

داده‌کاوی نمره‌های آزمون متمرکز دانشگاه جامع علمی کاربردی

اسرافیل علاء*

مریم ملباقر**

مصطفی یوسفی طزرجان***

چکیده

هدف از اجرای این پژوهش، شناسایی و کشف الگوها و روابط ناشناخته و شناسایی موارد خارج از الگوها، از حجم عظیم داده‌های سامانه آموزشی دانشگاه جامع علمی کاربردی بود. بدین‌منظور نمره‌های ۱۹۲۰۷ دانشجوی/ درس در ۸ درس، ۶ استان و ۱۲۰ مرکز آموزشی که به صورت متمرکز در سراسر کشور در نیم‌سال دوم تحصیلی ۹۸-۱۳۹۷ برگزار شده است، مطالعه و بررسی شد. تحلیل‌های آماری و داده‌کاوی با نرم‌افزار SPSS Modeler انجام گرفت. ابتدا آمارها محاسبه و رابطه ویژگی‌ها و درس‌های مختلف بررسی شد. سپس با خوشه‌بندی K-means، پنج خوشه حاصل شد. خوشه ۱، مربوط به دانشجویان ضعیف است، این دانشجویان نمره آزمون پایینی دارند. خوشه ۲، مربوط به دانشجویان ممتاز است که تعداد واحد بالا گرفتند و نمره آزمون بالایی دارند. خوشه ۴، مربوط به دانشجویان ممتازی است که تعداد واحد کمتری گرفته‌اند، این دانشجویان حتی از دانشجویان خوشه ۲ نیز نمره‌های آزمون بالاتری دارند. خوشه ۵، مربوط به تعداد واحد اخذ شده کم است که دانشجویان مشروط را دربردارد. دانشجویان این خوشه نسبت به دانشجویان ضعیف (خوشه ۱) نمره آزمون بهتری دارند. سپس رابطه متغیرهای مختلف بررسی شد. از بین متغیرها، بین نمره آزمون و معدل رابطه مستقیم وجود دارد. نتایج این پژوهش می‌تواند در پیشبرد هدف‌های آموزشی دانشگاه مؤثر باشد.

واژگان کلیدی: دانشگاه جامع علمی کاربردی، آزمون متمرکز، داده‌کاوی، خوشه‌بندی، کشف قوانین همبستگی

* مدیرکل دفتر سنجش و آزمون دانشگاه جامع علمی کاربردی، سازمان مرکزی، تهران، ایران
** عضو هیئت علمی دانشگاه جامع علمی کاربردی، سازمان مرکزی، تهران، ایران (نویسنده مسئول):
(mollabagher@gmail.com)

*** عضو هیئت علمی دانشگاه جامع علمی کاربردی، البرز، ایران

مقدمه

مؤسسات آموزش عالی برای ایجاد آموزش باکیفیت از راه‌های افزایش سطح آگاهی، دانش و ظرفیت ذهن انسان می‌توانند در دانشگاه‌ها تغییر ایجاد کنند (اولالی و اویوول^۱، ۲۰۱۶). کیفیت تعامل دانشجو و استاد، مشارکت دانشجویان و فرایندهای تعامل، شاخص‌های حیاتی است که بر دانشجویان، ناکامی و میزان موفقیت آنان تأثیر می‌گذارد. همچنین برای یادگیری مطلوب و اینکه یک نظام آموزشی بتواند روی شیوه‌ها و شکاف‌های خود تجدیدنظر کند، داشتن ابزاری برای گردآوری داده‌ها و تهیه بازخورد مربوط به عملکرد در این موضوع بسیار مهم است (دارادومیس و همکاران^۲، ۲۰۱۹).

داده‌کاوی استخراج اطلاعات مفید پنهان از یک مجموعه داده‌ها از راه آنالیز علمی و روش‌هایی است که روند داده‌ها را شناسایی می‌کند و الگوهای پنهان را در مجموعه داده‌های داده شده قرار می‌دهد و به همین ترتیب، داده‌کاوی را می‌توان کشف دانش دانست (حسین و همکاران^۳، ۲۰۱۸). دانش قابل کشف با داده‌کاوی در حوزه آموزش، نه تنها قابل استفاده صاحبان سیستم یعنی مدرسان و مسئولان آموزشی بلکه قابل استفاده برای کاربران سیستم یعنی دانشجویان نیز است (شفیع‌پور مطلق و نظری، ۱۳۹۲). از طرفی برای بهبود عملکرد تحصیلی دانشجویان، نیاز به اطلاعات در مورد وضعیت دانشجو است؛ زیرا دانشجویان سطوح مختلف انگیزه، درک متفاوتی از آموزش و یادگیری و شرایط محیطی متفاوت از یکدیگر دارند (کومار و همکاران^۴، ۲۰۱۱). با توجه به اینکه این پژوهش در دانشگاه جامع علمی کاربردی و با موضوع آزمون متمرکز صورت گرفته است، در این بخش اطلاعات کلی در مورد دانشگاه جامع علمی کاربردی و نیز آزمون متمرکز ارائه می‌شود.

دانشگاه جامع علمی کاربردی، بزرگ‌ترین متولی نظام آموزش عالی علمی کاربردی با حدود ۹۰۰ برنامه درسی، بیش از ۶۰۰ مرکز آموزشی فعال، بیش از ۴۰۰ هزار دانشجو، بیش از یک میلیون دانش‌آموخته در سراسر کشور با ۱۷ درصد سهم از کل دانشجویان دانشگاه‌های ایران است که زیر نظر وزارت علوم، تحقیقات و

1. Olaleye & Oyewole

2. Daradoumis et al.

3. Hussain et al.

4. Kumar et al.

فناوری اداره می‌شود. این دانشگاه با ساختار تشکیلاتی حوزه ستاد مرکزی، ۳۲ واحد استانی، بالغ بر ۶۲۰ مرکز آموزش علمی کاربردی تحت نظارت، به دنبال فراهم آوردن موجباتی است که مشارکت سازمان‌ها و دستگاه‌های اجرایی دولتی و غیردولتی را برای آموزش نیروی انسانی متخصص و مورد نیاز بخش‌های مختلف اقتصادی، اجتماعی و فرهنگی کشور ممکن سازد.

با توجه به تأکید رئیس دانشگاه مبنی بر اجرای آزمون‌های متمرکز در درس‌های منتخب در هر مقطع تحصیلی و به منظور حفظ و ارتقای کیفیت آموزش‌ها و نظارت بر حسن اجرای درست آن، مصوبه اجرای برگزاری آزمون‌های متمرکز پایان نیم‌سال در مردادماه ۱۳۹۷ اعلام شد. این آزمون‌ها در مرحله اول، صرفاً در نیم‌سال اول تحصیلی ۹۸-۱۳۹۷ در استان البرز به صورت آزمایشی اجرا شد و در نیم‌سال دوم تحصیلی ۹۸-۱۳۹۷ نیز در استان‌های البرز، آذربایجان شرقی، چهارمحال و بختیاری، سمنان، قم و کرمانشاه (۱۲۰ مرکز آموزشی، ۱۹۲۰۷ دانشجوی درس و ۲۳۵ رشته) با انتخاب یک استان در مناطق نه‌گانه در کل کشور و نیز تأیید رئیس واحد استانی انجام گرفت (در سه منطقه به دلیل وقوع حوادث طبیعی، این آزمون در هیچ استانی برگزار نشد). پژوهش حاضر، نتیجه بررسی داده‌های مربوط به برگزاری آزمون در این نیم‌سال است. برخی از هدف‌های برگزاری آزمون متمرکز عبارت‌اند از: حفظ و ارتقای کیفیت مطلوب آموزش‌های مهارتی، نظارت بر حسن اجرای صحیح آموزش‌ها منطبق بر الگوها، سنجش کیفیت عملکرد مدرسان، ارزیابی عملکرد آموزشی مؤسسات و مراکز آموزشی، اعتلای سطح دانش دانشجویان، سنجش یادگیری و میزان دانش کسب شده توسط دانشجویان. پرسش‌های این آزمون‌ها به صورت ۲۸ پرسش تستی هر کدام به ارزش ۰٫۵ نمره بود و توسط مدرسان علمی کاربردی این درس‌ها در نیم‌سال جاری، طراحی و بازبینی شده است. سطح پرسش‌ها در چهار بخش (بسیار آسان، آسان، متوسط، دشوار) و منطبق با سرفصل کتاب‌های معرفی شده، تعیین شده است. مدت‌زمان برگزاری آزمون ۴۰ دقیقه بود و نمره منفی نداشت. نمره آزمون (۲۰ نمره) شامل ۴ نمره آزمون متمرکز و ۶ نمره در اختیار مدرس برای ثبت عملکرد دانشجویان در فعالیت کلاسی و حل تمرین در نظر گرفته شد. هر چند سیاست توسعه کمی آموزش عالی در ایران به‌عنوان پاسخی به تقاضای رویه رشد اجتماعی در قالب توسعه‌های موازی آموزش عالی، زمینه‌توان آموزش عالی را فراهم کرده است، اما آموزش عالی ایران با روند کنونی رشد کمی و به سبب نبودن ابزار سیاست‌گذاری

مشخص برای تضمین کیفیت، الگویی مؤثر و مناسب برای استقرار و نگهداری کیفیت نداشته است (عباسی و شیره‌پز آرانی، ۱۳۹۰).

