

طراحی مدل انتخاب نیروی انسانی با رویکرد داده‌کاوی (مورد: استخدام داوطلبان آزمون‌های ورودی یک بانک تجاری در ایران)

عادل آذر^۱، پرویز احمدی^۲، محمد وحید سبط^۳

۱. دانشیار گروه مدیریت، دانشگاه تربیت مدرس، ایران

۲. استادیار گروه مدیریت، دانشگاه تربیت مدرس، ایران

۳. دانشجوی کارشناسی ارشد مدیریت صنعتی، دانشگاه تربیت مدرس، ایران

(تاریخ دریافت: ۱۳۸۸/۹/۴، تاریخ تصویب: ۱۳۸۹/۳/۲)

چکیده

موفقیت یا شکست سازمان، ارتباط مستقیمی با چگونگی جذب و نگهداری منابع انسانی آن دارد. اغلب در رابطه با برگزاری آزمون‌های ورودی و فرآیند جذب کارکنان، داده‌ها و اطلاعات فراوانی در سازمان‌ها وجود دارد که بدون استفاده قرار می‌گیرند. داده‌کاوی، به‌عنوان راه حل برای چنین مسائلی است. در این پژوهش که از حیث هدف، کاربردی و از جنبه ماهیت از نوع پژوهش‌های همبستگی و همخوانی محسوب می‌شود، سعی شده است که با استفاده از تکنیک‌های داده‌کاوی، قواعد و روابط بین نمرات آزمون‌های ورودی و سایر متغیرهای شخصی و شغلی (که قبل از ورود هر کس به سازمان مشخص می‌شود) و وضعیت کارکنان با عملکرد شغلی و وضعیت ارتقاء آنان شناسایی شود. در نتیجه با مطالعه و بررسی پایگاه‌های داده‌ی آزمون و منابع انسانی یک بانک تجاری برای ۲ سال متوالی (۱۳۸۳ و ۱۳۸۴)، شاخص‌های نیروی انسانی که بر عملکرد یا ارتقاء موثر بودند، شناسایی شدند. تکنیک داده‌کاوی مورد استفاده در این پژوهش، درخت تصمیم‌گیری است و استخراج قواعد نیز با استفاده از الگوریتم‌های CART، C5.0، CHAID، QUEST و مناسب؛ از بین قواعد به دست آمده، قواعد غیربدهی مشخص و علت وجود این قواعد با کمک خبرگان تبیین شده است. از جمله نتایج، حذف متغیر ارزیابی عملکرد به عنوان متغیر هدف در روند این پژوهش است که ناشی از عدم دقت تکمیل فرم‌های ارزیابی عملکرد در فرآیند ارزیابی بانک بوده است. همچنین در این پژوهش مشخص شده است از مجموع ۲۶ متغیر بررسی شده، پنج متغیر: "نمره کل آزمون"، "امتیاز مصاحبه"، "مقطع تحصیلی"، "تجربه حرفه‌ای" و "استان محل خدمت" بر ارتقای داوطلبان تاثیرگذار بوده‌اند. این نتایج منجر به دانشی شده است که امکان کاربردی نمودن آن‌ها وجود خواهد داشت.

واژه‌های کلیدی: داده‌کاوی، انتخاب نیروی انسانی، درخت تصمیم، دسته‌بندی، کشف دانش، مدیریت منابع انسانی

۱. مقدمه

مهم‌ترین سرمایه‌ی سازمان، نیروی انسانی آن است. در سال‌های اخیر توجه به نیروی انسانی در سازمان‌ها بخش عظیمی از زمان و سرمایه سازمان‌های پیشرو را به خود اختصاص داده است. موفقیت یا شکست سازمان، بستگی به چگونگی جذب و نگهداری منابع انسانی آن دارد. [۱۵]

هم‌چنین انتخاب و به‌کارگیری شایسته‌ترین افراد در هر شغل یکی از مهم‌ترین مسایل و تصمیم‌های مدیریتی محسوب می‌شود. [۱]

از سوی دیگر تولید فراوان داده‌هایی که در حوزه‌های سازمانی ایجاد می‌شوند به مثابه قطعات طلائی هستند که از لجن‌های حاصل از فرآوری مس، به‌دست می‌آیند. بنابراین در دنیای امروز داده‌ها و اطلاعات اولیه از اهمیت زیادی برخوردار نیستند اغلب به‌صورت گسترده و توزیعی در اختیار همه است اما فرآوری و بازپروری آن و تولید دانش و استخراج گزاره‌های آن در فرآیند غنی‌سازی داده‌ها از اهمیت زیادی برخوردار است. بنابراین سازمان‌ها باید به دانش و مهارت‌های مربوط به فرآوری اطلاعات همانند داده-کاوی تسلط داشته باشند. [۱۸]

باتوجه به وجود بانک داده‌های فراوان در سازمان‌ها (به‌ویژه در حوزه‌ی مدیریت منابع انسانی) در این پژوهش نیز سعی بر این است ضمن استفاده از فن داده‌کاوی به‌عنوان یکی از ابزارهای فناوری اطلاعات و با پرداختن به یکی از مهم‌ترین مسایل مدیریتی در حوزه مدیریت منابع انسانی (گزینش و انتخاب کارکنان)، عوامل و قواعد موثر بر عملکرد نیروی انسانی شناسایی شده و با کشف الگوهای پنهان بین وضعیت کارکنان با کارایی و اثربخشی آن‌ها و اعلام آن به مدیران، بتوان آن‌ها را برای تصمیم‌گیری در مورد استخدام و انتخاب صحیح کارکنان، هدایت نمود.

۲. مفاهیم

۲-۱. داده‌کاوی چیست؟

داده‌کاوی در اواخر دهه‌ی ۱۹۸۰ پدیدار گشته، در دهه‌ی ۱۹۹۰ گام‌های بلندی در این شاخه از علم برداشته شده و انتظار می‌رود در این قرن به رشد و پیشرفت خود ادامه دهد. [۱۲] داده‌کاوی فرآیند کشف روابط ناشناخته و الگوی درون داده است. [۴] داده‌کاوی در حقیقت کشف ساختارهای جالب توجه، غیرمنتظره و با ارزش از داخل مجموعه‌ی وسیعی

از داده‌ها است و فعالیتی است که به‌طوراساسی با آمار و تحلیل دقیق داده‌ها منطبق است. [۱۲]

پیش‌بینی می‌شود که داده‌کاوی در دهه‌ی آینده با توسعه‌ی انقلابی مواجه شود. [۱۱] در واقع تکنولوژی MIT آن را یکی از ده تکنولوژی برتری می‌داند که نقش چشمگیری در تحول جهان خواهد داشت. [۹]

مدل فرآیندی CRISP-DM فعالیت‌های داده‌کاوی را به شش مرحله دسته‌بندی می‌کند که هر یک نیز وظایف متنوعی دارند. [۷] و [۲۵]

- | | | |
|--------------|-------------|--------------------|
| ۱. فهم تجاری | ۲. فهم داده | ۳. آماده‌سازی داده |
| ۴. مدل‌سازی | ۵. ارزیابی | ۶. پیاده‌سازی |

۲-۲. کشف دانش

فرآیند کشف دانش شامل مجموعه‌ای از تکنیک‌های داده‌کاوی است که هدف آن حل مسایل معین و تصمیم‌گیری با استفاده از تکنیک‌های ریاضی و کامپیوتر و براساس تحلیل داده‌های موجود در یک پایگاه داده بزرگ، یافتن یک راه‌حل براساس الگوهای کشف شده در داده‌ها و به کار بردن راه‌حلی برای مسئله تعریف شده می‌باشد.

