کاربست فنون داده کاوی در تبیین صحت مدل های پیش بینی افت تحصیلی دانشجویان

آزاده فدوی رودسری[۱]، کیوان صالحی[۱*]، ابراهیم خدایی[۱]، علی مقدم زاده[۱]، محمد جوادی‌پور[۱]

۱) گروه روش ها و برنامه های آموزشی و درسی، دانشکده علوم تربیتی و روان شناسی، دانشگاه تهران، تهران، ایران

تاریخ دریافت: ۱۳۹۸/۷/۶ **تاریخ پذیرش: ۱۳۹۸/۹/۱۷**

چکیده

مقدمه: افت تحصیلی دانشجویان یکی از آفت های سیستم آموزش عالی است و وجود سیستمی برای پیش بینی و در نتیجه پیشگیری از وقوع آن، می تواند نقش تعیین کننده ای در ارتقای کیفیت آموزش باشد. استفاده از داده کاوی آموزشی، از موثرترین راهبردهای پایش و بهبود کیفیت آموزش است. این مطالعه با هدف مقایسه فنون مختلف داده کاوی و مشخص ساختن بهترین مدل پیش بینی انجام شده است.

مواد و روش ها: جامعه پژوهش، کلیه دانشجویان دوره کارشناسی رشته های علوم آزمایشگاهی، مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران در سال تحصیلی ۹۸-۱۳۹۷ بودند و نمونه ای به حجم ۵۰۰ نفر انتخاب شد. بعد از اجرا و کنار گذاشتن پرسش نامه های مخدوش، ۱۵۳ پرسش نامه از رشته علوم آزمایشگاهی، ۸۵ پرسش نامه مدیریت اطلاعات سلامت، و ۷۳ پرسش نامه رشته تکنولوژی علوم پرتوی مورد تحلیل قرار گرفت. ابزار جمع آوری داده ها، پرسش نامه سنجش افت تحصیلی دانشجویان با ضریب پایایی ۰/۹۷۱ بود که روایی آن توسط متخصصان تایید شد. داده های جمع آوری شده از دانشجویان با الگوریتم های رگرسیون لجستیک، درخت تصمیم، شبکه عصبی، شبکه بیزین و ماشین بردار پشتیبان و با استفاده از نرم افزار SPPSS vol.16، SMART_PLS-3 و SPPSS_MODELLER-18 کلمنتاین تحلیل شد.

یافته های پژوهش: یافته ها نشان داد، میزان افت تحصیلی در رشته های مختلف در سطح معناداری ۰/۰۵ متفاوت است. در رشته علوم آزمایشگاهی، میزان صحت مدل پیش بین داده ها در الگوریتم های متفاوت به ترتیب بهترین مدل پیش‌بینی عبارت است از: الگوریتم های درخت تصمیم و شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی و رگرسیون لجستیک. در رشته مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی میزان صحت مدل پیش بین داده ها در الگوریتم های متفاوت به ترتیب بهترین مدل پیش بینی عبارت است از: الگوریتم شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی و رگرسیون لجستیک. الگوریتم درخت تصمیم جوابی نداشت. هم چنین آزمون خی دو نشان داد تفاوت موجود بین الگوریتم های به دست آمده از لحاظ آماری در رشته علوم آزمایشگاهی در سطح ۰/۰۵ معنادار نیست، در رشته مدیریت اطلاعات و در رشته تکنولوژی علوم پرتوی در سطح ۰/۰۱ معنادار است.

بحث و نتیجه گیری: نتایج بیان کننده این‌است که در داده های واقعی مدل رگرسیون‌لجستیک توان پیش بین کمتری نسبت به مدل های شبکه ای داراست و در هر مجموعه از داده ها، الگوریتم پیش بین مناسب داده را باید کاوش‌کرد. الگوریتم های پیش بین، به خوبی توان پیش بینی افت تحصیلی را در هر سه رشته مورد بررسی در دانشکده پیراپزشکی را داراست و می توان از آن، در امر آموزش دانشجویان پزشکی و پیراپزشکی به منظور شناسایی دانشجویان در معرض خطر افت تحصیلی جهت انجام اقدامات مربوط به پیشگیری، استفاده کرد.

واژه های کلیدی: عملکرد تحصیلی، افت تحصیلی، داده کاوی، داده کاوی آموزشی

* نویسنده مسئول: گروه روش‌ها و برنامه های آموزشی و درسی، دانشکده علوم تربیتی و روان شناسی، دانشگاه تهران. تهران، ایران.

Email: keyvansalehi@ut.ac.ir

Copyright © 2019 Journal of Ilam University of Medical Science. This is an open-access article distributed under the terms of the Creative Commons Attribution international 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material, in any medium or format, provided the original work is properly cited.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

مقدمه

ثبت نام در آموزش عالی در سراسر جهان به میزان قابل توجهی افزایش یافته است(۱). و مراکز علمی و دانشگاهی به جهت این که مسئولیت پذیرش دانشجو و تربیت نیروی متخصص در رشته های مختلف را بر عهده دارند، از عوامل مهم رشد یک کشور محسوب می شوند(۲) و نظام آموزشی در دانشگاه های علوم پزشکی، در ارتقای کیفیت آموزش و تربیت نیروی انسانی پاسخ گو به نیازهای جامعه تلاش می کند(۳). زیرا نظام سلامت امروز به افراد متخصص توانمند و کارآمد جهت پاسخ گویی به نیازهای کمی و کیفی کشور نیاز دارد(۴). در آموزش علوم پزشکی آموزش و برنامه های آموزشی باید به گونه ای باشد که افرادی را تربیت نماید که به گسترش روزافزون دامنه این علم توجه داشته، و هم چنین از دانش، تجربه و مهارت کافی برخوردار باشند(۵). به همین جهت، موظف به ارائه حجم وسیعی از دانش، نگرش و مهارت‌هاست.

