

# تعیین شاخص‌های موثر بر گزینش منابع انسانی

## با رویکرد داده‌کاوی

دکتر عادل آذر\*

دکتر پرویز احمدی\*\*

محمدوحید سبط\*\*\*

### چکیده

در این مقاله تلاش شده است با به کارگیری فنون داده‌کاوی و بررسی الگوهای درون داده‌ها و عملکرد و ارتقای داوطلبان آزمونهای ورودی، دانش نهفته در آن کشف و برای جذب نیروهای جدید به کارگرفته شود.

به همین منظور، ضمن استفاده از بانک داده آزمونهای استخدامی از الگوریتمهای تکنیک درخت تصمیم، درحکم یکی از مؤثرترین فنون داده‌کاوی، بهره گرفته شده است. همچنین با تعریف "امتیاز ارزیابی عملکرد" و "امتیاز ارتقا" به عنوان متغیرهای هدف، دانش نهفته در زمینه عملکرد افراد با سایر متغیرهای شخصی و متغیرهای آزمون، کشف شده و خبرگان آن را تبیین کرده‌اند. همچنین ضمن اصلاح مدل مفهومی این فرآیند، از جمله نتایج حاصل شده، حذف متغیر ارزیابی عملکرد بوده است. در این پژوهش مشخص شده است که 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. نتایج به دست آمده در این پژوهش و روش ارائه شده برای کشف دانش، قابل تعمیم به سایر حوزه‌هایی است که نقش سرمایه انسانی در آنها حیاتی است. از جمله این حوزه‌ها می‌توان به حوزه "آموزش و پرورش" اشاره کرد که با توجه به اهمیت والای تعلیم و تربیت، محقق ساختن آن در گرو گزینش صحیح "معلمان" است. همچنین شناخت شاخصهای مؤثر بر گزینش معلمان و استادان از طریق آزمونهای استخدامی ادوار گذشته، به منظور استفاده در گزینشهای بهتر آینده از این طریق امکان‌پذیر خواهد بود.

**کلید واژه‌ها:** گزینش منابع انسانی، داده‌کاوی، آموزش و پرورش، کشف دانش، درخت تصمیم

---

دریافت مقاله: ۸۸/۱/۲۹ پذیرش نهایی: ۸۸/۹/۳۰

\*. استاد گروه مدیریت، دانشگاه تربیت مدرس azar@modas.ac.ir

\*\* استادیار گروه مدیریت، دانشگاه تربیت مدرس ahmadi@sadadco.com

\*\*\* کارشناس ارشد مدیریت صنعتی، دانشگاه تربیت مدرس sebt@modares.ac.ir

## مقدمه

باتوجه به وجود بانک داده‌های فراوان در سازمانها (به ویژه در حوزه مدیریت منابع انسانی)، در این تحقیق با استفاده از روشهای داده‌کاوی<sup>۱</sup>، به بررسی ارتباط میان داده‌های موجود حاصل از آزمونهای استخدامی با دو متغیر مهم عملکردی (امتیاز ارتقا و نمره ارزیابی عملکرد) اقدام شده است. با بهره‌گیری از این شیوه که بسیار موثرتر از روشهای آماری، دانش موجود در داده‌های گذشته را آشکار و قواعد نهفته در آن را با روابط اگر ... آنگاه ...<sup>۲</sup> بیان می‌نماید، ۶۸ قاعده از درون بانک داده‌ها استخراج شده است که در نهایت ۱۹ قاعده از میان آنها را خبرگان قابل اعتماد تشخیص داده‌اند. سپس بر مبنای این قواعد اقدام به تفسیر آنها، کشف دانش و در نهایت ارائه پیشنهادات مؤثر برای بهبود نظام کارمندیابی و ارزیابی جامعه هدف پژوهش شده است.

فارغ از جامعه آماری مورد استفاده در این پژوهش، به کارگیری چنین شیوه‌ای در گزینش منابع انسانی، به ویژه در علوم مدیریتی، نوین و حائز اهمیت تلقی می‌شود. در نتیجه تعمیم آن در هر سازمان که نیازمند به کارگیری منابع انسانی و در نتیجه گزینش بهترین و شایسته‌ترین افراد است، یکی از اهداف این مقاله است. سازمانی که در اینجا بیشتر مدنظر است، "آموزش و پرورش" و به تبع آن، انتخاب معلمان است.

یکی از دغدغه‌های اساسی آموزش و پرورش، در حکم نهادی حیاتی و تأثیر گذار در جامعه که وظیفه آموزش خواندن و نوشتن، و مهم‌تر از آن وظیفه ارزشمند انسان‌سازی و فرهنگ‌سازی را نیز بر عهده دارد، گزینش، و استخدام معلمان بوده است، زیرا معلمان سرمایه‌های اصلی آموزش و پرورش کشور هستند. به همین منظور، برگزاری آزمونهای ادواری استخدامی در این نهاد، امری طبیعی بوده است. در نتیجه بانک داده‌های ارزشمندی در این حوزه وجود دارد که امکان گزینش بر مبنای داده‌کاوی را برای گزینش معلمان فراهم می‌آورد.

## مفاهیم

### ۱. کشف دانش و داده‌کاوی

در اواخر سال ۱۹۸۰ اصطلاح جدید کشف دانش در پایگاه داده بکار برده شد و سریعاً متخصصان هوش مصنوعی، واژه فراگیری ماشین را برای پوشش فرآیند کلی استخراج دانش از

---

1. Data mining Methods  
2. If- Then Relationships

پایگاههای داده از بدو تعریف مسئله و اهداف آن، تا تحلیلهای نهایی نتایج، مورد استفاده قرار دادند. البته ممکن است طبق این تعریف چنین سوء تفاهمی پیش آید: زمانی که ابزارهای کاوش در پایگاههای داده به کار می‌روند، نتایج به گونه‌ای سحرآمیز ظاهر می‌شوند. اما فرآیند کلی کشف دانش از خودکار بودن تا حدودی به دور است و حتی دخالتهای زیاد انسان را نیز در بر دارد. بر همین اساس بسیاری از نویسندگان، فرآیند کشف دانش را ترکیبی از هنر و علم تلقی می‌کنند (وایس و ایندورخیا<sup>۱</sup>، ۱۹۹۸).

داده‌کاوی فرآیند کشف روابط ناشناخته و الگوی درون داده است (بری و اینوف<sup>۲</sup>، ۲۰۰۴). داده‌کاوی در حقیقت کشف ساختارهای جالب توجه، غیرمنتظره و با ارزش از داخل مجموعه‌ای وسیع از داده‌ها و فعالیتی است که اساسا با آمار و تحلیل دقیق داده‌ها منطبق است (هند، مانیلا و اسمیت<sup>۳</sup>، ۲۰۰۱).

پیش‌بینی می‌شود که داده‌کاوی در دهه آینده با توسعه‌ای انقلابی مواجه شود (دانیل و لاروس<sup>۴</sup>، ۲۰۰۶). در واقع مؤسسه تکنولوژی ماساچوست<sup>۵</sup> (MIT) آن را یکی از ده تکنولوژی برتری می‌داند که نقشی چشمگیر در تحول جهان خواهد داشت. (دانیل و لاروس، ۲۰۰۵).

