https://dorl.net/dor/20.1001.1.24235261.1397.11.2.8.1

مدیریت و برنامه‌ریزی در نظام‌های آموزشی
دوره ۱۱، شماره ۲ (پیاپی ۲۱)، پاییز و زمستان ۱۳۹۷، ۲۰۸-۱۸۷

Journal of Management and Planning in Educational Systems 2018, Vol. 11 (2), 187-208

تحلیل داده‌های آموزشی دانشجویان با هدف ارزیابی موفقیت تحصیلی با استفاده از رویکرد داده‌کاوی (نمونه موردی: دانشکده مدیریت و مهندسی صنایع دانشگاه شاهرود)

Analyzing Students' Educational Information to Evaluate Their Success via Using Data Mining Method (Case Study: Faculty of Management and Industrial Engineering, Shahrood University of Technology)

دریافت مقاله: ۱۳۹۴/۴/۵؛ **دریافت نسخه نهایی:** ۱۳۹۵/۱۰/۲۷؛ **پذیرش مقاله:** ۱۳۹۵/۱۱/۱۴

علی اکبر حسنی[۱] و مرتضی بذرافشان[۲]

چکیده: کسب دانش پیرامون الگوهای رفتاری وضعیت تحصیلی دانشجویان نقش بسزایی در مدیریت موفق مراکز آموزشی و ارائه خدمات مطلوب و متناسب با وضعیت‌های فعلی و آتی تحصیلی دانشجویان دارد. این دانش، در اطلاعات آموزشی نهفته بوده و می‌تواند با استفاده از ابزار داده‌کاوی استخراج شود. در این مقاله یک مدل داده‌کاوی مبتنی بر تکنیک خوشه‌بندی سلسله‌مراتبی توسعه‌یافته برای تحلیل اطلاعات مربوط به وضعیت تحصیلی دانشجویان شامل اطلاعات پیش از ورود به دانشگاه و وضعیت تحصیلی آنان ارائه شده است. علاوه بر آن، یک مدل پیش‌بینی مبتنی بر قرارگیری هر دانشجو با مشخصات خاص در یک خوشه و برآورد عملکرد تحصیلی وی ارائه شده است. مدل داده‌کاوی پیشنهاد شده در دانشکده مدیریت دانشگاه صنعتی شاهرود به‌عنوان نمونه موردی پیاده‌سازی شده است. نتایج تحقیق توسط خبرگان و مدیران آموزشی دانشگاه اعتبارسنجی شده است. پس از تحلیل داده‌ها و ارائه مدل پیش‌بینی، خوشه‌های دانشجویان مبتنی بر شاخص نشان‌دهنده وضعیت تحصیلی ارائه شده است. نتایج نشان می‌دهند که برخی از ویژگی‌های دانشجویان مربوط به دوران قبل و بعد از ورود آنان به دانشگاه بر تعلق آنان به خوشه‌های وضعیت تحصیلی تأثیر معنادار دارند. با استفاده از ابزار داده‌کاوی، مجموعه‌ای از قوانین برای ارتقای کیفیت تصمیمات مدیران آموزشی و انتقال دانشجویان از خوشه‌های با وضعیت تحصیلی نامناسب به وضعیت بهتر پیشنهاد شده است. هم‌چنین با توجه به شناسایی وضعیت تحصیلی دانشجویان، مدل پیش‌بینی‌کننده وضعیت آتی تحصیلی آنان نظیر رشد معدل و یا مشروطی تحصیلی ارائه شده است. تمامی نتایج با توجه به نظرات دریافت‌شده از خبرگان ارزیابی و اعتبار آنان تأیید شده است.

کلیدواژه‌ها: موفقیت تحصیلی، داده‌کاوی آموزشی، مدل پیش‌بینی، خوشه‌بندی سلسله‌مراتبی توسعه داده شده.

A. Hasani (PhD.) & M. Bazrafshan (PhD.)

Abstract: Knowledge acquisition about the behavior patterns of the educational status of students can have a significant impact on the successful management of educational institutes and provides appropriate services in accordance with current and future status of students. This knowledge is hidden in educational data and could be extracted via using data mining method. In this paper, a data mining model to analyze students' academic achievement based on the data related to before and after the students' entrance to university was proposed. To tackle this problem, an advanced hierarchical clustering method was developed. In addition, a forecasting model to determine the belonging of each student to the each considered cluster was provided. The proposed model was evaluated on a case study of the Faculty of Management in Shahrood University of Technology. The obtained results were validated by experts and university administrators. The students were classified in different clusters via considering theirs specific features. The results indicated that some of the considered features related to before and after entering university have significant impacts on determining the students' educational achievements and also their assigned cluster. Some decision rules are proposed for educational administrators to improve the overall students' educational achievements and transferring students from clusters with poor academic situations to better ones such as applied rules for transferring students with special status. In addition, an efficient forecasting model for assessment of students' achievement states was proposed. The obtained results were verified by experts.

Key words: educational achievement, educational data mining, forecasting model, advanced hierarchical clustering.

۱. نویسنده مسئول: استادیار دانشکده مهندسی صنایع و مدیریت، دانشگاه صنعتی شاهرود.
aa.hasani@shahroodut.ac.ir
۲. استادیار، دانشکده مهندسی صنایع، دانشگاه آزاد لاهیجان.

مقدمه

امروزه بسیاری از مراکز آموزش عالی برای مقابله و کاهش زیان‌های ناشی از افت تحصیلی دانشجویان، از ابزار کارآمدی به نام داده‌کاوی آموزشی استفاده می‌نمایند. در این فرآیند، استخراج داده‌های مرتبط با دانشجویان شامل اطلاعات فردی، آموزشی و موارد مشابه دیگر از میان انبوه داده‌های ذخیره شده در پایگاه‌های اطلاعاتی در اختیار مراکز آموزش عالی و تجزیه و تحلیل آن‌ها منجر به شناسایی الگوهای پنهان خواهد شد که ازجمله می‌توان به مواردی نظیر شناسایی دانشجویان درخطر افت تحصیل، ارائه الگوی مناسب برای حفظ دانشجو با استفاده مؤثر از خدماتی هم‌چون مشاوره، توصیه به شرکت در کلاس‌های تقویتی و بازآموزی و موارد مشابه دیگر اشاره نمود. این یافته‌ها (الگوهای پنهان) مزیت‌های ارزشمند فراوانی را در ابعاد مختلف مالی، مسئولیت اجتماعی و درنهایت پیشبرد اهداف مدنظر برای مراکز آموزش عالی به همراه خواهد داشت.

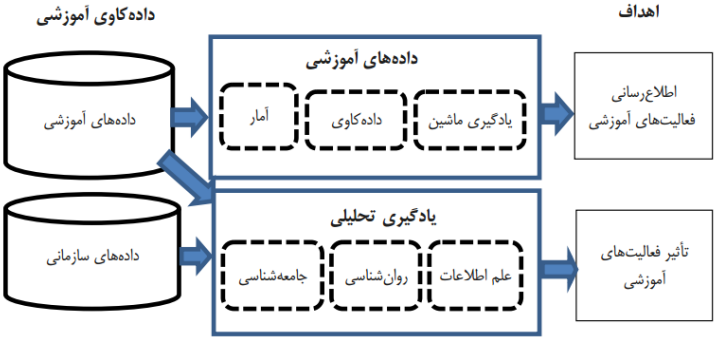
هدف از این پژوهش، شناسایی و معرفی عوامل مهم در موفقیت تحصیلی دانشجویان دانشکده مدیریت و مهندسی صنایع دانشگاه صنعتی شاهرود است. در این راستا، پاسخ‌گویی به این سؤال حائز اهمیت است که کدام‌یک از عوامل متعدد نظیر سن، جنسیت، بومی بودن، مقطع تحصیلی، سنوات تحصیلی، سهمیه قبولی، نوع دیپلم، معدل ورود و موارد مشابه دیگر در ایجاد وضعیت تحصیلی دانشجویان از اهمیت بیشتری برخوردار است؟ رویکرد پیش رو بر مبنای جمع‌آوری داده‌های مناسب از داده‌های ذخیره شده در پایگاه‌های اطلاعات دانشجویی و تجزیه‌وتحلیل آن‌ها، و درنهایت شناسایی علل تغییر در وضعیت تحصیلی است. به دنبال آن نیز، ارائه یک مدل پیش‌بینی برای تعیین سیاست‌های مداخله‌جویانه مناسب به‌منظور کاهش درصد افت تحصیلی دانشجویان و بهبود وضعیت تحصیلی آنان در راستای راهبردهای مناسب برای پشتیبانی از تصمیم‌های مدیران ارشد آموزش در دانشگاه صنعتی شاهرود به‌عنوان نمونه موردی و مراکز آموزش عالی شکل خواهد گرفت. با توجه به اینکه آموزش عالی همواره با داده‌ها و اطلاعات بسیار زیادی پیرامون دانشگاه‌ها، دانشجویان، اعضای هیأت‌علمی، کارکنان، منابع مادی و موارد مشابه دیگر روبرو است و در اکثر مواقع این داده‌ها می‌تواند حامل اطلاعات و الگوهای با ارزشی باشند، لذا به نظر می‌رسد یکی از مهم‌ترین کاربردهای داده‌کاوی نیز در آموزش عالی باشد. از رویکرد داده‌کاوی آموزشی می‌توان به منزله‌ی «یادگیری تحلیلی» به کمک مجموعه‌هایی از داده‌های حجیم آموزشی یاد کرد (بینکوفسکی، ۲۰۱۲). داده‌کاوی آموزشی از روش‌ها و تکنیک‌های آماری، یادگیری ماشین و داده‌کاوی، به‌منظور تجزیه‌وتحلیل داده‌های جمع‌آوری‌شده برای توسعه و بهبود بهره می‌برد و بر ایجاد مخازن اطلاعات ساختارمند به‌منزله منابع داده در داده‌کاوی تمرکز می‌کند (شکل ۱). در طی سالیان اخیر، تکنیک داده‌کاوی و

کاربرد آن بسیار موردتوجه محققان در حوزه‌های مختلف قرار گرفته و تبدیل بـه یکـی از اجـزای اصلی فرآیند کمک به تصمیم‌سازی در سازمان‌ها شده است (برادواج و پال، ۲۰۱۱).

