

مدلی برای پیش‌بینی آسیب‌پذیری تحصیلی مقطع کارشناسی مبتنی بر شبکه عصبی A Neural Network Based Model for Predicting Educational Vulnerability of Undergraduate Students

دریافت مقاله: ۱۳۹۴/۱۱/۲۴؛ دریافت نسخه نهایی: ۱۳۹۵/۶/۱؛ پذیرش مقاله: ۱۳۹۵/۶/۲۱

A. Aminbeidokhti (Ph.D.), M. Fathian
Boroojeni (Ph.D.) & A. Nameni (MA)

Abstract: The aim of this study as a part of a doctoral thesis was to develop a model for predicting educational vulnerability of undergraduate students in engineering disciplines in short term period (by semester). The method was data mining by using neural network algorithm. The statistical population, including all "Term-student" from the first semester in academic year 1390-91 till the second semester of 1393-94 in three Iranian technical-engineering universities (with a total of 53,422 records). The needed data were used in the model by direct exploitation of MISs in all three universities. The results indicate that by using the available data in educational systems of universities and engaging the neural networks algorithm, it is possible to make a prediction by more than 95 percent accuracy and with validity over 60, in terms of semester results for all students. "GPA (Grade Point Average) of the last semester", "total GPA", "the condition of the semesters in the case of being an odd or an even one", "type of units taken within the semester" and "engaging in extra activities", were identified as the most effective predictive variables.

Key words: educational vulnerability, neural networks, prediction.

علی اکبر امین بیدختی^۱، محمد فتحیان بروجنی^۲ و
احمد نامنی^۳

چکیده: هدف پژوهش حاضر به‌عنوان بخشی از رساله دکتری، توسعه مدلی برای پیش‌بینی آسیب‌پذیری تحصیلی دانشجویان مقطع کارشناسی رشته‌های مهندسی در کوتاه مدت (نیمسال تحصیلی) بوده که به روش داده‌کاوی و با استفاده از الگوریتم شبکه عصبی به اجرا درآمده است. جامعه آماری، دربرگیرنده کلیه "ترم- دانشجویان" از نیمسال اول سال تحصیلی ۹۱-۱۳۹۰ لغایت دوم ۹۴-۱۳۹۳ در سه مورد از دانشگاه‌های فنی-مهندسی کشور (مجموعاً ۵۳۴۲۲ رکورد) بوده که با بهره‌برداری مستقیم از سیستم‌های آموزش هر سه دانشگاه در مدل‌سازی وارد شده‌اند. نتایج حاکی از آن است که با بهره‌گیری از داده‌های موجود در سیستم‌های حوزه آموزش دانشگاه‌ها و به خدمت گرفتن شبکه عصبی می‌توان با صحت بیش از ۹۵ درصد و اعتبار بالاتر از ۶۰، نسبت به پیش‌بینی وضعیت نیمسال برای یکایک دانشجویان اقدام نمود. معدل نیمسال قبل، معدل کل، زوج یا فرد بودن نیمسال، نوع واحدهای اخذ شده و مبادرت به فعالیت‌های فوق-برنامه در نیمسال، به‌عنوان مؤثرترین متغیرهای پیش‌بین، تشخیص داده شدند.

کلید واژه‌ها: آسیب‌پذیری تحصیلی، پیش‌بینی، شبکه عصبی.

۱. استاد گروه مدیریت آموزشی دانشگاه سمنان.

۲. استاد دانشگاه علم و صنعت ایران.

۳. نویسنده مسئول: دانش آموخته دکتری مدیریت آموزشی دانشگاه سمنان namenia@iust.ac.ir

مقدمه

موضوع عملکرد تحصیلی^۱، یکی از زمینه‌های پژوهشی اصلی^۲ (مادر) در حوزه مدیریت آموزشی به‌شمار می‌رود. در این حوزه گسترده، طبقه‌بندی^۳ و خوشه‌بندی^۴ فراگیران، جزو موضوعات فرعی بسیار مطرح در سالیان اخیر به‌شمار می‌آیند. در هر یک از این زیر-حوزه‌های فرعی، یکی از رایج‌ترین پیش‌فرض‌های نظری که بسیاری از پژوهشگران (به‌عنوان مثال احمدوند، ۱۳۸۰)، به آن قائل هستند، طبقه‌بندی سه ناحیه‌ای است که طبق آن، اعضای یک جامعه آماری از فراگیران، معمولاً به سه طبقه یا خوشه شامل: قشر متوسط (اکثریت)^۵، قشر ضعیف^۶ و قشر ممتاز^۸ تقسیم یا فرض می‌شوند. گرچه در پژوهش‌های از نوع داده کاوی (به‌عنوان مثال: مطالعه اویلید^۹، اولادیپوپو^{۱۰} و اوباگیوآ^{۱۱}، ۲۰۱۰)، طبقه‌بندی ۵ طیفی نیز مرسوم است که در آن، دو طیف بسیار قوی و بسیار ضعیف به سه طیف قبلی اضافه می‌شوند.

در این فضای پژوهشی، به مدت نزدیک به یک قرن، توجه خاصی به دانشجویان ممتاز و نخبه مطرح بوده است. با این وجود، شواهد قابل توجهی وجود دارد که نشان می‌دهد از دو دهه پیش، توجه ویژه‌ای به طیف فراگیران ضعیف معطوف گردیده است.

در شکل ۱ نمودار حاوی مقایسه تعداد تحقیقات مرتبط با ۴ کلیدواژه بسیار شناخته شده در زمینه طیف‌بندی فراگیران، آورده شده است. هر سه جستجو در پایگاه انتشارات علمی ScinceDirect و با محدود ساختن دامنه جستجو به سال ۱۹۹۲ به بعد، انجام یافته است. جستجو در سایر پایگاه‌های علمی حاوی مقالات مدیریت آموزشی نیز نتایج مشابهی را نشان می‌دهد. پایین‌ترین نمودار، تعداد مقاله‌های احصاء شده براساس جستجوی عبارت Gifted Students را نشان می‌دهد. که در نظام آموزشی آمریکا به معنای دانش‌آموز/ دانشجوی ممتاز بوده و می‌تواند شاخصی از میزان توجه پژوهشگران به طیف ممتاز باشد. نمودار دوم، تعداد مقالات احصاء شده در همین سایت براساس جستجوی عبارت Educational Equity را نشان می‌دهد. این نمودار می‌تواند شاخصی باشد از میزان توجه پژوهشگران به بحث عدالت آموزشی که با آفت و خیز نه چندان چشمگیری همواره به‌عنوان یکی از موضوعات پژوهشی در حوزه مدیریت آموزشی مطرح بوده است. نمودار سوم از پایین، نتایج جستجو براساس عبارت honor students را نشان می‌دهد که در نظام‌های اروپایی به معنای دانش‌آموز/ دانشجوی ممتاز می‌باشد. اما بالاترین

