

بهینه‌سازی ارزیابی عملکرد منابع انسانی در شرکت‌های کوچک و متوسط (SMEs) با استفاده از

الگوریتم‌های داده‌کاوی (مطالعه موردی: شرکت‌های کاشی و سرامیک استان یزد)

امیر احسان اسحاقیه فیروزآبادی

پژوهشگر پسادکتری گروه مدیریت، دانشگاه میبد، میبد، ایران.

(ایمیل: dr.eshaghiyeh@gmail.com)

محمد زارعی محمودآبادی* (نویسنده مسئول)

دانشیار گروه مدیریت صنعتی، دانشگاه میبد، میبد، ایران.

(ایمیل: zareei.m@meybod.ac.ir)

چکیده

زمینه و هدف: این پژوهش با هدف طراحی یک چارچوب ترکیبی داده‌کاوی برای ارزیابی عملکرد منابع انسانی در شرکت‌های کوچک و متوسط (SMEs) صنعت کاشی و سرامیک یزد انجام شد. تمرکز اصلی بر رفع محدودیت‌های روش‌های سنتی از طریق پیاده‌سازی الگوریتم‌های نوین یادگیری ماشین است.

روش‌شناسی: در این تحقیق کاربردی-توسعه‌ای، از داده‌های ۲۰۰ کارمند در SMEs یزد با ۱۰ شاخص کلیدی ارزیابی عملکرد (استخراج‌شده از ادبیات پژوهش) استفاده شد. پس از پیش‌پردازش داده‌ها در پلتون (Google Colab) چهار الگوریتم داده‌کاوی شامل: درخت تصمیم (CART)، K-NN (با بهینه‌سازی فاصله اقلیدسی)، نایو بیز، ماشین بردار پشتیبان (SVM) با کرنل RBF و رگرسیون لجستیک، پیاده‌سازی و با معیارهای دقت (Accuracy)، صحت (Precision)، بازخوانی (Recall) و F1-Score ارزیابی شدند.

یافته‌های کلیدی: الگوریتم رگرسیون لجستیک با دقت ۹۸/۳٪ و F1-Score ۹۷٪ بهترین عملکرد را در طبقه‌بندی عملکرد کارکنان نشان داد. SVM با دقت ۹۶/۱٪ جایگاه دوم را کسب کرد، درحالی‌که نایو بیز با اختلاف معنادار (دقت ۸۷/۵٪) ضعیف‌ترین نتایج را ارائه کرد و همچنین شاخص‌های "توسعه حرفه‌ای" و "توانایی حل مسئله" بیشترین تأثیر را در مدل نهایی داشتند.

نتیجه‌گیری و ارزش پژوهش: مدل پیشنهادی مبتنی بر رگرسیون لجستیک، قابلیت کاهش ۳۰٪ خطای ارزیابی عملکرد در مقایسه با روش‌های مرسوم SMEs را دارد. این چارچوب نه تنها سازماندهی منابع انسانی را عینی‌سازی می‌کند، بلکه با تفسیرپذیری بالای مدل، امکان شناسایی نقاط بهبود عملکرد کارکنان جدید و موجود را فراهم می‌سازد.

کلمات کلیدی: مدل داده‌کاوی، ارزیابی عملکرد منابع انسانی، رگرسیون لجستیک، یادگیری ماشین کاربردی، شرکت‌های کوچک و متوسط (SMEs).

۱. مقدمه

در سال‌های اخیر، با رشد چشمگیر فناوری‌های اطلاعات و علوم کامپیوتر، داده‌کاوی به عنوان رویکردی نوین و کارآمد در پاسخ به چالش‌های مدیریت سنتی مورد توجه گسترده قرار گرفته است. داده‌کاوی به عنوان یکی از شاخه‌های پیشرفته تحلیل داده، قابلیت کشف دانش و اطلاعات نهفته در دسته‌های عظیم داده را فراهم می‌کند و فرصت‌های ارزنده‌ای برای استخراج بینش‌های جدید پیش روی تحلیلگران قرار می‌دهد (۱). فرآیند داده‌کاوی عمدتاً معطوف به کشف الگوها و روابط پنهان در سامانه‌های داده‌ای وسیع بوده و این قابلیت نقش مهمی در پیش‌بینی روندهای آتی و تصمیم‌گیری‌های مبتنی بر داده در سازمان‌ها ایفا می‌کند (۲). این فرآیند پس از طی مراحل پاک‌سازی و پیش‌پردازش داده‌ها آغاز شده و سرفصل‌هایی همچون شناسایی ناهنجاری‌ها، تحلیل ارتباطات، خوشه‌بندی، طبقه‌بندی، رگرسیون و خلاصه‌سازی را شامل می‌شود (۳).

در عرصه مدیریت منابع انسانی نیز، به دلیل حجم عظیم داده‌های حاصل از فرایندهای مختلف سازمانی، نیاز به بهره‌گیری از فناوری‌های تحلیل داده بیش از پیش احساس می‌شود. در حالی که سامانه‌های پایگاه داده فعلی عمدتاً عملیات پایه‌ای نظیر ورود، بازیابی و پردازش آماری داده‌ها را به شکل مطلوب انجام می‌دهند، اما در کشف روابط پیچیده و قوانین سیستماتیک نهفته در داده‌های پرسنلی ناتوان هستند (۲). به همین منظور، تکیه بر روش‌های داده‌کاوی برای تجزیه و تحلیل هوشمندانه این داده‌ها امری ضروری تلقی می‌شود. از جمله روش‌های کلیدی در این زمینه می‌توان به الگوریتم‌های طبقه‌بندی، خوشه‌بندی، تحلیل وابستگی، تحلیل رگرسیون و شبکه‌های عصبی اشاره کرد که هر کدام می‌توانند به شیوه‌های متفاوت در ارتقا مدیریت منابع انسانی نقش آفرین باشند (۳). برای نمونه، بهره‌گیری از یادگیری ماشین و شبکه‌های عصبی عمیق امکان اخذ تصمیمات دقیق‌تر و کارآمدتر را در این حوزه میسر می‌سازد (۴). کاربردهای داده‌کاوی در منابع انسانی همچنین در حوزه‌هایی چون مدیریت مشتری، تولید و مالی نمود می‌یابد. مدیریت مشتری بر شناسایی، جذب، حفظ و توسعه روابط مؤثر با مشتریان تمرکز دارد، در حالی که مدیریت تولید به مباحثی نظیر کنترل کیفیت، زمان‌بندی، تشخیص خطا و تحلیل نقایص فرایندی می‌پردازد. مدیریت مالی نیز حوزه مهمی برای کشف و پیشگیری از انواع تقلب‌های مالی فراهم آورده است که شامل تعرفه‌های بانکی، بیمه، اوراق بهادار و تخلفات متنوع مالی می‌شود (۵).

از سوی دیگر، داده‌کاوی این قابلیت را ایجاد می‌کند تا اطلاعات ارزشمند مرتبط با وضعیت شغلی کارکنان از پایگاه‌های گوناگون استخراج شده و برای اهدافی چون آموزش و جذب، استعدادیابی، تدوین نظام‌های حقوق و رفاه رقابتی و همچنین ارتقای سیستم ارزیابی عملکرد مورد استفاده قرار گیرد (۲). تحلیل دقیق داده‌ها می‌تواند مدیران منابع انسانی را در اتخاذ تصمیمات استراتژیک درباره مدیریت استعداد، جذب نیرو و سنجش عملکرد، به شکل قابل توجهی یاری دهد و الگوها و روندهایی را آشکار سازد که منجر به بهبود بهره‌وری و رضایت شغلی پرسنل خواهند شد (۶). با وجود این پتانسیل‌ها، مرور متون تخصصی نشان می‌دهد که تعداد پژوهش‌هایی که روی طراحی سیستم‌های مدیریت منابع انسانی مبتنی بر داده‌کاوی، به‌ویژه در شرکت‌های کوچک و متوسط (SMEs)، تمرکز داشته‌اند بسیار محدود است. از طرف دیگر، سیستم‌های مرسوم مدیریت منابع انسانی عمدتاً به نگهداری داده‌های ابتدایی نظیر سن و سوابق تحصیلی بسنده می‌کنند و تحلیل‌های عمیق و پشتیبانی هوشمندانه در فرآیند تصمیم‌گیری ارائه نمی‌دهند. این در حالی است که عملیات روزمره سازمان‌ها، روزانه حجم بسیار بالایی از داده‌های کلیدی را تولید می‌کند که حاوی اطلاعات اساسی برای موفقیت سازمان است (۷).

در نتیجه، مدیریت سنتی منابع انسانی نه تنها زمان‌بر و پرهزینه است، بلکه مستعد بروز خطاهای قابل توجهی نیز می‌باشد (۱) و کارایی لازم برای پاسخگویی به نیازهای روبه‌رشد بنگاه‌های اقتصادی را ندارد. از این رو، به‌روزرسانی و بهینه‌سازی سیستم‌های مدیریت منابع انسانی از مسیر داده‌کاوی، ضرورتی اجتناب‌ناپذیر است (۸). پژوهش کنونی بر آن است تا با بهره‌گیری از قابلیت‌های پیشرفته داده‌کاوی، نسبت به طراحی و پیاده‌سازی یک سیستم مدیریت منابع انسانی ویژه شرکت‌های کوچک و متوسط استان یزد اقدام نماید، که نتایج آن می‌تواند بستری مؤثر برای تدوین راهبردهای علمی و کاربردی، ارتقا فرآیندهای تصمیم‌گیری و افزایش رقابت‌پذیری سازمان‌ها فراهم آورد.

۲. چارچوب نظری

۲.۱. سیستم مدیریت منابع انسانی

در گذشته، مدیریت منابع انسانی معمولاً به صورت دستی و با اتکا بر پرونده‌های کاغذی انجام می‌گرفت که این شیوه‌ها علاوه بر زمان‌بر بودن، منجر به بروز خطاهای انسانی نیز می‌شدند. با رشد فناوری، نرم‌افزارهای مدیریت منابع انسانی توسعه یافتند و ابتدا به صورت جداگانه و محدود در سازمان‌ها ایفای نقش می‌کردند، اما امروزه سیستم‌های مدرن HRMS به شکل یکپارچه و عمدتاً مبتنی بر ابر (Cloud) پیاده‌سازی می‌شوند و طیف وسیعی از قابلیت‌ها را به سازمان‌ها عرضه می‌کنند (۹). سیستم مدیریت منابع انسانی (HRMS) یک بستر نرم‌افزاری جامع است که همه فرآیندهای مرتبط با نیروی انسانی را به شکل خودکار و یکپارچه مدیریت کرده و با ارائه ابزارهای پیشرفته، کنترل و نظارت بر حوزه‌هایی مانند جذب و استخدام، حقوق و دستمزد، حضور و غیاب، آموزش و توسعه، و ارزیابی عملکرد را میسر می‌سازد (۱۰). نگهداری متمرکز داده‌های کارکنان، خودکارسازی فرآیندهای تکراری و استفاده از گزارش‌های تحلیلی استراتژیک، نه تنها بهره‌وری را افزایش و هزینه‌ها را کاهش می‌دهد، بلکه رضایت و تجربه کارکنان را نیز بهبود می‌بخشد (۹).