یک الگو هنگامی می‌تواند نمایانگر دانش باشد که:

- به سادگی برای افراد قابل فهم باشد.
- با درجه‌ی اطمینانی قابل اعتبار باشد.
- کاربردی باشد.
- مطلب جدیدی که کاربر در گذشته از آن اطلاع نداشت یا به دنبال یافتن آن بود را ارائه دهد [۲۰].

۳-۲. گزینش و انتخاب کارکنان

بردول و رایت (۲۰۰۴) خاطر نشان می‌کنند که استخدام و گزینش فرآیندهایی معطوف به شناسایی جذب و ایمن‌سازی افراد مناسب برای رودرویی با نیازهای منابع انسانی یک سازمان است. دو اصطلاح "استخدام پایان یافت" و "گزینش شروع شد" اغلب به‌طور مداوم، به عنوان مبحثی به کار برده می‌شوند. [۳]

هر چند به‌منظور تجزیه و تحلیل، بهتر است که این دو را از هم جدا کنیم، ولی استخدام در بیشتر تعاریف ما نیمه نخست و گزینش نیمه دوم را شامل می‌شود. استخدام

معطوف به شناسایی و جذب داوطلبان موردنظر و گزینش معطوف به انتخاب مناسب‌ترین‌ها است. گزینش نمایانگر مرحله نهایی تصمیم‌سازی در فرآیند استخدام به شمار می‌رود [۸].

استخدام و گزینش، گذرگاهی برای به‌کارگیری افراد و ایجاد منابع در سازمان است. یک محیط تجاری که به‌طور فزاینده‌ای رقابتی و جهانی است، ضمن همراهی با نیاز به کیفیت و خدمت‌رسانی به مشتری، اهمیت استخدام و گزینش افراد مناسب را افزایش داده و به‌عنوان "کارفرمای انتخاب" مورد ملاحظه قرار گرفته است. [۲۲]

از سوی دیگر در بررسی به‌طورنسیب جامعی که در خصوص روش‌های موجود گزینش نیروی انسانی و دقت هر یک انجام شده است، می‌توان به نگاره‌ی ۱ اشاره نمود. [۳]

نگاره‌ی ۱. مقایسه دقت روش‌های مختلف گزینش (عدد ۱ معادل بهترین پیشگویی است)

دقت	روش‌های گزینش
۰.۷۲	ارزیابی میانی (ترفع) ^۱
۰.۶۳	آزمون‌های هوش و مصاحبه‌های ساخت‌یافته ^۲
۰.۶۲	آزمون‌های هوش و نمونه‌کار ^۳
۰.۵۶	آزمون‌های قابلیت ^۴
۰.۵۲	آزمون‌های نمونه‌کار ^۵
۰.۵۱	آزمون‌های هوش
۰.۵۱	مصاحبه‌های ساخت‌یافته ^۶
۰.۴۰	آزمون‌های شخصیت
۰.۳۵	رزومه
۰.۲۶	مصاحبه‌های نوعی ^۷
۰.۲۶	مقایسه
۰.۱۸	تعداد سال‌های تجربه کاری
۰.۱۰	مدت زمان تحصیل
۰.۰۲	خط‌شناسی ^۸
۰.۰۰	پیشگویی تصادفی

1. Assessment Centers (Promotion)
2. Intelligence Tests and Structured Interview
3. Intelligence Tests and Work Sampling
4. Ability Tests
5. Work Sample Tests
6. Structured Interview
7. Typical Interview
- 8 Graphology

داده‌کاوی می‌تواند در شناسایی ریشه‌ها و عوامل برخی از مشکلات نیروی انسانی سازمان‌ها به کار رود در این زمینه متاسفانه در ایران هیچ کاری انجام نشده است. ولی در سایر کشورها، شرکت‌ها و افرادی درباره این موضوع پیشنهادهایی داده‌اند و کارهایی نیز انجام شده است. هافمن [۱۳]، شروک [۲۴]، هوویدی [۱۴] و اولسان [۲۱] هر یک جداگانه گزارش‌هایی در مورد کاربرد داده‌کاوی در مدیریت نیروی انسانی ارائه نموده‌اند.

۲-۴. درخت تصمیم‌گیری

درخت تصمیم‌گیری یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد. درخت تصمیم‌گیری برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد. یعنی درخت تصمیم‌گیری پیش‌بینی خود را در قالب یک سری قوانین توضیح می‌دهد در حالی که در شبکه‌های عصبی تنها پیش‌بینی نهایی بیان می‌شود و چگونگی آن در خود شبکه پنهان باقی می‌ماند.

الگوریتم درخت تصمیم‌گیری با انتخاب آزمونی شروع می‌شود که بهترین جداسازی را برای دسته‌ها انجام دهد. مهم‌ترین هدف از انجام دسته‌بندی، به دست آوردن مدلی برای پیش‌بینی می‌باشد. بدین منظور از مجموعه‌ای از داده‌ها به نام "داده‌های آموزشی" که مجموعه‌ای از متغیرها و رکوردها است استفاده می‌کنیم. در مراحل بعدی، همین کار برای گره‌های پایین‌تر با داده‌های کمتر صورت می‌گیرد تا بهترین قانون‌ها حاصل شود. سرانجام درخت آن‌قدر بزرگ می‌شود تا دیگر نتوان جداسازی بهتری برای داده‌های گره انجام داد. در این مرحله باید اثربخشی درخت ایجاد شده اندازه‌گیری شود. برای این کار از یک مجموعه‌ی رکوردها یا داده‌های آزمایشی استفاده می‌شود که متفاوت با داده‌های اولیه‌ای که درخت را ایجاد کرده‌اند، می‌باشد. معیاری که اندازه‌گیری می‌شود عبارت است از: درصد داده‌هایی که به‌طور صحیح دسته‌بندی می‌شوند و دسته‌ی پیش‌بینی شده با دسته‌ی واقعی آن‌ها یکسان است.

به‌طور کلی می‌توان مزایای استفاده از روش درخت تصمیم نسبت به سایر تکنیک‌های داده‌کاوی را به شرح زیر بیان نمود:

- نسبت به سایر مدل‌های دسته‌بندی زودتر محاسبه می‌شود.
- به‌طور معمول دقت آن نسبت به سایر مدل‌ها بهتر است.
- برای یادگیری ساده و آسان است.