پیشینه پژوهش

امروز بانک‌های اطلاعاتی وسیعی از ویژگی‌های دانشجویان موجود است که اطلاعات مربوط به ویژگی‌های خانوادگی، تحصیلی و غیره را شامل می‌شود. پیدا کردن الگوها و دانش نهفته در این اطلاعات می‌تواند به تصمیم‌گیرندگان عرصه آموزش عالی کمک شایانی بکند (صباغیان و کاظمی، ۱۳۸۷، عیسایی و همکاران، ۱۳۹۰). استفاده از تکنیک‌های پیشرفته داده‌کاوی مانند خوشه‌بندی، طبقه‌بندی، و غیره می‌تواند در طبقه‌بندی دانشگاه‌ها، یافتن الگوهای خاص و با ارزش در مورد دانشجویان موفق، یافتن یک برنامه یا روش موفق تدریس، یافتن نقاط بحرانی در مدیریت مالی دانشگاه‌ها و موارد دیگر کاربرد داشته باشد. اهمیت این حوزه تحقیقاتی توسط صفاپور (۱۳۹۲) به‌خوبی نشان داده شده است. در این مقاله تکنیک‌های داده‌کاوی برای اصلاح کارآیی مؤسسات آموزش عالی مورد مطالعه وتجزیه‌وتحلیل قرارگرفته است. نتایج نشان داده است که اگر تکنیک‌های داده‌کاوی همانند دسته‌بندی، تصمیمات درختی، و سازمان‌دهی لازم بکار گرفته شود، عملکرد، مدیریت دوره تحصیلی، انتخاب واحدها، سرعت یادگیری و ارزیابی آن، و مدیریت کمک هزینه تحصیلی دانشجویان در یک مؤسسه اصلاح و به آن‌ها کمک می‌شود. هم چنین تحلیل داده‌ها نقش مؤثری را در تصمیم‌گیری‌ها داشته و این عمل می‌تواند رویکردی برای آزمودن تأثیر استفاده از تکنیک‌های داده‌کاوی در آموزش عالی باشد. داده‌کاوی آموزشی زمینه های مختلفی را پوشش می‌دهد که برخی از مهم‌ترین آن‌ها عبارت‌اند از:

- یافتن علل احتمالی تداوم افت تحصیلی یا انصراف از تحصیل دانشجو؛
- ساخت مدل عملکرد دانشجو نظیر پیش‌بینی وضعیت تحصیلی؛
- فراهم آوری زمینه کـار بـا حجـم زیـادی از داده‌هـای ذخیـره‌شـده توسـط کـاربران غیرمتخصص و کسب دانش عمیق درباره متدها، فرآیندها و الگـوریتم‌هـای مدیریت سیستم آموزشی (گارسیا، رومرو، ونتورا و کاسترو،۲۰۱۱)؛ و
- پشتیبانی مؤثر از دانشـجویان به‌واسطه تحلیل اطلاعات سیستم مدیریت آموزشی (پچنیزکی، کالدر، واسیلی اوا و برا، ۲۰۰۸).



شکل ۱. نظریه بینکوفسکی پیرامون داده‌کاوی آموزشی (احمدی و همکاران (۱۳۹۴))

برای نمونه، در حوزه تحلیل داده‌های آموزشی، تاری (۱۳۹۰) در پایان‌نامه خود اقدام به تحلیل مدل‌های داده‌کاوی برای به‌کارگیری در آموزش عالی و تعیین پارامترهای مناسب برای واردکردن به مدل نهایی کرده است. در این تحقیق از اطلاعات دانشجویان دانشگاه پیام نور قم به‌عنوان نمونه موردی مورد مطالعه استفاده شده است. به‌صورت مشابه طالبی (۱۳۹۲) در پایان‌نامه خود اقدام به ارائه یک سیستم پشتیبانی از تصمیم برای مدیران گروه جهت تشکیل سبد دروس مناسب کرده است. در این تحقیق نیز از اطلاعات دانشجویان مقطع کارشناسی دانشگاه قم به‌عنوان نمونه موردی تحقیق استفاده‌شده است. احمدی و همکارانش (۱۳۹۴) در تحقیق خود بروی دانشجویان انصرافی دانشگاه تهران تمرکز کرده‌اند. نویسندگان در این مقاله رویکرد پذیرش دانشجوی شهریه‌پرداز را نوعی کسب‌وکار و انصراف دانشجو را روی‌گردانی مشتری در نظر گرفته شده است و به دنبال بررسی عوامل انصراف دانشجویان و اتخاذ سیاست‌های مداخله‌جویانه‌ی بازدارنده هستند. در این پژوهش داده‌های کمی و کیفی بر مبنای روش پژوهش کریسپ از داده‌کاوی اطلاعات دانشجویان ورودی شهریه پرداز (۲۱۴۲۰ دانشجو دانشگاه تهران طی سال‌های ۱۳۹۲- ۱۳۸۸) استخراج‌شده از بانک‌های اطلاعاتی سیستم آموزش دانشگاه تهران، به‌دست‌آمده است. پس از تحلیل داده‌ها و ارائه‌ی مدل پیش‌بینی، جدول احتمال انصراف و مدل رگرسیونی انصراف، یافته‌های پژوهش ترم‌های اول و دوم (به‌ویژه ترم اول در رده‌ی سنی ۳۱-۲۴ سال) را به‌منزله‌ی پرخطرترین دوره‌ی زمانی، دانشجویان ارشد را مستعدترین مقطع و دوره‌ی شبانه را پرخطرترین دوره‌ی تحصیلی برای انصراف دانشجو (روی‌گردانی مشتری) شناسایی کرده‌اند. مرادی و نصیری (۱۳۹۰) دانش‌آموزان دوره دبیرستان را به‌عنوان حوزه مورد مطالعه خود انتخاب کرده‌اند. در این حوزه یکی از چالش‌های جدی مدیران پیش‌بینی دوره فارغ‌التحصیلی دانش‌آموزان در پایان سال تحصیلی به‌منظور شناسایی دانش‌آموزانی است که