1. Academic performance
3. Classifying
5. Sub- field
7. Poor spectrum
9. Oyelade
11. Obagbuwa

2. Main- field
4. Clustering
6. The middle majority
8. Honor spectrum
10. Oladipupo

نمودار، بیانگر تعداد مقالات احصاء شده براساس عبارت *poor students* را نشان می‌دهد. این واژه در ادبیات کلان مدیریت آموزشی، به معنای فراگیران مناطق فقیر و محروم به کار می‌رود و در سطح خرد (به‌عنوان مثال در پژوهش شلسمن و کلمن ۲۰۱۱)، به معنای دانشجویان ضعیف کاربرد دارد که هم ارز *weak student* می‌باشد. همان‌طور که در شکل، قابل ملاحظه است نمودار اخیر، به طرز چشمگیری بالاتر از سایر نمودارهای رقیب، قرار گرفته است.

اما علاوه بر پشتوانه‌های آماری، شواهد نظری مختلفی نیز برای نشان دادن توجه اندیشمندان و منتقدین جهانی به طیف فراگیران ضعیف قابل ارائه است. که از جمله آن‌ها می‌توان به توصیه‌های تند منتقدین آمریکائی (به‌عنوان مثال: کراتی^۱، ۲۰۱۱) مبنی بر ضرورت توجه دولتمردان این کشور به طیف فراگیران ضعیف اشاره نمود. در مجموع، استنباط می‌شود که در دو دهه اخیر، احیای سرمایه‌های انسانی در ناحیه فراگیران ضعیف، نزد سیاست‌گذاران نظام‌های آموزشی و به موازات آن، نزد پژوهشگران حوزه مدیریت آموزشی، جایگاه خاصی یافته است.

در کشور ما نیز مطالعات متعددی در خصوص بررسی عوامل و متغیرهای مؤثر در ضعف یا افت تحصیلی در میان دانش‌آموزان و دانشجویان به انجام رسیده که از جمله آن‌ها می‌توان به پژوهش مقیمی و همکاران (۱۳۸۹) اشاره نمود. در پژوهش یاد شده محرز می‌گردد که عوامل آموزشی، تأثیر معناداری در افت تحصیلی دانشجویان پسر در مقطع راهنمایی دارند.

در مقابل، پژوهش صباغیان و پورکاظمی (۱۳۸۷) به‌عنوان نمونه‌ای از پژوهش‌های داخل که بر موفقیت تحصیلی تمرکز داشته‌اند قابل ذکر است. نامبردگان در تحقیق خود نشان می‌دهند که موفقیت تحصیلی دانشجویانی که رشته کارشناسی آن‌ها با کارشناسی ارشد آن‌ها مرتبط بوده است موفق‌تر از کسانی هستند که رشته‌های تحصیلی دو مقطع آن‌ها نامرتب است.

در هر حال، در این فضای پژوهشی، بخش عمده‌ای از پژوهش‌ها از نوع پیش‌بینی و به خصوص داده‌کاوی می‌باشند. این‌طور فرض می‌شود که اگر بتوان پیش‌بینی دقیق و به موقعی از فراگیری که در معرض خطر ضعف آموزشی قرار دارند به عمل آورد می‌توان با مراقبت‌ها و حمایت‌های به موقع از افت تحصیلی آنان پیشگیری نمود.

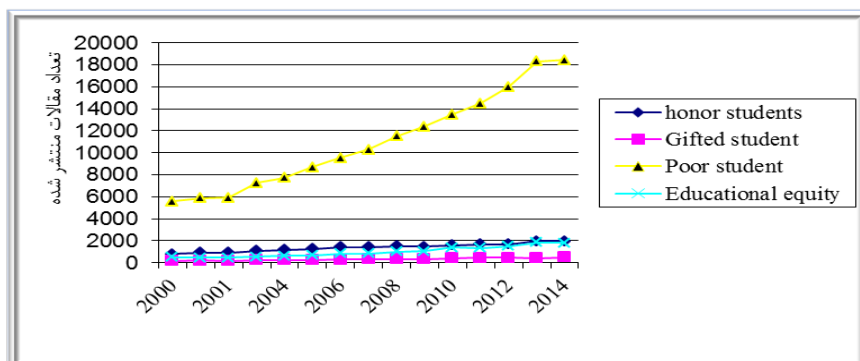
بیکر^۲ و یاسف^۳ (۲۰۰۹)، در پژوهشی مروری که بر روی مقالات منتشر شده جهانی طی سال‌های ۲۰۰۸ و ۲۰۰۹ در زمینه داده‌کاوی آموزشی^۴ به عمل آورده‌اند، پیش‌بینی را به‌عنوان پرکاربردترین حوزه‌های داده‌کاوی شناسایی کرده‌اند. شکل (۲) در این رابطه، عیناً از مقاله ایشان درج شده است.

1. Dorothyjean Cratty
3. Yacef

2. Baker
4. Educational data mining (EDM)

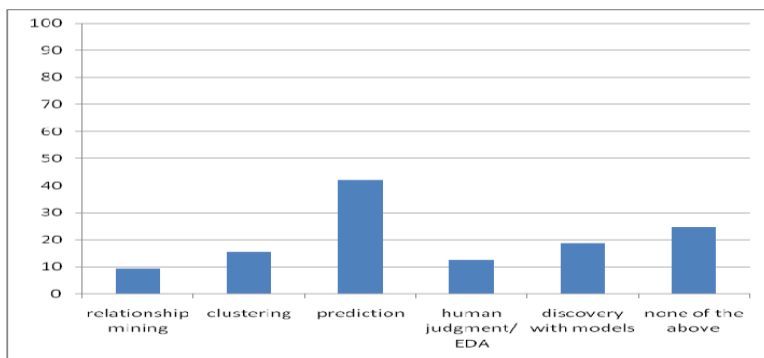
هم‌چنین پینا آیالا^۱ (۲۰۱۴) در پژوهش مروری جامعی که در حوزه داده‌کاوی آموزشی انجام داده است، مطالعات از نوع پیش‌بینی را به‌عنوان رایج‌ترین مطالعات این حوزه، شناسایی می‌نماید. شکل (۳) عیناً از مقاله نامبرده اقتباس گردیده است. در این شکل، نسبت تعداد پژوهش‌های پیش‌بینی در مقایسه با نوع توصیفی طی سه سال ۲۰۱۰، ۲۰۱۱ و ۲۰۱۲، به تصویر کشیده شده است.