- قانون‌های به‌دست آمده در آن راحت‌تر درک می‌شود.
- آزمون‌های برگزار شده در هریک از الگوریتم‌های شناخته شده درخت تصمیم، متفاوت است و انتخاب شاخه‌ها و جداسازی در هریک به روش متفاوتی انجام می‌گیرد.

۳. روش پژوهش

در این پروژه که از حیث هدف، کاربردی و از جنبه‌ی ماهیت از نوع پژوهش‌های همبستگی و همخوانی محسوب می‌شود و مدل آن توسط مطالعات موردی، بررسی و آزمایش شده است، سعی بر این است که با استفاده از تکنیک‌های داده‌کاوی، قواعد و روابط بین نمرات آزمون‌های ورودی و وضعیت کارکنان با عملکرد شغلی و وضعیت ارتقاء آنان شناسایی شود. ابتدا با مطالعه و بررسی پایگاه‌های داده‌ی آزمون و اطلاعات شغلی و شخصی پذیرفته شدگان یکی از بانک‌های تجاری ایران، شاخص‌های نیروی انسانی که بر عملکرد یا ارتقاء موثر بوده و داده‌های آنها نیز موجود و پوشا بودند، شناسایی شد. در گام بعدی و با توجه به مطالعات انجام شده پیشین در خصوص داده‌کاوی و عملکرد منابع انسانی، مدل مفهومی مناسبی برای اجرای داده‌کاوی طراحی شده است. براساس فنون داده‌کاوی، داده‌ها پیش پردازش و پاک‌سازی شدند، بدین‌معناکه متغیرهایی که برای هدف موردنظر مناسب نبودند، حذف و اطلاعات تکراری یکی شد. پس از اصلاحات متغیرها، رکوردها نیز مورد پالایش، حذف و اصلاح قرار گرفتند. در این فرآیند، رکوردهایی که فیلدهای خالی داشتند حذف و در صورت لزوم برای داده‌های گم‌شده، و با استفاده از تکنیک‌های داده‌کاوی، مقدار مناسب جایگزین گردید.

از آنجایی که در اکثر مطالعه‌های انجام شده در حوزه‌ی مدیریت منابع انسانی، تکنیک درخت تصمیم روش مورد استفاده در داده‌کاوی، به کار رفته است؛ ما نیز درخت تصمیم را برای قابل فهم و آسان بودن آن در این مطالعه مدنظر قرار داده‌ایم. هم‌چنین الگوریتم‌های مورد استفاده در این تکنیک، عبارتند از QUEST (Quick Unbiased and Efficient) (Statistical Tree) [۱۹]، CHAID (Chi-squared Automatic Interaction Detector) [۱۷]، C5.0 [۱۶] و CART [۵] که نحوه‌ی انتخاب بهترین الگوریتم نیز در ادامه آورده شده است. مجموعه داده‌ها با استفاده از نرم افزار Excel آماده‌سازی و استخراج قواعد نیز با استفاده از نرم افزار SPSS Clementine (ویرایش ۱۲) انجام شده است.

۳-۱. جامعه‌ی آماری و روش گردآوری اطلاعات

جامعه‌ی آماری این پژوهش که به‌منظور آزمایش مدل طراحی شده، مورد استفاده قرار گرفته، عبارت است از:

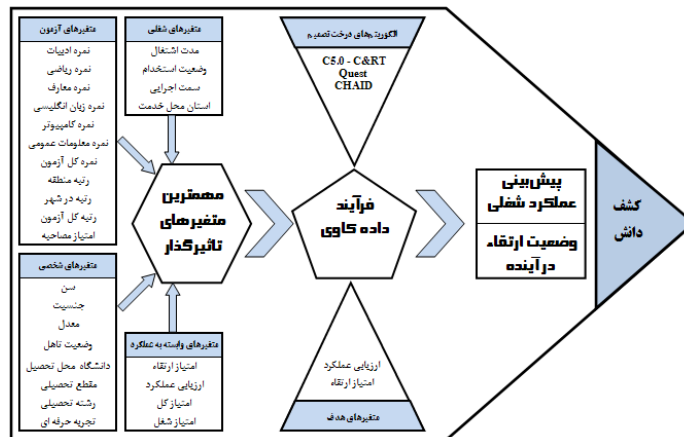
"پذیرفته‌شدگان و استخدام‌شدگان آزمون ورودی سراسری سال‌های ۱۳۸۳ و ۱۳۸۴ یکی از بانک‌های تجاری در ایران"

این جامعه‌ی آماری، خود از جامعه‌ای بزرگتر استخراج شده که در واقع عبارت است از کلیه‌ی شرکت‌کنندگان این آزمون‌ها که به‌طور متوسط بالغ بر ۳۰۰۰۰ نفر در هر سال بوده‌اند. ولی به‌واسطه‌ی نیاز به سنجش ارتباط نمرات آزمون‌ها با عملکرد و ارتقاء شغلی شرکت‌کنندگان، به‌ناچار تنها تعدادی از این جامعه، که درنهایت به استخدام بانک درآمده و حداقل یک سال از فعالیت آن‌ها در بانک گذشته، انتخاب شده‌اند. به‌طوری‌که امکان اخذ نمره‌ی ارزیابی عملکرد شغلی و امتیاز ارتقاء برای این افراد امکان پذیر بوده است.

با توجه به این توضیح‌ها، تعداد پذیرفته‌شدگان و استخدام‌شدگان طی سال‌های ۱۳۸۳ و ۱۳۸۴ بانک، بالغ بر ۹۴۰ نفر بوده‌اند که با توجه به در دسترس نبودن نمرات آزمون ورودی برای تمام این جامعه، در نهایت نمونه‌ای ۷۱۱ نفری که در مجموع دارای ۲۶ فیلد اطلاعاتی مفید، صحیح و کامل از اطلاعات پرسنلی، وضعیت استخدامی، نمرات آزمون و عملکرد شغلی بوده‌اند جامعه‌ی نهایی انتخاب شده است.

۳-۲. مدل مفهومی

باتوجه به متغیرهای تعریف شده و بانک داده‌های موجود، مدل مفهومی کشف دانش از نتایج آزمون‌های ورودی به صورت نمودار ۱ طراحی شده است:



نمودار ۱. مدل مفهومی کشف دانش از بانک داده آزمون ورودی داوطلبان

همان گونه که از مدل نیز مشخص است، نخست، اقدام به تعریف گروه های متغیر و شناسایی مهم ترین متغیرهای تأثیرگذار خواهد شد، سپس با ورود این متغیرها و هم چنین متغیرهای هدف انتخاب شده (ارزیابی عملکرد و امتیاز ارتقاء) به تفکیک هر الگوریتم، فرآیند داده کاوی اجرا و نتایج آن به منظور پیش بینی عملکرد شغلی و هم چنین وضعیت ارتقای داوطلبان در آینده به کار خواهد رفت. نتایج نهایی نیز به عنوان دانش سازمانی و به منظور افزایش کارایی و اثربخشی در مجموعه سازمان، مورد استفاده قرار خواهند گرفت.

