# تشخیص خطا با استفاده از ماشین بردار پشتیبان چندکلاسه

علی رنجبر[۱]، امیرحسین رحمانی[۲،*]

۱- دانشجوی کارشناسی‌ارشد، گروه برق، واحد دزفول، دانشگاه آزاد اسلامی، دزفول، ایران، Takay313@gmail.com

*۲- مربی، گروه برق، واحد دزفول، دانشگاه آزاد اسلامی، دزفول، ایران، A_h_rahmani@yahoo.com

تاریخ دریافت: ۹۸/۶/۳۰ تاریخ پذیرش: ۹۸/۹/۱۱

**چکیده: در محیط‌های صنعتی، مقدار زیادی از داده‌ها تولید می‌شود که به نوبه خود انبار پایگاه داده و داده‌ها را از همه مناطق مربوطه مانند برنامه‌ریزی، طراحی فرآیند، مواد، مونتاژ، تولید، کیفیت، کنترل فرآیند، برنامه‌ریزی، تشخیص خطا، خاموش کردن، مدیریت ارتباط با مشتری و غیره جمع‌آوری می‌کند. داده‌کاوی به ابزار مورد استفاده برای کسب دانش برای روند صنعتی ساخت آهن و فولاد تبدیل‌شده است. با توجه به رشد سریع داده‌کاوی، صنایع مختلف از فنّاوری داده‌کاوی برای جستجوی الگوهای پنهان که ممکن است بیشتر به سیستم با دانش جدیدی که مدل‌های جدید را برای بهبود کیفیت تولید، هزینه مطلوب بهره‌وری و تعمیر و نگهداری و غیره بپردازند استفاده کردند. بهبود مستمر تمام فرایند تولید فولاد با توجه به اجتناب از کمبود کیفیت و بهبود تولید مرتبط با آن، یک وظیفه اساسی تولیدکننده فولاد است. بنابراین، استراتژی نقص صفر امروزه محبوب است و برای حفظ آن، چندین تکنیک تضمین کیفیت استفاده می‌شود. در این مقاله سعی میشود با استفاده از داده کاوی حسگرهای موثر در وضعیت سیستم شناسایی شوند وسپس با استفاده از ماشین بردار پشتیبان مدل مناسبی برای پیشبینی وضعیت سیستم به دست آورده شود که در این مقاله با دقت بیش از ۹۵ درصد حالت خطا تشخیص داده شده است.**

**واژه‌های کلیدی: روش‌های گروهی ، تصمیم‌گیری، الگوها ، بردار پشتیبان ، نزدیکترین همسایگی.**

## ۱-مقدمه

حس‌گر المان، حس‌کننده‌ای است که کمیت‌های فیزیکی مانند فشار، حرارت، رطوبت، دما، و ... را به کمیت‌های الکتریکی پیوسته (آنالوگ) یا غیر پیوسته (دیجیتال) تبدیل می‌کند. درواقع حس‌گر یک وسیله الکتریکی است که تغییرات فیزیکی یا شیمیایی را اندازه‌گیری می‌کند و آن را به سیگنال الکتریکی تبدیل می‌نماید. حس‌گرها در انواع دستگاه‌های اندازه‌گیری، دستگاه‌های کنترل آنالوگ و دیجیتال مانند PLC مورداستفاده قرار می‌گیرند. عملکرد حس‌گرها و قابلیت اتصال آن‌ها به دستگاه‌های مختلف ازجمله PLC باعث شده است که حس‌گر بخشی از اجزای جدانشدنی دستگاه کنترل خودکار و رباتیک باشد. حس‌گرها اطلاعات مختلف از وضعیت اجزای متحرک سیستم را به واحد کنترل ارسال نموده و باعث تغییر وضعیت عملکرد دستگاه‌ها می‌شوند.

حس‌گرها جهت تبدیل عوامل فیزیکی مانند حرارت، فشار، نیرو، طول، زاویه چرخش، دبی و غیره به سیگنال‌های الکتریکی بکار برده می‌شوند و به همین منظور حس‌گرهای مختلفی که قابلیت تبدیل این عوامل را به جریان برق دارا می‌باشند، ساخته‌شده‌اند. با پیشرفت سریع تکنیک خودکارسازی و پیچیده‌تر شدن پروسه‌های صنعتی و کاربرد روزافزون این شاخه از تکنیک نیاز شدیدی به کاربرد حس‌گرهای مختلف که اطلاعات مربوط به عملیات تولید را درک و بر اساس این اطلاعات مقتضی صادر گردد، احساس می‌شود.

حس‌گرها به‌عنوان اعضای حسی یک سیستم، وظیفه جمع‌آوری و با تبدیل اطلاعات را به‌صورت که برای یک سیستم کنترل و با اندازه‌گیری قابل‌تجزیه و تحلیل باشد به عهده‌دارند. در مهروموم‌های اخیر حس‌گرها به‌صورت یک عنصر قابل‌تفکیک



[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

دستگاه‌های مختلف صنعتی مورداستفاده قرارگرفته و پیشرفت سریعی در جهت جوابگویی به تقاضاهای صنعت در این شاخه از علم الکترونیک انجام پذیرفته‌است.

نوآوری های این مقاله به شرح زیر است:

۱. در این مقاله هیچ محدودیتی برای سیستم مورد مطالعه در نظر گرفته نمی شود.
۲. با کاهش ابعاد پایگاه داده و انتخاب حسگرهای موثر در گزارش وضعیت سیستم سرعت پردازش بیشتر می‌شود.
۳. علاوه بر پایش سیستم، روش پیشنهادی میتواند به بهره بردار وضعیت سیستم را تنها با کارکرد چند حسگر گزارش دهد.

همان‌طور که اشاره شد، حسگرهای صنعتی نمونه های متفاوت و گسترده ای دارند که در صنعت کاربردهای بسیاری دارند. از آنجا که امروزه استفاده از حسگرها ما را با حجم زیادی از اطلاعات روبرو می کند، در این مقاله سعی می شود اطلاعات حسگرها دسته بندی شود و در نهایت بتوان بدون حضور اپراتور متخصص وضعیت سیستم را بررسی و پیش‌بینی کرد.

ساختار این مقاله به صورت زیر می باشد:

۱. در بخش ۱ اطلاعات اولیه در خصوص انواع حسگرها داده خواهد شد و نیز به مرور کارهای انجام‌شده در زمینه داده-کاوی و کاربردهای آن در صنعت و روش‌های پیش‌بینی وضعیت سیستم می‌پردازیم.
۲. در بخش ۲ به بیان مدل پیشنهادی و فلوچارت حل مساله می‌پردازیم.
۳. در بخش ۳ با معرفی سیستم موردمطالعه به بیان نتایج شبیه‌سازی روش پیشنهادی بر روی این سیستم می-پردازیم.
۴. در آخرین بخش به تحلیل نتایج به‌دست‌آمده و کاربردهای مقاله می‌پردازیم.

