

استفاده از تکنیک داده کاوی برای ساخت مدل های طبقه بندی برای پیش بینی عملکرد کارکنان

چکیده

سرمایه انسانی از نگرانی‌ها برای مدیریت شرکت‌هایی است که علاقه بسیاری برای استخدام پرسنل واجد شرایط دارند که انتظار می‌رود به خوبی کار کنند. به تازگی، علاقه رو به رشدی در داده کاوی به وجود آمده است، که در آن هدف کشف دانش درست و نفع بالا برای کاربران است. در این مقاله، تکنیک‌های داده کاوی برای ساخت یک مدل طبقه بندی برای پیش بینی عملکرد کارکنان استفاده می‌شود. برای ساخت مدل طبقه بندی، روش‌های داده کاوی CRISP-DM به تصویب رسیدند. درخت تصمیم به عنوان مهمترین ابزار داده کاوی برای ایجاد مدل طبقه بندی مورد استفاده قرار گرفت، که در آن چندین قانون طبقه بندی تولید شد. برای اعتباربخشی به الگوی تولید شده، چندین آزمایش با استفاده از داده‌های واقعی جمع آوری شده از چندین شرکت انجام شده است. مدل برای پیش بینی عملکرد متقاضیان جدید در نظر گرفته شده است.

کلمات کلیدی: داده کاوی؛ تقسیم بندی؛ درخت تصمیم گیری؛ عملکرد کار.

1. مقدمه

منابع انسانی تقریباً به یکی از نگرانی‌های اصلی مدیران در تمام کسب و کارها من جمله، شرکت‌های خصوصی، موسسات آموزشی و سازمان‌های دولتی تبدیل شده است. سازمان کسب و کار واقعاً علاقه مند به حل و فصل برنامه‌ای برای انتخاب

درست کارکنان مناسب است. بعد از استخدام کارکنان، مدیریت در مورد عملکرد این کارکنان که سیستم مدیریت ارزیابی در تلاش برای حفظ عملکرد خوب کارمندان است، نگران می‌شود (Chen و Chein، 2006).

داده‌کاوی یک میدان جوان و امیدوار از اطلاعات و کشف دانش است (Han و همکارانش، 2011). در آغاز به دلیل وجود داده‌های بزرگ حاوی مقادیر زیادی از دانش پنهان، به‌عنوان یک هدف مهم برای صنعت اطلاعات بود. با تکنیک‌های داده‌کاوی، مانند دانش می‌توان وظایف پایگاه داده را از ذخیره و بازیابی به یادگیری و استخراج دانش تبدیل کرد.

داده‌کاوی شامل مجموعه‌ای از تکنیک‌ها است که می‌تواند برای استخراج دانش مرتبط و جالب از داده‌ها مورد استفاده قرار گیرد. داده‌کاوی شامل چندین کار مانند رابطه‌ی کاوش در قوانین، طبقه‌بندی و پیش‌بینی و خوشه‌بندی است. تکنیک‌های طبقه‌بندی، تکنیک‌های یادگیری هستند که داده‌ها را بنا به برجسب کلاس از پیش تعریف شده طبقه‌بندی می‌کنند. این یکی از مفیدترین تکنیک‌های داده‌کاوی در برای ساخت مدل‌های طبقه‌بندی از مجموعه داده‌های ورودی است. تکنیک‌های طبقه‌بندی مورد استفاده معمولاً برای ساخت مدل‌هایی مورد استفاده قرار می‌گیرند که برای پیش‌بینی روند داده است. چندین الگوریتم برای طبقه‌بندی داده‌ها مانند درخت تصمیم‌گیری و طبقه‌بندی ساده‌ی بیزی وجود دارد. با طبقه‌بندی، مدل تولید شده قادر به پیش‌بینی یک کلاس برای اطلاعات داده شده، بنا به اطلاعات آموخته شده از تاریخچه‌ی داده‌ها است.

درخت تصمیم‌گیری یکی از تکنیک‌های مورد استفاده است، زیرا درخت تصمیمی از اطلاعات داده شده با استفاده از معادلات ساده بنا به نسبت افزایش محاسبه ایجاد می‌کند که به‌طور خودکار نوعی وزن به ویژگی‌ها نسبت می‌دهد و محقق به‌طور ضمنی می‌تواند ویژگی موثر بر روی هدف پیش‌بینی شده را تشخیص دهد. به‌عنوان نتیجه‌ی این روش، یک درخت تصمیم‌گیری با قوانین طبقه‌بندی ساخته شده از آن ایجاد خواهد شد (Han و همکارانش، 2011).

طبقه‌بندی بیز یک روش طبقه‌بندی دیگری است که برای پیش‌بینی یک کلاس هدف استفاده می‌شود. که به محاسبات بر روی احتمالات، یعنی قضیه بیزی بستگی دارد. به دلیل استفاده از این طبقه‌بندی، نتایج حاصل از این طبقه‌بندی دقیق و موثر هستند و نسبت به افزودن داده‌های جدید به مجموعه داده حساس هستند (Han و همکارانش، 2011).

مطالعات متعددی از داده‌کاوی برای استخراج قوانین و پیش‌بینی رفتارهای خاصی در علوم مختلفی نظیر، فن‌آوری اطلاعات، منابع انسانی، آموزش و پرورش، زیست‌شناسی و پزشکی استفاده می‌کنند.

به‌عنوان مثال، Beikzadeh و Delavari (2004) از تکنیک‌های داده‌کاوی برای پیشرفت در سیستم‌های آموزشی استفاده می‌کنند. Radaideh و همکارانش (2006) نیز از داده‌کاوی برای پیش‌بینی عملکرد دانشجویان دانشگاه استفاده کردند. از سوی دیگر، بسیاری از محققان پزشکی، از تکنیک‌های داده‌کاوی برای استخراج بالینی با استفاده از فایل‌های عظیم داده‌های بیماران استفاده می‌کنند، Lavrac (1999) یکی از این محققان بود. Mullins و همکارانش (2006) نیز بر روی داده‌های بیماران برای استخراج قوانین انجمن بیماری با استفاده از روش‌های بدون نظارت کار کردند.