۳-۳. محدودیت پژوهش

تنها موردی که محدودیت در این پژوهش تلقی می شود، الزام بانک داده ها و متغیرهای از پیش تعریف شده توسط بانک مورد مطالعه بوده است. در نتیجه امکان پیشنهاد متغیرهای بیشتر برای محدودیت داده ها، برای تیم کارشناسی وجود نداشته است.

۴. مطالعه موردی

۴-۱. آماده سازی داده ها

آماده سازی داده ها شامل کلیه مراحل است که رکوردها و متغیرها را برای ساخت مدل و درخت، آماده می نماید. اغلب این گام از فرآیند داده کاوی و کشف دانش، طولانی ترین و مهم ترین مرحله محسوب می شود. به طوری که هر قدر کیفیت آماده سازی داده ها بهتر انجام شود و زمان بیشتری برای آن صرف شود، مدل سازی و نتیجه گیری نیز بهتر و آسان تر

خواهد بود. در ادامه برای آشنایی با فرآیند طی شده در این فاز، فعالیت‌های طی شده به- صورت گام به گام ذکر شده است.

گام ۱) ابتدا با برگزاری جلسات توجیهی با مدیران و مسئولان منابع انسانی بانک، هدف از انجام پژوهش تشریح و کلیه داده‌ها و اطلاعات موجود در رابطه با پرسنل و کارکنان ارزیابی شده است.

گام ۲) در این مرحله نوع متغیرهای موردنیاز که شامل ۴ دسته (متغیرهای شخصی، متغیرهای آزمون، متغیرهای شغلی و متغیرهای وابسته به عملکرد) بودند و با توجه به فیلدهای موجود با همکاری کارشناسان منابع انسانی بانک تعریف شده است.

گام ۳) طبق آمار مدیریت منابع انسانی بانک، از میان شرکت‌کنندگان در آزمون سال‌های ۸۳ و ۸۴، در حدود ۹۷۰ نفر پذیرفته و استخدام شده‌اند. با توجه به پراکندگی و مجزا بودن بانک داده‌های آزمون و بانک داده‌های پرسنلی، با استفاده از انواع ابزارهای برنامه‌نویسی و قابلیت‌های برنامه Microsoft Excel این انطباق انجام و در نهایت ۷۱۷ رکورد با ۳۷ فیلد مطابقت داده شدند.

گام ۴) از آنجایی که هر فیلد نمایانگر یک متغیر است بنابراین در اولین گام پس از تشکیل بانک داده‌ها، اقدام به حذف و یا شناسایی متغیرهای مشابه و ترکیب و تبدیل آن‌ها به یکدیگر شده است. طی این اقدام، ۳۷ فیلد موجود، تبدیل به ۲۶ فیلد نهایی شده است.

گام ۵) در مرحله‌ی بعد، به بررسی و بازنگری مقادیر (کمی و کیفی) هر متغیر پرداخته شد که با ترکیب، تلفیق، تعریف دسته و طبقه، حذف و کاهش همراه بوده است. سپس اقدام به بررسی رکوردها شده است که با توجه به ناقص بودن اطلاعات اصلی ۶ رکورد، در کل از مجموعه‌ی رکوردها حذف و در نتیجه تعداد کل رکوردهای سالم به ۷۱۱ مورد رسید.

گام ۶) در خاتمه پس از انجام فعالیت‌های ذکر شده، بانک داده‌ها در قالب فایل Excel و با ۷۱۱ رکورد و ۲۶ متغیر جهت ساخت مدل و درخت آماده شده است. این متغیرها به نوبه‌ی خود نیز در چهار دسته تقسیم بندی شدند که این دسته‌ها عبارت است از:

- متغیرهای شخصی (شامل: سن، معدل، تجربه‌ی حرفه‌ای، وضعیت تأهل، جنسیت، مقطع تحصیلی، رشته‌ی تحصیلی و دانشگاه محل تحصیل)

- متغیرهای آزمون (شامل: نمره ادبیات، نمره معارف، نمره ریاضی، نمره زبان، نمره کامپیوتر، نمره معلومات عمومی، نمره کل آزمون، رتبه منطقه، رتبه کل آزمون و امتیاز مصاحبه)
- متغیرهای شغلی (شامل: مدت اشتغال، وضعیت استخدام، سمت اجرایی و استان محل خدمت)
- متغیرهای وابسته به عملکرد (شامل: امتیاز ارتقاء، ارزیابی عملکرد، امتیاز شغل و امتیاز کل)

۲-۴. تعریف متغیرهای هدف

باتوجه به مدل مفهومی ارائه شده (نمودار ۲)، متغیرهای هدف در این پژوهش، عبارت است از: امتیاز ارتقاء و ارزیابی عملکرد.

- **امتیاز ارتقاء:** باتوجه به توانمندی کارکنان در ارتقای شغل خود در مدت اشتغال، این متغیر عبارت است از تفاضل امتیاز شغل سمت فعلی از امتیاز شغل اولین سمت کارمند (داوطلب) در بدو استخدام. بدیهی است که در صورت عدم تغییر سمت کارمند در مدت اشتغال وی، این امتیاز صفر خواهد بود. همچنین حداکثر امتیاز موجود در بین داوطلبان که به این متغیر تعلق گرفته است، برابر ۱۱۵۰ است. با این وصف، و باتوجه به این که این متغیر یکی از متغیرهای هدف تعریف شده است، امتیاز محاسبه شده در ۴ دسته طبقه‌بندی شده که در نگاره ۲ آمده است.

نگاره ۲. گزینه‌ها و طبقه‌بندی امتیاز ارتقاء

امتیاز ارتقاء	برچسب	معیار
۷۰۰ - ۱۱۵۰	A	ارتقای عالی
۳۰۰ - ۵۰۰	B	ارتقای خوب
۵۰ - ۲۰۰	C	ارتقا داشته است.
صفر	D	بدون ارتقا

- **ارزیابی عملکرد:** این متغیر بر مبنای آیین‌نامه‌ی مشخصی در بانک، حداقل سالی ۲ بار برای هر کارمند و توسط مدیر مستقیم وی تکمیل و عدد آن بر مبنای ۱۰۰ ارائه می‌شود. برای آسانی اجرای داده‌کاوی بر اساس متغیر وابسته کلیه‌ی اعداد ارائه شده تبدیل به شاخص کیفی شده‌اند که مطابق نگاره ۳ در ۷ دسته طبقه‌بندی شده است.