این درحالیست که پژوهش‌های از نوع پیش‌بینی در داخل کشور هنوز عمدتاً از طریق شیوه‌های کلاسیک آماری از قبیل رگرسیون به انجام می‌رسند. به‌عبارت دیگر به لحاظ فنون و الگوریتم‌های پیش‌بینی یک شکاف حدوداً دو دهه‌ای بین پژوهش‌های داخل و خارج از کشور دیده می‌شود. خوشبختانه توسعه سیستم‌های اطلاعات مدیریت در حوزه‌های آموزش دانشگاه‌های کشور که منبع ارزشمندی برای ثبت و نگهداری داده‌ها محسوب می‌شوند و رایج شدن شیوه‌های نوین داده‌کاوی در حوزه‌های مهندسی و مدیریت، این نوید را می‌دهد که این شکاف فنی در حوزه‌ی مدیریت آموزشی نیز به زودی بر طرف شود.

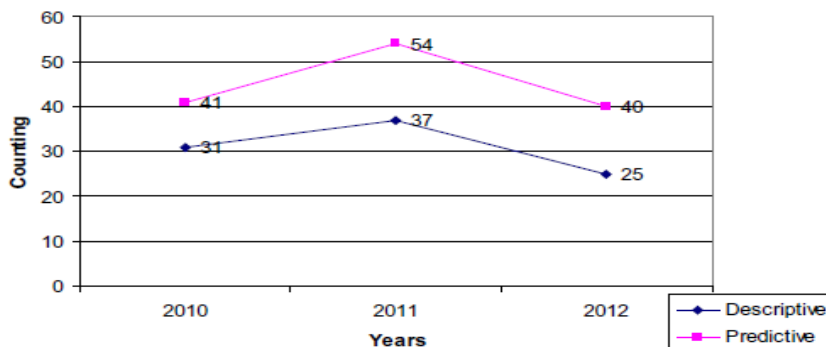


شکل ۱. مقایسه تعداد مقالات احصاء شده از طریق پایگاه Science Direct در مورد ۴ کلیدواژه منتخب در حوزه موضوعی پژوهش.

اما صرف‌نظر از فنون پیش‌بینی، بحث متغیرهای مورد استفاده در پیش‌بینی از اهمیت ویژه‌ای برخوردار است. مرور پژوهش‌های پیشین به شناسایی و گزینش متغیرهای کارآمد کمک می‌نماید. در مرور پژوهش‌های پیشین، مجموعاً ۱۱۴ پژوهش مرتبط داخلی و خارجی شناسایی گردیده که ذکر تمامی آن‌ها در این مقاله امکان‌پذیر نمی‌باشد. لذا خلاصه‌ای از آن‌ها به مقتضای بحث، مورد اشاره قرار خواهند گرفت.



شکل ۲. مقایسه حوزه‌های مختلف داده‌کاوی آموزشی طی سال‌های ۲۰۰۸ و ۲۰۰۹ براساس تعداد مقالات منتشر شده بین‌المللی، به نقل از بیکر و یاسف (۲۰۰۹)



شکل ۳. مقایسه تعداد پژوهش‌های جهانی داده‌کاوی آموزشی در دو زمینه توصیفی و پیش‌بینی (به نقل از پینا آیالا، ۲۰۱۴)

پژوهش ویلکاکسون^۱، جانسون^۲، هرش^۳، المر^۴ و برنستام (۲۰۰۹)، به‌عنوان نمونه‌ای از مطالعاتی که به‌طور کلی بر پیش‌بینی موفقیت تحصیلی با اتکا بر شاخص‌های پذیرشی، تمرکز داشته، قابل ذکر است. در حالی که شل‌سل‌من^۵ و کل‌من^۶ (۲۰۱۱) با تمرکز بر ضعف تحصیلی در دانشجویان رشته داروسازی دانشگاه کانکتیکوت^۷ تلاش می‌کنند با آمیزه‌ای از متغیرهای جمعیت‌شناختی^۸ و پذیرشی^۹، مهم‌ترین پیش‌بینی‌کننده‌های ضعف تحصیلی در این رشته را شناسایی نمایند. شیوه پژوهش ایشان، تجربی با کنترل‌های آماری بوده است. در این مطالعه موردی، عدم اتمام به موقع تحصیل و قرار گرفتن دانشجو به لحاظ معدل کل در یک سوم پایین جامعه آماری به‌عنوان نشانگرهای ضعف تحصیلی انتخاب گردیده بوده‌اند. کوآتا^{۱۰} (۲۰۱۲)، در تحقیقی کیفی از طریق پیمایش (پرسشنامه)، به بررسی علل افت تحصیلی دانشجویان رشته رادیوگرافی تشخیصی در دانشگاه فنی اوربان^{۱۱} می‌پردازد. پژوهش وی نشان می‌دهد که از دید دانشجویان ضعیف سال آخر، موارد زیر به‌عنوان عمده‌ترین دلایل ضعف تحصیلی به‌شمار می‌روند: مشکل در فهم مطالب درسی، فقدان آمادگی کافی (برای شروع دوره/ درس)، عدم مطالعه مستقل (توسط دانشجو بدون استاد یا همقطاران) و ضعف در فهم سؤالات آزمون‌ها. این پژوهشگر در نتیجه‌گیری خود، تغییر شیوه آموزش این دانشگاه به شیوه ارزیابی‌محور^{۱۲} را به‌عنوان یک استراتژی آموزشی برای پیشگیری از ضعف و افت تحصیلی دانشجویان این رشته توصیه می‌نماید.