Karatepe و همکارانش (2006) عملکرد یک کارمند را مانند بهره‌وری خود در مقایسه با هم‌نوعشان را تعریف کردند. از سوی دیگر، Schwab (1991)، عملکرد استادان دانشگاه را در مطالعه خود، به‌عنوان تعداد تحقیقات ذکر شده و یا منتشر شده توصیف کرد. به‌طور کلی، عملکرد معمولاً با واحدهای تولید شده توسط کارمند در کار خود در دوره معین از زمان اندازه‌گیری شد.

محققان مانند Chein و Chen (2006) در بهبود انتخاب کارمند با ساخت یک مدل، با استفاده از تکنیک داده‌کاوی برای پیش‌بینی عملکرد متقاضیان اقداماتی انجام دادند. که به ویژگی‌های انتخاب شده از CV، برنامه کار و مصاحبه‌های خود بستگی دارد. عملکرد آنها می‌تواند مانند یک پایگاه برای تصمیم‌گیرندگان برای تصمیم‌گیری خود در مورد استفاده از هر دوی آنها یا هیچکدام پیش‌بینی شود.

مطالعات قبلی، صفات موثر زیادی در عملکرد کارکنان را مشخص می‌کند. برخی از این ویژگی‌ها، ویژگی‌های شخصی و بعضی دیگر آموزشی و در نهایت ویژگی‌های حرفه‌ای در نظر گرفته شده‌اند. Chein و Chen (2006) از ویژگی‌های زیادی برای پیش‌بینی عملکرد کارکنان استفاده کردند. آنها سن، جنس، وضعیت تأهل، تجربه، آموزش و پرورش، موضوعات عمده را به‌عنوان عوامل بالقوه که ممکن است عملکرد را تحت تأثیر قرار دهد، مشخص کردند. سپس آنها سن، جنس و وضعیت تأهل را کنار گذاشتند، به‌طوری‌که هیچ تبیعی در فرآیند انتخاب شخصی وجود نداشته باشد.

در نتیجه در مطالعه خود، آنها دریافتند که عملکرد کارکنان به شدت با درجه آموزش و پرورش و سابقه کار تحت تاثیر قرار دارد.

Kahya (2007) نیز بر روی عوامل خاصی که عملکرد شغلی را تحت تاثیر قرار می دهد جستجو کردند. محقق، مطالعات قبلی را بررسی کرد و اثر تجربه، حقوق و دستمزد، آموزش و پرورش، شرایط کار و رضایت شغلی در عملکرد را توصیف کرد. به عنوان یک نتیجه از پژوهش، مشخص شد که عوامل متعددی عملکرد کارکنان را تحت تاثیر قرار می دهند. موقعیت یا درجه‌ی کارمند در این شرکت از اثر مثبت بالایی در عملکرد خود دارد. از سوی دیگر، شرایط و محیط کار، هر دو رابطه مثبت و منفی بر روی عملکرد نشان دادند. کارمندان با تحصیلات بالا و کارکنان واجد شرایط، نارضایتی از شرایط کاری بد را بیان کردند و در نتیجه بر عملکرد خود تاثیر منفی گذاشتند. از سوی دیگر، کارمندان با مدارک تحصیلی پایین، کارآیی بالایی به رغم شرایط بد نشان دادند. علاوه بر این، تجربه در اکثر موارد رابطه‌ی مثبتی نشان داده است، در حالی که آموزش و پرورش رابطه مشخصی با عملکرد ندارد.

در مطالعه آنها، Salleh و همکارانش (2011) تاثیر انگیزه بر عملکرد شغلی کارکنان دولت در مالزی مورد آزمایش قرار گرفت. این مطالعه یک رابطه مثبت بین انگیزه وابستگی و عملکرد شغلی را نشان داد. همانطور که مردم با انگیزه وابستگی بالاتر و روابط میان فردی قوی با همکاران و مدیران و تمایل به انجام خیلی بهتر شغل خود دارند.

Jantan و همکارانش (2010) در مقاله خود، سیستم معماری منابع انسانی (HR) را برای پیش بینی استعداد متقاضی براساس اطلاعات پر شده در نرم افزار منابع انسانی و تجربه‌ی گذشته، با استفاده از تکنیک‌های داده کاوی مورد بحث قرار دادند. هدف از این مقاله این بود که راهی برای پیش بینی استعداد در موسسات بالاتر مالزی پیدا کنند. بنابراین آنها عوامل خاصی را به عنوان صفات سیستم خود، از جمله، صلاحیت حرفه‌ای، آموزش و تعهد اجتماعی مشخص کردند. سپس، چندین روش داده کاوی (ترکیبی) برای پیدا کردن قوانین پیش بینی کردند. شبکه عصبی مصنوعی، درخت تصمیم و نظریه مجموعه نمونه‌هایی از تکنیک‌های انتخاب شده هستند.

همین نویسندگان، Jantan و همکارانش (b2010) از الگوریتم درخت تصمیم طبقه بندی 4.5C برای پیش بینی استعداد بشر در مدیریت منابع انسانی، با ایجاد قوانین طبقه بندی برای ثبت تاریخی و آزمایش آنها بر روی داده‌های

نهان با محاسبه‌ی دقت و صحت استفاده کردند. آنها قصد استفاده از این قوانین برای ایجاد یک سیستم DSS را داشتند که می‌تواند توسط مدیریت برای پیش‌بینی عملکرد کارکنان و تبلیغات بالقوه استفاده شود. به‌طور کلی، این مقاله تلاش اولیه برای استفاده از مفاهیم داده‌کاوی، به ویژه طبقه‌بندی، برای کمک به حمایت از منابع انسانی مدیران و تصمیم‌گیرندگان با ارزیابی داده برای مطالعه‌ی ویژگی‌های اصلی کارکنان را دارند که ممکن است عملکرد کارکنان را تحت تاثیر قرار دهد. در این مقاله مفاهیم داده‌کاوی به‌منظور توسعه‌ی یک مدل برای حمایت پیش‌بینی عملکرد کارکنان اعمال شده است. در بخش 2، شرح کامل مطالعه، تعیین روش، نتایج، بحث در مورد نتایج ارائه شده است.