نگاره‌ی ۳. گزینه‌ها و طبقه‌بندی امتیاز ارزیابی عملکرد

امتیاز	معیار
۱۰۰	عالی
۹۹ - ۱۰۰	خیلی خوب
۹۸ - ۹۹	خوب
۹۶ - ۹۸	معمولی
۹۳ - ۹۶	ضعیف
۹۰ - ۹۳	خیلی ضعیف
< ۹۰	بدون کارایی

۳-۴. تشکیل درخت‌ها

برای تشکیل درخت در گام نخست، پرسش‌های متعددی مطرح می‌شود که عبارت است از:

- آیا باید تمام متغیرها را هم‌زمان در تشکیل درخت دخالت داد؟
- آیا دقت درخت به تعداد متغیرها یا رکوردها وابسته است؟
- چه ترکیبی از متغیرها، بهترین درخت را تشکیل خواهد داد؟
- آیا انتخاب متغیر هدف نیز در تشکیل درخت اهمیت دارد؟
- کدام الگوریتم داده‌کاوی، بهترین نتیجه را حاصل خواهد کرد؟

برای پاسخ به پرسش‌های فوق، فرآیند طی شده در تشکیل درخت‌ها به شرح زیر می‌-

باشد:

گام ۱) باتوجه به شناسایی چهار دسته متغیر (شخصی، آزمون، شغلی و وابسته به عملکرد) و هم‌چنین وجود دو نوع متغیر هدف (ارزیابی عملکرد و امتیاز ارتقاء) و از سوی دیگر وجود الگوریتم‌های متفاوت برای تشکیل درخت تصمیم، ابتدا به بررسی کلیه‌ی مدل‌های ممکن که امکان تشکیل آن‌ها با بانک داده‌های موجود، میسر است، پرداخته شد. باتوجه به وجود ۴ حالت متغیر مستقل، ۲ متغیر هدف و ۶ نوع الگوریتم مطرح برای تکنیک درخت تصمیم (QUEST، CHAID، C5.0، CART، CART در حالت Towing و CART در حالت Ordered)، امکان تشکیل ۱۸۰ درخت متفاوت به صورت زیر شناسایی گردید:

$$(2^x - 1) * y * A$$

که در آن، x تعداد دسته‌های متغیرهای مستقل، y تعداد متغیر هدف و A تعداد

الگوریتم می‌باشد. در نتیجه:

$$180 = 6 * 2 * (1 - 2^4)$$

باتوجه به این که تعداد رکوردهای آموزشی (با انتخاب تصادفی)، ۵۵۰ مورد در نظر گرفته شدند، هریک از ۱۸۰ مدل شناسایی شده با این رکوردها و با استفاده از نرم افزار SPSS Clementine آموزش داده شدند و دقت هر کدام محاسبه و ثبت شد. شایان ذکر است که در ۷ حالت شناسایی شده با الگوریتم QUEST، امکان تشکیل درخت وجود نداشت. از سوی دیگر درخت‌هایی که تعداد سطوح آن‌ها بیش از ۴ سطح است، به دلیل طولانی شدن قاعده‌های استخراج شده، کارایی چندانی نخواهند داشت. به همین دلیل با اعمال محدودیت تعداد سطح (حداکثر ۴ سطح)، درخت‌های آن‌ها تشکیل شده است. گام ۲) باتوجه به نتایج حاصل شده از درخت‌های تشکیل شده، و بررسی مهم‌ترین متغیرهای تاثیرگذار در بین کل درخت‌ها (۱۷۳ درخت)، ۱۲ درخت نیز با مهم‌ترین متغیرهای حاصل از درخت‌های موجود، تشکیل شده است که جمع‌بندی خطای حاصل از هر الگوریتم و با توجه به ۱۸۵ درخت موجود در نگاره ۴ ارایه شده است.

نگاره ۴. میزان دقت هر الگوریتم بر مبنای نتایج ۱۸۵ درخت ساخته شده

نام الگوریتم	میانگین دقت الگوریتم در تولید درخت‌ها	میانگین دقت تشکیل درخت توسط کل الگوریتم‌ها
QUEST	٪۶۳.۴۹	٪۶۷.۲۹
CHAID	٪۶۰.۳۳	
C 5.0	٪۸۰.۴۳	
CART	٪۶۵.۱۱	
CART (Towing)	٪۶۶.۷۸	
CART (Ordered)	٪۶۶.۷۸	

گام ۳) در نهایت با توجه به نتایج، اقدام به تجزیه و تحلیل دقت درخت‌ها با توجه به نوع الگوریتم مورد استفاده و همچنین نوع متغیر هدف شده است که نتایج آن در نگاره ۵) خلاصه شده است.

تکاه‌ی ۵. میزان دقت هر متغیر هدف بر مبنای نتایج ۱۸۵ درخت ساخته شده

نام متغیر هدف	نام الگوریتم	میانگین دقت در الگوریتم‌ها	توضیحات	میانگین دقت کل متغیر هدف در تمام الگوریتم‌ها
ارزیابی عملکرد	QUEST	٪۳۰.۸۸	میانگین حاصل از ۱۲ درخت	٪۴۳.۱۱
	CHAID	٪۳۰.۹۴	میانگین حاصل از ۱۶ درخت	
	C 5.0	٪۶۸.۲۰	میانگین حاصل از ۱۶ درخت	
	CART	٪۳۹.۹۹	میانگین حاصل از ۱۶ درخت	
	CART (Towing)	٪۴۲.۷۸	میانگین حاصل از ۱۶ درخت	
	CART (Ordered)	٪۴۲.۷۸	میانگین حاصل از ۱۶ درخت	
امتیاز ارتقاء	QUEST	٪۹۳.۶۰	میانگین حاصل از ۱۳ درخت	٪۹۱.۲۱
	CHAID	٪۸۹.۷۳	میانگین حاصل از ۱۶ درخت	
	C 5.0	٪۹۲.۶۵	میانگین حاصل از ۱۶ درخت	
	CART	٪۹۰.۲۳	میانگین حاصل از ۱۶ درخت	
	CART (Towing)	٪۹۰.۷۷	میانگین حاصل از ۱۶ درخت	
	CART (Ordered)	٪۹۰.۷۷	میانگین حاصل از ۱۶ درخت	

۴-۴. انتخاب اولیه‌ی درخت‌های مناسب برای استخراج قواعد

باتوجه به کثرت درخت‌های تولید شده، بدیهی است که قواعد حاصل شده از این درخت‌ها نیز بسیار زیاد خواهند بود. به منظور نتیجه‌گیری مناسب از قواعد حاصل شده، نیاز است تا درخت‌های مناسب از سایر مدل‌های تولید شده، جدا شوند. برای انجام انتخاب بین درخت‌ها با نظر کارشناسی، معیارهایی تعیین شده است که عبارت است از:

۱. انتخاب درخت‌هایی با دقت بالای ۷۰ درصد.
۲. انتخاب درخت‌هایی که حداقل تعداد سطوح آن‌ها بیشتر از یک بوده است. (یعنی به طور حتم قاعده‌ای شکل گرفته است).
۳. انتخاب درخت‌هایی که حداقل یکی از گروه‌های "متغیر شخصی" یا "متغیر آزمون" در آن قرار داشته‌اند.
۴. انتخاب درخت‌هایی که مهم‌ترین متغیرهای تأثیرگذار آن‌ها یکی از متغیرهای عضو گروه متغیرهای شخصی یا متغیرهای آزمون بوده است.
۵. با توجه به اعمال شرایط گفته شده، در نهایت ۱۷ درخت اولیه برای مرحله‌ی آزمون انتخاب شده‌اند که از بین آن‌ها، ۳ درخت با متغیر هدف "ارزیابی عملکرد" وجود داشتند که الگوریتم مورد استفاده در هر سه آن‌ها، C5.0 بوده است.