عملکرد تحصیلی دانشجویان به معنای میزان دستیابی دانشجو به اهداف آموزشی است(۶). افت تحصیلی هنگامی اتفاق می افتد که فراگیری که از نظر نمره ضعیف است، به انتظار خود نرسد و معیارهای دانشگاهی را برآورده نکند(۷). با ملاک های مختلفی نظیر مشروطی(اخذ معدل کمتر از ۱۲)، تکرار درس(ناشی از عدم کسب نمره قبولی در مرتبه اول)، طولانی شدن مدت تحصیل، اخراج، انصراف، ترک تحصیل و ...قابل بررسی است(۸). افت تحصیلی نه تنها برای خود فراگیر و والدین ناامید کننده است، بلکه تاثیرات آن از نظر کمبود نیروی انسانی در همه زمینه های اقتصاد و سیاست برای جامعه به همان اندازه بسیار مهم است(۹).

دانشجویان خصوصاً دانشجویان رشته های مربوط به علوم پزشکی و پیراپزشکی، به واسطه رشته تحصیلی و حساسیت کاری در دوران تحصیلات دانشگاهی با خطرات و آسیب های فراوانی، از جمله استرس های شدید، افسردگی، اضطراب، تغییر در سبک زندگی، مشکلات شغلی رو به رو هستند(۱۰) و عدم کنترل افت تحصیلی در این دانشجویان، موجب پایین آمدن سطح علمی و کارایی ارائه کنندگان خدمات سلامت در سال های آینده خواهد شد(۱۱) و به علت حساسیت شغلی و ارتباط آن با سلامت افراد جامعه، این مشکل به صورت بحرانی تری خود را نشان می دهد(۱۲) و در تحقیقات متعددی عوامل مرتبط با افت تحصیلی در رشته های علوم پزشکی و پیراپزشکی بررسی شده است(۱۳).

جلوگیری از افت تحصیلی یک چالش جدی است(۱۴). برای دستیابی به بهترین نتایج و کاهش میزان افت، نظام های آموزشی به روش های خلاقانه ای برای بهبود کیفیت آموزش نیاز دارند(۱۵). محققان برای درک بهتر عملکرد دانشگاهی و سبک های یادگیری و هم چنین موضوع های مختلفی که مستقیماً با آن ها در ارتباط هستند، روش های داده کاوی آموزشی را تهیه کرده اند(۱۶).

داده کاوی آموزشی را می توان به عنوان کاربرد فنون داده کاوی سنتی در تجزیه و تحلیل داده های آموزشی با هدف حل مشکلات در زمینه آموزشی تعریف کرد(۱۷) زمانی که موسسات از داده کاوی برای تشخیص دانشجویان تحت ریسک استفاده می کنند، می توانند از شکست و حذف آن ها جلوگیری کنند، قبل از این که حتی دانشجویان خودشان از این که در معرض خطر افت تحصیلی هستند مطلع باشند(۱۸). در سال های اخیر، از فنون مختلف داده کاوی و الگوریتم های طبقه بندی استفاده شده است مانند بیز ساده، درخت تصمیم گیری، شبکه های عصبی، تشخیص دورتر و فنون آماری پیشرفته. این فنون برای به دست آوردن اطلاعات، کمک به سیستم های پشتیبانی تصمیم گیری و الگوی استخراج از داده های دانشجویی استفاده می شود(۱۹). نکته مهم در بررسی فنون و مدل های گوناگون پیش بینی، مقایسه میزان دقت آن ها، میزان تطبیق پذیری آن ها با داده های آینده و نیز صحت آن ها در مجموعه های مختلف است که در طول پروژه مورد توجه قرار می گیرد و در واقع اساس مقایسه و انتخاب بهترین مدل ها است(۲۰).

رگرسیون لجستیک، فرمی از رگرسیون می باشد و زمانی مورد استفاده قرار می گیرد که متغیر وابسته به

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

صورت دو یا چندسطحی بوده و متغیرهای مستقل از هر نوع دیگر باشند(۲۱). درخت تصمیم شامل مجموعه ای از گره های تصمیم است که توسط شاخه هایی به یکدیگر متصل شده اند. این گره های تصمیم از گره ریشه آغاز و تا رسیدن به گره های برگ بسط یافته اند(۲۲). در این تحقیق از الگوریتم درخت تصمیم CHAID استفاده شده است. شبکه های عصبی با پردازش داده های تجربی دانش یا قانون نهفته در ورای داده ها را به ساختار شبکه منتقل می کنند و فنونی هستند که می توانند برای مدل های با ارتباطات پیچیده مورد استفاده قرار گیرند(۲۳). شبکه های بیزی یک چارچوب مناسب و شهودی برای مشخص کردن توزیع های پیچیده با احتمال مشترک هستند(۲۴) و می توانند تعامل علی و روابط آماری متغیرهای یک سیستم را با یک نمایش گرافیکی نشان دهد، به طوری که مدل رسم شده به سادگی قابل تفسیر است(۲۵) و ماشین بردار پشتیبان فن نسبتاً جدیدی در حوزه داده کاوی است و در بسیاری از مسائل طبقه بندی به طور موفق استفاده شده است(۲۶).

در عرصه داده کاوی آموزشی در زمینه افت تحصیلی تحقیقات متعددی انجام شده است. که می توان مطالعه اتکاء بر شاخص های پذیرشی در پیش بینی موفقیت تحصیلی(۲۷)؛ مرور مقالات منتشرشده در زمینه داده کاوی آموزشی،(۱۷)؛ پیش بینی نمرات کلاسی دانشجویان با سه روش شبکه عصبی و درختان تصمیم و بیزی(۲۸) و استفاده از شبکه های بیزی برای مدل سازی رابطه علی و عوامل عملکرد دانشجویی(۲۵) را نام برد. در برخی از تحقیقات از رگرسیون لجستیک در پیش بینی افت تحصیلی دانشجویان پزشکی و پیراپزشکی استفاده شده است(۲۹) و در تحقیق دیگری روش خوشه بندی به کار برده شده است(۳۰). محققین پژوهشی که از فنون داده کاوی آموزشی در پیش بینی عملکرد تحصیلی(افت-پیشرفت) در میان دانشجویان رشته های علوم پزشکی و پیراپزشکی استفاده کرده باشد نیافتند. تحقیقات در زمینه افت تحصیلی دانشجویان رشته های علوم پزشکی و پیراپزشکی عموماً عوامل مرتبط با افت تحصیلی را مورد شناسایی قرار داده است(۳۱).