در اکثر بخش‌های فولادی، [۱] تولید بسیار رقابتی است. برای برطرف کردن چالش‌های فرا منطقه‌ای، یک شرکت باید تولید کم‌هزینه داشته باشد و هنوز کارکنان با مهارت بالا، انعطاف‌پذیر و کارآمد را حفظ کند که بتوانند محصولات با کیفیت بالا و ارزان قیمت را طراحی و تولید کنند. این می‌تواند با استفاده از تکنیک‌های داده‌کاوی به منظور بهبود تصمیم‌گیری به دست آید.

داده‌کاوی به‌طورکلی می‌تواند به‌عنوان یک تکنیک برای پیدا کردن الگوهای (استخراج) یا اطلاعات جالب در مقدار زیادی از داده‌ها باشد [۲]. این تکنیک به‌طور گسترده‌ای درزمینه تحقیقاتی مانند مهندسی، بازاریابی، کسب‌وکار، آموزش‌وپرورش و در حال حاضر به‌ویژه در صنایع مانند آهن، فولاد، لاستیک و غیره استفاده می‌شود. بااین‌حال، دانش می‌تواند انواع مختلفی داشته باشد و شناسایی نوع دانش برای استخراج هنگام آزمایش مقدار زیادی اطلاعات تولیدشده در طول تولید. داده‌کاوی فرآیند الگوهای جالب و دانش از مقدار زیادی از داده‌ها است. منبع داده می‌تواند شامل پایگاه‌های داده، انبار داده‌ها، وب، سایر مخازن اطلاعات، داده‌هایی است که به‌صورت خودکار به سیستم منتقل می‌شوند [۳].

فولاد شامل آلیاژ آهن، کربن و منگنز با مقدار کمی از سیلیکون، فسفر و گوگرد است. مراحل تولید فولاد: حرارت دادن، خنک کردن، ذوب کردن، انجماد. هند سومین گروه بزرگ از این گروه است (از هشتم در سال ۲۰۰۳) و انتظار می‌رود که دومین تولیدکننده در آینده نزدیک باشد.

بیشتر برنامه‌های کاربردی صنعتی استخراج داده‌ها در صنایع فولاد عبارت‌اند از مدل‌سازی سیستم، نزدیک شدن به فن‌آوری‌های تولید جدید و بهبود کیفیت محصولات، خواص ضد خوردگی، فولاد گالوانیزه یک محصول که تقاضای بیشتری را در بخش‌های مختلف تجربه می‌کند. مدل‌سازی غیرخطی شامل تکنیک‌های مهم کاربرد گسترده‌ای به دلیل کارایی در سرعت. هدف از تجزیه‌وتحلیل داده‌ها، پیش‌بینی طول ضریب عملکرد، ضریب نفوذ به‌عنوان توابع تعداد زیادی از متغیرها، ازجمله ترکیب شیمیایی، حرارت عملیات، سرعت نوار در فرایند انلینگ بود. همچنین در بهینه‌سازی و کنترل فرآیند صنعتی استفاده می‌شود. این مطالعه گسترده‌ای است که به بررسی تکنیک‌های موجود در کوره‌های انفجاری در صنایع فولاد می‌پردازد.

هدف اصلی از کوره انفجار، کاهش فرایند شیمیایی و تبدیل اکسید آهن به آهن مایع است که به نام فلز داغ شناخته می‌شود. فرایند تولید فلز داغ، حدود ۷۰ درصد کل انرژی تولیدشده از تولید فولاد را به مصرف می‌رساند. علاوه بر این نیازهای اجتماعی و صنعتی فولاد آهن، قیمت بالای مواد خام و عوامل کاهش‌دهنده نیز ضرورت مدل‌سازی این فرآیند پیچیده را برای افزایش بهره‌وری و کاهش هزینه افزایش داده است [۴]. برای اجرای این، تکنیک‌های داده‌کاوی در بسیاری از مراحل عملیات انفجار کوره مورد آزمایش قرارگرفته‌اند.

با توجه به این‌که G.J. Zheng، W. Zhang، P. Hu و D.Y Shi [۵] در این فرایند شکل‌گیری گرم، بسیاری از متغیرهای طراحی تأثیری بر نتیجه روش‌های مختلف و پیچیده مانند ویژگی‌های هندسه و پارامترهای فرایند تشکیل دارند. دشوار است درک ارتباط بین متغیرهای طراحی و نتایج، که برای هدایت طراحی بسیار مهم است. در این مقاله داده‌کاوی برای کشف تأثیر ویژگی‌های هندسی گذشته و پارامترهای شکل‌گیری گرم بر نتایج گرمایی وسیله مدل B استایل معرفی شد و محدوده پارامتر بهینه تعیین شد. ابتدا چند پارامتر متغیر انتخاب و ۱۰۰ گروه داده تجربی انتخاب شدند

با روش سوپر لاتین تولید شد و سپس نتایج تجزیه‌وتحلیل متشکل از روش‌های محدود به دست آمد. تجزیه‌وتحلیل بعدی و تکامل شبیه‌سازی با استفاده از الگوریتم‌های تصمیم‌گیری درخت


[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

(DT) انجام شد. درنهایت، یک سری از قوانین شکل‌گیری گرم پایه B اصلاح شد، ازجمله درجه حرارت اولیه فلز ورق باید بین c.۷۲۰ تا c۸۰۰ کنترل شود. بنابراین، یک مدل واقعی ستون b برای آزمودن قوانین طراحی‌شده و نتیجه درست و مؤثر بود.