2. ساخت مدل طبقه‌بندی

هدف اصلی از روش پیشنهادی ساخت مدل طبقه‌بندی است که ویژگی‌های خاصی را که ممکن است عملکرد شغلی را تحت تاثیر قرار دهد مورد بررسی قرار می‌دهند. برای انجام این کار، روش CRISP-DM (صنعت پردازش استاندارد برای داده‌کاوی) (CRISP-DM، 2007) برای ساخت یک مدل طبقه‌بندی استفاده شد. که شامل پنج گام است: درک کسب‌وکار، درک داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و استقرار است.

(A) رده‌بندی مقدماتی داده

به‌طور کلی، طبقه‌بندی داده‌ها یک فرایند دو مرحله‌ای است. در مرحله اول، که به نام مرحله یادگیری است، یک مدل که مجموعه‌ای از پیش‌تعیین شده از کلاس‌ها یا مفاهیم توصیف است، با تجزیه و تحلیل مجموعه‌ای از نمونه پایگاه داده ساخته شده است. هر نمونه فرض شده است که متعلق به یک کلاس از پیش تعریف شده است. در مرحله دوم، مدل با استفاده از مجموعه داده‌های مختلف برای برآورد صحت مدل طبقه‌بندی مورد آزمون قرار می‌گیرد. اگر دقت مدل قابل قبول تشخیص داده شود، مدل را می‌توان برای طبقه‌بندی نمونه داده‌های آینده که برچسب کلاس آن‌ها شناخته شده نیست مورد استفاده قرار داد. در پایان، این مدل به عنوان یک طبقه‌بندی در فرایند تصمیم‌گیری عمل می‌کند.

چندین تکنیک مختلف وجود دارد که می‌تواند برای طبقه‌بندی مانند درخت تصمیم‌گیری، روشهای بیزی، الگوریتم براساس قوانین و شبکه‌های عصبی استفاده شود.

طبقه‌بندی درخت تصمیم‌گیری از تکنیک‌های کاملاً محبوب است چون ساخت درخت به‌هیچ‌گونه دانش تخصصی یا پارامتر تنظیم نیاز ندارد و برای کشف دانش اکتشافی مناسب است. درخت تصمیم‌گیری می‌تواند یک مدل با قوانینی که برای انسان قابل خواندن و قابل تفسیر هستند تولید شود. درخت تصمیم‌مزایای استفاده از تفسیر آسان و درک برای تصمیم‌گیرندگان را با توجه به مقایسه با دامنه‌ی دانش خود برای اعتبارسنجی و تصمیم خود را توجیه می‌کند. برخی از طبقه‌بندی‌های درخت تصمیم NBTtree، 4.8/J 5.0/ C 4.5C، و غیره است.

روش 4.5C یکی از شاخه‌های درخت تصمیم‌گیری است که می‌تواند هر دو درخت تصمیم‌گیری و مجموعه قوانین را تولید کند و یک درخت به‌منظور بهبود دقت پیش‌بینی بسازد. طبقه‌بندی 48/J 5.0/ C 4.5C در میان محبوب‌ترین و قدرتمندترین طبقه‌بندی‌های درخت تصمیم‌گیری است. 4.5C یک درخت اولیه با استفاده از الگوریتم تقسیم‌و‌حل ایجاد می‌کند. شرح کامل الگوریتم را می‌توان در هر کتاب داده‌کاوی یا یادگیری ماشین مانند (Han و همکارانش، 2011) و (Witten و همکارانش، 2011) پیدا کرد.

ابزار WEKA (Witten و همکارانش، 2011) یک ابزار با استفاده‌ی گسترده برای یادگیری ماشین و داده‌کاوی است که در دانشگاه وایکاتو در نیوزیلند توسعه یافته است. که شامل یک مجموعه بزرگ از الگوریتم‌های یادگیری ماشین و داده‌کاوی نوشته شده در جاوا است. WEKA شامل ابزاری برای رگرسیون، طبقه‌بندی، خوشه‌بندی، قوانین انجمن، تجسم و پردازش اولیه است. WEKA به ابزاری بسیار محبوب با محققان دانشگاهی و صنعتی تبدیل شده است و نیز به طور گسترده برای اهداف آموزشی استفاده می‌شود. ابزار WEKA شناخته شده با نسخه‌های J48، J48.1 یک پیاده‌سازی بهینه‌سازی شده از 4.5REV C است.

B) پردازش مجموعه داده و درک اطلاعات

هنگامی که ایده‌ی مطالعه به ذهن آمد، منجر به استفاده از مدل طبقه‌بندی برای پیش‌بینی عملکرد بنا به یک مجموعه داده از یک شرکت خاص در نظر گرفته شد. به‌طوری‌که هر گونه عوامل دیگر با توجه به محیط کار، شرایط، مدیریت و همکاران اثر مشابهی بر همه کارکنان دارند، و تاثیر ویژگی‌های جمع‌آوری شده آشکارتر و آسان‌تر برای طبقه‌بندی است. متأسفانه، اطلاعات جمع‌آوری شده از شرکت اولی به اندازه کافی نبود و بر اساس یک مدل طبقه‌بندی نشده بود. در این مورد، یکی دیگر از تلاش‌ها، جمع‌آوری یک گروه از داده‌ها از یک شرکت دیگر گرفته شد. به‌منظور جمع‌آوری داده‌های مورد نیاز، پرسشنامه آماده شد و یا از طریق ایمیل و یا دستی به کارکنان هر دو شرکت توزیع شد. سپس، در اینترنت پخش شد تا توسط کارکنان شاغل در هر شرکت IT پر شود. پرسشنامه توسط 130 نفر، 37 نفر از شرکت اول، 38 نفر از دومی و بقیه از چند شرکت دیگر با استفاده از پرسشنامه اینترنت پر شده بود. چند ویژگی که عملکرد کلاس را پیش‌بینی می‌کرد در پرسشنامه مطرح شده است. این لیست از ویژگی‌های جمع‌آوری شده در جدول 1 ارائه شده است.