در فضای رقابتی کنونی، سیستم HRMS یک ضرورت برای سازمان‌های کوچک و بزرگ محسوب می‌شود، به گونه‌ای که با قابلیت‌های انعطاف‌پذیر خود، سازمان‌ها را در بهینه‌سازی بهره‌برداری از سرمایه انسانی و دستیابی به اهداف بلندمدت یاری می‌دهد (۹). این سیستم‌ها با خودکارسازی فرآیندها، ارتقای دسترسی به داده‌ها و پشتیبانی از تصمیم‌گیری‌های راهبردی، نقشی کلیدی در موفقیت سازمانی ایفا می‌کنند. در آینده نیز با تلفیق فناوری‌هایی چون هوش مصنوعی و تحلیل‌های پیش‌بینانه، HRMS به طور فزاینده‌ای هوشمندتر و کاربرمحورتر خواهد شد و بهره‌بردن کارآمد از آن، امتیاز رقابتی قابل توجهی برای سازمان‌ها ایجاد خواهد کرد (۱۱).

۲.۲. سیستم مدیریت منابع انسانی در شرکت‌های تولیدی کوچک

شرکت‌های کوچک تولیدی به عنوان بخش کلیدی اقتصاد، نقشی برجسته در ایجاد اشتغال و تولید محصولات باکیفیت دارند، اما با چالش‌هایی چون محدودیت‌های مالی، کمبود نیروی متخصص و رقابت شدید روبه‌رو هستند. اجرای سیستم‌های HRMS متناسب با مقتضیات این شرکت‌ها می‌تواند بهبود کارایی، کاهش هزینه‌ها و افزایش توان رقابت را به همراه داشته باشد (۱۰). در این شرکت‌ها، HRMS علاوه بر مدیریت فرآیندهای کلاسیک نظیر حقوق و دستمزد و جذب نیرو، باید فرآیندهای خاصی مثل برنامه‌ریزی شیفت‌ها، آموزش مهارت‌های فنی و ایمنی و مدیریت عملکرد افراد در خط تولید را نیز پوشش دهد. فناوری‌های نوین مانند سیستم‌های ابری، اینترنت اشیا و هوش مصنوعی می‌توانند تحول چشمگیری در این زمینه ایجاد کنند و شرکت‌های کوچک را به سمت بهینه‌سازی فرآیندها و رشد بلندمدت سوق دهند (۱۱).

۲.۳. اهمیت و ضرورت کاربرد داده‌کاوی در سیستم منابع انسانی شرکت‌های تولیدی کوچک

امروزه تحلیل داده‌ها و ابزارهای داده‌کاوی در مدیریت منابع انسانی اهمیت فزاینده‌ای یافته‌اند و مزایای ارزشمندی برای سازمان‌ها به همراه دارند. به کارگیری این فناوری‌ها، امکان جمع‌آوری، تحلیل و پیش‌بینی نیازهای منابع انسانی را فراهم می‌آورد و اثربخشی مدل‌های سنتی را بهبود می‌بخشد (۱۲). سیستم‌های اطلاعات منابع انسانی (HRIS) نیز به طور گسترده در سازمان‌ها برای ثبت داده‌های کارکنان و تهیه گزارش‌های مدیریتی کاربرد دارد (۱۳). ابزارهای تحلیلی نه تنها برای سنجش عملکرد، تصمیم‌گیری درباره ارتقاء و حقوق، بلکه برای افزایش نگهداشت کارکنان و توسعه مهارت‌ها بسیار مؤثر است (۱۴). با وجود شناخته شدن اهمیت تحلیل داده، هنوز پیاده‌سازی آن در عرصه عمل نوظاست و آینده این رویکرد در مدیریت منابع انسانی بسیار امیدوارکننده تلقی می‌شود (۱۵).

نظر به ساختار ویژه شرکت‌های کوچک و متوسط، یکپارچه‌سازی داده‌کاوی با سیستم‌های HRMS برای این کسب‌وکارها اهمیت مضاعفی دارد؛ چرا که مدل‌های HRM مبتنی بر شرکت‌های بزرگ برای آنها قابل استفاده نیست (۱۷). چالش‌هایی مانند جذب و نگهداشت استعداد، ارزیابی عملکرد و فرهنگ‌سازی مثبت، در نبود تحلیل داده تشدید می‌شود (۹). بهره‌گیری از داده‌کاوی در SME ها مزیت رقابتی و توان لازم برای تصمیم‌سازی‌های

راهبردی را ارتقا می‌بخشد و با شناسایی الگوهای مؤثر، تخصیص بهینه منابع و بهبود جایگاه رقابتی حاصل خواهد شد (۱۸). برای SME ها، ضرورت طراحی HRMS بر اساس داده کاوی به دلیل منابع محدود و چالش‌های منحصر به فرد آنها بسیار مهم است (۱۹). داده کاوی می‌تواند به این شرکت‌ها کمک کند تا با بهینه‌سازی شیوه‌های منابع انسانی مانند استخدام، انتخاب، حفظ و توسعه منابع انسانی بر کمبود منابع غلبه کنند و به یکپارچه‌سازی HRM با سیستم‌های اطلاعاتی یاری رساند (۱۹).

در نتیجه، طراحی HRMS بر اساس داده کاوی برای SME ها جهت افزایش قابلیت‌های پویا، ارتقای نوآوری و بهبود عملکرد کلی ضروری است. با استفاده از تکنیک‌های داده کاوی، SME ها می‌توانند رویکردهای استراتژیک‌تری را برای مدیریت منابع انسانی توسعه دهند که منجر به تصمیم‌گیری بهتر، تخصیص منابع بهینه و موقعیت رقابتی قوی‌تر در بازار گردد.

۲.۴. ارزیابی عملکرد کارکنان در شرکت‌های تولیدی کوچک

ارزیابی عملکرد کارکنان یکی از محورهای اصلی مدیریت منابع انسانی است که به سازمان‌ها امکان اندازه‌گیری عملکرد فردی، شناسایی نقاط قوت و ضعف و برنامه‌ریزی برای بهبود مستمر را می‌دهد (۲۰). در شرکت‌های تولیدی کوچک، این فرآیند به دلیل ارتباط مستقیم عملکرد کارکنان با کیفیت محصول و سودآوری، نقش اساسی دارد. اما با محدودیت منابع و تخصص مواجه است و نیازمند طراحی سیستمی متناسب می‌باشد.

۲.۴.۱. اهداف ارزیابی عملکرد

- **بهبود بهره‌وری:** شناسایی و تفکیک کارکنان فعال و غیرفعال و ارائه بازخورد مؤثر؛
- **ارتقای کیفیت محصولات:** تضمین رعایت استانداردهای کیفی در فرایند تولید؛
- **توسعه مهارت‌ها:** شناسایی نیازهای آموزشی و برنامه‌ریزی توسعه فردی؛
- **افزایش انگیزه:** ارج نهادن و پاداش‌دهی به عملکردهای ممتاز؛
- **پشتیبانی از تصمیم‌های مدیریتی:** ارائه داده‌های دقیق برای تصمیم در مورد ارتقاء، جابه‌جایی یا خاتمه همکاری با کارکنان.

۲.۴.۲. روش‌های ارزیابی عملکرد

- **ارزیابی مبتنی بر اهداف (MBO):** تعیین اهداف مشخص و ارزیابی تحقق آنها نزد هر کارمند؛
- **ارزیابی ۳۶۰ درجه:** دریافت بازخورد از همکاران، مدیران و - در صورت لزوم - مشتریان؛ به‌ویژه مناسب برای تیم‌های کوچک؛
- **ارزیابی مبتنی بر شاخص‌های کلیدی عملکرد (KPI):** سنجش بر مبنای شاخص‌های کمی مانند تعداد محصول تولیدی یا نرخ ضایعات؛
- **ارزیابی رفتاری:** بررسی و ارزیابی رفتارهایی نظیر رعایت ایمنی، همکاری تیمی و انضباط شغلی؛
- **ارزیابی مبتنی بر مهارت‌ها:** سنجش سطح دانش و مهارت فنی کارکنان و شناسایی نیازهای آموزشی مرتبط (۲۱).

۳. پیشینه پژوهش

خدایاری ابدی (۱۴۰۳)، در پژوهش خود به بهینه‌سازی مدیریت منابع انسانی از طریق یکپارچه‌سازی فناوری سیستم اطلاعات مدیریت (MIS) و اصول مدیریت منابع انسانی پرداخته است. نتایج این تحقیق نشان داد که رویکرد پیشنهادی مبتنی بر فناوری MIS می‌تواند سازمان‌ها را قادر سازد داده‌های منابع انسانی را به شکل یکپارچه ادغام کرده، روندهای مهم را تحلیل نموده و نیازهای آینده نیروی کار را با دقت بیشتری پیش‌بینی کند. همچنین، این رویکرد به شناسایی بهینه‌ترین استراتژی‌های توسعه نیروی کار در راستای مأموریت و چشم‌انداز سازمان کمک می‌کند و اهمیت آن فراتر از افزایش ساده عملکرد نیروی انسانی رفته و بینش‌های جدیدی در مدیریت کلی منابع انسانی ارائه می‌دهد (۲۲).

محبی و مومنی (۱۴۰۳) با تمرکز بر جایگاه مدیریت منابع انسانی در سیستم‌های تولید چابک دریافتند که پیاده‌سازی سیستم‌های مدیریت منابع انسانی از طریق جذب و انتخاب نیروی کار چابک، آموزش و توسعه چندمهارتی کارکنان، برقراری نظام پاداش و جبران خدمت اثربخش و ایجاد فرهنگ سازمانی حمایتی، زیرساخت‌های لازم برای تحقق تولید چابک را در سازمان‌ها فراهم می‌سازد (۲۳).

قاسمی و همکاران (۱۴۰۲) مزایا، معایب و کاربرد سیستم‌های خبره هوش مصنوعی در حوزه مدیریت منابع انسانی را بررسی کردند. یافته‌های این پژوهش نشان داد که استفاده از هوش مصنوعی در مدیریت منابع انسانی مزایا و چالش‌های متنوعی در پنج حوزه مختلف و کاربردهایی در شش حوزه تخصصی داشته و قابلیت‌های نوآورانه‌ای را در تصمیم‌سازی منابع انسانی وارد می‌کند (۲۴).