نانیس^{۱۳}، هادسون^{۱۴}، فیل هورز^{۱۵} و تنگ^{۱۶} (۲۰۰۵) و روزاندر^{۱۷}، بک استروم^{۱۸} و استنبرگ^{۱۹}، (۲۰۱۱)، عوامل متنوع دخیل در عملکرد تحصیلی دانشجویان را به دو دسته کلی تقسیم می‌نمایند: ۱- عوامل مربوط به دانشجو ۲- عوامل مرتبط با نظام آموزشی. ایشان عوامل مربوط به دانشجو را به‌نوبه خود به دو بخش، قابل تفکیک می‌دانند. یک دسته عوامل تأثیرگذار در طول دوران تحصیل که شامل مواردی از قبیل: انگیزه، پشتکار، تیپ شخصیتی، هوش و استعداد تحصیلی می‌باشد. و دوم عواملی که عقبه^{۲۰} تحصیلی دانشجو را نشان می‌دهند و در برگیرنده مواردی هم‌چون: رشته تحصیلی مقاطع قبلی، نوع و سطح آموزشگاه‌های مقاطع قبلی، نمرات

1. Willcockson

3. Hersh

5. Schlesselman

7. Connecticut

9. Admission criteria

11. Durban

13. Nonis

15. Philhours

17. Rosander

19. Stenberg

2. Johnson

4. Elmer

6. Coleman

8. Demographics

10. Gqweta

12. Assessment oriented

14. Hudson

16. Teng

18. Bäckström

20. Background

دانشجو در مقاطع قبل و نمرات کسب شده او در آزمون‌های ورودی و مصاحبه‌ها می‌باشند. ایشان تأکید دارند که کامل‌ترین و مؤثرترین راهکارهای پیش‌بینی و ارزیابی عملکرد تحصیلی، آن‌هایی هستند که به هر دو دسته از عوامل برشمرده شده مرتبط با دانشجو و نیز عوامل مرتبط با نظام آموزشی توجه دارند. با این حال یادآوری می‌نمایند که محدودیت‌های زمانی و هزینه‌ای، اغلب مانع از اتخاذ چنین رویکردی می‌شود و لذا دانشگاه‌ها در مواجهه با محدودیت‌ها، به جای اتخاذ رویکرد کامل‌گرا، ترجیح می‌دهند از ساز و کارهایی بهره ببرند که با حداقل متغیرها به بیشترین قدرت ارزیابی و پیش‌بینی برسند.

نصیر^۱، رشید^۲، احمد^۳ و نورشاه^۴ (۲۰۱۲)، به بحث خوشه‌بندی فراگیران و ارزیابی تکنیک‌های مرسوم در این حوزه پرداخته‌اند. از آن‌جا که این پژوهش در کشور مالزی (به‌عنوان کشوری در حال توسعه) به انجام رسیده و از این نظر که در ارزیابی تکنیک‌های داده‌کاوی و خوشه‌بندی، طیف دانشجویان ضعیف را مورد استفاده قرار داده، قابل توجه است. اما در داخل کشور، مطالعات زارع شاه‌آبادی (۱۳۷۱) و بهمن‌پور (۱۳۷۳) به‌عنوان نمونه‌های نسبتاً قدیمی و پژوهش‌های زارعی (۱۳۹۲)، بخشی (۱۱۳۹۲) و موسوی (۱۳۹۲) به‌عنوان نمونه‌هایی از پژوهش‌های نسبتاً جدید که از زوایای مختلف بر پیش‌بینی ضعف تحصیلی تمرکز داشته‌اند، قابل اشاره‌اند. هدف پژوهش بهمن‌پور، بررسی عوامل گوناگونی بوده است که در انصراف دانشجویان از ادامه تحصیل، مؤثر و تعیین‌کننده بوده است. روش وی، پیمایشی با استفاده از پرسشنامه بوده است. جامعه آماری، کلیه دانشجویان منصرف از تحصیل در سال‌های ۱۳۶۳ تا ۱۳۷۰ دانشگاه علامه طباطبایی بوده‌اند. یافته‌های وی نشان می‌دهد که یکی از مسائل پیچیده‌ی نظام آموزش عالی در ایران مسئله افت تحصیلی است که به انحاء مختلف از جمله اخراج، مشروطی، ترک تحصیل یا انصراف از تحصیل، هر ساله چهره خود را می‌نمایاند و بخش قابل ملاحظه‌ای از منابع انسانی و اقتصادی را به هدر می‌دهد. هم‌چنین طبق نتایج این پژوهش، عامل اشتغال به کار، مسئولیت خانواده و عدم علاقه به رشته تحصیلی در انصراف از تحصیل مؤثر بوده‌اند.

غفاری، کارشکی و رضائی (۱۳۸۹) در قالب پایان‌نامه کارشناسی ارشد به شیوه‌ای پیمایشی مبتنی بر نمونه‌گیری و پرسشنامه، در دانشگاه فردوسی مشهد با رویکردی مقایسه‌ای به بررسی علل افت تحصیلی دانشجویان ایرانی و افغانی در دانشگاه یاد شده و دانشگاه هرات می‌پردازند. نتایج پژوهش ایشان نشان می‌دهد که در رابطه با میزان افت تحصیلی در میان سه گروه آماری،

تفاوت معنا داری وجود دارد و دانشجویان افغانی دانشگاه فردوسی از دو گروه دیگر، آفت تحصیلی بیشتری داشته‌اند. لکن به بررسی علل آن نمی‌پردازد.

یقینی و حیدری (۱۳۸۹)، در پژوهشی مشابه با مطالعه نصیر و همکاران، تلاش می‌کنند مناسب‌ترین تکنیک‌های داده‌کاوی برای پیش‌بینی وضعیت تحصیلی دانشجویان داخل کشور را از طریق مقایسه قدرت پیش‌بینی روش‌های مختلف شناسائی و معرفی کنند. ایشان در پژوهش خود به این نتیجه می‌رسند که در خوشه‌بندی وضعیت تحصیلی، تکنیک نزدیک‌ترین همسایه و در پیش‌بینی وضعیت تحصیلی، شبکه عصبی کارآمدتر از سایر تکنیک‌ها می‌باشند. ایشان در این تحقیق از آمیزه‌ای از متغیرهای تحصیلی و جمعیت شناختی بهره می‌برند. محدود بودن تعداد متغیرهای ایشان و عدم تمایز بین متغیرهای تأثیرگذار بر دانشجویانی هر خوشه از عواملی است که تعمیم‌پذیری نتایج ایشان را به شدت محدود می‌سازد.