دچار افت تحصیلی شده و ادامه تحصیل آن‌ها با مشکل روبرو خواهد شد. در این تحقیق و با استفاده از فن‌های داده‌کاوی، وضعیت تحصیلی آتی دانش‌آموزان شامل دوره تحصیلی از تحصیلات متوسطه و وضعیت فارغ‌التحصیلی از مدارس روزانه یا بزرگسال پیش‌بینی شده است. در این تحقیق برای ساخت مدل‌های موردنظر از تکنیک‌های مختلفی نظیر درخت تصمیم و شبکه‌های عصبی استفاده شده است. نتایج به‌دست‌آمده از این تحقیق به مدیران آموزشی در جهت ارائه مشاوره لازم برای پیش‌گیری از رسیدن دانش‌آموزان به وضعیت بحرانی کمک شایان توجهی داشته است. نامنی و طالع‌پسند (۱۳۹۴) نیز در یک مطالعه جامع اقدام به شناسایی شاخص‌های پذیرشی مؤثر در موفقیت تحصیلی دانشجویان دوره‌های کارشناسی ارشد در رشته‌های مهندسی و مدیریت در دوره‌های آموزش الکترونیکی و تدوین آن‌ها در قالب یک مدل پیش‌بینی نموده‌اند. برای این منظور از طرح مطالعه مبتنی بر همبستگی استفاده نموده‌اند. معیارهایی هم‌چون شیوه ورود به دوره کارشناسی ارشد، نوع دانشگاه محل تحصیل مقطع کارشناسی، نوع رشته تحصیلی مقطع کارشناسی و معدل کل مقطع کارشناسی به‌عنوان متغیرهای پیش‌بینی و معدل کل واحدهای اخذ شده در مقطع کارشناسی ارشد به‌عنوان متغیر ملاک جمع‌آوری شده و داده‌های مذکور با استفاده از یک مدل رگرسیونی تحلیل شده‌اند. نتایج مطالعه انجام‌شده توسط نامنی و طالع‌پسند حاکی از آن است که از میان متغیرهای پیش‌بین واردشده در تحلیل، رشته تحصیلی مقطع کارشناسی (مهندسی)، پذیرش از طریق آزمون اختصاصی (با وزن منفی)، معدل کل مقطع کارشناسی، نوع دانشگاه محل تحصیل مقطع کارشناسی (دولتی) و رشته تحصیلی مقطع کارشناسی (علوم پایه)، به ترتیب، به‌عنوان مؤثرترین عوامل یا پیش‌بینی‌کننده‌های موفقیت در مقطع کارشناسی ارشد شناسایی‌شده‌اند. در ادامه، مجموعه الگوریتم‌های داده‌کاوی به کار گرفته شده در مطالعات داده‌کاوی آموزشی نمایش داده شده است (جدول ۱). همان‌طور که مشاهده می‌شود، الگوریتم‌های مبتنی بر شبکه‌های عصبی و درخت تصمیم‌گیری بیشترین کاربرد را در شناسایی الگوهایی موجود در داده‌های مطالعات آموزشی به خود اختصاص داده‌اند. علاقه‌مندان به مطالعه بیشتر پیرامون مطالعات انجام شده در حوزه داده‌کاوی آموزشی می‌توانند به مطالعه مروری انجام‌شده توسط رومرو و ونتارا (۲۰۰۷) مراجعه نمایند.

با توجه به مطالعات انجام‌شده در ادبیات موضوع داده‌کاوی آموزشی، می‌توان دریافت نمود علی‌رغم اهمیت و کارکرد تکنیک داده‌کاوی، همچنان مطالعات اندکی در این حوزه انجام شده است و هم‌چنان نیاز به ارائه و توسعه مدل‌های بومی برای مراکز آموزش عالی در سطح کشور احساس می‌شود. مدل داده‌کاوی توسعه داده شده با استفاده از نرم‌افزار آر پیاده‌سازی خواهد شد. نرم‌افزار آر یک نرم‌افزار با دسترسی رایگان است که به‌راحتی در دسترس افراد قرار می‌گیرد. هم

چنین این نرم‌افزار دارای کتابخانه‌های مختلفی است که توسط کاربران مختلف آن آماده و بر روی سایتی که برای این منظور تهیه شده است در دسترس عموم قرار می‌گیرد. از این‌رو می‌توان به‌راحتی از تجارب سایر کاربران این نرم‌افزار در پروژه‌های تحقیقاتی استفاده کرد. علاوه بر این مسئله می‌توان مشکلات احتمالی ایجاد شده را با سایر کاربران حرفه‌ای نرم‌افزار به اشتراک گذاشته و از تجارب آن‌ها استفاده کرد. از سوی دیگر با توجه به کنسول کد نویسی این نرم‌افزار، کاربر آن به‌راحتی قادر است هر نوع تغییر و یا خواسته‌ای را در پروژه خود ایجاد کند. به‌عبارت‌دیگر این امکان وجود دارد تا علاوه بر پیاده‌سازی توابع آماری و داده‌کاوی، اینترفیس مناسب برای ورود داده‌ها طراحی و ساخته شود. هم‌چنین می‌توان انواع ساختار داده (خروجی انواع دستگاه‌های اطلاعاتی) هم‌چون بانک اطلاعاتی اکسس و اِس. کیو. اِل و یا فایل‌های اکسل و متنی را دریافت کرده، عمل پالایش و پیش‌پردازش را انجام داد. به‌عبارت دیگر به بهترین نحو می‌توان ورودی مدل داده‌کاوی را تولید کرد. درنهایت می‌توان گفت با توجه به هسته مرکزی این نرم‌افزار، زمان اجرای قطعه کدهای طراحی‌شده بسیار سریع است. این خاصیت در پروژه‌های داده‌کاوی که در آن حجم بالایی از داده‌ها مورد تحلیل قرار می‌گیرد بسیار کارآمد است.

جدول ۱. الگوریتم‌های داده‌کاوی استفاده‌شده در مطالعات داده‌کاوی آموزشی

الگوریتم داده‌کاوی	فروانی استفاده	منبع
درخت تصمیم‌گیری	۱۸	وانگ (۲۰۰۵)؛ اسمیت (۲۰۰۵)؛ هرزوک (۲۰۰۶)، ساپربای و همکاران (۲۰۰۶)؛ ساجیتپاراپیتا (۲۰۰۶)؛ دلوری و همکاران (۲۰۰۸)؛ پیتمن (۲۰۰۸)؛ رومرو و همکاران (۲۰۰۹)؛ ناندشوار و چادری (۲۰۰۹)؛ دکر و همکاران (۲۰۰۹)؛ ژانگ و همکاران (۲۰۱۰)؛ وای یو و همکاران (۲۰۱۰)؛ کواچیچ (۲۰۱۰)؛ دلن (۲۰۱۰)؛ برادواج و پال (۲۰۱۱)؛ گویال و وهرا (۲۰۱۲)؛ کبکچیوا (۲۰۱۳)؛ احمدی و همکاران (۱۳۹۴)
شبکه‌های عصبی	۱۳	نپلاوا و اسنورک (۲۰۰۳)؛ بیکر و همکاران (۲۰۰۴)؛ وانگ (۲۰۰۵)؛ هرزوک (۲۰۰۶)؛ ساپربای و همکاران (۲۰۰۶)؛ ساجیتپاراپیتا (۲۰۰۶)؛ پیتمن (۲۰۰۸)؛ رومرو و همکاران (۲۰۰۹)؛ لیکورنزو و همکاران (۲۰۰۹)؛ وای یو و همکاران (۲۰۱۰)؛ دلن (۲۰۱۰)؛ کبکچیوا (۲۰۱۱)؛ بوگارد و همکاران (۲۰۱۲)
رگرسیون	۸	هرزوک (۲۰۰۵) و (۲۰۰۶)؛ وانگ (۲۰۰۵)؛ ساجیتپاراپیتا (۲۰۰۶)؛ پیتمن (۲۰۰۸)؛ دکر و همکاران

		(۲۰۰۹)؛ دلن (۲۰۱۰)؛ بوگارد و همکاران (۲۰۱۳)؛ نامنی و طالع‌پسند (۱۳۹۴)
ماشین بردار پشتیبان	۳	بیکر و همکاران (۲۰۰۴)؛ پیتمن (۲۰۰۸)؛ دکر و همکاران (۲۰۰۹)؛ لیکورنزو و همکاران (۲۰۰۹)؛ ژانگ و همکاران (۲۰۱۰)
بیزین	۳	پیتمن (۲۰۰۸)؛ دکر و همکاران (۲۰۰۹)؛ ژانگ و همکاران (۲۰۱۰)
جنگل تصادفی	۲	ساپربای و همکاران (۲۰۰۶)؛ دکر و همکاران (۲۰۰۹)
تک قانون	۱	دکر و همکاران (۲۰۰۹)؛ کبکچیوا و همکاران (۲۰۱۲)
تحلیل تشخیص خطی	۱	ساپربای و همکاران (۲۰۰۶)
الگوریتم‌های فرا ابتکاری نظیر الگوریتم ژنتیک	۱	رومرو و همکاران (۲۰۰۹)
*K*امین نزدیک‌ترین همسایه	۱	کبکچیوا (۲۰۱۲)
مارس	۱	وای یو و همکاران (۲۰۱۰)

روش‌شناسی پژوهش

در این مقاله به‌منظور استفاده از تکنیک‌های داده‌کاوی در تحلیل داده‌های مربوط به دانشجویان، در گام اول اقدام به طراحی یک فرآیند کلان برای دستیابی به داده‌ها و درنهایت تحلیل آن شده است. فرآیند استفاده شده در این رابطه در شکل ۲ نمایش داده شده است. در هر یک از مراحل این فرآیند از تکنیک‌ها و مدل‌های توسعه‌یافته مبتنی بر ادبیات موضوع داده‌کاوی استفاده شده است. در ادامه هر یک از مراحل این فرآیند تشریح خواهد شد.

مرحله ۱ (انتخاب منابع داده): در این مرحله محققین اقدام به انتخاب منابع داده خود از بین منابع مختلف موجود در دانشگاه کرده‌اند. در این ارتباط با توجه به نمونه موردی تحت بررسی، بانک اطلاعاتی سیستم آموزشی جامع دانشگاه صنعتی شاهرود به‌عنوان منبع داده‌های مورد استفاده در این تحقیق انتخاب شده است. جامع و کامل بودن بانک اطلاعات این سیستم آموزشی و نیز محدودیت دسترسی به بانک‌های اطلاعاتی دیگر موجود در دانشگاه منجر به این انتخاب شده است.