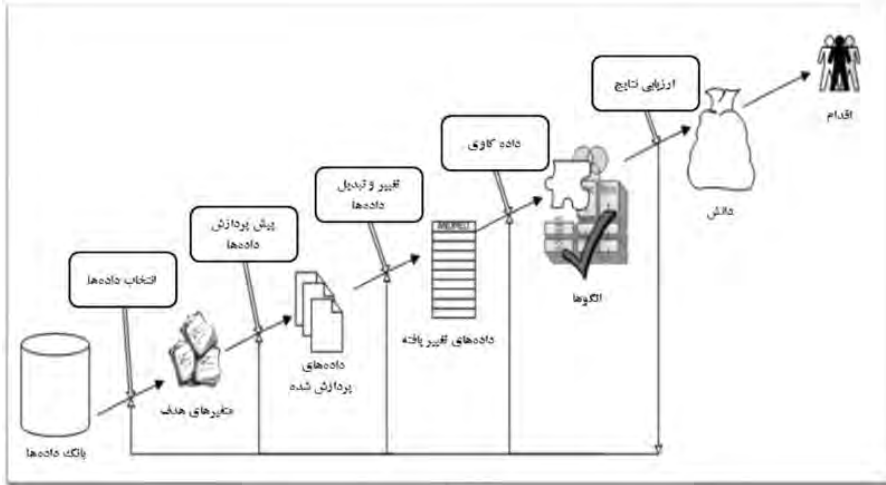
یک الگو هنگامی می‌تواند نمایانگر دانش باشد که:

- به سادگی برای افراد قابل فهم باشد.
- با درجه اطمینانی قابل اعتبار باشد.
- کاربردی باشد.
- مطلب جدیدی را ارائه دهد که کاربر قبلا از آن اطلاع نداشت یا در پی یافتن آن بود (مندونکا و ساندرهاف<sup>۶</sup>، ۱۹۹۹).

فرآیند کشف دانش طی مراحل طی و به صورت تکرارشونده صورت می‌گیرد که شکل زیر این فرآیند را به خوبی نشان می‌دهد.

---

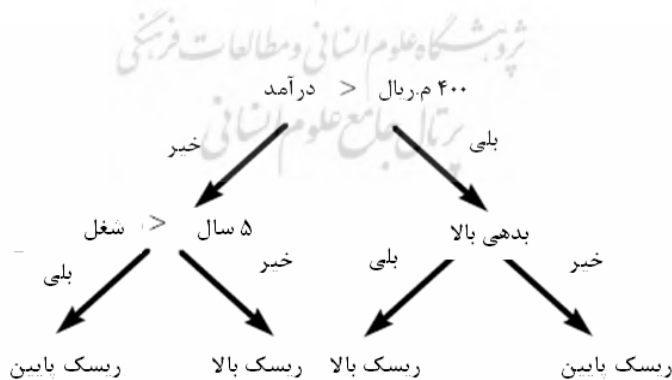
1. Weiss and Indurkha  
 2. Berry and Linoff  
 3. Hand, Mannila and Smyth  
 4. Daniel and Larose  
 5. Massachusetts Institute of Technology. (MIT)  
 6. Mendonca and Sunderhaft



شکل شماره ۱. فرآیند کشف دانش با رویکرد تکرارشونده

۲. درخت تصمیم

درخت تصمیم، روشی برای نمایش یک سری از قوانین است که منتهی به یک رده یا مقدار می‌شود. برای مثال، می‌خواهیم متقاضیان وام را به دارندگان ریسک اعتبار خوب و بد تقسیم کنیم. شکل شماره ۲ درخت تصمیمی را نشان می‌دهد که این مسئله را حل می‌کند و همه مؤلفه‌های اساسی یک درخت تصمیم در آن نشان داده شده است: نود تصمیم، شاخه‌ها و برگها (ادل‌اشتاین<sup>۱</sup>، ۱۹۹۹).



شکل شماره ۲. درخت تصمیم

1. Edelstein

در درخت تصمیم‌گیری تعدادی پرسش وجود دارد و با مشخص شدن پاسخ هر پرسش، سؤالی دیگر طرح می‌شود. اگر سؤاها درست و سنجیده پرسیده شوند، تعداد کمی از پرسشها، برای پیش‌بینی دسته رکورد جدید، کافی خواهد بود. عملکرد درخت تصمیم به این صورت است که یک گره ریشه در بالای آن کشیده شده و برگهای آن در پایین قرار دارند. یک رکورد در گره ریشه وارد می‌شود و در این گره یک تست (آزمون) صورت می‌گیرد تا روشن شود که این رکورد به کدام یک از گره‌های فرزند (شاخه پایین‌تر) خواهد رفت.

معمولاً روشهای متفاوت برای انتخاب این آزمون اولیه وجود دارد، ولی هدف همه آنها یکی است: "گزینش روشی که بهترین جداسازی را در کلاسهای هدف انجام دهد." این فرآیند آنقدر ادامه پیدا می‌کند تا رکورد به گره برگ برسد.

در اینجا باید اثربخشی درخت ایجاد شده اندازه‌گیری شود. برای این کار از یک مجموعه رکوردها یا داده‌های آزمایشی استفاده می‌شود که متفاوت است با داده‌های اولیه‌ای که درخت را ایجاد کرده‌اند. معیاری هم که اندازه‌گیری می‌شود، هم عبارت است از "درصد داده‌هایی که به طور صحیح دسته‌بندی می‌شوند و دسته پیش‌بینی شده با دسته واقعی آنها یکسان است." هر مسیر ایجاد شده از ریشه به یک برگ، معادل یک قاعده<sup>۱</sup> است.

درخت تصمیم دارای الگوریتم‌های متعدد است که هر کدام با شیوه‌ای متفاوت، اما با رویکردی یکسان به تشکیل درخت، می‌پردازد. برای آشنایی بیشتر با نحوه تشکیل درخت، توضیحی مختصر درباره شیوه شاخه زدن در یکی از معروف‌ترین الگوریتم‌های درخت تصمیم به نام CART ارائه شده است. الگوریتم CART را که یکی از معروف‌ترین روشهای انجام‌دادن درخت تصمیم است بریمن و همکارانش در سال ۱۹۸۴ ایجاد کردند (برایمن و همکاران<sup>۲</sup>، ۱۹۸۴).

روش CART شاخه‌های خود را به صورت دو تایی و تنها بر اساس یک فیلد (متغیر مستقل) ایجاد می‌کند. یعنی هر گروه غیر برگ آن، به دو گروه دیگر تفکیک می‌شود. نخستین کار این است که کدام یک از فیلدها بهترین شاخه را تولید می‌کند.

معیاری که برای ارزیابی شاخه‌ها به کار می‌رود، عبارت است از "گوناگونی"<sup>۳</sup>. برای محاسبه گوناگونی برای یک مجموعه از رکوردها روشهای بسیار وجود دارد. که در تمامی آنها "گوناگونی

---

1. Rule  
2. Breiman et al  
3. Diversity

زیاد" عبارت از مجموعه‌هایی است که کلاسهای گوناگونی در خود داشته باشند و "گوناگونی کم" عبارت است از مجموعه‌هایی که اعضای یک کلاس در آن بر سایر کلاسها غلبه کند و "بهترین شاخه زدن" آن است که گوناگونی در مجموعه‌ها را تا حد امکان کم کند. معیارهای محاسبه گوناگونی عبارتند از:

$$\min(P(c1), P(c2)) \\ 2 P(c1) P(c2) \\ [P(c1) \log P(c2)] + [P(c2) \log P(c1)]$$

در این روش، هدف اصلی حداکثرسازی عبارت زیر است:

$$\text{Diversity(befor split)} - ((\text{Diversity(Left child)} + \text{Diversity(Right child)}))$$

حال سعی می‌کنیم که برای هر فیلد با استفاده از یکی از فرمولهای داده شده برای محاسبه گوناگونی، حداقل مقدار گوناگونی را که می‌توان ایجاد کرد، به دست آوریم. سپس با مقایسه فرمول ارائه شده در بالا برای اندازه‌گیری گوناگونی قبل و بعد از شاخه زدن به وسیله هر یک از فیلدها بهترین فیلدی را برمی‌گزینیم که کمترین گوناگونی را در شاخه‌های حاصل ایجاد می‌کند و دو شاخه به وسیله آن می‌زنیم. در مرحله بعد دو شاخه داریم که هر کدام دارای یکسری رکورد هستند (هریک از رکوردهای گره بالاتر در یکی از شاخه‌ها قرار گرفته است). حال برای هر شاخه مانند قبل عمل می‌کنیم.

