



مدلی برای پیش‌بینی عملکرد دانشجویان در آزمون متمرکز مبتنی بر داده‌کاوی

مصطفی یوسفی طزرجان *

اسرافیل علاء **

مریم ملباقر ***

چکیده

هدف این پژوهش، ارائه الگویی جهت پیش‌بینی نمرات دانشجویان دانشگاه جامع علمی کاربردی در آزمون‌های متمرکز، در نیم‌سال‌های آتی دانشگاه بوده است. بدین منظور وضعیت نمرات ۱۹/۲۰۷ دانشجو/درس در ۸ عنوان درسی در ۶ استان و ۱۲۰ مرکز آموزشی که به صورت متمرکز در مقطع کاردانی و کارشناسی و هم‌زمان در سراسر کشور در نیم‌سال دوم تحصیلی ۹۸-۱۳۹۷ برگزار شده، مورد مطالعه قرار گرفته است و با استفاده از روش انتخاب ویژگی، مؤثرترین آن‌ها انتخاب شدند. برای روشن تر شدن روابط بین ویژگی‌های انتخاب شده با استفاده از مدل درخت تصمیم‌گیری و الگوریتم C5.0، با استفاده از نرم‌افزار SPSS Modeler و ۱۰ شاخص مؤثر، مدلی برای پیش‌بینی نمرات دانشجویان در نیم‌سال آتی در دروس مصوب آزمون متمرکز ارائه شده است. این الگوی پیش‌بینی می‌تواند برای کارآمدتر ساختن فرآیند یادگیری در سیستم دانشگاهی مؤثر باشد. از نتایج این مدل می‌توان به پیشنهادهایی برای اصلاح فرآیند آزمون، یافتن دانشجویان و مراکز و شرایط خارج از الگو جهت نظارت بیشتر و شناسایی مراکزی که میانگین معدل دانشجویان‌شان بالا بوده اما در آزمون متمرکز عملکرد ضعیف‌تری داشتند، اشاره کرد.

واژگان کلیدی

آزمون متمرکز، داده‌کاوی، پیش‌بینی، دانشگاه جامع علمی کاربردی

* عضو هیات علمی دانشگاه جامع علمی کاربردی، البرز، ایران Tezerjan@yahoo.com

** مدیرکل دفتر سنجش و آزمون دانشگاه جامع علمی کاربردی، تهران، ایران alasanjesh@yahoo.com

*** عضو هیات علمی دانشگاه جامع علمی کاربردی، تهران، ایران mollabagher@gmail.com

نویسنده مسؤول یا طرف مکاتبه: مریم ملباقر

تاریخ پذیرش مقاله: ۱۳۹۹/۰۲/۲۸

تاریخ دریافت مقاله: ۱۳۹۹/۰۱/۱۱

مقدمه

امروزه در اکثر دانشگاه‌های ایران، بانک‌های اطلاعاتی وسیعی از ویژگی‌ها و عملکرد دانشجویان موجود است که حجم بالایی از اطلاعات مربوط به سوابق آموزشی، تحصیلی و غیره را شامل می‌شود. نرم‌افزارهای رایانه‌ای به کار گرفته شده برای این منظور، غالباً فقط برای مکانیزه کردن وضع موجود، اجرای پرس‌وجوهای معمولی و برنامه‌ریزی کوتاه مدت اداری پاسخ‌گو هستند در حالی که در عمق درون این حجم داده‌ها، الگوها و روابط بسیار جالبی میان پارامترهای مختلف به صورت پنهان باقی می‌ماند. داده‌کاوی^۱ یک تکنیک میان‌رشته‌ای برای اکتشاف این الگوها است که از علوم یادگیری ماشین، تشخیص الگو^۲، آمار، پایگاه داده و بصری‌سازی^۳ به منظور استخراج اطلاعات از پایگاه‌های داده بزرگ استفاده می‌کند (Cabena et al., 1998). دانش قابل کشف از طریق داده‌کاوی در حوزه آموزش، نه تنها قابل استفاده صاحبان سیستم؛ یعنی مدرسين و مسئولین آموزشی بلکه قابل استفاده برای کاربران سیستم یعنی دانشجویان نیز است (Shafiepourm & Nazari, 2013). از طرفی برای بهبود عملکرد تحصیلی دانشجویان، نیاز به اطلاعات در مورد وضعیت دانشجویان است؛ زیرا دانشجویان دارای سطوح مختلف انگیزه، درک متفاوتی از آموزش و یادگیری و شرایط محیطی متفاوت از یکدیگر هستند (Kumar et al., 2011). از آنجایی که این پژوهش در بستر دانشگاه جامع علمی کاربردی و در خصوص آزمون متمرکز صورت گرفته است ارائه توضیحاتی در معرفی این دانشگاه و نیز آزمون متمرکز مفید خواهد بود. دانشگاه جامع علمی کاربردی، بزرگترین متولی نظام آموزش عالی علمی کاربردی با حدود ۹۰۰ برنامه درسی، بیش از ۶۰۰ مرکز آموزشی فعال، بیش از ۴۰۰/۰۰۰ دانشجو، بیش از ۱/۰۰۰/۰۰۰ دانش‌آموخته در سراسر کشور با ۱۷ درصد سهم از کل دانشجویان دانشگاهی کشور است که زیر نظر وزارت علوم، تحقیقات و فن‌آوری 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 و اجتماعی و فرهنگی کشور ممکن سازد (Uast, 2020).

-
1. Data Mining
 2. Pattern Recognition
 3. Visual Modeling

با توجه به تأکید رییس دانشگاه مبنی بر اجرای آزمون‌های متمرکز در دروس منتخب در هر مقطع تحصیلی و به منظور حفظ و ارتقاء کیفیت آموزش‌ها و نظارت بر حسن اجرای صحیح آن، مصوبه اجرای برگزاری آزمون‌های متمرکز پایان نیم‌سال در مردادماه ۱۳۹۷ اعلام گردید. در مرحله اول صرفاً در نیم‌سال اول تحصیلی ۹۸-۱۳۹۷ در استان البرز به صورت آزمایشی انجام شد و در نیم‌سال دوم تحصیلی ۹۸-۱۳۹۷ نیز در استان‌های البرز، آذربایجان شرقی، چهارمحال و بختیاری، سمنان، قم و کرمانشاه (۱۲۰ مرکز آموزشی، ۱۹/۲۰۷ نفر درس و ۲۳۵ رشته) انجام شد. معیار انتخاب این استان‌ها، انتخاب یک استان از هر یک از مناطق ۹ گانه دانشگاه (به استثناء مناطق درگیر حوادث غیرطبیعی) و اعلام آمادگی رئیس واحد استان می‌باشد. با توجه به وقوع بلایای طبیعی از جمله سیل و زلزله در ۳ منطقه از مناطق مربوط، هیچ استانی برای اجرای آزمایشی در سطح دوم انتخاب نشد. برخی از اهداف برگزاری آزمون متمرکز عبارتند از:

- حفظ و ارتقاء 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

در این پژوهش هدف اصلی، ارائه مدلی جهت پیش‌بینی نمرات دانشجویان مبتنی بر روش‌های داده‌کاوی و با استفاده از داده‌های مربوط به برگزاری آزمون در نیم‌سال دوم سال تحصیلی ۹۸-۱۳۹۷ و نیز بهره‌برداری از نتایج آن در نیم‌سال‌های آتی است. اهداف فرعی این پژوهش عبارتند از:

اصلاح فرآیندهای مربوط آزمون متمرکز بر پایه نتایج حاصله از پژوهش؛
علمی‌تر و دقیق‌تر شدن اجرا و نظارت بر آزمون متمرکز (کشف داده‌های خارج از الگو).

پژوهش‌های بسیاری در ایران و نیز در سطح جهان صورت گرفته که برخی از آن‌ها به اختصار معرفی می‌شوند:

جدول ۱: پیشینه پژوهش در منابع داخلی

نویسنده/نویسندگان	توضیحات
(Ghoddosi et al., 2019)	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، دانشجویان رشته مهندسی صنایع دانشگاه تربت حیدریه، پرداختند.
(Hasani & Bazrafshan, 2018)	در این پژوهش یک مدل داده‌کاوی مبتنی بر تکنیک 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.
(Shamloo et al., 2015)	در این پژوهش، برای پیش‌بینی رفتار دانشجویان مؤسسات آموزش عالی از تکنیک‌های مختلف درخت تصمیم از جمله CART, CHAID, و C5.0 استفاده شده است.
(Afrash et al., 2014)	پژوهشگران در پژوهشی با عنوان "روش‌های داده‌کاوی برای سیستم مدیریت سطح علمی دانشجویان و مؤسسات آموزش عالی" به بررسی انواع داده‌های آموزشی که در اختیار مؤسسات آموزشی عالی است پرداخته‌اند. پژوهشگران دریافته‌اند که روش‌های داده‌کاوی برای این کار به کار بردند. این پژوهش روی ۳۲ دانشجوی شرکت‌کننده در کلاس انفورماتیک دوره کارشناسی ارشد اقتصاد در جامعه نوین در بازه سال‌های ۹۰-۱۳۸۹، ۳۲ دانشجو در سال تحصیلی ۹۱-۹۰ و ۳۲ دانشجو در سال تحصیلی ۹۲-۱۳۹۱ انجام شده است. نتایج این پژوهش بر این موضوع که داده‌کاوی به طور کلی محدود به مجموعه داده‌های با مقیاس بزرگ نیست، صحنه می‌گذارد. با این همه اکثر محققان در تحقیقات پیشین، این یافته را پیش‌بینی کرده بودند.
(Ghadirimodares, 2013)	نویسندگان ابتدا به بررسی اجمالی مفاهیم و روش‌های داده‌کاوی آموزشی پرداختند و سپس مروری بر کارهای انجام شده در این زمینه در سال‌های اخیر انجام شده است. در ادامه با بررسی موردی و با کمک تکنیک‌های داده‌کاوی از جمله تکنیک شبکه عصبی، خوشه‌بندی و درخت تصمیم به ارزیابی نمرات نهایی دانشجویان پرداختند و در نهایت با مقایسه نتایج حاصل، میزان دقت تکنیک‌ها با هم مقایسه شده است.
(Yaghini et al., 2013)	در این پژوهش با استفاده از تکنیک‌های داده‌کاوی، وضعیت تحصیلی آتی دانشجویان شامل معدل نیم‌سال آینده، معدل کل در زمان فارغ‌التحصیلی و وضعیت فارغ‌التحصیلی در دانشگاه علم و صنعت ایران با استفاده از شبکه‌های عصبی، درخت‌های تصمیم و SVM پیش‌بینی شده است.
(Minaee et al., 2013)	در این پژوهش با استخراج قوانین انجمنی و تحلیل، خوشه‌بندی دانشجویان سال اول دانشگاه قم مورد بررسی تحلیلی قرار گرفته است و ارتباط و همبستگی وضعیت تحصیلی دانشجویان قبل از ورود به دانشگاه و تاثیر آن بر افت تحصیلی وی بعد از ورود به دانشگاه بررسی شده است.
(Heidari & Yaghini, 2010)	پژوهشی با عنوان "دسته‌بندی و پیش‌بینی وضعیت تحصیلی دانشجویان با استفاده از تکنیک‌های داده‌کاوی" با استفاده از داده‌های دموگرافیک و سوابق تحصیلی دانشجویان با کمک تکنیک‌های دسته‌بندی درخت تصمیم، رگرسیون لجستیک، نزدیکترین همسایگی و شبکه‌های عصبی مدل‌های مختلفی را برای پیش‌بینی نمرات دانشجویان دانشگاه علم و صنعت انجام شده است. در نهایت مدل نزدیک‌ترین همسایگی را به عنوان بهترین مدل معرفی نمودند.