شکل ۲. فرآیند طراحی‌شده برای داده‌کاوی

مرحله ۲ (انتخاب و گردآوری داده): پس از مشخص شدن پایگاه داده اطلاعاتی در دسترس، اقدام به انتخاب داده‌های مربوط به دانشجویان از این بانک اطلاعاتی مدنظر شده است. در این ارتباط و با توجه به محدوده تحقیق، اطلاعات دانشجویان رشته‌های مدیریت و حسابداری در مقاطع کارشناسی و کارشناسی ارشد از دانشکده مدیریت در خلال سال‌های ۱۳۸۴ تا ۱۳۹۴ گردآوری شده است. دلیل انتخاب این محدوده کامل نبودن و پراکنده بودن داده‌های دانشجویان پیش از سال ۱۳۸۴ است. علاوه بر این مسئله، تغییرات گسترده‌ای که در نظام آموزشی پذیرش دانشجو در سال‌های اخیر ایجاد شده است، محدوده مناسب برای انتخاب داده‌های را کوچک‌تر کرده است. پس از انتخاب محدوده داده‌های موردنظر، اطلاعات تحصیلی و کلی دانشجویان در سه بخش کلی به شرح زیر گردآوری شد.

- اطلاعات جمعیت شناختی دانشجویان ازجمله سال و محل تولد، سطح درآمد خانواده، مجرد و یا متأهل بودن، تعداد فرزندان و موارد مشابه دیگر.
- اطلاعات تحصیلی پس از ورود به دانشگاه صنعتی شاهرود ازجمله رشته دیپلم و یا کارشناسی، نحوه ورود به دانشگاه (کنکور، استعداد درخشان، انتقالی) و جزییات آن (معدل دیپلم، رتبه کنکور ورودی، منطقه امتحانی، ریزنمرات کنکور ورودی به دانشگاه و موارد مشابه دیگر).
- اطلاعات تحصیلی پس از ورود به دانشگاه صنعتی شاهرود ازجمله ریزنمرات ترم‌های تحصیلی، معدل کل و هر ترم دانشجویان، تعداد مشروطی‌های متوالی و غیر متوالی.

اطلاعات موردنظر در این بخش در سه بانک اطلاعاتی مجزا در اختیار محققین قرار داده شد. با ایجاد دسترسی به داده‌های مزبور امکان بررسی دقیق‌تر و درنهایت اعمال مدل‌های داده‌کاوی بروی آن فراهم گردید.

مرحله ۳ (آماده‌سازی داده): با توجه به گردآوری داده‌ها در سه بانک اطلاعات مجزا لازم است تا داده‌ها تجمیع شده و در صورت وجود داده‌های ناقص از بانک اطلاعات حذف شده و یا تکمیل شوند. این عمل از طریق کلید اصلی شماره دانشجویی دانشجویان انجام شده است. در این مرحله

با تجمیع اولیه داده‌ها مشاهده گردید که بخشی از داده‌ها (عمدتاً اطلاعات پیش از ورود به دانشگاه) برای تعدادی از دانشجویان دارای نقص بوده و امکان استفاده از آن وجود ندارد. در این موقعیت بانک‌های اطلاعات دانشگاه برای تکمیل این داده‌ها مجدداً مورد بررسی قرار گرفت که مشاهده گردید امکان تکمیل اطلاعات دارای نقص وجود ندارد. از این‌رو اطلاعات این دانشجویان از بانک اطلاعاتی نهایی حذف گردید. پس از این عمل، داده‌ها مجدداً مورد بررسی قرار گرفت. در این مرحله متغیر محل تولد و زندگی دانشجو مورد بررسی قرار گرفت. با بررسی انجام شده و به دلیل پراکندگی بسیار زیاد این متغیر (تقریباً گسترده کل کشور) این متغیر تبدیل به متغیری معنادار که امکان تحلیل‌های بعدی را ایجاد کند، شده است. برای این منظور این متغیر به متغیر بومی/ غیربومی بودن دانشجویان تبدیل شده است. پس از متغیر مربوط به محل زندگی، محققین بروی اطلاعات پیش از ورود دانشجویان به دانشگاه صنعتی شاهرود تمرکز نموده‌اند. اهمیت این بررسی از آنجایی ایجاد می‌شود که داده‌های موجود در بانک اطلاعات شامل دانشجویان کارشناسی و کارشناسی ارشد می‌شود. از این‌رو آزمون‌ها و درنهایت نتیجه آزمون‌ها قابل‌مقایسه با یکدیگر نیست. ازاین‌رو محققین اقدام به حذف جزئیات امتحانات ورودی به دانشگاه کرده و تنها از نمره نهایی آزمون ورودی همانند رتبه کنکور ورودی و نیز معدل دوره تحصیلی قبل از ورود به دانشگاه صنعتی شاهرود در تحلیل‌های نهایی خود استفاده نموده‌اند. پس از نهایی کردن جزئیات داده‌های پیش از ورود به دانشگاه صنعتی شاهرود، نوبت به داده‌های حین تحصیل دانشجویان در دانشگاه صنعتی شاهرود می‌رسد. برای این منظور، در اولین گام داده‌های یکتا به ازای هر دانشجو حذف گردید تا از ورود آن به مدل نهایی جلوگیری شود. در این بخش فیلدهای شماره دانشجویی و شماره ملی حذف شده است. در این بخش نیز همانند داده‌های مربوط به پیش از ورود به دانشگاه، اطلاعات جزئی دروس گذرانده در طول ترم‌های مختلف و نمرات اخذ شده به دلیل متفاوت بودن در بین رشته‌های مختلف امکان مقایسه وجود ندارد. از این‌رو به‌منظور همسنگ کردن داده‌ها، از ورود داده‌های جزئی جلوگیری شده و به‌جای آن از متغیر معدل دانشجویان، وضعیت مشروطی دانشجویان به‌صورت متوالی و غیرمتوالی استفاده شده است. با انجام اصلاحات شرح داده‌شده امکان مقایسه بین رکوردهای داده‌های موجود در پایگاه داده نهایی فراهم شده است.

مرحله ۴ (اعمال مدل داده‌کاوی بر داده‌ها): پس از نهایی شدن داده‌ها امکان اعمال مدل‌های داده‌کاوی بروی آن‌ها فراهم شده است. در این بخش از دو مدل مختلف استفاده شده است. در مرحله اول با انجام خوشه‌بندی بروی داده‌ها امکان مقایسه درون و بین خوشه‌ای بروی دانشجویان و متغیرهای تأثیرگذار بروی آن فراهم می‌شود. در مرحله دوم از روش‌های مبتنی بر درخت تصمیم‌گیری برای ایجاد قوانین و رویه‌های تصمیم‌گیری استفاده شده است. درخت

تصمیم‌گیری در کنار اطلاعات به‌دست‌آمده از خوشه‌های مختلف دانشجویان کمک شایانی به سیاست‌گذاری و اتخاذ تصمیمات مدیریتی در سطح دانشگاه خواهد کرد. مدل‌های استفاده‌شده در این بخش به شرح زیر است.

- خوشه‌بندی داده‌ها: به‌منظور انجام خوشه‌بندی بروی داده‌های مورد بررسی لازم است تا در گام اول تعداد خوشه‌های مناسب برای تحلیل تعیین گردد. برای این منظور می‌توان از نمودار آرنج استفاده کرد. در این نمودار نقطه شکست نشان‌دهنده تعداد خوشه‌های مناسب برای تحلیل است. در ادامه و پس از تعیین تعداد خوشه‌ها می‌توان مدل خوشه‌بندی را بر داده‌ها اعمال کرد. در این تحقیق از روش خوشه‌بندی سلسله‌مراتبی استفاده شده است. در این روش داده‌ها در مرحله اول همگی در یک خوشه قرار داده شده و سپس به‌صورت مرحله‌ای داده‌ها در خوشه‌های مختلف قرار داده می‌شوند (روکاخ و میمون، ۲۰۰۵)؛ باوجود مزایای بسیار زیاد این روش، زمان‌بر بودن و نیز حساسیت بسیار زیاد خروجی آن نسبت به وجود داده‌های پرت دقت خروجی این روش را تحت تأثیر قرار می‌دهد؛ از سوی دیگر حجیم بودن حجم داده‌های مورد استفاده در این تحقیق و نیز وجود داده‌های به‌ظاهر پرت (دانشجویان بسیار ضعیف و یا بسیار قوی) نگرانی نسبت نتایج این الگوریتم را ایجاد می‌کند. به‌منظور رفع این مشکل از الگوریتم پیشنهادی توسط گاگولوسکی و همکاران (۲۰۱۶) استفاده شده است.
- استخراج قانون: پس از انجام خوشه‌بندی و تحلیل‌های متناسب با آن لازم است تا با استفاده از روش‌های موجود در استخراج قوانین، قواعد و قانون‌های برای ایجاد تصمیم‌گیری و سیاست‌گذاری‌های مدیریت بیان شود. این امر با استفاده از خروجی‌های روش خوشه‌بندی سلسله‌مراتبی ممکن خواهد شد، چراکه در این روش درنهایت یک ساختار درختی متشکل از خوشه‌های مختلف ایجاد می‌شود؛ ازاین‌رو می‌توان از ساختار مربوطه برای بیان دقیق قوانین و قواعد حاکم بر دسته‌بندی دانشجویان استفاده کرد.