در مرجع[۶] نویسندگان یک سیستم پشتیبانی تصمیم را ایجاد کردند که می‌تواند خوردگی را تعیین کند و زمان رشد خوردگی را برای نگهداری با استفاده از رویکرد مدل‌سازی احتمالی طراحی کند. در این کار، فرایند کنترل خوردگی بر مبنای داده کاوی معرفی‌شده است، این فرایند به‌جای عملیات پیشگیرانه و پیشگیرانه که با استفاده از بازرسی‌های منظم برنامه‌ریزی‌شده و استفاده از حس‌گرهای خوردگی به‌طور مداوم برای تسهیل اقدامات پیشگیرانه انجام می‌شود، به‌جای نگهداری پیشگیرانه و پیش‌بینی‌شده است. استفاده از فرآیند داده کاوی شامل تجزیه‌وتحلیل داده‌های تاریخی که در طول زمان جمع‌آوری می‌شوند در مورد ضخامت فلز معروف که قبلاً در معرض عوامل محیطی مانند بارندگی، دما، رطوبت و غیره قرار دارد. نتایج تجزیه‌وتحلیل، تصمیم‌گیری‌های پیشگیرانه و دانش مبتنی بر کنترل خوردگی را برای تصمیم‌گیری فراهم می‌کند. روش طبقه‌بندی مانند روش قضیه بیز، شبکه‌های عصبی برای مدل‌سازی عدم قطعیت در وقوع خوردگی با توجه به عدم اطمینان دانش و عدم اطمینان داده‌ها برای تصمیم‌گیری‌های اطلاعات استفاده‌شده است. مدل‌سازی احتمالاتی برای توسعه یک روش معادله داده استفاده می‌شود که می‌تواند برای به دست آوردن اطلاعات از یک پایگاه داده بر روی فلزات که قبلاً در معرض محیط جو قرارگرفته‌اند استفاده شود. نویسندگان محمد ساعیه، مهدی مقیمی و ایوب بقیری [۷] پیشنهاد می‌کنند که فرایند انحلال یکی از عملیات مهم تولید ورق‌های فولادی سرد و نورد است که کیفیت نهایی محصولات نورد سرد را تحت تأثیر قرار می‌دهد. در فرایند خود، کویل های سرد نورد گرم به‌آرامی به دمای موردنظر گرم می‌شوند و سپس خنک می‌شوند. مدل‌سازی فرایند خنک‌سازی (پیش‌بینی زمان گرمایش و خنک‌کننده و پیش‌بینی روند گرمایش سیم‌پیچ کویل) کار بسیار پیچیده و گران است. مدل‌سازی فرایند انحلال را می‌توان با استفاده از مدل‌های حرارتی انجام داد. در این مقاله مدل‌سازی فرایند انحلال با استفاده از تکنیک‌های داده‌کاوی به دلیل سرعت‌بالا در پردازش داده‌ها، نتایج قابل‌قبول و سادگی آن‌ها برای استفاده از آن پیشنهاد شده است. در این مطالعه، آن‌ها روش مدل‌سازی فرایند انلینگ را با استفاده از تکنیک‌های داده‌کاوی پیشنهاد کردند. پس از آزمودن تکنیک‌های مختلف داده‌کاوی، شبکه عصبی پروتکل انتقال خورده پیش‌بینی‌شده برای پیش‌بینی زمان گرمایش و خنک‌کننده و درجه حرارت هسته کویل در طی فرایند انحلال انتخاب‌شده است. نتیجه پیش‌بینی‌شده شبکه‌های عصبی برای مدل‌سازی فرایند انحلال به‌اندازه کافی دقیق بود، اما درحالی‌که ما از دقت بیشتری استفاده می‌کنیم، دقت پیش‌بینی کننده‌ها بهبود می‌یابد. روش حاضر برای پیش‌بینی رفتار فرایندهایی است که نمی‌توان با هر معادلی تحلیلی یا فیزیکی توصیف کرد. تکنیک‌های دیگر مانند طبقه‌بندی، رگرسیون‌های مختلف و یا روش خوشه‌بندی می‌تواند برای بهینه‌سازی پارامترهای ورودی فرایند انحلال برای به حداکثر رساندن بهره‌وری از عملیات خنثی استفاده شود.

به گفته سید مهران شریفی و حمیدرضا اسماعیلی [۸]، استفاده از روش‌های داده‌کاوی برای پیش‌بینی نقص در سطح فولاد یک چالش بود. در صنعت فولاد، به‌خصوص فولاد آلیاژی، ایجاد یک محصول متفرقه متفاوت می‌تواند هزینه‌های زیادی را برای تولیدکنندگان فولاد تحمیل کند. یکی از نقایص مشترک در تولید نمرات فولاد کم‌کربن، نقص حفره و پلک است. نقص آن‌یک اتلاف وقت و هزینه برای از بین بردن این نقص است، ما باید سطح محصول را تمیز کنیم. در بعضی موارد، شدت نقص‌ها ممکن است منجر به قطعه‌قطعه شدن محصول شود. سنگ‌زنی باعث تلف شدن زمان و هزینه تولید خواهد شد. فراوانی نقص‌ها به عوامل متعددی مربوط است

تجزیه و تحلیل مواد و فرایندهای تولید. در این مطالعه، نویسندگان یک مدل برای پیش‌بینی این گسل با روش داده‌کاوی، ازجمله درخت تصمیم‌گیری، شبکه عصبی، ایجاد کردند. آن‌ها اعمال کردند

این تکنیک‌ها به داده‌های جمع‌آوری‌شده از شرکت آلیاژ فولاد ایران است. مدل ایجادشده با استفاده از درخت تصمیم‌گیری دارای دقت بالاتر است. مزایای این مدل عبارت‌اند از:

* کاهش روند پردازش و هزینه انرژی با حذف مرحله سنگ‌زنی.
* یک ابزار بهینه برای پیش‌بینی نقص و درخواست برای تولید محصولات بدون نقص وجود دارد.

بنا به گفته نویسندگان [۹] این مقاله یک روش داده‌کاوی برای انتخاب متغیر و استخراج دانش از مجموعه داده‌ها را نشان داد. این رویکرد مبتنی بر رگرسیون نمادین غلط است (هر متغیر موجود در مجموعه داده به‌عنوان متغیر هدف در رگرسیون چندگانه رفتار می‌شود) و متغیر رگرسیون متغیر جدید برای برنامه‌نویسی ژنتیکی. ارتباط هر متغیر ورودی محاسبه‌شده و مدل تقریبی متغیر هدف ایجادشده است. پیکربندی‌های برنامه‌نویسی ژنتیکی با متغیرهای مختلف هدف چندین بار اجرا می‌شود تا اثرات تصادفی را کاهش دهد و نتایج جمع‌آوری‌شده به‌عنوان شبکه تعامل متغیر نمایش داده می‌شود. این شبکه تعامل، مؤلفه‌های مهم سیستم و روابط ضمنی بین متغیرها را برجسته می‌کند. کل رویکرد به دلیل پیچیدگی کوره انفجار و بسیاری از ارتباطات بین متغیرها بر روی یک مجموعه داده کوره انفجار آزمایش‌شده است.

بسیاری از متغیرها در فرآیند کوره انفجاری به‌طور ضمنی مرتبط هستند، به علت روابط فیزیکی و یا به دلیل کنترل خارجی


[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

پارامترهای کوره انفجار. مثال‌هایی برای متغیرهایی با رابطه ضمنی با متغیرهای دیگر عبارت‌اند از: شعله یا پارامترهای انفجار داغ. معمولاً، چنین رابطه ضمنی پیشینی در سناریوهای مدل‌سازی مبتنی برداده شناخته‌نشده است، اما می‌تواند از اطلاعات مربوط به متغیر جمع‌آوری‌شده از پرونده‌های متعدد معیار سنجش استخراج شود.

با استفاده از رویکرد داده‌کاوی رگرسیون نامتجانس، چندین مدل شناسایی‌شده‌اند که مقادیر مشاهده‌شده در فرآیند کوره انفجاری را تقریباً دقیق نشان می‌دهند. این آزمایش‌ها همچنین منجر به تعداد زیادی از مدل‌های مختلف اجزای کوره انفجار می‌شود. مدل تولیدشده برای استخراج اطلاعات در مورد روابط ضمنی در مجموعه داده استفاده می‌شود تا مجموعه‌ای از متغیرهای ورودی مربوط را کاهش داده و متقارن کند.