جدول ۲: پیشینه پژوهش در منابع خارجی

نویسنده / نویسندگان	توضیحات
(Waheed et al., 2020)	در این پژوهش با استفاده از روش شبکه عصبی مصنوعی عمیق، مدلی جهت پیش‌بینی عملکرد آموزشی دانشجویان و مشخص کردن زمان مداخله احتمالی در فرآیند آموزشی دانشجویان به منظور جلوگیری از افت تحصیلی، ارائه شده است.
(Adekitan & Salau, 2019)	در این پژوهش، داده‌های اطلاعات دانشجویان دانشگاه نیجریه در سه سال اول تحصیل با استفاده از ۶ الگوریتم انجام شده است. از نتیجه این تحقیق برای پیش‌بینی امکان فارغ‌التحصیلی این دانشجویان استفاده شد.
(Khanbaba ei et al., 2018)	هدف این پژوهش، ایجاد یک چارچوب یکپارچه برای استفاده از تکنیک‌های داده‌کاوی برای بهبود فرآیندهاست. یک چارچوب در سه بخش و پنج مرحله تدوین شده و پیشنهاداتی برای بهبود فرآیندهای کاری در حوزه داده‌کاوی آموزشی ارائه شد.
(yahya a., 2017)	در این پژوهش، طبقه‌بندی دانشجویان در حوزه داده‌کاوی آموزشی با استفاده از روش PSC ^۱ انجام شده است.
(Harwati et al., 2015)	هدف این پژوهش، تمرکز بر روی نگاشت دانشجویان با استفاده از الگوریتم خوشه‌بندی K به منظور تشخیص الگوی پنهان و طبقه‌بندی دانشجویان بر اساس اطلاعات جمعیت‌شناسی (جنسیت، محل اقامت، GPA ^۲ ، درجه دوره‌های خاص و به طور متوسط حضور در کلاس) است. در این پژوهش اطلاعات حدود ۳۰۰ دانشجو مورد استفاده قرار گرفته است.
(Mujib et al., 2013)	هدف این پژوهش، طبقه‌بندی دانشجویان برای پیش‌بینی قبولی دانشجویان است. متغیرهای پژوهش نیز شامل نمرات، میانگین نمرات و جنسیت بود. این طبقه‌بندی بر اساس الگوریتم NBC ^۳ انجام شد و جداول پیش‌بینی احتمال فارغ‌التحصیلی دانشجویان، بر مبنای آن ساخته شد.
(Dejaeger et al. (2011)	پژوهشگران با بررسی مدل‌های داده‌کاوی، ویژگی‌های اصلی رضایت دانشجویان را با استفاده از تکنیک‌های داده‌کاوی، شناسایی و نتایج را به منظور حمایت از مدیریت در فرآیند تصمیم‌گیری مورد بررسی قرار دادند.
(Varghese et al (2010)	الگوریتم خوشه‌بندی k-means روی اطلاعات ۸/۰۰۰ دانشجو انجام شد. این خوشه‌بندی بر اساس پنج متغیر: میزان حضور، نمره آزمون، نمرات سمینارها، تکالیف و نمرات دانشجویان صورت گرفت.
(Kotsiantis, 2009)	این پژوهش با استفاده از روش‌های مدرن یادگیری ماشین و تحلیل هزینه مؤثر انجام شد، هدف اصلی پژوهشگر، پیش‌بینی افت تحصیلی دانشجویان جدید با دقت قابل قبول بود.
Romero and Ventura (2007)	پژوهشگران در سال‌های ۱۹۹۵ تا ۲۰۰۵، یک نظرسنجی انجام دادند که در آن به بررسی انواع مختلف سیستم‌های آموزشی و همچنین نحوه استفاده از تکنیک‌های داده‌کاوی برای هر یک از آن‌ها پرداختند.

1. Particle Swarm Classification
2. Grade Point Average
3. Naive Bayes classifier

روش

با توجه به طیف گسترده‌ای از تکنیک‌های داده‌کاوی و تنوع انواع مختلف داده در پایگاه داده به منظور نیل به اهداف داده‌کاوی، باید فرآیند انجام پژوهش به صورت کاملاً دقیق تعریف گردد. با در نظر گرفتن این که پژوهش حاضر در فضای دانشگاهی و از طریق اطلاعات موجود در سامانه‌های اینترنتی دانشگاه صورت گرفته، لذا در مرحله کسب اطلاعات، مشکلی وجود نداشته است. روش پژوهش بر حسب هدف، از نوع کاربردی و از متدولوژی داده‌کاوی برای آن استفاده شده است و سه روش کاوش، توصیف و پیش‌بینی را با ابزارهای داده‌کاوی در بردارد. روش پژوهش از نظر روش جمع‌آوری اطلاعات در بخش مرور ادبیات، روش کتابخانه‌ای است و در بخش تفسیر نتایج داده‌کاوی، روش پژوهش میدانی است. این پژوهش به صورت مطالعه موردی روی داده‌های آزمون متمرکز دانشگاه جامع علمی کاربردی اجرا شده است.

پایگاه داده مورد استفاده SQL Server است که خروجی داده‌های دانشجویان، مدرسان و نمرات در فایل‌های مجزا در قالب نرم افزار Excel در اختیار محققین قرار گرفت. برخی از اطلاعات تجمیعی و گزارشی با استفاده از Query های طراحی شده در نرم‌افزار Access به دست آمد. سپس داده‌ها در نرم‌افزار SPSS Modeler وارد شد و پیش‌پردازش و نیز تحلیل‌های داده‌کاوی و نمودارهای ارائه خروجی نتایج با استفاده از آن صورت پذیرفت. برخی از نمودارها نیز در نرم‌افزار Excel ترسیم شده است.

پس از دریافت کل داده‌های مورد نیاز در ۶ استان، مراحل انجام کار به شرح زیر می‌باشد:

- ۱- پیش‌پردازش داده‌ها: حذف اطلاعات غیر مرتبط و استخراج متغیرهای کلیدی در شناخت الگو، پیش‌پردازش نامیده می‌شود (Kennerdy et al., 1998). این مرحله شامل انتخاب ویژگی‌ها، پاک‌سازی داده‌ها و تبدیل داده‌ها است.
- ۲- انتخاب ویژگی: به دلیل این که مدل‌های ناپارامتریک در طبقه‌بندی مبتنی بر داده هستند نیاز به صرف زمان و هزینه زیاد برای کسب داده‌های مدل است لذا بهتر است ویژگی‌ها و داده‌هایی جمع‌آوری شود که از اهمیت بیش‌تری در ساخت مدل طبقه‌بندی برخوردار هستند. در ساخت مدل مناسب طبقه‌بندی، نیاز به داده‌های آموزشی با کیفیت مناسب است. (Salappa et al., 2007). 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 شکل ۱ نشان داده شده است.