C) آماده‌سازی داده

پس از این که پرسشنامه جمع‌آوری شد، روند آماده‌سازی داده‌ها انجام شد. ابتدا، اطلاعات مندرج در پرسشنامه به اکسل منتقل شد. پس از آن، انواع داده‌ها بازبینی و اصلاح شد. برخی از ویژگی‌ها مانند سال تجربه و دوره خدمت، در مقادیر پیوسته وارد شدند. بنابراین، تغییر آنها بر اساس محدوده نشان داده شده است. ویژگی‌های دیگر مانند تخصص، عنوان شغلی و رتبه، با مقادیر گسسته نشان داده شدند. برای مثال، در تخصص، عنوانی مانند مهندسی برق و مهندسی کامپیوتر بودند که به‌عنوان یک عبارت در نظر گرفته شده‌اند. CIS و MIS به‌عنوان فناوری اطلاعات در نظر گرفته شده‌اند.

این فایل‌ها آماده شدند و به فرمت (ARFF) سازگار با ابزار داده‌کاوی WEKA (Witten و همکارانش، 2011) تبدیل شدند، که در ساخت مدل استفاده می‌شود.

TABLE I. FULL DESCRIPTION OF ATTRIBUTES USED FOR PREDICTING THE PERFORMANCE CLASS

Attribute	Description	Possible Values
*Age	Employee's Age	a, b, c, d, e
Gender	Employee's Gender	Male, Female
MStatue	Employee's Marital Status	Single, Married with kids, Married without kids, Other
NKids	No. of Kids	0, 1, 2, 3, 4

UnivType	The type of university of graduation	Public, Private
GSpecial	General Specialization	Business, IT, English Literature, Engineering, CS, Other
Degree	Employee Education Degree	Diploma, Bachelor, High Diploma, Master, PhD
Grade	Employee Graduation Grade	Excellent, Very good, Good, Acceptable, Other
Country	Country of University	This attribute was eliminated, since majority of the values were Jordan
**Expyears	No. of Working Experience Years	a, b, c, d, e
PrevCo	No. of Previous Companies the employee worked for	0,1,2,3,4,5
JobTitle	Employee's Job Title in the current company	Developer, Officer, QA, Data Entry, System Administrator, Office Manager, Technical Writer, Technical Manager, Software Engineering, Accountant, Infrastructure Engineer, Department Manager, software Architect, Analyst, Designer, Trainer, PM, Consultant, Customer Support, GM.
Rank	Employee's Rank in the current company	Junior, Senior, Team Leader, Manager, Architect
***ServPeriod	Service Period in the current company (in years)	a, b, c, d
Working hours	No of working hours	Full Time, Part Time, Other. This attribute was eliminated, since most instances has Full time value.
****SalRange	Employee's Range of Salary	a, b, c, d, e
UncomWorkcond	Working in uncomfortable conditions (in employee's perspective)	Yes, No
Dissatsalrank	Existence of dissatisfaction in either salary or rank	Yes, No
Performance	Employee's performance, either as informed or predicted. This is a class	Accomplish, Exceed, Far Exceed

Notes: * Age range: a) Under 25 b) 25-29 c) 30-34 d) 35-40 e) Over 40
 ** Experience Years Range: a) None b) Less than a year c) 1-5 d) More than 5 - 10 e) More than 10
 *** Service Period Range: a) Less than 1 year b) 1-5 c) More than 5-10 d) More than 10
 **** Salary Range: a) Less than 250 b) 250-499 c) 500-999 d) 1000-2000 e) Over 2000

TABLE II. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E1

Method	10-Fold Cross Validation	Hold-out (60%)
<i>ID3</i>	36.9%	36.5%
<i>C4.5 (J4.8)</i>	42.3%	48.1%
<i>Naïve Bayes</i>	40.7%	44.2%

همانطور که قبلا گفته شد، داده به سه مجموعه داده تقسیم می شود. اولی شامل داده های کارمندان اولین شرکت فناوری اطلاعات است. مرحله دوم شامل داده های کارکنان شرکت دوم IT است. سومی شامل تمام داده های جمع آوری شده از سه منبع هستند. هر مجموعه داده دارای دو فایل ARFF حاوی داده های آن، با ویژگی کلاس (عملکرد) است. هر یک از این مجموعه داده ها در یک آزمایش جداگانه استفاده شده اند.

3. مدل سازی و آزمایشات

بعد از این که داده‌ها تهیه شد، مدل‌های طبقه‌بندی ساخته شدند. با استفاده از روش درخت تصمیم‌گیری، یک درخت برای هر یک از این آزمایش‌ها ساخته شده است. در این روش، مقدار نسبت افزایش برای نشان دادن اثربخشی وزن هر یک از ویژگی‌ها در کلاس مورد آزمایش، استفاده شده است و بر اساس ترتیب گره‌های درخت مشخص شده است. نتایج به دست آمده در بخش‌های زیر بحث شده است.

با اشاره به بحث مطالعات قبلی و همانطور که در جدول (1) شرح داده شد، یک گروه از صفات انتخاب شده است که در برابر اثر خود بر عملکرد کارکنان مورد آزمایش قرار می‌گیرد.

این صفات شامل، (1) سن، جنس، وضعیت تاهل و تعداد بچه‌ها (در صورت وجود) است، (2) اطلاعات آموزش و پرورش مانند: نوع دانشگاه، تخصصی عمومی، درجه و نمره، (3) اطلاعات حرفه‌ای مانند: تعداد سال‌های تجربه، تعدادی از شرکت‌های قبلی که کار می‌کردند، عنوان شغلی، رتبه، دوره خدمت در شرکت فعلی، حقوق و دستمزد کار، پیدا کردن شرایط کار ناراحت کننده و نارضایتی از حقوق و دستمزد و یا رتبه است. این ویژگی‌ها برای پیش‌بینی عملکرد کارکنان مورد استفاده قرار می‌گرفت.