۴-۵. آزمون درخت‌های تولید شده

به منظور آزمون ۱۷ مدل انتخاب شده برای استخراج قواعد، از داده‌های آزمایشی استفاده شده است. ۱۶۱ رکورد از کل داده‌های موجود بانک داده‌های نهایی، داده‌ی آزمایشی و به صورت تصادفی انتخاب شده است و برای این مرحله کنار گذاشته شدند. به عبارت دیگر، هیچ‌یک از این ۱۶۱ رکورد در آموزش دادن درخت و تشکیل مدل، دخالت نداشته‌اند. در این مرحله هر ۱۷ درخت انتخاب شده در مرحله قبل، با ورود داده‌های آزمایشی مورد آزمون قرار گرفته و دقت آن‌ها تعیین شده است.

باتوجه به این که مطابق نگاره‌ی ۶، دقت حاصل از سه درخت تولید شده با متغیر هدف "ارزیابی عملکرد"، بسیار کاهش یافتند، بنابراین این سه درخت و در نتیجه متغیر هدف "ارزیابی عملکرد" قبل از ورود به مرحله‌ی استخراج قواعد حذف شده است.

نگاره‌ی ۶. میزان دقت آزمون درخت‌های حاصل شده با متغیر هدف "ارزیابی عملکرد"

کد درخت	میزان دقت با داده‌های آموزشی	میزان دقت با داده‌ی آزمایشی	فاضل دقت حاصل از آزمون داده‌ها
C5-A-2	٪۶۰.۹۱	٪۲۳.۶۰	٪۳۷.۳۱
C5-A-5	٪۹۷.۶۴	٪۲۱.۱۲	٪۷۶.۵۲
C5-A-11	٪۸۴.۰۰	٪۲۱.۷۴	٪۶۲.۲۶

در نهایت تعداد مدل‌های نهایی انتخاب شده برای استخراج قواعد به ۱۴ مورد رسید که باتوجه به یکسانی مدل و قواعد تولید شده توسط دو الگوریتم CART (Ordered) و CART (Towing) به طور صرف یکی از آن‌ها در این انتخاب در نظر گرفته شده است. که نتایج آزمون آن‌ها در نگاره‌ی ۷ خلاصه شده است.

نگاره‌ی ۷. میزان دقت هر الگوریتم بر مبنای نتایج ۱۴ درخت منتخب

نام الگوریتم	میانگین دقت الگوریتم در تولید درخت‌ها	میزان دقت با داده‌های آزمایشی	خطای آزمون با داده‌های آزمایشی
QUEST	٪۶۹.۰۹	٪۷۰.۱۹	٪۱.۱۰-
CHAID	٪۷۱.۸۲	٪۷۷.۰۲	٪۵.۲۰-
C 5.0	٪۷۲.۳۲	٪۶۶.۴۶	٪۵.۸۶
CART	٪۸۱.۷۴	٪۷۹.۳۳	٪۲.۴۱
CART (Ordered) & (Towing)	٪۷۲.۶۱	٪۷۰.۵۲	٪۲.۰۹
میانگین کل	٪۷۵.۴۸	٪۷۲.۹۵	٪۲.۵۳

جمع بندی آنالیزهای ارایه شده نشان دهنده این است که در مجموع درخت‌های منتخب، متغیر هدف "امتیاز ارتقاء" و الگوریتم CART بیشترین دقت را حاصل کرده‌اند. علاوه بر آن که متغیر هدف "ارزیابی عملکرد" در هیچ کدام از مدل‌های نهایی مشاهده نمی‌شود و مدل‌های ساخته شده توسط آن از خطای بسیار بالایی برخوردار هستند.

۴-۶. استخراج قواعد

پس از انتخاب نهایی درخت‌ها که طی فرآیندهای توضیح داده شده در بخش‌های گذشته انجام شده است، نوبت به استخراج قواعد و درخت‌های تشکیل شده می‌رسد. منظور از استخراج قواعد، بررسی تک تک درخت‌های تشکیل شده و تعریف هر قاعده از زبان ریاضی و منطقی به زبان قابل فهم و قابل برداشت برای خبرگان است.

از سوی دیگر، هر قاعده‌ی تعیین شده، دارای احتمال وقوعی خواهد بود که براساس تعداد رکورد منطبق با قاعده تعریف شده از بین کل رکوردهای آن شاخه به دست می‌آید:

$$P = n_c / n$$

که در آن: P احتمال وقوع قاعده در بین سایر رکوردهای بانک داده، n تعداد کل رکوردهای بررسی شده در شاخه و n_c تعداد رکوردهای منطبق با قاعده است.

بدیهی است که هر میزان شاخه‌ای عمیق تر باشد (به عبارت دیگر در سطوح پایین تری از درخت قرار گیرد)، مقدار n کمتر خواهد شد. البته این مطلب به طور لزوم به معنای کاهش P نیست. در نتیجه می‌توان به قواعد استخراج شده اطمینان کرد.

از بین ۸۹ قاعده استخراجی و با توجه به تکرار قواعد در بعضی از درخت‌ها، در نهایت ۶۸ قاعده از فرآیند داده کاوی به دست آمد. هم چنین متغیرهایی که با توجه به قواعد استخراج شده، دارای همبستگی با متغیر هدف (امتیاز ارتقاء) بوده‌اند عبارت است از: "استان محل خدمت، مقطع تحصیلی، رشته تحصیلی، دانشگاه محل تحصیل، نمره کل آزمون، امتیاز مصاحبه، نمره ریاضی، نمره زبان، نمره معلومات عمومی و تجربه حرفه‌ای".

بدیهی است که سایر متغیرها نیز ارتباط معناداری با امتیاز ارتقاء نداشته‌اند. در نهایت و در بین ۶۸ قاعده‌ی تولید شده، فقط ۱۰ متغیر، تأثیر گذار بوده‌اند.

از طرف دیگر، عدم وجود متغیر هدف "ارزیابی عملکرد" در بین قواعد نهایی، ناشی از بی تأثیر بودن این متغیر در نتایج حاصل شده است. در حقیقت این متغیر در مراحل قبل

از استخراج قواعد، حذف شده است.