برای ارزیابی درخت ایجاد شده به روش CART یا هر روش دیگر معیارهایی وجود دارند. یکی از مهمترین و اصلی‌ترین این معیارها "نرخ خطا" در درخت است. برای محاسبه نرخ خطا در درخت ابتدا باید نرخ خطا در هر شاخه را به دست آوریم. نرخ خطا در هر برگ عبارت است از نسبت تعداد رکوردهایی که کلاس یا دسته آنها درست انتخاب و پیش‌بینی نشده است. مثلا اگر در یک برگ ۱۰ رکورد وجود داشته باشد و برای این رکوردها کلاس A پیش‌بینی شده باشد، در حالی که تنها ۸ عدد از این رکوردها واقعا دارای کلاس A باشند و دوتای آنها در کلاس دیگر باشند، نرخ خطا ۰/۲۰ خواهد بود.

کیفیت درخت حاصله نیز مهم است. مثلا اگر هدف پیش‌بینی قد افراد باشد که دو کلاس کوتاه و بلند برای افراد وجود داشته باشد، فرض کنید یک مجموعه ۱۱ نفری وجود دارد که همگی کوتاه قد هستند بجز کامران. اگر این گره را به دو شاخه تقسیم کنیم ممکن است قاعده‌ای مانند زیر ایجاد شود:

"افراد دارای کمتر از ۲۸ سال که نام آنها کامران است، بلند قد هستند"

این شاخه زدن با آنکه نرخ خطای درخت را برای مجموعه آموزشی کاهش می‌دهد، سبب ایجاد یک قاعده بدون کیفیت مانند قانون بالا می‌شود.

برای جلوگیری از ایجاد چنین قاعده‌هایی در بعضی از شاخه‌ها که چنین شرایطی وجود دارد، هرس یا برش<sup>۱</sup> صورت می‌گیرد. این کار با آنکه نرخ خطا را افزایش می‌دهد، از ایجاد بعضی قاعده‌های ناکارا جلوگیری می‌کند.

پس از توضیح چگونگی شیوه دسته‌بندی روش CART باید به این نکته اشاره کرد که روشهای دیگر درخت تصمیم مانند C5.0 و Chaid نیز برای دسته بندی، ساختار تقریباً مشابهی دارند و هدف همه آنها به دست آوردن درختی با کیفیت بالا و نرخ خطای کم برای دسته‌بندی داده‌هاست و بیشتر تفاوتها در شیوه شاخه زدن و برش شاخه‌هاست.

### ۳. گزینش و انتخاب کارکنان

(بردول و رایت<sup>۲</sup>، ۲۰۰۴) خاطر نشان می‌کنند که کارمندیابی<sup>۳</sup> و گزینش<sup>۴</sup> فرآیندهایی معطوف به شناسایی و جذب افراد مناسب برای رودرویی با نیازهای منابع انسانی یک سازمان است. گزینش نمایانگر مرحله نهایی تصمیم‌سازی در فرآیند استخدام به شمار می‌رود (کولینگ و مایر<sup>۵</sup>، ۱۹۹۰).

کارمندیابی و گزینش، گذرگاهی برای به کارگیری افراد و ایجاد منابع انسانی در سازمان است. محیط تجاری که به طور فزاینده رقابتی و جهانی است، ضمن همراهی با نیاز به کیفیت و خدمت‌رسانی به مشتری، اهمیت کارمندیابی و گزینش افراد مناسب را افزایش داده و به عنوان "کارفرمای انتخاب"<sup>۶</sup> مورد ملاحظه قرار گرفته است (پورتر، اسمیت و فاگ<sup>۷</sup>، ۲۰۰۶).

آنچه در این مقاله مدنظر بوده، کمک به "گزینش" شایسته‌ترین افراد با استفاده از داده‌های گذشته است. لذا فرآیند "استخدام" که پس از این مرحله خواهد بود، در این فرآیند مورد نظر نیست.

- 
1. Pruning
  2. Beardwell & Wright
  3. Recruitment
  4. Selection
  5. Cowling and Mailer
  6. Employer of Choice
  7. Porter, Smith and Fagg

#### ۴. دقت گزینش<sup>۱</sup>

به طور کلی، "دقت گزینش" بر حسب درجه تطابق میان پیشگویی کننده و کارآیی شغلی آینده تعیین می‌شود و میزان صحت و اعتبار روش انتخاب شده را در بر می‌گیرد. جدول زیر که مطابق تحقیقات انجام شده درباره روشهای متفاوت گزینش و اندازه‌گیری همبستگی عامل مشترک میان کارآیی شغلی پیش‌بینی شده و واقعی آنهاست، بیانگر میزان دقت هر یک از این روشها و به نوعی میزان استفاده کارفرمایان از هر کدام از آنهاست (اسمیت<sup>۲</sup>، ۲۰۰۲).

جدول شماره ۱. مقایسه دقت روشهای متفاوت گزینش (عدد ۱ معادل بهترین پیشگویی است)

دقت	روشهای گزینش
۰٫۷۲	ارزیابی میانی (ترفیع) <sup>۳</sup>
۰٫۶۳	آزمونهای هوش و مصاحبه های ساخت یافته <sup>۴</sup>
۰٫۶۲	آزمونهای هوش و نمونه کار <sup>۵</sup>
۰٫۵۶	آزمونهای قابلیت <sup>۶</sup>
۰٫۵۲	آزمونهای نمونه کار <sup>۷</sup>
۰٫۵۱	آزمونهای هوش
۰٫۵۱	مصاحبه های ساخت یافته <sup>۸</sup>
۰٫۴۰	آزمونهای شخصیت
۰٫۳۵	سوابق شغلی و تحصیلی (رزومه)
۰٫۲۶	مصاحبه های نوعی <sup>۹</sup>
۰٫۲۶	مقایسه
۰٫۱۸	تعداد سالهای تجربه کاری
۰٫۱۰	مدت زمان تحصیل
۰٫۰۰	پیشگویی تصادفی

البته نوع شغل بیشترین تأثیر را در انتخاب هر یک از این روشهای گزینش برای پر کردن مشاغل خالی دارد. به عنوان نمونه، روش ارزیابی میانی (ترفیع) معمولاً برای سمتهای مدیریتی و رتبه‌ای مورد استفاده قرار می‌گیرد.

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



## پیشینه کاربرد داده‌کاوی در مدیریت منابع انسانی

داده‌کاوی را می‌توان در شناسایی ریشه‌ها و عوامل برخی از مشکلات منابع انسانی سازمانها به کار برد. در این زمینه متاسفانه در ایران هیچ پژوهشی انجام نشده است. در حالی که در سایر کشورها، شرکتها و افرادی درباره این موضوع پیشنهادهایی داده‌اند و کارهایی نیز انجام شده است. هافمن<sup>۱</sup> (1998)، شروک<sup>۲</sup> (1999)، هوویدی<sup>۳</sup> (2002) و اولسان<sup>۴</sup> (2002) هریک جداگانه گزارشهایی در مورد کاربرد داده‌کاوی در مدیریت منابع انسانی ارائه کرده‌اند.

داده‌کاوی برای تشخیص کارکنان غیر سودمند و خطا کار به منظور اخراج از شرکت، کشف کارکنان با استعداد و استفاده از آنها در واحدهای تجاری دارای اولویت بالا و سوددهی زیاد، مشخص کردن کارکنان درستکار و تشویق آنها، کشف استعدادها در زمینه نیازهای ملی و بررسی رقبا پیشنهاد شده است (شرکت Spss، ۲۰۰۴).

مشاور شرکت بیمه "متلایف" برای تحقیقات واحدهای بازرسی ویژه، از قواعد "اگر ... آنگاه ..." استخراج شده از میان داده‌های پیشین، استفاده کرده است و نمونه‌ای از نتایج این تحقیق را که به حوزه مدیریت منابع انسانی مرتبط می‌شود، ارائه نموده است (دنکر<sup>۵</sup>، ۲۰۰۰).