هدف از انجام این تحقیق، مقایسه میزان صحت فنون مختلف داده کاوی و مشخص ساختن بهترین مدل پیش بینی است و سوال اصلی تحقیق، این است که از میان مدل های داده کاوی رگرسیون لجستیک، درخت تصمیم، شبکه عصبی، شبکه بیزین و ماشین بردار پشتیبان، کدام مدل در پیش بینی افت تحصیلی دانشجویان رشته های علوم آزمایشگاهی، مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی مناسب تر است؟

مواد و روش ها

جامعه پژوهش، کلیه دانشجویان دوره کارشناسی دانشکده پیراپزشکی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران در سال تحصیلی ۹۸-۱۳۹۷ هستند و با استفاده از روش نمونه گیری در دسترس، ابتدا از میان رشته های دوره کارشناسی دانشکده پیراپزشکی، رشته های علوم آزمایشگاهی، مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی انتخاب شدند. آن گاه ۵۰۰ نفر از دانشجویان این رشته ها در تحقیق شرکت کردند، به این صورت که با مراجعه به کلاس های دانشجویان از آنان خواسته شد تا به سوالات پرسش نامه پاسخ دهند. بعد از اجرا و کنار گذاشتن پرسش نامه های مخدوش، ۱۵۳ پرسش نامه از رشته علوم آزمایشگاهی، ۸۵ پرسش نامه مدیریت اطلاعات سلامت، و ۷۳ پرسش نامه رشته تکنولوژی علوم پرتوی مورد تحلیل قرارگرفت. برای گردآوری داده ها از ابزار سنجش افت تحصیلی دانشجویان استفاده شده است. سوالات این ابزار با توجه به نتایج تحقیق کیفی، نوشته شده است که با استفاده از روش پدیدارشناسی، ضمن تحلیل ادراک و تجربه زیسته دانشجویانی که با افت تحصیلی مواجه شده بودند، ۱۳۳ نشانگر و ۱۶ ملاک مرتبط با افت تحصیلی دانشجویان شناخته شد. یافته های بخش کیفی به دلیل گستردگی نشانگرها، مضمون ها و زیر مضمون ها در مقاله مجزایی در حال انتشار است. از آن جا که نیاز بود تا ابزاری متناسب با موقعیت محیطی کشور ساخته شود،

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

با توجه به نشانگرهای یافت شده، ۱۴۱ گویه تدوین گردید هم چنین، ویژگی های روان سنجی ابزار مورد نظر شامل همسانی درونی و روایی محتوا و روایی سازه با استفاده از ضریب آلفای کرونباخ(۰/۹۷۱) و نظر متخصصان(استادان روش شناسی و محتوایی)، و روایی سازه آن با تحلیل عامل اکتشافی و تاییدی مورد بررسی و تایید قرار گرفت(۳۲). در این ابزار دانشجویان ابتدا به پرسش هایی در مورد معدل دوره دبیرستان، معدل دوره کارشناسی تا زمان تکمیل پرسش نامه، تعداد واحد افتاده و تعداد ترم های مشروطی، متغیر هدف«افت تحصیلی» پاسخ دادند و در ادامه به سوال هایی چهار گزینه ای با طیف لیکرت(خیلی زیاد، زیاد، کم و خیلی کم) و در برگیرنده عواملی شامل مسائل ناشی از انتخاب رشته، مسائل فردی، استاد، قوانین و امکانات دانشگاه، مشکلات اجتماعی-اقتصادی، سبک یادگیری و مطالعه، معاشرت نامناسب گروه همسالان، تفاوت فرهنگی و ناپختگی، دوره دبیرستان، فرهنگ خانواده، فقدان شادی و نشاط، شیوه گزینش، مشکلات اقتصادی و کار به هنگام تحصیل، محیط زندگی، برنامه ریزی درسی نامناسب و مشکلات خانوادگی پاسخ دادند. تحلیل داده ها با استفاده از نرم افزار های SPSS vol.16، SMART PLS 3 و SPSS MODELLER 18 انجام گرفت. به طوری که، بررسی معناداری میزان تفاوت افت تحصیلی رشته های مختلف و هم چنین بررسی معناداری میزان صحت مدل ها در هر رشته با استفاده از آزمون آماری خی دو و با نرم افزار SPSS vol.16 انجام شد و جهت انجام تحلیل عامل تاییدی و محاسبه نمرات عاملی از نرم افزار SMART PLS 3 استفاده شد. هم چنین جهت به دست آوردن مدل پیش بین، از الگوریتم های شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی، رگرسیون لجستیک، الگوریتم درخت تصمیم و از نرم افزار SPSS MODELLER 18(کلمنتاین) استفاده شد.

یافته های پژوهش

با توجه به تعداد بالای سوالات پرسش نامه و عوامل شناخته شده در بخش تحلیل عامل تاییدی، ابتدا با استفاده از نرم افزار SMART PLS 3 برای هر عامل، نمرات عاملی محاسبه شد و این عوامل به عنوان داده های ورودی در مدل های پیش بینی مورد استفاده قرار گرفتند. برای ساخت الگوریتم های پیش بین از نرم افزار SPSS MODELLER 18(کلمنتاین) استفاده شد. کلمنتاین یکی از بهترین نرم افزارها برای پیاده سازی الگوریتم های داده کاوی است. اعتبارسنجی الگوریتم های یادگیری، بر این اساس است که در ابتدا داده ها به دو بخش تقسیم می شود و بخش اول به عنوان داده های آموزشی و بخش دوم به عنوان داده های آزمایشی مورد استفاده قرار گیرند. هدف از انجام این کار این است که اعتبار نتایج حاصل از کاربرد داده های آموزشی با داده های آزمایشی مورد بررسی قرار گیرد. در نرم افزار کلمنتاین تقسیم داده ها به صورت خودکار و تصادفی انجام می شود و ۷۰ درصد داده ها در بخش آموزشی و ۳۰ درصد در بخش آزمایشی استفاده می شود تا اعتبار مدل مورد بررسی قرار گیرد.