(A) آزمایش اول (E1): استفاده از کل داده (130 مورد)

سه روش طبقه‌بندی بر مجموعه داده برای ساخت مدل طبقه‌بندی استفاده شده است. تکنیک‌ها عبارتند از: درخت تصمیم با دو نسخه، 3ID و 4.5C (در WEKA 4.8J) و طبقه‌بندی ساده بیز. برای هر آزمایش، دقت با استفاده از اعتبار 10-برابر متقاطع، و روش نگه داشتن مورد بررسی قرار گرفت. جدول 2 درصد دقت هر یک از این روش‌ها را نشان می‌دهد.

درخت تولید شده توسط الگوریتم 3ID بسیار عمیق بود، زیرا از صفت JobTitle که دارای 20 ارزش متفاوت است آغاز شده است.

JobTitle حداکثر نسبت وزن را دارد، که از گره شروع و موثرترین ویژگی ساخته شده است. دیگر ویژگی‌ها که در درخت تصمیم‌گیری شرکت کردند UnivType، SalRange، ExpYears، درجه، سن، MStatue، جنسیت، GSspecial و رتبه بودند. ویژگی‌های دیگر مانند: PrevCo، Nkids، uncomworkcond، dissatsalrank و درجه در سایر نقاط درخت تصمیم‌گیری ظاهر شدند.

درخت نشان داد که تمامی این ویژگی‌ها به نوعی بر عملکرد کارکنان تاثیر دارند، اما ویژگی‌های موثر عبارتند از: JobTitle، UnivType و سن. نکات دیگری می‌تواند از درخت استخراج شود که کارمندان جوان عملکرد بهتری نسبت به آنهایی که مسن‌تر هستند دارند. هر کجا که جنس در نظر گرفته شود، کارکنان مرد دارای عملکرد بالاتری از زنان هستند. علاوه بر این، کارمندان با نمرات فارغ التحصیلی بالاتر عملکرد بالاتری دارند. در نهایت، کارکنان با رتبه بالاتر عملکرد کمتری از خود نشانه دادند که مدیران کمتر از کارکنان با رتبه پایین کار می‌کنند.

درخت تولید شده با استفاده از الگوریتم 4.5C نیز نشان داد که ویژگی JobTitle ویژگی موثری است. طبقه‌بندی ساده بیز وزن هر یک از ویژگی‌های موجود در طبقه‌بندی را نشان نمی‌دهد، اما برای مقایسه نتایج حاصل از 3ID و 4.5C که پیش از این در جدول 2 بیان شد استفاده می‌شود. می‌توان با دقت بیان کرد که محدوده درصد از حدود 36 درصد به 45 درصد، درصد کمی است.

ب) آزمایش دوم (2E): استفاده از مجموعه داده جمع آوری شده از اولین شرکت IT (37 مورد)

با استفاده از روش یکسان 1E، جدول 3 دقت پیش‌بینی برای هر الگوریتم اعمال شده به این مجموعه داده را نشان می‌دهد.

TABLE III. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E2

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	37.8 %	26.7%
C4.5 (J4.8)	48.6%	53.3%
Naive Bayes	37.8%	46.7%

درخت تصمیم‌گیری ساخته شده توسط الگوریتم ID3، روند متفاوتی از تولید درخت توسط 1E را نشان می‌دهد، چون در این درخت، گره شروع PrevCo بود، و صفات با بالاترین نسبت وزن NKids، GSpecial، JobTitle، و سن بودند، در حالی که ویژگی‌های با تاثیر کمتر MStatus، Uncomworkcond، UnivType، درجه و Dissatsalwork بود. درخت تصمیم ساخته شده با استفاده از الگوریتم 4.5C آنقدر هرس شد که تنها شامل سه ویژگی است؛ با PrevCo به عنوان گره شروع و Dissatsalrank و درجه به عنوان ویژگی‌های دیگر.

برای ویژگی PrevCo، می‌توان متوجه شد که عملکرد کارکنان از حد معمول متفاوت است و به مراتب بیشتر است اگر PrevCo کمتر از 3 باشد، پس از آن انجام می‌شود و دوباره وقتی که PrevCo بیش از 3 می‌شود بالا می‌رود. علاوه‌براین، هنگامی که dissatsalrank به است، وظیفه‌ی کارمند انجام می‌شود، در حالی که مقداری را نشان ندهد رضایت از یک کارمند از انجام کار تجاوز می‌کند. این مسئله یک واکنش طبیعی از نارضایتی در حقوق و دستمزد یا رتبه را نشان می‌دهد.

ویژگی درجه یک نتیجه جالب توجه است، که نشان می‌دهد که کارکنان با نمره خوب، به مراتب در اجرا قوی هستند، در حالی که نمرات دیگر تنها تجاوز می‌کنند. این می‌تواند نشان دهد که فارغ‌التحصیلان با معدل بالا لزوماً نشانگر داشتن کارمندان خوب نیست. به‌عنوان مثالی از درخت 4.5C، شکل 1 درخت تولید شده برای 2E را نشان می‌دهد و جدول 4 قوانین طبقه‌بندی تولید شده با تعداد نمونه که هر قانون حمایت می‌کند را نشان می‌دهد.