رحمتی، لسانی و خلیل‌زاده (۱۳۹۱)، در پژوهشی در قالب پایان‌نامه کارشناسی ارشد به شیوه‌ای پیمایشی (پرسشنامه) و با بهره‌گیری از تکنیک تحلیل عاملی اکتشافی به کشف و تبیین برخی از عوامل و فاکتورهای مؤثر بر مشروط شدن دانشجویان دانشگاه شهید باهنر کرمان می‌پردازند. نتایج پژوهش ایشان حاکی از آن است که در دانشگاه شهید باهنر، تعداد ترم‌های مشروطی دانشجویان با عوامل؛ عدم حضور دانشجویان در کلاس، تنگناهای فرهنگی، اجتماعی و نامناسب بودن خدمات دانشگاه، درگیر شدن دانشجویان با مسائل حاشیه‌ای در دانشگاه (گروه‌های سیاسی یا انجمن‌ها)، عدم آشنایی دانشجویان با رشته تحصیلی و مشکلات مربوط به آن، برقراری رابطه عاطفی دانشجویان با جنس مخالف و مشکلات عاطفی با او، نامناسب بودن خدمات مشاوره‌ای دانشگاه، عدم برنامه‌ریزی تحصیلی درست و بی‌برنامگی دانشجویان، مشکلات محتوایی دروس و مطالعه ناکارآمد، محیط فیزیکی نامناسب کلاس‌های درس دانشگاه، بی‌انگیزگی تحصیلی دانشجویان، اضطراب و عدم اعتماد به نفس دانشجویان، محل اخذ مدرک پیش دانشگاهی، محل سکونت خانواده، میزان تحصیلات مادر، تاریخ اخذ مدرک دیپلم یا پیش دانشگاهی، معدل کل ترم و سال ورود به دانشگاه دانشجویان رابطه دارد.

در مجموع پژوهش‌های مرور شده، عناوین متغیرهای وابسته (هدف) در حوزه موضوعی پژوهش، بسیار محدود و رایج‌ترین آنها عبارت‌اند از: عملکرد تحصیلی، پیشرفت تحصیلی، آفت تحصیلی و ضعف تحصیلی. همچنین نشانگرهای مورد استفاده در این پژوهش‌ها برای اندازه‌گیری متغیر وابسته نیز بسیار محدودند که رایج‌ترین آن‌ها عبارت‌اند از: معدل کل، تعداد ترم‌های مشروطی، اتمام یا عدم اتمام به موقع تحصیل. اما تعداد متغیرهای مستقل (پیش‌بین) در این حوزه بسیار متنوعند.

حد و حدود جوامع آماری در اکثر قریب به اتفاق پژوهش‌های مرور شده داخل در حد یک یا چند رشته تحصیلی و یا چند دانشکده از یک یا چند دانشگاه و بر مبنای نمونه‌گیری بوده است که تعداد آن‌ها غالباً از ۵۰۰ نفر تجاوز نمی‌کند. این محدودیت احتمالاً بدین علت است که در اکثر پژوهش‌ها، داده‌های لازم از طریق پرسشنامه تهیه می‌شوند که این امر، انتخاب و به کارگیری نمونه‌های بزرگ‌تر از ۵۰۰ نفر را برای پژوهشگر، غیرممکن یا بسیار دشوار می‌سازد. در واقع پژوهش‌هایی هم‌چون: حیدری (۱۳۸۹) در داخل کشور با جامعه ۳۸۷۲ نفر که از طریق داده کاوی به انجام رسیده است بسیار معدودند. این درحالیست که پژوهش‌هایی از نوع داده کاوی در حوزه موضوعی پژوهش درخارج از کشور، از سال‌ها پیش رایج بوده است به‌عنوان مثال، وینترز^۱ (۲۰۰۶) با جامعه آماری ۱۱۰۰۰ نفر.

و بالاخره اینکه تعداد پژوهش‌هایی که جوامع آماری آن‌ها رشته‌های فنی و مهندسی باشند در مقایسه با پژوهش‌هایی که جامعه آماری آن‌ها از مجموعه رشته‌های علوم پزشکی یا علوم انسانی باشند به نسبت پایین است (سه مورد در بین کل پژوهش‌های مرور شده).

به‌عنوان جمع‌بندی یافته‌های حاصل از مرور پژوهش‌های پیشین می‌توان: "کمبود پژوهش‌های داده‌کاوی در حوزه مدیریت آموزشی"، "عدم بهره‌گیری از داده‌های ارزشمند موجود در سیستم‌های اطلاعات آموزش دانشگاه‌ها"، "ضرورت توجه خاص به دانشجویان ضعیف" و "کمبود مطالعات انجام شده در رشته‌های مهندسی" را به‌عنوان شکاف تحقیقاتی و ضرورتی دانست که پژوهش حاضر قصد پوشش دادن به آن را داشته است.

سؤالات پژوهش

۱. متغیرها، داده‌ها و اطلاعات موجود در سیستم‌های اطلاعاتی دانشگاه‌های مورد مطالعه تا چه حد در پیش‌بینی مشروطی دانشجو در یک نیمسال تحصیلی قابل استفاده‌اند؟
۲. اهمیت نسبی هر یک از این متغیرها در هر یک از دانشگاه‌های مورد مطالعه چگونه است؟
۳. صحت و کارایی مدل‌های حاصله چقدر است؟

متغیرهای پژوهش

متغیر وابسته (هدف) در این پژوهش عبارت بوده از "وضعیت نهایی نیمسال تحصیلی" که دارای دو حالت "مشروط" و "عادی" بوده است. ملاک تعیین وضعیت مشروط مطابق با آیین‌نامه‌های آموزشی دانشگاه‌ها کسب معدل کمتر از ۱۲ در نیمسال بوده است. متغیرهای مستقل (پیش‌بین) به شرح جدول شماره (۱) بوده‌اند. در این مجموعه از به کارگرفتن متغیرهای جمعیت‌شناختی و پذیرشی خودداری گردیده است. زیرا وضعیت این متغیرها برای هر دانشجو در طول دوران

تحصیل، معمولاً ثابت است. در عوض تلاش گردیده از متغیرهایی بهره برده شود که وضعیت آن‌ها برای هر دانشجو در ترم‌های مختلف متفاوت است. هم‌چنین موجود بودن داده‌های مربوط به متغیر در سیستم اطلاعات آموزش دانشگاه، شرط اساسی برای انتخاب متغیر پیش‌بین بوده است.