علی و کلاچ (۲۰۲۴)، قابلیت‌های هوش مصنوعی در استخدام منابع انسانی را با تحلیل ۶۳۵ مقاله و بررسی دقیق ۳۵ مورد منتخب مورد مطالعه قرار دادند. این پژوهش مبتنی بر روش مرور نظام‌مند کتابخانه‌ای بوده و منجر به شناسایی و دسته‌بندی نقاط قوت، ضعف و قابلیت‌های کلیدی هوش مصنوعی در فرایندهای مدیریت منابع انسانی شد (۲۵).

ماروما و مویلوا (۲۰۱۷) بررسی جامعی پیرامون کاربرد تکنیک‌های داده‌کاوی در مدیریت منابع انسانی انجام دادند. آن‌ها جهت ارزیابی عملکرد سیستم منابع انسانی از شاخص‌هایی چون سطح رضایتمندی، تعداد پروژه‌ها، میانگین ساعت ماهانه و فعالیت‌های تصادفی استفاده نمودند و الگوریتم‌های طبقه‌بندی داده‌کاوی مانند درخت تصمیم و رگرسیون لجستیک را به کار بردند. نتایج پژوهش نشان داد که مدل درخت تصمیم با کارایی ۹۷ درصد برتری معنیداری نسبت به رگرسیون لجستیک دارد و برای گردآوری، ارزیابی و تحلیل داده‌های مدیریت منابع انسانی از اثربخشی مناسبی برخوردار است (۲۶).

بررسی پیشینه پژوهشی نشان می‌دهد که در حوزه سیستم‌های مدیریت منابع انسانی، شکاف‌های علمی مهمی وجود دارد که توسعه مدل‌های نوآورانه برای پر کردن این خلأها را ضروری می‌سازد. یکی از این چالش‌ها، ارزیابی و تحلیل عملکرد در سیستم‌های مدیریت منابع انسانی - به‌ویژه در مطالعات داخلی - است. در اغلب پژوهش‌های ایرانی، سنجش و تحلیل کارایی این سیستم‌ها عمدتاً با رویکردهای آماری، کیفی و تصمیم‌گیری چندمعیاره انجام شده و این رویکردها غالباً باعث کندی فرایند و ناسازگاری داده‌های جدید با سیستم‌های موجود می‌شوند. در حالی که پژوهش حاضر با محوریت بهره‌گیری از روش‌های نوین دسته‌بندی و یادگیری ماشین و پیاده‌سازی الگوریتم‌ها در زبان‌های برنامه‌نویسی پیشرفته مانند پایتون، به دنبال افزایش سرعت، دقت و بهینه‌تر ساختن پیاده‌سازی داده‌کاوی در حوزه مدیریت منابع انسانی است. این رویکرد راه را برای تجزیه و تحلیل دقیق‌تر، سریع‌تر و اثربخش‌تر عملکرد نیروی انسانی سازمان‌های کوچک و متوسط فراهم می‌کند.

۴. داده‌کاوی

داده‌کاوی به فرآیند بررسی و تحلیل حجم عظیمی از داده‌ها جهت کشف الگوها و قوانین پنهان و معنادار در درون داده‌ها اطلاق می‌شود و عموماً با ساخت مدل‌های مرتبط همراه است. یک مدل در این حوزه، در واقع الگوریتم یا مجموعه‌ای از قوانین است که مجموعه‌ای از ورودی‌ها را به هدف خاصی مرتبط می‌سازد (۲۷).

در این پژوهش، داده‌های کارکنان استخراج شده از بانک اطلاعاتی نیروی انسانی شرکت‌های کوچک تولیدی کاشی و سرامیک شهر یزد مبنای کار قرار گرفت. الگوریتم‌های مختلف داده‌کاوی، به شکل عملی در محیط زبان برنامه‌نویسی پایتون پیاده‌سازی و ارزیابی شد و بهترین الگوریتم برای پیش‌بینی ارزیابی عملکرد کارکنان انتخاب گردید. در این تحقیق، پنج الگوریتم داده‌کاوی زیر بررسی و تحلیل شده است:

۱. الگوریتم درخت تصمیم؛

۲. الگوریتم ک-نزدیک‌ترین همسایگان (K-NN)؛

۳. الگوریتم نایو بیزین؛

۴. الگوریتم ماشین بردار پشتیبان (SVM)؛

۵. الگوریتم رگرسیون لجستیک.

۴.۱. الگوریتم درخت تصمیم

درخت تصمیم در یادگیری ماشین، یک مدل پیش‌بینی‌کننده است که از اطلاعات مشاهده‌شده برای استنتاج مقدار هدف بهره می‌برد. این ساختار، از رایج‌ترین روش‌های داده‌کاوی بوده و در هر گره داخلی، یک متغیر تصمیم قرار می‌گیرد؛ شاخه‌ها، مقادیر ممکن این متغیر را نشان می‌دهند و گره برگ نمایانگر نتیجه یا برچسب هدف است. با تقسیم مکرر داده‌ها بر اساس ویژگی‌ها، درختان تصمیم قادرند توصیفات قابل درک برای انسان را ارائه دهند و دسته‌بندی و پیش‌بینی داده‌ها را امکان‌پذیر سازند. این الگوریتم ضمن پشتیبانی از صفات گسسته و پیوسته، در حل مسائل چندبعدی و برخورد با حجم بزرگ داده‌ها عملکرد قابل توجهی دارد. مزایایی همچون تولید توصیف ساده و قابل تفسیر، امکان اضافه شدن داده جدید و انعطاف‌پذیری در انتخاب ویژگی‌ها، از نقاط قوت این روش محسوب می‌شود (۲۸).

۴.۲. الگوریتم ک-نزدیک‌ترین همسایگان (K-NN)

الگوریتم K-NN نوعی روش آماری ناپارامتری برای طبقه‌بندی و رگرسیون است. در این الگوریتم، برای هر نمونه جدید، k نمونه آموزشی با کوتاه‌ترین فاصله (معمولاً با معیار فاصله اقلیدسی) لحاظ شده و برچسب نهایی بر اساس اکثریت برچسب‌های این همسایگان تخصیص می‌یابد. در حالت رگرسیون نیز میانگین مقدار خروجی‌های همسایگان نزدیک مدنظر قرار می‌گیرد. نرمال‌سازی داده‌ها، به افزایش دقت این الگوریتم کمک می‌کند و پیاده‌سازی صحیح مراحل آن می‌تواند بازدهی مناسبی حاصل کند (۲۹).

۴.۳. الگوریتم نایو بیزین

در رویکرد یادگیری ماشینی، الگوریتم نایو بیزین مبتنی بر احتمال و نظریه بیز تعریف می‌شود. هدف اصلی، یافتن محتمل‌ترین فرضیه (طبقه) برای داده‌های آموزشی است که با استفاده از احتمال قبلی و داده‌های مشاهده‌شده، احتمال پسین محاسبه می‌شود. در این الگوریتم، فرض استقلال ویژگی‌ها ساده‌سازی بااهمیتی در مدل‌سازی به وجود می‌آورد که خصوصاً برای داده‌های متنی و دسته‌بندی سریع بسیار مفید است (۲۸):

$$P(h|D) = \frac{P(h) \times P(D|h)}{P(D)}$$

در این فرمول، $P(h)$ احتمال پیشین فرضیه، $P(D)$ احتمال مشاهده داده و $P(D|h)$ درست‌نمایی داده با فرض صحت فرضیه است.

۴.۴. الگوریتم ماشین بردار پشتیبان (SVM)

این الگوریتم از روش‌های یادگیری با نظارت در دسته‌بندی و رگرسیون است که با ایجاد یک ابرصفحه با بیشترین فاصله حاشیه‌ای، داده‌ها را از هم تفکیک می‌کند. معماری الگوریتم بر پایه یافتن خط یا ابرصفحه‌ای است که بتواند داده‌ها را با بیشترین اطمینان جداسازی نماید. برای داده‌های غیرخطی از توابع هسته (کرنل) جهت نگاشت داده‌ها به فضای ابعاد بالاتر بهره گرفته می‌شود. تاریخچه الگوریتم SVM به سال ۱۹۶۴ در مؤسسه علوم کنترل آکادمی علوم روسیه بازمی‌گردد و با معرفی هسته‌های مختلف توسعه قابل ملاحظه‌ای یافته است (۳۰). فرمول ابرصفحه در SVM به صورت زیر تعریف می‌شود:

$$W \cdot X - b = 0$$

که در آن W بردار نرمال و b مقدار بایاس بوده و داده‌ها بر اساس رابطه زیر برچسب می‌گیرند:

$$\begin{cases} W \cdot X - b \geq 1 : y = 1 \\ W \cdot X - b \leq -1 : y = -1 \end{cases}$$

۴,۵. رگرسیون لجستیک

رگرسیون لجستیک یک روش آماری برای دسته‌بندی است که عمدتاً برای پیش‌بینی احتمال عضویت نمونه‌ها در یک کلاس خاص (دو یا چند کلاس) کاربرد داشته و علیرغم نامش، روشی برای طبقه‌بندی است. توابع کلیدی در این مدل عبارتند از:

۴,۵,۱. تابع سگموئید

خروجی را به بازه $[0, 1]$ نگاشت می‌کند و بیانگر احتمال حضور نمونه در کلاس موردنظر است:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

که در آن:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

۴,۵,۲. آستانه تصمیم‌گیری

معمولاً مقدار 0.5 به عنوان آستانه جهت تفکیک بین کلاس‌ها استفاده می‌شود:

$$\begin{cases} \sigma(z) \geq 0.5 \rightarrow \{\text{کلاس } 1\} \\ \sigma(z) < 0.5 \rightarrow \{\text{کلاس } 0\} \end{cases}$$

۴,۵,۳. تابع هزینه و آموزش مدل

برای آموزش مدل، از تابع هزینه لاگ-احتمال استفاده می‌شود و با بهینه‌سازی آن (اغلب گرادینت کاهشی)، ضرایب مدل (β) تعیین می‌گردند (۳۰):

$$J(\beta) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

۵. اهداف پژوهش

۱. شناسایی و بررسی ابعاد و شاخص‌های کلیدی در طراحی سیستم مدیریت منابع انسانی ویژه شرکت‌های کوچک و متوسط، با توجه به اقتضائات و نیازهای خاص این سازمان‌ها؛
۲. طراحی و ارزیابی یک سیستم مدیریت منابع انسانی مبتنی بر داده‌کاوی و استفاده از روش‌های نوین دسته‌بندی به منظور ارتقای دقت و اثربخشی سامانه؛
۳. سنجش اعتبار و صحت مدل پیشنهادی بر پایه الگوریتم‌های پیشرفته داده‌کاوی، با اتکا به معیارهای کیفی و کمی معتبر در حوزه مدیریت منابع انسانی؛
۴. انتخاب بهینه‌ترین روش دسته‌بندی داده‌کاوی به منظور ارائه مدل نهایی برای پیاده‌سازی سیستم مدیریت منابع انسانی کارآمد در شرکت‌های کوچک و متوسط.