جدول شماره ۱ نشان دهنده میزان مشروطی در رشته های مختلف است. همین طور که جدول شماره ۱ مشخص است از میان دانشجویان مشارکت کننده در تحقیق، میزان افت تحصیلی در رشته علوم آزمایشگاهی ۲۸/۷۵ درصد در رشته مدیریت اطلاعات سلامت، ۳۱/۷۶ درصد و در رشته تکنولوژی علوم پرتوی ۴۵/۲ درصد است.

در جدول شماره ۲، نتایج آزمون خی دو جهت مقایسه میزان معناداری تفاوت موجود در رشته ها بررسی شده است. جهت انجام این امر از نرم افزار SPSS vol.16 استفاده شد. همان گونه که جدول شماره ۲ نشان می دهد، مقدار خی دو برابر با ۶/۱۵۴ با درجه آزادی ۲ و در سطح ۰/۰۴۶ دو دامنه معنادار است. شکل شماره ۱، مسیر ساخته شده با استفاده از نرم افزار کلمنتاین، جهت اجرای الگوریتم داده کاوی، در داده های رشته های علوم آزمایشگاهی، مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی است. همان گونه که در شکل شماره ۱ مشخص است، الگوریتم درخت تصمیم CHAID در داده های رشته های مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی موفق به پیش بینی نبوده است.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

جدول شماره ۳، نشان دهنده میزان صحت مدل در الگوریتم های مورد استفاده در رشته علوم آزمایشگاهی است: نتایج جدول شماره ۳ نشان می دهد، در رشته علوم آزمایشگاهی، تمامی الگوریتم ها با میزان صحت بالاتر از ۷۱ درصد می توانند وقوع افت تحصیلی را پیش بینی کنند. هم چنین الگوریتم های درخت تصمیم و شبکه بیز با میزان ۸۲/۲۴ درصد بیشترین صحت پیش بینی افت تحصیلی را دارند و ماشین بردار پشتیبان با میزان پیش بینی ۷۶/۹۷ درصد دومین رتبه را در پیش بینی دارا بوده و بعد از آن شبکه عصبی با میزان صحت پیش بینی ۷۳/۶۸ درصد است و رگرسیون لجستیک کمترین صحت پیش بینی را داراست.

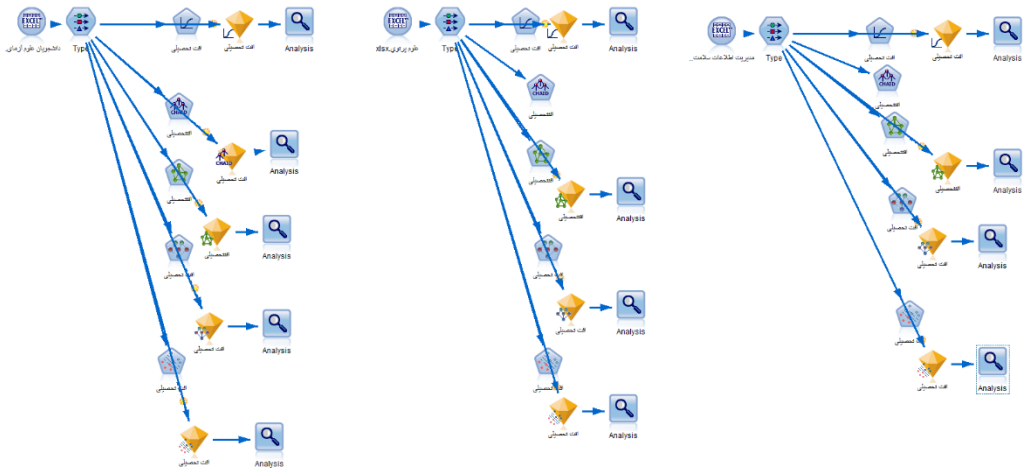
در رشته مدیریت اطلاعات سلامت، میزان صحت پیش بینی الگوریتم ها به ترتیب شبکه بیز با میزان ۹۷/۲۶ درصد بهترین الگوریتم پیشبین، ماشین بردار پشتیبان با میزان ۸۴/۷۱ درصد در جایگاه دوم، شبکه عصبی با اختلاف کم و به مقدار ۸۳/۵۳ درصد در جایگاه بعدی و کمترین میزان صحت الگوریتم در رگرسیون لوجستیک به مقدار ۷۷/۶۵ درصد است.

و در رشته تکنولوژی علوم پرتوی شبکه بیز با میزان ۹۸/۸۲ درصد بهترین الگوریتم پیشبین، ماشین بردار پشتیبان با میزان ۷۱/۲۳ درصد در جایگاه دوم، شبکه عصبی و رگرسیون لجستیک هر دو به مقدار ۶۴/۳۸ درصد در جایگاه آخر هستند.

جدول شماره ۴، در رشته علوم آزمایشگاهی به بررسی معناداری تفاوت میزان صحت الگوریتم های متفاوت با استفاده از آزمون خی دو، پرداخته است. همان گونه که نتایج جدول شماره ۴ نشان می دهد تفاوت میان الگوریتم های متفاوت در پیش بینی و تبیین میزان افت تحصیلی در سطح ۰/۰۵ معنی دار نیست.

جدول شماره ۵، در رشته مدیریت اطلاعات سلامت به بررسی معناداری تفاوت میزان صحت الگوریتم های متفاوت با استفاده از آزمون خی دو، پرداخته است. همان گونه که نتایج جدول شماره ۵ نشان می دهد تفاوت میان الگوریتم های متفاوت در پیش بینی و تبیین میزان افت تحصیلی در سطح ۰/۰۱ معنی دار است و می توان گفت الگوریتم شبکه بیز بهترین مدل برای پیش بینی افت تحصیلی است.

