

به کارگیری روش‌های یادگیری ماشین جهت مدل‌سازی کاهش واردات بخار با هدف مدیریت منابع انرژی

احسان شریف‌آرا^۱، مجید عباسپور^{۲*} و علیرضا سرائی^۳

(۱) دانشجوی دکتری رشته مهندسی سیستم‌های انرژی، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.
(۲) استاد دانشگاه صنعتی شریف، دانشکده مهندسی مکانیک، تهران، ایران. * رایانامه نویسنده مسئول مکاتبات: abbpor@sharif.edu
(۳) استادیار گروه آموزشی مکانیک، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران.

تاریخ پذیرش: ۱۴۰۳/۰۲/۳۱

تاریخ دریافت: ۱۴۰۲/۱۲/۱۱

چکیده

واحدهای تولید اتیلن اکساید، اتیلن گلیکول یکپارچه، مصرف‌کنندگان مهم حامل‌های انرژی به‌ویژه بخار پرفشار هستند. در این پژوهش، چارچوب مدل‌سازی دقیق برای واردات بخار پرفشار در یک واحد پتروشیمی پیشنهاد گردید. بدین منظور از زبان برنامه‌نویسی پایتون نسخه ۳ برای آنالیز داده‌های خام کارکردی روزانه واحد پتروشیمی مورد مطالعه در بازه زمانی کارکرد ۵ ساله تحت یک نوع کاتالیست استفاده شد. گزینش‌پذیری کاتالیست و ظرفیت تولید واحد به‌عنوان شاخص‌ها و واردات بخار پرفشار به‌عنوان هدف، جهت مدل‌سازی در نظر گرفته شد. الگوریتم‌های رگرسیون متفاوتی برای گزینش بهترین مدل مورد ارزیابی قرار گرفتند و فرایارامترهای آنها با استفاده از الگوریتم‌های جستجوی شبکه‌ای و بیزین تنظیم گردیدند. روش درخت تصادفی بهترین عملکرد را در شاخص‌های ارزیابی نسبت به سایر روش‌های رگرسیون، اما با بیشترین مدت زمان پردازش به خود اختصاص داد. با در نظر گرفتن مدت زمان پردازش و عملکرد، روش رگرسیون رنج چندجمله‌ای، انتخاب مناسبی تشخیص داده شد. با استفاده از مدل رگرسیون درخت تصادفی تنظیم‌شده، میزان تقریبی ۲۹۱ تن در روز صرفه‌جویی واردات بخار پرفشار (۱۶ درصد میانگین واردات بخار) معادل با انتشار تقریبی روزانه ۳۸ تن دی‌اکسیدکربن (۰/۳۷ درصد انتشار معادل ظرفیت اسمی تولید بخار تامین‌کننده) حاصل شد. روش پیشنهادی در سیستم مدیریت انرژی واحد مورد مطالعه یکپارچه‌سازی شده است و می‌تواند در دیگر واحدهای مشابه نیز به‌منظور پایش مداوم رفتار واردات بخار پرفشار بر حسب عملکرد کاتالیست و تولیدات واحد پیاده‌سازی شود.

واژه‌های کلیدی: پیش‌بینی مصرف بخار، فرصت‌های صرفه‌جویی انرژی، مدل‌سازی انرژی، یادگیری ماشین.

مقدمه

فعالیت می‌کنند، وادار کرده است تا بهینه‌سازی مصرف سرویس‌های جانبی و تولید و افزایش بهره‌وری انرژی را بیش از پیش مورد توجه قرار دهند. در میان فرآیندهای تولید مختلف در صنعت پتروشیمی، فرآیند تولید اتیلن‌اکساید و اتیلن‌گلیکول در دسته صنایع انرژی برقرار دارند. در نتیجه، حتی بهبود جزئی در کارایی این فرآیندها می‌تواند تاثیر مثبت قابل توجهی بر پایداری آنها داشته باشد.

سرعت بالای صنعتی شدن و رشد جمعیت جهانی، باعث ایجاد چالش‌های اساسی در مقیاس جهانی به‌ویژه در بخش انرژی و آب شده است (Ahmadi et al., 2021). این چالش‌ها همراه با کاهش قابل توجه ذخایر انرژی و افزایش قابل توجه قیمت انرژی، اکثر مدیران صنایع، به‌ویژه صنعت پتروشیمی را که در تولید اتیلن‌اکساید^۱ و اتیلن‌گلیکول^۲

اجرای روش‌های مدل‌سازی جعبه سیاه و جعبه خاکستری در صنایع نفت، گاز، پتروشیمی و انرژی شامل می‌شوند و پذیرش و ادغام این فناوری‌ها به‌ویژه در صنعت پتروشیمی روند افزایشی داشته است (Mhlanga, 2021).

Rahimpour و همکاران (۲۰۱۱) برای شبیه‌سازی دینامیکی رفتار یک راکتور با بستر ثابت که برای تولید اتیلن اکساید استفاده می‌شود از یک مدل ناهمگن تک‌بعدی استفاده کردند. آنها از یک شبکه عصبی پرسپترون چندلایه^۸ برای تعیین غلظت بهینه اتیلن‌دی‌کلراید استفاده کردند. علاوه بر این، آنها بررسی کردند که چگونه سرعت گاز ورودی بر کسر مولی اتیلن‌اکساید در هر دو فاز گاز و جامد تاثیر می‌گذارد. آنها مدل خود را در یک دوره عملیاتی سه ساله با مقایسه پیش‌بینی‌های آن با شرایط صنعتی در حالت واقعی اعتبارسنجی کردند. نتایج شبیه‌سازی‌های مدل پویا و داده‌های فرآیند تاریخی، سطح رضایت‌بخشی از همخوانی را در شبیه‌سازی با عدم دقت کمتر از ۵ درصد نشان دادند.

Zahedi و همکاران (۲۰۱۱) یک مدل جعبه خاکستری^۹ پویا را برای راکتور اتیلن‌اکساید با بستر ثابت توسعه دادند. در ابتدا آنها یک مدل کینتیک برای واکنش‌ها را با استفاده از رویکرد شبکه عصبی مصنوعی ایجاد کردند. این مدل بر اساس داده‌های عملیاتی یک راکتور اتیلن‌اکساید معمولی با ورودی‌های شامل زمان، کسر مولی CO_2 ، $\text{C}_2\text{H}_4\text{O}$ ، C_2H_4 ، H_2O و O_2 و خروجی‌هایی که نشان‌دهنده حاصلضرب سرعت واکنش و غیرفعال شدن کاتالیست (r^*a) است، ایجاد کردند. پس از آزمایش الگوریتم‌های آموزشی مختلف، Zahedi و همکاران (۲۰۱۱) نشان دادند الگوریتم شبکه تابع پایه شعاعی^{۱۰} دقیق‌ترین تخمین داده‌ها را به همراه دارد. نتایج تولیدشده توسط شبکه عصبی با داده‌های پیش‌بینی نشده مطابقت نزدیکی داشت و توانایی آن را برای تعمیم نشان داد. آنها با استفاده از شبکه عصبی، (r^*a) را برآورد کردند و مدل راکتور هیبریدی را استخراج نمودند. آنها به این نتیجه رسیدند مدل جعبه خاکستری، $1/437$ برابر دقت بیشتری نسبت به مدل مکانیستیک نشان می‌دهد.

اتیلن‌اکساید به‌عنوان یک محصول میانی مهم در صنعت پتروشیمی و سومین ماده خام شیمیایی آلی مهم در میان مشتقات اتیلن به شمار می‌رود. این محصول به طیف متنوعی از کالاهای با ارزش از جمله ضد یخ، مواد شوینده، مواد دارویی و پلاستیک تبدیل می‌شود. بر اساس داده‌های سایت استاتیتسا^۱، حجم بازار جهانی اکسیداتیلن در سال ۲۰۲۲ به ۳۱/۵۸ میلیون تن رسید و پیش‌بینی می‌شود تا پایان سال ۲۰۳۰ به ۳۶/۴۵ میلیون تن برسد (Statista, 2023). علی‌رغم تلاش‌های متعدد برای معرفی راه‌حل‌های صرفه‌جویی در مصرف انرژی در این فرآیند صنعتی، روش‌های بهینه‌سازی پیشنهادی اغلب هنگام اجرا در فرآیندهای تولید واقعی با چالش‌هایی مواجه می‌شوند (Qian et al., 2016).

بخش قابل توجهی از اتیلن‌اکساید به اتیلن‌گلیکول تبدیل می‌شود. در نتیجه، واحدهای اتیلن‌اکساید/اتیلن‌گلیکول یکپارچه به‌عنوان یکی از سودآورترین واحدها در صنعت پتروشیمی به‌شمار می‌روند. محصول اولیه این واحدها که به نام مونواتیلن‌گلیکول^۲ شناخته می‌شود، در تولید الیاف پلی-استر، پلی‌اتیلن ترفتالات^۳ و همچنین ضدیخ و داروهای آرایشی و بهداشتی کاربرد دارد. با این حال، این واحدها همچنین مصرف‌کنندگان قابل توجه انرژی به‌ویژه به شکل بخار پرفشار هستند. این بخار اهداف مختلفی را در سیستم‌های بخار پیچیده این تاسیسات انجام می‌دهد. واحدهای اتیلن‌اکساید/اتیلن‌گلیکول علاوه بر این، بخار پرفشار را در سمت پوسته راکتورهای اتیلن‌اکساید خود و بخار کم‌فشار را در بخش تولید اتیلن‌گلیکول تولید می‌کنند.

در سال‌های اخیر، پیشرفت فناوری، مجموعه‌ای از اصطلاحات و ایده‌های جدید را معرفی کرده است. مفاهیم نوظهور مانند داده‌های بزرگ^۴، صنعت چهارم^۵، هوش مصنوعی^۶ و شاخه خاص آن، یادگیری ماشین^۷ در صنایع مختلف کاربرد پیدا کرده‌اند. این کاربردها طیف گسترده‌ای را نظیر پیش‌بینی نتایج مقاوم‌سازی ساختمان‌های تجاری تا

- 1 Statista
- 2 Mono Ethylene Glycol (MEG)
- 3 Polyethylene Terephthalate (PET)
- 4 Big Data
- 5 Industry 4.0
- 6 Artificial Intelligence
- 7 Machine Learning

- 8 Multi-Layer Perceptron
- 9 Gray Box Model
- 10 Radial Basis Function Network

Zendehboudi و همکاران (۲۰۱۸) تاکید کردند مدل های هیبریدی در رسیدگی به چالش های مرتبط با فرآیندهای پیچیده موثر بوده اند. آنها همچنین نمونه هایی از ترکیب های مختلف مدل های جعبه سیاه و جعبه سفید را ارائه کردند که برای کاهش پیچیدگی مدل طراحی شده بودند. علاوه بر این، آنها بر اهمیت مدل های جعبه خاکستری ترکیبی که ویژگی های تعمیم پذیری و تفسیر پذیری فیزیکی مدل های جعبه سفید را با قابلیت های کاهش پیچیدگی مدل های جعبه سیاه ترکیب می کنند، تاکید کردند. آنان نشان دادند چگونه این ادغام، فرآیند مدل سازی را ساده کرده و بهبود می بخشد.