همچنین از داده‌کاوی به منظور کاهش زمان، کاهش هزینه و گسترش دستیابی به داوطلبان دارای بالاترین کیفیت در استخدام نیروهای جدید استفاده شده است (موسسه SAS<sup>۶</sup>، ۲۰۰۱).

در این میان، بهترین تحقیقی که برای انتخاب کردن کارکنان با استفاده از رویکرد داده‌کاوی انجام شده، پژوهشی است با عنوان "داده‌کاوی برای گزینش بهتر کارکنان و افزایش سرمایه انسانی" که "چین" و "چن" ارائه کرده و مطالعه موردی آن در صنایع پیشرفته انجام گرفته است. در همین زمینه مدلی نیز برای انتخاب کردن کارکنان از طریق داده‌کاوی ارائه شده است. این پژوهش با هدف

---

1. Hoffmann  
2. Schroeck  
3. Schroeck  
4. Olson  
5. Denker  
6. SAS Institute Inc

ایجاد دانشی به منظور جذب و گزینش بهتر کارکنان در صنایع الکترونیک صورت پذیرفته است (چین و چن<sup>۱</sup>، ۲۰۰۷).

در مجموع آنچه از مطالعات پیشین قابل استحصال است، در زمینه روش مورد استفاده محققان حوزه داده‌کاوی و منابع انسانی است که همگی از روش "درخت تصمیم" برای این منظور استفاده کرده‌اند. اما نوع الگوریتم مورد استفاده در آنها متفاوت است.

### روش فرآیند تحقیق

در این مقاله سعی بر این است که با استفاده از فنون داده‌کاوی، قواعد و روابط میان نمرات آزمونهای ورودی و وضعیت کارکنان با عملکرد شغلی و وضعیت ارتقای آنان شناسایی شود. ابتدا با مطالعه و بررسی پایگاههای داده آزمون و اطلاعات شغلی و شخصی پذیرفته شدگان یکی از بانکهای تجاری ایران، شاخصهای منابع انسانی که بر عملکرد یا ارتقا آنها مؤثر و داده‌های آنها نیز موجود و پوشا بودند، شناسایی شد. در گام بعدی نیز باتوجه به مطالعات انجام شده پیشین درباره داده‌کاوی و عملکرد منابع انسانی، مدل مفهومی مناسب برای اجرای داده‌کاوی طراحی شد. در این زمینه کوشیدند تا ضمن الگو برداری مناسب از مطالعات گذشته، مشکلات و نواقص مدل‌های پیشین نیز برطرف شود.

سپس انبار داده تشکیل شد. بر اساس فنون داده‌کاوی، داده‌ها پیش پردازش و پاکسازی شدند، یعنی 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 مرحله بعد با فنون درخت تصمیم، داده‌ها دسته‌بندی و تجزیه و تحلیل شدند و نتایج اولیه به دست آمد. در نهایت از میان قواعد به دست آمده، قواعد غیربدهی مشخص و علت وجود این قواعد تبیین شد. در این تحقیق که بر پایه روش‌شناسی داده‌کاوی صورت گرفته است، مجموعه داده‌ها با استفاده از نرم افزار "اکسل" ۲ آماده‌سازی و قواعد نیز با استفاده از نرم افزار "کلمنتاین" ۳ (ویرایش ۱۲) استخراج شده است. در نهایت پس از تایید روایی و پایایی قواعد استخراجی، مدل

- 
1. Chien and Chen
  2. Excel
  3. SPSS Clementine

کشف دانش و به کارگیری آن برای گزینش و استخدام داوطلبان آزمونهای ورودی با استفاده از داده‌کاوی، ارائه شده است.

### جامعه آماری و روش گردآوری اطلاعات

جامعه آماری این تحقیق که به منظور آزمایش مدل طراحی شده، مورد استفاده قرار گرفته، عبارت است از:

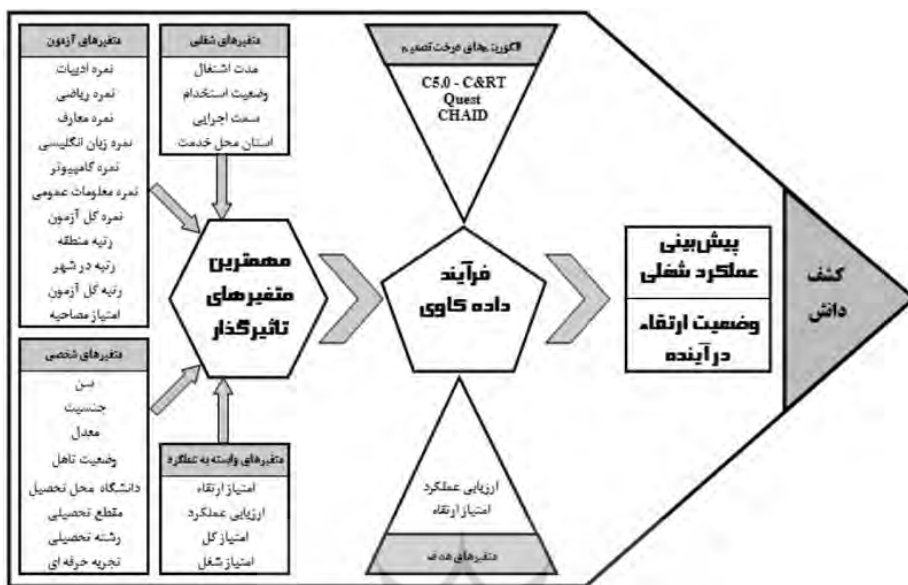
"پذیرفته‌شدگان و استخدام‌شدگان آزمون ورودی سراسری سالهای ۱۳۸۳ و ۱۳۸۴ یکی از بانکهای تجاری در ایران"

به دلیل نیاز به سنجش ارتباط نمرات آزمونها با عملکرد و ارتقای شغلی شرکت‌کنندگان، عده‌ای از این جامعه که در نهایت به استخدام بانک درآمده و حداقل یک سال از فعالیت آنها در بانک گذشته است، انتخاب شده‌اند. به طوری که امکان اخذ نمره ارزیابی عملکرد شغلی و امتیاز ارتقا برای این افراد امکان پذیر بوده است.

با توجه به این توضیحات، عده پذیرفته‌شدگان و استخدام‌شدگان طی سالهای ۱۳۸۳ و ۱۳۸۴ بانک، بالغ بر ۹۴۰ نفر بوده است. با توجه به در دسترس نبودن نمرات آزمون ورودی برای تمام این جامعه، در نهایت نمونه‌ای ۷۱۱ نفری که مجموعاً دارای ۲۶ فیلد اطلاعاتی مفید، صحیح و کامل از اطلاعات شخصی، وضعیت استخدامی، نمرات آزمون و عملکرد شغلی بوده‌اند به عنوان جامعه آماری نهایی انتخاب شده‌اند.

### مدل مفهومی

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



شکل شماره ۳. مدل مفهومی کشف دانش از بانک داده آزمون ورودی داوطلبان

همان گونه که از مدل مشخص است، در گام نخست، اقدام به تعریف گروه‌های متغیر و شناسایی مهم‌ترین و با اهمیت‌ترین متغیرهای تأثیرگذار خواهد شد، سپس با ورود این متغیرها و همچنین متغیرهای هدف انتخاب شده (ارزیابی عملکرد و امتیاز ارتقا) به تناسب هر الگوریتم، فرآیند داده‌کاوی اجرا خواهد شد و نتایج آن به منظور پیش‌بینی عملکرد شغلی و همچنین وضعیت ارتقای داوطلبان در آینده به کار خواهد رفت. نتایج نهایی حاصل شده نیز به عنوان دانش سازمانی و به منظور افزایش کارایی و اثربخشی در مجموعه سازمان، مورد استفاده قرار خواهد گرفت.