جدول شماره ۶، در رشته تکنولوژی علوم پرتوی به بررسی معناداری تفاوت میزان صحت الگوریتم های متفاوت با استفاده از آزمون خی دو، پرداخته است. همان گونه که نتایج جدول شماره ۶ نشان می دهد تفاوت میان الگوریتم های متفاوت در پیش بینی و تبیین میزان افت تحصیلی در سطح ۰/۰۱ معنی دار است و می توان گفت شبکه بیز بهترین مدل پیش بین افت تحصیلی است.



شکل شماره ۱. مسیر ساخته شده با استفاده از نرم افزار کلمنتاین، جهت اجرای الگوریتم داده کاوی

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

جدول شماره ۱. میزان مشروطی در رشته های مختلف

رشته تحصیلی	تعداد اعضا نمونه	تعداد دانشجویان دچار افت تحصیلی	درصد میزان افت تحصیلی
علوم آزمایشگاهی	۱۵۳	۴۴	۲۸/۷۵
مدیریت اطلاعات سلامت	۸۵	۲۷	۳۱/۷۶
تکنولوژی علوم پرتوی	۷۳	۳۳	۴۵/۲

جدول شماره ۲. آزمون خی دو، بررسی معناداری تفاوت میزان افت تحصیلی در رشته ها

-	ارزش	درجه آزادی	معنی داری دو دامنه
خی دو	۶/۱۵۴	۲	۰/۰۴۶
نسبت درست نمایی	۵/۹۸۹	۲	۰/۰۵۰
ارتباط خطی با خطی	۵/۴۱۰	۱	۰/۰۲۰

جدول شماره ۳. نتایج صحت الگوریتم های پیش بین

نوع الگوریتم	میزان صحت پیش بینی رشته علوم آزمایشگاهی	میزان صحت پیش بینی مدیریت اطلاعات سلامت	میزان صحت پیش بینی تکنولوژی علوم پرتوی
رگرسیون لوجستیک	٪ ۷۱/۷۱	٪ ۷۷/۶۵	٪ ۶۴/۳۸
درخت تصمیم CHAID	٪ ۸۲/۲۴	-	-
شبکه عصبی	٪ ۷۳/۶۸	٪ ۸۳/۵۳	٪ ۶۴/۳۸
شبکه بیز	٪ ۸۲/۲۴	٪ ۹۸/۸۲	٪ ۹۷/۲۶
ماشین بردار پشتیبان	٪ ۷۶/۹۷	٪ ۸۴/۷۱	٪ ۷۱/۲۳

جدول شماره ۴. معناداری آزمون خی دو، تفاوت میزان صحت الگوریتم های متفاوت در رشته علوم آزمایشگاهی

-	ارزش	درجه آزادی	معنی داری دو دامنه
خی دو	۸/۰۸۶	۴	۰/۰۸۸
نسبت درست نمایی	۸/۱۳۸	۴	۰/۰۸۷
ارتباط خطی با خطی	۰/۹۶۱	۱	۰/۳۲۷

جدول شماره ۵. معناداری آزمون خی دو، تفاوت میزان صحت الگوریتم های متفاوت در رشته مدیریت اطلاعات سلامت

-	ارزش	درجه آزادی	معنی داری دو دامنه
خی دو	۱۷/۲۵۸	۳	۰/۰۰۱
نسبت درست نمایی	۲۳/۲۰۸	۳	۰/۰۰۰۱
ارتباط خطی با خطی	۴/۷۳۱	۱	۰/۰۳۰

جدول شماره ۶. معناداری آزمون خی دو، تفاوت میزان صحت الگوریتم های متفاوت در رشته تکنولوژی علوم پرتوی

-	ارزش	درجه آزادی	معنی داری دو دامنه
خی دو	۲۸/۰۴۹	۳	۰/۰۰۰۱
نسبت درست نمایی	۳۶/۶۳۹	۳	۰/۰۰۰۱
ارتباط خطی با خطی	۵/۴۳۹	۱	۰/۰۲۰

بحث و نتیجه گیری

یافته های این تحقیق نشان داد که میزان افت تحصیلی در رشته های مختلف متفاوت است. در رشته علوم آزمایشگاهی، میزان صحت مدل پیش بین داده ها در الگوریتم های متفاوت به ترتیب بهترین مدل پیش بینی عبارت است از: الگوریتم های درخت تصمیم و شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی و رگرسیون لجستیک. در رشته مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی میزان صحت مدل پیش بین داده ها در الگوریتم های متفاوت به ترتیب بهترین مدل پیش بینی عبارت است از: الگوریتم شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

و رگرسیون لجستیک. در این بخش، الگوریتم درخت تصمیم جوابی نداشت.

نتایج تحقیق نشان داد که میزان افت تحصیلی در دانشجویان دوره کارشناسی رشته های علوم آزمایشگاهی، مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی با هم متفاوت است. البته با توجه به این که روش نمونه گیری در این تحقیق از نوع تصادفی نبوده است که از کاستی ها و محدودیت های این تحقیق است، هم چنین تعداد زیادی از نمونه انتخاب شده به طور کامل به سوالات پاسخ نداده و از تحقیق حذف شدند، این نکته باعث می شود با وجود معنادار بودن نتیجه آزمون معناداری، تعمیم این نتیجه به کل دانشجویان محتاطانه باشد. در تحقیقات متفاوتی تفاوت میزان افت تحصیلی در دانشجویان رشته های مختلف پزشکی و پیراپزشکی مورد بررسی قرار گرفته است. در این تحقیقات میزان افت تحصیلی در دانشجویان علوم آزمایشگاهی بیشتر از سایر رشته ها بوده است که با نتایج تحقیق حاضر هم خوانی ندارد(۳۳،۳۴).

در رشته علوم آزمایشگاهی، تمامی الگوریتم های مورد بررسی پیش بینی قابل قبولی از وقوع افت تحصیلی دارند. بین میزان صحت پیش بینی افت تحصیلی در الگوریتم های مختلف تفاوت وجود دارد و شبکه بیزی و ماشین بردار پشتیبان بهترین مدل های پیشبین در این داده ها هستند و مدل رگرسیون لجستیک کمترین میزان صحت پیش بین را داراست و تفاوت موجود از لحاظ آماری نیز معنادار نیست.