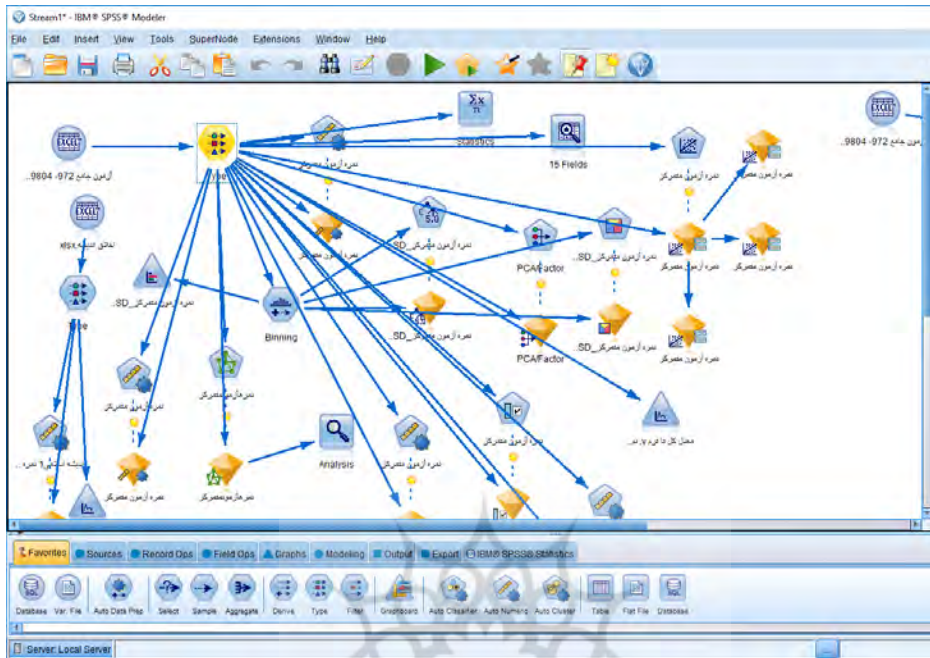
۳- پاک‌سازی داده‌ها: تمیز کردن شامل: شناسایی مقادیر گمشده، متناقض یا اشتباه است. ابزارهای به کار رفته در این مرحله از فرآیند، شامل ابزارهای گرافیکی برای تهیه تصویری از توزیع‌های آماری مانند: ماکسیمم، حداقل‌ها، میانگین مقادیر و اندازه‌های کوچک است. برخی از ورودی‌ها به وضوح نامعتبر بودند که ناشی از خطای انسانی یا تکامل سیستم گزارش‌دهی مشکل شده است. در این مرحله داده‌های NULL در فایل اکسل به مقدار تهی " " تبدیل شدند؛ زیرا نرم‌افزارهای مورد استفاده، آن‌ها را به عنوان کاراکترهای متنی در نظر می‌گرفتند.

شمای کلی مدل توسط نرم‌افزار (SPSS Modeler) Clementine، در شکل ۲ نشان داده شده است.

Field	Measurement	Values	Missing	Check	Role
استان	Nominal	"آذربایجان شرقی"، "آ..."		None	Input
کد استان	Continuous	[10.0,40.0]		None	Input
نام شهر	Nominal	"اسلام آباد عرب"...		None	Input
کد مرکز	Continuous	[13004.0,940...		None	Input
نام مرکز	Nominal	مجتمع آموزشی و بی...		None	Input
شماره دانشجویی	Continuous	[9.313100319...		None	Input
نام دانشجو	Typeless			None	None
نام خانوادگی دانشجو	Typeless			None	None
جنسیت	Flag	مردان		None	Input
تاریخ تولد	Continuous	[1948-11-19,2...		None	Input
رشته تحصیلی	Nominal	"مهندسی فناوری اس..."		None	Input
کد رشته	Continuous	[961111.0,9.3...		None	Input
کد درس	Continuous	[26971.0,321...		None	Input
نام درس	Nominal	"اخلاق اسلامی" "اند..."		None	Input
نمره کل	Continuous	[0.0,27.0]		None	Input
نمره آزمون متمرکز	Continuous	[0.0,14.0]		None	Both
نمره استاد	Continuous	[-10.0,20.0]		None	Both
نرم	Continuous	[972.0,972.0]		None	Input
تعداد واحد اخذ شده	Continuous	[0.0,24.0]		None	Input
تعداد واحد پاس شده	Continuous	[0.0,24.0]		None	Input
معدل نرم	Continuous	[0.0,27.0]		None	Input
معدل کل تا نرم	Continuous	[0.0,19.88]		None	Input
نام مدرس	Nominal	آرش، امین، اباصلت، ا...		None	Input
نام خانوادگی مدرس	Typeless			None	None
کد مدرس	Continuous	[1071.0,9.510...		None	Input
جنسیت مدرس	Nominal	زن، مرد، نامشخص		None	Input
تاریخ تولد مدرس	Continuous	[1934-03-21,1...		None	None

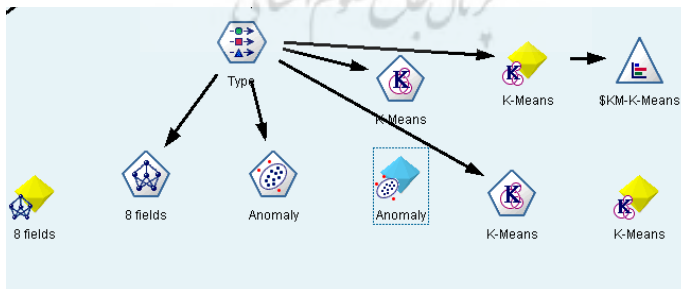
شکل ۱: ویژگی‌های انتخاب شده از دانشجو، مدرس و مرکز آموزشی و نوع داده‌ای هر یک

۴- تهیه آمارهای اولیه و کلی: در این مرحله آمارهای کلی و دموگرافیک مربوطه دانشجویان، درس و استان‌ها ارائه شده است. یافته‌ها در بخش بعدی ارائه می‌گردد.



شکل ۲ : شمای کلی مدل

۵- انجام داده کاوی و ارائه مدل: انجام داده کاوی و ارائه مدل با استفاده از یافته‌های پژوهش و در نهایت ارائه نتایج پژوهش در این مرحله انجام شده است. بعد از انجام تحلیل‌های عمومی، مدل‌های داده کاوی به کار گرفته شد. مدل‌های مختلفی مورد استفاده، در شکل ۳ نمایش داده شده است.



شکل ۳ : مدل‌های مورد بررسی

سؤالات این آزمون‌ها به صورت ۲۸ سؤال تستی هر کدام به ارزش ۰,۵ نمره و توسط مدرسان علمی کاربردی این دروس در ترم جاری، طراحی و بازبینی شده است. سطح سؤالات در چهار بخش (بسیار آسان، آسان، متوسط و دشوار) و منطبق با سرفصل کتاب‌های معرفی شده، تعیین شده است. مدت زمان برگزاری آزمون ۴۰ دقیقه و آزمون، فاقد نمره منفی بوده است. نمره آزمون (۲۰ نمره) شامل ۱۴ نمره آزمون متمرکز و ۶ نمره در اختیار مدرس، جهت ثبت عملکرد دانشجویان در انجام فعالیت کلاسی و حل تمرین در نظر گرفته شده است.

یافته‌ها

با توجه به داده‌های اطلاعات، آمار اولیه تعداد دانشجو-آزمون و نیز معدل نمره آزمون به تفکیک استان در جدول ۳ نشان داده شده است.