در مرجع بعدی همچنین استفاده از رگرسیون نمادین در کوره انفجار با استفاده از داده‌های آسیاب خورشیدی [۱۰] توصیف می‌شود. این نشان می‌دهد که استفاده از یک سیستم رگرسیون سمبلیک اقتباس‌شده در دو مجموعه داده‌های مختلف، شامل اندازه‌گیری از فرآیند کوره انفجاری است که شایع‌ترین تولید فلز گرم (آهن مایع) است. اگرچه فرایند شیمیایی و فیزیکی به‌خوبی درک شده است، از دست دادن گرما در مناطق خاص کوره به‌طور کامل درک نمی‌شود. دانش در مورد چنین روابط می‌تواند برای بهینه‌سازی فرایند کوره انفجاری استفاده شود و بنابراین مدل‌سازی فرایند کوره انفجار بر اساس داده‌های واقعی دنیای جمع‌آوری‌شده از اهمیت خاصی برخوردار است. تجزیه‌وتحلیل رگرسیون یک فرعی از داده‌کاوی است که تلاش می‌کند تا دانش موجود در مجموعه داده را آشکار سازد. در این کار، رگرسیون نمادین با استفاده از یک درخت بر اساس یک سیستم برنامه‌نویسی ژنتیکی برای تکمیل فرمول‌های ریاضی انجام می‌شود. برنامه‌نویسی ژنتیکی [۱۱] یک الگوریتم تکاملی است که برنامه‌هایی را برای حل یک مسئله ارائه می‌دهد. در این روش مدل‌سازی، سه جنبه الگوریتمی گنجانده‌شده و با یک رویکرد رگرسیون سمبلیک استاندارد مقایسه می‌شود. دقیقاً نویسندگان اثرات انتخاب به‌صورت آفلاین را با استفاده از ضریب همبستگی 2R به‌عنوان عملکرد تناسب‌اندام مورد آزمایش قراردادند و نمونه‌های ارزیابی‌شده یا موارد تناسب‌اندام را نیز نمونه‌برداری کردند. انتخاب بیرونی یک انتخاب اضافی در الگوریتم ژنتیک و برنامه‌نویسی ژنتیکی است که بعد از نوترکیب و جهش اعمال می‌شود. تناسب فرد در تجزیه‌وتحلیل رگرسیون نماد معمولاً به‌عنوان خطای میانگین میدان بین ارزش پیش‌بینی‌شده مدل و ارزش مشاهده‌شده متغیر هدف محاسبه می‌شود. الگوریتم ژنتیک سریع‌تر از الگوریتم ژنتیک بدون انتخاب فرزند است. در این سهم، سه سازگاری الگوریتمی، انتخاب فرزند، ضریب تعیین 2R به‌عنوان تابع تناسب و نمونه‌برداری، به یک سیستم رگرسیون نمادین نسبتاً استاندارد موردبررسی قرارگرفته است. اثرات ترکیب این سازوکارها بر مجموعه داده‌های واقعی در دو فرآیند تولید فولاد نشان داده‌شده است. بهترین پیشرفت‌ها ازلحاظ کیفیت به علت استفاده از عملکرد تناسب‌اندام as2R به دست آمد.

با توجه به Jong-Hag Jeon [۱۲]، این ارائه نمونه‌ای از کاربرد داده‌کاوی در کارخانه فولاد است. مهندسان مسئول کنترل کیفیت با دو نوع مشکلات در حل مشکلات با استفاده از روش‌های آماری برخورد می‌کنند. کوره‌های انفجاری داغ برای مصرف‌کنندگان انرژی مهم هستند. گرمایش انبساط داغ ۱۰٪ -۱۵٪ از کل انرژی مصرف‌شده توسط کوره‌های انفجاری است. کاهش مصرف انرژی برای گرمایش انفجاری داغ تأثیر مهمی در هزینه تولید گرما دارد.

تکنیک‌های داده‌کاوی در صنایع فولادی نتایج بسیار عالی‌ای را برای اهداف زیر به دست می‌آورند. نظارت بر کیفیت محصول نظارت بر فرآیند نظارت بر استراتژی‌های نگهداری.

همچنین، به نظر می‌رسد روشن است که داده‌کاوی با استفاده از شبیه‌سازی مدل‌سازی شمارشی که داده‌های زمان واقعی در دسترس نیست بهبود می‌یابد. در تمام موارد، تأثیر این فناوری‌ها در فرایند صنعتی مدرن، ضروری است که شرکت‌ها از این امر آگاهی یابند تا دانش را از فرایند استخراج کنند و بتوانند آن‌ها را بهبود ببخشند. همانطور که در مرور مطالعات قبلی گفته شد بیشتر روش های موجود سعی در مدل کردن سیستم و پایش برخط سیستم داشتند. در بسیاری از موارد به دلیل بالا بودن هزینه امکان پیاده سازی طرح از نظر اقتصادی فراهم نبوده است. بعضی از مقالات بررسی شده سیستم را در حالات خاص بررسی میکردند و روش پیشنهادی انها تنها مختص به شرایط کار خاص سیستم می باشد. از طرفی بعضی از روش های بالا به دلیل حجم بالای محاسبات توانایی پاسخگویی سریع ندارند. در ادامه سعی می‌شود با روش پیشنهادی خود معایب گفته شده در بالا را تا حد امکان پوشش داد.

## ۲-مدل‌سازی روش پیشنهادی

ماشین‌های بردار پشتیبانی (SVMs) روش‌های یادگیری مبتنی بر هسته است که برای حل مشکلات رگرسیون موفقیت‌آمیز است. بااین‌حال، استفاده از آن‌ها در برنامه‌های قابلیت اطمینان به‌طور گسترده‌ای موردبررسی قرار نگرفته است. در این مقاله، یک تحلیل مقایسه‌ای برای ارزیابی اثربخشی SVM در پیش‌بینی زمان به شکست و قابلیت اطمینان قطعات مهندسی بر اساس داده‌های سری زمانی ارائه‌شده است. کارایی در مورد مطالعات موردی رگرسیون SVM بر روی دیگر روش‌های پیشرفته یادگیری مانند عملکرد پایه شعاعی ، مدل Box-Jenkins یکپارچه متحرک و شبکه‌های محلی عادی مجازی محاسبه می‌شود. مقایسه نشان می‌دهد که در موارد مورد تجزیه‌وتحلیل SVM بهتر از روش‌های دیگر قابل‌مقایسه است.



[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

در ادامه ابتدا روش انتخاب ویژگی با توجه به نزدیکترین همسایگی معرفی میشود و سپس مدل سازی بردار پشتیبان بیان می شود.

جستجوی نزدیکترین همسایه (یاNNS)، که همچنین با نام‌های جستجوی مجاورت، جستجوی همسانی یا جستجوی نزدیک‌ترین نقطه شناخته می‌شود، یک مسئله بهینه سازی برای پیدا کردن نزدیک‌ترین نقطه‌ها در فضاهای متریک است. مسئله بدین صورت است که: مجموعه S شامل تعدادی نقطه در یک فضای متریک مانند M و نیز یک نقطه پرس و جوی q∈M داده شده‌است، هدف پیدا کردن نزدیک‌ترین نقطه در S به q است. در بسیاری از موارد، فضای M به صورت یک فضای اقلیدسی d-بعدی و فاصله بین نقاط با معیار فاصله اقلیدسی، فاصله منهتن یا دیگر فاصله‌های متریک سنجیده می‌شود. دونالد کنوت در جلد ۳ از هنر برنامه نویسی کامپیوتر (۱۹۷۳) این مسئله را "مسئله دفتر پست" نامید. این نام‌گذاری اشاره یک به برنامه کامپیوتری دارد که برای اختصاص نزدیکترین اداره پست به یک محل اقامت توشته شده بود.