۴. مدل داده‌کاوی پیشنهاد برای تحلیل داده‌های آموزشی دانشجویان

در این تحقیق به‌منظور خوشه‌بندی اطلاعات دانشجویان از روش خوشه‌بندی سلسله مراتبی استفاده شده است (گاگولوسکی و همکاران، ۲۰۱۶). در روش خوشه‌بندی سلسله مراتبی از دو رویکرد زیر می‌توان استفاده کرد.

- رویکرد بالا به پایین: در این رویکرد تمامی مشاهدات در یک خوشه قرار داده شده، سپس با مقایسه بین آن‌ها عمل تفکیک و ایجاد خوشه‌های ریزتر انجام می‌شود.

- رویکرد پایین به بالا: در این رویکرد هر یک از مشاهدات یک خوشه را ایجاد کرده، سپس با عمل مقایسه بین آن‌ها عمل ادغام و کاهش تعداد خوشه‌ها انجام می‌شود.

در روش الگوریتم خوشه‌بندی سلسله‌مراتبی نیاز به یک روش برای اندازه‌گیری فاصله (تفاوت) بین مشاهدات و نیز یک رویکرد مشخص برای تعیین میزان تفاوت بین خوشه‌ها براساس تفاوت بین مشاهدات وجود دارد. در ادبیات روش‌ها و معیارهای مختلفی برای این دو نیازمندی ارائه شده است. برای مثال روش‌های فاصله اقلیدسی، مجذور فاصله اقلیدسی، فاصله منهتن، فاصله ماهالنوبیس برای اندازه‌گیری فاصله و روش‌های کمترین فاصله، بیشترین فاصله، فاصله میانگین، فاصله مرکز خوشه برای ارزیابی تفاوت بین خوشه‌ها ارائه شده است. روش خوشه‌بندی سلسله مراتبی روشی زمان‌بر است، به‌گونه‌ای که رویکرد بالا به پایین دارای درجه پیچیدگی $O(2^n)$ و رویکرد پایین به بالا دارای درجه پیچیدگی $O(n^3)$ است. روش مورد استفاده در اندازه‌گیری فاصله تفاوت بین مشاهدات و نیز تفاوت بین خوشه‌ها به‌صورت مستقیم بر کیفیت و سرعت این الگوریتم اثرگذار است. از سوی دیگر وجود داده‌های پرت در بین مشاهدات می‌تواند نتیجه نهایی را به‌صورت قابل‌توجه تحت تأثیر خود قرار دهد؛ ازاین‌رو لازم بود تا نسبت به انتخاب معیار مناسب برای محاسبه تفاوت و فاصله بین مشاهدات و بین خوشه‌ها اقدام شود. به‌منظور رفع این مشکل از رویکرد معرفی‌شده توسط گاگولوسکی و همکاران (۲۰۱۶) استفاده شده است. در این رویکرد به‌منظور فائق آمدن بر وجود داده‌های پرت و نیز افزایش دقت و سرعت محاسبات در دسته‌بندی از شاخص جینی برای مقایسه بین خوشه‌ها و درنهایت ایجاد تصمیم برای ترکیب آن‌ها استفاده شده است. با وجود نوآوری انجام‌شده در این تحقیق بروی معیار مقایسه بین خوشه‌ای و ایجاد تصمیم برای ترکیب آن‌ها، در این تحقیق بحثی بروی چگونگی محاسبه بین مشاهدات انجام نشده است. به‌منظور حل این مسئله در هنگام اجرای این الگوریتم از چهار رویکرد شامل؛ فاصله اقلیدسی، مجذور فاصله اقلیدسی، فاصله منهتن و فاصله ماهالنوبیس در محاسبه فاصله بین مشاهدات استفاده شد که درنهایت روش مجذور فاصله اقلیدسی بهترین نتیجه را برای محققین حاصل نمود. این مقایسه با استفاده از دو معیار آنتروپی و ضریب واریانس انجام شده و نتایج آن در جدول ۲ آورده شده است.

جدول ۲. مقایسه روش‌های اندازه‌گیری فاصله بین مشاهدات

معیار	فاصله اقلیدسی	مجذور فاصله اقلیدسی	فاصله منهتن	فاصله ماهالنوبیس
آنتروپی	۰/۳۵	۰/۳۲	۰/۴۱	۰/۳۹
ضریب واریانس	۰/۳۲	۰/۳۳	۰/۳۸	۰/۳۵

در این تحقیق، محققین به‌منظور اطمینان از مناسب بودن روش انتخابی در خوشه‌بندی داده‌ها، روش پیشنهادی در این بخش را با چند روش دیگر نیز مورد مقایسه قرار داده که درنهایت مناسب بودن روش مزبور مشخص گردید. لازم به ذکر است در این بررسی مواردی که نیاز به مشخص کردن تعداد خوشه‌ها پیش از آغاز الگوریتم وجود داشت، تعداد خوشه‌ها از روش‌های سلسله مراتبی استخراج و در نظر گرفته می‌شود. همانند مقایسه مربوط به روش اندازه‌گیری فاصله مقایسه بین روش‌های خوشه‌بندی نیز با استفاده از دو معیار آنتروپی و ضریب واریانس انجام شده است (جدول ۳).

جدول ۳. مقایسه روش‌های خوشه‌بندی

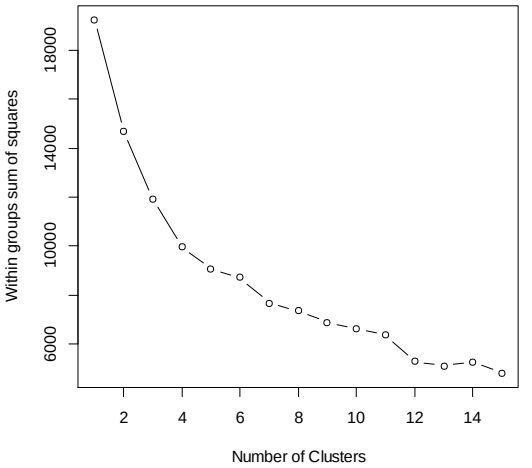
معیار	سلسله مراتبی بازبینی‌شده	کامینز	حداقل درخت گسترده
آنتروپی	۰/۳۲	۰/۶۷	۰/۴۹
ضریب واریانس	۰/۳۳	۰/۴۸	۰/۵۴

یافته‌های پژوهش

همان‌طور که پیش‌ازاین نیز بیان گردید، به‌منظور آغاز ارزیابی داده‌های به‌دست آمده از سیستم آموزشی دانشگاه در گام اول سه سری داده به‌دست آمده در یک فایل اطلاعاتی تجمیع شده و پس از آن اقدام به حذف فیلدهای تکراری و نیز دانشجویان غیرمرتبط در این پروژه شده است. در این پروژه تنها دانشجویان رشته‌های حسابداری و مدیریت در نظر گرفته شده‌اند. با توجه به اینکه تعدادی از مشخصه‌های موجود در بانک اطلاعاتی تهیه شده به‌صورت یکتا برای هر دانشجو بوده (همانند شماره دانشجویی) و نیز دارای رنج تغییرات بسیار زیاد هستند (همانند شهر محل تولد) لازم است تا وجود چنین متغیرهایی بررسی‌شده تا نسبت به حذف متغیرهای مزبور اقدام شود. برای این عمل از روش تحلیل مؤلفه‌های اصلی استفاده شد. پس از اجرای تحلیل مؤلفه‌های اصلی، مشاهده گردید که متغیرهای شماره دانشجویی، شهر محل سکونت، رتبه در سهمیه و نمره اول آزمون دارای تغییرات بسیار بالایی هستند. از این‌رو متغیرهای مزبور از بانک اطلاعات حذف شده‌اند. پس از آماده شدن داده‌ها، الگوریتم خوشه‌بندی پیشنهادی بر روی داده‌ها اعمال شده است. پس از اعمال الگوریتم مورد نظر مشاهده گردید که شش خوشه مجزا در درون داده‌ها تشکیل شده است. به‌منظور اطمینان از مناسب بودن تعداد خوشه‌ها، نمودار آرنجی نیز برای داده‌ها رسم شده است (شکل ۳). با توجه به ادبیات موضوع محل ایجاد شدن آرنج، نشان‌دهنده

تعداد خوشه‌های مناسب برای داده‌ها است. همان‌طور که در این نمودار مشاهده می‌شود تعداد شش عدد خوشه برای داده‌های در دست مناسب است.

