

ارائه مدلی جهت پیش‌بینی و بهبود کیفیت تولید با استفاده از الگوریتم‌های درخت تصمیم و برنامه‌ریزی خطی (مورد مطالعه: شرکت‌های تولیدکننده موج‌گیرتیا در ایران)

نادره سادات راست قلم

گروه مهندسی صنایع، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران

رویا محمدعلی پور اهری (مسئول مکاتبات)

گروه مهندسی صنایع، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران

roya.ahari@gmail.com

شکرچی زاده، احمد رضا

گروه مدیریت، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران

امین دوست، عاطفه

گروه مهندسی صنایع، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران

تاریخ دریافت: ۱۴۰۰-۰۱-۲۴ تاریخ پذیرش: ۱۴۰۰-۱۲-۰۳

چکیده

امروز در اکثر صنایع و کارخانه‌های کشور از ابزار کنترل کیفیت آماری جهت بهبود کیفیت محصولات استفاده می‌نمایند ولی با توجه به حجم بالای داده‌ها، در حال حاضر نیاز به ابزاری قوی‌تر می‌باشد که بتواند فرایندهای کنترل کیفیت آماری را تحت کنترل قرار دهد، با توجه به گستردگی الگوریتم‌های داده‌کاوی و توانایی آن در کشف قوانین در این تحقیق از ابزار داده‌کاوی جهت بهبود فرایند کنترل کیفیت و افزایش آن استفاده شده است و همچنین هدف اصلی تحقیق ارائه روشی جهت کشف اقلام معیوب قبل از تولید کامل آن‌ها و جلوگیری از تولید اقلام معیوب می‌باشد. روش کار بدین گونه است که در ابتدا پایگاه داده خرابی تشکیل می‌گردد و پس از جمع‌آوری داده‌های کنترل کیفیت، با استفاده از الگوریتم‌های مختلف درخت تصمیم دقت پیش‌بینی کیفیت قطعات تعیین می‌گردد و در مرحله بعد با استفاده از مدل تحلیل پوششی داده‌ها هر یک از قوانین ارزیابی می‌گردد و در نهایت با استفاده از قوانینی که در هر ایستگاه کاری صدق می‌کنند ارزیابی ایستگاه‌های کاری انجام می‌گیرد. بر این اساس در این پژوهش جامعه آماری تمامی قطعات تولیدی موج‌گیرتیا در سال ۱۳۹۸ می‌باشد. خصیصه‌ها شامل ۹ ایستگاه کاری هستند، بر اساس نتایج بدست آمده بهترین الگوریتم در پیش‌بینی خرابی C5 می‌باشد و مهم‌ترین خصیصه‌های انتخابی توسط آن نیز به‌عنوان مهم‌ترین خصیصه‌ها تعیین می‌گردند که عبارت‌اند از: کیفیت خنک‌کاری، کیفیت سوراخ‌کاری و کیفیت برش. همچنین ارزیابی قوانین نیز با استفاده از مدل تحلیل پوششی داده‌ها انجام شده و مهم‌ترین قوانین استخراج شده‌اند و در نهایت بر اساس حل مدل دستگاه‌هایی که در اولویت اصلاحی برای سال جاری قرار گیرند عبارتند از: رولینگ، لحیم‌کاری و برش

کلمات کلیدی: کنترل کیفیت، کاهش ضایعات، داده‌کاوی، مدل تحلیل پوششی داده‌ها، الگوهای درخت تصمیم.

مقدمه:

با شروع انقلاب صنعتی در اواسط قرن هجدهم، ماشین‌های تولیدی کم‌کم جایگزین ابزار و مهارت‌های فردی افراد هنرمند و صنعتگر گردید. با پیدایش روش‌های جدید و پیچیده، اشتیاق به تولید بیشتر افزایش یافت و بدین ترتیب نیاز به کنترل کیفیت نیز ابعاد تازه‌ای پیدا نمود. پایه و اساس کنترل کیفیت آماری به مفهوم امروزی آن، طی سال‌های دهه ۱۹۲۰ گذاشته شد. استفاده از روش‌های کنترل کیفیت آماری، کلیه صنایع کارخانه‌ای را در راستای بهبود بهره‌وری، جلوگیری از تولید اقلام معیوب، جلوگیری از تنظیم‌های غیرضروری و آگاهی از فرآیند یاری می‌رساند. (آندرسن و دیگران، ۲۰۱۱) طبق تجربیات متممادی، واحدهای تولیدی که از روش‌های کنترل کیفیت آماری بهره می‌برند، از بازده قابل توجهی نسبت به سایرین برخوردارند. امروز در اکثر صنایع و کارخانه‌های کشور از ابزار کنترل کیفیت آماری جهت بهبود کیفیت محصولات استفاده می‌نمایند ولی با توجه به حجم بالای داده‌ها، در حال حاضر نیاز به ابزاری قوی‌تر می‌باشد که بتواند فرایندهای کنترل کیفیت آماری را تحت کنترل قرار دهد، با توجه به گستردگی الگوریتم‌های داده‌کاوی و توانایی آن در کشف قوانین این سؤال به وجود می‌آید که چگونه می‌توان از ابزار داده‌کاوی جهت بهبود فرایند کنترل کیفیت و افزایش آن استفاده نمود و چگونه می‌توان به ارائه روشی جهت کشف اقلام معیوب قبل از تولید کامل آنها و جلوگیری از تولید اقلام معیوب پرداخت. از طرفی استفاده از تکنیک‌هایی مانند تکنیک‌های آماری شاید نتواند به حل مساله فوق کمک کند چرا که علاوه بر اینکه ممکن است متغیرها (اسیتگاه‌های کاری خط تولید) از نوع غیرکمی باشد، احتمال تاثیرگذاری متغیرها بنحوی است که دارای تاثیرات متقابل بوده و یا ممکن است تاثیرات عملا خطی نباشند لذا نیاز است تا از سایر الگوریتم‌هایی که دارای پیش نیاز تعریف شده‌ای (از جمله نرمال بودن) ندارند و توانایی پیش بینی مبتنی بر شرایط خاص را دارند استفاده نمود. همچنین استفاده از تکنیک‌هایی مانند DOE نیز نمی