۴-۷. انتخاب قواعد

باتوجه به تعداد زیاد قواعد، بدیهی است که نمی‌توان تمام این قواعد را، منطقی دانست. این گفته دارای ۳ دلیل واضح است:

۱. تمام قواعد ارایه شده از احتمال و تعداد بالایی برخوردار نیستند.
 ۲. بعضی از قواعد ممکن است باهم در تضاد باشند.
 ۳. الگوهای استخراج شده از بین داده‌ها، به‌طور لزوم به معنای دانش موجود در آن‌ها نیست و ممکن است برحسب تصادف بوجود آمده باشند.
- باتوجه به این که تحلیل و بررسی همه‌ی این قواعد، می‌تواند مشکل باشد، در نتیجه براساس نظر خبرگان و بر مبنای دو اصل زیر، اقدام به تعیین قواعد نهایی شده است.
- قواعدی که هم‌زمان در آن‌ها $n \geq 100$ و $P \geq 60\%$ است.
 - قواعدی که هم‌زمان در آن‌ها $n \geq 50$ و $P \geq 70\%$ است.
- با اعمال قواعد فوق و اظهار نظر خبرگان مدیریت منابع انسانی بر روی قواعد انتخاب شده با روش بالا، در نهایت ۱۹ قاعده از ۶۸ قاعده باقی ماند که نمونه‌ای از آن‌ها به شرح نگاره‌ی زیر است:

نگاره‌ی ۸. نمونه‌ای از قواعد انتخاب شده نهایی

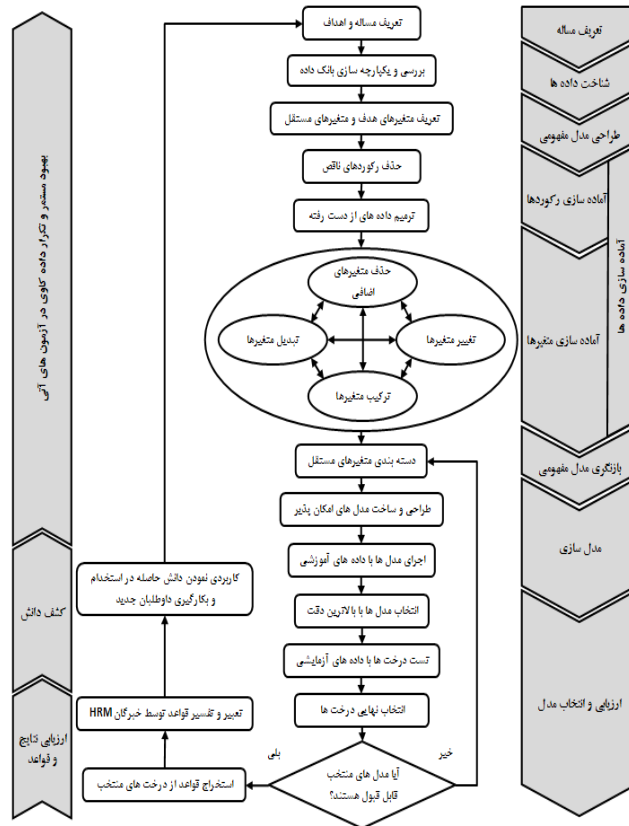
شرح قاعده‌ی استخراج شده	احتمال (P)	تعداد در شاخه (n)	کد قاعده
داوطلبانی که نمره‌ی کل آزمون و نمره‌ی مصاحبه‌ی آن‌ها به ترتیب کمتر یا مساوی ۶۴۴۹.۵ و ۸۷.۵ باشد، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۸۲.۰۰	۲۲۸	R12
داوطلبانی که نمره‌ی کل آزمون آن‌ها بیشتر از ۶۴۴۹.۵ ولی امتیاز مصاحبه‌ی آن‌ها کمتر یا مساوی ۸۴.۵ باشد، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۶۳.۵۰	۱۹۲	R15
داوطلبانی که قبل از استخدام تجربه‌ی حرفه‌ای نداشته‌اند و مدرک کارشناسی دارند، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۸۹.۷۰	۱۱۷	R27
داوطلبانی که قبل از استخدام تجربه‌ی کاری خیلی کمی داشته‌اند و نمره‌ی کل آزمون آن‌ها کمتر یا مساوی ۶۴۴۹.۵ است، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۸۴.۵۰	۲۲۰	R36
داوطلبانی که امتیاز مصاحبه‌ی آن‌ها کمتر یا مساوی ۸۴.۵ و نمره‌ی کل آزمون آن‌ها بیشتر از ۶۴۴۹.۵ باشد، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۶۲.۶۰	۱۸۲	R40
داوطلبانی که امتیاز شغل آن‌ها در حال حاضر بیشتر از ۲۰۷۵ بوده و دارای مدرک تحصیلی دیپلم یا کاردانی هستند، وضعیت ارتقای آن‌ها پس از ۳ سال، B است.	٪۱۰۰.۰۰	۱۴۸	R53
داوطلبانی که دارای مدرک کارشناسی هستند، وضعیت ارتقای آن‌ها پس از ۳ سال، D است.	٪۸۳.۲۰	۱۳۷	R60

هم چنین متغیرهایی که با توجه به قواعد انتخاب شده‌ی نهایی، دارای همبستگی با متغیر هدف (امتیاز ارتقاء) بوده‌اند به ۵ متغیر کاهش یافتند که عبارت است از:

- مقطع تحصیلی (تکرار شده در ۱۱ قاعده)
- نمره ی کل آزمون (تکرار شده در ۸ قاعده)
- امتیاز مصاحبه (تکرار شده در ۸ قاعده)
- استان محل خدمت (تکرار شده در ۶ قاعده)
- تجربه‌ی حرفه‌ای (تکرار شده در ۲ قاعده)

۵. مدل نهایی پژوهش

در نهایت و باتوجه به مراحل طی شده در پژوهش و نتایج، مدل نهایی این پژوهش به صورت نمودار ۲ارایه شده است:



نمودار ۲. مدل کشف دانش حاصل از آزمون‌های ورودی داوطلبان با روش داده کاوی

همان گونه که ملاحظه می‌شود، کلیه مراحل توضیح داده شده در قسمت‌های قبلی، در مدل گفته شده لحاظ شده است و توضیح‌های آن‌ها نیز پیش‌تر ارائه شده است. تنها قسمتی که در این مدل قابل تأمل است، آخرین مرحله از مراحل ۱۱ گانه‌ی فرآیند آرایه شده است که تأکیدی بر اهمیت تکرارپذیری فرآیند داده‌کاوی و استفاده از دانش کشف شده در انجام مجدد این مراحل می‌باشد. به طوری که امکان بهبود نتایج داده‌کاوی و استخراج قواعد در هربار برگزاری آزمون فراهم خواهد شد و داده‌کاوی با رویکردی هدایت شده انجام خواهد گرفت.