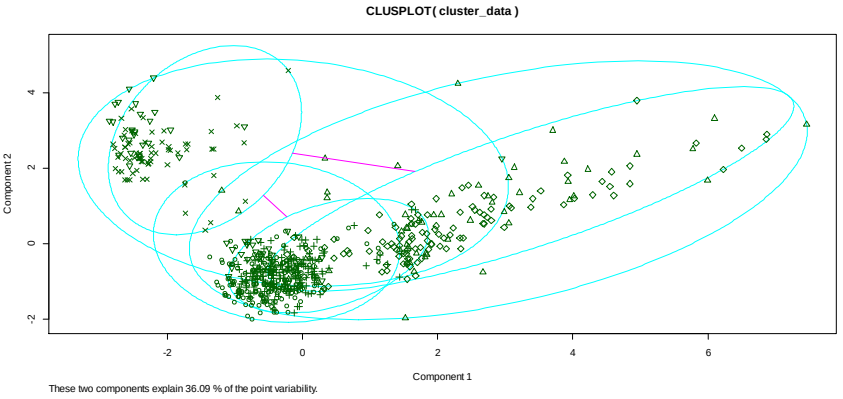
پایگاه داده مورد استفاده در این تحقیق دارای ۶۷۲ عضو است که تعداد اعضای تخصیص داده شده به هر یک از گروه‌ها به شرح جدول ۴ است. همان‌طور که مشاهده می‌شود گروه ۵ و ۱ با ۲۷۸ و ۹ عضو به ترتیب دارای بیشترین و کمترین تعداد اعضا می‌باشند. همپوشانی داده‌ها که در خوشه‌های مختلف قرار داده شده نیز در شکل ۴ نمایش داده شده است. چنانچه بخواهیم هر یک از خوشه‌های ایجادشده را نام‌گذاری کنیم، می‌تواند از شاخص معدل کل دانشجویان حاضر در هر یک از خوشه‌های استفاده شود. برای این منظور میانگین معدل هر یک از خوشه‌ها باید محاسبه شود. نتایج این محاسبه در جدول ۵ آورده شده است.



شکل ۳. نمودار تعیین تعداد خوشه‌ها (نمودار آرنجی)

جدول ۴. تعداد اعضای هر یک از خوشه‌ها

شماره خوشه	تعداد اعضا	شماره خوشه	تعداد اعضا
۱	۹	۴	۷۰
۲	۱۹۹	۵	۲۷۸
۳	۵۸	۶	۵۸



شکل ۴. همپوشانی داده‌ها در خوشه‌های مختلف

جدول ۵. میانگین معدل دانشجویان هر خوشه

شماره خوشه	میانگین معدل	تعداد اعضای هر خوشه	شماره خوشه	میانگین معدل	تعداد اعضای هر خوشه
۱	۱۶/۳۸۱	۹	۴	۱۶/۹۴۲	۷۰
۲	۱۳/۶۶۴	۱۹۹	۵	۱۶/۱۱۹	۲۷۸
۳	۱۸/۱۵۲	۵۸	۶	۹/۲۴۳	۵۸
کل	۱۵/۰۶۴	۶۷۲			

براساس میانگین معدل محاسبه‌شده می‌توان گفت خوشه ۳ دانشجویان عالی، خوشه‌های۱، ۴ و ۵ دانشجویان خوب، خوشه ۲ دانشجویان متوسط و خوشه ۶ دانشجویان ضعیف را تشکیل داده‌اند. در این تقسیم‌بندی اولین سؤالی که ایجاد می‌شود وجه تمایز خوشه ۱ با خوشه‌های ۴ و ۵ است. برای این منظور سایر ویژگی‌های دانشجویان مورد بررسی قرار می‌گیرد. با بررسی انجام شده می‌توان مشاهده کرد دانشجویان خوشه ۴ دانشجویان غیربومی بوده، اما دانشجویان خوشه ۵ غالباً دانشجویان بومی هستند. این تفاوت در بین خوشه‌های ۱ و ۶ نیز وجود دارد. با بررسی داده‌ها مشخص می‌شود دانشجویان خوشه ۱ دانشجویان انتقالی از سایر دانشگاه‌ها به دانشگاه صنعتی شاهرود می‌باشند. از سوی دیگر چنانچه همراه با این خوشه دانشجویان موجود در گروه ۶ نیز در نظر گرفته شود مشاهده می‌شود که غالب دانشجویان این خوشه نیز دانشجویان انتقالی از سایر دانشگاه‌ها و یا دانشجویان نوبت دوم هستند که دارای معدل بسیار پایینی می‌باشند. از این‌رو می‌توان گفت خوشه‌های ۱ و ۶ دربرگیرنده دانشجویان انتقالی از سایر دانشگاه و نوبت

دومی است که به دلیل وجود تفاوت معنادار در معدل آن‌ها این دو خوشه از یکدیگر جدا شده است. هم‌چنین تعدادی از دانشجویان روزانه با معدل پایین نیز در خوشه ۶ قرارگرفته‌اند. در ادامه این بررسی می‌توان مشاهده کرد که بالغ بر ۵۰ درصد دانشجویان حاضر در خوشه عالی دانشجویان پذیرفته شده در دوره‌های روزانه می‌باشند. هم‌چنین هیچ دانشجویی که با سهمیه خاص مانند ایثارگران و یا خانواده شهدا در دانشگاه پذیرفته شده است، در این خوشه قرار ندارد. این امر نشان‌دهنده اختلاف سطح علمی دانشجویان مزبور با متوسط سطح علمی دانشجویان پذیرفته‌شده در دانشگاه است که نشان‌دهنده نیازمندی آن‌ها به ارائه خدمات خاص برای بهبود سطح علمی است.

چنانچه بررسی گروه‌های ضعیف و متوسط را بروی تعداد مشروطی‌های متوالی تمرکز کند مشاهده می‌شود ۵۳ درصد دانشجویان خوشه متوسط و ۷۲ درصد دانشجویان ضعیف حداقل دو مشروطی متوالی را تجربه کرده‌اند. مشروطی متوالی از آنجایی دارایی اهمیت است که نشان‌دهنده ممتد بودن وضعیت نامناسب تحصیلی دانشجو است. در ادامه این بررسی چنانچه سال‌هایی که این دو گروه دارای بیشترین مشروطی بوده‌اند بررسی شود مشاهده می‌شود که در سال ۹۰ و ۹۱ اوج این مسئله بوده است هرچند که پس از سال ۹۱ این مسئله روندی کاهشی داشته است. در ادامه بررسی اگر سال ورود دانشجویان هر یک از خوشه‌های به‌عنوان متغیر نشان‌گر مورد استفاده قرار گیرد مشاهده می‌شود که خوشه‌های ۱، ۳، ۴ و ۵ که دانشجویان خوب و عالی هستند همگی در سال‌های ۹۰ و پس از آن وارد دانشگاه شده‌اند از سوی دیگر دانشجویانی که سال ورود آن‌ها مربوط به دهه ۸۰ است همگی در خوشه‌های ۲ و ۶ (دانشجویان متوسط و ضعیف) قرار دارند. این تفاوت معنادار نشان‌دهنده ایجاد یک تغییر مهم و اثرگذار در فضای آموزشی دانشگاه در خلال سال‌ها ۹۰، ۹۱ و پس از آن است. با مشخص شدن ویژگی‌های کلی داده‌های قرارگرفته در هر یک از خوشه‌ها لازم است تا از آن در جهت ایجاد دانش در قالب روال‌های تصمیم‌گیری استفاده شود. قوانین استخراج شده از داده‌ها در شکل ۵ ارائه شده است. علاوه بر قوانین استخراج‌شده در این بخش با توجه به تفاوت محسوس وضعیت دانشجویان قبل و بعد از سال ۹۰ در دانشگاه، لازم است تا مسئولان دانشگاه تغییرات ایجادشده سال‌های مزبور را بررسی کرده و از آن به‌عنوان تجارب برتر استفاده کنند.

شکل ۵. قوانین تصمیم‌گیری استخراج‌شده

هم‌چنین، استفاده از نتایج به‌دست‌آمده از تحلیل‌های مربوط به دسته‌بندی دانشجویان می‌توان در پیش‌بینی وضعیت دانشجویان جدیدالورود استفاده کرد. در گام اول برای بررسی این موضوع مشاهده می‌شود که نسبت دانشجویان بومی به غیربومی در دانشگاه مورد بررسی تقریباً برابر است. اما با وجود برابری تقریبی دانشجویان، مشاهده می‌شود که دانشجویان بومی به‌صورت متوسط دارای تعداد مشروطی بیشتری نسبت به دانشجویان غیربومی هستند. بر این اساس هر دانشجوی بومی ۰,۴۷ مشروطی و غیربومی ۰,۳ مشروطی هستند؛ با این وجود متوسط معدل دانشجویان بومی ۱۵/۲۳ و غیربومی ۱۴/۸۹ است. این امر با وجود بالاتر بودن نرخ مشروطی دانشجویان بومی می‌تواند زنگ خطری در خصوص وضعیت تحصیلی دانشجویان بومی باشد. در نظر گرفتن این دو متغیر در کنار یکدیگر نشان از پراکندگی بالا در بین دانشجویان بومی است. به‌عبارت‌دیگر این دو متغیر نشان‌دهنده آن است که دانشجویان مشروطی دارای تعداد مشروطی بالا اما دانشجویان غیربومی دارای تعداد مشروطی کمتر اما فراگیرتر هستند. از این‌رو دو سیاست زیر در خصوص دانشجویان بومی و غیربومی پیشنهاد می‌شود.