در این پژوهش، قصد بر آن است تا با سنجش متغیرهای مؤثر در آزمون متمرکز و نیز کشف رابطه بین این متغیرها و درس‌ها، گامی در جهت کیفی‌سازی این آزمون‌ها و سنجش عملکرد دانشجویان فراهم شود و ابزاری در جهت تصمیم‌سازی سنجش کیفیت در اختیار مدیران ارشد دانشگاه قرار گیرد.

مبانی نظری و پیشینه پژوهش

داده‌کاوی: تعریف یکسانی درباره داده‌کاوی وجود ندارد. اما به عبارت ساده، داده‌کاوی دانش استخراج اطلاعات مفید از پایگاه داده‌ها یا مجموعه داده‌ها است (هند و همکاران، ۲۰۰۱).

خوشه‌بندی: خوشه‌بندی که به آن قطعه‌قطعه‌سازی نیز می‌گویند، موارد درون یک گروه در آن با هم یکسان ولی با موارد خارج از آن گروه متفاوت هستند. خوشه‌بندی شامل فرایند گروه‌بندی داده‌ها به طبقات و خوشه‌هایی است که موضوعات درون هر خوشه، شباهت بسیار بالایی با یکدیگر داشته و با دیگر خوشه‌ها متفاوت هستند (زوی و همکاران، ۲۰۰۹). یکی از معروف‌ترین فنون در خوشه‌بندی، شبکه‌های عصبی است (نای و همکاران، ۲۰۰۹).

همبستگی یا کشف قواعد وابستگی: هدف همبستگی، برقراری روابط میان مقوله‌هایی است که با یکدیگر در یک کارکرد، مسلم فرض شده‌اند. محاسبات آماری از ابزارهای مورد استفاده در تحلیل همبستگی است (احمد، ۲۰۰۴).

با توجه به اهمیت فراوان داده‌کاوی آموزشی، پژوهش‌های بسیاری در کشور و نیز در سطح جهان صورت گرفته که برخی از آنها در قالب جدول‌های (۱ و ۲) به اختصار معرفی می‌شوند:

1. Hand et al.

2. Zoe et al.

3. Ngai et al.

4. Ahmed

جدول (۱) پیشینه برخی پژوهش‌ها در کشور

نویسنده/ نویسندگان	توضیحات
(حیدری و یقینی، ۱۳۸۹)	پژوهشی با عنوان «دسته‌بندی و پیش‌بینی وضعیت تحصیلی دانشجویان با استفاده از فنون داده‌کاوی» با استفاده از داده‌های دموگرافیک و سوابق تحصیلی دانشجویان با کمک فنون دسته‌بندی درخت تصمیم، رگرسیون لجستیک، نزدیک‌ترین همسایگی و شبکه‌های عصبی مدل‌های مختلفی برای پیش‌بینی نمره‌های دانشجویان دانشگاه علم و صنعت صورت گرفته است. در نهایت، مدل نزدیک‌ترین همسایگی به‌عنوان بهترین مدل معرفی شده است. بر همین اساس می‌توان از مدل‌های پیشنهادی به‌عنوان یک ابزار پشتیبان تصمیم‌گیری در نظام‌های آموزشی استفاده کرد.
(زاهدی و همکاران، ۱۳۹۲)	در این پژوهش مدلی ارائه شده است که با استفاده از داده‌های موجود دانشجویان فارغ‌التحصیل رشته حسابداری دانشگاه شاهرود، موفقیت یا عدم موفقیت دانشجویان در حال تحصیل این رشته را بررسی کرده است. همچنین مدل پیشنهاد شده افزون بر یاری رساندن به مدیران آموزشی به دانشجویان در حال تحصیل نیز کمک می‌کند. پس از داده‌کاوی اطلاعات موردنظر و در نظر گرفتن جدول‌ها و مقایسه‌های انجام گرفته، این نتیجه حاصل شد که معدل کل دانشجویان این رشته به میانگین معدل دو ترم اول و میانگین معدل درس‌های پایه دو ترم اول وابسته است.
(فرهادی و همکاران، ۱۳۹۲)	در این مقاله، داده‌کاوی روی داده‌های آموزشی دانشگاه صنعتی شاهرود اجرا شده است. هدف اصلی در این پژوهش، شناسایی عوامل مؤثر در موفقیت یا عدم موفقیت دانشجویان رشته ریاضیات در فارغ‌التحصیلی (دانش‌آموخته، اخراج، انصراف و ...) است. دانشجویان می‌توانند با استفاده از این متغیرها، موفقیت و کارایی خود را مورد سنجش قرار داده و در بهبود وضعیت تحصیلی‌شان کوشش کنند. همچنین این مدل می‌تواند به ارتقای سطح آموزشی و علمی دانشگاه کمک کند.
(رستمی و همکاران، ۱۳۹۴)	پژوهشی با عنوان «پیش‌بینی میزان پیشرفت تحصیلی دانشجویان با روش خوشه‌بندی فازی در محیط‌های آموزشی» در یک نیم‌سال روی دانشجویان ورودی ۸۵ تا ۹۱ دانشکده کامپیوتر و فناوری اطلاعات امیرکبیر انجام گرفته است. پژوهشگران بر مبنای الگوی به دست آمده دریافتند که می‌توان با توجه به ویژگی‌های تأثیرگذار (سن، جنسیت، وضعیت اشتغال، وضعیت خوابگاهی بودن، تأهل، وضعیت سلامت، تعداد افراد خانواده، وضعیت تغذیه، قدرت محاسبات ریاضی، تعداد واحد درسی گذرانده و نمره‌های میان‌ترم درس‌ها) بر دانشجویان از ابتدای نیم‌سال راهنمایی‌های لازم را به دانشجویان، ارائه و آنان را از محدوده نمره‌های آبی خود در سال‌های بعد آگاه کرد تا بر این اساس برنامه‌ریزی مناسب تحصیلی صورت گیرد.
(برومندی و فاموری، ۱۳۹۵)	پژوهشگران در دانشگاه آزاد مرودشت به بررسی وضعیت تحصیلی دانشجویان دانشگاه آزاد مرودشت با رویکرد داده‌کاوی اقدام کردند. در این پژوهش با استفاده از فن داده‌کاوی با روش دسته‌بندی به کمک الگوریتم درخت ID3، بیزی و KNN روی داده‌های مربوط به ۱۲۳۳۲ دانشجوی دانشگاه آزاد اسلامی مرودشت در بین سال‌های ۱۳۸۸ تا ۱۳۹۴ وضعیت تحصیلی دانشجویان را بر اساس مشخصات فردی و گذشته تحصیلی آنان بررسی کردند. با استفاده از نتایج این تحلیل، مسئولان دانشگاه آزاد مرودشت می‌توانند در برنامه‌ریزی برای نوع ورود

نویسنده / نویسندگان	توضیحات
	دانشجو و نیز تشخیص گروه‌های آموزشی موفق، در نحوه ارائه خدمات برای بهبود کیفیت وضعیت تحصیلی دانشجویان در رشته‌های مختلف، تصمیم‌گیری‌های لازم را به عمل آوردند.
(حسینی و بذرافشان، ۱۳۹۷)	در این پژوهش، یک مدل داده‌کاوی مبتنی بر فن خوشه‌بندی سلسله‌مراتبی توسعه‌یافته برای تحلیل اطلاعات مربوط به وضعیت تحصیلی دانشجویان در دانشکده مدیریت دانشگاه صنعتی شاهرود، شامل اطلاعات پیش از ورود به دانشگاه و وضعیت تحصیلی آنان ارائه شده است. این خوشه‌بندی براساس شاخص‌های نشان‌دهنده وضعیت تحصیلی دانشجویان صورت گرفت. نتایج نشان داد، ویژگی‌های دانشجویان مربوط به دوران قبل از ورود به دانشگاه نیز تأثیر بسزایی در عملکرد آموزشی آنان در دانشگاه دارد. بر این اساس مجموعه‌ای از قوانین برای ارتقای تصمیم‌گیری مدیران، پیشنهاد شد.
(قدوسی و همکاران، ۱۳۹۸)	پژوهشگران در مقاله‌ای با عنوان «پیش‌بینی و تحلیل عملکرد دانشجویان به کمک تکنیک‌های داده‌کاوی به منظور بهبود عملکرد تحصیلی» به ارائه مدلی بر مبنای ۱۳ شاخص و الگوریتم Logit Boost، بر روی اطلاعات دانشجویان رشته مهندسی صنایع دانشگاه تربیت مدرس، پرداختند. بر مبنای این پژوهش، مدل‌های مختلفی برای پیش‌بینی وضعیت تحصیلی دانشجویان در نیم‌سال بعدی ارائه شد و پژوهشگران به این نتیجه رسیدند که می‌توان مدل پیشنهادی را به‌عنوان یک ابزار پشتیبان تصمیم‌گیری در نظام‌های آموزشی استفاده کرد. در نهایت، با توجه به نتایج به دست آمده و نظرخواهی از خبرگان دانشگاهی، فرایند انتخاب واحد، بازطراحی شد.