جدول ۱. متغیرهای پیش‌بین وارد شده به مدل

نوع متغیر	عنوان/ شرح	مقیاس	وضعیت های قابل اطلاق
عملکردی	معدل کل	نسبتی	پیوسته (بین ۰ تا ۲۰)
	معدل نیمسال قبل	نسبتی	پیوسته (بین ۰ تا ۲۰)
	مجموع واحدهای گذرانده	نسبتی	پیوسته (بزرگتر، مساوی ۰)
	تعداد نیمسال‌های مشروطی قبلی	گسسته	بزرگتر مساوی صفر
	تعداد نیمسال‌های مشروطی متوالی	گسسته	بزرگتر مساوی صفر
ترمی (متغیرهای اختصاصی نیمسال مورد نظر)	مجموع تعداد واحدهای اخذ شده	گسسته	۱ تا ۲۴
	تعداد واحدهای اخذ شده عمومی	گسسته	۰ تا ۲۴
	تعداد واحدهای اخذ شده پایه	گسسته	۰ تا ۲۴
	تعداد واحدهای اخذ شده اصلی	گسسته	۰ تا ۲۴
	تعداد واحدهای اخذ شده تخصصی	گسسته	۰ تا ۲۴
	تعداد واحدهای اخذ شده اختیاری	گسسته	۰ تا ۲۴
	اخذ یا عدم اخذ پروژه	دو دویی	دارد / ندارد
	فعالیت فوق برنامه در نیمسال	دو دویی	دارد / ندارد
	زوج یا فرد بودن نیمسال	دو دویی	زوج / فرد

روش‌شناسی پژوهش

روش پژوهش حاضر، داده‌کاوی^۱ بوده که با بهره‌گیری از الگوریتم شبکه‌های عصبی^۲ در حوزه داده‌کاوی آموزشی درآمده است. با وجود مشابهت در کلیات داده‌کاوی نزد مراجع مختلف، تعاریف زیر، بیشترین تأثیر را بر مراحل اجرائی این پژوهش داشته‌اند:

- فایاد^۳، پایتتسکی شاپریو^۴ و اسمیت^۵ (۱۹۹۶): داده‌کاوی عبارت است از فرآیند تشخیص الگوهای معتبر^۶، بدیع^۷، بالقوه مفید^۸ و قابل فهم^۹ از داده‌ها.
- جی و همکاران (۲۰۰۹)، ترجمه علی خانزاده (۱۳۹۲): داده‌کاوی به فرآیند جستجو و کشف مدل‌های گوناگون، مختصرسازی‌ها و اخذ مقادیر از مجموعه‌های از داده‌های معلوم اطلاق می‌گردد.

- شهرابی (۱۳۹۲): داده‌کاوی، فرآیند کشف دانش پنهان درون داده‌های انبوه می‌باشد. تقاطع زمینه‌ای حوزه‌های داده‌کاوی و مدیریت آموزشی، قلمرو "داده‌کاوی آموزشی" را شکل می‌دهد که درصدد کسب دانش از داده‌های آموزشی و خدمات آموزشی است. تعریف مورد استناد پژوهش حاضر، از داده‌کاوی آموزشی، تعریف رسمی انجمن جهانی داده‌کاوی آموزشی^{۱۰} به قرار زیر می‌باشد:

داده‌کاوی آموزشی، یک رشته^{۱۱} در حال ظهور^{۱۲} است که توجه اصلی آن: ۱- توسعه روش‌های کاوش در داده‌های انحصاری و در حال رشد مجموع‌های آموزشی ۲- استفاده از آن روش‌ها برای شناخت بهتر فراگیران و مجموعه‌ای که در آن یاد می‌گیرند، می‌باشد.

نرم‌افزار پژوهش

نرم‌افزار مورد استفاده در این پژوهش، نرم افزار SPSS Modeler نسخه ۱۲، محصول شرکت IBM می‌باشد که به نام تجاری (12) Clementine مشهور است.

الگوریتم پژوهش (شبکه عصبی)

مدل‌سازی در پژوهش حاضر بر مبنای "الگوریتم شبکه‌های عصبی مصنوعی" که در ادبیات داده‌کاوی به اختصار، "شبکه عصبی" نامیده می‌شود انجام گردیده است. این الگوریتم یکی از قوی‌ترین ابزارهای پیش‌بینی و در عین حال، مبهم‌ترین الگوریتم به لحاظ بیان فرآیند محاسباتی که

1. Data mining

3. Fayyad

5. Smyth

7. Novel

9. Understandable

10. International educational data mining society

11. Discipline

2. Neural networks

4. Piatetsky-Shapiro

6. Valid

8. Potentially useful

12. Emerging

انجام می‌دهد می‌باشد. به عبارت دیگر با وجود آن که اندازه عددی یا طبقه متغیر هدف را با دقت و صحت بسیار بالایی پیش‌بینی می‌کند در خصوص چگونگی پیش‌بینی، اطلاعات چندانی نمی‌دهد. تا جایی که در برخی از مراجع، از این الگوریتم به عنوان جعبه سیاه^۱ یاد کرده‌اند. به همین دلیل در پژوهش‌هایی که هدف آن‌ها بیشتر، نفس پیش‌بینی می‌باشد (همانند پژوهش حاضر) و یا این که اصل پیش‌بینی نسبت به چرایی و چگونگی تأثیرگذاری متغیرهای پیش‌بین بر یکدیگر و بر متغیر هدف، اولویت داشته باشد این الگوریتم در مقایسه با الگوریتم‌های رقیب از جمله: رگرسیون کلاسیک، رگرسیون لجستیک، درخت تصمیم، تئوری بیزین و غیره ارجحیت داده می‌شود.

به‌طور کلی، منطق شبکه عصبی بدین صورت است که ابتدا همه متغیرهای پیش‌بین را در یک لایه و متغیر یا متغیرهای هدف را در یک لایه دیگر، فهرست می‌نماید. سپس بین این دو لایه که شامل متغیرهای واقعی پژوهش می‌باشند چند لایه متغیر مجازی ایجاد می‌نماید. سپس از سمت لایه متغیرهای پیش‌بین تا لایه متغیرهای هدف تمامی متغیرها ارتباط دو بر دو برقرار می‌نماید. مجموعه حاصله که شامل لایه متغیرهای پیش‌بین، لایه متغیرهای هدف و لایه‌های متغیرهای مجازی می‌باشد عملاً تشکیل یک شبکه عصبی اولیه را می‌دهند. آن‌گاه نرم‌افزار، به هر یک از متغیرهای مجازی که خود، ایجاد کرده یک وزن تصادفی می‌دهد. سپس بر اساس مقادیر واقعی متغیرهای پیش‌بین و مقادیر وزن‌های تصادفی که به متغیرهای مجازی داده است اندازه متغیر هدف را پیش‌بینی نموده و سپس میزان فاصله (خطای) بین اندازه واقعی متغیر هدف با اندازه‌ای که خود محاسبه نموده را مشخص می‌نماید.