Zendehboudi و همکاران (۲۰۱۴) در یک مطالعه جداگانه یک مدل ریاضی را معرفی کردند و از قدرت شبکه های عصبی مصنوعی برای شبیه سازی و بهینه سازی یک واحد تولید اوره استفاده کردند. مدل ریاضی آنها تعادل های بخار-مایع را برای سیستم $\text{NH}_3\text{-CO}_2\text{-H}_2\text{O-(NH}_2)_2\text{CO}$ در هر دو چارچوب ترمودینامیکی و واکنش یکپارچه می کند. ایشان پیش بینی های شبکه عصبی مصنوعی را با داده های واقعی و نتایج تولید شده توسط مدل ریاضی مقایسه کردند و سطح رضایت بخشی از توافق بین این روش ها آشکار شد. Zendehboudi و همکاران (۲۰۱۴) از طریق تحلیل حساسیت سیستماتیک، شرایط فرآیندی بهینه برای تولید اوره را شناسایی کردند و دریافتند که رویکرد شبکه عصبی توسعه یافته ابزار بسیار موثری برای مدل سازی تعادل های فازی پیچیده شامل واکنش ها در یک واحد تولید اوره صنعتی است. Moghadasi و همکاران (۲۰۲۱) یک چارچوب جامع شامل پیش پردازش داده ها و مدل سازی برای ایجاد یک مدل یادگیری ماشینی فوق العاده دقیق با هدف پیش بینی استفاده از بخار در فرآیند شیرین سازی گاز ابداع کردند. آنها طیف گسترده ای از پیشرفته ترین الگوریتم های رگرسیون را برای شناسایی راه حل بهینه متناسب با مطالعه موردی منحصر به فرد خود بررسی کردند. هدف نهایی آنها استفاده از این مدل به عنوان وسیله ای برای یافتن فرصت های بالقوه صرفه جویی در انرژی در این نوع واحدها بود. آنها برای تقویت تحقیقات خود، داده های تاریخی از سال ۲۰۱۷ تا ۲۰۱۹ را از تصفیه خانه گاز در مجتمع گاز پارس جنوبی به دست آوردند. Moghadasi و همکاران (۲۰۲۱) این داده ها را تحت یک

Zampf و Wallek (۲۰۲۲) از تکنیک های یادگیری ماشین برای ایجاد مدل های فرآیندی با استفاده از داده های استخراج شده از یک پالایشگاه استفاده کردند. سپس این مدل ها در یک فرآیند فراگیر ادغام شدند که با استفاده از روش متوالی-مدولار حل شدند. اهداف بهینه سازی بر به حداکثر رساندن حاشیه تولید و در عین حال به حداقل رساندن انتشار CO_2 متمرکز بود که منجر به یک راه حل عددی بسیار پایدار برای بهینه سازی چندهدفه بعدی گردید. مطالعه آنها شامل استفاده از هر دو مدل جعبه سیاه و جعبه خاکستری بود که با رویکرد LP به خوبی تثبیت شده مقایسه شدند. Zampf و Wallek (۲۰۲۲) به این نتیجه رسیدند ادغام دانش مهندسی با تکنیک های داده محور برای واحدهای فرآیندی، استفاده موثر از اطلاعات غیرمستقیم را ممکن می سازد و می تواند قابلیت های پیش بینی مدل های واحد را در مقایسه با مدل هایی که از این ترکیب استفاده نمی کنند، به طور قابل توجهی افزایش دهد. به عنوان یک توصیه کلی برای مدل سازی تجهیزات فرآیندهای شیمیایی، آنها از ادغام مدل های جعبه سفید در کنار مدل های جعبه سیاه برگرفته از روش های یادگیری ماشین برای دستیابی به مدل های جعبه خاکستری قوی و قابل اعتماد استفاده کردند و پیشنهاد کردند که تحقیقات بیشتر باید در حوزه توسعه و گسترش مدل های جعبه خاکستری مبتنی بر داده به ویژه در زمینه راکتورهای شیمیایی صورت پذیرد.

Zendehboudi و همکاران (۲۰۱۸) یک بررسی جامع از مدل های هیبریدی^۱ در زمینه های نفت، گاز، فرآیندهای شیمیایی و سیستم های انرژی انجام دادند. تحقیقات آنها شامل تجزیه و تحلیل مدل های فرعی، راهبردهای هیبریداسیون، طرح های ساختاری، معیارهای غربالگری و روندهای نوظهور در مدل سازی ترکیبی با تاکید خاص بر این بخش های صنعتی بود. آنها تاکید کردند ماهیت پیچیده فرآیندهای شیمیایی و انرژی، استفاده از ابزارهای ریاضی پیشرفته را به دلیل پیچیدگی های ریاضی و درک محدود این فرآیندها که شامل غیرخطی بودن، دینامیک گسترده و چندمقیاسی، بازه های زمانی طولانی، عدم قطعیت ها و ابعاد بالا را ناگزیر می سازد.

کاهش ۷،۰۲۷ تن انتشار دی‌اکسیدکربن در سال را نشان داد. آنها تاکید کردند با بهینه‌سازی حالت‌های عملیاتی واحد تحت بهترین شیوه‌ها، می‌توان به این نتایج مطلوب دست یافت.

همانطور که تحقیقات فوق نشان می‌دهند، در چند سال گذشته، تجزیه و تحلیل‌های مرتبط با پیش‌بینی مبتنی بر یادگیری ماشین توجه بسیاری را در صنعت پتروشیمی به خود جلب کرده است. با این حال، کاربرد این روش‌ها برای پیش‌بینی و بهینه‌سازی مصرف انرژی، به‌ویژه در مورد واردات بخار پرفشار در واحد‌های پتروشیمی اتیلن‌اکساید / اتیلن-گلایکول یکپارچه مورد بررسی قرار نگرفته است. این شکاف تحقیقاتی انگیزه مطالعه حاضر است که هدف آن ارزیابی یک چارچوب تجزیه و تحلیل داده‌ها و مدل‌سازی جامع با استفاده از ابزارهای یادگیری ماشین و پیش‌بینی رفتار واردات بخار فشار بالای یک واحد پتروشیمی اتیلن‌اکساید/ اتیلن‌گلایکول یکپارچه با استفاده از داده‌های عملیاتی پنج‌ساله استخراج‌شده از سیستم کنترل گسترده^۴ این واحد است. هدف بعد، استفاده از مدل تولیدشده به‌عنوان مرجعی برای واردات بخار پرفشار واحد است که متعاقباً امکان ارزیابی عملکرد واردات بخار واحد را به‌عنوان بخشی جدایی‌ناپذیر از سیستم مدیریت انرژی (ISO-50001) واحد مورد مطالعه فراهم می‌کند. این روش به نوبه خود شناسایی فرصت‌های بالقوه صرفه‌جویی انرژی را تسهیل می‌کند. همچنین، این مدل قابلیت یکپارچه‌سازی با سیستم اینترنت اشیا صنعتی^۵ واحد را دارد که منجر به هوشمند شدن آن می‌گردد. نتایج این تحقیق این پتانسیل را دارد که به‌عنوان سنگ بنای محققان و متخصصان صنعت پتروشیمی در راستای افزایش کارایی عملیاتی و عملکرد انرژی واحدهای پتروشیمی اتیلن‌اکساید/ اتیلن‌گلایکول از طریق رویکردهای داده-محور باشد.

مواد و روش‌ها

واحدها و شیمی فرآیند

فرآیند تولید اتیلن‌اکساید/ اتیلن‌گلایکول شامل دو بخش اصلی است: بخش تولید اتیلن‌اکساید (EO) و اتیلن‌گلایکول (EG). در بخش اول، اتیلن تحت اکسیداسیون مستقیم روی

روش پردازش داده چندمرحله‌ای قرار دادند که شامل تجزیه و تحلیل چندخطی و انتخاب ویژگی مبتنی بر مدل بود. در فرآیند انتخاب مدل، آنها مجموعه گسترده‌ای از تکنیک‌های رگرسیون را مورد بررسی قرار دادند. آنان مشخص کردند مدل ماشین‌های تقویت‌کننده گرادیان^۱ نسبت به رقبا بهتر عمل کرده و کمترین خطای اعتبارسنجی متقاطع^۲ را ارزیابی می‌دهد. در یک سناریوی آزمایشی، مدل آنها به دقت قابل توجه ۹۸ درصد ضریب تعیین^۳ برای پیش‌بینی مصرف بخار دست یافت. علاوه بر این، مدل آنها پتانسیل کاهش ۲ درصد در مصرف بخار را به میزان صرفه‌جویی سالانه ۲۴،۰۰۰ تن شناسایی کردند.

Moghadasi و همکاران (۲۰۲۱) در مطالعه‌ای جداگانه، یک چارچوب مبتنی بر داده‌های نوآورانه را معرفی کردند که از مدل‌های یادگیری ماشین برای ایجاد سیستم‌های مدیریت انرژی کارآمد در صنایع انرژی‌بر، در راستای استانداردهای ISO 50001:2018 استفاده می‌کرد. از طریق اجرای این چارچوب سه لایه، آنها با موفقیت الزام‌های استاندارد شامل موارد پیش‌رو را برآورده کردند: (الف) شناسایی متغیرهایی که به‌طور قابل توجهی بر عملکرد انرژی تاثیر می‌گذارند؛ (ب) توسعه شاخص‌های عملکرد انرژی و خط پایه انرژی؛ (ج) تجزیه و تحلیل عملکرد انرژی موجود؛ (د) کمی‌سازی فرصت‌های صرفه‌جویی انرژی؛ (ه) تعیین اهداف انرژی؛ و (و) تدوین یک برنامه اقدام انرژی. رویکرد مدیریت انرژی مبتنی بر داده آنها در یک واحد کربن‌زدایی اتان در مقیاس صنعتی واقع در مجتمع گاز پارس جنوبی در ایران به‌کار گرفته شد. مجموعه داده آنها شامل ۵۲،۵۶۰ رکورد از بخش‌های واحد است که در طول سال ۲۰۲۰ جمع‌آوری شده است و به‌عنوان دوره پایه عمل می‌کند. با به‌کارگیری الگوریتم ماشین‌های تقویت‌کننده گرادیان، آنان یک مدل بسیار دقیق با دقت قابل توجه ۹۹ درصد ضریب تعیین برای پیش‌بینی مصرف انرژی واحد به‌عنوان خط پایه انرژی ایجاد کردند. یافته‌های آنها پتانسیل قابل توجهی برای صرفه‌جویی در انرژی، معادل ۱۰/۳۱ درصد یا ۶۳،۱۱۹ تن مصرف بخار سالانه و

1 Gradient Boosting

2 Cross Validation

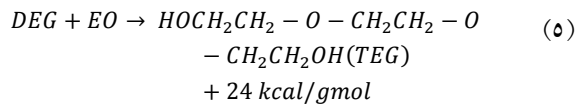
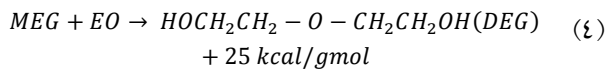
3 R-Squared (Coefficient of Determination)

4 Distributed Control System

5 Industrial Internet of Things

به کارگیری روش های یادگیری ماشین جهت مدل سازی کاهش واردات بخار با هدف مدیریت منابع انرژی / ۳۱

واکنش های جانبی رقابتی با واکنش رابطه (۳) منجر به تشکیل دی اتیلن گلیکول، تری اتیلن گلیکول و گلیکول های سنگین تر بر اساس واکنش های رابطه (۴) و (۵) می شود:



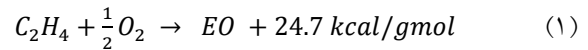
محصول راکتور گلیکول تحت یک فرآیند تبخیر سه مرحله ای قرار می گیرد تا اکثریت محتوای آب آن حذف شود و به دنبال آن آب باقی مانده در ستون آب گیری حذف می شود. در بخش خالص سازی گلیکول، گلیکول های خام استخراج شده از سیستم حذف آب به مونواتیلن گلیکول^۳، دی اتیلن گلیکول^۴، تری اتیلن گلیکول^۵ و گلیکول های سنگین تر جدا می شوند.