جدول (۲) پیشینه برخی پژوهش‌ها در خارج کشور

نویسنده / نویسندگان	توضیحات
(کسیانتیس ^۱ ، ۲۰۰۹)	این پژوهش با استفاده از روش‌های مدرن یادگیری ماشین و تحلیل هزینه مؤثر اجرا شد. هدف اصلی پژوهشگر، پیش‌بینی افت تحصیلی دانشجویان جدید با دقت قابل قبول بود. در پایان، پژوهشگر با معرفی راهکارهایی از پیاده‌سازی آموزش مجازی، سعی در کاهش آمار افت تحصیلی دانشجویان تازه‌وارد داشت.
(ورقسه و همکاران ^۲ ، ۲۰۱۰)	در این پژوهش، الگوریتم خوشه‌بندی k-means روی اطلاعات ۸۰۰۰ دانشجویان اجرا شد. این خوشه‌بندی بر اساس پنج متغیر: میزان حضور، نمره آزمون، نمره سمینارها، تکالیف و نمره‌های دانشجویان صورت گرفت. نتایج نشان داد، رابطه مستقیمی بین سطح حضور دانشجویان با نمره‌های کسب شده توسط آنها وجود دارد.
(مجیب و همکاران ^۱ ، ۲۰۱۳)	هدف از اجرای این پژوهش، طبقه‌بندی دانشجویان برای پیش‌بینی قبولی دانشجویان بود. متغیرهای پژوهش نیز شامل نمره‌ها، میانگین نمره‌ها و جنسیت بود. این طبقه‌بندی بر اساس الگوریتم NBC ^۲ انجام گرفت و جدول‌های پیش‌بینی احتمال فارغ‌التحصیلی دانشجویان بر

1. Kotsiantis

2. Varghese et al.

توضیحات	نویسنده/ نویسندگان
مبنای آن ساخته شد. کسب اطلاعات لازم برای فراهم ساختن امکان فارغ‌التحصیلی در مناسب‌ترین زمان از هدف‌های این پژوهش بود.	
در این پژوهش، عوامل متعددی که بر عملکرد دانشجویان در آموزش عالی تأثیر می‌گذارد، بررسی شده است و با استفاده از الگوریتم‌های درخت تصمیم‌گیری و نیز فن بیزی، یک مدل کیفی ارائه شده است که به بهترین نحو، دانشجویان را طبقه‌بندی و عملکرد آنان را بر اساس عوامل فردی و اجتماعی مرتبط پیش‌بینی می‌کند. نتیجه این پژوهش نشان داد که کارایی دانشجویان صرفاً براساس عملکرد آموزشی آنان نیست و متغیرهای بسیاری خارج از محیط دانشگاهی بر کارایی تحصیلی آنان تأثیرگذار است.	(سا، ۲۰۱۶)
در این پژوهش، طبقه‌بندی دانشجویان در حوزه داده‌کاوی آموزشی با استفاده از روش PSC ^۵ انجام گرفته است. ابتدا مجموعه‌ای از پرسش‌های کلاسی درس استادان گردآوری شد و به صورت دستی با یادداشت‌های Bloom ^۶ حاشیه‌نویسی شد و با پاسخ‌های دانشجویان، طبقه‌بندی براساس سطوح شناختی Bloom ^۶ ، انجام گرفت.	(یحیی، ^۴ ۲۰۱۷)
با توجه به اینکه سیستم‌های مدیریت فرایند سنتی، قادر به مدیریت و بهبود فرایندها با حجم بالایی داده نیستند. هدف از اجرای این پژوهش، ایجاد یک چهارچوب یکپارچه برای استفاده از فنون داده‌کاوی به منظور بهبود فرایندهاست. در این پژوهش از دو روش خوشه‌بندی و طبقه‌بندی، برای کشف الگوهای ارزشمند استفاده شده است. یک چهارچوب در سه بخش و پنج مرحله تدوین شده و پیشنهادهایی برای بهبود فرایندهای کاری در حوزه داده‌کاوی آموزشی ارائه شد.	(خانبابایی و همکاران ^۷ ، ۲۰۱۸)
در این پژوهش، داده‌کاوی اطلاعات دانشجویان دانشگاه نیجریه در سه سال اول تحصیل با استفاده از ۶ الگوریتم در روش رگرسیون درجه دوم انجام گرفته است. از نتیجه این پژوهش برای پیش‌بینی امکان فارغ‌التحصیلی دانشجویان استفاده شد و همچنین با شناسایی دانشجویان ضعیف، مداخله زودهنگام برای اصلاح فرایندهای آموزشی این دانشجویان صورت گرفت.	(ادکیتان و سالو ^۸ ، ۲۰۱۹)
در این پژوهش با استفاده از روش شبکه عصبی مصنوعی عمیق، مدلی جهت پیش‌بینی عملکرد آموزشی دانشجویان با دقتی بین ۸۴ تا ۹۳ درصد و مشخص کردن زمان مداخله احتمالی در فرایند آموزشی دانشجویان به منظور جلوگیری از افت تحصیلی، ارائه شده است.	(واحد و همکاران ^۸ ، ۲۰۲۰)

1. Mujib et al.

2. Naive Bayes classifier

3. Saa

4. yahya

5. Particle Swarm Classification

6. Khanbabaei et al.

7. Adekitan & Salau

8. Waheed et la.

روش پژوهش

داده‌کاوی یکی از پیشرفت‌های اخیر در فناوری‌های مدیریت داده است. داده‌کاوی، مجموعه‌ای از فنون است که برای اشخاص امکان حرکتی فراتر از تحلیل و استخراج اطلاعات پنهان انبوه داده‌ها را ایجاد می‌کند (کوتوفری و استوفل^۱، ۲۰۰۷).

داده‌کاوی یک فن میان‌رشته‌ای برای اکتشاف الگوهای جذاب و ناشناخته است که از علوم یادگیری ماشین، تشخیص الگو، آمار، پایگاه داده‌ها و بصری‌سازی به منظور استخراج اطلاعات از پایگاه داده‌های بزرگ استفاده می‌کند. با توجه به تعداد زیاد «دانشجو درس» برگزار شده و پیچیدگی داده‌ها و نیاز به پاک‌سازی و استخراج اطلاعات پنهان، الگوها و روابط مشخص در حجم انبوهی از داده‌ها با هدف پیش‌بینی رویدادها و نتایج آتی، در پژوهش حاضر از داده‌کاوی استفاده شده است. داده‌کاوی آموزشی ابزار کارآمدی برای شناسایی و کشف الگوها و روابط ناشناخته، جذاب و مفید از حجم عظیم داده‌های ثبت شده در سامانه‌های دانشگاه‌هاست.

پیشرفت در سرعت و پردازش رایانه‌ها امکان تجزیه و تحلیل‌های خودکار، سریع و آسان را ایجاد کرده است. هر چه مجموعه داده‌های گردآوری شده پیچیده‌تر شود، کشف بینش‌های مربوطه نیز دشوارتر خواهد شد. در بیشتر دانشگاه‌های ایران، بانک‌های اطلاعاتی وسیعی از ویژگی‌ها و عملکرد دانشجویان موجود است که شامل حجم زیادی از اطلاعات مربوط به سوابق آموزشی، تحصیلی و غیره می‌شود. با توجه به طیف گسترده‌ای از فنون داده‌کاوی و تنوع انواع مختلف داده در پایگاه داده، به منظور نیل به هدف‌های داده‌کاوی باید فرایند اجرای پژوهش به صورت کاملاً دقیق تعریف شود. با در نظر گرفتن اینکه این پژوهش در فضای دانشگاهی و با استفاده از اطلاعات موجود در سامانه‌های دانشگاه صورت گرفته، بنابراین در کسب اطلاعات مشکلی وجود نداشته است. پس از دریافت داده‌های مورد نیاز مراحل اجرای کار به شرح زیر بود:

۱- پیش‌پردازش داده‌ها: حذف اطلاعات غیرمرتبط و استخراج متغیرهای کلیدی در شناخت الگو، پیش‌پردازش نامیده می‌شود (کندی و همکاران^۲، ۱۹۹۸). این مرحله شامل انتخاب ویژگی‌ها، پاک‌سازی داده‌ها و تبدیل داده‌ها است.

1. Cotofrei & Stoffel

2. Kennedy et al.

۲- انتخاب ویژگی: به دلیل اینکه مدل‌های ناپارامتری در طبقه‌بندی مبتنی بر داده‌ها هستند، به صرف زمان و هزینه زیاد برای کسب داده‌های مدل نیاز است. بنابراین، بهتر است ویژگی‌ها و داده‌هایی گردآوری شود که از اهمیت بیشتری در ساخت مدل طبقه‌بندی برخوردار هستند. در ساخت مدل مناسب طبقه‌بندی، نیاز به داده‌های آموزشی با کیفیت مناسب است (سالاپا و همکاران^۱، ۲۰۰۷). موفقیت نتیجه پردازش داده‌ها، هر مرحله از انتخاب داده‌ها تعیین می‌شود و با توضیح ناهنجاری‌های موجود در نتایج، پایان می‌یابد. ویژگی‌های انتخاب شده از دانشجو، مدرس و مرکز آموزشی و نوع داده‌ای هر یک در جدول (۳) نشان داده شده است.

^۱. Salappa et al.