راه حل‌های متعددی برای مسئله NNS پیشنهاد شده‌اند. کیفیت و سودمندی این الگوریتم‌ها بر اساس پیچیدگی زمانی و همچنین پیچیدگی حافظه مصرفی ساختمان داده‌هاهای آنها، ارزیابی می‌شود. مشاهدات غیررسمی که معمولاً آن را با عنوان نفرین ابعاد یاد می‌کنند، بیان می‌دارد که در واقع هیچ راه حل دقیق و همه منظوره‌ای برای مسئله NNS در فضای اقلیدسی با ابعاد بالا وجود ندارد که از پیش پردازش با مرتبه چندجمله‌ای و زمان جستجوی چندلگاریتمی O((log n)^k) بهره ببرد.

## ۲-۱- جستجوی خطی

ساده‌ترین راه حل برای مسئله NNS، محاسبه فاصله نقطه پرسش شده تا تمامی نقاطِ پایگاه داده، همراه با نگهداری "بهترین جواب پیدا شده تا کنون" است. این الگوریتم که به آن الگوریتم سرراست نیز گفته می شود، دارای مرتبه زمانی O(nd) است که در آن N اندازه مجموعه S و d تعداد ابعاد فضای M است. در این روش هیچ داده ساختاری از جستجوها ذخیره نخواهد شد و الگوریتم بجز ذخیره نقاط پایگاه داده، هزینه حافظه‌ای دیگری ندارد. روش ذکر شده، بطور میانگین، از روش‌های پارتیشن‌بندی فضا، در ابعاد بالاتر، بهتر عمل می‌کند.

## ۲-۲- پارتیشن‌بندی فضا

از سال ۱۹۷۰، روش انشعاب و تحدید بر روی مسئله NNS اعمال شده‌است. در فضای اقلیدسی، این کار با نام شاخص‌های مکانی یا روش دسترسی مکانی شناخته می‌شود. روش‌های متعددی برای پارتیشن بندی فضا به منظور حل مسئله NNS ارائه و توسعه داده شده‌اند. در این بین، ساده‌ترین روش، استفاده از ساختمان داده درخت کی‌دی (به انگلیسی: k-d tree) است که به صورت تکراری فضای جستجو را در یکی از ابعاد، به دو نیم‌بخش تقسیم می‌کند به گونه‌ای که هر نیم‌بخش تقریباً نیمی از نقاط بخشِ بزرگتر را در برخواهد گرفت. پرس و جو با پیمایش درخت از ریشه تا برگ صورت می‌پذیرد به گونه‌ای که در هر مرحله، مقدار بردار گره در بُعد مشخصی، با همان بُعد از نقطه پرس و جو، مقایسه می‌شود؛ اگر مقدار نقطه پرس وجو بیشتر بود به زیر درخت سمت راست و در غیر این‌صورت به زیردرخت چپ حرکت می‌کنیم. همچنین بسته به اینکه فاصله این دو نقطه چقدر است، شاخه‌های کناری نیز ممکن است نیاز به بررسی داشته باشند. برای مسائلی که تعداد بعد فضا ثابت است، میانگین پیچیدگی زمانی این روش از مرتبه O(log N) است. در مورد نقاط راندم توزیع شده تحلیل پیچیدگی زمانی مربوط به بدترین حالت صورت می گیرد. روش دیگر، استفاده از ساختمان داده درخت مستطیلی است که برای استفاده NNS در زمینه‌های پویا بکار گرفته می‌شود. این ساختمان داده دارای الگوریتم‌های مؤثری برای درج و حذف از درخت می‌باشد.

در رابطه با فضاهای متریک کلی، روش انشعاب و تحدید با نام درخت متریک شناخته می‌شود. به عنوان نمونه می‌توان به vp-tree و BK-tree اشاره کرد.

با استفاده از مجموعه‌ای از نقاطی از یک فضای ۳ بعدی و قرار دادن آن‌ها در یک درخت BSP و نیز با داشتن یک نقطه پرس و جو از همان فضا، یک راه حل ممکن برای مسئله پیدا کردن نزدیک‌ترین نقطه در ابری از نقاط، به نقطه پرس و جو، در زیر در قالب یک الگوریتم توضیح داده می شود. گفتنی است، ممکن است هیچ نقطه‌ای به عنوان جواب وجود نداشته باشد، زیرا ممکن است این جواب، یکتا نباشد اما در عمل، معمولاً ما تنها دنبال پیدا کردن یکی از زیر مجموعه های تمام ابر-نقاط داده شده، که در کوتاه‌ترین فاصله نسبت به نقطه پرس و جو، واقع شده هستیم.

ایده این است که، برای هر شاخه از درخت، حدس می‌زنیم که نزدیک‌ترین نقطه در ابر، در نیم فضای شامل نقطه پرس و جو قرار دارد. اگر چه این حدس ممکن است درست نباشد، اما ابتکار خوبی است. پس از پیمایش بازگشتی نیم فضای حدس زده شده، فاصله محاسبه شده را با کمترین فاصله نقطه پرس و جو تا صفحه پارتیشن‌کننده فضا مقایسه می‌کنیم. این فاصله، کمترین فاصله ایست که بین نقطه پرس و جو و نقاط نیم فضای جستجو نشده می‌تواند وجود داشته باشد. اگر فاصله نزدیکترین نقطه در این نیم-فضا کمتر بود، پاسخ قطعاً در همین نیم-فضاست و نیازی به جستجوی نیم فضای مجاور نیست، در غیر اینصورت باید آن نیم-فضا نیز به صورت بازگشتی جستجو شده و پاسخِ آن با کوتاهترین فاصله بدست آمده تاکنون مقایسه گردد و مقدار درست برگردانده شود. عملکرد این الگوریتم زمانی که نقطه پرس و جو در نزدیکی ابر باشد، بیشتر به زمان لگاریتمی نزدیکتر است تا زمان خطی. زیرا هنگامی که فاصله بین نقطه پرس و جو و نزدیک‌ترین نقطه در ابر


[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

به صفر میل کند، الگوریتم تنها با استفاده از جستجوی نقطه پرس و جو به عنوان یک کلید می‌تواند به نتیجه درست برسد.

فاصله دو نقطه p و q اندازه پاره‌خطی‌ست که آنها را به هم متصل می‌کند (pq vector). در مختصات دکارتی اگر:

$$p = (p_1, p_2 \dots p_n)$$
$$q = (q_1, q_2 \dots q_n) \quad (۱)$$

دو نقطه در فضای اقلیدسی n بعدی باشند، آنگاه فاصله بین آنها به صورت زیر تعریف می‌شود:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n -} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (۲)$$

حال با استفاده از الگوریتم ژنتیک ابتدا بهترین مراکز دسته انتخاب می‌شود. به صورتی که مجموع فواصل همه اعضای هر دسته تا مرکز کمترین مقدار خود باشد.

سپس با حذف داده ها به صورت تصادفی و تکرار سعی می‌شود حداقل تعداد ویژگی ها را به صورتی پیدا کرد که نماینده دسته بندی اولیه باشند.