در این مطالعه موردی، کاربران انرژی قابل توجه حرارتی، بخار مصرف می کنند که از مرز واحد به شکل بخار پرفشار در فشار ۴۰ بار نسبی^۶ و دمای ۴۱۰ درجه سانتی گراد تامین می شود. سپس این بخار با استفاده از یک سیستم توزیع بخار پیچیده در سراسر واحد توزیع می شود (Shell EO/EG Process Design Package, 2023).

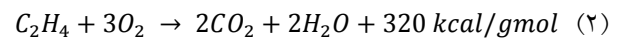
ویژگی های مرتبط با واردات بخار پرفشار

رفتار واردات بخار واحد تحت شرایط مختلف خوراک و تولید، تغییر پذیری را نشان می دهد. این تنوع را می توان از طریق مدل های رگرسیونی که خطوط پایه واردات بخار نیروگاه را تعیین می کند، تجزیه و تحلیل کرد. همان طور که قبلا در بخش شیمی فرآیند مورد بحث قرار گرفت، ورودی واحد شامل اتیلن و اکسیژن است که نتیجه آن تولید اتیلن-اکساید در راکتورهای اتیلن-اکساید واحد به عنوان یک محصول واسطه است. متعاقبا، هنگامی که این محصول میانی با آب ترکیب می شود، مونواتیلن گلیکول، دی اتیلن گلیکول و تری اتیلن گلیکول تولید می کند. هدف طراحی این واحد، تولید مداوم مقدار مشخصی گلیکول در هر روز در طول عمر

یک کاتالیست حاوی نقره بر پایه آلومینا قرار می گیرد که در دماهای معمولاً از ۲۲۰ تا ۲۷۰ درجه سانتی گراد از طریق واکنش شیمیایی رابطه (۱) رخ می دهد:



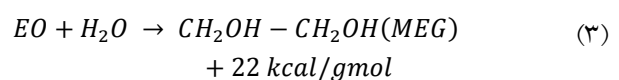
واکنش جانبی در رابطه (۲) که غیرانتخابی و بسیار گرمازا است نیز با واکنش اصلی رابطه (۱) رقابت و دی-اکسیدکربن و آب تولید می کند:



سرکوب واکنش رابطه (۲) با استفاده از اتیل کلراید^۱ انجام می شود. اتیلن اکساید خود مقاومت نسبتاً بالایی در برابر اکسیداسیون بیشتر از خود نشان می دهد. گرمای تولید شده در طی واکنش با تولید بخار فشار بالا در پوسته راکتور و بخار با فشار متوسط در خنک کننده محصول اولیه کنترل می شود. دی اکسیدکربن تولید شده به عنوان محصول جانبی در واکنش جانبی رابطه (۲) از طریق یک واکنش شیمیایی با پتاسیم-کربنات محلول در آب در بخش حذف CO₂ از سیکل گاز حذف می شود.

اتیلن اکساید تولید شده در راکتورها با جذب در آب در بخش بازیابی اتیلن اکساید از گاز محصول راکتور جدا می شود. متعاقبا، اتیلن اکساید از حلال در بخش جداسازی استخراج می شود. سیستم بازیافت ترکیبات سبک^۲ به منظور حذف CO₂ و ترکیبات سبک است که در جذب کننده اتیلن اکساید جذب شده و سپس در استریپر اتیلن اکساید خارج می شود. حذف ترکیبات سبک برای اطمینان از عملکرد طبیعی واکنش گلیکول و به حداقل رساندن خوردگی ناشی از حضور CO₂ ضروری است (Shell EO/EG Process Design Package, 2023).

در بخش واکنش اتیلن گلیکول و آب زدایی، یک واکنش هیدرولیز حرارتی فاز مایع EO در حضور آب اضافی بر اساس واکنش رابطه (۳) رخ می دهد:



3 Mono Ethylene Glycol (MEG)
4 Diethylene Glycol (DEG)
5 Tri ethylene Glycol (TEG)
6 Bar Gauge (barg)

1 Ethyle Chloride
2 Light Ends Recovery

معیارها بینش مستقیمی را در مورد وضعیت عملیاتی واحد و کارایی کلی آن ارائه می‌کنند که به نوبه خود تاثیر قابل توجهی بر واردات بخار پرفشار واحد دارد. تصمیم به گنجاندن آنها به‌عنوان ویژگی‌ها^۲ در رویکرد یادگیری ماشین، ریشه در درک عمیق نویسنده‌گان این پژوهش از عملکرد واحد مورد مطالعه و تحقیقات قبلی آنها در مورد رفتار واردات بخار فشار بالای واحد در سناریوهای مختلف تولید و بارگیری کاتالیست دارد (Sharifara et al., 2023).

روش پژوهش

ابتدا داده‌های تاریخی خام برای مطالعه موردی از سیستم کنترل گسترده واحد جمع‌آوری گردید. داده‌ها شامل واردات روزانه بخار پرفشار، خوراک اکسیژن و اتیلن و تولید گلیکول‌ها برای پنج سال کارکرد واحد تحت یک نوع کاتالیست اتیلن اکساید با فعالیت بالا^۳ بود. برای محاسبه گزینش‌پذیری روزانه، از مقادیر روزانه اتیلن و اکسیژن وارداتی و برای محاسبه ظرفیت تولید روزانه از مقادیر تولید روزانه گلیکول استفاده شد. بنابراین، مجموعه داده‌ها شامل گزینش‌پذیری و ظرفیت تولید روزانه به‌عنوان متغیرهای ویژگی و واردات روزانه بخار پرفشار به‌عنوان متغیر هدف^۴ بود. در این بخش، رویکرد و روش پژوهش برای پاک‌سازی، مدل‌سازی و ارزیابی این مجموعه داده‌ها شرح داده می‌شود. مراحل پاک‌سازی و آماده‌سازی مجموعه داده‌ها برای مدل‌سازی در بخش «پیش‌پردازش داده‌ها» و مراحل مدل‌سازی مجموعه داده‌های پاک‌سازی شده با استفاده از روش‌های یادگیری ماشین در بخش «پردازش داده‌ها» و ارزیابی عملکرد مدل‌های آموزش دیده با داده‌های آزمایشی در بخش «ارزیابی داده‌ها» شرح داده شده است. از زبان برنامه‌نویسی پایتون^۵ و کتابخانه سایکیت-لرن^۶ آن برای اعمال روش در مجموعه داده‌ها استفاده گردید. فلوجارت مراحل مربوط به این فرآیند در شکل (۱) نشان داده شده است.

عملیاتی کاتالیست اتیلن‌اکساید است. این مقدار از پیش تعیین شده به‌عنوان مبنایی برای محاسبه ظرفیت تولید واحد تحت سناریوهای عملکردی مختلف با استفاده از رابطه (۶) به کار می‌رود (Shell EO/EG Process Design Package, 2023).

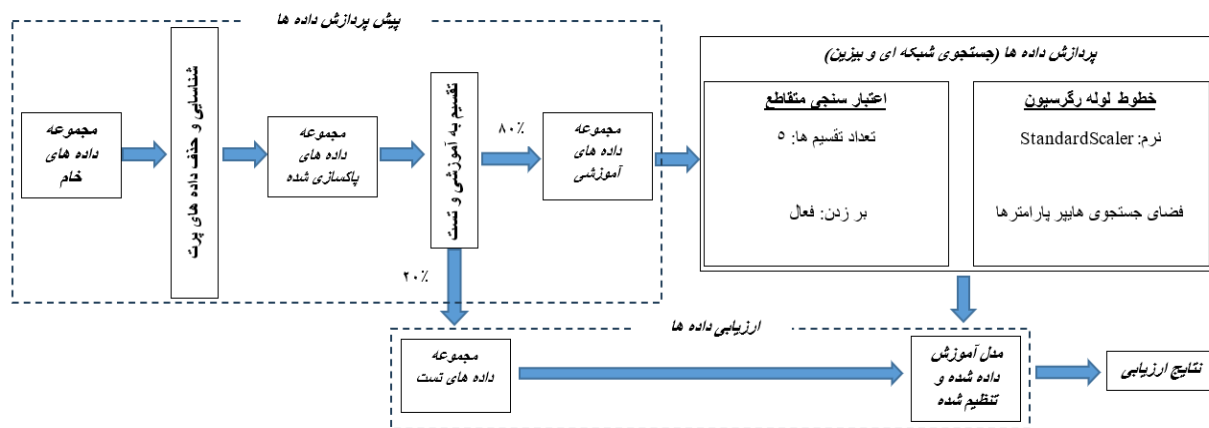
$$(6) \quad \text{ظرفیت تولید (\%)} = \frac{\text{تولید واقعی}}{\text{تولید طراحی}} * 100$$

برای در نظر گرفتن تاثیر مقدار مصرف خوراک بر واردات بخار پرفشار واحد، می‌توان از پارامتر گزینش‌پذیری^۱ کاتالیست استفاده کرد. این پارامتر نشان‌دهنده نسبت مقدار اکسیژن و اتیلن است که در راکتورهای اتیلن‌اکساید واکنش می‌دهد. این شاخص ارزشمند برای سنجش عملکرد راکتورهای اتیلن‌اکساید و در نتیجه عملکرد کلی واحد به‌شمار می‌رود و تسلط واکنش اصلی را در راکتورهای اتیلن‌اکساید بر واکنش جانبی غیرانتخابی بسیار گرمازا به‌طور کمی نشان می‌دهد. محاسبه گزینش‌پذیری را می‌توان با استفاده از مقدار خوراک اکسیژن و اتیلن ورودی به واحد با استفاده از رابطه (۷) انجام داد (Shell EO/EG Process Design Package, 2023).

$$(7) \quad \text{گزینش‌پذیری (\%)} = \left(1.2 - 0.35 \frac{\text{مقدار تن اکسیژن واکنش داده}}{\text{مقدار تن اتیلن واکنش داده}} \right) * 100$$

در معادله داده شده، ثابت‌های ۱/۲ و ۰/۳۵ نقش مهمی ایفا می‌کنند، زیرا ارتباط خاصی بین نسبت واکنش‌دهنده‌ها و گزینش‌پذیری حاصل برقرار می‌کنند. این ثابت‌ها از بینش‌های نظری در مورد مکانیسم‌های واکنش تثبیت‌شده و سینتیک‌های حاکم بر تولید اتیلن‌اکساید مشتق شده‌اند. ذکر این نکته ضروری است که این معادله بر اساس اندازه‌گیری اکسیژن و اتیلن واکنش نشان داده است. در این مطالعه موردی خاص به‌دلیل استفاده از خوراک با خلوص بالا، این موارد را می‌توان با نرخ‌های خوراک ورودی واقعی تقریب زد (Shell EO/EG Process Design Package, 2023).