۶. نتیجه‌گیری و پیشنهادها

به کارگیری صحیح و علمی کارکنانی که از طریق آزمون‌های ورودی مؤسسه‌ها و شرکت‌ها انتخاب می‌شوند و به‌ویژه در سازمان‌های خدمات محور، دارای اهمیت خاصی است که استفاده از علم داده‌کاوی و کشف دانش نهفته در این زمینه بسیار موثر خواهد بود. در این مقاله ضمن بیان رویکردهای پیشین به حوزه‌ی استفاده از داده‌کاوی در مدیریت منابع انسانی، تلاش شد تا مدل‌های پیشین توسعه داده شود و مدل جدیدی با نگرشی پویا و تأکید بر انتخاب بهترین درخت (های) ممکن متناسب با بانک داده‌ها ارائه شود. هم‌چنین با توجه به رویکرد مفهومی در خصوص متغیرهای تأثیرگذار و متغیرهای هدف، مشخص شده است که متغیر هدف "ارزیابی عملکرد"، گزینه‌ی مناسبی برای برقراری ارتباط با سایر متغیرها نبوده و هم‌چنین از بین ۲۶ متغیر تأثیرگذار فقط پنج متغیر: "استان محل خدمت، مقطع تحصیلی، نمره‌ی کل آزمون، امتیاز مصاحبه و تجربه‌ی حرفه‌ای" بر متغیر هدف "امتیاز ارتقاء" تأثیرگذار هستند.

در همین رابطه پیشنهادهایی به سازمان مورد مطالعه ارائه شده است که نتیجه‌ی تجزیه و تحلیل خروجی‌های مدل طراحی شده بوده است. از جمله این پیشنهادها می‌توان به موارد زیر اشاره نمود:

- با توجه به حذف متغیرهای آزمون مانند (نمره‌ی ادبیات، معارف، ریاضی و معلومات عمومی) ضروری است نوع دروس انتخابی و یا محتوای آنان تغییر یابد و با هدف اندازه‌گیری پتانسیل‌های آینده (نه فقط اطلاعات گذشته) طراحی شوند.
- روش‌های امتیازدهی ارزیابی عملکرد تغییر یافته و برای تکمیل اطلاعات آن از ابزارهای

آموزشی و فرهنگ‌سازی استفاده شود.

• می‌توان به بررسی و در صورت امکان‌پذیر بودن به روش‌هایی که برگزاری قسمتی از آزمون را به صورت تشریحی فراهم می‌نماید نیز پرداخت.
در نهایت، آنچه بیشتر حایز اهمیت است، استفاده‌ی بهینه و گسترده از علم داده‌کاوی در سایر حوزه‌های مدیریتی است که می‌تواند ضمن استفاده‌ی کم هزینه از داده‌های موجود در سازمان‌ها، افق جدیدی را از دانش‌های نهفته در سازمان برای مدیران روشن نماید.

منابع

۱. محمدی محمود، امین‌ناصری محمدرضا، آذر عادل. الگوی انتخاب و به‌کارگیری نیروی انسانی در سازمان‌ها با استفاده از شبکه‌های عصبی مصنوعی، فصلنامه مدرس (ویژه‌نامه مدیریت) ۱۳۸۳، ۸(۳) (پیاپی ۳۶): ۱۳۵-۱۵۶
2. Anderson A. Effective Personnel Management: A Skills and Activity-based Approach. Oxford: Blackwell Business; 1994.
3. Beardwell I., L. Holden, T. Claydon. Human Resource Management. Harlow: FT/Prentice Hall; 2004.
4. Berry Michael J. A., Gordon S. Linoff. Data Mining Techniques for Marketing Sales and Customer Relationship Management. 2 Ed. John.Wiley.&.Sons. Publishing Inc.; 2004.
5. Breiman L., Friedman J. H., Olshen R. A., Stone C. J. Classification and Regression Trees. New York: Chapman & Hall; 1984.
6. Chien Chen-Fu, Li-Fei Chen. Data Mining To Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry. Expert Systems With Applications 2007; 34: 280-290.
7. Clifton C., Thuraisingham. Emerging Standards for Data Mining. Computer Standard & Interface 2001; 23(3): 187-193.
8. Cowling A., C. Mailer. Managing Human Resources. 2nd Edition. London: Edward Arnold; 1990.
9. Daniel T., John Larose. Discovery Knowledge In Data, An Introduction To Data Mining. Wiley Interscience; 2005.
10. Denker Robert. Audites For Human Resource Applications; 2000. [Http://www.Wizsoft.Com/App4-Wr.Html](http://www.Wizsoft.Com/App4-Wr.Html).

11. Daniel T., John Larose. Data Mining Methods And Models. Wiley & Sons Publishing Inc.; 2006
12. Hand David, Heikki Mannila, Padhraic Smyth. Principles of Data Mining. The MIT Press; 2001.
13. Hoffmann Carl C. Decision Support & Data Warehousing in Human Resource; 1998.
[Http://www.Cob.Vt.Edu/Mgmt/Carlson/AHRIS/Ihrim.Ppt](http://www.Cob.Vt.Edu/Mgmt/Carlson/AHRIS/Ihrim.Ppt).
14. Howedy Rodney. First Door Uses Web Trends to Create a Human Resource Database; 2002. [Http://www.Dmreview.Com/DM Review First Door Uses Web Trends To Create A Human Resource. Htm/ DM Review Magazin Article](http://www.Dmreview.Com/DM%20Review%20First%20Door%20Uses%20Web%20Trends%20To%20Create%20A%20Human%20Resource.Htm/).
15. Jazani N., Human Resource Management, Tehran: Ney Publishing Inc.; 2000
16. Kantardzic Mehmed. Data Mining: Concept, Models and Algorithms. Wiley: IEE Press; 2003.
17. Kass G. V., An exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics 1980; 29(2): 119-127.
18. Khaef E., Ahmad.A., Mottaghi, P., Sebt, M.V., Examining the Influence of Employment Model Utilization Based on Data Mining, on the Employees' Replacement Rate, First Iran Data Mining Conference, Nov; 2007.
19. Loh W.-Y., Shih Y.-S., Split Selection Methods for Classification Trees, Statistica Sinica 1997; 7: 815-840
20. Mendonca M., N. L. Sunderhaft. Mining Software Engineering Data Survey; 1999. www.Dacs.Ditc.Mil/Techs/Datamining/Index.Html
21. Olson David L, Durusun Delen. Advanced Data Mining Techniques. Hidelberg: spring. 2008;
[Http://www.Ait.Unl.Edu/Dolson/Datamining.ppt](http://www.Ait.Unl.Edu/Dolson/Datamining.ppt).
22. Porter K., P. Smith, R. Fagg. Leadership and Management for HR Professionals. Oxford: Butterworth Heinemann; 2006.
23. SAS Institute Inc; 2001. <http://www.Sas.Com>
24. Schroeck Michael J. DW In Human Resource Management Systems; 1999. [http://www.Dmreview.Com/DM Review Insights From The Front Line DW In Human Resource.Htm/ DM Review Magazin Article](http://www.Dmreview.Com/DM%20Review%20Insights%20From%20The%20Front%20Line%20DW%20In%20Human%20Resource.Htm/).
25. Shearer C. The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing 2000; 5: 4 13-22.