### تعریف متغیرها

به منظور سهولت بخشیدن به تشکیل بانک داده‌ها و انجام دادن مراحل داده‌کاوی، متغیرهای مستقل در ۴ دسته تقسیم‌بندی شدند که عبارتند از:

۱. **متغیرهای شخصی** (شامل: سن، معدل، تجربه حرفه‌ای، وضعیت تأهل، جنسیت، مقطع تحصیلی، رشته تحصیلی و دانشگاه محل تحصیل)
۲. **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)

۳. **متغیرهای شغلی** (شامل: مدت اشتغال، وضعیت استخدام، سمت اجرایی و استان محل خدمت)

۴. **متغیرهای وابسته به عملکرد** (شامل: امتیاز ارتقا، ارزیابی عملکرد، امتیاز شغل و امتیاز کل)

در ادامه تعریف متغیرهای مستقلی که نیازمند توضیح بیشتر هستند، ارائه شده است:

✦ **تجربه حرفه‌ای:** متغیری است که نشان‌دهنده کل تجربه شغلی کارمند از شروع فعالیت حرفه‌ای وی است (چه در بانک و چه در سایر شرکتها یا مؤسسات) و با واحد "سال" نمایش داده شده است. این متغیر، عددی میان ۳ تا ۱۰ سال را شامل می‌شود و قطعاً بزرگتر یا مساوی متغیر "مدت اشتغال" خواهد بود.

✦ **دانشگاه محل تحصیل:** آخرین دانشگاه محل تحصیل داوطلبان در ۵ گروه دسته‌بندی شده است که عبارتند از دانشگاههای سراسری تهران، دانشگاههای سراسری شهرستانها، دانشگاههای آزاد تهران، دانشگاههای آزاد شهرستان و دانشگاههای پیام نور

✦ **نمره نرمال ادبیات، ریاضی، معارف، زبان، کامپیوتر و معلومات عمومی:** عبارت است از نمره نرمال شده درس ادبیات در آزمون ورودی کتبی که پس از محاسبه تعداد پاسخهای صحیح، غلط و نزده و بر مبنای کسر یک پاسخ صحیح به ازای هر سه پاسخ غلط، محاسبه شده است.

✦ **نمره کل آزمون:** عبارت است از نمره نهایی آزمون کتبی ورودی که باتوجه به نمره نرمال و ضریب اهمیت دروس ششگانه (ادبیات، ریاضی، معارف اسلامی، زبان انگلیسی، کامپیوتر و معلومات عمومی) محاسبه شده است.

✦ **رتبه منطقه:** رتبه داوطلب در منطقه<sup>۱</sup> انتخابی که پس از مرتب کردن نزولی نمره کل آزمون تمامی داوطلبان آن منطقه تعیین شده است.

✦ **رتبه کل آزمون:** رتبه داوطلب در کل کشور و در همان سال که پس از مرتب کردن نزولی نمره کل آزمون تمامی داوطلبان تعیین شده است.

✦ **رتبه در شهر:** رتبه داوطلب در شهر انتخابی که پس از مرتب کردن نزولی نمره کل آزمون تمامی داوطلبان آن شهر منطقه تعیین شده است.

۱. باتوجه به محدود بودن حوزه‌های برگزاری آزمون، کل کشور به چندین منطقه تقسیم شده است که هر منطقه چند شهر را در بر می‌گیرد. لذا داوطلبان شهرهای گوناگون باید برای شرکت در آزمون، به شهر منطقه خود مراجعه کنند.

✦ **امتیاز مصاحبه:** برای ۴۲۰ رکورد (داوطلب) امتیاز مصاحبه حضوری نیز ثبت شده است که امتیازی میان ۵۸ تا ۹۴ (از ۱۰۰) را در بر می‌گیرد. این مصاحبه قبل از استخدام نهایی برگزار شده است.

✦ **مدت اشتغال:** عبارت است از تفاضل سال جاری از نخستین سال استخدام قطعی داوطلب پس از پذیرفته شدن وی. کم‌ترین میزان این متغیر برابر با ۳ سال است.

✦ **وضعیت استخدام:** این متغیر از نوع "دو حالتی" و عبارت است از نوع استخدام کارمند (داوطلب) که دارای دو حالت رسمی و قراردادی است.

✦ **سمت اجرایی:** جدا از عنوان سمت هریک از کارکنان، هر شغل به سه دسته اجرایی، تخصصی و مدیریتی تقسیم می‌شود که این متغیر معرف نوع سمت فعلی کارمند در هریک از این حالات است.

✦ **استان محل خدمت:** این متغیر معرف استان محل خدمت (پذیرفته شدن) داوطلب است که مجموعاً ۲۷ استان از ۳۰ استان کشور را در بر گرفته است. ضمن آنکه استانهای خراسان رضوی، شمالی و جنوبی در این تقسیم‌بندی، با عنوان "خراسان" مشخص شده‌اند. همچنین برای تحلیل کردن بهتر نتایج، استانهای کشور به شش دسته (مطابق جدول زیر) تقسیم شده‌اند.

جدول شماره ۲. تقسیم‌بندی استان محل خدمت (آزمون)

گروه استانی	عنوان در جدول داده‌ها	جمعیت (هزار نفر)	تعداد شعب	نام استانهای گروه
استان تهران	Tehran S.	۱۳۴۲۲	۵۲۲	تهران
استانهای شمالی	North S.	۹۳۱۶	۲۷۹	اردبیل، قزوین، گلستان، گیلان، مازندران
استانهای غربی	West S.	۱۵۵۸۵	۴۶۴	آذربایجان شرقی، آذربایجان غربی، ایلام، چهارمحال و بختیاری، زنجان، کردستان، کرمانشاه، لرستان، همدان
استانهای جنوبی	South S.	۶۵۶۵	۲۶۹	بوشهر، خوزستان، هرمزگان
استانهای شرقی	East S.	۱۲۰۹۹	۳۴۲	خراسان جنوبی، خراسان رضوی، خراسان شمالی، سیستان و بلوچستان، کرمان
استانهای مرکزی	Central S.	۱۳۵۰۹	۵۳۴	اصفهان، سمنان، فارس، قم، مرکزی، یزد، کهگیلویه و بویر احمد

✦ **امتیاز کل:** کل امتیازی که هر کارمند کسب می‌کند، با توجه به امتیاز شغل، مدت اشتغال، سابقه حرفه‌ای، ارزیابی عملکرد و سمت اجرایی وی محاسبه می‌شود که مبنای پرداخت حقوق ماهانه وی خواهد بود. این متغیر، عددی میان ۲۸۱۸ تا ۵۰۸۹ را شامل می‌شود.

✦ **امتیاز شغل:** عبارت است از امتیاز پایه‌ای که به هر سمت تعلق می‌گیرد و هر میزان که بیشتر باشد، اهمیت آن شغل نیز بیشتر خواهد بود. این متغیر، عددی میان ۱۸۰۰ تا ۳۱۰۰ را شامل می‌شود.

### تعریف متغیرهای هدف

باتوجه به مدل مفهومی ارائه شده، متغیرهای هدف در این پژوهش، عبارتند از: ارزیابی عملکرد و امتیاز ارتقا.

✦ **امتیاز ارتقا:** باتوجه به توانمندی کارکنان در ارتقای شغل خود در مدت اشتغال، این متغیر عبارت است از تفاضل امتیاز شغل سمت فعلی از امتیاز شغل نخستین 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 میان داوطلبان که به این متغیر تعلق گرفته، برابر ۱۱۵۰ است. با این وصف، و باتوجه به اینکه این متغیر به عنوان یکی از متغیرهای هدف تعریف شده است، امتیاز محاسبه شده در ۴ دسته طبقه‌بندی شده که در جدول زیر آمده است.

جدول شماره ۳. گزینه‌ها و طبقه‌بندی امتیاز ارتقا

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

✦ **ارزیابی عملکرد:** این متغیر بر مبنای آیین‌نامه مشخصی در بانک، حداقل سالی ۲ بار برای هر کارمند و توسط مدیر مستقیم وی تکمیل و عدد آن بر مبنای ۱۰۰ رایجه می‌شود. جهت سهولت اجرای داده‌کاوی بر اساس متغیر وابسته کلیه اعداد رایجه شده تبدیل به شاخص کیفی گردیده‌اند که مطابق جدول زیر در ۷ دسته طبقه‌بندی شده است.