جدول (۳) داده‌های مورد استفاده و وضعیت مقیاس اندازه‌گیری آنها

ردیف	نام شاخص	نوع داده	وضعیت استفاده
۱	استان	اسمی	ورودی
۲	کد استان	پیوسته	ورودی
۳	نام شهر	اسمی	ورودی
۴	کد مرکز	پیوسته	ورودی
۵	نام مرکز	اسمی	ورودی
۶	شماره دانشجویی	پیوسته	ورودی
۷	نام دانشجو	حذف	حذف
۸	نام خانوادگی دانشجو	حذف	حذف
۹	جنسیت دانشجو	اسمی (دوتایی)	ورودی
۱۰	تاریخ تولد	پیوسته	ورودی
۱۱	رشته تحصیلی	اسمی	ورودی
۱۲	کد رشته	پیوسته	ورودی
۱۳	کد درس	پیوسته	ورودی
۱۴	نام درس	اسمی	ورودی
۱۵	نمره کل	پیوسته	ورودی
۱۶	نمره آزمون متمرکز	پیوسته	ورودی/ خروجی
۱۷	نمره استاد	پیوسته	ورودی/ خروجی
۱۸	ترم	پیوسته	ورودی
۱۹	تعداد واحد اخذ شده	پیوسته	ورودی
۲۰	تعداد واحد پاس شده	پیوسته	ورودی
۲۱	معدل ترم	پیوسته	ورودی
۲۲	معدل کل تا ترم	پیوسته	ورودی
۲۳	نام مدرس	حذف	حذف
۲۴	نام خانوادگی مدرس	حذف	حذف
۲۵	کد مدرس	پیوسته	ورودی
۲۶	جنسیت مدرس	اسمی	ورودی
۲۷	تاریخ تولد مدرس	پیوسته	حذف

همان‌طور که در جدول (۳) دیده می‌شود شاخص‌های نام و نام خانوادگی دانشجو و نام و نام خانوادگی مدرس و تاریخ تولد مدرس حذف شده‌اند. البته تاریخ تولد دانشجو به عنوان شاخص ورودی باقی مانده است زیرا در پیش‌پردازش داده‌ها، تعداد مدرسان با تاریخ تولد نادرست یا نامشخص زیاد بوده است ولی تاریخ تولد دانشجویان اکثراً درست بوده است. شاخص‌های نمره آزمون متمرکز و نمره استاد نیز هم به‌عنوان شاخص ورودی و هم به‌عنوان شاخص خروجی در داده‌کاوی وارد شده‌اند. البته برخی از شاخص‌ها در مراحل بعدی داده‌کاوی در مدل‌ها حذف شده‌اند.

۱- پاک‌سازی داده‌ها: تمیز کردن شامل شناسایی مقادیر گمشده، متناقض یا اشتباه است (تانگ و مک‌لینان^۱، ۲۰۰۵). ابزارهای به کار رفته در این مرحله از فرایند شامل ابزارهای گرافیکی برای تهیه تصویری از توزیع‌های آماری مانند ماکسیمم، حداقل‌ها، میانگین مقادیر و اندازه‌های کوچک است. برخی از ورودی‌ها به وضوح نامعتبر بودند، که ناشی از خطای انسانی یا تکامل سیستم گزارش‌دهی مشکل شده است. در این مرحله ابتدا داده‌های NULL در فایل اکسل به مقدار تهی ("")، تبدیل شدند. زیرا نرم‌افزارهای مورد استفاده، آنها را به‌عنوان کاراکترهای متنی در نظر می‌گرفتند. تعداد کمی داده‌های پرت مانند ثبت نمره‌های منفی در بخش نمره استاد و نمره بالاتر از ۲۰ در نمره آزمون و نمره بالاتر از ۲۰ در نمره کل و معدل ترم، در داده‌ها وجود داشت.

۲- تهیه آمارهای اولیه و کلی: در این مرحله آمارهای کلی و دموگرافیک مربوطه دانشجویان، درس‌ها و استان‌ها ارائه شده است.

۳- داده‌کاوی و ارائه نتایج: اجرای داده‌کاوی و ارائه نتایج آن با استفاده از یافته‌های پژوهش در این مرحله انجام گرفته است. پس از انجام تحلیل‌های عمومی، فنون مختلف داده‌کاوی از جمله تحلیل آماری، نمودارها و مصورسازی، شبکه عصبی، خوشه بندی کا میانگین، انومالی و ... به کار گرفته شده که به برخی از مهم‌ترین نتایج حاصل در بخش یافته‌های پژوهش، اشاره شده است.

یافته‌های پژوهش

با توجه به داده‌کاوی اطلاعات، آمارهای اولیه به شرح زیر است:

¹. Tang & Maclennan

طبق جدول (۴)، مردان سهم بیشتری از دانشجویان را نسبت به زنان در این آزمون‌ها به خود اختصاص داده‌اند.

جدول (۴) تعداد دانشجو-آزمون به تفکیک جنسیت

ردیف	جنسیت	تعداد
۱	زن	۷۰۳۶
۲	مرد	۱۲۱۷۱
	جمع کل	۱۹۲۰۷

همان‌طور که در جدول (۵) دیده می‌شود، بیشترین دانشجویان شرکت‌کننده در آزمون‌های متمرکز مربوط به استان البرز و کمترین آنان مربوط به استان سمنان هستند. بیشترین میانگین مربوط به استان قم و کمترین میانگین مربوط به استان کرمانشاه است. البته با توجه به بلایای طبیعی مانند زلزله در این استان در نیم‌سال تحصیلی می‌توان این موضوع را از عوامل ناکامی دانشجویان این استان به شمار آورد.

جدول (۵) تعداد دانشجو-آزمون به تفکیک استان

ردیف	استان	تعداد دانشجو-آزمون	معدل نمره آزمون (از ۱۴)
۱	آذربایجان شرقی	۳۴۵۶	۷/۴۵
۲	البرز	۷۸۳۹	۷/۸۴
۳	چهارمحال و بختیاری	۱۴۷۶	۷/۵۱
۴	سمنان	۱۷۰۲	۷/۶۴
۵	قم	۲۳۹۱	۷/۸۸
۶	کرمانشاه	۲۳۴۳	۷/۲۸
	جمع کل	۱۹۲۰۷	۷/۶۶

جدول (۶) تعداد دانشجو-آزمون و معدل نمره آزمون متمرکز به تفکیک درس

ردیف	عنوان درس	تعداد دانشجو- آزمون	معدل نمره آزمون متمرکز (از ۱۴)
۱	اخلاق اسلامی	۲۲۲۳	۶/۸۶
۲	اندیشه اسلامی ۱	۳۲۳۴	۷/۶۶
۳	اندیشه اسلامی ۲ (نبوت و امامت)	۲۲۵۴	۶/۰۰
۴	تاریخ تحلیلی صدر اسلام	۲۳۵۶	۷/۳۲
۵	تفسیر موضوعی قرآن	۲۶۲۱	۸/۸۰
۶	دانش خانواده و جمعیت	۲۹۶۹	۸/۹۸
۷	مهارت‌ها و قوانین کسب و کار	۱۶۱۷	۷/۵۵
۸	مهارت‌های مسئله‌یابی و تصمیم‌گیری	۱۹۲۱	۷/۵۰
	جمع کل	۱۹۲۰۷	۷/۶۶

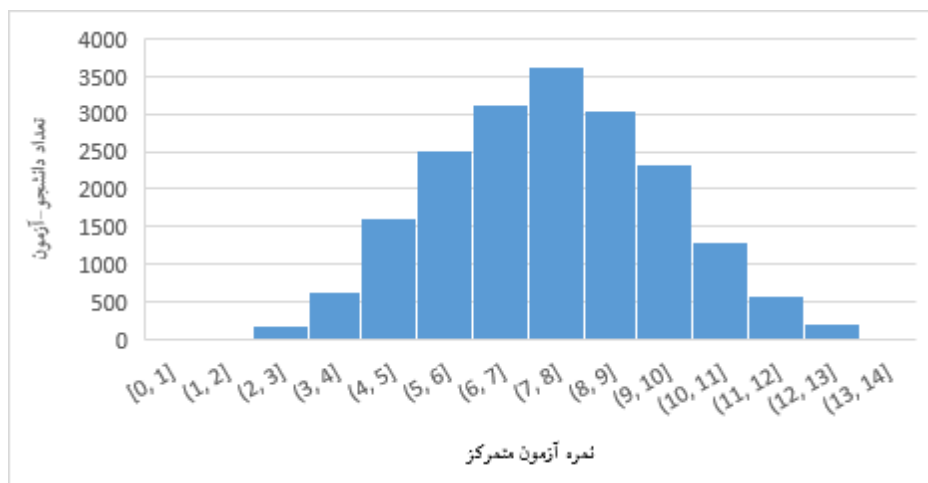
مطابق جدول (۶) در درس اندیشه اسلامی ۱ بیشترین دانشجو شرکت داشتند و در درس مهارت‌ها و قوانین کسب‌وکار با توجه به اختصاصی بودن آن کمترین دانشجو شرکت کرده است. بالاترین نمره درس «دانش خانواده و جمعیت» و کمترین نمره درس «اندیشه اسلامی ۲ (نبوت و امامت)» بوده است.