در این مقاله برای انتخاب ویژگی بیشترین تغییرات در خروجی در نظر گرفته می‌شود. در ادامه روش بردار پشتیبان را به‌اختصار توضیح می‌دهیم.

ماشین بردار پشتیبانی (SVMs) یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده‌است. مبنای کاری دسته‌بندی کننده SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. قبل از تقسیمِ خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را به وسیله تابعِ phi به فضای با ابعاد خیلی بالاتر می‌بریم. برای اینکه بتوانیم مسئله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کنیم از قضیه دوگانی لاگرانژ برای تبدیلِ مسئله مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده phi که ما را به فضایی با ابعاد بالا می‌برد، تابعِ ساده‌تری به نامِ تابع هسته که ضرب برداری تابع phi است ظاهر می‌شود استفاده می‌کنیم. از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموید می‌توان استفاده نمود.

ما مجموعه داده‌های آزمایش D شامل n عضو (نقطه) را در اختیار داریم که به صورت زیر تعریف می‌شود:

$$D = \{(X_i, y_i)\, X_i \in \mathbb{R}^p, y_i \in \{-1; 1\}\}_{i-1}^{n} \quad (۳)$$

جایی که مقدار y برابر ۱ یا ۱ - و هر X یک بردار حقیقی p-بعدی است. هدف پیدا کردن ابرصفحه جداکننده با بیشترین فاصله از نقاط حاشیه‌ای است که نقاط با =۱y را از نقاط ۱-جدا کند. هر ابر صفحه می‌تواند به صورت مجموعه‌ای از نقاط $\{x\}$ که شرط زیر را ارضا می‌کنند نوشته شود:

$$W * X - b = 0 \quad (۴)$$

بردار نرمال است. که به ابرصفحه عمود است. ما می‌خواهیم w,b را طوری انتخاب کنیم که بیشترین فاصله بین ابر صفحه‌های موازی که داده‌ها را از هم جدا می‌کنند، ایجاد شود. این ابرصفحه‌ها با استفاده از رابطه زیر توصیف می‌شوند.

اگر داده‌های آموزشی جدایی پذیر خطی باشند، ما می‌توانیم دو ابر صفحه در حاشیه نقاط به‌طوری‌که هیچ نقطه مشترکی نداشته باشند، در نظر بگیریم و سپس سعی کنیم، فاصله آن‌ها را، حداکثر کنیم. با استفاده از هندسه، فاصله این دو صفحه است؛ بنابراین ما باید را مینیمم کنیم. برای اینکه از ورود نقاط به حاشیه جلوگیری کنیم، شرایط زیر را اضافه می‌کنیم: برای هر I

$$w.x_i - b \geq 1, \text{if } y_i = 1$$
$$w.x_i - b \leq -1, \text{if } y_i = -1 \quad (۵)$$

با کنار هم قرار دادن این دو یک مسئله بهینه‌سازی به دست می‌آید:

$$\min_{(w,b)} ||w||$$
$$s.t. \forall 1 \leq i \leq n \quad (۶)$$
$$y_i(w.x_i - b) \geq 1.$$

همان‌طور که در بالا توضیح داده شد روش پیشنهادی ما شامل مراحل زیر می باشد که باید به ترتیب انجام شوند:

۱. جمع آوری داده های سیستم به تعداد زیاد. هرچه پایگاه داده بزرگتر داشته باشیم شرایط و موقعیت های بیشتری از سیستم در نظر گرفته می شود و نتیجه کامل‌تر خواهد بود.
۲. دسته بندی داده‌ها با استفاده از الگوریتم نزدیکترین همسایگی.در این قسمت داده‌ها را به چند دسته دلخواه تقسیم کرده و مراکز دسته‌هایی که بهترین جواب مساله همسایگی را به ما بدهند انتخاب می‌شوند.
۳. با مراکز داده شده در مرحله قبل و استفاده از الگوریتم ژنتیک سعی می‌شود دسته‌بندی مرحله ۲ را با تعداد کمتری از ویژگی ها به دست آورد.


[ DOR: 20.1001.1.23222468.1398.8.3.3.8 ]
Journal of Novel Researches on Electrical Power - Vol. 8-No. 3- Autumn 2019

۴. داده‌های به دست آمده از مرحله ۳ را با استفاده از ماشین بردار پشتیبان دسته بندی کرده و سعی می‌شود صفحات جداکننده داده‌ها، را با بیشترین حاشیه به دست آورد.

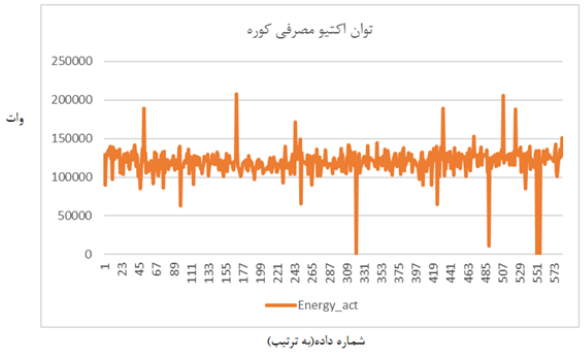
در ادامه بر روی نمونه آزمایشی مراحل گفته شده در بالا را انجام می‌دهیم و نتایج را بررسی می‌کنیم.

## ۳-نتایج شبیه‌سازی

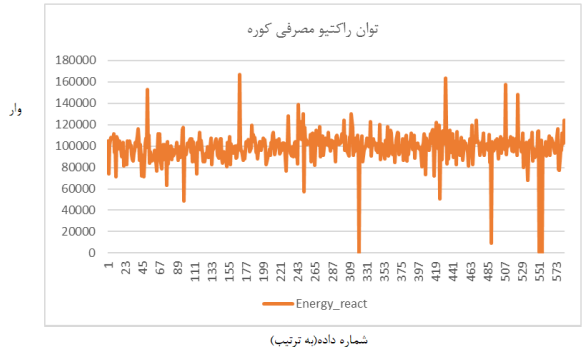
شرکت فولاد خوزستان، دومین تولیدکننده فولاد خام در ایران بعد از شرکت فولاد مبارکه اصفهان می‌باشد. کارخانه مرکزی این شرکت با وسعت ۳/۸ کیلومتر مربع، در مجاورت شهر اهواز (کیلومتر۱۰جاده اهواز بندرامام خمینی (ره)) واقع شده است و دفتر مرکزی آن نیز در اهواز قرار دارد.

شرکت فولاد خوزستان در سال ۱۲ فروردین ۱۳۶۸ به‌عنوان نخستین مجتمع تولید آهن و فولاد ایران، به روش احیاء مستقیم و کوره قوس الکتریکی راه‌اندازی شد.