جدول شماره ۴. گزینه‌ها و طبقه‌بندی امتیاز ارزیابی عملکرد

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

## مطالعه موردی

### ۱. آماده سازی داده‌ها

آماده‌سازی داده‌ها شامل کلیه مراحل است که رکوردها و متغیرها را برای ساخت مدل و درخت، آماده می‌نماید. معمولاً این گام از فرآیند داده‌کاوی و کشف دانش، طولانی‌ترین و مهمترین مرحله محسوب می‌شود.

به منظور آماده‌سازی داده‌ها، ابتدا با برگزاری جلسات توجیهی با مدیران و مسئولان منابع انسانی موسسه مورد مطالعه، هدف از انجام پژوهش تشریح و کلیه داده‌ها و اطلاعات موجود در رابطه با پرسنل و کارکنان ارزیابی گردید. طبق آمار مدیریت منابع انسانی بانک، از میان شرکت کنندگان در آزمون سالهای ۸۳ و ۸۴، حدوداً ۹۷۰ نفر پذیرفته و استخدام شده‌اند که با توجه به پراکندگی و مجزا بودن بانک داده‌های آزمون و بانک داده‌های پرسنلی و پس از انطباق، نهایتاً ۷۱۷ رکورد با ۳۷ فیلد مطابقت داده شدند. از آنجاییکه هر فیلد نمایانگر یک متغیر می‌باشد لذا در اولین گام پس از تشکیل بانک داده‌ها، اقدام به حذف و یا شناسایی متغیرهای مشابه و ترکیب و تبدیل آنها به یکدیگر گردید. طی این اقدام، ۳۷ فیلد موجود، تبدیل به ۲۶ فیلد نهایی گردیدند که بیشتر به آنها اشاره شد. در مرحله بعد، به بررسی و بازنگری مقادیر (کمی و کیفی) هر متغیر پرداخته شد که با ترکیب، تلفیق، تعریف دسته و طبقه، حذف و کاهش متغیرها همراه بوده است؛ به طوری که در نهایت تعداد کل رکوردهای سالم به ۷۱۱ مورد رسید.

### ۲. تشکیل درخت تصمیم

باتوجه به شناسایی چهار دسته متغیر (شخصی، آزمون، شغلی و وابسته به عملکرد) و همچنین وجود دو نوع متغیر هدف (ارزیابی عملکرد و امتیاز ارتقا) و وجود الگوریتمهای متفاوت برای تشکیل دادن درخت تصمیم، ابتدا پرداخته شد به بررسی همه مدلهای ممکن که امکان تشکیل آنها با بانک داده‌های موجود، میسر است. 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 "دنباله دار" ۱ و "سفرشی" ۲، امکان تشکیل ۱۸۰ درخت متفاوت به صورت رابطه (۱) شناسایی شد:

1. Towing
2. Ordered



$$\text{رابطه (۱)} \quad (2^x - 1) * y * A$$

که در آن،  $x$  تعداد دسته‌های متغیرهای مستقل،  $y$  تعداد متغیر هدف و  $A$  تعداد الگوریتم است، در نتیجه:

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

باتوجه به اینکه 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 "کلمتاین" آموزش داده شدند و دقت هر کدام محاسبه و ثبت شد که به تفکیک هر متغیر هدف در جدول زیر ارایه شده است. شایان ذکر است که در ۷ حالت شناسایی شده با الگوریتم QUEST، امکان تشکیل درخت وجود نداشت.

جدول شماره ۵. میزان دقت هر متغیر هدف بر مبنای نتایج حاصله از 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 ارتقا	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	

### ۳. انتخاب درختهای مناسب برای استخراج قواعد

باتوجه به کثرت درختهای تولید شده، بدیهی است که قواعد حاصل شده از این درختها نیز بسیار زیاد خواهد شد. به منظور نتیجه‌گیری مناسب از قواعد حاصل شده، نیاز است تا درختهای مناسب از سایر مدل‌های تولید شده، منفک گردند. به منظور گزینش میان درختها، ابتدا با نظر کارشناسی و تعیین معیارهای کمی و اعمال شرایطی در نوع متغیرهای مستقل، درختهایی با دقت بالای ۷۰ درصد انتخاب شدند و در نهایت ۱۷ درخت اولیه را برای مرحله آزمون برگزیدند.

از میان درختهای منتخب، ۳ درخت با متغیر هدف "ارزیابی عملکرد" وجود داشتند که الگوریتم مورد استفاده در هر سه آنها، C5.0 بود. همچنین ۱۴ درخت با متغیر هدف "امتیاز ارتقا" متشکل از تمام الگوریتمهای موجود نیز انتخاب شدند.

باتوجه به اینکه پیشتر ۱۶۱ رکورد از کل داده‌های موجود بانک داده‌های نهایی، به عنوان داده آزمایشی و به صورت تصادفی انتخاب و برای مرحله آزمون کنار گذاشته شدند؛ در مرحله بعد هر ۱۷ درخت انتخاب شدند و با ورود داده‌های آزمایشی، مورد آزمون قرار گرفتند و دقت آنها تعیین گردید.

مطابق جدول زیر، دقت حاصل از سه درخت تولید شده با متغیر هدف "ارزیابی عملکرد"، بسیار کاهش یافت، لذا این سه درخت و در نتیجه متغیر هدف "ارزیابی عملکرد" قبل از ورود به مرحله استخراج قواعد حذف شدند.

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

کد درخت	میزان دقت با داده‌های آموزشی	میزان دقت با داده آزمایشی	تفاضل 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	٪۳۱٫۷۴	٪۵۲٫۲۶

در نهایت مدل‌های نهایی انتخاب شده برای استخراج قواعد به ۱۴ عدد رسید که باتوجه به یکسانی مدل و قواعد تولید شده دو الگوریتم CART (Ordered) و CART (Towing) صرفاً یکی از آنها در نظر گرفته شده که نتایج آزمون آنها در جدول زیر خلاصه شده است.

جدول شماره ۷. میزان دقت هر الگوریتم بر مبنای نتایج حاصله از 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

جمع‌بندی تحلیلهای فوق حاکی از این است که در مجموع درختهای منتخب، متغیر هدف "امتیاز ارتقا" و الگوریتم CART بیشترین دقت را حاصل کرده‌اند. ضمن آنکه متغیر هدف "ارزیابی عملکرد" در هیچ کدام از مدل‌های نهایی مشاهده نمی‌شود و مدل‌های ساخته شده آن از خطای بسیار بالا برخوردارند.

#### ۴. استخراج و انتخاب قواعد

پس از انتخاب نهایی درختها، نوبت به استخراج قواعد از آنها می‌رسد. منظور از استخراج قواعد، بررسی تک‌تک درختهای تشکیل شده و تعریف هر قاعده از زبان ریاضی و منطقی به زبان قابل فهم و قابل برداشت برای خبرگان است. همچنین، هر قاعده تعیین شده دارای احتمال وقوعی خواهد بود که براساس تعداد رکورد منطبق با قاعده تعریف شده از میان کل رکوردهای آن شاخه به دست می‌آید. رابطه (۲) احتمال وقوع را نشان می‌دهد.

$$P = n_c / n \quad \text{رابطه (۲)}$$

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

از میان ۸۹ قاعده استخراجی و باتوجه به تکرار قواعد در بعضی از درختها، نهایتاً ۶۸ قاعده از فرآیند داده‌کاوی به دست آمد. همچنین متغیرهایی که با توجه به قواعد استخراج شده، دارای همبستگی با متغیر هدف (امتیاز ارتقا) بوده‌اند عبارتند 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 رابطه‌ای معنادار با امتیاز ارتقا نداشته‌اند. درنهایت و در میان ۶۸ قاعده تولید شده، فقط ۱۰ متغیر، تاثیرگذار بوده‌اند.