جدول ۳: تعداد دانشجو-آزمون را به تفکیک استان

ردیف	استان	تعداد دانشجو-آزمون	معدل نمره آزمون
۱	آذربایجان شرقی	۳/۴۵۶	۷/۴۵
۲	البرز	۷/۸۳۹	۷/۸۴
۳	چهارمحال و بختیاری	۱/۴۷۶	۷/۵۱
۴	سمنان	۱/۷۰۲	۷/۶۴
۵	قم	۲/۳۹۱	۷/۸۸
۶	کرمانشاه	۲/۳۴۳	۷/۲۸
	جمع کل	۱۹/۲۰۷	۷/۶۶

همان‌طور که در جدول ۳ مشاهده می‌شود بیش‌ترین دانشجویان شرکت‌کننده در آزمون‌های متمرکز مربوط به استان البرز و کم‌ترین آنان مربوط به استان سمنان می‌باشد. بالاترین میانگین مربوط به استان قم و کم‌ترین میانگین مربوط به استان کرمانشاه می‌باشد. البته با توجه به بلایای طبیعی در این استان نظیر زلزله در نیم‌سال تحصیلی می‌توان این موضوع را یکی از عوامل ناکامی دانشجویان این استان به شمار آورد که علی‌رغم بلایای طبیعی در این استان، رئیس واحد استانی آمادگی خود را مبنی بر برگزاری این آزمون اعلام داشت.

جدول ۴: تعداد دانشجو-آزمون را به تفکیک جنسیت

ردیف	جنسیت	تعداد
۱	زن	۷/۰۳۶
۲	مرد	۱۲/۱۷۱
جمع کل		۱۹/۲۰۷

همان‌طور که در جدول ۴ مشاهده می‌شود مردان سهم بیش‌تری از دانشجویان را نسبت به زنان در این آزمون‌ها به خود اختصاص داده‌اند.

جدول ۵: تعداد دانشجو-آزمون و معدل نمره آزمون متمرکز را به تفکیک درس

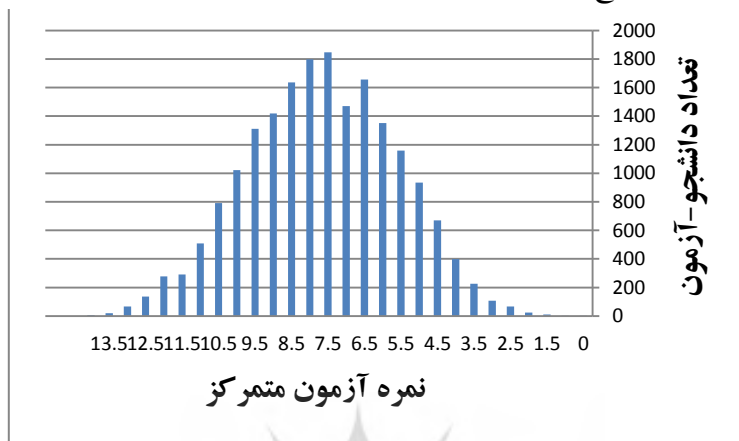
ردیف	عنوان درس	تعداد دانشجو-آزمون	معدل نمره آزمون متمرکز
۱	اخلاق اسلامی	۲/۲۲۳	۶/۸۶
۲	اندیشه اسلامی ۱	۳/۲۳۴	۷/۶۶
۳	اندیشه اسلامی ۲	۲/۲۵۴	۶/۰۰
۴	تاریخ تحلیلی صدر اسلام	۲/۳۵۶	۷/۳۲
۵	تفسیر موضوعی قرآن	۲/۶۲۱	۸/۸۰
۶	دانش خانواده و جمعیت	۲/۹۶۹	۸/۹۸
۷	مهارت‌ها و قوانین کسب و کار	۱/۶۱۷	۷/۵۵
۸	مهارت‌های مسئله‌یابی و تصمیم‌گیری	۱/۹۲۱	۷/۵۰
جمع کل		۱۹/۲۰۷	۷/۶۶

مطابق جدول ۵ درس اندیشه اسلامی ۱ دارای بیش‌ترین دانشجوی شرکت‌کننده بوده است و درس مهارت‌ها و قوانین کسب و کار با توجه به اختصاصی بودن آن از کم‌ترین دانشجوی برخوردار شده است. بالاترین نمره درس "دانش خانواده و جمعیت" و کم‌ترین نمره درس "اندیشه اسلامی ۲" بوده است.

جدول ۶ تعداد دانشجو-آزمون را به تفکیک درس و وضعیت نمره کل

تعداد دانشجو	درس و وضعیت نمره کل
۲/۲۲۳	اخلاق اسلامی
۱/۷۲۷	۱۵ تا ۱۰
۲۷۱	۲۰ تا ۱۵
۲۲۵	مردود
۳/۲۳۴	اندیشه اسلامی ۱
۲/۱۳۴	۱۵ تا ۱۰
۸۲۹	۲۰ تا ۱۵
۲۷۱	مردود
۲/۲۵۴	اندیشه اسلامی ۲
۱/۶۰۷	۱۵ تا ۱۰
۱۷۲	۲۰ تا ۱۵
۴۷۵	مردود
۲/۳۶۸	تاریخ تحلیلی صدر اسلام
۱/۸۲۳	۱۵ تا ۱۰
۴۲۲	۲۰ تا ۱۵
۱۲۳	مردود
۲/۶۲۱	تفسیر موضوعی قرآن
۱/۳۵۹	۱۵ تا ۱۰
۱/۱۹۱	۲۰ تا ۱۵
۷۱	مردود
۲/۹۶۹	دانش خانواده و جمعیت
۱/۴۴۷	۱۵ تا ۱۰
۱/۴۶۱	۲۰ تا ۱۵
۶۱	مردود
۱/۶۱۷	مهارت‌ها و قوانین کسب و کار
۲/۳۳۲	۱۵ تا ۱۰
۲۷۸	۲۰ تا ۱۵
۱۰۷	مردود
۱/۹۲۱	مهارت‌های مسئله‌یابی و تصمیم‌گیری
۱/۴۶۲	۱۵ تا ۱۰
۳۴۵	۲۰ تا ۱۵
۱۱۴	مردود
۱۹/۲۰۷	تعداد کل

در شکل ۶ توزیع نمره آزمون متمرکز نشان داده شده است. همان طور که در شکل نیز قابل مشاهده است، دارای توزیع نرمال می باشد.