۶. سؤالات پژوهش

۱. ابعاد و شاخص‌های اصلی موثر در طراحی سیستم مدیریت منابع انسانی در شرکت‌های کوچک و متوسط کدام‌اند؟
۲. چه فرآیندی می‌تواند برای تدوین و استقرار یک سیستم مدیریت منابع انسانی مبتنی بر روش‌های دسته‌بندی داده‌کاوی اتخاذ شود؟
۳. اعتبار و میزان صحت مدل ارائه شده، مبتنی بر الگوریتم‌های دسته‌بندی داده‌کاوی، چگونه قابل ارزیابی است؟

۴. در میان روش‌های مختلف دسته‌بندی داده‌کاوی، کدام روش برای ارائه مدل نهایی سیستم مدیریت منابع انسانی در شرکت‌های کوچک و متوسط کارآمدتر و برتر است؟

۷. روش‌شناسی پژوهش

این تحقیق، از نظر هدف، ماهیتی بنیادی-کاربردی دارد. به این معنا که ضمن توسعه دانش علمی در حوزه طراحی سیستم‌های مدیریت منابع انسانی، راهکارهایی عملی برای حل یک مسئله واقعی در شرکت‌های کوچک و متوسط ارائه می‌دهد. پژوهش‌های بنیادی به دنبال تبیین روابط میان متغیرها و آزمون نظریه‌ها هستند، در حالی که تحقیقات کاربردی بر تولید دانش برای رفع نیازهای عملی و ارتقای کارایی روش‌ها در موقعیت‌های واقعی تأکید دارند. رویکرد پژوهش ترکیبی (کیفی-کمی) و مبتنی بر منطق قیاسی-استقرایی است. به جای تصویب یا رد فرضیه، پژوهش حاضر با طرح سوالات و پاسخگویی نظام‌مند، به کشف و ارائه مدل مناسب می‌پردازد. پژوهش کیفی (شناسایی شاخص‌ها و ابعاد مدل) به صورت مقدماتی اجرا، و سپس در مرحله کمی (آزمایش مدل و الگوریتم‌های داده‌کاوی)، ادامه یافته است. روش‌های گردآوری داده‌ها در پژوهش حاضر، در دو بخش کتابخانه‌ای (جهت استخراج پیشینه و ادبیات موضوع، شناسایی ابعاد و شاخص‌های کلیدی مدل) و میدانی (جمع‌آوری داده‌های واقعی از شرکت‌های هدف) انجام شده است. برای گردآوری داده‌های میدانی، اطلاعات مرتبط با منابع انسانی و ارزیابی عملکرد کارکنان از بانک اطلاعاتی ۲۰۰ نفر از نیروهای انسانی شرکت‌های کوچک تولیدکننده کاشی و سرامیک مستقر در شهر یزد اخذ و مورد تحلیل قرار گرفت.

با توجه به ماهیت اکتشافی پژوهش، هدف اصلی کشف ابزار و مدل بهینه مدیریت منابع انسانی، متناسب با ویژگی‌ها و چالش‌های شرکت‌های کوچک و متوسط، با بهره‌گیری از روش‌های یادگیری ماشین و دسته‌بندی داده‌کاوی است. تجزیه و تحلیل داده‌ها و ارزیابی مدل پیشنهادی به صورت پیمایشی و با استفاده از الگوریتم‌های داده‌کاوی منتخب در بستر نرم‌افزاری Python اجرا گردید.

فرآیند انجام پژوهش مطابق مراحل زیر صورت گرفته است:

- مرور ادبیات و پیشینه پژوهش و استخراج شاخص‌های کلیدی؛
- جمع‌آوری داده‌های میدانی از بانک اطلاعات شرکت‌های هدف؛
- پیش‌پردازش و پاک‌سازی داده‌ها جهت آماده‌سازی برای تحلیل داده‌کاوی؛
- پیاده‌سازی و ارزیابی الگوریتم‌های منتخب داده‌کاوی (درخت تصمیم، K-NN، نایو بیزین، ماشین بردار پشتیبان و رگرسیون لجستیک) جهت ارزیابی عملکرد کارکنان و اعتبار مدل پیشنهادی؛
- تحلیل نتایج، انتخاب بهینه‌ترین الگوریتم و استخراج مدل نهایی.

این ساختار، امکان اعتبارسنجی و مقایسه الگوریتم‌ها از منظر معیارهای دقت، صحت، بازخوانی و F1-Score را در ارزیابی عملکرد منابع انسانی فراهم آورده و زمینه را برای توسعه عملی راه‌حل‌های داده‌محور در مدیریت منابع انسانی شرکت‌های کوچک و متوسط فراهم می‌کند.

۸. یافته‌های پژوهش

در این پژوهش، با اتکا به مبانی نظری و پیشینه مطالعات داخلی و خارجی، ابعاد و شاخص‌های کلیدی برای طراحی سیستم مدیریت منابع انسانی در شرکت‌های کوچک و متوسط شناسایی شد. خروجی این بخش در جدول ۱ ارائه شده است:

جدول ۱: شاخص‌های ارزیابی عملکرد کارکنان

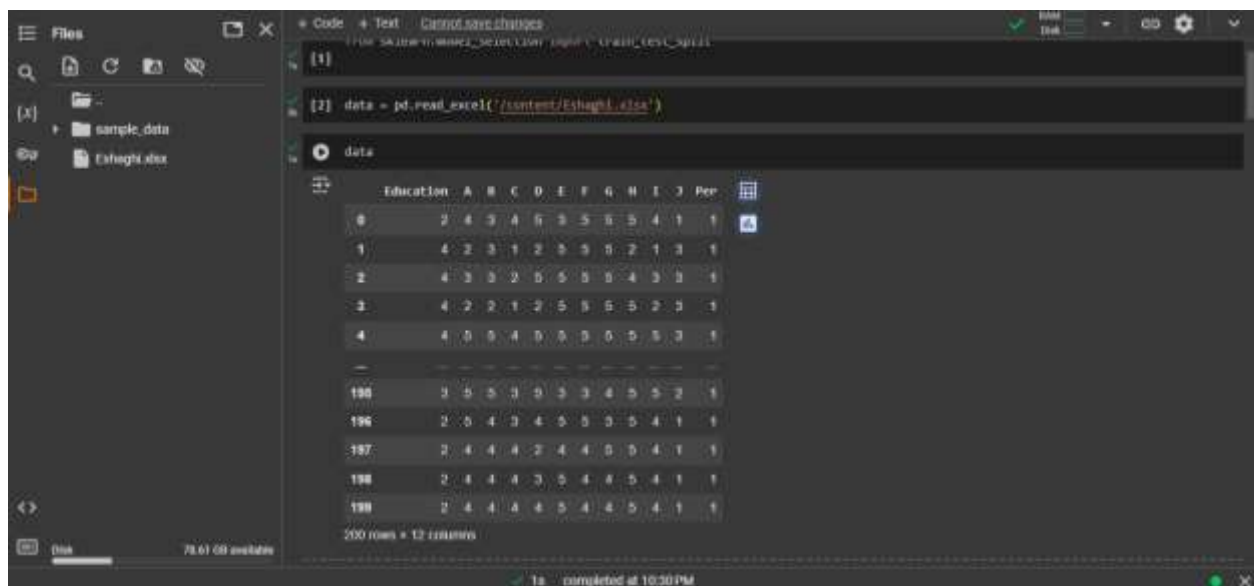
ردیف	شاخص‌های ارزیابی عملکرد کارکنان	منبع
۱	سطح تحصیلات	(۲۲)، (۲۳)، (۲۴).
۲	توسعه حرفه‌ای شامل تخصص، آموزش و یادگیری	(۲۲)، (۲۵)، (۲۷).
۳	تعهد و اخلاق حرفه‌ای	(۲۶)، (۲۸)، (۲۹)

۴	نوآوری و خلاقیت	(۲۸)،(۲۴).
۵	انضباط و بهداشت حرفه‌ای	(۲۲)،(۲۸)،(۳۰).
۶	اهتمام، اراده و پشتکار	(۲۷)،(۲۸)،(۳۰).
۷	حل مسائل با دقت	(۲۴)،(۲۵).
۸	تعامل، تابعیت و فرمان‌پذیری	(۲۷)،(۲۸)،(۳۰).
۹	توانایی کار تیمی	(۲۶)،(۲۷)،(۲۸).
۱۰	رضایت شغلی	(۲۹)،(۳۰).

بر مبنای همین شاخص‌ها، داده‌های مورد نیاز از بانک‌های اطلاعاتی شرکت‌های تولیدی کوچک و متوسط حوزه لعاب کاشی و سرامیک در شهر یزد گردآوری شد. این داده‌ها پس از کدگذاری، امتیازدهی و استانداردسازی، برای اجرای الگوریتم‌های داده کاوی آماده شدند.

فرآیند پیاده‌سازی و تحلیل داده‌ها در محیط زبان برنامه‌نویسی پایتون و از طریق سرویس ابری Google Colab انجام گرفت. در مرحله پیش‌پردازش، داده‌ها وارد فایل اکسل شده و سپس به محیط Colab منتقل گردید. برای ارزیابی الگوریتم‌های یادگیری ماشین، متغیر «ارزیابی عملکرد» به عنوان متغیر هدف تعیین شد. تقسیم‌بندی داده‌ها برای آموزش و آزمون به ترتیب ۷۰٪ برای آموزش مدل و ۳۰٪ برای آزمون مدل در نظر گرفته شد.

پیکربندی سخت‌افزاری انجام تحلیل‌ها شامل یک رایانه با پردازنده پنج هسته‌ای، فرکانس ۱/۴ گیگاهرتز و حافظه ۴ گیگابایت بود. داده‌ها پس از پیش‌پردازش در محیط Google Colab طبق شکل ۲، در قالب یک دیتاست ساخت‌یافته در دسترس قرار گرفتند.



شکل ۲: نمایشی از داده‌ها در محیط Google Colab

Google Colab محیطی ابری مبتنی بر Jupyter Notebook است که امکان اجرای کدهای پایتون را به شکل تعاملی و در بستر اینترنت مهیا می‌کند. این محیط خصوصاً برای پروژه‌های یادگیری ماشین، تحلیل داده و هوش مصنوعی بسیار کارآمد و محبوب است. همان‌گونه که در شکل ۲ نیز نمایش داده شده، داده‌های جمع‌آوری شده به صورت اکسل وارد محیط Colab شده و برای اجرای الگوریتم‌های داده کاوی مورد استفاده قرار گرفتند.