توانند کمک چندانی به انجام این تحقیق نمایند زیرا تعداد عوامل باعث می‌گردد که انجام آزمایشات مستقل با دشواری همراه باشد چنانچه بعنوان مثال ۸ ایستگاه کاری وجود داشته باشد و در این صورت ۸! آزمایش باید صورت پذیرد تا حالات مختلف مورد تست قرار گیرد و اگر هر ایستگاه دارای بیش از ۲ حالت باشد در اینصورت تعداد حالات افزایش زیادتری می‌یابد و لذا نمی‌توان از این تکنیک استفاده نمود. بر این اساس در این پایان‌نامه متغیرهای مستقل داده‌های مربوط به فرایندهای کنترل کیفیت آماری در مراحل مختلف تولید از ایستگاه کاری اول تا آخرین ایستگاه کاری می‌باشند و متغیر وابسته میزان کیفیت قطعه در مرحله تست نهایی کنترل کیفیت می‌باشد و هدف از انجام این تحقیق تعیین الگوهای منجر به خرابی قطعات و تعیین خصیصه‌های مؤثر بر آن با استفاده از الگوریتم‌های مختلف درخت تصمیم می‌باشد. همچنین مقایسه تطبیقی الگوریتم‌های مختلف درخت تصمیم از جمله CHILD، QUEST TREE، C&R TREE، TREE و C5 در این تحقیق مدنظر است که این الگوریتم‌ها هر یک قوانینی را استخراج می‌کنند ولی مقایسه این قوانین و ارزیابی قوانین (قوانینی که در مورد هر یک از الگوریتم‌های درخت تصمیم استخراج شده است) بر اساس معیارهای مؤثر تا کنون در تحقیقات مشابه دیده نشده است. چنانچه می‌دانیم الگوریتم‌های درخت تصمیم هر یک بر اساس قواعد و قوانین خود به پیش بینی یک متغیر (کلاس) می‌پردازند که این پیش بینی بر اساس قوانینی خاص انجام می‌گردد.

در معبود تحقیقات انجام شده در زمینه بهینه سازی تولید با استفاده از الگوریتم‌های درخت تصمیم، تنها قوانین سطوح کیفیت تعیین شده است و بررسی نشده است که این قوانین را با چه مکانیزمی می‌توان استفاده نمود. به بیان دیگر در بهینه سازی کیفیت محصولات نمی‌توان به تنهایی و بدون توجه به متغیرهای مؤثر بر تصمیم‌گیری از جمله هزینه، دشواری اجرا، امکان بهبود کارکرد هر ایستگاه کاری، نمی‌توان نسبت به بهینه سازی خطوط تولید امیدوار

یک درخت تصمیم به طور معمول با یک گره شروع می شود، که به نتایج احتمالی منشعب می شود. هر یک از این نتایج منجر به گره های اضافی می شود که به سایر امکانات منشعب می شوند. به طور ویژه در این الگوریتم ها برای جست و جوی فضای نمایی از مدل های ممکن به کار می روند.

روش مبتنی بر تئوری داده ها، بر روی کمینه سازی تعداد آزمون هایی که به یک نمونه برای طبقه بندی در یک پایگاه داده اجازه می دهند، تأکید دارند. بخش انتخاب ویژگی از ID3 بر پایه این فرض استوار است که پیچیدگی درخت تصمیم به طور قوی با مقدار داده ها منتقل شده توسط مقدار ویژگی معین مرتبط است. داده ها ذهنی، مشخصه ای را انتخاب می کند که بیشترین داده ها به دست آمده را فراهم می کند یعنی ویژگی ای که داده ها مورد نیاز در زیر درخت ناشی از طبقه بندی نمونه را کمینه سازی می کند. بسط ID3، الگوریتم C5 می باشد که دامنه ی طبقه بندی را از صفات رده بندی به مشخصه های عددی توسعه می دهد، برآورد و سنجش مربوط به ویژگی هایی می شوند که از تقسیم داده ها به زیرمجموعه هایی که آنروپی گروهی پایینی به دست می آیند؛ یعنی، وقتی اکثریت مثال ها در آن، متعلق به یک تک گروه و کلاس می باشند. اساساً الگوریتم، مشخصه ای را انتخاب می کند که درجه ماکزیمم تمایز بین طبقه ها را به طور موضعی فراهم می کند.

برای اعمال بعضی از روش ها که بر پایه ی روش یادگیری استقرایی استوار هستند، چندین پیش نیاز اصلی باید برآورده شود:

۱. تمام داده ها در مورد یک موضوع یا مثال باید برحسب مجموعه ی ثابتی از ویژگی ها یا صفات بیان شوند. هر ویژگی و صفت ممکن است، مقادیر گسسته یا مقادیر عددی داشته باشد، اما ویژگی های به کاررفته برای توصیف نمونه ها نباید از یک مورد به مورد دیگر متفاوت باشد.
۲. گروه ها و کلاس های از پیش مشخص شده. کلاس ها و گروه هایی که نمونه ها به آن ها منسوب شده اند،