در رشته مدیریت اطلاعات سلامت و تکنولوژی علوم پرتوی الگوریتم درخت تصمیم CHAD جوابی در بر نداشت و میزان صحت سایر الگوریتم ها به ترتیب بیشترین میزان صحت الگوریتم عبارت است از: الگوریتم شبکه بیز، ماشین بردار پشتیبان، شبکه عصبی و رگرسیون لجستیک. هم چنین آزمون خی دو نشان داد تفاوت موجود بین الگوریتم های به دست آمده از لحاظ آماری معنادار است. مرور مقالات منتشر شده در زمینه داده کاوی آموزشی توسط بیکر و یاسف(۲۰۰۹) بیانگر این مطلب بود که پیش بینی، پرکاربردترین روش حوزه داده کاوی است(۱۷). همراستا با نتایج این تحقیق، حامدی و دیرین(۲۰۱۸) و میلان و همکاران (۲۰۱۰) از الگوریتم شبکه بیزی در حوزه داده کاوی آموزشی استفاده کرده اند(۲۴،۲۵). از نظر چاوار (۲۰۱۱)، استفاده از فنون پیش بینی مانند رگرسیون خطی و چندگانه یا شبکه عصبی می تواند در فرآیند ارزیابی دانشجویی و بر مجموعه داده های آنان اعمال شود و الگوهای موفقیت دانشجویان استخراج گردد و در نتیجه دانشگاه ها می توانند عملکرد دانشجویان را پیش بینی کنند(۳۵). اما نتایج این پژوهش هم راستا با نتایج پژوهش کاباکچیوا(۲۰۱۳) میزان صحت الگوریتم های متفاوت بر داده های یکسان می تواند متفاوت باشد(۱۹). وجود نتایج متناقض در میان تحقیقات بررسی شده می تواند به این علت باشد که در هر تحقیق، پژوهشگران بنا بر نظر خود متغیرهای پیش بین متفاوتی انتخاب کردند. در نتیجه تفاوت موجود در نتایج حاصل از تحقیقات، نه تنها به علت استفاده از مدل های پیش بین متفاوت بلکه به علت انتخاب متغیرهای پیش بین متفاوت نیز یکسان نیست. با این وجود اهمیت پیش بینی وقوع افت تحصیلی به حدی است که هر نوع مطالعه ای در این زمینه می تواند مثمرثمر تلقی شود.

بنا بر نتایج این تحقیق شبکه بیزی بهترین مدل های پیشبین در این داده ها هستند و مدل رگرسیون لجستیک کمترین میزان صحت پیش بین را داراست. بر اساس نتایج مطالعه کاباکچیوا(۲۰۱۳) مدل شبکه بیز کمترین دقت را در پیش بینی افت تحصیلی داشت(۱۹). با وجود اهمیت افت تحصیلی دانشجویان رشته های مربوط به علوم پزشکی و پیراپزشکی به علت نقش آینده ساز این دانشجویان در بخش سلامت آینده کشور، تحقیقات تنها در زمینه عوامل مربوط به افت تحصیلی انجام شده است(۱۱،۱۲) و تنها از رگرسیون لجستیک و خوشه بندی جهت پیش بینی افت تحصیلی استفاده شده است(۲۹) همان گونه که نتایج این تحقیق نشان داد روش های داده کاوی که از الگوریتم یادگیری، جهت پیش بینی استفاده می کنند در پیش بینی موفق تر عمل می کنند. این نتایج گواه بر این است که در ایران، داده کاوی آموزشی در آموزش دانشجویان رشته های علوم پزشکی و پیراپزشکی مورد استقبال قرار نگرفته است. حال آن که

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

رشته های علوم پزشکی و پیراپزشکی با روش های داده کاوی نا آشنا نیستند و از روش های متفاوت داده کاوی در عرصه سلامت و پیش بینی وقوع بیماری در تحقیقات متعدد استفاده می کنند(۳۶) و تنها در بخش آموزش است که داده کاوی آموزشی استفاده نشده است.

امروزه شناسایی علل افت تحصیلی و راه های پیشگیری از آن، از مهم ترین دغدغه های مسئولین و دست اندکاران نظام آموزشی است. جلوگیری از افت تحصیلی از نظر فردی و اجتماعی امری ضروری است. تحلیل و الگوبرداری از عملکرد دانشجویان به موسسات آموزشی کمک می کند تا کیفیت ارائه دوره خود را بهبود بخشند. هم چنین بستری را فراهم می نماید تا دانشجویان در مطالعات خود بهتر عمل کرده و از شایستگی ها و صلاحیت های بهتری برخوردار شوند(۲۵). این در حالی است که همان گونه که شاناک و همکاران(۲۰۱۰) بیان کرده اند(۲۸)، داده کاوی ابزار تحلیل قدرتمندی است که به موسسات آموزشی این امکان را می دهد تا منابع مالی و کارکنان را به بهترین وجه به امر آموزش تخصیص دهند و پیامدهای ناگوار افت تحصیلی دانشجویان را کنترل کنند. داده کاوی آموزشی با کشف الگوهای پنهان در پایگاه های داده های بزرگ به محققان این امکان را می دهد تا بتوانند مدل هایی بسازند که با دقت بالایی عملکرد تحصیلی دانشجویان را پیش بینی کنند. همیشه در هر امری که نتایج ناگواری در پی دارد پیشنهاد می گردد از وقوع آن پیشگیری شود و ناگواری وقوع افت تحصیلی و هزینه های هنگفت آن برای شخص دانشجو، خانواده و خصوصاً اجتماع(به علت پدید آمدن فاجعه ای با عنوان پرورش متخصص ناکارآمد) بر کسی پوشیده نیست، از این رو با توجه به موفقیت روش های داده کاوی آموزشی، خصوصا الگوریتم شبکه بیز در پیش بینی بروز افت تحصیلی، استفاده از آن در سیستم های آموزش عالی می تواند ثمربخش باشد. عدم همکاری و مشارکت دانشجویان پزشکی برای شرکت در این پژوهش و عدم تکمیل پرسش نامه توسط آن ها منجر به محدود شدن جامعه مورد هدف مطالعه گردید که متعاقب آن نتایج قابل تعمیم به‌این گروه از دانشجویان نیست لذا این مورد جزء محمدودیت های مطالعه حاضر است. در این راستا محققان پیشنهاد می کنند:

۱-نتایج مطالعه حاضر محدود به نمونه دردسترس از دانشجویان دانشکده پیراپزشکی است و به علت عدم همکاری تمامی دانشجویان رشته های مورد بررسی در این مطالعه، محققین نتوانستند از روش نمونه گیری تصادفی جهت تعمیم نتایج استفاده کنند، بنا بر این پیشنهاد می شود تا در تحقیقات بعدی تفاوت میزان افت تحصیلی رشته های مختلف با استفاده از نمونه گیری تصادفی بررسی شود.

۲-با توجه به این که یکی از عوامل مربوط به افت تحصیلی، رشته تحصیلی است و مدل حاصل از الگوریتم شبکه بیز بهترین صحت پیش بینی‌کننده افت تحصیلی بود، پیشنهاد می شود میزان صحت هر الگوریتم در رشته های مختلف مورد بررسی قرار گیرد.

۳-با توجه به میزان بالای حساسیت وقوع افت تحصیلی در رشته های علوم پزشکی و پیراپزشکی و اهمیت پیش بینی وقوع آن در این جوامع، پیشنهاد می شود از داده کاوی آموزشی در رشته های علوم پزشکی و پیراپزشکی دانشگاه های دیگر نیز استفاده شود.

کد اخلاق: IR.UT.PSYEDU.REC.1398.024

References

1.Vanklaveren C, Kooiman K, Cornelisz, I, Meeter M. [The higher education enrollment decision: feedback on expected study success and updating behavior]. J Res Edu Eff 2019; 12: 1-23. doi. 10.1080/19345747.2018.1496501

2.Vahabi A, Khatrei A, Vahabi B, Mirzaee M, Ahmadian M. [The perspective of public health students of Kurdistan University of medical sciences regarding field of study and future career and its related factors in 2014]. J Nurs Mid Pramed 2016; 1: 33-45. (Persian) doi. 10.29252/sjnmp.1.3.5

3.Emamghorashi F, Heydari ST, Najafipour S.[Evaluation of effecting factors on

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

educational status of medical students in Jahrom Medical University during 1994-2003]. JBUMS 2010; 12: 40-45. (Persian)

4.Atashnafas M, Atashnafas E. [GIS and Irans holistic scientific map on health domain]. HBIJ 2011; 8: 305-313. (Persian)

5.Ehsanpour S. [Achieving minimum learning requirements from the viewpoints of midwifery students in Isfahan school of nursing and midwifery]. Iranian J Med Edu2006; 6:17-25. (Persian)

6.Banik P, Kumar B. [Impact of information literacy skill on students academic performance in Bangladesh]. Int J European Stud 2019; 3: 27-33. doi. 10.11648/j.ijes.20190301.15

7.Widyastuti T, Kurniawan A, Chandra NP. [Coping strategies on students after experiencing academic failure an indigenous study in javanese context]. Work Pap Ser 2017; 3:22-6.

8.Motlagh M, Elhampour H, Shakurnia A. [Factors affecting students' academic failure in Ahvaz Jundishapur University of medical sciences in 2005]. Iran J Med Edu 2008; 8: 91-9.

9.Aremu A, Sokan B. [A multi causal evaluation of academic performance of Nigerian learners issues and implications for national development]. Dep Guid Coun Uni Ibadan 2003;2:124-9.

10.Shams B, Farshidfar M, Hassanzadeh A. [Effect of counseling on the achievement of university students with dropout]. Iranian J Med Edu 2000; 1: 36-41.

11.Jabraili M, Mousavivaezi SJ, Mikaeli P, Saadatyanov R, Aqlmad S. [Evaluating the influencing factors of educational underachievement in gifted and talented students of Urmia University of medical sciences]. Horiz Med Edu Dev 2011; 4: 21-5.

12.Moniri R, Ghalebtarash H, Mussavi G. [The reasons of educational failure among paramedical students in Kashan University of Medical Sciences]. Iranian J Med Edu 2006; 1: 13540. (Persian)

13.Lancia L, Petrucci C, Giorgi F, Dante A, Cifone MG. [Academic success or failure in nursing students, results of a retrospective observational study]. Nurs Edu Today 2013; 33: 1501-5. doi:10.1016/j.nedt.2013.05.001

14.Rafoth MA. Academic failure prevention of encyclopedia of applied psychology. 1th ed. Spielberger CD Acad Publication. 2004; P. 1-7

15.Ashraf A, Anwer S, Khan MK. [A comparative study of predicting student performance by use of data mining techniques]. ASRJETS 2018; 44: 122-36.

16.Baker RSJD. [Data mining for education]. Int Enc Edu 2010; 7: 112-18. doi.org/10.1016/B978-0-08-044894-7.01318-X

17.Baker RS, Yacef K. [The state of educational data mining in 2009: A review and future visions]. J Edu Data Min 2009; 1: 3-17. doi.org/10.5281/zenodo.3554657

18.Delavari N. Beikzadeh MR. [A new analysis model for data mining processes in higher educational systems]. Int Sym Inf Com Tech Malaysia 2004;2:23-8.