۱.۸. پیاده‌سازی الگوریتم درخت تصمیم

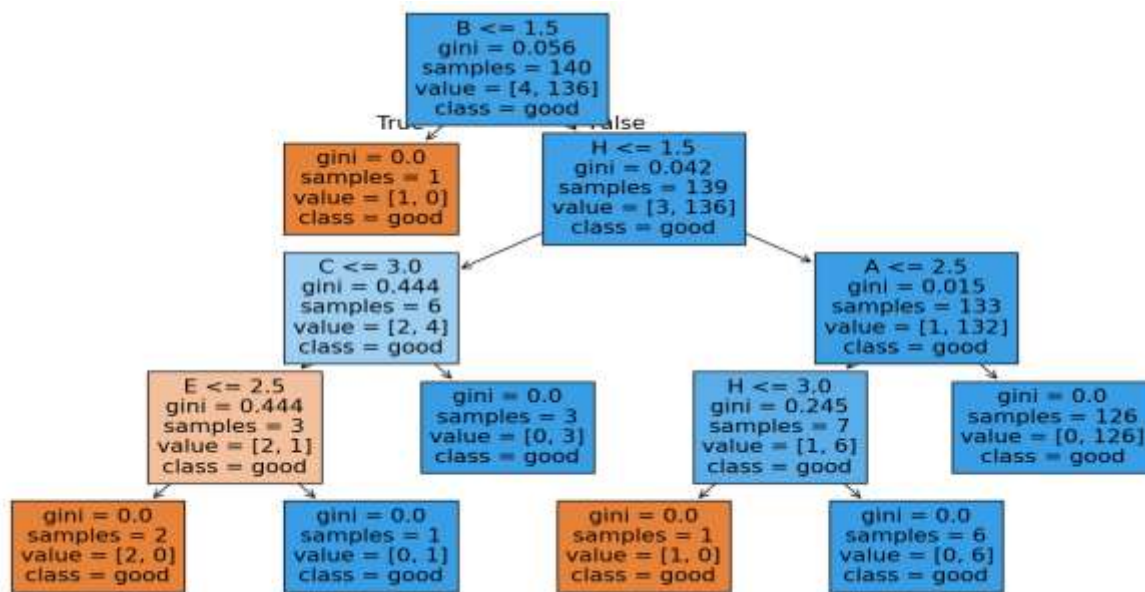
در این بخش، الگوریتم درخت تصمیم به منظور ارزیابی عملکرد کارکنان در شرکت‌های کوچک و متوسط تولیدکننده کاشی و سرامیک در شهر یزد، بر روی ۲۰۰ رکورد مربوط به کارکنان پیاده‌سازی و تحلیل شد. شاخص‌های کلیدی ارزیابی، مطابق جدول ۱، پیش‌تر تعیین و داده‌ها پس از پیش‌پردازش جهت کار با الگوریتم مذکور آماده‌سازی گردیدند.

در فرآیند اجرا، پارامترهایی همچون کارایی الگوریتم، دقت (Accuracy)، صحت عملکرد (Precision) و معیار F (F1-Score) به طور خودکار توسط کدهای پایتون محاسبه شد. خروجی الگوریتم درخت تصمیم مطابق جدول ۲ ارائه می‌شود:

جدول ۲: نتایج الگوریتم درخت تصمیم

الگوریتم داده کاوی	کارایی الگوریتم	دقت	صحت عملکرد	معیار F1
درخت تصمیم	۰/۸۹	۰/۹	۰/۹۲	۰/۹۴

همچنین ساختار درخت تصمیم تولیدشده برای ارزیابی عملکرد کارکنان، در شکل ۳ نمایش داده شده است.

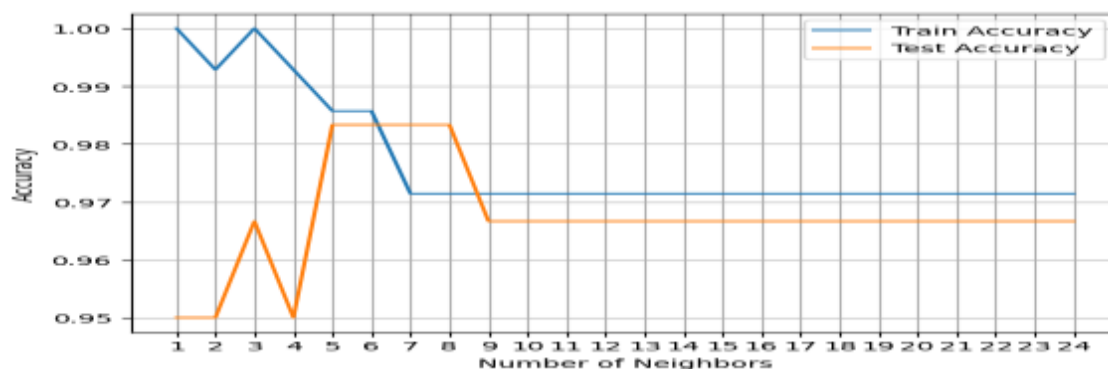


شکل ۳: نمودار حاصل از پیاده‌سازی الگوریتم درخت تصمیم جهت ارزیابی سیستم مدیریت منابع انسانی

تحلیل نتایج به دست آمده نشان می‌دهد که الگوریتم درخت تصمیم، با دقت و صحت قابل قبول، توزیع طبقه‌بندی صحیحی از عملکرد کارکنان شرکت‌های تولیدی کوچک و متوسط در این نمونه موردی ارائه کرده است. این یافته‌ها بیانگر آن است که استفاده از الگوریتم درخت تصمیم می‌تواند ابزاری مؤثر در مدیریت و ارزیابی عملکرد منابع انسانی در صنایع هدف مذکور باشد.

۲.۸. پیاده‌سازی الگوریتم نزدیک‌ترین همسایه (K-NN)

در گام بعدی پژوهش، الگوریتم نزدیک‌ترین همسایه (K-NN) به منظور ارزیابی عملکرد کارکنان شرکت‌های کوچک و متوسط مورد استفاده قرار گرفت. با توجه به ماهیت پارامتریک این الگوریتم نسبت به مقدار K، ابتدا لازم بود مقدار بهینه K تعیین شود تا بهترین عملکرد الگوریتم حاصل گردد. برای این منظور، الگوریتم K-NN در محیط برنامه‌نویسی پایتون، با مقادیر مختلف K در بازه ۱ تا ۲۵ پیاده‌سازی و ارزیابی شد. نتایج حاصل از تکرار اجرای الگوریتم با مقادیر گوناگون K در شکل ۴ ارائه شده است.



شکل ۴: تعیین مقدار بهینه K برای اجرای الگوریتم K-NN

بر اساس روند نتایج نمایش داده شده در شکل ۴، مقادیر ۵ و ۶ برای متغیر K عملکرد بهتری را نشان دادند. لذا در فرآیند پیاده سازی نهایی، مقدار K برابر با ۵ انتخاب و الگوریتم K-NN با این مقدار اجرا گردید. جدول ۳، شاخص های ارزیابی عملکرد الگوریتم K-NN را نشان می دهد.

جدول ۳: نتایج الگوریتم نزدیک ترین همسایه (K-NN)

الگوریتم داده کاوی	کارایی الگوریتم	دقت	صحت عملکرد	معیار F1
نزدیک ترین همسایه (K-NN)	۰/۹۵	۰/۹۶	۰/۹۷	۰/۹۸

۳.۸. پیاده سازی الگوریتم نایو بیز

سومین الگوریتم دسته بندی اجرا شده در این پژوهش، الگوریتم نایو بیز (Naive Bayes) است. این الگوریتم با هدف تحلیل و پیش بینی عملکرد کارکنان، بر روی داده های جمع آوری شده از شرکت های کوچک و متوسط، در محیط برنامه نویسی پایتون پیاده سازی گردید. همان گونه که در جدول ۴ مشاهده می شود، شاخص های کلیدی عملکرد الگوریتم نایو بیز، شامل دقت (Accuracy)، صحت (Precision)، بازیابی (Recall) و معیار F1 برحسب نتایج خروجی به شرح زیر است:

جدول ۴: نتایج الگوریتم نایو بیز

الگوریتم داده کاوی	کارایی الگوریتم	دقت	صحت عملکرد	معیار F1
نایو بیز	۰/۸۲	۰/۸۴	۰/۸۵	۰/۸۷

تحلیل نتایج نشان می دهد که الگوریتم نایو بیز گرچه قادر به دسته بندی عملکرد کارکنان با سطح مناسبی از صحت و دقت بوده است، اما در مقایسه با الگوریتم های دیگر مورد استفاده در این پژوهش (مانند K-NN و درخت تصمیم)، با موفقیت نسبی پایین تری عمل کرده و دقت و کارایی آن در سطح متوسط قرار دارد. با این حال، سرعت اجرای بالا و سادگی پیاده سازی این الگوریتم، می تواند آن را به گزینه ای مناسب برای تحلیل های مقدماتی و کسب رویکرد اولیه نسبت به داده ها در پروژه های مدیریت منابع انسانی تبدیل کند.

۴.۸. پیاده سازی الگوریتم ماشین بردار پشتیبان (SVM)

چهارمین الگوریتم دسته بندی مورد استفاده در این پژوهش، ماشین بردار پشتیبان (SVM) است که برای طبقه بندی داده های مربوط به عملکرد کارکنان شرکت های تولیدی کوچک و متوسط، با بهره گیری از زبان برنامه نویسی پایتون پیاده سازی گردید. SVM به دلیل توان بالای خود در جداسازی بهینه طبقات، به ویژه در داده های با ابعاد بالا، یکی از قدرتمندترین روش های یادگیری ماشین محسوب می شود. عملکرد این الگوریتم بر اساس معیارهای کلیدی همانند دقت (Accuracy)، صحت (Precision)، بازیابی (Recall) و معیار F1 مورد ارزیابی قرار گرفت که نتایج به دست آمده در جدول ۵ ارائه شده است.

جدول ۵: نتایج الگوریتم ماشین بردار پشتیبان (SVM)

الگوریتم داده کاوی	کارایی الگوریتم	دقت	صحت عملکرد	معیار F1
ماشین بردار پشتیبان (SVM)	۰/۹۶	۰/۹۸	۰/۹۷	۰/۹۸

بر اساس یافته‌های حاصل از تحلیل داده‌ها، الگوریتم SVM توانسته است عملکرد بسیار مطلوبی در طبقه‌بندی صحیح و دقیق کارکنان از منظر ارزیابی عملکرد ارائه دهد. میزان بالای معیارهای دقت و معیار F1 نشان‌دهنده آن است که این الگوریتم نه تنها توان تمایز بالا در شناسایی دسته‌های مختلف عملکرد را دارد، بلکه نرخ خطای طبقه‌بندی را نیز به حداقل رسانده است. نتایج این بخش حاکی از آن است که SVM به عنوان یکی از گزینه‌های برتر در میان مدل‌های یادگیری ماشین مورد بررسی، قابلیت پیاده‌سازی مؤثر در سیستم‌های هوشمند مدیریت منابع انسانی شرکت‌های کوچک و متوسط را دارد و با توجه به معیارهای ارزیابی، در رقابت با سایر الگوریتم‌ها از پایداری و دقت بسیار بالایی برخوردار است.