بود که این موارد در تحقیقات پیشین قابل ملاحظه نیست لذا در این تحقیق پس از استخراج قوانین، با استفاده از مدل برنامه ریزی مناسب تصمیم گیری می شود که کدامیک از قوانین در اولویت قرار گرفته اند و بر آن اساس بهینه سازی تولید انجام می گردد. بررسی تحقیقات پیشین نشان می دهد که ارتباط بین الگوریتم های داده کاوی که توانایی تولید قوانین را دارند با مدل های برنامه ریزی خطی انجام شده است و در مورد قوانین انجمنی پس از تعیین قوانین با استفاده از الگوریتمی مانند اپریوری با استفاده از معیارهایی مانند **support, confidence** و حتی **lift** رتبه بندی قوانین با استفاده از تکنیک های تصمیم گیری چندمعیاره و یا مدل های برنامه ریزی خطی انجام شده است اما در مورد قوانین درخت تصمیم حتی معیارهای ارزیابی قوانین در تحقیقات حاضر مشخص نشده است چه برسد به ارتباط آن با مدل های برنامه ریزی خطی، همچنین لازم به ذکر است که از تکنیک های تصمیم گیری دیگر (چه جبرانی و چه غیرجبرانی) مانند تاپسیس، تحلیل پوششی داده ها و ... نیز نمی توان استفاده نمود چرا که در مساله ای که در این تحقیق عنوان شده است متغیرهایی مانند هزینه و ... وجود دارد که بصورت محدودیت باید به مدل وارد شوند که تکنیک های دیگر بجز تکنیک های **OR** توانایی مدلسازی را در این زمینه ندارند. از طرفی در این تحقیق بهینه سازی دستگاهها با توجه به دو فاز انجام می گردد مه فاز نخست بررسی قوانینی است که به یک ایستگاه کاری منجر می گردند و فاز دوم در نظر گرفتن متغیرهای مناسب برای بهبود عملکرد ایستگاه کاری از نظر بودجه ای و دشواری اجرا و ... می باشد.

مبانی نظری و پیشینه تحقیق

درخت تصمیم نقشه ای از نتایج احتمالی مجموعه ای از انتخاب های مرتبط است. این به یک فرد یا سازمان اجازه می دهد تا اقدامات احتمالی را بر اساس هزینه ها، احتمالات و مزایای آنها با یکدیگر بسنجد. می توان از آنها برای هدایت بحث غیررسمی یا ترسیم الگوریتمی استفاده کرد که بهترین انتخاب را از نظر ریاضی پیش بینی می کند.

باید زودتر ایجاد شوند. در اصطلاح یادگیری ماشین این یادگیری نظارتی می باشد.

۳. داده‌های کافی. تعمیم استقرایی ارائه شده به شکل درخت تصمیم با شناسایی الگو در داده‌ها محقق می‌شود. این روش اگر تعداد کافی از الگوهای قوی از انطباق تصادفی قابل تمایز باشند، معتبر خواهد بود. همان‌گونه که این تفاوت معمولاً وابسته به آزمون‌های آماری است، تعداد کافی از نمونه‌ها برای اینکه این آزمون‌ها مؤثر باشند، باید وجود داشته باشند.

۴. مدل‌های طبقه‌بندی منطقی. این روش‌ها تنها آن دسته از رده‌بندها را ایجاد می‌کنند که بتوانند به‌عنوان درختان تصمیم یا قوانین تصمیم بیان شوند. این شکل‌ها اساساً توصیف یک کلاس به یک عبارت منطقی که پایه‌های آن عباراتی درباره‌ی مقادیر ویژگی‌ها خاص هستند، محدود می‌کند. (بایسال و دیگران، ۲۰۱۸)

الگوریتم C5، تولید درخت تصمیم

مهم‌ترین بخش الگوریتم C5، فرآیند ایجاد یک درخت اولیه‌ی تصمیم از مجموعه نمونه‌های آموزشی است. به‌عنوان یک نتیجه، این الگوریتم یک رده‌بند به شکل یک درخت تصمیم ایجاد می‌کند. یک ساختار با دو نوع گره: یک گره برگ که یک کلاس را نشان می‌دهد یا یک گره تصمیم که برخی آزمون‌ها را برای اجرا بر روی مقدار یک صفت و ویژگی با یک شاخه و زیر درخت برای هر نتیجه‌ی ممکن از آزمون تعیین می‌کند.

یک درخت تصمیم می‌تواند برای طبقه‌بندی یک نمونه‌ی جدید با شروع از ریشه‌ی درخت و حرکت در آن تا رسیدن به یک گره برگ مورد استفاده قرار گیرد. در هر گره‌ی تصمیم غیر برگ، نتیجه‌ی مشخصه‌ها برای آزمایش در گره تعیین می‌شود و توجه به ریشه‌ی زیر درخت انتخاب شده منتقل می‌شود.

چارچوب الگوریتم C5 بر پایه‌ی روش CLS هانت برای ایجاد و ساخت یک درخت تصمیم از یک مجموعه‌ی T از نمونه‌های آموزشی می‌باشد. اجازه دهید گروه‌ها و کلاس‌ها به صورت $\{C_1, C_2, \dots, C_k\}$ نمایش داده

شوند. در اینجا سه امکان برای محتوای مجموعه‌ی T وجود دارد:

۱. T شامل یک نمونه یا بیشتر می‌باشد که همه متعلق به یک تک گروه (کلاس) Cj می‌باشد، درخت تصمیم برای T، یک برگ است که گروه Cj را شناسایی می‌کند.

۲. T در اینجا حاوی هیچ نمونه‌ای نیست و درخت تصمیم دوباره یک برگ است، اما گروهی که با برگ متناظر می‌باشند، باید از داده‌ها دیگر T از قبیل گروه اکثریت کلی در T معین شود. الگوریتم C5 که به‌عنوان یک معیار برای اغلب گروه‌های عادی در اصل گره معین به کار می‌رود.

۳. T شامل نمونه‌هایی است که به ترکیبی از گروه‌ها تعلق دارد. در این شرایط، ایده اصلی پالایش T در زیرمجموعه از نمونه‌هایی است که به سمت یک مجموعه‌ی تک گروهی از نمونه‌ها پیشروی می‌کند بر پایه‌ی تک ویژگی، یک آزمون مناسب که نتیجه یا نتایج اختصاصی متقابلی دارد $\{O_1, O_2, \dots, O_n\}$ انتخاب می‌شود. T به زیر درخت‌های T_1, T_2, \dots, T_n تقسیم می‌شود به نحوی که T_1 شامل تمام نمونه‌ها در T است که دارای نتایج O_i از آزمون انتخابی می‌باشد. درخت تصمیم برای T شامل یک گره تصمیم است که آزمون لازم و یک شاخه برای هر نتیجه‌ی ممکن را شناسایی می‌کند.