- دانشجویان بومی نیاز به برنامه‌ریزی عمومی نداشته و تنها دانشجویانی که برای اولین ترم مشروط می‌شوند نیاز به مراقبت بسیار بالایی دارند تا از تکرار مشروطی آن‌ها جلوگیری شود.
- دانشجویان غیربومی به‌صورت عمومی نیاز به برنامه‌ریزی دارند تا از مشروطی آن‌ها جلوگیری شود.

نتایج تحلیل نوع ورود افراد به دانشگاه براساس وضعیت بومی بودن نشان می‌دهد که بیشترین تعداد دانشجویان از طریق آزمون سنجش به دانشگاه وارد شده‌اند اما نکته دارای اهمیت، مربوط به ورود از طریق سهمیه دانش‌آموختگان نخبگان است. در میان دانشجویان بررسی‌شده تنها تعداد بسیار محدودی دانشجو از این طریق به دانشگاه وارد شده‌اند که دانشجویان مزبور بومی هستند. ازاین‌رو می‌توان گفت دانشگاه در زمینه روابط عمومی و شناساندن خود به‌عنوان یک دانشگاه با شرایط مناسب شاید چندان مؤثر و قوی عمل نکرده است. از سوی دیگر، با بررسی وضعیت تأهل دانشجویان و میانگین تعداد مشروطی و معدل آنان مشاهده می‌شود که متوسط تعداد مشروطی دانشجویان متأهل بالاتر از دانشجویان مجرد بوده و لازم است تا شرایط مناسبی برای این دسته از دانشجویان ایجاد شود. این مسئله با توجه به میانگین معدل دانشجویان متأهل جدی‌تر است؛ چرا که بالاتر بودن معدل آن‌ها نسبت به دانشجویان مجرد نشان از توان بالای این دسته از دانشجویان است، با این وجود متوسط بالای مشروطی نشان‌دهنده مشروطی‌های متوالی این دانشجویان است. درنهایت، بررسی وضعیت تحصیلی دانشجویان با توجه به مقطع تحصیلی، نشان می‌دهد که میانگین معدل دانشجویان دوره کارشناسی پایین‌تر از میانگین معدل

دانشجویان دوره کارشناسی ارشد است. هم‌چنین میانگین تعداد مشروطی‌های دانشجویان کارشناسی از دانشجویان دوره کارشناسی ارشد بالاتر است. بالاتر بودن میانگین معدل دانشجویان کارشناسی ارشد نسبت به دانشجویان کارشناسی تا حدودی قابل قبول است که درنهایت منجر به کمتر شدن تعداد مشروطی دانشجویان کارشناسی ارشد نسبت دانشجویان کارشناسی می‌شود. در این رابطه روند غیرمنطقی در بین داده‌ها مشاهده نمی‌شود؛ با این وجود تفاوت فاحشی که بین میانگین تعداد مشروطی دانشجویان آموزش‌محور و سایر دانشجویان وجود دارد می‌تواند برای مسئولین دانشگاه به‌عنوان یک زنگ خطر در نظر گرفته شود.

بحث و نتیجه‌گیری

همراه با توسعه آموزش عالی در طی سال‌های اخیر و افزایش توجه به مدیریت کارآمد و اثربخش مراکز آموزش عالی با توجه به اهمیت و جایگاه آن در توسعه منابع انسانی یک جامعه، ابزارها و تکنیک‌های تحلیلی با هدف ارتقای کیفیت تصمیمات مدیران آموزشی مورد توجه محققان و فعالان این عرصه قرار گرفته است. در این میان، تکنیک داده‌کاوی آموزشی با فراهم نمودن امکان تحلیل حجم انبوه داده‌ها در یک مرکز آموزش عالی به‌خوبی می‌تواند رفتارها و الگوهای پنهان در متغیرهای مدنظر را شناسایی نموده و زمینه را برای تحلیل و تصمیم‌سازی درست فراهم کند. ازاین‌رو در این مقاله، با استفاده از تکنیک داده‌کاوی آموزشی سعی شده است تا مسئله مهم تحلیل وضعیت تحصیلی دانشجویان در طی دوران تحصیل و ارائه قواعد تصمیم‌سازی جهت بهبود کیفیت کلی آن مورد بررسی و ارزیابی قرار گیرد. برای این منظور از مجموعه‌ای از تکنیک‌های توسعه‌یافته داده‌کاوی هم‌چون خوشه‌بندی سلسله‌مراتبی استفاده شده است. جهت تحلیل و پیاده‌سازی تکنیک داده‌کاوی پیشنهادی، از اطلاعات دانشجویان رشته‌های مدیریت و حسابداری دانشگاه صنعتی شاهرود به‌عنوان نمونه موردی استفاده شده است. نتایج تحلیل حاکی از اثرگذاری معنادار برخی از ویژگی‌های دانشجویان مبتنی بر وضعیت آنان به هنگام ورود به دانشگاه و حین تحصیل بر وضعیت تحصیلی آنان است. بر این اساس، دانشجویان در خوشه‌های متفاوت از نظر شاخص وضعیت تحصیلی قرار می‌گیرند که با توجه به مجموعه ویژگی‌های غالب هر دسته می‌توان اقدام به معرفی راه‌کارهای مناسب برای مدیریت وضعیت تحصیلی هرکدام از خوشه‌ها نمود. هم‌چنین، مجموعه‌ای از قواعد تصمیم‌گیری جهت مدیریت آینده تحصیلی دانشجویان و بهبود وضعیت آنان و اجتناب از ورود به مناطق پرخطر پیشنهاد شده است. با توجه به ارزیابی مطلوب خبرگان آموزشی از کیفیت نتایج، اعتبار نتایج مورد بررسی و تأیید قرار گرفته است. درنهایت، با توجه به اثرگذاری وضعیت تحصیلی دانشجویان بر وضعیت آینده شغلی آنان پس از فارغ‌التحصیل شدن از دانشگاه، به‌عنوان فرصت تحقیقات آتی پیشنهاد می‌شود که علاوه

بر لحاظ نمودن اطلاعات قبل و بعد از ورود به دانشگاه، اطلاعات شغلی و وضعیت زندگی پس از فارغ‌التحصیلی نیز در تحلیل‌های انجام‌شده موردبررسی قرار گیرد. به‌طور کلی برخی از نتایج کلیدی پیاده‌سازی مدل داده‌کاوی پیشنهادی جهت تحلیل اطلاعات آموزشی دانشجویان عبارت است از:

- معدل کل دانشجویان می‌تواند شاخص مناسبی جهت ارزیابی و پیش‌بینی وضعیت تحصیلی دانشجویان معرفی شود. با استفاده از این شاخصه کلیدی در کنار دیگر شاخصه‌های توصیف‌کننده دانشجو، می‌توان قوانین اجرایی مناسبی جهت ارتقای اثربخشی برخی از تصمیمات آموزشی مانند پذیرش انتقالی و برنامه‌ریزی تحصیلی دانشجویان عادی و یا در آستانه افت تحصیلی ارائه نمود. برای نمونه، دانشجویان بومی نیاز به برنامه‌ریزی خاص نداشته و تنها دانشجویانی که برای اولین ترم مشروط می‌شوند نیاز به مراقبت بسیار بالایی دارند تا از تکرار مشروطی آن‌ها جلوگیری شود. دانشجویان غیربومی به‌صورت عمومی نیاز به برنامه‌ریزی خاص دارند تا از مشروطی احتمالی آن‌ها در طی تحصیل جلوگیری شود. همچنین بر اساس تحلیل‌های انجام شده معدل ۱۳/۵ به‌عنوان حد هشدار بوده و لازم است تا متولیان برنامه‌ریزی آموزشی دانشگاه تدابیر و اقدامات محدودکننده تعریف شده برای دانشجویان مشروطی (با معدل کمتر از ۱۲) را به‌صورت تعدیل شده برای دانشجویان با معدل کمتر از ۱۳/۵ نیز اعمال نمایند. هم‌چنین برآورد می‌شود که ارتقای معدل دانشجو به مقدار ۱۶ و بالاتر از آن، می‌تواند به معنای پیشرفت تحصیلی دانشجو و رشد و شکوفایی وی در آینده تحصیلی باشد. از این‌رو لازم است برای این دسته از افراد نیز مکانیزم‌های تشویقی-آموزشی در نظر گرفته شود. در این میان دانشجویان غیربومی با معدل بالاتر از ۱۶ مستعد پیشرفت و شکوفایی بیشتر بوده و می‌توانند در برنامه‌های آتی دانشگاه دارای جایگاه ویژه‌ای باشند.
- برخی از ویژگی‌ها همانند وضعیت بومی یا غیربومی بودن، نحوه پذیرش (مانند انتقالی از سایر دانشگاه‌ها و یا استفاده از سهمیه خاص)، نوع قبولی (روزانه و شبانه)، وضعیت تأهل و مقطع تحصیلی، تأثیر معناداری بر وضعیت تحصیلی دانشجویان و قرار گرفتن آنان در دسته‌های مختلف وضعیت تحصیلی مانند عادی و یا پرخطر خواهد داشت. برای نمونه، دانشجویان کارشناسی ارشد به میزان کمتری درخطر افت تحصیلی نسبت به مقطع کارشناسی قرار دارند و تنها دسته‌های خاصی از آنان نظیر دانشجویان آموزش‌محور از ریسک افت تحصیلی بالاتر برخوردار هستند. از این‌رو نیاز است تا در برنامه‌ریزی آموزشی و هدایت تحصیلی گروه‌های مذکور، توجه خاصی را مبذول نمایند. و یا با توجه به آسیب‌پذیری بیشتر دانشجویان متأهل از نظر وضعیت تحصیلی، بسته‌های حمایتی دانشگاه