۸.۵. پیاده‌سازی الگوریتم رگرسیون لجستیک

پنجمین و آخرین الگوریتم مورد استفاده در این پژوهش، رگرسیون لجستیک است که برای طبقه‌بندی داده‌های ارزیابی عملکرد کارکنان شرکت‌های کوچک و متوسط با استفاده از زبان برنامه‌نویسی پایتون پیاده‌سازی و آزمایش شد. این الگوریتم به دلیل ماهیت خطی و توانایی مدل‌سازی احتمالاتی، یکی از پرکاربردترین ابزارها در مسائل دسته‌بندی دودویی و چندکلاسه در حوزه مدیریت منابع انسانی است. ارزیابی عملکرد الگوریتم رگرسیون لجستیک نیز بر اساس چهار معیار کلیدی شامل دقت (Accuracy)، صحت (Precision)، بازیابی (Recall) و معیار F1 انجام گرفت. نتایج به‌دست‌آمده در جدول ۶ ارائه شده است:

جدول ۶: نتایج الگوریتم رگرسیون لجستیک

الگوریتم داده کاوی	کارایی الگوریتم	دقت	صحت عملکرد	معیار F1
رگرسیون لجستیک	۰/۹۸	۰/۹۹	۰/۹۹	۰/۹۹

طبق نتایج فوق، رگرسیون لجستیک بالاترین معیارهای عملکرد را در میان الگوریتم‌های مورد بررسی کسب نموده است. دقت، صحت و معیار F1 همگی در سطح ۹۹٪ هستند که نشان‌دهنده توانایی بسیار بالای این مدل در طبقه‌بندی صحیح و جامع رکوردهای ارزیابی عملکرد کارکنان است. این یافته‌ها بیانگر آن است که رگرسیون لجستیک علاوه بر سادگی و قابلیت تفسیر، کارایی بسیار بالایی نیز در مسائل ارزیابی منابع انسانی دارد و می‌تواند به عنوان یک انتخاب مطلوب برای طراحی سیستم‌های هوشمند ارزیابی عملکرد در شرکت‌های کوچک و متوسط مورد استفاده قرار گیرد. نتایج به‌دست‌آمده مؤید کارکرد قوی این مدل در زمینه مدیریت داده‌های منابع انسانی می‌باشد.

۹. انتخاب الگوریتم برتر جهت ارزیابی سیستم عملکرد منابع انسانی

۹.۱. دقت اندازه‌گیری الگوریتم‌ها

یکی از مهم‌ترین معیارها برای سنجش کارایی الگوریتم‌های طبقه‌بندی، دقت طبقه‌بندی (Accuracy) است. دقت بیانگر درصد رکوردهای آزمایشی است که طبقه‌بندی به درستی انجام شده و به عنوان رایج‌ترین شاخص عملکرد الگوریتم‌های یادگیری ماشین، مبنای مقایسه مدل‌ها قرار می‌گیرد. به عبارت دیگر، دقت معیاری است که نشان می‌دهد مدل یادگیری ماشین، چه میزان از کل نمونه‌های داده را به درستی شناسایی و طبقه‌بندی کرده است. در این پژوهش، نتایج پیاده‌سازی و ارزیابی پنج الگوریتم اصلی دسته‌بندی (درخت تصمیم، K-NN، نایو بیز، ماشین بردار پشتیبان و رگرسیون لجستیک) حاکی از آن است که الگوریتم رگرسیون لجستیک با دقت ۹۸ درصد و الگوریتم K-NN با دقت ۹۵ درصد، بهترین عملکرد را در میان الگوریتم‌ها داشته‌اند؛ در مقابل، الگوریتم نایو بیز با دقت ۸۲ درصد، پایین‌ترین میزان دقت را ارائه داده است. این یافته‌ها در نمودار شماره ۵ به تصویر کشیده شده است. مفاهیم کلیدی برای ارزیابی دقت عبارت‌اند از:

- **TP (True Positive)**: تعداد نمونه‌هایی که به درستی به عنوان مثبت پیش‌بینی شده‌اند.
- **FN (False Negative)**: تعداد نمونه‌هایی که به اشتباه به عنوان منفی پیش‌بینی شده‌اند (در حالی که مثبت بوده‌اند).
- **FP (False Positive)**: تعداد نمونه‌هایی که به اشتباه مثبت پیش‌بینی شده‌اند (در حالی که منفی بوده‌اند).
- **TN (True Negative)**: تعداد نمونه‌هایی که به درستی به عنوان منفی پیش‌بینی شده‌اند.

فرمول محاسبه دقت طبقه‌بندی به شرح زیر است:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

در محیط زبان برنامه‌نویسی پایتون، کارایی (Accuracy) الگوریتم با دستور زیر از کتابخانه scikit-learn محاسبه گردید:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
accuracy = accuracy_score(y_true, y_pred)
```



شکل ۵: مقایسه دقت طبقه‌بندی الگوریتم‌های مورد استفاده در پژوهش

با توجه به نتایج به دست آمده و شواهد ارائه شده، می‌توان نتیجه گرفت رگرسیون لجستیک و K-NN به لحاظ دقت و معیارهای کلیدی دیگر نسبت به سایر الگوریتم‌ها برتری داشته‌اند و گزینه‌های اصلی برای پیاده‌سازی سیستم هوشمند ارزیابی عملکرد منابع انسانی در شرکت‌های کوچک و متوسط محسوب می‌شوند. این مسئله با یافته‌های پژوهش‌های مشابه نیز هم‌راستا است.

۹.۲. دقت تکرار الگوریتم

دقت تکرار الگوریتم (Recall) معیاری کلیدی برای ارزیابی عملکرد مدل‌های طبقه‌بندی محسوب می‌شود و نشان می‌دهد که در اندازه‌گیری‌های متوالی یا آزمون‌های متعدد، چه تعداد از نمونه‌های مثبت واقعی توسط مدل به درستی شناسایی و طبقه‌بندی شده‌اند. به زبان دیگر، دقت تکرار مشخص می‌کند که مدل تا چه حد در بازشناسی موارد مثبت واقعی موفق عمل کرده است.

فرمول دقت تکرار به صورت زیر تعریف می‌شود:

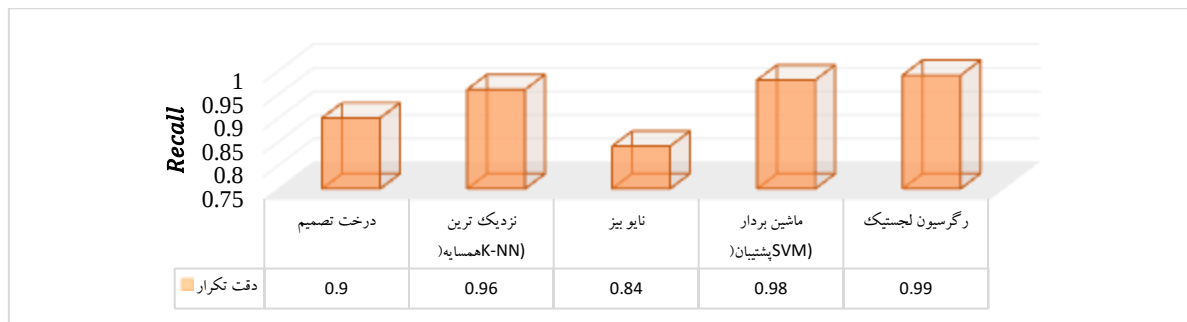
$$Recall = \frac{TP}{TP + FN}$$

که در آن:

- **TP (True Positive)**: تعداد نمونه‌هایی که به درستی به عنوان مثبت پیش‌بینی شده‌اند.
- **FN (False Negative)**: تعداد نمونه‌هایی که به اشتباه به عنوان منفی پیش‌بینی شده‌اند، در حالی که واقعاً مثبت بوده‌اند.

در محیط برنامه‌نویسی پایتون، مقدار دقت تکرار هر الگوریتم با استفاده از دستور زیر از کتابخانه scikit-learn قابل محاسبه است:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
recall = recall_score(y_true, y_pred)
```



شکل ۶: میزان دقت تکرار طبقه‌بندی الگوریتم‌های دسته‌بندی مورد استفاده

بر مبنای نتایج ترسیم‌شده در شکل ۶، الگوریتم رگرسیون لجستیک با دقت تکرار ۹۹ درصد، بالاترین عملکرد را در شناسایی نمونه‌های مثبت واقعی از خود نشان داده است. در مقابل، الگوریتم نایو بیز با دقت تکرار ۸۴ درصد، پایین‌ترین مقدار را بین الگوریتم‌های مورد بررسی دارد. این یافته‌ها بر برتری نسبی الگوریتم رگرسیون لجستیک نسبت به سایر روش‌ها در بازیابی نمونه‌های مثبت تأکید دارد و نشان‌دهنده مناسب بودن آن برای کاربردهای مدیریت منابع انسانی در محیط‌های داده‌محور است.

۹.۳. صحت الگوریتم

صحت الگوریتم (Precision) به معنای نسبت تعداد مواردی است که توسط الگوریتم به درستی به عنوان عضو یک کلاس شناسایی شده‌اند، به مجموع کل مواردی که الگوریتم به آن کلاس تخصیص داده است. به عبارت دیگر، صحت الگوریتم نشان می‌دهد چه سهمی از پیش‌بینی‌های مثبت مدل واقعاً صحیح بوده‌اند. این شاخص یکی از معیارهای مهم ارزیابی در مدل‌های طبقه‌بندی به شمار می‌رود و در تحلیل نتایج، نقش کلیدی دارد. فرمول صحت الگوریتم به صورت زیر تعریف می‌شود:

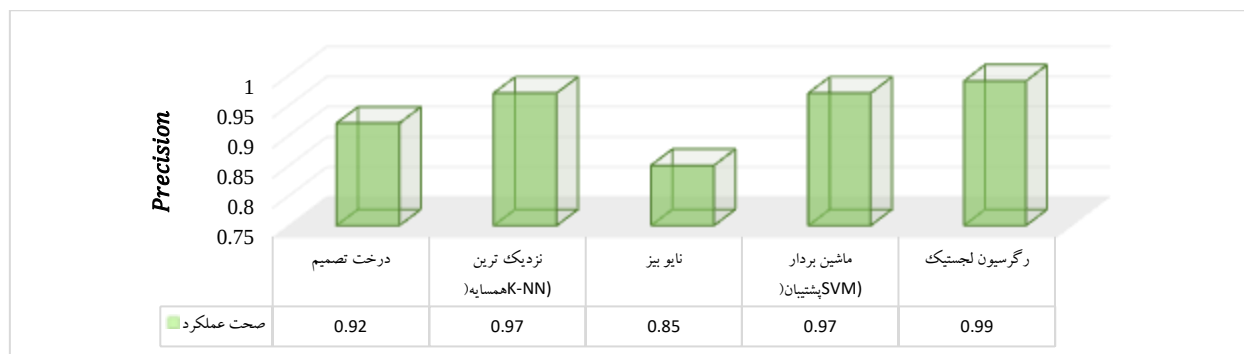
$$Precision = \frac{TP}{TP + FP}$$

که در آن:

- **TP (True Positive)**: تعداد نمونه‌هایی که به درستی به عنوان مثبت پیش‌بینی شده‌اند.
- **FP (False Positive)**: تعداد نمونه‌هایی که به اشتباه به عنوان مثبت پیش‌بینی شده‌اند، در حالی که واقعاً منفی بوده‌اند.