به‌طور کلی C5 راهکارهایی را برای سه نوع از آزمون‌ها ارائه می‌کند:

۱. آزمون "استاندارد" بر اساس ویژگی گسستگی، با یک نتیجه و یک شاخه برای هر مقدار ممکن آن ویژگی
۲. اگر مشخصه Y، دارای مقادیر عددی پیوسته باشد، یک آزمون دودویی با نتایج و پیشامدهای $Y \leq Z$ و $Y > Z$ با مقایسه مقدار آن در مقابل یک مقدار آستانه Z می‌تواند، تعریف شود.

۳. یک آزمون پیچیده‌تر بر پایه‌ی یک ویژگی گسسته است که در آن مقادیر ممکن به یک تعداد معتبر از گروه‌ها با یک نتیجه و پیشامد و شاخه برای هر گروه اختصاص داده می‌شود (دوی و دیگران، ۲۰۱۸)

تحلیل پوششی داده ها:

$$\text{کارایی واحد } j = \frac{\sum_{r=1}^s U_r Y_{rj}}{\sum_{i=1}^m V_i X_{ij}}$$

برای محاسبه کارایی هر DMU باید بیشینه شاخص کارایی یعنی نسبت مجموع خروجی های موزون (خروجی مجازی) به مجموع ورودی های موزون (ورودی مجازی) را بدست آورد. در نتیجه برای واحد مورد ارزیابی که آنرا واحد صفر می نامیم خواهیم داشت:

$$\text{MaxZo} = \frac{\sum_{r=1}^s U_r Y_{ro}}{\sum_{i=1}^m V_i X_{io}}$$

: Subject to

$$U_r \geq 0 \quad r = 1, 2, \dots, s$$

$$V_i \geq 0 \quad i = 1, 2, \dots, m$$

در رابطه فوق u_r و v_i ها متغیرهای مسئله و وزن ها می باشند و جواب مسئله مناسب ترین مقدار را برای وزن های واحد صفر ارائه و کارایی آن را اندازه گیری می نماید. در مدل فوق ضرایب خروجی خیلی بزرگ و ضرایب ورودی خیلی کوچک خواهند بود. بنابراین جهت جلوگیری از چنین مشکلی تمامی نسبت های کارایی واحدها را کوچکتر یا مساوی یک در نظر می گیریم و به عبارتی مدل را محدود می نماییم. لازم به ذکر است که در محدودیت ها به جای عدد یک هر عدد مثبت دلخواه دیگری مانند k را می توان قرار داد. بنابراین برنامه ریزی فوق به صورت زیر تغییر می نماید.

$$\text{MaxZo} = \frac{\sum_{r=1}^s U_r Y_{ro}}{\sum_{i=1}^m V_i X_{io}}$$

:Subject to

$$\leq 1 \frac{\sum_{r=1}^s U_r Y_{rj}}{\sum_{i=1}^m V_i X_{ij}}$$

$$\geq 0, v_i \geq 0 ; r = 1, 2, \dots, s; i = 1, 2, \dots, m; u_r$$

$$j = 1, 2, \dots, n$$

مدل فوق را مدل نسبت CCR می نامند که در آن حداکثر کارایی ممکن برای واحد مورد ارزیابی (واحد صفر) است. برای خطی کردن مدل فوق می توان از روش برنامه ریزی کسری CCR استفاده نمود. (پورتلا و دیگران، ۲۰۱۴)

پیشینه تحقیق:

تحلیل پوششی داده ها روشی مبتنی بر برنامه ریزی ریاضی جهت برآورد کارایی تکنیکی و ناکارایی های واحد است. این روش بدون تعیین فرضی از شکل تابع تولید، با حل مدل های ریاضی برای مجموعه ای از واحدهای تصمیم گیرنده و با استفاده از اطلاعات مربوط به میزان ورودی های و خروجی های واقعی آن واحدها، تابع تولید یا تابع هزینه را به صورت یک پوش قطعه ای برآورد می نماید. چارلز و کوپر در مقاله معروف خود، DEA را یک مدل برنامه ریزی ریاضی معرفی نمودند که با استفاده از داده های قابل مشاهده و انجام یک سری بهینه سازی ها جهت سنجش و ارزیابی کارایی هر واحد، کارایی نسبی مجموعه ای از واحدهای تصمیم گیری با ورودی ها و خروجی های چندگانه مشابه و قطعی را تعیین و آنها را با یکدیگر مورد مقایسه قرار می دهد

واحدهای تصمیم گیرنده مورد بررسی واحدهایی مستقل اند که ورودیهای مشابهی را برای تولید خروجیهای مشابه بکار می برند. لزوم همگونی ورودیها و خروجیها شرط اول در انتخاب واحدهای مورد ارزیابی است، زیرا تمامی این واحدها و پوشش به دست آمده در یک فضا مورد بررسی قرار می گیرند. مرز به دست آمده همان مرز کارایی است که نقاط واقع بر آن نقاط کارا هستند. سایر واحدها که در سطح پوششی قرار می گیرند ناکارا هستند و برای کارا شدن باید بر روی مرز قرار گیرند.

در اندازه گیری نسبی واحدها فارل برای ساختن یک واحد مجازی بر مجموع واحدها تمرکز نموده و به عنوان یک وسیله سنجش متداول برای اندازه گیری کارایی فنی رابطه ذیل را پیشنهاد نمود.

مجموع موزون ورودی ها / مجموع موزون خروجی ها = کارایی

در صورتی که هدف بررسی کارایی n واحد که هر کدام دارای m ورودی و s خروجی است باشد، کارایی واحد j ام به صورت زیر محاسبه می شود.