شکل ۶: توزیع نمره آزمون متمرکز را بر اساس تعداد دانشجو-آزمون

به طور کلی دو رویکرد عمده در داده کاوی وجود دارد: آنهایی که الگوها را استخراج می کنند و آنهایی که از اطلاعات کاوش شده، مدل تولید می کنند. تکنیک هایی مانند شبکه های عصبی و درخت طبقه بندی، مستقیماً مدل های برآورد یا تخمین را از داده ها تولید می کنند. این مدل ها دانش را در یک فرم قابل استفاده برای سازمان کد می کنند. در ادامه به ارائه آمار کلی و ویژگی های مؤثر در ارائه مدل پیش بینی در قالب جدول ۷ می پردازیم:

جدول ۷: آمار کلی ویژگی های مؤثر در مدل پیش بینی

ردیف	عنوان	Count	Mean	Min	Max	Range	Variance	Median
۱	نمره کل	۱۹/۲۰۱	۱۳/۱۸	۰	۲۷/۰۰	۲۷/۰۰	۵/۶۵	۱۳/۵۰
۲	نمره آزمون متمرکز	۱۹/۲۰۷	۷/۶۶	۰	۱۴/۰۰	۱۴/۰۰	۴/۲۴	۷/۵۰
۳	نمره استاد	۱۹/۲۰۱	۵/۵۱	-۱۰	۲۰/۰۰	۳۰/۰۰	۰/۹۹	۶/۰۰
۴	تعداد واحد اخذ شده	۱۹/۲۰۷	۱۶/۹۶	۰	۲۴/۰۰	۲۴/۰۰	۱۱/۴۰	۱۸/۰۰
۵	تعداد واحد پاس شده	۱۹/۲۰۷	۱۵/۶۳	۰	۲۴/۰۰	۲۴/۰۰	۱۸/۳۵	۱۷/۰۰
۶	معدل ترم	۱۹/۱۶۲	۱۴/۸۹	۰	۲۷/۰۰	۲۷/۰۰	۶/۱۲	۱۵/۰۵
۷	معدل کل تا ترم	۱۹/۱۸۳	۱۵/۱۰	۰	۱۹/۸۸	۱۹/۸۸	۵/۱۴	۱۵/۱۷

نتایج در درخت تصمیم، به صورت قواعد "اگر ... آنگاه ..." نیز قابل ارائه است. در ادامه برخی از قواعد استخراج شده از درخت تصمیم، ارائه شده است. در این مدل از الگوریتم چاید^۱، استفاده شده است و فیلد هدف، نمره آزمون متمرکز بوده است. درخت تا عمق ۳ با ۸۸ گره ارائه شده است و ۱۹/۲۰۷ رکورد را مورد بررسی قرار داده است.

and 13.03 < معدل ترم

or نام درس = تاریخ تحلیلی صدر اسلام (

or نام درس = مهارت‌ها و قوانین کسب و کار

and) نام درس = مهارت‌های مسأله‌یابی و تصمیم‌گیری

or استان = آذربایجان شرقی (

or استان = البرز

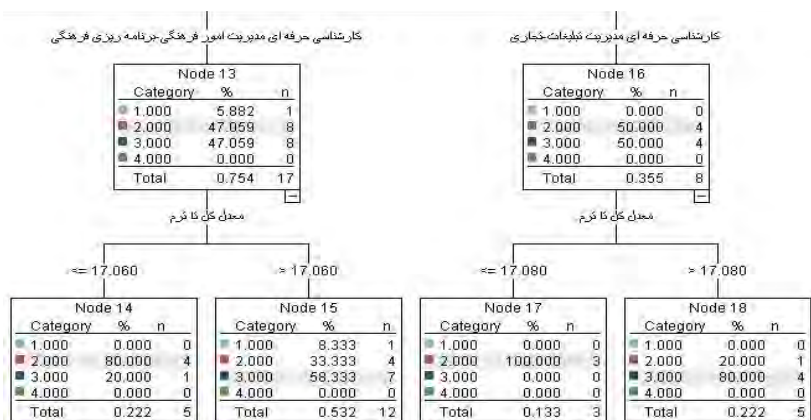
or استان = سمنان

(استان = قم)

Mean Rule=6.634
Record percentage=4.7
Record count=896

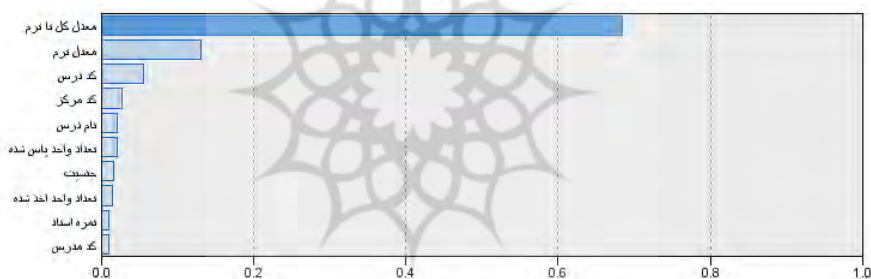
برای مثال، در تشریح عملکرد درخت تصمیم، مطابق قاعده استخراج شده بالا، دانشجویانی که دارای معدل ترم کم‌تر از ۱۳/۰۳ بوده‌اند برای درس تاریخ تحلیلی صدر اسلام، مهارت‌ها و قوانین کسب و کار یا مهارت‌های مسأله‌یابی و تصمیم‌گیری در استان‌های آذربایجان شرقی، البرز، سمنان یا قم که ۴/۷ درصد کل رکوردها را پوشش داده است و شامل ۸۹۶ رکورد است دارای میانگین نمره ۶,۶۳۴ هستند.

همان‌طور که در قواعد بالا دیده می‌شود معدل ترم بالاتر منجر به اخذ نمره آزمون متمرکز بالاتر شده است. این قواعد می‌تواند قابل استفاده برای پیش‌بینی نمرات آزمون متمرکز سایر دانشجویان باشد. شکل ۷ نمایی از خروجی یکی از درخت‌های تصمیم‌گیری استفاده شده در این پژوهش نشان داده شده است.



شکل ۷: نمایی از یکی از خروجی‌های درخت تصمیم‌گیری

در شکل ۸ اهمیت شاخص‌های پیش‌بینی‌کننده نمره آزمون متمرکز، در یکی از درخت‌های تصمیم‌گیری به صورت گرافیکی ارائه شده است.