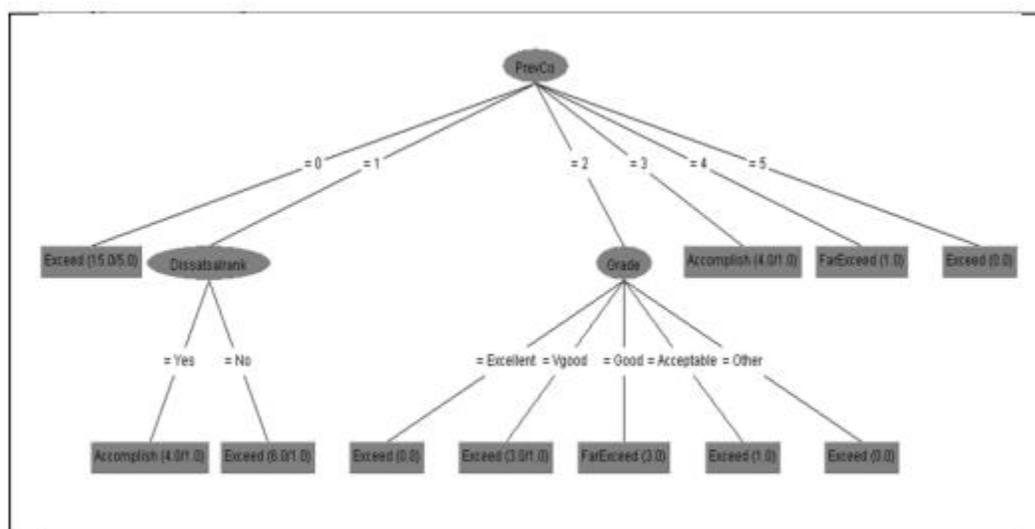


Figure 1. A decision tree generated by C4.5 algorithm for E2 for predicting performance

TABLE IV. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E2 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF PrevCo = 0 ==>	Exceed	15
2	IF PrevCo = 1 & Dissatsalrank = Yes ==>	Accomplish	4
3	IF PrevCo = 1 & Dissatsalrank = No ==>	Exceed	6
4	IF PrevCo = 2 & Grade <> Good ==>	Exceed	4
5	IF PrevCo = 2 & Grade = Good ==>	FarExceed	3
6	IF PrevCo = 3 ==>	Accomplish	4
7	IF PrevCo = 4 ==>	FarExceed	1

(C) آزمایش سوم (3E): استفاده از مجموعه داده جمع‌آوری شده از دومین شرکت فناوری اطلاعات (38 مورد)

جدول 5 درصد دقت ناشی از استفاده از الگوریتم‌های 3ID، 4.5C و ساده بیز در مجموعه داده‌ی شرکت دوم را نشان می‌دهد. توجه داشته باشید که درصد دقت در این آزمایش افزایش یافته است.

درخت تصمیم ساخته شده با استفاده از الگوریتم 3ID برای این آزمایش با JobTitle آغاز شده است و پس از آن برای صفاتی مانند GSpecial، رتبه، درجه و ExpYears و ویژگی‌های بیش از حد مانند SalRange، درجه و PrevCo وزن بیشتر شده است.

TABLE V. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E3

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	50%	43.7%
C4.5 (J4.8)	60.5%	56.2%
Naïve Bayes	65.8%	68.7%

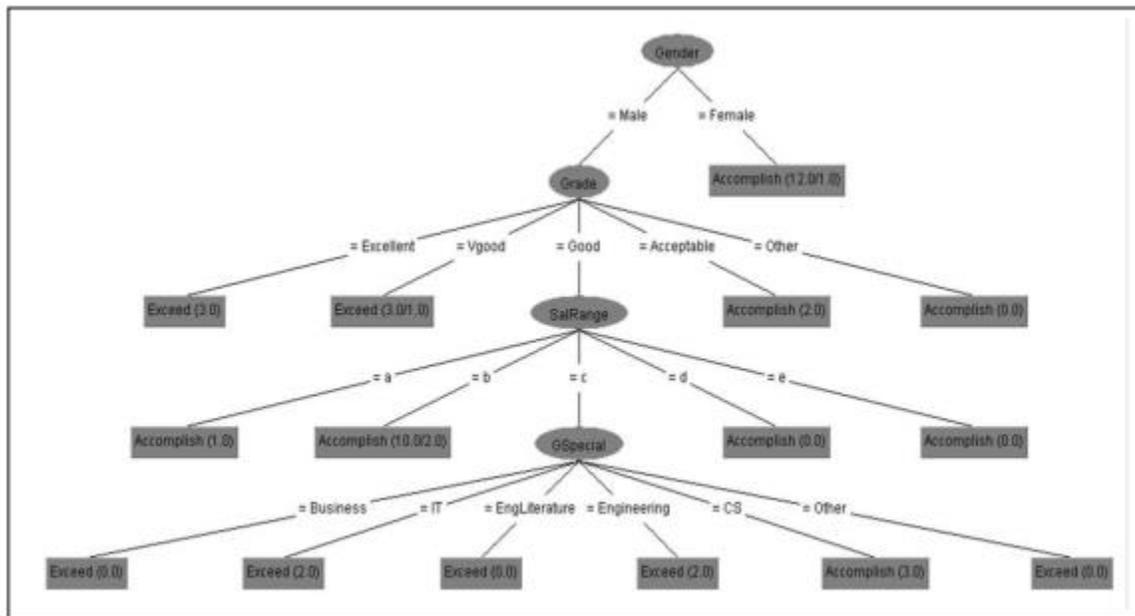


Figure 2. Decision Tree resulted from C4.5 algorithm for E3 to predict performance

TABLE VI. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E3 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF Gender = Male & Grade = Excellent ==>	Exceed	3
2	IF Gender = Male & Grade = Vgood ==>	Exceed	3
3	IF Gender = Male & Grade = Good & SalRange = a ==> Performance =	Accomplish	1
4	IF Gender = Male & Grade = Good & SalRange = b ==> Performance =	Accomplish	10
5	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = IT ==>	Exceed	2
6	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = Engineering ==>	Exceed	2
7	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = CS ==>	Accomplish	3
8	IF Gender = Male & Grade = Acceptable ==>	Accomplish	2
9	IF Gender = Female ==>	Accomplish	12

با استفاده از الگوریتم 4.5C، نسبت سود حاصل از جنسیت نسبت به شروع با درخت بالاتر بود و سپس درجه،

SalRange و GSpecial می‌آید. شکل 2 درخت تولید شده توسط WEKA برای E3 را نشان می‌دهد و جدول 6

قوانین طبقه‌بندی تولید شده را نشان می‌دهد.