پرداختند. این مقاله یک روش ترکیبی برای حل برنامه ریزی نگهداری تولید ارائه می دهد. مدل ریاضی پیشنهادی اجرا شده و راه حل بهتر کاهش هزینه ۱۰.۷۲٪ کمتر از برنامه پایه است. کلمیکاوای و همکاران^۵ (۲۰۱۶) پژوهشی تحت عنوان طراحی درخت تصمیم بر مبنای ویژگی عملکردی در زمینه شرایط تصادفی مورد استفاده برای تشخیص خطا در داده های مشاهده شده در اقیانوس را انجام داده اند. برای پیش بینی آب و هوا از داده های اقیانوس استفاده می شود. سنگ و همکاران^۶ (۲۰۱۶) پژوهشی تحت عنوان کنترل کیفیت الکترونیکی: روش ماشین های برداری پشتیبان (SVM) انجام داده اند. در این پژوهش روشهای بازرسی مبتنی بر سنسور، اتوماتیک و کامپیوتری جایگزین کنترل کیفیت سنتی مبتنی بر تکنیک های نمونه گیری شده است. تجهیزات تولیدی به شبکه متصل شده اند، وضعیت ماشین آلات تحت نظارت قرار داده می شود. سنسورهای بصری ابعاد را اندازه می گیرند و کلیه داده ها را در سیستم ثبت می کنند. نتایج تجزیه و تحلیل توسط الگوریتم های کامپیوتری تولید می شوند. هوانگ و همکاران^۷ (۲۰۱۶) در تحقیقی به پیش بینی فروش با استفاده از تکنیک های داده کاوی پرداختند. نتایج حاکی از آن است که دقت مدل ارائه شده بالاتر از یک مدل واحد است. فروشندگان می توانند از سیستم پیشنهادی برای پیش بینی دقیق فروش کالاهای مختلف استفاده کنند.

روش شناسی مقاله

این پژوهش از نظر هدف کاربردی است چون مجموعه ای از قواعد و ابزارهای معتبر (قابل اطمینان) و نظام یافته برای بررسی واقعیت ها و کشف مجهولات و دستیابی به راه حل مشکلات است. از آنجایی که این پژوهش به پیش بینی کیفیت می پردازد در قلمرو پژوهش های علی است و از آنجایی که از ابزار میدانی جهت گردآوری داده ها استفاده می شود یک پژوهش پیمایشی است.

دوی و همکاران^۱ (۲۰۱۸) در تحقیقی به پیش بینی کیفیت آب پرداختند. آنها در این تحقیق از نرم افزار R جهت ارائه تحلیل هایی برای پیش بینی کیفیت آب معدنی پرداختند. در این تحقیق از ابزارهایی مانند پای چارت و اسکات پلات و همچنین از خوشه بندی برای بررسی عواملی که موثر بر کیفیت بالا و پایین می شوند استفاده شده است. بر اساس نتایج بدست آمده از این تحقیق با استفاده از ابزارهای بیان شده می توان تا حد قابل ملاحظه ای پیش بینی کیفیت را به درستی انجام داد. لوکاس^۲ (۲۰۱۸) به پیش بینی کیفیت تماس های تلفنی و راهکارهای بهبود آن با استفاده از الگوریتم های شبکه بیزین پرداخته است در این تحقیق می توان با استفاده از متغیرهای مختلف رضایت مشتری از تماس را پیش بینی نمود و قوانینی که منجر به رضایت بالا و پایین نیز می گردند، مورد تعیین قرار گرفته اند. بوردین و دیگران (۲۰۱۷) به ارائه مدلی با استفاده از مدل های برنامه ریزی خطی برای تجزیه و تحلیل ضعف باطری در سیستم های برق خورشیدی پرداخته اند. بر این اساس در این مقاله هدف جایگزینی باتری های جدید برای جلوگیری از افزایش هزینه های تعمیراتی می باشد بر این اساس مدل برنامه ریزی خطی تشکیل می گردد. بایسال و همکاران^۳ (۲۰۱۷) پژوهشی تحت عنوان تحلیل درخت تصمیم گیری از انتشارات گذشته در مورد اصلاح بخار کاتالیزوری برای توسعه روش های تحقیقی برای بهبود عملکرد: یک مقاله مروری آماری انجام داده اند. در این پژوهش پایگاه داده شامل ۵۵۰۸ مقاله داده آزمایشی برای اصلاح بخار متان با استفاده از ۸۱ مقاله (از ۴۵۳ مقاله ای که در ابتدا بررسی شدند) منتشر شده بین سال های ۲۰۰۴ و ۲۰۱۴ فراهم شده که این داده ها توسط الگوریتم های درخت تصمیم مورد تحلیل قرار گرفتند و قوانین استخراج شدند. الویس و همکاران^۴ (۲۰۱۷) در تحقیقی به بررسی برنامه ریزی خطی و الگوریتم ژنتیک برای برنامه ریزی نگهداری و هیدروترمال تولید

^۵ Kamikawaji et al

^۶ Seng et al

^۷ Huang et al

^۱ Devi et al

^۲ Loukas

^۳ Baysal et al

^۴ Alves et al

خصیصه‌ها: خصیصه‌ها شامل معیارهایی است که بر روی کیفیت موج گیر تأثیر می‌گذارند که بر این اساس خصیصه‌ها در این پژوهش عبارت‌اند از: ۱. کیفیت برش ۲. کیفیت ماشین کاری ۳. کیفیت رولینگ ۴. کیفیت خنک کاری ۵. کیفیت لحیم کاری ۶. کیفیت ترک یابی ۷. کیفیت قابگیری ۸. کیفیت فرمینگ ۹. کیفیت سوراخ کاری کلاس (برچسب): این معیار نشان‌دهنده کیفیت و قابل‌استفاده بودن محصول است که پس از تأیید واحد کنترل کیفیت در ۳ دسته به شرح زیر قرار می‌گیرد: ۱. کیفیت بالا: که محصول قابل‌استفاده می‌باشد. ۲. کیفیت متوسط: که محصول نیاز به انجام اقدام اصلاحی دارد. ۳. کیفیت پایین: ضایعات تولید محسوب می‌گردد که در جدول ۱ توصیف کیفیت تولید شده قابل ملاحظه است.