در محیط زبان برنامه‌نویسی پایتون، صحت الگوریتم با استفاده از دستور زیر از کتابخانه scikit-learn محاسبه شد:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
precision = precision_score(y_true, y_pred)
```



شکل ۷: میزان صحت عملکرد الگوریتم‌های دسته‌بندی مورد استفاده در پژوهش

بررسی نتایج حاصل از شکل ۷ نشان می‌دهد که الگوریتم رگرسیون لجستیک با صحت ۹۹ درصد برترین عملکرد را در بین الگوریتم‌های بررسی شده داشته است. در مقابل، الگوریتم نایو بیز با صحت ۸۵ درصد پایین‌ترین میزان صحت را در میان الگوریتم‌ها به خود اختصاص داده است. این یافته‌ها بیانگر توان بالای الگوریتم رگرسیون لجستیک در شناسایی صحیح موارد مثبت پیش‌بینی شده و دقت بالای آن در کاربردهای مدیریت داده‌محور منابع انسانی است.

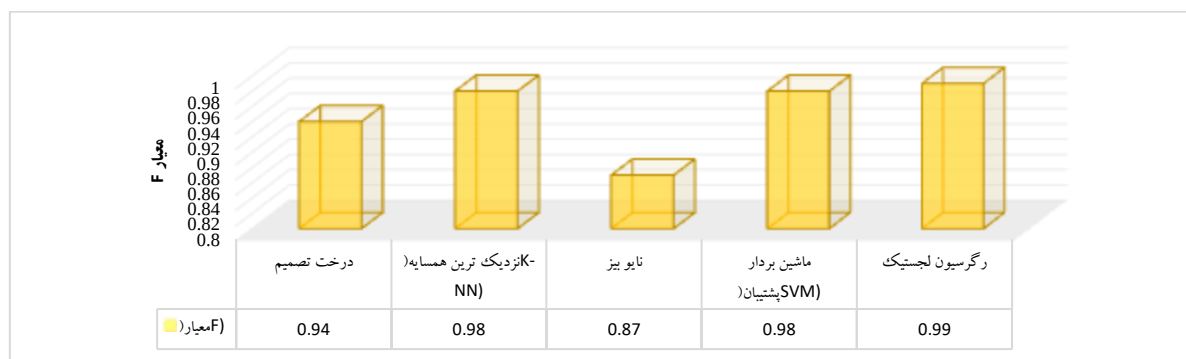
۹.۴. معیار F

معیار $F1$ (F1-Score) یکی از مهم‌ترین شاخص‌های ارزیابی کیفیت الگوریتم‌های طبقه‌بندی است که میانگین وزنی دقت تکرار (Recall) و صحت عملکرد (Precision) را به صورت همزمان در نظر می‌گیرد. این معیار به ویژه در پروژه‌هایی با داده‌های نامتوازن یا زمانی که هر دو بُعد صحت و کامل بودن پیش‌بینی‌ها اهمیت دارند، اهمیت ویژه‌ای دارد. مقدار F1-Score برای یک طبقه‌بندی کننده ایده‌آل برابر با ۱ و در ضعیف‌ترین حالت به صفر میل می‌کند. فرمول محاسبه معیار F به صورت زیر است:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

در محیط زبان برنامه‌نویسی پایتون، محاسبه این معیار با استفاده از تابع زیر از کتابخانه scikit-learn انجام شد:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
f1 = f1_score(y_true, y_pred)
```



شکل ۸: مقایسه مقدار معیار F الگوریتم‌های دسته‌بندی مورد استفاده در پژوهش

بر اساس نتایج به‌دست آمده و مطابق با شکل ۸، الگوریتم رگرسیون لجستیک با معیار F معادل ۹۹ درصد، بالاترین عملکرد را در میان الگوریتم‌های مورد بررسی نشان داده است. در مقابل، الگوریتم نایو بیز با معیار F برابر با ۸۷ درصد، دارای پایین‌ترین امتیاز بوده است. این مقایسه نشان می‌دهد که رگرسیون لجستیک علاوه بر دقت و صحت بالا، توانایی مناسبی در ایجاد تعادل بین شناسایی صحیح نمونه‌های مثبت و منفی دارد و از این نظر برای پیاده‌سازی سیستم‌های ارزیابی عملکرد منابع انسانی در شرکت‌های کوچک و متوسط گزینه‌ای ایده‌آل محسوب می‌شود.

۱۰. نتیجه‌گیری و پیشنهادها

این پژوهش با هدف طراحی و پیاده‌سازی سیستم هوشمند ارزیابی عملکرد منابع انسانی در شرکت‌های کوچک و متوسط (SMES) با بهره‌گیری از الگوریتم‌های داده‌کاوی انجام شد. یافته‌ها نشان داد که تکنیک‌های داده‌کاوی، به ویژه دسته‌بندی، ابزار مؤثری برای استخراج دانش ارزشمند از داده‌های منابع انسانی و تحلیل کاربردی عملکرد کارکنان به شمار می‌آیند. مرحله پیش‌پردازش، استانداردسازی و نرمال‌سازی داده‌ها نقش پررنگی در افزایش دقت و کیفیت مدل‌های یادگیری ماشین داشت و موجب افزایش اعتبار نتایج شد.

در پاسخ به سؤال اول تحقیق، مبنی بر شناسایی ابعاد و شاخص‌های کلیدی طراحی سیستم مدیریت منابع انسانی در شرکت‌های کوچک و متوسط، مطالعه چارچوب نظری و پیشینه داخلی و خارجی، ده شاخص اصلی شامل سطح تحصیلات، توسعه حرفه‌ای، تعهد و اخلاق حرفه‌ای، نوآوری و خلاقیت، انضباط و بهداشت حرفه‌ای، اراده و پشتکار، دقت در حل مسئله، تعامل و تابعیت، کار تیمی و رضایت شغلی را شناسایی و تأیید کرد.

در رابطه با سؤال دوم پژوهش، نحوه تدوین سامانه مدیریت منابع انسانی مبتنی بر دسته‌بندی داده‌کاوی با پیاده‌سازی الگوریتم‌های تصمیم‌گیری، نایو بیز، نزدیک‌ترین همسایه (K-NN)، ماشین بردار پشتیبان (SVM) و رگرسیون لجستیک در پایتون بر روی داده‌های جمع‌آوری شده از ۲۰۰ نفر از کارکنان شرکت‌های تولیدی نمونه ارزیابی شد. مدل حاصل، فرآیند ارزیابی و انتخاب بهینه نیروی انسانی را در بستر واقعی تسهیل نمود.

در پاسخ به سؤالات سوم و چهارم، یعنی اعتبارسنجی و انتخاب بهترین روش، نتایج تجربی بیانگر آن است که رگرسیون لجستیک بالاترین دقت (۹۸ درصد)، دقت تکرار (۹۹ درصد)، صحت (۹۹ درصد) و معیار F1 (۹۹ درصد) را کسب کرده و بهترین الگوریتم برای تحلیل شایستگی و ارزیابی عملکرد کارکنان بوده است. در مقابل، الگوریتم نایو بیز کمترین میزان دقت و صحت را در فرآیند ارزیابی عملکرد ارائه کرده است. این یافته با مطالعات علی و کلاچ (۲۰۲۴) و ماروما و موپلوا (۲۰۱۷) و سایر پیشینه‌های علمی هم‌راستا است و کاربرد عملی قابل توجهی در محیط‌های واقعی سازمانی SMEs دارد.

از جمله مهم‌ترین محدودیت‌های پژوهش می‌توان به نبود بانک اطلاعات جامع منابع انسانی در شرکت‌های کوچک اشاره کرد که البته با تلاش پژوهشگران، داده‌های ۲۰۰ نفر گردآوری و مورد استفاده قرار گرفت. بدیهی‌ست، دسترسی به داده‌های گسترده‌تر می‌تواند موجب افزایش اعتبار نتایج و بهبود کارایی الگوریتم‌های داده‌کاوی در آینده گردد.

۱.۱۰. پیشنهادها برای تحقیقات آتی:

- بهره‌گیری از داده‌های عملیاتی خط تولید (نظیر سرعت کار، دقت، میزان خطا و زمان توقف) برای مدل‌سازی دقیق‌تر ارزیابی عملکرد کارکنان و استخراج الگوها از طریق داده‌کاوی..
 - استفاده از روش‌های خوشه‌بندی برای شناسایی ترکیب‌های بهینه تیمی بر اساس مهارت‌ها و عملکرد، و تعیین معیارهای بهتر برای استخدام نیروی جدید با تحلیل رفتار کارکنان موفق.
 - ارزیابی تأثیر سیاست‌های منابع انسانی همچون برنامه‌های پاداش، تغییر شیفت یا آموزش‌های هدفمند بر پایش عملکرد کارکنان با استفاده از تکنیک‌های داده‌کاوی.
 - توصیه به شرکت‌های کوچک: بهره‌گیری از ابزارهای متن‌باز داده‌کاوی مانند Weka، RapidMiner، یا کتابخانه‌های Scikit-learn در Python به دلیل مقرون به صرفه بودن و انعطاف‌پذیری بالا.
 - توسعه پژوهش‌ها در حوزه‌هایی چون استعدادسنجی، ارتقاء شغلی، فرآیند جذب و گزینش و نیز آموزش نیروی انسانی با استفاده از الگوریتم‌های داده‌کاوی و آموزش کاربردی کارکنان در حوزه علم داده و فناوری‌های مرتبط.
 - تحلیل بانک اطلاعات سوابق استخدام و شناسایی ویژگی‌های کلیدی موفقیت با ابزارهای یادگیری ماشین به منظور بهبود دقت و کیفیت فرآیند استخدام و نگهداشت نیرو در شرکت‌های کوچک و متوسط.
- به طور کلی، پیاده‌سازی سیستم‌های مبتنی بر یادگیری ماشین و داده‌کاوی در مدیریت منابع انسانی می‌تواند سهم مؤثری در کارآمدی و افزایش کیفیت کل زنجیره تصمیم‌گیری در شرکت‌های کوچک و متوسط ایفا کند و پایه علمی و عملی مناسبی برای بهبود رقابت‌پذیری فراهم آورد.