نمونه‌ای از این قواعد که برای تحلیل و استخراج دانش به خبرگان ارائه شده، به شرح جدول

زیر است:

## جدول شماره ۷. نمونه‌ای از قواعد انتخاب شده نهایی

شرح قاعده استخراج شده	احتمال (P)	کد قاعده
داوطلبانی که قبل از استخدام تجربه کاری داشته‌اند و مدرک کارشناسی دارند و نمره معلومات عمومی آنها بیشتر از ۵۹۵۳ است، در صورتی که مجرد باشند، وضعیت ارتقای آنها پس از ۳ سال، عالی (A) است.	٪۵۰,۰۰	R03
داوطلبانی که قبل از استخدام تجربه کاری داشته‌اند و مدرک کارشناسی دارند و نمره معلومات عمومی آنها بیشتر از ۵۹۵۳ است در صورتی که متاهل باشند، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۸۷,۵۰	R04
داوطلبانی که نمره کل آزمون و نمره مصاحبه آنها به ترتیب کمتر یا مساوی ۶۴۴۹,۵ و ۸۷,۵ باشد، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۸۲,۰۰	R12
داوطلبانی که نمره کل آزمون آنها بیشتر از ۶۴۴۹,۵ ولی امتیاز مصاحبه آنها کمتر یا مساوی ۸۴,۵ باشد، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۶۳,۵۰	R15
داوطلبانی که قبل از استخدام تجربه حرفه‌ای نداشته‌اند و مدرک کارشناسی دارند، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۸۹,۷۰	R27
داوطلبانی که قبل از استخدام تجربه کاری خیلی کمی داشته‌اند و نمره کل آزمون آنها کمتر یا مساوی ۶۴۴۹,۵ است، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۸۴,۵۰	R36
داوطلبانی که امتیاز مصاحبه آنها کمتر یا مساوی ۸۴,۵ و نمره کل آزمون آنها بیشتر از ۶۴۴۹,۵ باشد، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۶۷,۶۰	R40
داوطلبانی که امتیاز شغل آنها در حال حاضر بیشتر از ۲۰۷۵ است و دارای مدرک تحصیلی دیپلم یا کاردانی هستند، وضعیت ارتقای آنها پس از ۳ سال، خوب (B) است.	٪۱۰۰,۰۰	R53
داوطلبانی که دارای مدرک کارشناسی هستند، پس از ۳ سال، بدون ارتقا (D) هستند.	٪۸۳,۲۰	R60

همچنین متغیرهایی که با توجه به قواعد انتخاب شده نهایی، دارای همبستگی با متغیر هدف (امتیاز ارتقا) بوده‌اند به ۵ متغیر کاهش یافتند که عبارتند از:

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

## دانش حاصل از قواعد نهایی و مراحل داده‌کاوی

باتوجه به اینکه قواعد استخراجی به شکل خروجیهای مطرح شده، قابل استفاده نیستند؛ با ارائه آنها به کارشناسان و خبرگان حوزه منابع انسانی، ضمن تعیین قواعد بی‌معنا (که صرفاً ناشی از تصادف بوده است)، دانش حاصل در سایر قواعد مشخص شد که نمونه‌ای از آن ارائه شده است.

- تفسیر قاعده‌های R03 و R04:

در تفسیر این قاعده، خط مشی‌ها و استراتژیهای منابع انسانی سازمان بسیار تاثیرگذار است. اگر سازمان در چرخه رشد قرار دارد و در پی نوآوری در ارائه خدمات است، طبیعی است که این گروه از افراد، خلاق و خطرپذیر باشند و مایل به اینکه مسیر ترقی شغلی را به صورت جهشی ببینند. بنابراین، سازمان باید برای استخدام افراد با برنامه حرکت کند و نوع کیفیت منابع انسانی موردنیاز خود را براساس استراتژیهای حوزه مدیریت منابع انسانی تعیین نماید. همچنین تفاوت قاعده ۴ با قاعده ۳ در این است که افراد متأهل، کمتر از افراد مجرد، انگیزه و میل به ارتقا دارند.

### ● تفسیر قاعده R27:

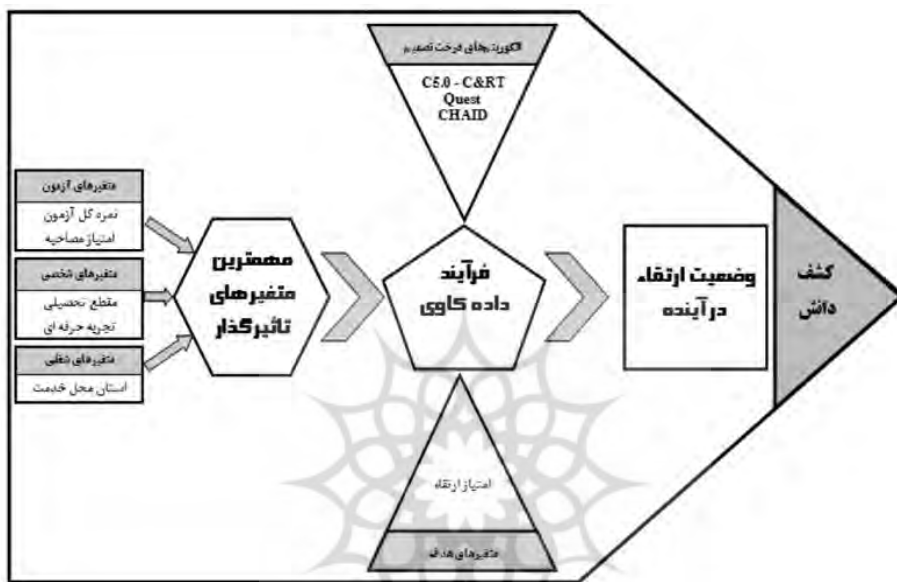
در تفسیر نتیجه مزبور، می‌توان گفت که سازمان باید، قبل از هرگونه جذب نیروی جدید، نسبت به تحلیل بافت منابع انسانی موجود خود اقدام کند. به این معنی که اولاً زنجیره مشاغل به درستی تعریف شود، ثانياً مشخص شود که در چه مشاغلی کمبود منابع انسانی وجود دارد؟ و آیا این کمبودها را می‌توان از میان نیروهای موجود تأمین کرد؟

پس از تحلیلهای فوق، نخستین عامل مهم در تحلیل منابع انسانی، وضعیت ارتقا و انتصاب کارکنان است. لذا باید تعیین شود که در حال حاضر چند درصد مشاغل سازمان بلا تصدی هستند؟ آیا مسیر ترقی شغلی به صورت اثربخش و صحیح تدوین و طراحی شده است؟ شایان ذکر است که تعمیم چنین نتایجی در حوزه آموزش و پرورش نیز امکان‌پذیر است. به عنوان نمونه اگر انتخاب شاخصها براساس نیازهای استخدامی معلمان و همچنین دانش‌آموزان صورت گیرد، می‌توان الگوهای مناسب را از طریق این روش استخراج نمود.

### نتیجه‌گیری و پیشنهادات

به کارگیری صحیح و علمی کارکنانی که از طریق آزمونهای ورودی مؤسسات و شرکتهای انتخاب می‌شوند و به ویژه در سازمانهای خدمات - محور، دارای اهمیت خاص است که استفاده از علم داده‌کاوی و کشف دانش نهفته در این زمینه بسیار مؤثر خواهد بود. در این مقاله ضمن بیان رویکردهای پیشین به حوزه استفاده از داده‌کاوی در مدیریت منابع انسانی، تلاش شد تا از این فن در گزینش داوطلبان آزمونهای ورودی بهره گرفته شود. همچنین باتوجه به رویکرد مفهومی در زمینه متغیرهای تاثیرگذار و متغیرهای هدف، مشخص گردید که از میان ۲۶ متغیر تاثیرگذار فقط پنج متغیر "استان محل خدمت، مقطع تحصیلی، نمره کل آزمون، امتیاز مصاحبه و تجربه حرفه‌ای" بر متغیر هدف "امتیاز ارتقا" تاثیرگذار هستند.