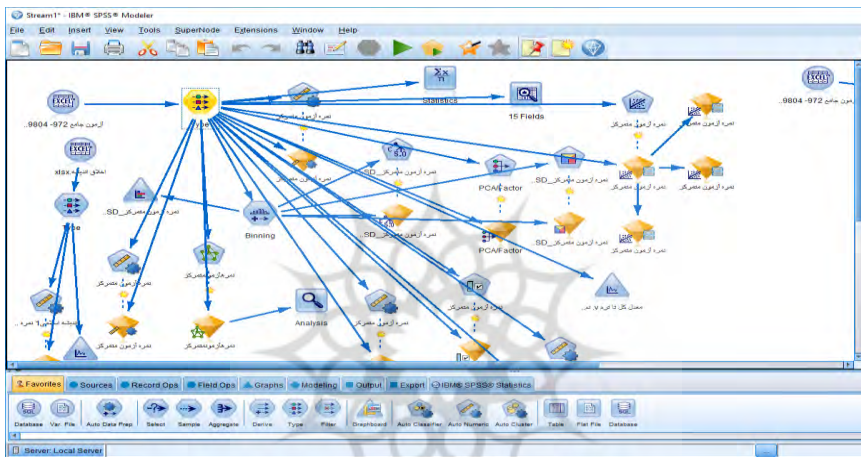


شکل ۸: اهمیت شاخص‌های پیش‌بینی‌کننده نمره آزمون متمرکز در یکی از مدل‌های درخت تصمیم‌گیری

همان‌طور که در شکل ۸ دیده می‌شود مهم‌ترین شاخص‌های پیش‌بینی‌کننده نمره آزمون متمرکز در این درخت تصمیم‌گیری به ترتیب عبارتند از: معدل کل تا ترم، معدل ترم، کد درس، کد مرکز، نام درس و ... و مواردی که معدل ترم پایین و نمره آزمون بالا دارند به تقلب مشکوک هستند لذا این دانشجویان باید در آزمون بعدی به طور دقیق‌تری نظارت شوند.

مدل‌سازی مبتنی بر داده، عمده‌ترین روش و وجه تمایز داده‌کاوی محسوب می‌شود. تکنیک‌های این روش شامل طبقه‌بندی و پیش‌گویی، خوشه‌یابی، مصورسازی داده‌ها، شبکه‌های عصبی، قواعد جمعی، درخت تصمیم‌گیری و ... است. با استفاده از این روش‌ها دانش

پنهان در داده‌ها کشف می‌شود (Wong & Leung, 1999; Han & Kamber, 2001). شکل ۹ نمای کلی مدل داده‌کاوی را از ورود داده‌ها تا پیش پردازش، تکنیک‌های آماری، مدل‌های شبکه عصبی، درخت تصمیم‌گیری، رگرسیون و ... را نشان می‌دهد. همان‌طور که در شکل ۹ دیده می‌شود از مدل‌های مختلفی از جمله: آمار، شبکه عصبی، درخت تصمیم‌گیری، تحلیل عاملی، PCA، خوشه‌بندی و ... برای تحلیل‌های مختلف استفاده شده است ولی به دلیل کمبود فضای تقریر صرفاً نتایج دو مدل C5 و درخت تصمیم‌گیری در این مقاله ارائه شده است.



شکل ۹: نمای کلی مدل

برای سنجش اعتبار نتایج به دست آمده، تحلیل‌ها با روش‌های مختلف انجام گردیده است که نتایج این تحلیل‌ها یکدیگر را تأیید می‌نمایند. هم‌چنین استفاده از روش‌های مختلف، باعث خواهد شد که نتایج روش‌های مختلف، بررسی و مورد استفاده قرار گیرد. به جای استفاده از یک الگوریتم بایستی محیطی را فراهم ساخت تا بتوان از الگوریتم‌ها یا متغیرهای مختلف استفاده کرد تا فرد بررسی‌کننده توان استفاده از مناسب‌ترین الگوریتم را داشته باشد علاوه بر آن، نتایج به مدیران و خبرگان دانشگاه ارائه و توسط آنان مورد تأیید قرار گرفت.

تحلیل‌های داده‌کاوی، نتایج ارزشمندی داشت که البته باید مورد تفسیر قرار گیرد. در ادامه برخی از نتایج اشکالات جزئی که در اجرا و نظارت بر آزمون‌های متمرکز بود ارائه شده‌اند.

با توجه به بررسی‌های آماری و تحلیل‌های اولیه داده‌کاوی پیشنهادهای زیر برای بهبود و رفع مشکلات آزمون‌های متمرکز اعلام می‌شود:

۱. مهم‌ترین شاخص‌های پیش‌بینی کننده نمره آزمون متمرکز در این درخت تصمیم‌گیری به ترتیب عبارتند از: معدل کل تا ترم، معدل ترم، کد درس، کد مرکز، نام درس و ...
۲. نمره آزمون با معدل کل، رابطه مستقیم اما ضعیف دارد.
۳. متوسط نمره دخترها ۸/۴ پسرها ۷/۶ است یعنی از منظر آزمون متمرکز برای این دروس، دانشجویان دختر موفق‌تر از دانشجویان پسر هستند. در تمام دروس میانگین نمرات دانشجویان زن بالاتر از میانگین نمرات دانشجویان مرد است.

بحث و نتیجه‌گیری

این پژوهش با بررسی پژوهش‌های پیشین در داخل و خارج کشور به صورت مطالعه موردی در دانشگاه جامع علمی کاربردی انجام شد. در این مقاله یک مدل پیش‌بینی نمرات آزمون متمرکز مبتنی بر داده‌کاوی و بر اساس ویژگی‌های معدل کل تا ترم، معدل ترم، کد مرکز، کد درس، تعداد واحد اخذ شده، نمره استاد ارائه شد. (Adekitan & Salau, 2019) به پیش‌بینی فارغ‌التحصیلی و (Mujib et al., 2013) به پیش‌بینی قبولی دانشجویان و (Kotsiantis, 2009) به پیش‌بینی افت تحصیلی پرداختند؛ اما پژوهش حاضر، در حین تحصیل ارزیابی دانشجو را بر عهده دارد و هم‌زمان مشکلات طراحی سؤالات را نیز شناسایی می‌کند. (yahya a., 2017) از روش PSC استفاده نموده، در حالی که پژوهش حاضر از روش‌های پیش‌تری استفاده کرده است. (Harwati et al., 2015) روی ۳۰۰ نفر و (Varghese et al., 2010) روی ۸/۰۰۰ نفر و (Hasani & Bazrafshan, 2018) فقط در یک دانشکده از خوشه‌بندی استفاده کرده‌اند، و (Afrash et al., 2014) روی ۳۲ دانشجو و (Minaee et al., 2013) نیز صرفاً روی دانشجویان سال اول دانشگاه قم تحلیل کرده است. (Shamloo et al., 2015) از انواع درخت تصمیم‌گیری برای پیش‌بینی رفتار دانشجویان استفاده کرده است. (Ghoddosi et al., 2019) هم صرفاً داده‌کاوی را روی دانشجویان مهندسی صنایع دانشگاه بیرجند انجام دادند؛ اما پژوهش حاضر هم از لحاظ تعداد دانشجو و هم از لحاظ روش‌های مورد استفاده بسیار کامل‌تر بوده است.