گزینش‌پذیری و ظرفیت تولید به‌عنوان دقیق‌ترین معیارها برای ارزیابی عملکرد کاتالیست و خوراک واحد و شرایط مربوط به تولید گلایکول به‌شمار می‌روند. در نتیجه، این



شکل ۱. فلوچارت مراحل مدل سازی

پیش پردازش داده ها

مرتبط با ظرفیت های تولید زیر ۷۰ درصد نیز از مجموعه داده ها حذف گردید. این تصمیم بر اساس مستندات واحد است که بیان می کند نسبت کاهش تولید^۳ واحد ۵۰ درصد ظرفیت تولید است و همچنین سابقه کارکردی واحد که نشان می دهد واحد مورد مطالعه معمولاً بالای ۷۰ درصد ظرفیت تولید کار می کند. بنابراین، ظرفیت های تولید روزانه کمتر به - عنوان فازهای انتقالی در نظر گرفته شد که طی آن واحد به سمت ظرفیت های تولید بالای ۷۰ درصد حرکت می کند. مصرف بخار پرفشار در طی این فازهای انتقالی، رفتار مصرف بخار نرمال واحد را منعکس نمی کند، بنابراین برای تمرکز بر سناریوهای پایدار، داده های مربوط به این ظرفیت های تولید پایین تر در مرحله پیش پردازش داده ها حذف گردید.

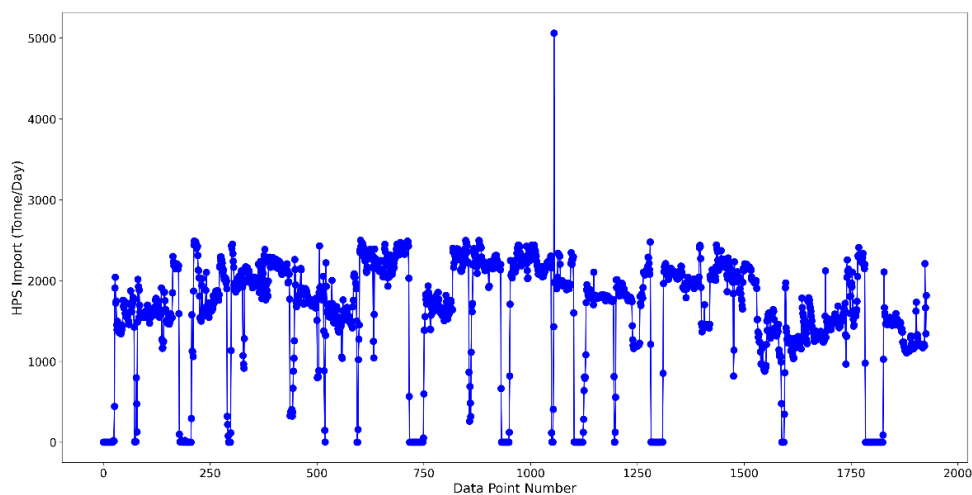
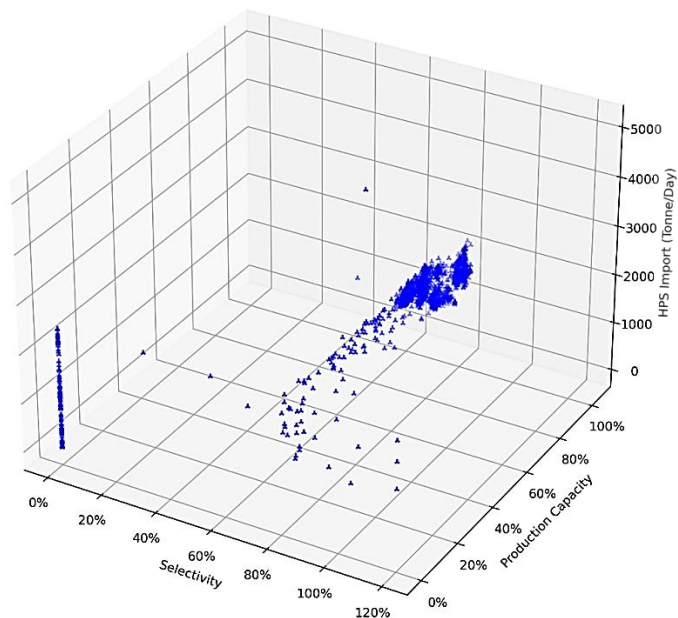
در مجموع ۶۴۸ نقطه داده به عنوان نقاط پرت شناسایی شد و از مجموعه داده های خام حذف گردید. همان طور که در جدول (۱) نشان داده شده است، حذف نقاط پرت به طور قابل توجهی انحراف استاندارد را برای همه زیرمجموعه های داده کاهش داد. تعداد زیاد نقاط داده در مجموعه داده ها، این امکان را فراهم آورد تا این نقاط پرت به راحتی حذف گردد. مجموعه داده های تمیز شده در شکل (۳) نشان داده شده است.

داده های صنعتی اغلب دارای نقاط پرت^۱، نادرست یا ناقص هستند. هنگامی که این داده ها برای تجزیه و تحلیل یا یادگیری ماشین استفاده می شوند، می توانند منجر به نتایج گمراه کننده شوند. برای جلوگیری از این امر، مهم است که داده ها پیش از پردازش، از نقاط پرت پاکسازی و ساختاردهی شوند. ترسیم داده های خام می تواند در شناسایی مشکلات بالقوه، مانند نقاط پرت یاری رسان باشد. در شکل (۲)، مشاهده می گردد که مجموعه داده حاوی مقادیر پرت بسیاری است که پیش از ادامه مرحله پردازش داده ها باید حذف می شدند.

پاکسازی داده های پرت

برای حذف نقاط پرت از مجموعه داده خام، از روش محدوده بین چارکی^۲ برای شناسایی نقاط پرت در هر یک از زیرمجموعه های داده های گزینش پذیری، ظرفیت تولید و واردات بخار پرفشار استفاده شد. برای روش IQR، چارک اول و سوم به ترتیب ۲۵ و ۷۵ درصد و آستانه ۱/۵ در نظر گرفته شد. سپس اتحاد شاخص های داده های پرت شناسایی شده از هر زیرمجموعه داده از مجموعه کلی داده خام اصلی حذف گردید.

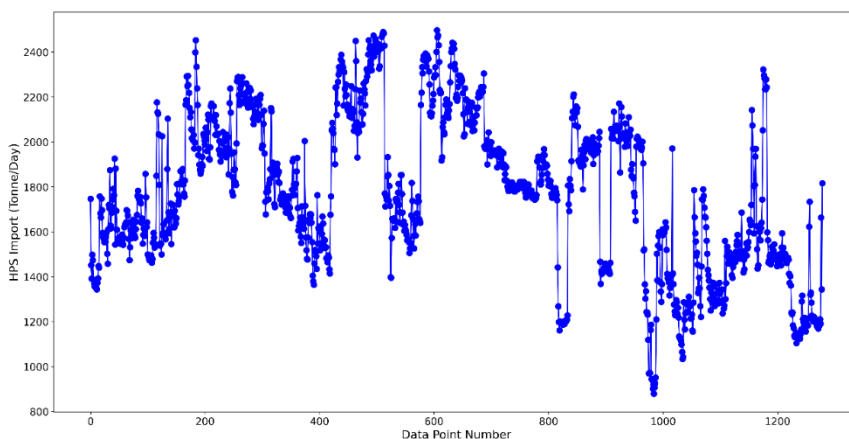
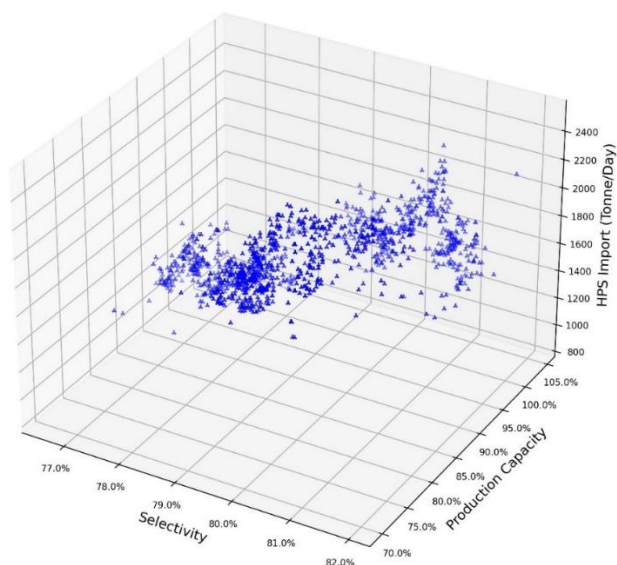
هدف این پژوهش، ایجاد بهترین مدل برای پیش بینی واردات بخار و شناسایی فرصت های صرفه جویی در شرایط عملیاتی پایدار است. برای دستیابی به این هدف، داده های



شکل ۲. داده‌های خام واردات بخار پرفشار واحد (بالا) و تغییرات واردات بخار پرفشار (پایین)

جدول ۱. توضیحات مجموعه داده‌های اصلی و پاک‌سازی شده

داده‌های پاک‌سازی شده			مجموعه داده‌های اصلی			
بخار پرفشار	ظرفیت تولید	گزینش پذیری	بخار پرفشار	ظرفیت تولید	گزینش پذیری	
۱۲۷۹	۱۲۷۹	۱۲۷۹	۱۹۲۷/۰۰	۱۹۲۷/۰۰	۱۹۲۷/۰۰	تعداد
۱۷۷۴/۸۸	۰/۸۸	۰/۸	۱۵۹۶/۰۵	۰/۶۷	۰/۶۶	میانگین
۳۴۳/۹۴	۰/۱	۰/۰۱	۷۰۴/۷۸	۰/۳۵	۰/۳۰	انحراف معیار
۸۷۸/۳۱	۰/۷	۰/۷۷	۰/۰۰	۰/۰۰	۰/۰۰	کمینه
۱۵۱۸/۷۶	۰/۷۸	۰/۷۸	۱۳۷۲/۹۱	۰/۶۵	۰/۷۸	%۲۵
۱۷۷۳/۸۱	۰/۸۸	۰/۸	۱۷۶۴/۱۹	۰/۷۸	۰/۸۰	%۵۰
۲۰۳۸/۲۲	۰/۹۸	۰/۸۱	۲۱۲۱/۸۸	۰/۹۳	۰/۸۱	%۷۵
۲۴۹۶/۸۹	۱/۰۴	۰/۸۲	۵۰۶۰/۵۶	۱/۰۴	۱/۲۰	بیشینه



شکل ۳. مجموعه داده های پاک سازی شده واردات بخار واحد (بالا) و تغییرات واردات بخار (پایین)

آنالیز همبستگی^۱

پس از پاک سازی مجموعه داده ها، ضرایب همبستگی پیرسون^۲ و اسپیرمن^۳ برای درک بهتر روابط بین متغیرها محاسبه گردید. ضریب همبستگی پیرسون اندازه گیری قدرت رابطه خطی بین دو متغیر است. ضریب اسپیرمن از سوی دیگر معیاری از رابطه یکنواخت بین دو متغیر است و زمانی استفاده می شود که توزیع داده ها ضریب همبستگی پیرسون را نامطلوب یا گمراه کننده کند (Hauke & Kossowski, 2011). نتایج آنالیز همبستگی برای مجموعه داده پاک شده در جداول (۲) و (۳) نشان داده شده است.