به منظور اجرای مدل مفهومی و امکان بررسی آن، با عنایت به نتایج حاصل از خروجیهای داده‌کاوی، مدل اولیه به شرح شکل (۴) عملیاتی شده است. از آنجایی که در قواعد نهایی، فقط ۵ متغیر مستقل حضور داشتند و همچنین باتوجه به حذف متغیر هدف "ارزیابی عملکرد" در مراحل داده‌کاوی، این مسئله در مدل نهایی، لحاظ شده است.



شکل شماره ۴. مدل نهایی کشف دانش از بانک داده آزمون ورودی داوطلبان

عدم وجود متغیر هدف "ارزیابی عملکرد" در میان قواعد نهایی، ناشی از بی‌تأثیر بودن این متغیر در نتایج حاصل شده است. در حقیقت این متغیر در مراحل قبل از استخراج قواعد، حذف شد که به اتفاق نظر همه خبرگان منابع انسانی، ناشی از عدم اجرای صحیح نظام ارزیابی عملکرد در مؤسسه مورد مطالعه است.

نتایج به دست آمده در این پژوهش و روش ارائه شده برای کشف دانش، قابل تعمیم به سایر حوزه‌هایی است که نقش سرمایه انسانی در آنها حیاتی است. از جمله این حوزه‌ها می‌توان به حوزه "آموزش و پرورش" اشاره کرد که با توجه به اهمیت والای تعلیم و تربیت، محقق ساختن آن در گرو گزینش صحیح "معلمان" است. از مهم‌ترین ابزارهای گزینش و استخدام معلمان در کشور ما،

آزمونهای ادواری است که وزارت آموزش و پرورش برگزار می‌کند. در نتیجه، شناخت شاخصهای مؤثر بر گزینش معلمها و اساتید از طریق آزمونهای استخدامی ادوار گذشته، برای استفاده در گزینشهای بهتر در آینده از این طریق امکان‌پذیر خواهد بود.

به همین دلیل، شرط نخست، دسترسی به بانک آزمونهای استخدامی گذشته معلمان است. در گام بعد مطابق با روش ارائه شده در این مقاله، ضمن تعریف متغیرهای مناسب برای انجام داده‌کاوی، متغیر هدفی مناسب مانند "میانگین نمرات دانش‌آموزان هر معلم در طول دوران تدریس" را می‌توان انتخاب کرد و سایر مراحل را مطابق با فنون داده‌کاوی ادامه داد. در نتیجه خبرگان حوزه آموزش و پرورش را خواهند یافت.

در نهایت، آنچه بیشتر حایز اهمیت است، استفاده بهینه و گسترده از علم داده‌کاوی در سایر حوزه‌های مدیریتی است که می‌تواند ضمن استفاده کم هزینه از داده‌های موجود در سازمانها، افقی جدیدی از دانشهای نهفته در سازمان را برای مدیران روشن کند. همچنین رویکرد داده‌کاوی، می‌تواند یکی از رویکردهای بسیار مفید در ارزیابی داوطلبان و پذیرفته‌شدگان در آزمونهای ورودی دانشگاههای کشور در مقاطع تحصیلی متفاوت باشد. افزون بر آن، از این فن می‌توان در تحلیل وضعیت تحصیلی یا انتخاب رشته در دوران متوسطه آموزش و پرورش نیز استفاده کرد و در سایر حوزه‌های تعلیم و تربیت و سنجش برای بررسی متغیرهای عملکردی آموزش با شاخصهای مؤثر به کار گرفت.

## منابع

- Beardwell, I, L. Holden, and T. Claydon (2004). *Human Resource Management*. Harlow: FT/Prentice Hall.
- Berry, Michael J. A, and Gordon S. Linoff (2004). *Data Mining Techniques for Marketing Sales and Customer Relationship Management*. 2 Ed. John.Wiley. & Sons. Publishing Inc, 2004.
- Breiman, L, Friedman, J. H, Olshen, R. A, and Stone, C. J, *Classification and Regression Trees*. New York: Chapman & Hall, 1984.
- Chien, Chen-Fu, and Li-Fei Chen. (2007). Data Mining To Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry. *Expert Systems With Applications* 34: Pp. 280–290.
- Cowling, A, and C. Mailer (1990). *Managing Human Resources*. 2nd Edition. London: Edward Arnold.
- Daniel, T, and John Larose (2005). *Discovery Knowledge In Data, An Introduction To Data Mining*. Wiley Interscience.
- Daniel, T, and John Larose (2006). *Data Mining Methods And Models*. Wiley & Sons Publishing Inc.
- Denker, Robert (2000). *Audites For Human Resource Applications*. Available [Http://www.Wizsoft.Com/App4-Wr.Html](http://www.Wizsoft.Com/App4-Wr.Html).
- Edelstein, Herber A (1999). *Introduction To Data Mining And Knowledge Discovery*. Two Crows Corporation.
- Hand, David, Heikki Mannila, and Padhraic Smyth (2001). *Principles of Data Mining*. The MIT Press.
- Hoffmann, Carl C (1998). *Decision Support & Data Warehousing in Human Resource*. [Http://www.Cob.Vt.Edu/Mgmt/Carlson/AHRIS/Ihrim.Ppt](http://www.Cob.Vt.Edu/Mgmt/Carlson/AHRIS/Ihrim.Ppt).
- Howeedy, Rodney (2002). *First Door Uses Web Trends to Create a Human Resource Database*. [Http://www.Dmreview.Com/DM\\_Review\\_First\\_Door\\_Uses\\_Web\\_Trends\\_To\\_Create\\_A\\_Human\\_Resource.Htm/](http://www.Dmreview.Com/DM_Review_First_Door_Uses_Web_Trends_To_Create_A_Human_Resource.Htm/) DM Review Magazin Article.
- Mendonca, M, and N. L. Sunderhaft (1999). *Mining Software Engineering Data Survey*. [www.Dacs.Ditc.Mil/Techs/Datamining/Index.Html](http://www.Dacs.Ditc.Mil/Techs/Datamining/Index.Html)
- Olson, David L. *Data Mining* (2002). [Http://www.Ait.Unl.Edu/Dolson/Datamining.ppt](http://www.Ait.Unl.Edu/Dolson/Datamining.ppt).
- Porter, K, P. Smith, and R. Fagg (2006). *Leadership and Management for HR Professionals*. Oxford: Butterworth Heinemann.
- SAS Institute Inc. 2001. <http://www.Sas.Com>
- 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/](http://www.Dmreview.Com/DM_Review_Insights_From_The_Front_Line_DW_In_Human_Resource.Htm/) DM Review Magazin Article.
- Spss Inc. *Effectively Manage Human Resource Using Data Mining*. 2004. [http://www.Spss.Com/Applications/Government/Human\\_Resource\\_Planning.Htm](http://www.Spss.Com/Applications/Government/Human_Resource_Planning.Htm).
- Smith, K (2002). Personnel Seleccion Research. *International Journal of Organizational and Occupational Psychology* Vol.2: pp. 441- 472.
- Weiss, S. M., and N. Indurkha (1998). *Predictive Data Mining: A Practical Guide*. Morgan Kaufmann.

## پی نوشت

جهت آشنایی و مفاهیم تخصصی داده‌کاوی و نحوه تشکیل درخت تصمیم، به خوانندگان گرامی پیشنهاد می‌شود به منابع زیر مراجعه نمایند.

“Daniel, T ”., and “Kantardzic, Mehmed. *Data Mining: Concept, Models and Algorithms*. Wiley: IEE Press, 2003.” “and John Larose . *Data Mining Methods And Models*. Wiley & Sons Publishing Inc., 2006.