در ادامه، آمار کلی (پیش از پیش پردازش داده‌ها) و ویژگی‌های مؤثر در آزمون متمرکز در قالب جدول (۷) و نیز تعداد دانشجو-آزمون به تفکیک درس و وضعیت نمره کل ارائه می‌شود:

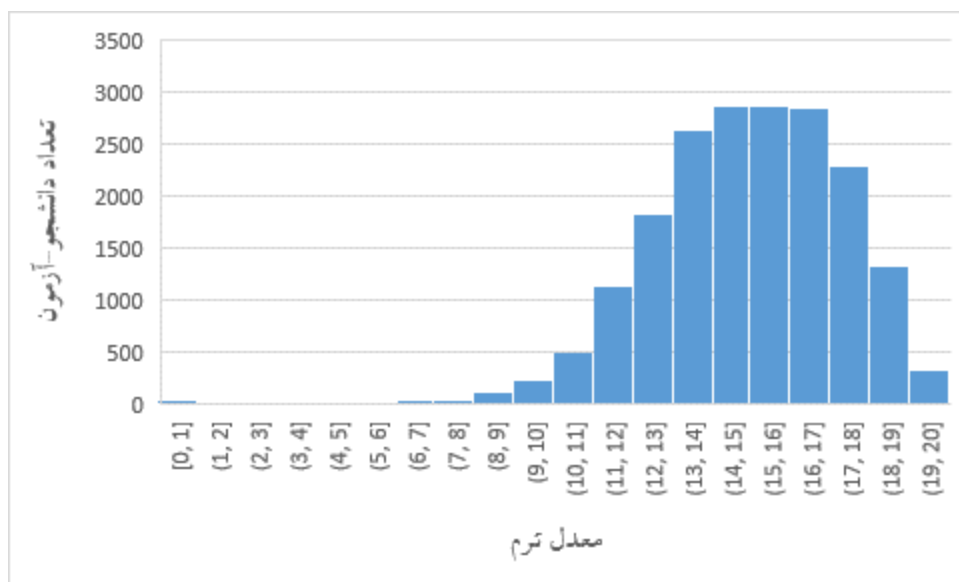
جدول (۷) آمار کلی ویژگی‌های مؤثر در مدل پیش‌بینی

ردیف	عنوان	Count	Mean	Min	Max	Range	Variance	Median
۱	نمره کل	۱۹۲۰۱	۱۳/۱۸	۰	۲۷/۰۰	۲۷/۰۰	۵/۶۵	۱۳/۵۰
۲	نمره آزمون متمرکز	۱۹۲۰۷	۷/۶۶	۰	۱۴/۰۰	۱۴/۰۰	۴/۲۴	۷/۵۰
۳	نمره استاد	۱۹۲۰۱	۵/۵۱	۱۰-	۲۰/۰۰	۳۰/۰۰	۰/۹۹	۶/۰۰
۴	تعداد واحد اخذ شده	۱۹۲۰۷	۱۶/۹۶	۰	۲۴/۰۰	۲۴/۰۰	۱۱/۴۰	۱۸/۰۰
۵	تعداد واحد پاس شده	۱۹۲۰۷	۱۵/۶۳	۰	۲۴/۰۰	۲۴/۰۰	۱۸/۳۵	۱۷/۰۰
۶	معدل ترم	۱۹۱۶۲	۱۴/۸۹	۰	۲۷/۰۰	۲۷/۰۰	۶/۱۲	۱۵/۰۵
۷	معدل کل ترم	۱۹۱۸۳	۱۵/۱۰	۰	۱۹/۸۸	۱۹/۸۸	۵/۱۴	۱۵/۱۷

همان‌طور که در جدول (۷) دیده می‌شود، داده‌های پرت در پایگاه داده، وجود داشته است که با استفاده از پیش‌پردازش شناسایی و حذف شد. از جمله داده‌های پرت می‌توان به ثبت نمره‌های منفی در بخش نمره استاد و نمره بالاتر از ۲۰ در نمره آزمون و نمره بالاتر از ۲۰ در نمره کل و معدل ترم، اشاره کرد. پس از استخراج آمار اولیه در این بخش، آمار تحلیلی از نتایج داده‌کاوی ارائه می‌شود. نخست، توزیع نمره‌های آزمون بررسی می‌شود. در شکل‌های (۱) و (۲)، توزیع نمره آزمون متمرکز و توزیع معدل ترم براساس تعداد دانشجو-آزمون نشان داده شده است. همان‌طور که در شکل دیده می‌شود این توزیع‌ها هر دو نرمال است. نتایج تست‌های آماری نیز این موضوع را تأیید کرده است.



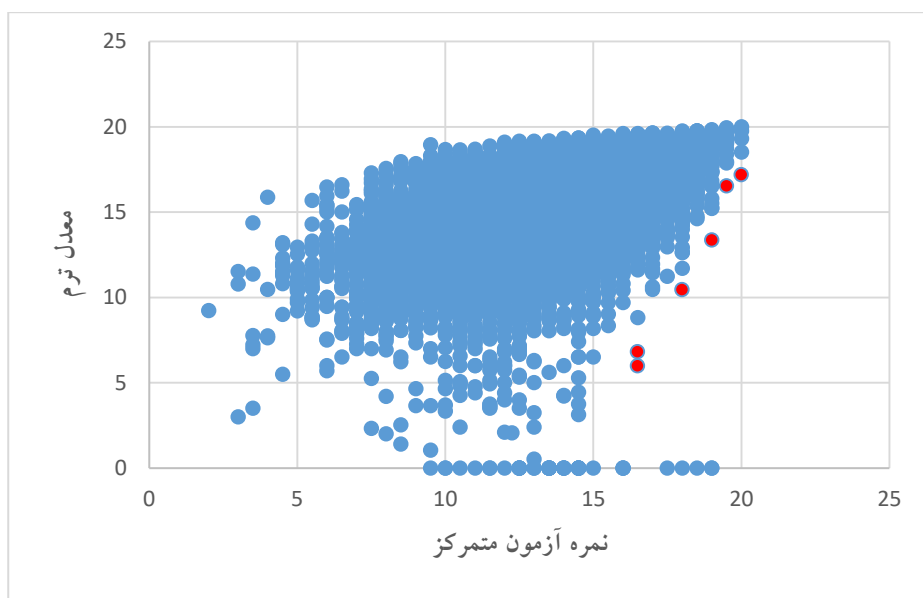
شکل (۱) توزیع نمره آزمون متمرکز بر اساس تعداد دانشجو-آزمون



شکل (۲) توزیع معدل ترم بر اساس تعداد دانشجو-آزمون

شکل (۱)، توزیع نمره آزمون متمرکز و شکل (۲)، توزیع معدل ترم را نشان می‌دهد؛ نرمال بودن نمودار نمره آزمون متمرکز در مقایسه با معدل که میانگین سایر آزمون‌ها است.

شکل (۳) رابطه معدل ترم با نمره آزمون را نشان می‌دهد. مواردی که معدل ترم پایین و نمره آزمون بالا دارند (به‌عنوان نمونه نقاط مشخص شده با رنگ قرمز) به تقلب مشکوک هستند پس این دانشجویان باید در آزمون بعدی به طور دقیق‌تری نظارت شوند.

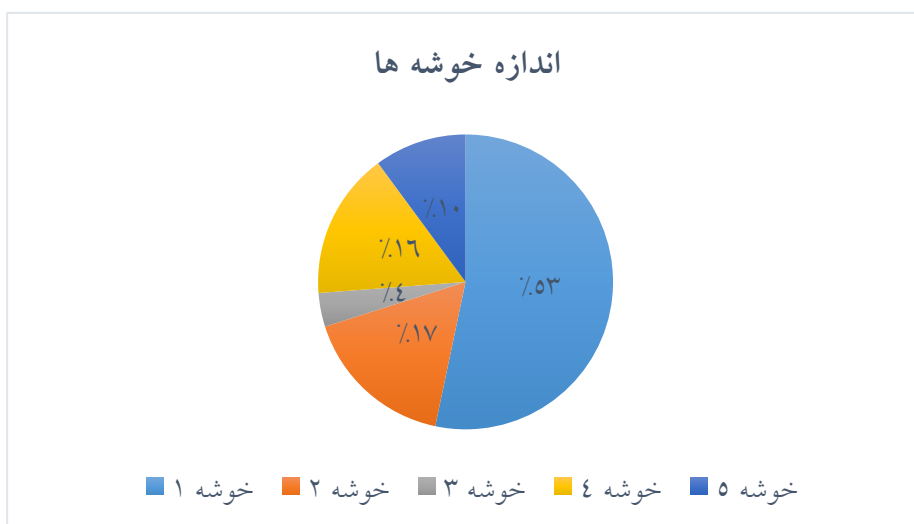


شکل (۳) رابطه معدل ترم و نمره آزمون

با توجه به بررسی آماری اولیه در این بخش به خوشه‌بندی اقدام می‌شود که هدف اصلی این پژوهش است. از آنجایی که در خوشه‌بندی، هدف، یافتن گروه‌هایی است که با یکدیگر، بسیار متفاوت ولی اعضای این گروه‌ها بسیار شبیه به هم هستند. آن چیزی که خوشه‌بندی را از طبقه‌بندی متمایز می‌کند، این است که در خوشه‌بندی، دیگر هیچ کلاس از پیش تعریف شده‌ای، وجود ندارد. رکوردها بر اساس شباهت‌های

خودشان گروه‌بندی می‌شوند و افرادی که خیره در آن کار هستند، باید این خوشه را تفسیر کنند.