نتایج پژوهش نقاط قوت و ضعف آزمون متمرکز و مراکز آموزش علمی کاربردی و دانشجویان را نیز مشخص نمود. برخی از این نتایج عبارتند از:

در بین بالاترین نمرات، تعدادی تک نفره هستند که به احتمال زیاد معرفی به استاد هستند 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 ناظرین اهمیت کم‌تری قایل می‌شوند و احتمال تقلب بالاتر می‌رود (مقایسه میانگین نمره در حوزه‌های کم تعداد و پر تعداد) لذا پیشنهاد می‌شود به جهت کاهش هزینه‌های نظارت و اجرا آزمون دروسی 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) لذا دستورالعمل و نظارتی بر نحوه تخصیص نمره مدرس، تدوین و ارائه شود.

از آنجا که داده‌کاوی یک فرآیند مستمر است پیشنهاد می‌گردد که در پژوهش‌های آتی نسبت به اجرای سایر تکنیک‌های داده‌کاوی اقدام شود تا با ایجاد سیستم‌های پشتیبان تصمیم‌مبتنی بر این الگوها و تحلیل حساسیت آن‌ها کمک مؤثری در تصمیم‌گیری بهتر مدیران باشد. با توجه به این موضوع، پیشنهادهای زیر برای تحقیقات آتی ارائه می‌گردد:

بررسی رابطه نمرات آزمون متمرکز با ارزیابی دانشجویان از مدرسین؛

بررسی رابطه نمرات آزمون‌ها با مباحث نظارتی مراکز آموزش علمی کاربردی؛

توسعه اجرای مدل در سطح کشور و ارزیابی سایر استان‌ها؛

اصلاح فرآیندهای مختلف با استفاده از نتایج تحلیل‌های انجام گرفته و بررسی

تغییرات اصلاحات انجام گرفته در نیم‌سال‌های بعد.

References

- 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.
- Afrash, M.R., Khodamoradi, M., & Forozan Monfared, M. (2014). Data Mining Methods for the Academic Level Management System of Students and Higher Education Institutions, First National Conference on Technology Advances in Electrical, *Electronics and Computer Engineering*, Neyshabour, Iran. (in Persian)
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining: From Concepts to Implementation*, Upper Saddle River, NJ: Prentice Hall.
- Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., & Baesens, B. (2011). Gaining insight into student satisfaction using comprehensible data mining techniques, *European Journal of Operational Research*, 218(2), 548 – 562.
- Ghadiri modarres, Z., Vahman, N., & ahmadzadeh, M.R. (2013). Comparison of the accuracy of data mining techniques in training data mining, *7th Iranian Data Mining Conference*, Tehran, Iran. (in Persian)
- Ghoddosi, M., Mirsaeeedi, F., & kosha, H.R. (2019). Predicting and analyzing student performance using data mining techniques to improve academic performance, *Training technology*, 14(1), 1-13. (in Persian)
- Han, J. & Kamber, M., *Data Mining Concepts and Techniques*, First Edition, Morgan Kaufmann Publishers, (2001).
- Harwatia, Alfiania A.P., & Wulandari F.A. (2015). Mapping Student's Performance Based on Data Mining Approach (A Case Study), *The 2014 International Conference on Agro-industry (ICoA): Competitive and sustainable Agroindustry for Human Welfare*, 173 – 177.
- Hasani, A.A., & Bazrafshan, M. (2018). Analysis of Students' Educational Data to Evaluate Academic Success Using Data Mining Approach (Case Study: Shahroud University School of Management and Engineering), *Management and planning in the educational education system*, 11 (2), 187-208. (in Persian)
- Heidari, S., & Yaghini, M. (2010). Classifying and Predicting Students' Academic Status Using Data Mining Techniques, *Higher Education Letter*, new course, 3(12), 107-124. (in Persian)
- Kennerdy, R.L., Lee, Y., Roy, B.V., Reed, C.D., & Lippmann, R.P. (1998). *Solving Data Mining Problems through Pattern Recognition*, S.I., Prentice Hall.
- Khanbabaei M., Movahedi Sobhani F., Alborzi M. and Radfar R. (2018). Developing an integrated framework for using data mining techniques and ontology concepts for process improvemen. *Journal of Systems and Software*, 137, March 2018, 78-95.
- Kotsiantis, S. (2009). Educational data mining: a case study for predicting dropout-prone students, *International Journal of Knowledge Engineering and Soft Data Paradigms*, 1(2), 101.

- Kumar, B.B., & Saurabh P. (2011). Data Mining: A Prediction for Performance Improvement Using Classification. (*IJCSIS*) *International Journal of Computer Science and Information Security*, 9(4), April 2011.
- Minaee, B., Mirfazl, S.S. & Hani, S.H. (2012). Identifying Factors Influencing Students' Academic Failure Using Associative Laws and Cluster Analysis (Case Study: Qom University), *6th Iranian Data Mining Conference, Tehran, Iran*. (in Persian)
- Mujib, R., Hadi S., & Sarosa M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Jurnal EECCIS*, 7(1), 59–64.
- Romero, C. Ventura, S. (2007). Educational data mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33, 135-146.
- Salappa, A., Doumpos, m., & Zopoundidis, C. (2007). Feature selection algorithms in classification problems: an experimental evaluation, *optimizing Methods and Software*, 22(1), 2-5.
- Shafiepourm, F., & Nazari, H. (2013). Designing a Modified Model to Evaluate the Effectiveness of E-Learning Strategies on Students' Academic Success, *Technology of education*, 8(1), 31-37. (in Persian)
- Shamloo, R., Omidvari, M., & Aminfar, F. (2015). Predicting Students' Educational Behavior with Data Mining Approach in Higher Education Institutions (Case Study of Booyen Zahra Azad University), *Third International Conference on Modern Research in Management - Economics and Accountin*, Istanbul, Turke. (in Persian)
- Varghese M.B., Jose, T., & Paulose, J. (2010), Clustering Student Data to Characterise Performance Pattern, *International Journal of Advanced Computer Science and Applications*, *Special Issue on Artificial Intelligence*, 138-140.
- 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.
- Wong M.L., & K.S Leung, (1999). *Data Mining Using Grammer Based Genetic Programming and Applications*, Kluwer Academic Publication.
- Yaghini, M., Akbari, A., & Sharifi, S.M.M. (2008). Predicting students' educational status using data mining techniques, *2th Iranian Data Mining Conference*, Tehran, Iran. (in Persian)
- Yahya, A.A. (2017). Swarm intelligence-based approach for educational data classification, *Journal of King Saud University - Computer and Information Sciences*, 31(1), January 2019, 35-51.