‌توان با مدل‌سازی رگرسیونی کوره، با استفاده از داده‌های آموزشی، بهترین میزان باز یا بسته بودن این شیر را (باتوجه به مصرف سوخت و بازده کوره) که به وسیله یک موقعیت‌یاب کنترل می‌شود، به 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.

۴. نتیجه‌گیری

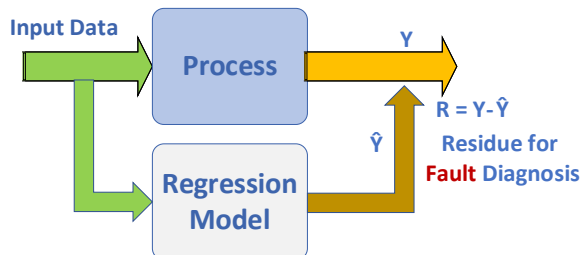
در پالایشگاه‌های گاز، فرایندهای متنوع و پیچیده‌ای وجود دارد که اغلب فاقد مدل‌های دقیق فیزیکی است. تشخیص سریع عیب در این سامانه‌ها همواره چالشی بزرگ بوده که تاکنون به روش‌های سنتی و باتکیه بر تحلیل کارشناسان باتجربه و صرف زمان زیاد انجام شده است که این روش‌ها به دلیل زمان‌بر بودن، باعث هدررفت انرژی و یا مواد با ارزش می‌شود. برای تشخیص عیب دقیق هر سامانه، نیاز به داشتن مدل دقیقی از آن سامانه است؛ این مقاله امکان مدل‌سازی رگرسیونی سامانه‌های 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، دارد. در این پژوهش، مدل‌سازی رگرسیونی سامانه‌های مرتبه اول و دوم در شرایط حضور نوفه و داده‌های ناقص یا پرت بررسی و خطاهای مرتبط تحلیل شد. همچنین، نشان داده شد که برای سامانه‌های 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 رگرسیونی فرایند گاوسی و رگرسیون ماشین بردار پشتیبان نسبت به سایر روش‌های

نم‌گیر پالایش گاز و یا موارد مشابه استفاده می‌شود. کنترل این کوره‌ها به وسیله یک شیر کنترل سوخت انجام می‌شود که بسته به میزان دمای مورد نظر، میزان سوخت ورودی را کم یا زیاد می‌کند.



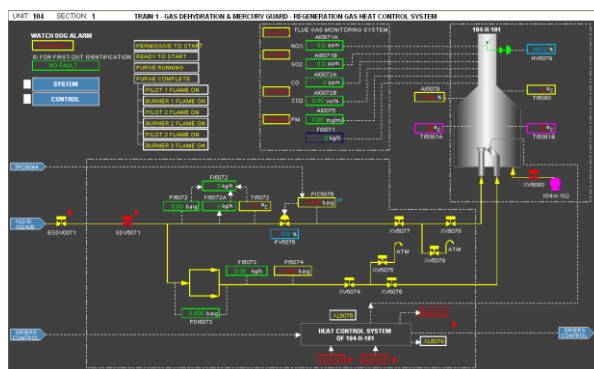
شکل ۱۵. مدل توربو اکسپندر با روش‌های رگرسیون ماشین بردار پشتیبان و فرایند گاوسی (دقت بیشینه با مدل فرایند گاوسی یک درجه سلسیوس دقت ۹۸/۷٪).

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.



شکل ۱۷. کوره با دریچه دودکش قابل تنظیم برای بهترین بازده.

Figure 17. Furnace with an adjustable chimney valve for optimal efficiency.

- [7] 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).
- [8] Donald R. Woods. (2021). *Troubleshooting Guide for Process Engineers: A Complete Course with Applied Examples and Glossary*.
- [9] 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*, 135, 877–896, Elsevier Ltd.
- [10] M. J. Khosrowjerdi, & Mahdi Aliyari-Shoorehdeli. (2014). *Fault Diagnosis and Compensation in Control Systems*. Sahand University of Technology (SUT), [In Persian].
- [11] Mohsen Maleki Douz-Zadeh, & Hamid, M. (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*, 12(66), 58-65, [In Persian].
- [12] Norman P. Lieberman - *Troubleshooting Process Plant Control_ A Practical Guide to Avoiding and Correcting Mistakes*-John Wiley & Sons (2024).
- [13] Sarmast Alizadeh, F., & Esfandiari, N. (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, [In Persian].
- [14] Sayadmosleh, E., Shirazi, L., Soleimani, M., & Sarmad, M. (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, [In Persian].
- [15] Tidriri, 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*, 42, 63-81, Elsevier Ltd.
- [16] 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*, 101(8), 4506-4522, John Wiley and Sons Inc.
- [17] 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.
- یادگیری ماشین، از جمله شبکه‌های عصبی، عمل کرد بهتری در حضور نوفه و داده‌های پرت دارند. این روش‌ها می‌توانند مدل‌های رگرسیونی دقیق و قابل‌اعتماد برای فرایندهای پیچیده ارائه کنند و از این مدل‌ها به راحتی می‌توان برای کاربردهای تشخیص عیب در فرایندهای پیچیده استفاده کرد. به عنوان کار آینده، پیشنهاد می‌شود که برنامه‌آپلیکیشن‌های نوشته‌شده که عیب‌های تشخیص داده‌شده در سامانه کنترل گسترده را از راه یک شبکه امن پیام‌رسان به اطلاع بهره‌بردار و یا رئیس واحد برساند تا عملیات تشخیص عیب با سرعت و دقت بیشتری انجام پذیرد.

۵. تقدیر و تشکر

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

مراجع

- [1] Ahmadi, S. H., & Khosrowjerdi, M. J. (2021). Fault detection Automation in Distributed Control Systems using Data-driven methods: SVM and KNN. *TechRxiv. Preprint*.
- [2] Akbarpour, S., & Khosrowjerdi, M. J. (2024). Estimation of Combustor and Compressor Faults in Industrial Gas Turbines by Using Multiple Model Approach. *Iranian Chemical Engineering Journal*, 23(134), [In Persian].
- [3] 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.
- [4] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16–28.
- [5] Cheliotis, M., Lazakis, I., & Theotokatos, G. (2020). Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering*, 216.
- [6] 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.