جدول ۱: توصیف کیفیت تولید شده

درصد	تعداد	
۸۳/۸۵	۳۴۹۶	بالا
۵/۷۸	۲۴۱	پایین
۱۰/۳۶	۴۳۲	متوسط
۱۰۰	۴۱۶۹	مجموع

در داده‌کاوی قبل از پیاده‌سازی الگوریتم‌ها معمولاً دقت پایگاه داده را آنالیز می‌کنند. در این راستا دقت الگوریتم طبقه‌بندی مورد استفاده را مورد تحلیل قرار می‌دهند که در جدول ۲ دقت الگوریتم درخت دیده می‌شود.

جدول ۲: دقت الگوریتم‌های درخت تصمیم

الگوریتم	تعداد پیش‌بینی صحیح	تعداد پیش‌بینی اشتباه	دقت (درصد)
QUEST TREE	۳۶۷۶	۴۹۳	۸۸
C&R TREE	۳۷۶۳	۴۰۶	۹۰
C5	۳۹۲۱	۲۴۸	۹۴

همان‌طور که در جدول ۲ دیده می‌شود الگوریتم دارای دقت بیش از ۸۰٪ است بنابراین می‌توان نتیجه گرفت که داده‌ها برای داده‌کاوی مناسب هستند. در این مقاله از ۳ الگوریتم درخت تصمیم برای پیاده‌سازی و تعیین قوانین استفاده شده است که نتایج آنها در اشکال ۱ الی ۳ قابل ملاحظه است.

روش جمع‌آوری داده‌ها در این پژوهش به ۲ صورت بوده است که به شرح زیر می‌باشد:

ابتدا با استفاده از مطالعات کتابخانه‌ای داده‌ها لازم در مورد پژوهش جمع‌آوری و انتخاب گردید که بر اساس این مدل‌های مربوط به داده‌کاوی و روش‌های کنترل کیفیت آماری مورد بررسی قرار گرفتند. از روش‌های میدانی نیز جهت جمع‌آوری داده‌های لازم برای مراحل مختلف داده‌کاوی استفاده می‌گردد. ابزار گردآوری داده‌ها به روش‌هایی اطلاق می‌شود که از طریق آن‌ها داده‌های مورد نیاز جمع‌آوری می‌گردد. در این پایان‌نامه ابزار گردآوری داده‌ها فیش‌ها و جداول موجود در واحد تضمین کیفیت شرکت زرسایپا می‌باشد. بر این اساس در این پژوهش جامعه آماری تمامی قطعات تولیدی موج‌گیر تپا در سال ۱۳۹۸ می‌باشد. همچنین نمونه آماری این پژوهش تولیداتی هستند که در شرکت زرسایپا مورد کنترل آماری در خط تولید قرار گرفته‌اند.

مراحل تجزیه و تحلیل داده‌ها:

مرحله ۱: تشکیل پایگاه داده مرحله ۲: پیاده‌سازی الگوریتم‌های درخت تصمیم و تعیین الگوهای درخت تصمیم مرحله ۳: تعیین دقت پیش‌بینی هر الگوریتم با استفاده از اعتباریابی ۱۰ تایی مرحله ۴: تعیین کارایی قوانین در سطح هر درخت با تشکیل مدل تحلیل پوششی داده‌ها مرحله ۵: تعیین کارایی قوانین بر اساس دقت الگوریتم‌های درخت تصمیم با استفاده از مدل تحلیل پوششی داده‌ها مرحله ۶: تعیین درجه عضویت هر ایستگاه کاری در هر یک از قوانین مرحله ۷: تشکیل مدل برنامه ریزی خطی برای ارزیابی ایستگاه‌های کاری برای اصلاح

نتایج:

در این بخش پایگاه داده‌ها به شرح زیر توضیح داده خواهد شد:

رکورد‌ها: شامل ۴۱۶۹ موج‌گیر تپا هستند که از لحاظ کنترل کیفیت مورد بررسی قرار گرفته و با توجه به کیفیت قطعه در دسته‌های مختلف تقسیم‌بندی شده‌اند.

	۳	۰/۳۵	۳۶
	۴	۰/۹۹	۱۴
	۵	۰/۲۹	۴۰
	۶	۱	۱
	۷	۰/۰۹	۴۸
	۸	۰/۳۵	۳۵
	۱۳	۱	۱
	۱۴	۰/۶۰	۳۰

	۱	۱
	۰/۰۹	۱۰
	۰/۳۶	۷
	۱	۱
	۰/۶۰	۶

در این مرحله ارزیابی قوانین با تشکیل مدل تحلیل پوششی داده ها انجام شده است. بطور کلی در مدل تشکیل شده ورودی خروجی به شرح زیر است:

ورودی: عدد ثابت ۱ است.

خروجی: خروجی شامل ۴ متغیر مختلف است که عبارتند از: تعداد معیوب، تعداد دوباره کاری، احتمال معیوب و احتمال دوباره کاری که در مقدار دقت الگوریتم درخت تصمیم ضرب شده است که در جدول ۴ نتیجه مدل تحلیل پوششی داده ها در مورد درخت **C&R TREE** قابل ملاحظه است.

جدول ۵: مدل کلی تحلیل پوششی داده ها برای کلیه قوانین

الگوریتم های ۳ گانه درخت تصمیم

درخت	دقت الگوریتم	شماره نود	تعداد معیوب	تعداد دوباره کاری	احتمال معیوب	احتمال دوباره کاری
C&R TREE	۰/۹	۱	۵۴	۳۲۷	۵	۶۵
		۲	۱۸۷	۱۰۵	۶	۳
		۳	۲۰	۱۱۰	۳	۱۴
		۴	۳۴	۲۱۷	۱۱	۷۲
		۵	۱۸	۹۳	۱	۴
		۶	۱۶۹	۱۲	۲۴	۲
		۷	۴	۲۶	۱	۶
		۸	۱۶	۸۴	۵	۲۴
		۱۳	۷۰	۴	۷۷	۴
		۱۴	۹۹	۸	۱۶	۱

با استفاده از نرم افزار دیپ مدل تحلیل پوششی داده ها حل می شود که نتیجه ارزیابی قوانین در جدول ۶ قابل ملاحظه می باشد.