1 Correlation Analysis
2 Pearson
3 Spearman

جدول ۲. ضرایب همبستگی پیرسون برای داده‌های پاک‌سازی شده

پیرسون	گزینش پذیری	ظرفیت تولید	بخار پرفشار
گزینش پذیری	۱/۰۰	۰/۰۵	۰/۶۳
ظرفیت تولید	۰/۰۵	۱/۰۰	-۰/۵۸
بخار پرفشار	۰/۶۳	-۰/۵۸	۱/۰۰

جدول ۳. ضرایب همبستگی اسپیرمن برای داده‌های پاک‌سازی شده

اسپیرمن	گزینش پذیری	ظرفیت تولید	بخار پرفشار
گزینش پذیری	۱/۰۰	۰/۱۷	۰/۵۳
ظرفیت تولید	۰/۱۷	۱/۰۰	-۰/۵۶
بخار پرفشار	۰/۵۳	-۰/۵۶	۱/۰۰

جداسازی داده‌های آموزشی - آزمایشی

سپس مجموعه داده‌های پاک‌سازی شده به دو مجموعه تقسیم شدند: یک مجموعه آموزشی (۸۰ درصد) و یک مجموعه آزمایشی (۲۰ درصد). مجموعه آزمایشی برای ارزیابی خطای تعمیم مدل و انجام ارزیابی نهایی مدل برای ارزیابی عملکرد مدل بر روی داده‌های جدیدی که قبلاً هرگز ندیده است کنار گذاشته شد. جدول (۴) توضیحات مجموعه داده‌ها را برای مجموعه‌های آموزشی و آزمایشی نشان می‌دهد.

همان‌طور که ضرایب به وضوح در جداول (۲) و (۳) نشان می‌دهد، بین بخار پرفشار و گزینش‌پذیری و بخار پرفشار و ظرفیت تولید همبستگی خوبی وجود دارد، اما بین پارامترهای گزینش‌پذیری و ظرفیت تولید همبستگی بسیار کم وجود دارد. هر دو ضرایب پیرسون و اسپیرمن برای ظرفیت تولید - بخار پرفشار، منفی و برای گزینش‌پذیری - بخار پرفشار مثبت بودند که نشان‌دهنده همبستگی معکوس بین بخار پرفشار و ظرفیت تولید و یک رابطه مستقیم بین بخار پرفشار و انتخاب‌پذیری است.

جدول ۴. توضیحات مجموعه داده‌های آموزشی و آزمایشی

تعداد	مجموعه داده‌های آموزشی			مجموعه داده‌های آزمایشی		
	گزینش‌پذیری	ظرفیت تولید	بخار پرفشار	گزینش‌پذیری	ظرفیت تولید	بخار پرفشار
۱۰۲۳/۰۰	۱۰۲۳/۰۰	۱۰۲۳/۰۰	۲۵۶/۰۰	۲۵۶/۰۰	۲۵۶/۰۰	۲۵۶/۰۰
میانگین	۰/۸۰	۰/۸۸	۱۷۷۸/۸۴	۰/۸۰	۰/۸۹	۱۷۵۹/۰۵
STD	۰/۰۱	۰/۱۰	۳۴۲/۷۹	۰/۰۱	۰/۱۰	۳۴۸/۷۰
کمینه	۰/۷۷	۰/۷۰	۸۷۸/۳۱	۰/۷۷	۰/۷۰	۹۰۰/۷۸
%۲۵	۰/۷۹	۰/۷۸	۱۵۲۶/۵۴	۰/۷۸	۰/۷۹	۱۴۸۹/۵۸
%۵۰	۰/۸۰	۰/۸۸	۱۷۷۹/۴۱	۰/۸۰	۰/۸۹	۱۷۴۷/۲۰
%۷۵	۰/۸۱	۰/۹۸	۲۰۴۳/۶۶	۰/۸۱	۰/۹۸	۲۰۲۴/۳۹
بیشینه	۰/۸۲	۱/۰۴	۲۴۹۶/۸۹	۰/۸۲	۱/۰۴	۲۴۷۳/۷۲

استانداردسازی

استاندارد ۱ داشته باشند. این کار با کم کردن میانگین از هر نقطه داده و سپس تقسیم بر انحراف استاندارد انجام می‌شود. استانداردسازی می‌تواند به مقایسه بیشتر ویژگی‌های مجموعه داده با یکدیگر کمک کند و یادگیری رابطه بین ویژگی‌ها و

استانداردسازی یک روش پیش‌پردازش داده است که می‌تواند برای بهبود عملکرد الگوریتم‌های یادگیری ماشین استفاده شود. این روش یک نوع تبدیل داده است که داده‌ها را به گونه‌ای مقیاس‌بندی می‌کند که میانگین صفر و انحراف

روش های یادگیری ماشین و تنظیم فرآیندها

برای یافتن بهترین روش یادگیری ماشین برای مدل سازی داده های پاک سازی شده، روش های یادگیری ماشین زیر در نظر گرفته شد. رگرسیون حداقل مربعات معمولی خطی^۳، رگرسیون چندجمله ای ریب^۴، رگرسیون بردار پشتیبان^۵، رگرسیون تقویت کننده گرایان^۶ و رگرسیون جنگل تصادفی^۷. تمامی روش های یادگیری ماشین ذکر شده به جز رگرسیون خطی حداقل مربعات معمولی دارای فرآیندهایی هستند که باید برای بهینه سازی مدل و دستیابی به بهترین نتایج تنظیم شوند. فرآیندها پارامترهای مدل یادگیری ماشین هستند که نمی توان از داده ها یاد گرفت. آنها باید به صورت دستی یا با استفاده از روش های مختلف از جمله جستجوی شبکه ای^۸ یا بیزین^۹ تنظیم شوند.

برای تنظیم فرآیندهای خطوط لوله مدل سازی، روش جستجوی شبکه ای برای روش رگرسیون چندجمله ای ریب انتخاب گردید، زیرا یک روش نسبتاً ساده با تنها دو فرآیند است. روش جستجوی بیزین برای سایر روش های یادگیری ماشین انتخاب شد، زیرا آنها مدل های پیچیده تری با تعداد فرآیندهای بیشتری هستند. از روش های تنظیم فرآیند و اعتبارسنجی متقابل برای یافتن بهترین مقادیر فرآیندها با استفاده از کلاس های GridSearchCV و BayesSearchCV در کتابخانه Scikit-Learn استفاده گردید. فضاهای جستجوی فرآیند و مقادیر بهینه برای روش های یادگیری ماشینی فوق به تفصیل در بخش نتایج و بحث مورد اشاره قرار گرفته اند.

جهت پایبندی به API امتیازدهی استاندارد شده معرفی شده در Scikit-learn، هر دو الگوریتم جستجو از پارامتر امتیازدهی میانگین مربعات خطای منفی^{۱۰} استفاده می کنند. اما حتی برای معیارهایی مانند MSE که باید به حداقل برسد، این مدل ها برای به حداکثر رساندن معیارهای عملکرد طراحی شده اند. در نتیجه، از آنجایی که به حداقل رساندن MSE معادل به

متغیر هدف را برای الگوریتم آسان تر کند. استاندارد سازی همچنین می تواند با سخت تر کردن شرایط منجر به جلوگیری از برازش بیش از حد مدل روی داده های آموزشی شود. بنابراین در این پژوهش، تمام خطوط لوله^۱ مدل سازی شامل کلاس StandardScaler می باشند که کلاسی در کتابخانه Scikit-Learn نرم افزار پایتون برای استاندارد سازی ویژگی های مجموعه داده ها قبل از مرحله پردازش داده ها است.

پردازش داده ها و آنالیز رگرسیون روی مجموعه داده های

واردات بخار پاک سازی شده

اعتبارسنجی مقاطع K لایه^۲

اثر بخشی مدل های یادگیری ماشین آموزش دیده و عملکرد آنها باید قبل از استقرار به طور کامل ارزیابی شود. در این راستا، اعتبارسنجی مقاطع K لایه یک تکنیک ارزشمند است. در این تکنیک، مجموعه داده های آموزشی به طور تصادفی به گروه های k ، به نام لایه ها با اندازه تقریباً مساوی تقسیم می شوند. سپس در تقسیم های k ، یک لایه به عنوان مجموعه اعتبارسنجی عمل می کند، در حالی که مدل یادگیری ماشین روی لایه های $k-1$ باقی مانده آموزش داده می شود. سپس عملکرد مدل با میانگین گیری عملکرد آن در تمام لایه ها محاسبه می شود (James et al., 2013).

با آموزش و ارزیابی مدل بر روی چندین زیرمجموعه متنوع از داده های آموزشی، درک عمیق تری از عملکرد آن بر روی داده های دیده نشده به دست می آید. این اطلاعات ارزشمند می تواند برای بهینه سازی فرآیندهای مدل مورد استفاده قرار گیرد، در نتیجه قابلیت های تعمیم آن را افزایش داده و آن را قادر می سازد تا با در نظر گرفتن مبادله بایاس-واریانس، به طور موثر بر روی داده های دیده نشده قبلی عمل کند. در این مطالعه از اعتبارسنجی مقاطع با تعداد لایه ۵ برای ارزیابی عملکرد خطوط لوله مدل سازی در تقسیم بندی های مختلف داده های آموزشی استفاده گردید.