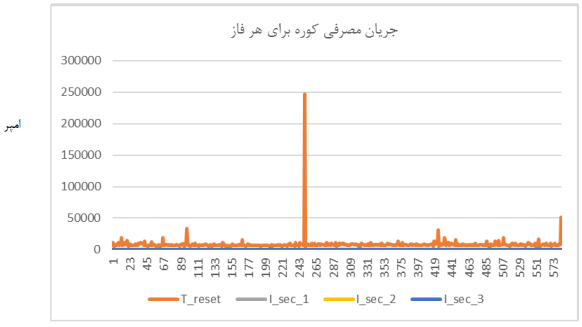
هدف از انجام این مرحله کاهش ابعاد پایگاه داده از ماتریس ۱۵۰ ستونی به ماتریسی با تعداد ستون‌های کمتر می‌باشد. اطلاعات استفاده‌شده در این مقاله از داده‌های کوره شماره۲ واحد کارخانه‌های فولادسازی است. اطلاعات شامل پایش های الکتریکی و حرارتی کوره می‌باشند. در ادامه قسمتی از این اطلاعات نشان داده می‌شود.



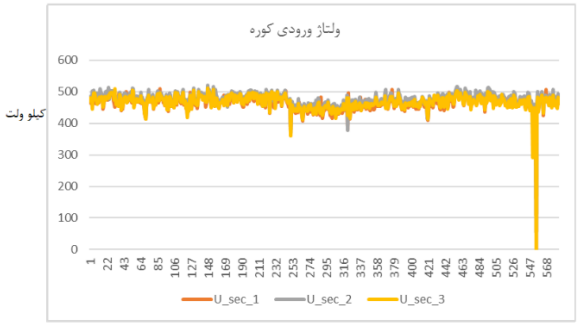

شکل (۱): توان اکتیو مصرفی کوره




شکل (۲): توان راکتیو مصرفی کوره




شکل (۳): جریان مصرفی کوره برای هر فاز




شکل (۴): ولتاژ ورودی کوره


شکل (۵): شباهت نمودار جریان و توان راکتیو فاز۳


شکل (۶): شباهت نمودار جریان و توان فاز ۲

شاید به‌نظرآید طبق قوانین برقی توان به جریان وابسته است و باید گفت این مطلب درست است ولی در فرایند ویژگی رابطه مستقیم در نظر گرفته می‌شود و اثبات می‌شود درحالی‌که توان اکتیو به ولتاژ و زاویه بار نیز بستگی دارد.

همان‌طور که به‌صورت نمونه در بالا دیده می‌شود دمای کوره و جریان ورودی باهم هیچ شباهتی ندارند .درنتیجه نمی‌توان هیچکدام از این دو را از ویژگی‌ها حذف کرد.

پس از انجام مرحله انتخاب ویژگی و یا کاهش ابعاد ویژگی‌های زیر انتخاب می‌شوند.

جدول (۱): ویژگی های انتخاب شده

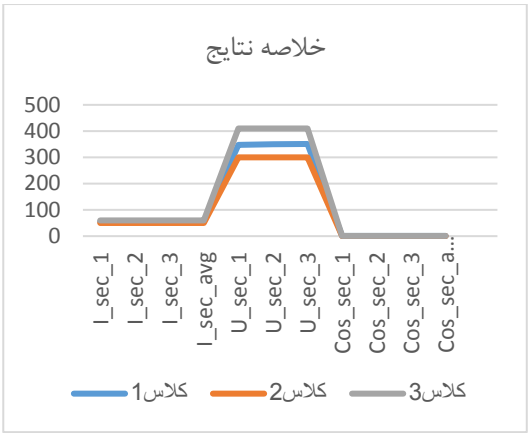
| شماره | پارامتر | معرفی |
|---|---|---|
| ۱ | I_sec_1 | جریان فاز۱ |
| ۲ | I_sec_2 | جریان فاز۲ |
| ۳ | I_sec_3 | جریان فاز۳ |
| ۴ | I_sec_avg | جریان متوسط |
| ۵ | U_sec_1 | ولتاژ فاز۱ |
| ۶ | U_sec_2 | ولتاژ فاز۲ |
| ۷ | U_sec_3 | ولتاژ فاز۳ |
| ۸ | T_reset | دمای کوره |
| ۹ | Cos_sec_1 | زاویه بین ولتاژ و جریان فاز۱ |
| ۱۰ | Cos_sec_2 | زاویه بین ولتاژ و جریان فاز۲ |
| ۱۱ | Cos_sec_3 | زاویه بین ولتاژ و جریان فاز۳ |
| ۱۲ | Cos_sec_avg | متوسط زاویه بین ولتاژ و جریان |

بردارهای پشتیبان به زبان ساده، مجموعه‌ای از نقاط در فضای n بعدی داده‌ها هستند که مرز دسته ها را مشخص می‌کنند و مرزبندی و دسته‌بندی داده‌ها بر اساس آن‌ها انجام می‌شود و با جابجایی یکی از آن‌ها، خروجی دسته‌بندی ممکن است تغییر کند. در این مقاله بعد از اعمال مرحله انتخاب ویژگی به دنبال بردار های پشتیبان ۱۰ بعدی می‌گردیم.

بعد از انجام الگوریتم ماشین بردار پشتیبان به نتایج زیر می‌رسیم.

جدول (۲): نتایج ویژگی های انتخابی

| پارامتر | کلاس۱ | کلاس۲ | کلاس۳ |
|---|---|---|---|
| I_sec_1 | ۵۵ | ۵۰ | ۶۰ |
| I_sec_2 | ۵۵ | ۵۰ | ۶۰ |
| I_sec_3 | ۵۵ | ۵۰ | ۶۰ |
| I_sec_avg | ۵۵ | ۵۰ | ۶۰ |
| U_sec_1 | ۳۴۷ | ۳۰۰ | ۴۱۰ |
| U_sec_2 | ۳۵۰ | ۳۰۰ | ۴۱۰ |
| U_sec_3 | ۳۵۱ | ۳۰۰ | ۴۱۰ |
| T_reset | ۵۰۰۰۰ | ۱۰۰۰۰۰ | ۲۰۰۰۰ |
| Cos_sec_1 | ۰/۸ | ۰/۵ | ۰/۸۸ |
| Cos_sec_2 | ۰/۸ | ۰/۵ | ۰/۸۸ |
| Cos_sec_3 | ۰/۸ | ۰/۵ | ۰/۸۸ |
| Cos_sec_avg | ۰/۸ | ۰/۵ | ۰/۸۸ |




شکل (۷): خلاصه نتایج

## ۴-نتیجه گیری

همان‌طور که می بینیم کلاس ۱ که کلاس میانی نیز است کلاس ناحیه خطر می باشد که معمولاً بعد از این مرحله احتمال خطای سیستم بالا می رود.

نمودار کلاس ۲ که کلاس خطا نیز نامیده می‌شود مرزبندی ناحیه خطا را نشان می‌دهد. با بررسی داده‌ها درصورتی‌که هرکدام از پارامتر ها به زیر نمودار کلاس ۲ بیایند می‌توان به این نتیجه رسید که در سیستم خطایی رخ داده است.

نمودار کلاس ۳ که کلاس نرمال نیز نامیده می‌شود مرز بندی کارکرد نرمال را نشان می‌دهد و هر پارامتری که در روی این خط یا بالای آن قرار داشته باشد در حالت نرمال به سر می برد.