در نهایت با تغییرات مکرر و متوالی در وزن‌های تصادفی که به متغیرهای مجازی می‌دهد تا آن‌جا پیش می‌رود که به کمترین خطا (بهترین پیش‌بینی) در متغیر هدف برسد. در این حالت، الگوریتم شبکه عصبی به شکل بهینه خود رسیده و قابل بهره‌برداری در پیش‌بینی‌های واقعی است.

قابل انتظار است که پارامترهایی هم‌چون تعداد لایه‌های مجازی، تعداد تکرارهای نرم‌افزار در بهبود شبکه و غیره، توسط پژوهشگر قابل کنترل باشد.

روش توسعه شبکه عصبی در این پژوهش، روش exhaust prune بوده که قوی‌ترین شیوهی توسعه شبکه عصبی در نرم‌افزار کلمنتاین می‌باشد. تعداد لایه‌های پنهان شبکه، برابر سه، مجموع لایه‌های شبکه (با احتساب لایه‌های پیش‌بین و هدف) برابر ۵، و تعداد دورهای بازنگری در اوزان گره‌های شبکه، مطابق با پیش فرض نرم‌افزار، برابر ۲۵۰ دور تنظیم گردیده است.

جامعه آماری پژوهش

جامعه آماری برای توسعه مدل پژوهش در هر یک از ۳ دانشگاه مورد مطالعه به قرار زیر بوده است: کلیه رکوردهای (ترم- دانشجویهای) مقطع کارشناسی در رشته‌ای مهندسی از نیمسال اول سال تحصیلی ۹۲-۱۳۹۱ لغایت نیمسال دوم سال تحصیلی ۹۴-۱۳۹۳.

این رکوردها جمعاً در شش نیمسال تحصیلی اصلی (نیمسال‌های اول و دوم بدون احتساب نیمسال‌های تابستانه) جای می‌گرفته‌اند. طبیعتاً در هر یک از این نیمسال‌ها، رکوردهای مربوط به دانشجویان ترم اول، ترم دوم، ... الی ترم آخر موجود بوده است. در هر یک از دانشگاه‌های مورد مطالعه، از ۵۰٪ رکوردها برای یادگیری مدل، ۳۰٪ برای آزمون مدل و ۲۰٪ برای اعتبارسنجی استفاده شده است. بخش‌بندی فوق با استفاده از گره پارتیشن^۱ نرم‌افزار و به صورت کاملاً تصادفی صورت پذیرفته است.

اقدامات پیش از مدل‌سازی: مطابق معمول پروژه‌های داده‌کاوی: حذف فیله‌های زائد، حذف داده‌های پرت، نرمال‌سازی دامنه پراکندگی و اندازه عددی متغیرها، کدگذاری داده‌ها، ارزیابی اولیه فیله‌ها و حذف متغیرهای ناکارآمد در این پژوهش قبل از اقدام به مدل‌سازی انجام شده است.

یافته‌های پژوهش

به‌طور کلی مدل‌های حاصل از شبکه عصبی، ۳ نوع یافته ارائه می‌نمایند. اول، گزارش پیش‌بینی مدل در خصوص اندازه عددی یا کلاس متغیر هدف برای یکایک اعضای جامعه آماری یا جامعه هدف پژوهش. دوم، اهمیت نسبی متغیرهای پیش‌بین به کار رفته در مدل و سوم، شاخص‌های صحت و کارایی مدل.

در مقاله حاضر از ارائه گزارش نوع اول یعنی گزارش پیش‌بینی مورد به مورد اعضای جامعه آماری صرفنظر گردیده است ولی گزارش‌های نوع دوم (جدول شماره ۲) و سوم (جدول شماره ۳)، حاصل از مدل پژوهش، ارائه می‌گردند.

جدول ۲. اهمیت نسبی متغیرها (فیلدهای) ورودی مدل پژوهش به تفکیک دانشگاه

دانشگاه ۳	دانشگاه ۲	دانشگاه ۱	برچسب	عنوان متغیر(فیلد)
۰,۰۷۹	۰,۴۳۶	۰,۳۵۴	Past semester ave	معدل نیمسال قبل
۰,۲۵۴	۰,۱۸۷	۰,۲۲۷	Total ave	معدل کل تا قبل از شروع نیمسال
۰,۰۴۵	۰,۰۳۸	۰,۱۴۲	Odd/ even	زوج یا فرد بودن نیمسال
۰,۰۹۷	۰,۰۳۴	۰,۰۳۵	Special units	تعداد واحدهای تخصصی در نیمسال
۰,۰۶۳	۰,۰۴۳	۰,۰۳۳	Minor	رشته تحصیلی
۰,۱۱۳	۰,۰۳۸	۰,۰۲۶	Extra activities	داشتن یا نداشتن فعالیت‌های فوق برنامه در نیمسال
۰,۰۳۵	۰,۰۴۲	۰,۰۳۳	Optional units	تعداد واحدهای اختیاری در نیمسال
۰,۰۴۵	۰,۰۲۶	۰,۰۱۴	Principal units	تعداد واحدهای علوم پایه در نیمسال
وارد نشده	بی اهمیت	۰,۰۰۳	Pre university units	تعداد واحدهای پیش‌دانشگاهی در نیمسال
۰,۰۱۱	۰,۰۴۱	۰,۰۱۱	Passed units	تعداد واحدهای گذرانده تا قبل از شروع نیمسال
۰,۰۲۷	۰,۰۳۹	۰,۰۰۹	General units	تعداد واحدهای عمومی در نیمسال
وارد نشده	بی اهمیت	۰,۰۰۸	Main units	تعداد واحدهای اصلی در نیمسال
۰,۱۶۸	۰,۰۷۶	۰,۰۰۷	Semester units	مجموع تعداد واحدهای نیمسال
وارد نشده	فاقد صلاحیت	فاقد صلاحیت	Project	داشتن/ نداشتن پروژه در نیمسال
وارد نشده	فاقد صلاحیت	فاقد صلاحیت	Semester kind	نوع نیمسال (عادی، مهمان)

در مدل‌های حاصل از شبکه عصبی، اهمیت نسبی هر یک از متغیرهای پیش‌بین با یک مقدار عددی بین صفر تا ۱ بیان می‌شود. به گونه‌ای که اندازه مجموع آن‌ها برابر با یک بشود. همان‌طور که ملاحظه می‌شود در هر سه دانشگاه، "معدل کل" و "معدل نیم سال قبل"، به‌عنوان پراهمیت‌ترین متغیرهای ورودی در پیش‌بینی مشروطی، کاربرد داشته‌اند. گرچه در دانشگاه‌های ۱ و ۲، معدل نیمسال قبل، و در دانشگاه ۳، معدل کل از بیشترین اهمیت نسبی برخوردار بوده‌اند. در دانشگاه ۱، "زوج یا فرد بودن نیمسال" با ضریب اهمیت نسبی ۰,۱۴۲ از سومین درجه اهمیت برای پیش‌بینی مشروطی برخوردار بوده در حالی‌که این متغیر در دو دانشگاه دیگر از ضریب اهمیت بالایی برخوردار نبوده است.