3 Ordinary Least Squares Linear Regression

4 Polynomial Ridge Regression

5 Support Vector Regression

6 Gradient Boosting Regression

7 Random Forest Regression

8 Grid Search

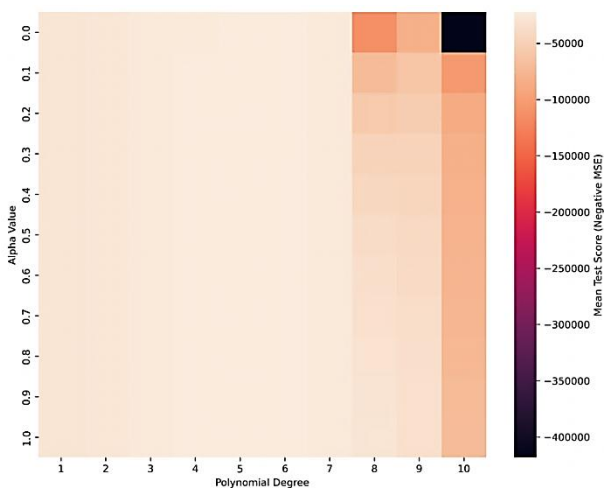
9 Bayesian Search

10 MSE

1 Pipeline

2 K-Fold Cross Validation

نشان داده شده‌اند که نشان‌دهنده اثربخشی آنها در افزایش عملکرد پیش‌بینی مدل رگرسیون چندجمله‌ای ریج است. شکل‌های (۵)، (۶) و (۷) بهبود تدریجی میانگین نمرات آزمون را در هر تکرار از الگوریتم جستجوی بیزین نشان می‌دهند و کارایی الگوریتم را در دستیابی به فرآیندهای بهینه در فضای جستجوی مشخص شده برای همه روش‌های رگرسیون به‌کار گرفته شده برجسته می‌کنند. الگوریتم جستجوی بیزین به خوبی در حداکثر تکرارهای مجاز ۱۰۰ برای همه مدل‌ها به همگرایی دست یافت. فرآیندهای بهینه شناسایی شده در فضاهای جستجوی مدل‌ها با استفاده از الگوریتم جستجوی بیزین در جداول (۶)، (۷) و (۸) آورده شده‌اند.



شکل ۴. نقشه حرارتی میانگین امتیاز آزمون^۴ برای روش رگرسیون چندجمله‌ای ریج

جدول ۵. فضای جستجوی فرآیندها و مقادیر تنظیم شده برای روش

رگرسیون چندجمله‌ای ریج		
مقدار تنظیم شده	فضای جستجو	فرآیندها
۵	۱-۱۰	درجه چندجمله‌ای
۰/۱	۱-۰	آلفا

حداکثر رساندن MSE منفی است، مقادیر MSE منفی می‌شوند.

خطوط لوله پس از تنظیم، با کل مجموعه داده آموزشی، آموزش داده شدند. سپس مدل‌های به‌دست آمده روی مجموعه داده‌های آزمایشی با استفاده از معیارهای مناسب، ارزیابی شدند. این معیارهای ارزیابی با جزئیات بیشتری در زیربخش ارزیابی مدل‌ها مورد بحث قرار گرفتند.

ارزیابی داده‌ها

ارزیابی داده‌ها مرحله نهایی یادگیری ماشینی است که در آن عملکرد یک مدل آموزش‌دیده بر روی داده‌های دیده‌نشده ارزیابی می‌شود. این کار برای تعیین میزان تعمیم مدل به داده‌های جدید انجام می‌شود. معیارهای مختلفی را می‌توان برای ارزیابی عملکرد یک مدل یادگیری ماشین، بسته به نوع مشکلی که در تلاش برای حل آن است، استفاده کرد. علاوه بر مبادله بایاس/واریانس، که با استفاده از اعتبارسنجی متقاطع در مرحله مدل‌سازی در مجموعه داده آموزشی به آن پرداخته شد، مدل‌های آموزش‌دیده نهایی نیز با استفاده از پارامترهای ضریب تعیین^۱، ریشه میانگین مجذور خطا^۲ و میانگین درصد خطای مطلق^۳ ارزیابی شدند.

نتایج

نتایج پردازش داده‌ها

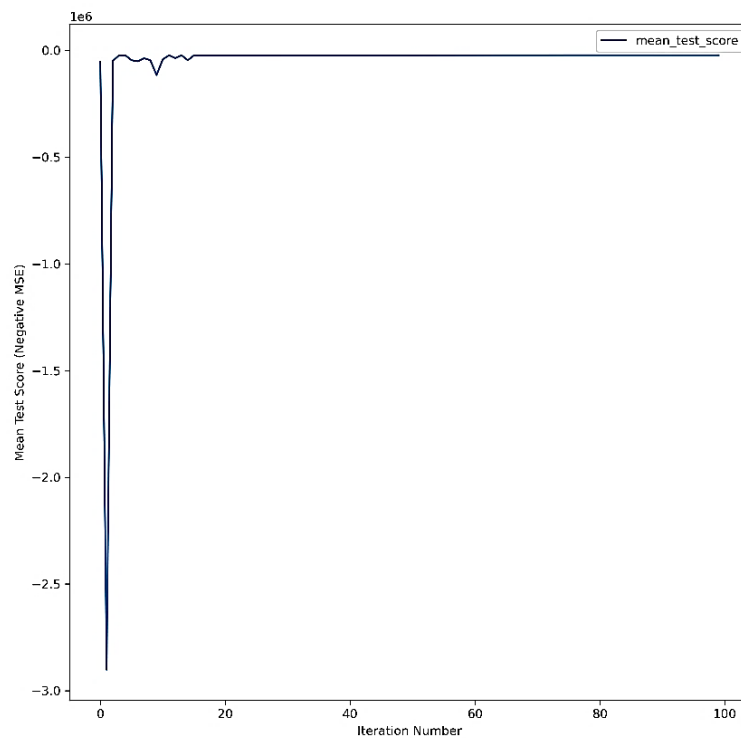
در شکل (۴) فضای جستجوی شبکه‌ای درجه چندجمله‌ای و مقدار آلفا در قالب یک نقشه حرارتی از میانگین نمرات آزمون برای روش رگرسیون چندجمله‌ای ریج نشان داده شده است که در آن رنگ‌های روشن‌تر نشان‌دهنده میانگین نمرات آزمون بالاتر و رنگ‌های تیره‌تر نشان‌دهنده میانگین نمرات آزمون پایین‌تر است. این نقشه حرارتی به‌طور موثر مناطقی از فضای جستجوی شبکه را برجسته می‌کند که عملکرد برتر را برای روش رگرسیون چندجمله‌ای ریج نشان می‌دهند. مقادیر فرآیندهای بهینه شناسایی شده با استفاده از این رویکرد برای تولید بالاترین میانگین نمرات آزمون در جدول (۵)

1 Coefficient of Determination (R²)

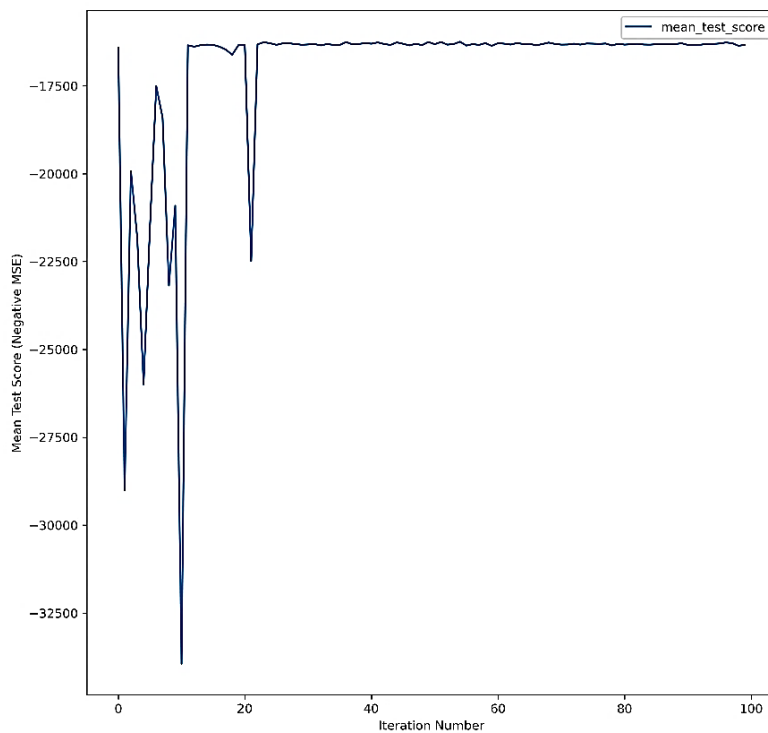
2 Root Mean Squared Error (RMSE)

3 Mean Absolute Percentage Error (MAPE)

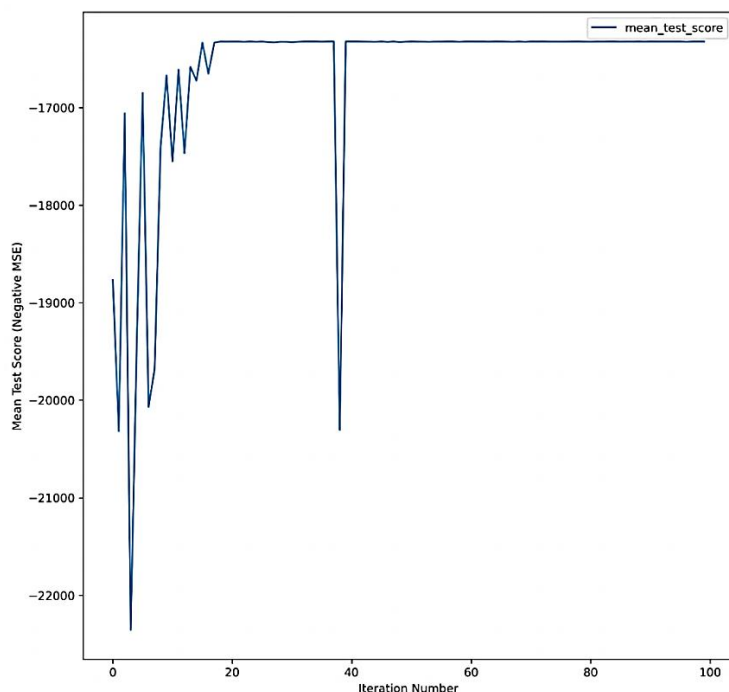
4 Mean Test Score



شکل ۵. تکامل میانگین امتیاز آزمون طی فرآیند بهینه‌سازی فرایارامترهای SVR



شکل ۶. تکامل میانگین امتیاز آزمون طی فرآیند بهینه‌سازی فرایارامترهای رگرسیون تقویت‌کننده گرادیان



شکل ۷. تکامل میانگین امتیاز آزمون طی فرآیند بهینه‌سازی فرآپارامترهای رگرسیون جنگل تصادفی

جدول ۶. فضای جستجوی فرآپارامترها و مقادیر تنظیم‌شده برای روش رگرسیون بردار پشتیبان

مقدار تنظیم‌شده	فضای جستجو	فرآپارامتر
۴۱۸/۷۶	۰/۰۱-۶۰۰	C
۰/۹۵	۰/۰۰۵-۱	اپسیلون
RBF	Linear, Poly, RBF, Sigmoid	کرنل
Scale	Scale, Auto	گاما

جدول ۷. فضای جستجوی فرآپارامترها و مقادیر تنظیم‌شده برای روش رگرسیون تقویت‌کننده گرادینت

مقدار تنظیم‌شده	فضای جستجو	فرآپارامتر
۳۲۱	۱۰-۵۰۰	تعداد تخمین‌گرها
۰/۰۱۰۷	۰/۰۱-۱	نرخ یادگیری
۶	۲-۲۰	عمق بیشینه
۱۰	۲-۲۰	کمینه جداسازی نمونه‌ها