مانند ارائه تسهیلات مالی مناسب و موارد مشابه دیگر، در کنار توجه خاص آموزشی به ایـن گروه می‌تواند از اثربخشی بالاتری برخوردار باشد.

منابع

Ahmadi, S., Karimzadgan, D., Khairati Kazerooni, T. (2015). Data Mining Of Students Withdrawal at University of Tehran, Focusing on Fee Paid Students (To Prevent Customer Churn). *Journal of Information Technology Management*, 7(2), 217-238 .

Baker, R. & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1): 3-17.

Baradwaj, B.K. and Pal, S. (2011). Mining Educational Data to Analyze Students Performance. *International Journal of Advanced Computer Science and Applications*, 2, 6, 63-71.

Barker, K., Trafalis, T. & Rhoads, T. (2004). IEEE Systems and Information Engineering Design Symposium, University of Oklahoma.

Beck, J. & Mostow, J. (2008). A Case Study Empirical Comparison of Three Methods. *9th international conference on Intelligent Tutoring Systems*, June 23-27, Montreal Canada.

Bienkowski, M., Feng M. & Means B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Available in: http://www.cra.org/ccc/files/docs/learning-analytics-ed.pdf.

Bogard, M. (2013). A Data Driven Analytic Strategy for Increasing Yield and Retention at Western Kentucky University Using SAS Enterprise BI and SAS Enterprise Miner. Available in: http://support.sas.com/resources/papers/ proceedings13/044-2013.pdf.

Bogard, M., James, C., Helbig, T. & Huff, G. (2012). Using SAS® Enterprise BI and SAS®Enterprise Miner TM to Reduce Student Attrition. SAS Conference Proceedings: SAS Global Forum 2012, April 22-25, Orlando, Florida.

Burez, J. & Van den Poel, D. (2007). Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2): 277–288.

Buttle, F. (2008). Customer Relationship Management, UK, Routledge. Campbell, J.P., Oblinger, D.G. (2007). Academic Analytics. Available in: http://net.educase.edu/ir/library/pdf/PUB6101.pdf.

Dekker, G.W., Pechenizkiy, M. & Vleeshouwers J.M. (2009). Predicting student drop out: A case study. 2nd International Educational Data Mining Conference, July 1-3, Cordoba Spain.

Delavari, N., Phon-Amnuaisuk, S., Beikzadeh, M. R. (2008). Data Mining Application in Higher Learning. *Informatics in Education*, 7(1): 31–54.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49 (4): 498-506.

Eesaee Cheshme Mahi., M., Bagherian., F., & Hasanzadeh Tavakkoli., M.R. (2011). Relation between personality dimensions with students' academic performance.

Biennial Journal of Management and Planning in Educational Systems, 4, 6, 115-130.

Gagolewski, M., Bartoszuk, M. and Cena, A. (2016). Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. Information Sciences.

Gagolewski, M., Bartoszuk, M., & Cena, A. (2016). Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. *Information Sciences*, 363, 8–2.

García, E., Romero, C., Ventura, S. & Castro, C. (2011). A collaborative educational association rules mining tool. *The Internet and Higher Education*, 14(2): 77-88.

Goyal, M., and Vohra, R. (2012). Applications of Data Mining in Higher Education. *International Journal of Computer Science Issues*, 9, 2(1): 113-21.

Hadden, J., Tiwari, A., Roy, R. & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, 34(10): 2902–2917.

Herzog, S. (2005). Measuring Determinants of Student Return VS. Dropout/ Stopout VS. Transfer: A First-to-Second Year Analysis of New Freshmen. *Research in Higher Education*, 46(8): 883–928.

Honigman, B. (2013). 5 Secrets to Increasing Customer Retention -- and Profits, Available in: http://www.entrepreneur.com/article/227946

Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies*, 13(1): 61–72.

Kabakchieva, D., Stefanova, K. & Kisimov, V. (2011). Analyzing University Datafor Determining Student Profiles and Predicting Performance, 4th International Conference on Educational Data Mining, Eindhoven, July 6-8, The Netherlands.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G. & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 53(3): 950-965.

Murphy, E.C. & Murphy, M.A. (2013). Leading on the Edge of Chaos, John Wiley & Sons, Canada. Available in: http://www.impactlearning.com/resources /metrics /customer-retention.

Namni, A. & Talepasand, S. (2016). Academic Achievement Predicting Model for Admission in Master Levels Case Study: E-Learning Center for Iran University of Science and Technology. *Journal of Management and Planning in Educational System*, 8, 15, 71-82.

Nandeshwar, A. & Chaudhari, S. (2009). Enrollment prediction models using data mining. Available in: http://nandeshwar.info/wp-content/uploads/2008/11/ DM WVU_Project.pdf.

Nandeshwar, A., Menzies, T. & Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, 38(12): 14984–14996.

Náplava, P. & Šnorek, M. (2003). Modeling of Student's Quality by Means of GMDH Algorithms. System Analysis Modeling Simulation, 43(10): 1415- 1426.

Pechenizkiy, M., Calders, T., Vasilyeva, E. & Bra, P. D. (2008). Mining the student assessment data: Lessons drawn from a small scale case study. in 1st International Educational Data Mining Conference (EDM2008). Montreal Canad, June 20-21 2008.

Pittman, K. (2008). Comparison of data mining techniques used to predict student retention. Ph.D. thesis, Nova Southeastern University, USA.

Rokach, Lior & Oded Maimon (2005). "*Clustering methods*." Data mining and knowledge discovery handbook. Springer US, 2005. 321-352.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1): 135–146.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to

Romero, C. (2008). Data mining algorithms to classify students. in 1st International Educational Data Mining Conference (EDM2008), June 20-21 Canada.

Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33, 135–146.

Romero, C., Ventura, S., Barnes, T. & Desmarais, M. (2009). Recommendation in higher education using data mining techniques, 2nd International Conference on Educational Data Mining, EDM 2009, July 1-3, Cordoba Spain.

Sabaghian, Z. & Pourkazemi., M. (2009). The Comparison of University Students Who Had Consecution In Their Field of Study At M.S. Level, With Those Who Had Not In Terms of Their Educational Success. *Journal of Management and Planning in Educational Systems*, 1 2, 7-21.

Safapour, P., (2014). Application of Data Mining In Higher Education, National Conference of Computer Engineering and Information Technology, Islamic Azad University, Shoshtar Branch.

Scott, G., Shah, M., Grebennikov, L. & Singh, H. (2008). Improving student retention: A University of Western Sydney case study. *Journal of Institutional Research*, 14(1): 9–23.

Smith, W. (2005). Applying Data Mining to Scheduling Courses at a University. *Communications of the Association for Information Systems*, 16, 463-479.

Sujitparapitaya, S. (2006). Considering student mobility in retention outcomes. *New Direction for Institutional Research*, 131, 35-51.

Superby, J., Vandamme, J. & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006), Hong kong, 37-44.

Talebi, M. (2014). Decision Support System for Faculty Managers for Lessen Presentation via using Data mining (Case study: Bachelor Degree of Payam Noor University, Qom Branch), Master Thesis, Payam Noor University, Tehran Branch, Computer and Information Technology.

Tantuco N. & Uy R. (2014). Creating Long-term Loyalty Relationships. Available in:http://catalogue.pearsoned.co.uk/assets/hip/gb/hip_gb_pearsonhighered/sample chapter/0273755021.pdf.

Tari, M. (2012). Effectiveness Factors in Higher Education Management, Payam Noor University- Tehran Branch- Department of Computer and Information Technology, Master Thesis.

Toyserkani, M. & Nasiri, M. (2012). Period of Student Graduation Prediction via Using Data Mining, First National conference of Scholars of Computer and Information Technology, Tabriz university, Tabriz.