در این پژوهش از خوشه‌بندی KMeans استفاده شده است. زیرا ساده‌ترین الگوریتم خوشه‌بندی و در بسیاری از مواقع از بهترین الگوریتم‌های خوشه‌بندی نیز هست. این الگوریتم می‌تواند یک گروه‌بندی ذاتی برای داده‌ها بسازد، بدون اینکه برچسب داده‌ها یا نوع آنها را بداند. در شکل (۴) اندازه خوشه‌ها در یکی از مدل‌های خوشه‌بندی مورد استفاده، ارائه شده است.



شکل (۴) اندازه خوشه‌ها

همان‌طور که در شکل (۴) دیده می‌شود، خوشه ۱ بیشترین تعداد رکوردها و خوشه ۳ کمترین تعداد رکوردها را داشته است. طبق شکل (۴)، داده‌ها در ۵ خوشه قرار گرفته‌اند که کوچک‌ترین خوشه ۳/۷ درصد و بزرگ‌ترین خوشه ۵۳/۳ درصد از رکوردها را در بر دارد. نسبت بزرگ‌ترین خوشه به کوچک‌ترین خوشه برابر ۱۴/۴۵ است. با بررسی اهمیت و وضعیت شاخص‌های ورودی، جدول (۸) حاصل شد.

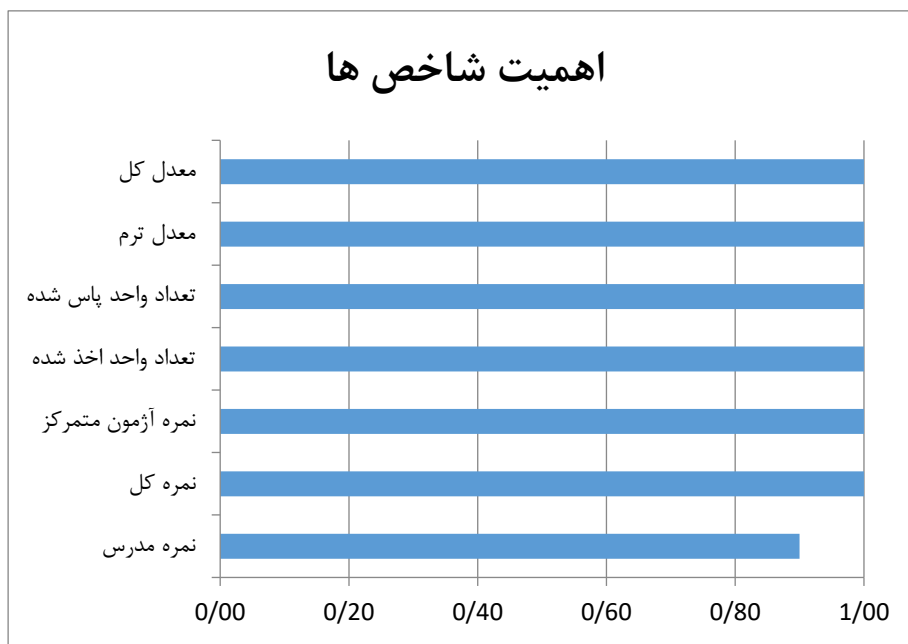
جدول (۸) اهمیت و وضعیت شاخص‌های ورودی

خوشه ۵	خوشه ۴	خوشه ۳	خوشه ۲	خوشه ۱	خوشه
تعداد واحد اخذ شده ۱۰/۶۵	تعداد واحد اخذ شده ۱۴/۷۷	تعداد واحد اخذ شده ۱۸/۸۷	تعداد واحد اخذ شده ۱۹/۲۸	تعداد واحد اخذ شده ۱۵/۴۰	ورودی‌ها
تعداد واحد پاس شده ۶/۱۵	تعداد واحد پاس شده ۱۳/۹۰	تعداد واحد پاس شده ۱۸/۰۲	تعداد واحد پاس شده ۱۹/۰۲	تعداد واحد پاس شده ۱۳/۰۷	
معدل ترم ۱۱/۴۸	معدل ترم ۱۶/۷۴	معدل ترم ۱۴/۴۹	معدل ترم ۱۶/۳۵	معدل ترم ۱۳/۳۹	
معدل کل ۱۲/۴۹	معدل کل ۱۶/۸۴	معدل کل ۱۴/۶۳	معدل کل ۱۶/۴۴	معدل کل ۱۳/۶۱	
نمره آزمون متمرکز ۶/۶۷	نمره آزمون متمرکز ۹/۶۹	نمره آزمون متمرکز ۶/۰۶	نمره آزمون متمرکز ۹/۱۸	نمره آزمون متمرکز ۶/۵۱	
نمره کل ۱۱/۶۱	نمره کل ۱۵/۴۳	نمره کل ۱۱/۵۸	نمره کل ۱۴/۹۱	نمره کل ۱۱/۸۱	
نمره مدرس ۴/۹۴	نمره مدرس ۵/۷۴	نمره مدرس ۵/۵۲	نمره مدرس ۵/۷۳	نمره مدرس ۵/۳۰	

جدول (۸) در هر خوشه وضعیت شاخص‌های مهم را نشان می‌دهد با استفاده از این جدول و تفصیل آن، وضعیت خوشه ارزیابی شده است که توسط تجربه تفسیر و برچسب‌گذاری می‌شود. خوشه ۱، مربوط به دانشجویان ضعیف است. این دانشجویان نمره آزمون پایینی دارند. همچنین، خوشه ۵ مربوط به دانشجویان مشروط است دانشجویان این خوشه نسبت به دانشجویان ضعیف یعنی خوشه ۱ نمره آزمون بهتری دارند. خوشه ۲ مربوط به دانشجویان ممتاز است که تعداد واحد بالا گرفتند و نمره آزمون بالایی دارند. خوشه ۴ مربوط به دانشجویان ممتازی است که تعداد واحد کمتری گرفته‌اند این دانشجویان حتی از دانشجویان خوشه ۲ نیز نمره‌های آزمون بالاتری دارند.

با استفاده از این نتایج می‌توان به دانشجویان خوشه‌های ۱ و ۲ پیشنهاد کرد که تعداد واحدهای کمتری را در هر نیمسال اخذ کنند تا کیفیت آموزشی بالاتری داشته باشند. همچنین مواردی که رفتاری متفاوت با خوشه خود دارند برای بررسی ویژه اعلام می‌شوند. این موارد دانشجویانی هستند که عملکردشان با خوشه‌ای که در آن قرار دارند متفاوت است. این تفاوت می‌تواند به وجود مشکلات روحی و روانی باشد لذا مرکز مشاوره می‌تواند به بهبود عملکرد این دانشجویان کمک کند. دانشجویانی که عملکرد بالاتر از خوشه خود دارند نیز می‌توانند در بررسی‌های تقلب مورد ارزیابی دقیق‌تر قرار گیرند.

در شکل (۵) نیز اهمیت شاخص‌های پیش‌بینی‌کننده در خوشه‌ها نشان داده شده است. همان‌طور که در این شکل دیده می‌شود مهم‌ترین شاخص‌های پیش‌بینی‌کننده خوشه‌ها به ترتیب عبارت‌اند از: معدل کل تا ترم، معدل ترم، تعداد واحد پاس شده، تعداد واحد اخذ شده، نمره آزمون متمرکز، نمره کل، نمره مدرس ...

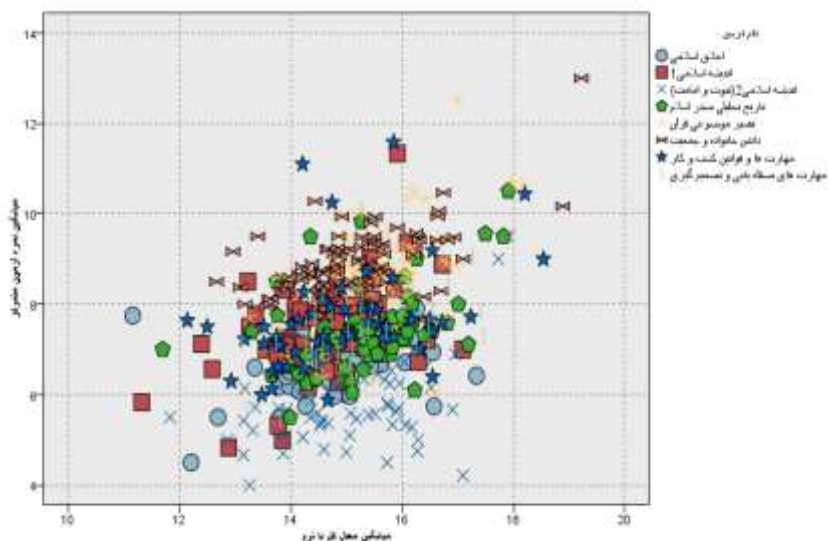


شکل (۵) اهمیت شاخص‌های پیش‌بینی‌کننده در خوشه‌ها

	تعداد واحد اخذ شده ۱۵/۴۰
	تعداد واحد پاس شده ۱۳/۰۷
	معدل ترم ۱۳/۳۹
	معدل کل ۱۳/۶۱
	نمره آزمون متمرکز ۶/۵۱
	نمره کل ۱۱/۸۱
	نمره مدرس ۵/۳۰