4. نتایج و بحث

این مطالعه عوامل متعددی را که ممکن است اثر زیادی بر عملکرد کارکنان داشته باشند نشان می‌دهد. یکی از عوامل موثر عنوان شغلی است.

روند اثربخشی عنوان شغلی در نتایج بسیار روشن نیست، از آنجا که حدود 20 عنوان شغلی مورد مطالعه وجود دارد، اما می‌توان آن را به نوع پیچیدگی کار و مسئولیت‌های مربوط و مطرح شده دانست. مسئولیت بالا گاهی اوقات بر انگیزه کارکنان و در نتیجه عملکرد در راه مثبت تاثیر می‌گذارد.

ویژگی نوع دانشگاه، در سه آزمایش، تاثیر عملکرد مثبتی دارد زمانی که کارمند از یک دانشگاه دولتی به جای یک دانشگاه خصوصی فارغ التحصیل شده باشد. این مسئله می‌تواند به این دلیل باشد که دانشگاه‌های دولتی، در اکثر موارد، دانش‌آموزان با نمرات بالا در دبیرستان را نسبت به دانشگاه‌های خصوصی پذیرش می‌کنند.

دیگر عوامل آموزشی مانند درجه و نمره تاثیر عملکرد کمی دارند، اما با روند نه چندان روشنی، ممکن است به عوامل دیگری مانند شخصیت کارکنان، که در این مطالعه در نظر گرفته نشده است بستگی داشته باشند. چنین، عوامل شخصیتی را می‌توان با تصمیم‌گیرندگان در مصاحبه به رسمیت شناخت، به طوری که آنها می‌توانند دانش خود را در مورد متقاضی کامل کنند. تخصص عمومی دانشگاه دارای اثر بسیار نزدیکی به عملکرد به عنوان ویژگی شغلی دارد. این می‌تواند به علت ارتباط بین این دو عامل باشد.

برخی از اطلاعات شخصی مانند سن، وضعیت تأهل و جنسیت نیز بر عملکرد تاثیر می‌گذارند. با این وجود، سن اثر مشخصی بر عملکرد ندارد، چرا که گاهی اوقات عملکرد با عامل تجربه و افزایش سن، اضافه می‌شود. از سوی دیگر، وضعیت تأهل، به روشنی مشخص است، کارکنان واحد در همه آزمایش‌ها، عملکرد بهتری از کارکنان متأهل و حتی آنهایی که با بچه‌های کارکنان ازدواج کرده بودند نشان داده‌اند.

اما، شگفت آور، در آزمایش 2E، روند عجیب و غریب با توجه به تعداد بچه‌ها است، که نشان داد که تعداد بیشتر بچه‌ها منجر به عملکرد بالاتری می‌شود. این می‌تواند یک پرت تصادفی باشد، زیرا مجموعه داده 2E به اندازه کافی برای اعلام

این قاعده بزرگ نیست. از سوی دیگر جنسیت، هیچ اثری بر آزمایش‌های 1E و 2E نداشت، چرا که نسبت زن در هر دو مجموعه داده قابل توجه نیست. اما در آزمایش 3E، عملکرد بالاتری را برای کارمندان مرد از زن نشان داد. به نظر می‌رسد عوامل متعددی بر عملکرد تاثیر دارد. حقوق و دستمزد، یکی از عوامل مثبت بر عملکرد است، این اثرات در آزمایش‌های 1E و 3E نشان داده شده است، در حالی که در 2E معنی‌دار نبود. تعدادی از شرکت‌های قبلی در 1E، رابطه مثبت و منفی با عملکرد نشان دادند. این می‌تواند به علت کارمندان تازه‌کاری که تجربه کار در شرکت‌های دیگر را ندارد باشد؛ آنها بهترین کار را برای به دست آوردن موقعیت‌های بهتر انجام می‌دهند. از سوی دیگر، کارکنانی که در برخی از شرکت‌های قبلی کار می‌کردند ممکن است تجربه زیادی نداشته باشند تا عملکرد خود را تحت تاثیر قرار دهند. در آزمایش‌های 2E و 3E تنها رابطه‌ی مثبت مشاهده می‌شود. همانطور که سال‌ها تجربه، عملکرد مثبت در 1E و 2E را تحت تاثیر قرار می‌دهد، در حالی که در 3E از اهمیت زیادی برخوردار است. ویژگی رتبه تاثیر جالبی در عملکرد، به ویژه در 1E و 3E نشان می‌دهد. همانطور که در 2E به‌عنوان عامل موثر گنجانده نشده است. در آزمایش‌های 1E و 3E متوجه شده بودند که، عملکرد کارکنان ارشد بیش از جوانان است. این به دلیل تجربه کارکنان بود که رتبه را تحت تاثیر قرار می‌داد. اما، شگفت‌آور، رهبران تیم و مدیران هستند که عملکرد کمتری از کارکنان ارشد دارند.

در نهایت، رضایت شغلی و محیط کار راحت تاثیر چندانی بر عملکرد دارد. برای 3E آنها به‌عنوان عوامل موثر شامل نمی‌شوند. در حالی که در 1E و 2E آنها به‌عنوان عوامل کم تاثیر بر عملکرد وزن در نظر گرفته شده‌اند. این را می‌توان این چنین تفسیر کرد که شرکت 3E دارای شرایط رضایت بخش‌تری از شرکت در 2E دارد. به‌عنوان نکته نهایی در مورد دقت مدل طبقه‌بندی ساخته شده برای سه آزمایش، می‌توان متوجه شد که برای الگوریتم‌های مختلف استفاده می‌شود، دقت طبقه‌بندی در آزمایش 2E و 3E نسبت به 1E بسیار بیشتر بود. این ممکن است به دلیل وجود کارکنان مختلف از شرکت‌ها، از جمله در 1E باشد، که عوامل مختلفی بر کلاس‌ها در آزمایش ایجاد می‌شود. در حالی که در 2E و 3E، به‌رغم مجموعه داده‌های کوچک آن‌ها، اما کارکنان مورد مطالعه، شرایط کار، محیط کار، مدیریت و همکاران یکسانی دارند که بر مطالعه بیشتر بر روی ویژگی‌های قابل اندازه‌گیری تمرکز دارند.