در بررسی مناطق حاشیه ای ناحیه بین نمودار کلاس ۱ و کلاس۲ همان‌طور که انتظار می رود نه می‌توان به این ناحیه ناحیه خطا نام داد نه ناحیه خطر.ولی هرچه به محدوده ناحیه کلاس ۲ نزدیک می شویم احتمال خطا بیشتر می‌شود.

ناحیه بین نمودار کلاس ۳ و کلاس۲ همان‌طور که انتظار می رود نه می‌توان به این ناحیه ناحیه نرمال نام داد نه ناحیه خطر.ولی هرچه به محدوده ناحیه کلاس۱نزدیک میشویم احتمال خطر بیشتر می‌شود.

ازآن‌جاکه دستگاه‌های الکتریکی حفاظتی اجازه عبور جریان یا ولتاژ بیشتر از مقدار خاصی را نمی‌دهند می‌توان تمام منطقه بالای نمودار کلاس ۳ را منطقه کارکرد عادی سیستم در نظر گرفت و همان‌طور که دیدیم روش پیشنهادی به خوبی توانست داده‌های حا صل از سال پایش سیستم را به خوبی خلا صه و کلاس بندی کند و با دقتی بیش از ۹۵ در صد حالت خطا را تشخیص دهد. در انجام این مقاله وضعیت سیستم را به ۳ قسمت تقسیم کردیم که می‌توان بنا به تشخیص بهره بردار این دست بندی کاهش و یا افزایش یابد.

## مراجع


[1] Doyen L, Gaudoin O. ۲۰۰۴Classes of imperfect repair models based on reduction of failure intensity or virtual age. Reliability Engineering & System Safety ;۸۴:۴۵_.۵۶

[2] Fuxing Yu, Yina Suo, Xin Zang, Aidind Yan, Fulong Liu, (۲۰۱۳), Data mining in blast furnace smelting parameter, Applied mechanics and materials, vol. ۳۰۳-۳۰۶, pp .۱۰۹۳-۱۰۹۶

[3] Hand D. J, Manila H, & Smyth, (۲۰۰۱): Principlesof Data mining, MIT press, Cambridge, Massachusetts. ISN-۲۶۲-۰۸۲۹۰-X

[4] Jiawei Han, Micheline Kamber, Jian Pei, (۲۰۱۲), Data Mining: Concepts and Techniques, Third Edition, USA, Morgan Kaufmann Publishers.

[5] John R. Koza, ۱۹۹۲. Consulting Associate Professor in computer science department Stanford university, Genetic programming: On the programming of computers by means of natural selection, the MIT press.

[6] Jong-Hag Jeon, POSCO, Pohang South Korea, Data mining application of six-sigma project, SUGI ۲۹ solutions, paper .۱۸۶-۲۹

[7] Mahamad saraee school of computing, science and Eng., university of Salford, greater Manchester, UK, Mehdi Moghimi, Dept. of Elec. & computer Eng., Islamic Azad university, Najafabad branch, Isfahan, Iran, Ayoub bagheri, Dept. of Elec and computer Eng, Isfahan university of technology, Isfahan Iran, (۲۰۱۱), Modeling Batch Annealing Process using Data Mining Techniques. ACM journal.

[8] Manisha Verma, Mauly Srivastava, Neha Chack, Abul Kumar Diswar, Nidhi Gupta, (۲۰۱۲), 'Comparative study of various clustering algorithms in data mining'. International journal for engineering research and applications, Vol ۲, Issue ۳, pp. .۱۳۷۹-۱۳۸۴

[9] Michael Kommenda, Gabriel Kronberger Christoph Feilmayr and Michael Affenzeller, ( ۲۳Sep ۲۰۱۳), Data mining using unguided symbolic regression on a blast furnace dataset, arXiv;۱۳۰۹,۵۹۳۱v ۱[cs.NE].

[10] Michael Kommenda, Gabriel Kronberger, Christoph Feilmayr, Leonhard Schickmair, Michael Affenzeller, Stephan Winkler and Stefan Wagner, Application of symbolic regression on blast furnace and temper mill datasets, [ n.d].

[11] Moura MC, Droguett EL. Mathematical formulation and numerical treatment based on transition frequency densities and quadrature methods for non-homogeneous semi-Markov processes. Reliability Engineering & System

[12] Nine law of Data Mining by Tom Khabaza (http://wwwkdnuggets.com/۱۶/۲۰۱۵/nine datamining-part-۱:html).


## رزومه

علی رنجبر در اهواز متو لد شـــده اســـت (۱۳۷۲). تحصـــیلات دانشـــگاهی خود را در مقطع کار شنا سی مهند سی برق- کنترل از دانشگاه آزاد اسلامی واحد ماهشهر(۱۳۹۵)، کار شنا سی‌ار شد مهند سی برق- کنترل از دانشـــگاه آزاد اســـلامی واحد دزفول(۱۳۹۸) ســـپری کرده اســـت. فعالیت‌های پژوهشـــی و علاقه‌مندی ایشـــان در زمینه داده کاوی، یادگیری ماشـــین، اتوماســـیون صـــنعتی، ابزار دقیق، کنترل بهینه، کنترل مقاوم و کنترل فازی است و در حال حاضر کارشناس ارشد برق-کنترل و مدیر پروژه ریومپ ســـیســـتم برق و ابزار دقیق پمپ خانه احیا۲ شرکت فولاد خوزستان می باشد.

**امیرحسین رحمانی** در اصفهان متولد شده است(۱۳۴۷). مدار ک کارشناسی و کارشناسی ارشد خود را در سال‌های (۱۳۷۰) و (۱۳۷۴) به ترتیب در رشته‌های مهندسی برق و کنترل از دانشگاه تهران اخذ نموده است. ایشان هم اکنون دانشجوی دکترای کنترل دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران و همچنین عضو هیأت علمی دانشگاه آزاد اسلامی واحد دزفول است. زمینه تحقیقاتی ایشان تحلیل سیستم‌های کنترل و روش‌های بهینه‌سازی هوشمند است.

# Fault Diagnosis by Using the Multi-class Support Vector Machine

**Ali Ranjbar[1], Amirhossein Rahmani[2,*]**

1- Department of Electrical Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran, Takay313@gmail.com
*2- Department of Electrical Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran

**Abstract:** In industrial environments, a large amount of data is generated which in turn stores the database and data from all relevant areas such as planning, process design, materials, assembly, production, quality, process control, scheduling, error detection, shutdown, relationship management. Collects with the customer, etc. Data mining has become the tool used to gain knowledge of the industrial process of iron and steel making. Due to the rapid growth of data mining, various industries have been using data mining technology to search for hidden patterns that may be more relevant to the new windshield system, which will introduce new models to improve production quality, optimal cost of productivity and maintenance, and so on. Continuous improvement of the entire steel production process due to the avoidance of quality deficiencies and associated production improvement is an essential task of the steel producer. Therefore, the zero defect strategy is popular today and several quality assurance techniques are used to maintain it. This article attempts to identify the effective state-of-the-art sensors in the system using data mining and then to obtain a suitable model using a support vector machine to predict the system status in this article, the accuracy of more than 95% of the error states is detected.

**Keywords: Group Methods, Decision Making, Patterns, Support Vector, Nearest Neighbor.**