تقدیر و تشکر

بدین وسیله مراتب سپاس و قدردانی صمیمانه خود را از مدیریت محترم شرکت کاشی امین میبد بابت حمایت مالی بی‌دریغ و همراهی ارزشمند در فرایند اجرای این پژوهش ابراز می‌داریم. حمایت‌های ایشان، به‌ویژه در تسهیل دسترسی به داده‌های تخصصی منابع انسانی و فراهم‌سازی بستر مناسب برای انجام مطالعات میدانی، نقشی تعیین‌کننده در ارتقای کیفیت علمی تحقیق ایفا نمود.

همچنین وظیفه خود می‌دانیم از همکاری مدیران و کارکنان سایر شرکت‌های کوچک و متوسط که در روند گردآوری داده‌ها و انجام بخش‌های عملی این پژوهش مشارکت فعال داشتند، صمیمانه قدردانی نماییم. مشارکت سازنده و حسن اعتماد این مجموعه‌ها انگیزه‌ای مضاعف برای نیل به اهداف علمی و کاربردی پژوهش فراهم آورد.

امید است نتایج و دستاوردهای این تحقیق بتواند بازتاب‌دهنده گوشه‌ای از تلاش جمعی فعالان حوزه صنعت کاشی و سرامیک بوده و نقشی مؤثر در بهبود و ارتقای سیاست‌های مدیریت منابع انسانی این بخش مهم از اقتصاد کشور ایفا کند.

منابع و مآخذ:

- (۱) قاسمی، آذر و قاسمی، مهدی و قیصری منوچهری، علیرضا، ۱۴۰۲، مزایا، معایب و کاربرد سیستم های خبره هوش مصنوعی در حوزه مدیریت منابع انسانی، سومین کنفرانس بین المللی تفکر سیستمی در عمل، مشهد، <https://civilica.com/doc/1873821>.
- (۲) حاجی حیدری، نسترن، خیرری، سیدحسین، تلافی داریانی، مجتبی، (۱۳۹۶)، چارچوب به کارگیری رویکرد داده کاوی در حوزه مدیریت منابع انسانی، فصلنامه علوم مدیریت ایران، سال ۱۲، شماره ۴۷، صص، ۸۱-۵۷.
- 3) Zhang Y, Xu S, Zhang L, Yang M. Big data and human resource management research: An integrative review and new directions for future research. *Journal of Business Research*. 2021;133:34-50.
- 4) Xing X, Wen Q. A Human Resource Evaluation and Recommendation System based on Big Data Mining. *Scalable Computing: Practice and Experience*. 2024;25.
- 5) Chen H, Cui X. Design and implementation of human resource management system based on B/S mode. *Procedia Computer Science*. 2022;208:442-9.
- 6) Šušnjar G, Slavic A, Berber N, Leković B. The Role of Human Resource Management in Small and Medium Sized Companies in Central-Eastern Europe. 2016. p. 205-29.
- (۷) اسلامی نژاد، رسول و چراغین، مسلم و کریمی، محمد و داودی، علیرضا، ۱۴۰۱، ارائه الگوی عوامل موثر بر پذیرش و اثربخشی سیستم مدیریت منابع انسانی الکترونیک در دانشگاه آزاد اسلامی مشهد، <https://civilica.com/doc/2020874>.
- (۸) حمید، عکافان، محمد، تاج الدین، مهدی، (۱۳۹۷)، بررسی الگوریتم های داده-کاوی در پیش بینی سمت شغلی کارکنان و پیشنهاد الگوریتم مناسب، فصلنامه پژوهش های حفاظتی-امنیتی دانشگاه امام حسین (ع)، سال هفتم، شماره ۲۷، صص، ۱۶۲-۱۲۹.
- 9) Wu Y, Wang Z, Wang S. Human Resource Allocation Based on Fuzzy Data Mining Algorithm. *Complexity*. 2021;2021:1-11.
- 10) Mishra P, Mishra P. Challenges and Opportunities of Big Data Analytics for Human Resource Management in Mining and Metal Industries. *Journal of Mines, Metals and Fuels*. 2023;1747-53.
- 11) Okinyi OM, Kelvin OK, Margaret KK. The Value Of Human Resource Information Systems In Human Resource Management. *INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY*. 2021;11(10).
- 12) Bandi G, Rao TS, Saadiq Ali S. Data Analytics Applications for Human Resource Management 2021. 1-5 p.
- 13) Zhang A. Influence of data mining technology in information analysis of human resource management on macroscopic economic management. *Plos one*. 2021;16(5):e0251483.
- 14) Liu J. Design and Application of Human resource management system Based on Data Mining Technology. *Procedia Computer Science*. 2023;228:241-52.
- 15) Fu H, Liu J, editors. Human Resource Management System for Small and Medium-Sized Enterprises based on SSM Frameworks 2024: Atlantis Press.
- 16) Shen H. Research on Human Resource Management Optimization and Decision Support System Construction Based on Electric Power Enterprise Strategy. *Applied Mathematics and Nonlinear Sciences*. 2024;9.
- 17) Dwivedi VK. Human Resource Management in the Power Industry Using Fuzzy Data Mining Algorithm.
- 18) Blazquez D, Domenech J. Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*. 2018;130:99-113.
- 19) Ma X, Zhang Y, Song Y, Wang E, Yao F, Zhang Z. Application of Data Mining in the Field of Human Resource Management: A Review 2019.
- 20) Novo Melo P, Machado CF, Brewster C. Human resource management in small and medium-sized enterprises: a performance model definition. *Strategic Management*. 2023;28(2):4-20.
- 21) Mishra S. 720-Degree Performance Appraisal -The Most Recently Introduced Concept & An Integrated Method in Performance Management System. *Journal of Scientific Research and Development*. 2022.
- (۲۲) خدایاری ابلی، حمیدرضا، ۱۴۰۳، بهینه سازی مدیریت منابع انسانی از طریق یکپارچه سازی فناوری سیستم اطلاعات مدیریت و اصول مدیریت منابع انسانی، هشتمین کنفرانس بین المللی مدیریت، حسابداری، بانکداری و اقتصاد ایران، مشهد، <https://civilica.com/doc/2056299>.

۲۳) محبی، حسین و مومنی، زینب، ۱۴۰۳، جایگاه مدیریت منابع انسانی در سیستم‌های تولید چابک، سومین کنفرانس بین‌المللی آزمایشگاه مدیریت و رویکردهای نوآورانه در مدیریت و اقتصاد، تهران، <https://civilica.com/doc/2026102>.

۲۴) قاسمی، آذر و قاسمی، مهدی و قیصری منوچهری، علیرضا، ۱۴۰۲، مزایا، معایب و کاربرد سیستم‌های خبره هوش مصنوعی در حوزه مدیریت منابع انسانی، سومین کنفرانس بین‌المللی تفکر سیستمی در عمل، مشهد، <https://civilica.com/doc/1873821>.

25) Ali,O., Kallach,L.,(2024), Artificial Intelligence Enabled Human Resources Recruitment Functionalities: A Scoping Review, *Procedia Computer Science*,Vol (232), 3268–3277.

26) Marumo,O.R, Mmopelwa,T.A., (2017), USE OF DATA MINING TECHNIQUES IN HUMAN RESOURCE MANAGEMENT, *ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL*, Vol(1),PP,41-46.

۲۷) پارسا، حمید، عکافان، محمد، تاج‌الدین، مهدی،(۱۳۹۷)، بررسی الگوریتم‌های داده‌کاوی در پیش‌بینی سمت شغلی کارکنان و پیشنهاد الگوریتم مناسب، فصلنامه پژوهش‌های حفاظتی-امنیتی دانشگاه جامع امام حسین(ع)، سال هفتم، شماره ۲۷، صص، ۱۶۲-۱۳۹.

28) 49-Mishra P, Mishra P. Challenges and Opportunities of Big Data Analytics for Human Resource Management in Mining and Metal Industries. *Journal of Mines, Metals and Fuels*. 2023;1747-53.

29) 50-Wang H, Yang Y, Zhang Y, editors. A macro human resource management platform enabled by big data technology2020: Springer.

30) Shaout A, Yousif MK. Performance evaluation–Methods and techniques survey. *International Journal of Computer and Information Technology*. 2014;3(5):966-79.

Optimizing Human Resource Performance Evaluation in Small and Medium Enterprises (SMEs) Using Data Mining Algorithms (Case Study: Tile and Ceramic Companies in Yazd Province)

1. Amir Ahsan Eshaghieh Firouzabadi

Postdoctoral Researcher, Department of Management, Meybod University, Meybod, Iran.

(Email: dr.eshaghieh@gmail.com)

2. Mohammad Zarei Mahmoudabadi* (Corresponding Author)

Associate Professor, Department of Industrial Management, Meybod University, Meybod, Iran.

(Email: zareim@meybod.ac.ir)

Abstract:

Background & Objective: The objective of this research is to design a hybrid data mining framework for evaluating human resource performance in SMEs within Yazd's tile and ceramic industry. The primary focus is on overcoming limitations of traditional methods by implementing advanced machine learning algorithms.

Methodology: In this applied-developmental research, data from 200 employees in Yazd SMEs, evaluated based on ten key performance indicators (extracted from the literature), were used. Following preprocessing in Python (Google Colab), four data mining algorithms—Decision Tree (CART), K-Nearest Neighbors (K-NN, optimized for Euclidean distance), Naive Bayes, Support Vector Machine (SVM) with RBF kernel, and Logistic Regression—were implemented and evaluated according to accuracy, precision, recall, and F1-score metrics.

Key Findings: Logistic Regression achieved the best performance for employee classification, with an accuracy of 98.3% and F1-score of 97%. SVM ranked second with 96.1% accuracy, while Naive Bayes yielded the weakest results (87.5% accuracy). Among input variables, “Professional Development” and “Problem-Solving Ability” had the greatest impact on the final model.

Conclusion and Research Value: The logistic regression-based model is capable of reducing performance evaluation errors by 30% compared to conventional SME methods. This framework not only provides an objective structure for human resource management but, thanks to the interpretability of the model, enables identification of performance improvement areas for both current and new employees.

Keywords: Data Mining Model, Human Resource Performance Evaluation, Logistic Regression, Applied Machine Learning, Small and Medium Enterprises (SMEs).