5. نتیجه‌گیری و کارهای آتی

این مقاله بر امکان ساخت یک مدل طبقه‌بندی برای پیش‌بینی عملکرد کارکنان متمرکز شده است. در حال کار بر روی عملکرد، بسیاری از ویژگی‌ها آزمایش شده است، و برخی از آنها در پیش‌بینی عملکرد موثر بودند. عنوان شغلی قویترین ویژگی و سپس نوع دانشگاه، با اثر کمی از درجه و نمره بود.

ویژگی سن هیچ اثر روشنی را نشان نمی‌دهد در حالی که وضعیت تاهل و جنس تاثیر کمی در برخی از آزمایشات برای پیش‌بینی عملکرد نشان داده‌اند. حقوق و دستمزد تعدادی از شرکت‌های قبلی، سال آزمایش و رضایت شغلی، هر کدام درجه‌ای تاثیر در پیش‌بینی عملکرد دارند. برای مدیریت شرکت و بخش منابع انسانی، این مدل، و یا نسخه‌ی افزایش یافته، می‌تواند در پیش‌بینی عملکرد کارکنان تازه متقاضی، استفاده شود. اقدامات چندی را می‌توان در این مورد برای جلوگیری از هر گونه خطر مربوط به استخدام کارمند ضعیف انجام داد.

به‌عنوان کارهای آتی، جمع‌آوری داده مناسب‌تر از چند شرکت توصیه می‌شود. پایگاه داده برای کارکنان فعلی و حتی آنهایی که قبلاً بودند به‌عنوان یک نرخ عملکرد صحیح برای هر یک از آنها می‌تواند مورد استفاده قرار گیرد. هنگامی که مدل مناسب تولید می‌شود، نرم‌افزار می‌تواند توسط HR از جمله قوانین تولید شده برای پیش‌بینی عملکرد کارکنان استفاده شود.

REFERENCES

- [1] Allen, M.W., Armstrong, D.J., Reid, M.F., Riemenschneider, C.K. (2009). "IT Employee Retention: Employee Expectations and Workplace Environments", SIGMIS-CPR'09, May 2009, Limerick, Ireland.
- [2] Al-Radaideh, Q. A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology (ACIT 2006), Dec 2006, Jordan.
- [3] Chein, C., Chen, L. (2006) "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", Expert Systems with Applications, In Press.
- [4] Cho, S., Johanson, M.M., Guchait, P. (2009). "Employees intent to leave: A comparison of determinants of intent to leave versus intent to stay", International Journal of Hospitality Management, 28, pp374-381.
- [5] CRISP-DM, (2007). Cross Industry Standard Process for Data Mining: Process Model. <http://www.crisp-dm.org/process/index.htm>. Accessed 10th May 2007.
- [6] Delavari, N., PHON-AMNUAISUK S., (2008). Data Mining Application in Higher Learning
- [7] Dreher, G.F. (1982) "The Role of Performance in the Turnover Process", Academy of Management Journal, 25(1), pp. 137-147.
- [8] Han, J., Kamber, M., Jian P. (2011). Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers.
- [9] Ibrahim, M.E., Al-Sejini, S., Al-Qassimi, O.A. (2004). "Job Satisfaction and Performance of Government Employees in UAE", Journal of Management Research, 4(1), pp. 1-12.
- [10] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010a) "Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application", International Journal of Humanities and Social Science, 5(11), pp. 694-702.
- [11] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010b). "Human Talent Prediction in HRM using C4.5 Classification Algorithm", International Journal on Computer Science and Engineering, 2(08-2010), pp. 2526-2534.
- [12] Karatepe, O.M., Uludag, O., Menevis, I., Hadzimehmedagic, L., Baddar, L. (2006). "The Effects of Selected Individual Characteristics on Frontline Employee Performance and Job Satisfaction", Tourism Management, 27 (2006), pp. 547-560.
- [13] Kayha, E. (2007). "The Effects of Job Characteristics and Working Conditions on Job Performance", International Journal of Industrial Ergonomics, In Press.
- [14] Khilji, S., Wang, X. (2006). "New evidence in an old debate: Investigating the relationship between HR satisfaction and turnover", International Business Review, In Press.
- [15] Lavrac, N. (1999). "Selected Techniques for Data Mining in Medicine", Artificial Intelligence in Medicine, 16(1999), pp. 3-23.
- [16] Mullins, I., Siadaty, M., Lyman, J., Scully, K., Garrett, C., Millar, W., Mullar, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S., Knaus, W. (2006). "Data Mining and Clinical Data Repositories: Insights from 667,000 Patient Data Set", Computers in Biology and Medicine, 36, pp. 1351-1377.
- [17] Russ, F.A., McNeilly, K.M. (1995). "Links among Satisfaction, Commitment, and Turnover Intentions: The Moderating Effect of Experience, Gender and Performance", Journal of Business Research, 34, pp. 57-65.
- [18] Salleh, F., Dzulkifli, Z., Abdullah, W.A. and Yaakob, N. (2011). "The Effect of Motivation on Job Performance of State Government Employees in Malaysia", International Journal of Humanities and Social Science, 1(4), pp. 147-154.
- [19] Schwab, D. (1991). "Contextual variables in employee performance/turnover relationships", Academy of Management Journal, 34 (4), pp. 966-975.
- [20] Sexton, R.S., McMurtrey, S., Michalopoulos, J.O., Smith, A.M. (2005). "Employee Turnover: A Neural Network Solution", Computer & Operations Research, 32, pp. 2635-2651.
- [21] Witten I. Frank E., and Hall M. (2011). Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition, Morgan Kaufmann Publishers.