جدول ۸. فضای جستجوی فرآپارامترها و مقادیر تنظیم‌شده برای روش رگرسیون جنگل تصادفی

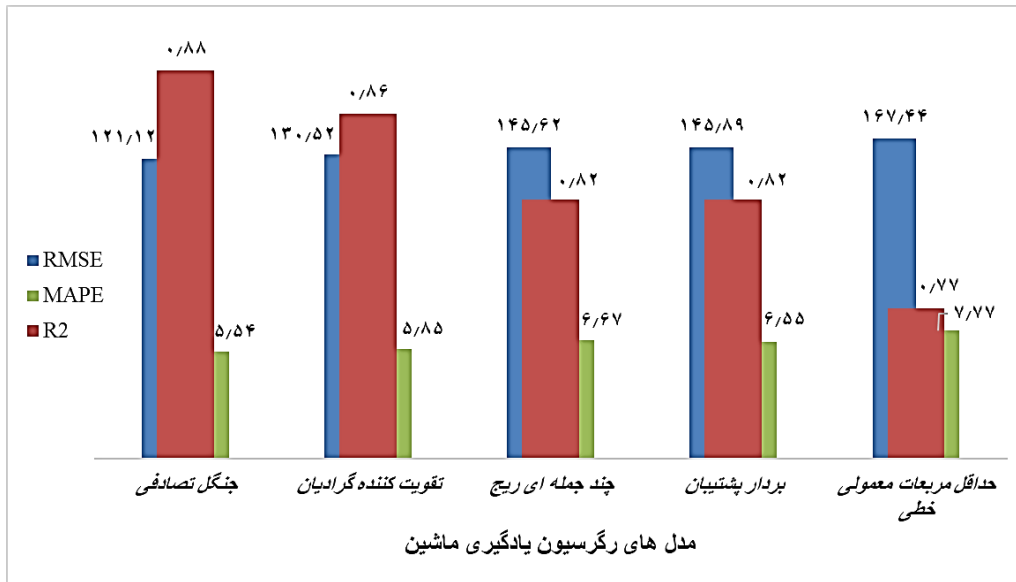
مقدار تنظیم‌شده	فضای جستجو	فرآپارامتر
۹۲۰	۱۰-۱۰۰۰	تعداد تخمین‌گرها
۱۷	۱-۲۰	بیشینه عمق
۲	۲-۲۰	کمینه جداسازی نمونه‌ها
۵	۱-۲۰	کمینه نمونه برگ
Log2	Sqrt, Log2, None	بیشینه تعداد ویژگی‌ها
False	True, False	بوت استرپینگ ^۱

ارزیابی فرصت های صرفه جویی انرژی

برای شناسایی فرصت های صرفه جویی در مصرف انرژی در مجموعه داده های واردات بخار پرفشار، از مدل رگرسیون جنگل تصادفی تنظیم شده استفاده شد، زیرا بالاترین کارایی را در بین پنج مدل رگرسیون ارزیابی شده نشان داد. سپس تفاوت بین هر مقدار واردات بخار پرفشار و مقدار پیش بینی شده آن توسط مدل رگرسیون جنگل تصادفی محاسبه گردید. سپس، مقادیر واردات بخار پرفشار با درصد تغییری بیش از ۵ درصد از مقادیر پیش بینی شده آنها استخراج شد و به عنوان مقادیر واردات بخار پرفشار غیرکارآمد طبقه بندی گردید. به طور مشابه، آنهایی که درصد تغییر کمتر از ۵- درصد از مقادیر پیش بینی شده را داشتند، به عنوان مقادیر واردات بخار پرفشار کارآمد استخراج شدند. سپس میانگین تفاوت های واردات بخار پرفشار برای مجموعه داده های وارداتی بخار پرفشار غیرکارآمد و کارآمد به مقدار ۱۵۱/۱۱ و ۱۴۰/۱۱- تن در روز محاسبه گردید. با جمع کردن مقدار مطلق این دو میانگین، فرصت صرفه جویی در واردات بخار پرفشار برابر با ۲۹۱/۲۲ تن در روز معادل ۱۶ درصد از میانگین واردات بخار پرفشار واحد با توجه به توضیحات داده پاک سازی شده (جدول ۱) محاسبه شد. اگر این فرصت صرفه جویی در بخار پرفشار محقق شود، منجر به جلوگیری از انتشار تقریباً ۳۸ تن CO_2 در روز توسط تامین کننده بخار پرفشار معادل ۰/۳۷ درصد از ظرفیت اسمی تولید بخار پرفشار تامین کننده خواهد شد. این صرفه جویی را می توان با مدیریت بهتر واردات بخار واحد مورد مطالعه و نزدیک کردن مقادیر واقعی واردات بخار پرفشار به مقادیر کارآمد به دست آورد. مقدار ۵ درصد به عنوان حد پذیرش تغییر درصد قابل قبول انتخاب گردید تا محدودیت های کلی مدل های یادگیری ماشین در نظر گرفته شود.

همان طور که در شکل (۸) و جدول (۹) نشان داده شده است، روش های مدل سازی مجموعه ای^۱، مانند رگرسیون جنگل تصادفی و رگرسیون تقویت کننده گرادیان، در هر سه معیار عملکرد (R^2 ، RMSE و MAPE) از سایر تکنیک ها بهتر عمل کردند. در بین مدل های رگرسیونی، مدل رگرسیون جنگل تصادفی بهترین و مدل رگرسیون خطی حداقل مربعات معمولی بدترین عملکرد را داشتند. عملکرد مدل های رگرسیون چندجمله ای ریح و رگرسیون بردار پشتیبان درجه ای از شباهت را نشان دادند. مقایسه عملکرد بهترین و بدترین مدل های یادگیری ماشینی در شکل (۹) با استفاده از ۵۰ نقطه نمونه تصادفی از مجموعه داده های آزمایشی نشان داده شده است.

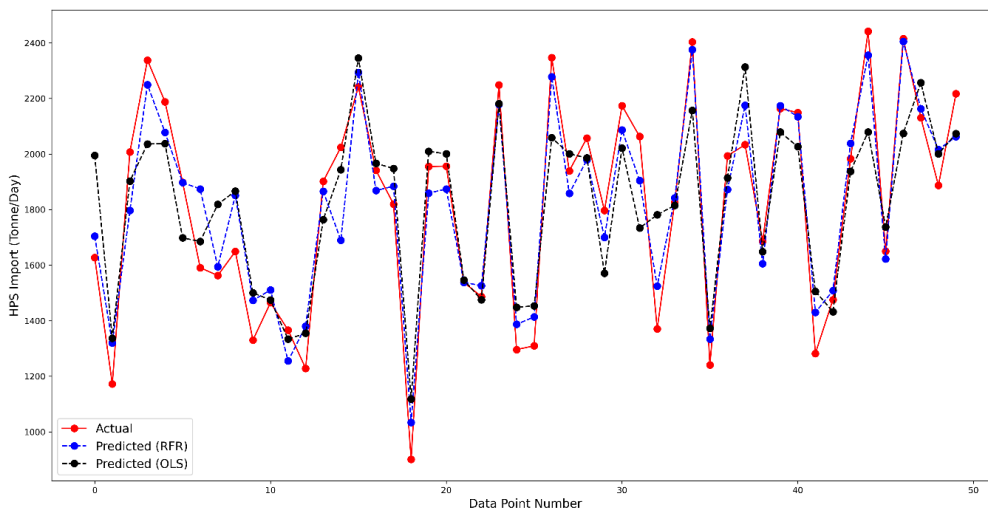
همانطور که در شکل (۸) و جدول (۹) نشان داده شده است، رگرسیون جنگل تصادفی به عنوان روش با بهترین عملکرد در بین روش های رگرسیون در نظر گرفته شده، ظاهر شد و بهترین امتیازها را در هر سه معیار عملکرد به دست آورد. با این حال، همان طور که در شکل (۱۰) و جدول (۹) نشان داده شده است، همچنین طولانی ترین زمان آموزش را نیز به خود اختصاص داد و به دنبال آن رگرسیون تقویت کننده گرادیان، تکنیک مدل سازی مجموعه ای دیگر بهترین عملکرد و بالاترین زمان آموزش را داشت. در مقابل، مدل های رگرسیون چندجمله ای ریح و رگرسیون خطی حداقل مربعات معمولی زمان های آموزش بسیار سریع تری را در مقایسه با مدل های مجموعه ای و مدل رگرسیون بردار پشتیبان نشان دادند. با این وجود، رگرسیون خطی حداقل مربعات معمولی بدترین عملکرد را نیز در سه معیار نشان داد. برعکس، رگرسیون چندجمله ای ریح عملکرد قابل قبولی را نشان داد که قابل مقایسه با تکنیک های مدل سازی مجموعه ای و کاملاً مشابه با مدل رگرسیون بردار پشتیبان بود. بنابراین در سناریوهایی که مدت زمان اجرا یک محدودیت حیاتی محسوب می شود، روش رگرسیون چندجمله ای ریح می تواند جایگزین مناسبی برای رگرسیون جنگل تصادفی در نظر گرفته شود.



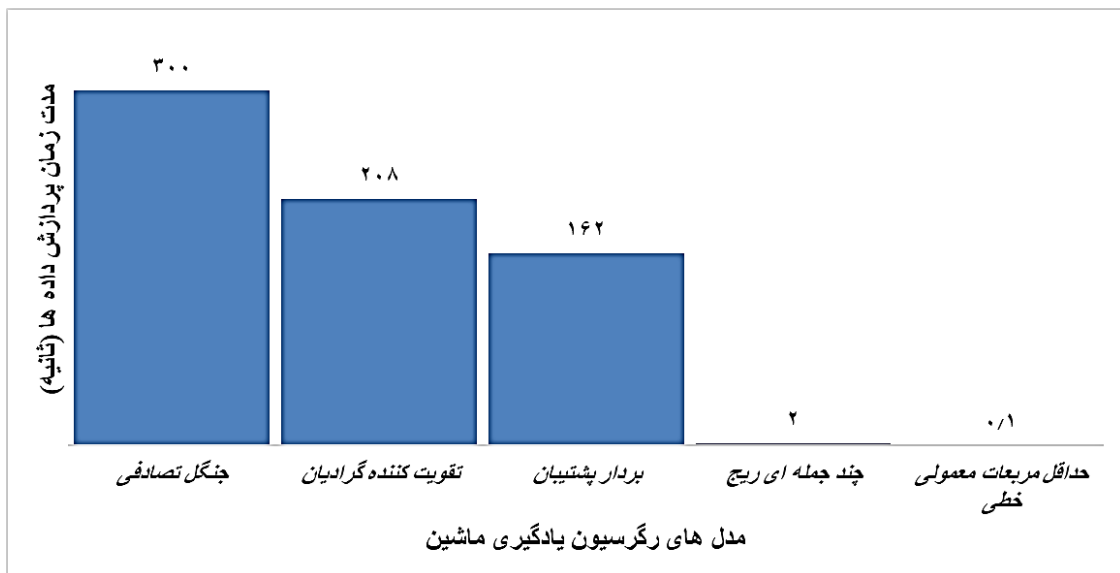
شکل ۸. نتایج پارامترهای ارزیابی برای مدل‌های رگرسیون یادگیری ماشین