19.Kabakchieva D. [Predicting student performance by using data mining methods for classification]. Cyb Inf Tech 2013; 13: 61-72. doi. 10.2478/cait-2013-0006

20.Luan J. Data Mining and knowledge management in higher education potential applications. Pres AIR Canada2002; 4:218-23.doi.org/10.1002/ir.35

21.Antonogeorgos G, Panagiotakos DB, Priftis KN, Tzonou A. [Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10- to 12-years-old children: divergence and similarity of the two statistical methods]. Int J Pediatr2009; 3:952042. doi. 10.1155/2009/952042

22.Larose DT. Discovering knowledge in data an introduction to data mining. John Wiley Son Inc Publication 2005;P.142-7. doi. 10.1002/0471687545

23.Collins GS, Mallett S, Omar O. [Developing risk prediction models for type 2 diabetes a systematic review of methodology and reporting]. BMC Med2011; 5:146-59. doi.org/10.1186/1741-7015-9-103

24.Millan E, Loboda T, Perez JL. [Bayesian networks for student model engineering]. Comp Edu 2010; 55:1663-83. doi. 10.1016/j.compedu.2010.07.010

25.Hamedi A, Dirin A. [A bayesian approach in students' performance analysis]. 10th Int Con Edu New Learn Technol 2018; Spain. doi. 10.21125/edulearn.2018.2498

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

26.Bellotti T. Crook J. [Support vector machines for credit scoring and discovery of significant features]. Exp Sys Appl 2009; 36: 3302-30. doi. 10.1016/j.eswa.2008.01.005

27.Willcockson IU, Johnson CW, Hersh W, Bernstam EV. Predictors of student success in graduate biomedical informatics training introductory course and program success. J Am Med Inf Ass 2009;16: 837-46. doi. 10.1197/jamia.M2895

28.Shannaq B, Rafael Y, Alexandro V.[Student relationship in higher education using data mining techniques]. Glob J Comp Sci Technol2010; 10: 54-9.

29.Eskandari M, rostami A, Hojati A, Kolifarhood G. [Trends of academic failure and its demographic determinants among medical students in Zanjan University of medical sciences]. Iranian J Med Edu 2014; 14:174-83.

30.Asif R, Merceron A, Ali S, Haider N. [Analyzing undergraduate student performance using educational data mining]. Comp Edu 2017; 113: 177-94. doi.org/10.1016/j.compedu.2017.05.007

31.Motlagh, M., Elhampour, H. Shakurnia A. [Factors affecting student academic failure in Ahvaz Jundishapur University of medical sciences in 2005]. Iran J Med Edu 2008; 8: 91-9.

32.Fadavi roodsari A, Salehi K, Khodaii E, Mogadamzadeh A, Javadipoor M. [Factors associated with bachelor student academic failure in University of Tehran]. Iran J Med Edu2016;2:72-7.

33.Tagharrobi Z, Fakharian E, Mirhoseini F, Rasooli Nejad SA, Akbari H, Ameli H. [Survey of educational drop-out indexes and realate factor in alumni of paramedical faculty of Kashan University of Medical Sciences]. J Babol Uni Med Sci 2010; 12 :76-89.

34.Edalatkhah H, Jahangiri S, Khanbabazadeh M, Amani F, Hashimilir M. [The educational situation of graduates in Ardebil University of Medical Sciences]. IJME 2005; 5: 193-5.

35.Chaware A. [Educational data mining an emerging trends in education]. Int J Adv Res Comp Sci 2011; 2: 117-22.

36.Sudhakar K, Manimekalai DM. [Study of heart disease prediction using data mining]. J Adv Res Comp Sci Soft Eng 2014; 4: 1157-60.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]

Application of Data Mining Techniques in Determining the Accuracy of the Models Predicting the Student Academic failure

Fadaviroodsari A[1], Salehi K[1], Khodaii E[1], Alimogadamzadeh A[1], Javadipoor M[1]*

(Received: September 28, 2019 Accepted: December 8, 2019)

Abstract

Introduction: Student academic failure is one of the shortcomings of higher education system. A system with the ability to predict and prevent the academic failure can play a decisive role in enhancing the quality of education. The utilization of educational data mining is one of the most effective strategies in monitoring and improving the quality of education. This study aimed to compare different data mining techniques in order to determine the best prediction model.

Materials & Methods: The study population consists of all undergraduate students studying Medical Laboratory Sciences, Health Information Technology Management, and Radiation Sciences Technology in Tehran University of Medical Sciences, Tehran, Iran in the academic year of 2018-2019. In total, 500 were selected to participate in this study. After administering and excluding the invalid questionnaires, 153, 85, and 73 questionnaires from the students of Medical Laboratory Sciences, Health Information Technology Management, and Radiation Sciences Technology were analyzed, respectively. The data were collected using students' academic failure questionnaire with a reliability Cronbach's alpha coefficient of 0.971 and confirmed content validity by experts. Furthermore, the data were analyzed in SPSS (version 16), SmartPLS 3, and SPSS-Modeler (Clementine) (version 18) software through logistic regression, decision tree, neural network, Bayesian network, and the support vector machine algorithms. *Ethics code:* IR.UT.PSYEDU.REC.1398.024

Findings: The results showed a significant difference among the majors in terms of the level of academic failure (P=0.046). Regarding the Medical Laboratory Sciences, the descending order of the accuracy of the data prediction model using each algorithm includes decision tree and Bayesian network, support vector machine, neural network, and logistic regression. In addition, considering the Health Information Technology Management and Radiation Sciences Technology, the best prediction models were Bayesian Network, Support Vector Machine, Neural Network, and logistic regression in a descending order. It is worth mentioning that the decision tree algorithm had no answer. Furthermore, the Chi-square test showed a statistically significant difference among the students of Medical Laboratory Sciences (0.88), Health Information Technology Management (0.001), and Radiation Science Technology (0.000) regarding the obtained algorithms.

Discussion & Conclusions: The results indicated that the logistic regression model had less predictive power in real data, compared to the network models. Therefore, the appropriate data prediction algorithm must be explored in any dataset. Predictive algorithms are well capable to predict academic failure in all three majors at the departments of allied medical sciences. Moreover, it can be used for medical and allied medical science students to identify students at the risk of failing to take preventive measures.

Keywords: Academic performance, Academic failure, Data mining, Educational data mining

1. Dept of Curriculum Planning, Faculty of Education and Psychology, University of Tehran, Tehran, Iran
**Corresponding author Email: keyvansalehi@ut.ac.ir*

[Downloaded from sjimu.medilam.ac.ir on 2026-11-11]

[DOI: 10.29252/sjimu.28.1.36]