جدول ۶: ارزیابی کلی قوانین

درخت	شماره نود	کارایی	رتبه
C&R TREE	۱	۱	۱
	۲	۱	۱

در این مرحله طبق متدولوژی در ابتدا ماتریس قوانین دستگاه ها تشکیل می گردد که در جدول ۷ قابل ملاحظه می باشد.

جدول ۷: ماتریس ارزیابی دستگاه ها بر اساس قوانین

درخت	شماره نود	خنک کاری	رولینگ	لحیم کاری	برش	قابگیری
C&R TREE	۱	۱	۰	۰	۰	۰
	۲	۱	۰	۰	۰	۰
	۳	۱	۰/۳۵	۰	۰	۰
	۴	۱	۰/۹۹	۰	۰	۰
	۵	۰/۲۹	۰	۰	۰	۰
	۶	۱	۰	۰	۰	۰
	۷	۱	۰/۳۵	۰	۰	۰/۰۹
	۸	۱	۰/۳۵	۰	۰	۰/۳۵
	۱۳	۱	۱	۰	۰	۰
	۱۴	۱	۰/۶۰	۰	۰	۰

با توجه به اینکه ۵ دستگاه برای اصلاح انتخاب شده اند مدل برنامه ریزی برای اصلاح برای این ۴ دستگاه تشکیل می گردد که عبارت است از:

$$\begin{aligned} \text{MODEL:} \\ \text{MAX} &= 0.53*x_1 + 0.55*x_2 + 0.18*x_3 + \\ & 0.02*x_4 + 0.01*x_5 ; \\ x_1+x_2+x_3+x_4+x_5 &= 3; \end{aligned}$$

$$50*x_1+35*x_2+17*x_3+1*x_4+2.5*x_5 < 70;$$

$$\text{@BIN}(x_1);$$

$$\text{@BIN}(x_2);$$

$$\text{@BIN}(x_3);$$

$$\text{@BIN}(x_4);$$

$$\text{@BIN}(x_5);$$

END

در این مدل تابع هدف بر اساس درجه عضویت دستگاه ها در قوانین تعریف می گردد. قید اول در مورد قابلیت فنی اصلاح دستگاه ها می باشد. قید دوم در مورد بودجه اصلاح هر یک از دستگاه ها می باشد. بر اساس حل مدل دستگاه هایی که در اولویت اصلاحی برای سال جاری قرار گیرند عبارتند از: رولینگ، لحیم کاری و برش بحث و نتیجه گیری:

در بحث کنترل کیفیت آماری نمودارها و فرآیندهای متنوعی وجود دارند ولی تمامی این فرایندها و نمودارها حالت توصیفی دارند و هیچ یک از آنها خاصیت پیش بینی کنندگی ندارد. بر این اساس در این پژوهش با دیدگاهی جدید به این مسئله نگاه کرده و سعی در پیش بینی معیوبی محصولات نمودیم. بنا بر مطالب بیان شده در فصل اول ابزار داده کاوی به دلیل خاصیت پیش بینی کنندگی توانایی مناسبی در زمینه های مختلف دارد. بر این اساس در این پایان نامه با استفاده از داده کاوی ۳ هدف عمده را دنبال نمودیم، یکی مهم ترین عوامل مؤثر بر خرابی قطعات و دیگری پیش بینی خرابی قطعات و در نهایت مهمترین راهکار برای جلوگیری از خرابی است. در مورد اول با ۳ الگوریتم مهم ترین دلایل خرابی را تعیین نمودیم که از آنجائی که دقت پیش بینی الگوریتم ۵C بیش از سایر الگوریتم ها می باشد لذا مهم ترین خصیصه های انتخابی توسط آن نیز به عنوان مهم ترین خصیصه ها تعیین می گردند که عبارت اند از: کیفیت خنک کاری، کیفیت سوراخ کاری و کیفیت برش. همچنین قوانین منجر به خرابی نیز می تواند کمک شایانی به افزایش کیفیت کند. بر این اساس مهمترین قوانین خرابی عبارتند از: ۱. کیفیت خنک کاری خیلی کم ۲. کیفیت خنک کاری کم و کیفیت رولینگ خیلی کم ۳. کیفیت خنک کاری کم و کیفیت رولینگ متوسط و کیفیت لحیم کاری غیر زیاد ۴. کیفیت خنک کاری کم و کیفیت رولینگ زیاد و کیفیت لحیم کاری غیر زیاد ۵. کیفیت خنک کاری متوسط و کیفیت رولینگ زیاد و کیفیت لحیم کاری غیر زیاد ۶. کیفیت خنک کاری متوسط و کیفیت رولینگ کم و کیفیت لحیم کاری غیر زیاد ۷. کیفیت خنک کاری متوسط

و کیفیت رولینگ خیلی کم و کیفیت لحیم کاری غیر زیاد ۸. کیفیت خنک کاری متوسط ۹. کیفیت خنک کاری غیر متوسط ۱۰. کیفیت خنک کاری خیلی کم و کم ۱۱. کیفیت خنک کاری کم و خیلی کم و کیفیت رولینگ متوسط و زیاد

همچنین در بحث بهبود اساسی ایستگاه های کاری نیز اولویت بندی راهکارها ارائه شده است.

بر اساس نتایج بدست آمده از این پژوهش می توان پیشنهادات عملی را ارائه نمود که عبارتند از:

در راستای مهم ترین خصیصه ها سه خصیصه به عنوان خصیصه مهم تر تشخیص داده شده اند که پیشنهادات در راستای بهبود کیفیت قطعات در هر یک از ایستگاه های کاری به شرح زیر مطرح می گردد:

کیفیت خنک کاری: در راستای بهبود عملکرد این ایستگاه کاری پیشنهاد می گردد که ماشین کاری در دو مرحله انجام گردد تا از این طریق کیفیت تا حد زیادی بهبود یابد.