شکل (۶) وضعیت شاخص‌ها را برای خوشه ۱

شکل (۶) نیز وضعیت شاخص‌ها را برای خوشه ۱ را نشان می‌دهد. با توجه به شکل (۶) در خوشه ۱ تعداد واحد اخذ شده کمتر از میانگین، تعداد واحد پاس شده کمتر از میانگین، معدل کمتر از میانگین، نمره کل کمتر از میانگین، نمره کمتر از میانگین، نمره استاد کامل است. مطابق این شکل، مهم‌ترین شاخص جداکننده خوشه‌بندی انجام گرفته به ترتیب عبارت‌اند از: تعداد واحد اخذ شده، تعداد واحد پاس شده، معدل، نمره کل، و... خوشه ۲ مربوط به تعداد واحد اخذ شده بالا است. در این خوشه نمره میانگین آزمون متمرکز ۹,۱۸ است. خوشه ۵ مربوط به تعداد واحد اخذ شده کم است. در این خوشه نمره میانگین آزمون متمرکز ۶,۶۷ است. همچنین سایر خوشه‌ها قابل تفسیر است.



شکل (۷) نمودار پراکنش نمره‌های آزمون در مقابل معدل تا ترم برای درس‌های مختلف

همان‌طور که در شکل (۷) دیده می‌شود، نمره آزمون متمرکز همه درس‌ها برای دانشجویانی که در بیش از یک درس آزمون داشته‌اند، رابطه مستقیمی با هم دارند. برای بررسی مشکلات احتمالی از جمله تقلب، مشکلات روحی دانشجویان و ... نیز می‌توان نقاط خارج از الگو را بیشتر بررسی کرد. برخی از نقاط خارج از الگو هستند که می‌تواند برای شناسایی مشکلات آموزشی و روحی یا مواردی مانند تقلب و ... بررسی ویژه شود. همچنین رنگ هر درس و میزان پراکندگی آن نشان‌دهنده آن است که درس دانش خانواده و جمعیت، برای دانشجویان آسان‌تر بوده و دانشجویان نمره‌های بالاتری نسبت به سایر درس‌ها در آن گرفته‌اند. همچنین درس اندیشه اسلامی ۲ (نبوت و امامت) دشوارترین درس بوده است. به مدرسان درس اندیشه اسلامی ۲ (نبوت و امامت) پیشنهاد می‌شود که مباحث را بیشتر و ساده‌تر آموزش دهند. همچنین به آموزش دانشگاه پیشنهاد می‌شود که این درس را در ساعات مناسب‌تر برای یادگیری مانند ساعات صبح ارائه دهند. در طراحی پرسش‌های آزمون متمرکز نیز پیشنهاد می‌شود که تعداد پرسش‌های سخت، کمتر شود. در این میان، نمره

آزمون متمرکز در درس مهارت‌های مسئله‌یابی و تصمیم‌گیری نیز هماهنگی بیشتری با معدل کل تا ترم دارد.

بحث و نتیجه‌گیری

این پژوهش با بررسی پژوهش‌های پیشین در داخل و خارج کشور به صورت مطالعه موردی در دانشگاه جامع علمی کاربردی انجام گرفت. در مقایسه این پژوهش با سایر پژوهش‌ها که در بخش پیشینه پژوهش ارائه شد، موارد زیر قابل ذکر است:

- پژوهش حاضر هم از لحاظ تعداد دانشجوی و هم از لحاظ روش‌های مورد استفاده بسیار کامل‌تر بوده است. ورقسه و همکاران (۲۰۱۰) تعداد ۸۰۰۰ نفر و حسنی و بذرافشان (۱۳۹۷) در یک دانشکده، به کار داده‌کاوی اقدام کردند. اما پژوهش حاضر به صورت فرا استانی و در ۶ استان در سطح کشور با تعداد قابل ملاحظه‌ای دانشجوی اجرا شده که بر کیفیت کار می‌افزاید. حسنی و بذرافشان (۱۳۹۷)، شاخص‌های پیش از ورود به دانشگاه را مطالعه کردند اما از آنجا که این شاخص‌ها معمولاً به درستی اعلام نمی‌شود در پژوهش حاضر از شاخص‌های دقیق سال‌های قبل مربوط به دوران دانشگاه استفاده شده و نتایج با دقت بالاتری به دست آمده است.
- خانابایی و همکاران (۲۰۱۸) چهارچوبی را طراحی کردند. ادکیتان و سالا (۲۰۱۹) به پیش‌بینی فارغ‌التحصیلی، مجیب و همکاران (۲۰۱۳) به پیش‌بینی قبولی دانشجویان و کوتسیانتیس (۲۰۰۹) به پیش‌بینی افت تحصیلی اقدام کردند. در حالی که در پژوهش حاضر، هم ارزیابی دانشجوی در حین تحصیل و هم شناسایی مشکلات در طراحی سؤال‌ها مد نظر بوده است.
- یحیی (۲۰۱۷) از PSC استفاده کرد و فقط در یک دانشکده از خوشه‌بندی استفاده شده است ولی پژوهش حاضر از لحاظ تعداد دانشجوی و هم از لحاظ روش‌های مورد استفاده بسیار کامل‌تر بوده است.
- در پژوهش زاهدی و همکاران (۱۳۹۲)، معدل کل دانشجویان وابسته به معدل دو ترم اول بوده است. همچنین، بر اساس نتایج پژوهش‌ها نمره آزمون متمرکز به معدل نیم‌سال قبل و معدل کل تا نیم‌سال قبل دانشجویان وابسته است. این پژوهش نشان داد دانشجویان دارای معدل بالا چنانچه در نیم‌سال بعد تعداد واحد کمتری را بگیرند از نمره‌های بالاتری برخوردار خواهند شد.

- این پژوهش برخلاف پژوهش رستمی و همکاران (۱۳۹۴)، نشان داد مهم‌ترین عوامل مؤثر در نمره دانشجویان، تعداد واحد و معدل سال قبل و ... است. البته در پژوهش‌های مشابه، جنسیت و سن رابطه قوی با نمره آزمون داشتند اما سایر پارامترهای آن پژوهش در کار حاضر یا مصداق نداشت یا مؤثر نبود.
- پژوهش حاضر همانند پژوهش قدوسی و همکاران (۱۳۹۸)، از شاخص‌های نیم‌سال جاری برای پیش‌بینی وضعیت نیم‌سال بعدی می‌تواند استفاده کند.
- پژوهش حاضر مشابه پژوهش مجیب و همکاران (۲۰۱۳)، نمره‌ها و جنسیت را در پیش‌بینی نمره‌ها مؤثر شناخته است. البته پژوهش مجیب و همکاران در خصوص احتمال فارغ‌التحصیلی اجرا شده است.
- با توجه به در دسترس نبودن نمره‌های سمینارها، تکالیف و میزان حضور دانشجویان این پژوهش با پژوهش ورقسه و همکاران (۲۰۱۰)، از لحاظ ورودی‌ها متفاوت بوده است و تنها در مدل خوشه‌بندی مشابهت دارد.
- ازجمله نتایج پژوهش، مشخص شدن نقاط قوت و ضعف آزمون متمرکز و مراکز آموزش علمی کاربردی و دانشجویان است. برخی از این نتایج عبارت‌اند از:
 - کشف داده‌های پرت در ثبت نمره‌های آزمون و نمره مدرس، که با اطلاع پژوهشگران مشکلات سامانه برطرف شد.
 - در بین بالاترین نمره‌ها، تعدادی تک‌نفره هستند که به احتمال زیاد معرفی به استاد بوده‌اند و قاعدتاً نباید نمره بالایی داشته باشند و مشکوک (نه لزوماً) به تقلب هستند چون نظارت بر این موارد هم احتمالاً کمتر بوده است.
 - در مواردی تعداد آزمون‌دهنده در حوزه آزمون کمتر از تعداد عوامل اجرایی و نظارتی آزمون هستند (تحلیل تعداد دانشجو-آزمون به تفکیک درس و مرکز). در این موارد ناظران اهمیت کمتری قائل می‌شوند و احتمال تقلب بالاتر می‌رود (مقایسه میانگین نمره در حوزه‌های کم‌تعداد و پرتعداد). از این‌رو، پیشنهاد می‌شود به جهت کاهش هزینه‌های نظارت و اجرا آزمون درس‌هایی که تعداد کمتر از ۱۰ دانشجو در مرکز دارند در نزدیک‌ترین مرکز آموزش علمی کاربردی دارای جمعیت آزمون‌دهنده بالاتر از ۱۰ نفر آزمون دهند.