جدول ۹. مقایسه نتایج پارامترهای ارزیابی برای مدل‌های رگرسیون یادگیری ماشین

مدت زمان پردازش (ثانیه)	MAPE	RMSE	R ²	مدل رگرسیون یادگیری ماشین
۳۰۰	۵/۵۴	۱۲۱/۱۲	۰/۸۸	جنگل تصادفی
۲۰۸	۵/۸۵	۱۳۰/۵۲	۰/۸۶	تقویت کننده گرادیان
۱۶۲	۶/۵۵	۱۴۵/۸۹	۰/۸۲	بردار پشتیبان
۲	۶/۶۷	۱۴۵/۶۲	۰/۸۲	چند جمله ای ریج
۰/۱	۷/۷۷	۱۶۷/۴۴	۰/۷۷	حداقل مربعات معمولی خطی



شکل ۹. مقایسه عملکرد بهترین و بدترین مدل یادگیری ماشین با مقادیر واقعی در یک نمونه شامل ۵۰ نقطه داده



شکل ۱۰. مدت زمان پردازش داده ها برای روش های یادگیری ماشین

بحث و نتیجه گیری

در مرحله پردازش داده ها، روش های رگرسیون مختلفی از جمله رگرسیون جنگل تصادفی، رگرسیون تقویت کننده گرادیان، رگرسیون بردار پشتیبان، رگرسیون چندجمله ای ریج و حداقل مربعات معمولی خطی برای انتخاب بهترین مدل در نظر گرفته شد و فرآیندهای آنها با استفاده از الگوریتم های جستجوی شبکه ای و بیزین با کمک کتابخانه های ماژول سایکیت لرن تنظیم گردید. نتایج ارزیابی داده ها نشان داد روش رگرسیون جنگل تصادفی در سه معیار عملکرد R^2 ، MAPE و RMSE بهتر از سایر روش ها عمل می کند. با این حال، مدل رگرسیون جنگل تصادفی طولانی ترین زمان را برای آموزش به خود اختصاص داد. از سوی دیگر، رگرسیون چندجمله ای ریج عملکرد قابل قبولی، قابل مقایسه با تکنیک های مدل سازی مجموعه ای و کاملاً مشابه رگرسیون بردار پشتیبان نشان داد. بنابراین، رگرسیون چندجمله ای ریج می تواند جایگزین مناسبی برای رگرسیون جنگل تصادفی در پروژه هایی باشد که مدت زمان پردازش یک محدودیت حیاتی است. همچنین، با به کارگیری مدل رگرسیون جنگل تصادفی بهینه شده، فرصت تقریبی ۲۹۱ تن در روز صرفه جویی در واردات بخار پرفشار معادل ۱۶ درصد از میانگین واردات بخار پرفشار واحد از طریق راهبردهای مدیریت واردات بخار پیشرفته شناسایی گردید. دستیابی به این فرصت صرفه جویی در بخار پرفشار منجر به جلوگیری از انتشار تقریباً ۳۸ تن CO_2 در روز توسط تامین کننده بخار پرفشار

با توجه به سوابق تحقیقاتی بررسی شده در این پژوهش و شناسایی خلاء تحقیقاتی در حوزه مدل سازی کاهش واردات بخار در واحدهای پتروشیمی و به طور خاص واحدهای تولید اتیلن اکساید/اتیلن گلیکول، یک چارچوب جدید پردازش داده و مدل سازی با دقت بالا برای پیش بینی مبتنی بر یادگیری ماشین واردات بخار پرفشار در یک واحد پتروشیمی EO/EG ارائه گردید. هدف این چارچوب، دستیابی به پیش بینی های با دقت بالا با استفاده از روش یادگیری ماشین بود. مجموعه داده های تاریخی خام، شامل پنج سال داده عملیاتی روزانه واحد تحت یک کاتالیست EO با فعالیت بالا، از سیستم کنترل گسترده واحد^۱ بازیابی شد. این مجموعه داده ها شامل تولید روزانه گلیکول، واردات روزانه اکسیژن و اتیلن و واردات روزانه بخار پرفشار بود. با استفاده از نرم افزار پایتون ۳، مجموعه داده جدیدی ساخته شد که شامل ظرفیت تولید روزانه واحد محاسبه شده از تولید روزانه گلیکول و گزینش-پذیری کاتالیست روزانه محاسبه شده از واردات روزانه اکسیژن و اتیلن به عنوان متغیرهای ویژگی و همچنین واردات روزانه بخار پرفشار به عنوان متغیر هدف بود. سپس مجموعه داده ها تحت یک مرحله پیش پردازش جامع قرار گرفت تا از کیفیت و ثبات مجموعه داده ها پیش از ورود به مرحله پردازش اطمینان حاصل شود.

Sources, Part A: Recovery, Utilization, and Environmental Effects, Dec 20, pp.1-8.

Moghadasi, M., Ozgoli, H.A. and Farhani, F. (2021) Steam consumption prediction of a gas sweetening process with methyldiethanolamine solvent using machine learning approaches. *International Journal of Energy Research*, 45(1): 879-93.

Qian, Y.Y., Chen, R.L., Chen, J.J. and Ding, G.R. (2016) Analysis of the production and market of Ethylene Oxide in China. *Chemical Industry*, 34(2): 46-48.

Rahimpour, M.R., Shayanmehr, M. and Nazari, M. (2011) Modeling and simulation of an industrial Ethylene Oxide (EO) reactor using Artificial Neural Networks (ANN). *Industrial and Engineering Chemistry Research*, 50(10): 6044-6052.

Sharifara, E., Abbaspour, M. and Saraei, A. (2024) Analysis of the high- pressure steam import behavior of an integrated ethylene oxide/ethylene glycol petrochemical plant under different production scenarios. *The Canadian Journal of Chemical Engineering*, 102(4): 1538-1557.

Shell International Chemicals, Shell Research and Technology Centre Amsterdam, EO/EG Process and Site Support. (2023) Shell EO/EG Process Design Package.

Statista. (2023) Retrieved from <https://www.statista.com/statistics/1245260/ethylene-oxide-market-volume-worldwide/>

Zahedi, G., Lohi, A. and Mahdi, K.A. (2011) Hybrid modeling of Ethylene-to-Ethylene Oxide heterogeneous reactor. *Fuel Processing Technology*, 92(9): 1725-1732.

Zapf, F. and Wallek, T. (2022) Case-study of a flowsheet simulation using deep-learning process models for multi-objective optimization of petrochemical production plants. *Computers and Chemical Engineering*, 162: 107823.

Zendehboudi, S., Rezaei, N. and Lohi, A. (2018) Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied Energy*, 228: 2539-2566.

Zendehboudi, S., Zahedi, G., Bahadori, A., Lohi, A., Elkamel, A. and Chatzis, I. (2014) A dual approach for modelling and optimisation of industrial urea reactor: Smart technique and grey box model. *The Canadian Journal of Chemical Engineering*, 92(3): 469-485.

می‌شود که معادل ۰/۳۷ درصد از ظرفیت اسمی تولید بخار پرفشار است.

در حال حاضر روش مدل‌سازی پیشنهاد شده در این پژوهش در حال یکپارچه‌سازی در چارچوب سیستم مدیریت انرژی^۱ واحد مورد مطالعه می‌باشد تا واردات بخار پرفشار کارآمد روزانه واحد بر اساس گزینش‌پذیری و ظرفیت تولید روزانه واحد مورد مطالعه به‌طور دقیق محاسبه گردیده و با مقدار بخار پرفشار واقعی روزانه مقایسه و درصد تغییر از مقدار کارآمد محاسبه گردد. با نظارت بر درصد تغییرات روزانه، افزایش میزان واردات بخار پرفشار قابل تشخیص بوده و امکان اجرای سریع اقدامات اصلاحی در سیستم بخار برای کاهش واردات بخار و نزدیک‌تر کردن آن به مقدار کارآمد، مسیر خواهد گردید.

به‌عنوان بخشی از تلاش‌های تحقیقاتی آینده، ادغام این چارچوب مدل‌سازی در سیستم اینترنت اشیا صنعتی^۲ واحد به‌طور کامل مورد بررسی قرار خواهد گرفت تا قابلیت‌های شناختی و هوشمند، به‌ویژه با استفاده از روش رگرسیون چندجمله‌ای رییج به‌دلیل مزیت آن در کاهش قابل توجه زمان پردازش با حفظ کارایی قابل قبول، به آن داده شود.

منابع

Ahmadi, A., Noorpoor, A.R., Kani, A.R. and Saraei, A. (2021) Modeling and Economic analysis of MED-TVC desalination with Allam Power Plant Cycle in Kish Island. *Iran. Iranian Journal of Chemistry and Chemical Engineering*, 40(6): 1882.

Hauke, J. and Kossowski, T. (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographical*, 30(2): 87-93.

James, G.M., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An Introduction to Statistical Learning*, Springer, p. 181.

Mhlanga, D. (2021) Artificial intelligence in the industry 4.0, and its impact on poverty, innovation, infrastructure development, and the sustainable development goals: Lessons from emerging economies. *Sustainability*, 13(5788): 1-16.

Moghadasi, M., Izadyar, N., Moghadasi, M. and Ghadamian, H. (2021) Applying machine learning techniques to implement the technical requirements of energy management systems in accordance with ISO 50001: 2018, an industrial case study. *Energy*

1 ISO-50001

2 Industrial Internet of Things

Application of Machine Learning Methods for Modeling Steam Import Reduction aimed at Energy Resources Management

Ehsan Sharifara¹, Madjid Abbaspour^{2*} and Alireza Saraei³

- 1) Ph.D. Candidate in Energy Systems Engineering, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.
- 2) Professor at School of Mechanical Engineering, Sharif University of Technology, Tehran, Iran.
*Corresponding Author Email Address: abbpor@sharif.edu
- 3) Assistant Professor at Department of Mechanical Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran.

Date of Submission: 2024/03/01

Date of Acceptance: 2024/05/20

Abstract

Integrated Ethylene Oxide/Ethylene Glycol (EO/EG) plants are major energy consumers, particularly in High-Pressure Steam (HPS) usage. This study proposes a data processing and modeling framework for high-accuracy machine learning modeling to predict the HPS import of an EO/EG petrochemical plant. The study employed Python3 and analyzed raw historical data from the plant's DCS, spanning five years of operation under the same EO catalyst. Daily feed and glycol production data were used to calculate catalyst selectivity and the plant's production capacity as models' input features, while HPS import served as the models' output target. Various regression algorithms were evaluated to select the best model for this case study, with their hyperparameters tuned using Grid Search and Bayesian Search algorithms. Random forest regression outperformed other methods in R^2 , MAPE, and RMSE metrics but had the longest training time. Polynomial ridge regression was a suitable choice considering both time consumption and performance. The tuned random forest regression model revealed an approximate 291 Tonne/Day potential for HPS import savings equivalent to 16% of the plant's average HPS import through enhanced steam import management strategies. Adopting this HPS saving measure would enable the HPS supplier to avert 38 tonnes of CO_2 emissions daily equivalent to 0.37% of its nominal HPS generation capacity. Our methodology in this paper can be applied to other EO/EG plants and is currently being integrated into the plant's energy management system, enabling continuous monitoring of steam import behavior relative to catalyst and plant performance.

Keywords: Energy conservation opportunity, Energy modeling, Machine learning, Steam consumption prediction.