کیفیت سوراخ کاری: با توجه به این که در حال حاضر کیفیت سوراخ کاری با فرآیندهای نادقیق و غیراستاندارد انجام می گردد که تاثیر منفی بر کیفیت قطعات می گذارد پیشنهاد می گردد فرآیند این ایستگاه کاری تغییر نماید.

کیفیت برش: در این راستا پیشنهاد می گردد استفاده از روش های برش ترکیبی از جمله برش هوا و آرگون و یا برش با ترکیب پودر گچ از جمله روش های مناسب برای بالا بردن کیفیت برش پیشنهاد می گردد.

منابع:

presented at the International Conference on Computer Supported Education

Chang, K.-M., Beck, J., Mostow, J., and Corbett, A. (2006). A bayes net toolkit for student modeling in intelligent tutoring systems. In: Paper presented at the Intelligent tutoring systems.

Charnes. A., Cooper. W.W., and Rhodes., E. (1978). Measuring the efficiency of making units. *European Journal Of Operational Research*. (2), 429-444.

Chen, N.-S., Wei, C.-W., and Chen, H.-J. (2008). Mining e-Learning domain concept map from academic articles. *Comput. Educ*, 50 (3), 1009–1021.

Chen, Y.-L., and Weng, C.-H. (2009). Mining fuzzy association rules from questionnaire data. *Knowledge-Based Syst*. 22 (1), 46–56.

Devi, S., Jothi, S., and Devi, A. (2018). Data Mining Case Study for Water Quality Prediction using R Tool. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 3, 262-269.

Huang, W., Zhang, Q., Xu, W., Fu, H., Wang, M., and Liang, X,A. (2016). Novel Trigger Model for Sales Prediction with Data Mining Techniques, *Data Science Journal*. Volume 14, p.15. DOI: <http://doi.org/10.5334/dsj>

Kamikawaji, Y., Matsuyama, H., Fukui, K., Hosoda, S., and Ono, S. (2016). Decision Tree-based Feature Function Design in Conditional Random Field Applied to Error Detection of Ocean Observation Data, *Computational Intelligence (SSCI)*, Volume 4, 1-8.

Kim, S., Jipitaklert, W., Park, S., and Hwang, S. (2014). Data mining model-based control charts for multivariate and auto correlated processes. *Expert Systems*

Alves, M., Ramirez, M., Guimaraes, F., and Escobar, A. (2017). Linear Programming And Genetic Algorithm For Generation Maintenance Scheduling And Hydrothermal Dispatch Considering Uncertainties In Multicriteria Decision Making. *XLIX Simpósio Brasileiro de Pesquisa Operacional, Blumenau-SC*, Volume 27 a 30, 1-12.

Andersen, K., Cook, G.E and Ramaswamy, K. (2011). Artificial Neural Networks Applied to Arc Welding Process Modeling and Control. *IEEE Transaction on Industrial Application*, Volume 26, pp. 824-830.

Baysal, M., Erdem,Gu'nay, M., and Yildirim, R. (2017). Decision tree analysis of past publications on catalytic steam reforming to develop heuristics for high performance: A statistical review. *International Journal Of Hydrogen Energy*, **2017, Volume 4 2 , 2 4 3-2 5 4**.

Bordin, CH., Anuta, H., Crossland, A., Gutierrez, L., Dent, Ch and Vigo, D. (2017). A linear programming approach for battery degradation analysis and optimization in offgrid power systes with solar energy integration. *Renewable Energy*, Elsevier, vol. 101©, pages 417-430.

Bouslah, B., Gharbi, A., and Pellerin, R. (2016). Integrated production, sampling quality control and maintenance of deteriorating production systems with AOQL constraint. *Omega*, 61(4), 110–126.

Buldu, A., and Üçgün, K., (2010). Data mining application on students' data. *Proc. Social Behav. Sci*. 2 (2), 5251–5259.

Campagni, R., Merlini, D., and Verri, M.C. (2014). An Analysis of Courses Evaluation Through Clustering. In: Paper

Wang, Y. M., and Chin, K. S. (2011). Fuzzy data envelopment analysis: A fuzzy expected value approach. *Expert Systems with Applications*, 38(9), 11678-11685.

Wu, D. D., Yang, Z., and Liang, L. (2006). Efficiency analysis of cross-region bank branches using fuzzy data envelopment analysis. *Applied Mathematics and Computation*, 181(1), 271-281.

Zhou, K., Yao, P., and Cai, L. (2015). Constant current vs. constant power control in AC resistance spot welding, *J. Mater. Process. Technol.* 223, 299–304.

with Applications, Volume 39, Issue 2, Pages 2073–2081.

Kumar, D., Rahman, Z., and Chan, F. (2016). A fuzzy AHP and fuzzy multi-objective linear programming model for order allocation in a sustainable supply chain: A case study, *International Journal of Computer Integrated Manufacturing*, Volume 30, Issue 6, 535-551.

Loukas, K. (2018). Quality improvement calls data mining: the case of the seven new quality tools, *Benchmarking: An International Journal*, Volume, 47-75.

PodrżajP., and Simončič, S. (2011). Resistance spot welding control based on fuzzy logic. *Int. J. Adv. Manuf. Technol.* 52 (9–12) 959–967.

Portela, M. S., Thanassoulis, E., and Simpson, G. (2004). Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society*, 55(10), 1111-1121.

Rayward-Smith, V.J. (2007). Statistics to measure correlation for data mining applications. *Comput. Stat. Data Anal.* 51 (8), 3968–3982.

Shaleena, K., and Paul, S. (2015). Data mining techniques for predicting student performance. In: Paper presented at the Engineering and Technology (ICETECH), 2015 IEEE International Conference.

Tseng, Y.L., Lin, K.Sh., and Jaw, F.Sh. (2016). Comparison of support-vector machine and sparse representation using a modified rule-based method for automated myocardial ischemia detection. *Computational and Mathematical Method in Medicine.* (6):1-8.

Tzu-Liang Tseng, B., Kalyan Reddy, A., Zhonghua, H., and Yongjin, K. (2016). E-quality control: A support vector machines approach. *Journal of Computational Design and Engineering*, Volume 3, 91–101.