- توزیع نمره مدرس با نمره آزمون، همبستگی ضعیف دارد (بررسی توزیع نمره مدرس در آزمون متمرکز و نمودار توزیع نمره آزمون در آزمون متمرکز و جدول میزان همبستگی متغیرها). بنابراین، دستورالعمل و نظارتی بر نحوه تخصیص نمره مدرس، تدوین و ارائه شود.
- معدل کل تا ترم و معدل ترم، رابطه همبستگی متوسطی با نمره آزمون دارند. بنابراین، آزمون متمرکز بیانگر و شاخصی برای سایر آزمون‌ها نیست و با توجه به رابطه همبستگی قوی بین معدل ترم و معدل کل تا ترم، نشانگر نیاز به تغییر نگرش در آزمون‌های متمرکز یا سایر آزمون‌ها است.
- مدل خوشه‌بندی نشان داد که برخی مراکز از الگوی حاکم پیروی نمی‌کنند که این مراکز باید با دقت مورد نظارت قرار گیرند تا علت انحراف نتیجه آنها مشخص شود.
- متوسط نمره دخترها $8/4$ و پسرها $7/6$ است یعنی از منظر آزمون متمرکز برای این درس‌ها دانشجویان دختر موفق‌تر از دانشجویان پسر هستند. در تمام درس‌ها میانگین نمره‌های دانشجویان زن بالاتر از میانگین نمره‌های دانشجویان مرد است.
- برای سنجش اعتبار نتایج به دست آمده، تحلیل‌ها با روش‌های مختلف انجام گرفته است که نتایج این تحلیل‌ها یکدیگر را تأیید می‌کنند. همچنین استفاده از روش‌های مختلف، باعث خواهد شد که نتایج روش‌های مختلف، بررسی و استفاده شود. همچنین، نتایج در جلسه حضوری به مدیران و خبرگان دانشگاه، ارائه و تأیید شد. در خوشه‌بندی دانشجویان، دانشجویان خارج از خوشه که عملکرد بسیار ضعیف‌تری از خوشه خود داشتند می‌توانند با رعایت ضوابط و محرمانگی اطلاعات به مراکز مشاوره معرفی شوند. در خوشه‌بندی عملکرد مراکز آموزش علمی کاربردی، مراکز خارج از خوشه در دو بخش مورد توجه قرار گرفتند و برای بررسی دقیق‌تر به اداره نظارت دانشگاه معرفی شدند. مراکز دارای عملکرد ضعیف‌تر از خوشه، برای بررسی نظارتی و بررسی عملکرد آزمون‌های غیرمتمرکز و مراکز دارای عملکرد قوی‌تر از خوشه، برای نظارت دقیق‌تر بر تقلب‌های احتمالی، به اداره نظارت دانشگاه معرفی شدند.

- از آنجا که داده‌کاوی، فرایندی مستمر است، پیشنهاد می‌شود که در پژوهش‌های آینده، سایر فنون داده‌کاوی اجرا شود تا با ایجاد سیستم‌های پشتیبان تصمیم مبتنی بر این الگوها و تحلیل حساسیت آنها کمک مؤثری در تصمیم‌گیری بهتر مدیران باشد. با توجه به این موضوع، پیشنهادهای زیر برای پژوهش‌های آینده ارائه می‌شود.
- بررسی رابطه نمره‌های آزمون متمرکز با ارزیابی دانشجویان از مدرسان
- تحلیل داده‌های سامانه مدرسان دانشگاه
- بررسی رابطه نمره‌های آزمون‌ها با مباحث نظارتی مراکز آموزش علمی کاربردی.

منابع

- برومندی، الهام؛ فاموری، فرزانه (۱۳۹۵). بررسی وضعیت تحصیلی دانشجویان دانشگاه آزاد مرودشت با رویکرد داده‌کاوی. همایش بین‌المللی افق‌های نوین در علوم پایه و فنی و مهندسی، تهران، انجمن افق نوین علم و فناوری.
- حسینی، علی‌اکبر؛ بذرافشان، مرتضی (۱۳۹۷). تحلیل داده‌های آموزشی دانشجویان با هدف ارزیابی موفقیت تحصیلی با استفاده از رویکرد داده‌کاوی (نمونه موردی: دانشکده مدیریت و مهندسی صنایع دانشگاه شاهرود). مدیریت و برنامه‌ریزی در نظام‌های آموزشی، ۱۱(۲)، ۱۸۷-۲۰۸.
- حیدری، سمیه؛ یقینی، مسعود (۱۳۸۹). دسته‌بندی و پیش‌بینی وضعیت تحصیلی دانشجویان با استفاده از تکنیک‌های داده‌کاوی. نامه آموزش عالی، ۳(۱۲)، ۱۰۷-۱۲۴.
- رستمی، محمد؛ آیت، سید سعید؛ صاغری، فرید؛ یعقوبی، فاطمه (۱۳۹۴). پیش‌بینی میزان پیشرفت تحصیلی دانشجویان با روش خوشه‌بندی فازی در محیط‌های آموزشی. فناوری آموزش، ۱۰(۱)، ۲۳-۳۶.
- زاهدی، آتنا؛ فاضلی، فائزه السادات؛ فرهادی، محسن (۱۳۹۲). بررسی عوامل موثر بر موفقیت دانشجویان رشته حسابداری در فارغ‌التحصیلی با استفاده از ابزار داده‌کاوی. هفتمین کنفرانس داده‌کاوی ایران، ۱۹ و ۲۰ آذرماه ۱۳۹۲، تهران.
- شفیع‌پور مطلق، فرهاد؛ نظری، حسین (۱۳۹۲). طراحی یک مدل تعدیل شده برای ارزیابی اثربخشی استراتژی‌های یادگیری الکترونیک بر موفقیت تحصیلی دانشجویان. فناوری آموزش، ۸(۱)، ۳۱-۳۷.
- فرهادی، محسن؛ باقری، فهیمه؛ فاطمی، فریده‌سادات (۱۳۹۲). به کارگیری ابزارهای داده‌کاوی جهت پیش‌بینی موفقیت یا عدم موفقیت دانشجویان رشته رباتیک در فارغ‌التحصیلی. هفتمین کنفرانس داده‌کاوی ایران، ۱۹ و ۲۰ آذرماه ۱۳۹۲، تهران، ایران.
- قدوسی، محمد؛ میرسعیدی، فاطمه؛ کوشا، حمیدرضا (۱۳۹۸). پیش‌بینی و تحلیل عملکرد دانشجویان به کمک تکنیک‌های داده‌کاوی به منظور بهبود عملکرد تحصیلی. فناوری آموزش، ۱۴(۱)، ۱-۱۳.

عباسی، محمدرضا؛ شیره‌پز آرانی، علی اصغر (۱۳۹۰). سیاست‌گذاری کیفیت آموزش عالی: چالش‌ها و چشم‌اندازها. پنجمین همایش ارزیابی کیفیت در نظام دانشگاهی، تهران، دانشگاه تهران (پردیس دانشکده‌های فنی).

Adekitan A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2), Article No: we01250.

Ahmed, S. R. (2004). Applications of Data mining in retail business. *Information Technology: Coding & Computing*, 2, 455-459.

Cotofrei, P., & Stoffel, K. (2007). *Stochastic Processes and Temporal Data Mining*. California: ACM Digital Library.

Daradoumis, T., Puiga, J. M., Arguedas, M., Linana L. C. (2019). Analyzing students' perceptions to improve the design of an automated assessment tool in online-distributed programming. *Computers & Education*, 128, 159-170.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.

Hussain, S., Dahan, N. A., Ba-Alwib, F.M., Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering & Computer Science*, 9(2), 447-459. DOI: 10.11591/ijeecs.v9.i2.pp447-459

Kennedy, R. L., Lee, Y., Roy, B. V., Reed, C. D. & Lippmann, R. P. (1998). *Solving Data Mining Problems through Pattern Recognition*. S. l., Prentice Hall.

Khanbabaie, M., Movahedi Sobhani, F., Alborzi M. & Radfar, R. (2018). Developing an integrated framework for using data mining techniques and ontology concepts for process improvemen. *Journal of Systems & Software*, 137, 78-95.

Kotsiantis, S. (2009). Educational data mining: a case study for predicting dropout-prone students. *International Journal of Knowledge Engineering & Soft Data Paradigms*, 1(2), 101-111.

Kumar, B. B., & Saurabh P. (2011). Data Mining: A Prediction for Performance Improvement Using Classification. *International Journal of Computer Science & Information Security*, 9(4), 136-140.

Mujib, R., Hadi, S., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Journal EECCIS*, 7(1), 59-64.

- Ngai, E. W., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592-2602.
- Olaleye, F. O., & Oyewole, B. K. (2016). Quality assurance in Nigerian university education: the role of the National universities commission (NUC) as a regulatory body. *International Journal of Academic Research in Progressive Education & Development*, 6(12), 2222-6990.
- Saa, A. A. (2016). Educational Data Mining & Students' Performance Prediction. *International Journal of Advanced Computer Science & Applications*, 7(5), 9-22.
- Salappa, A., Doumpos, m. & Zopoundidis, C. (2007). Feature selection algorithms in classification problems: an experimental evaluation. *Optimizing Methods & Software*, 22(1), 2-5.
- Tang, Z. M., & Maclennan, J. (2005). *Data Mining with SQL Server. Indianapolis*. Wiley Publishing Inc.
- Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104, March 2020, 106189.
- Zoe, Y. Z., Leonid, C., Frada, B., & Kenm, S. (2009). Combining Data Mining and case based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research*, 195, 662-675.