گرچه، همان‌طور که در فصل سوم توضیح داده شد، شبکه عصبی نشان نمی‌دهد که جهت تأثیرگذاری هر متغیر ورودی بر متغیر هدف، مثبت است یا منفی. اما بررسی جداگانه فراوانی‌ها نشان می‌دهد که مشروطی ترم‌های زوج بیشتر از ترم‌های فرد است.

بخش عمده‌ای از سایر متغیرهای این جدول به نوع و تعداد واحدهای اخذ شده دانشجوی در نیمسال اختصاص دارد. مجموع واحدهای اخذ شده دانشجوی در نیمسال مورد پیش‌بینی، در دانشگاه شماره ۳ با ضریب اهمیت ۰,۱۶۸، از اهمیت نسبی قابل توجهی برخوردار است. در حالی که این متغیر در دانشگاه‌های ۱ و ۲ از اهمیت چندان بالایی برخوردار نیست.

در کنار تعداد و نوع واحدهای درسی دانشجوی، پرداختن یا نپرداختن دانشجوی به فعالیت‌های رسمی فوق برنامه از نظر پژوهشگر و مدیریت دانشگاه‌های مورد مطالعه حائز اهمیت فراوان بوده است.

نتایج حاصل از اجرای مدل‌ها نشان می‌دهد که این متغیر در دانشگاه‌های ۱ و ۲، با اهمیت نسبی پایین‌تر از ۰,۰۵، تأثیر چندانی در مشروط شدن یا نشدن دانشجوی ندارد. این یافته را می‌توان این‌گونه تفسیر نمود که در این دو دانشگاه، چه دانشجوی به فعالیت‌های فوق برنامه بپردازد و چه نپردازد تقریباً در یک حد در معرض خطر مشروط شدن در نیمسال مربوطه قرار دارد.

اما در دانشگاه ۳ اهمیت نسبی این متغیر، بالاتر از ۰,۱ بوده و گرچه شبکه عصبی، توضیحی در مورد جهت آن نمی‌دهد، بررسی فراوانی‌ها نشان می‌دهد که پرداختن به فعالیت‌های فوق برنامه با مشروط شدن دانشجوی رابطه مثبت دارد. لازم به ذکر است در این پژوهش، عضویت در تیم‌های ورزشی، تشکل‌های دانشجویی و گروه‌های هنری، به‌عنوان مصادیق فعالیت‌های فوق برنامه در نظر گرفته شده‌اند.

و بالآخره اینکه ضریب اهمیت نسبی مجموع تعداد واحدهای گذرانده دانشجوی در مشروط شدن دانشجوی در یک نیمسال، در هر سه دانشگاه از اهمیت نسبی پائینتر از ۰,۰۵ برخوردار است. این یافته را می‌توان این‌گونه تفسیر نمود که دانشجویهای قدیمی‌تر و جدیدتر که تعداد واحدهای متفاوتی را گذرانده‌اند، صرف‌نظر از سایر متغیرها، به یک اندازه در معرض خطر مشروط شدن قرار دارند.

لازم به ذکر است درج عبارت "فاقد صلاحیت" در جدول ۲ بدین معناست که فیلد مربوطه برای دانشگاه مورد نظر در هنگام ارزیابی اولیه داده‌ها به‌عنوان متغیرهای کارآمد، شناسایی^۱ گردیده و لذا اصلاً در مدل‌سازی وارد نشده‌اند. و درج عبارت "وارد نشده" نیز بدین معناست که داده‌های مربوط به فیلد مورد نظر در سیستم اطلاعات دانشگاه مربوطه موجود نبوده است.

در اینجا ذکر یک نکته پنهان، ضروری است و آن هم وضعیت این دو متغیر برای ورودی‌های جدید (دانشجویان ترم اول) است. واقعیت این است که برای این دو متغیر بسیار پراهمیت در کارنامه دانشجویانی که اولین ترم تحصیل خود را شروع می‌نمایند داده‌ای وجود ندارد. به همین دلیل مدل حاصل از این پژوهش برای پیش‌بینی مشروطی دانشجویان ترم اول، مجبور است بدون بهره‌گیری از این دو متغیر مهم، صرفاً به سایر متغیرها بسنده نماید. هم‌چنین برای دانشجویان ترم دوم، معدل کل و معدل نیمسال قبل دانشجو عملاً با هم یکی است. لذا این دو متغیر برای دانشجویان ترم دوم از هم خطی کامل برخوردارند و لذا مجموعاً در حد یک متغیر کارآیی دارند.

اما با وجود توجه به این نکات پنهان در این پژوهش، از حذف رکوردهای ترم اول و دوم، اجتناب گردید و ترجیح داده شد که از اطلاعات کلیه رکوردهای موجود در جامعه آماری (که شامل دانشجویان نیمسال اول و دوم هم می‌شد) بهره برده شود، چراکه نگه داشتن رکوردهای مربوط به دانشجویان نیمسال اول و دوم به لحاظ حفظ داده‌های مربوط به سایر فیلدها حائز ارزش بود.

با توجه به واقعیت برشمرد شده می‌توان دو تفسیر متفاوت را ارائه داد. با نگاهی بدبینانه می‌توان گفت که مدل‌های توسعه داده شده در این پژوهش برای پیش‌بینی مشروطی دانشجویان نیمسال‌های اول و دوم در مقایسه با دانشجویان قدیمی‌تر از توان پایینتری برخوردار است، چراکه بر مبنای متغیرهای کمتری در مورد آن‌ها پیش‌بینی می‌نماید. در عین حال با نگاهی خوشبینانه می‌توان گفت مدل درست عمل می‌کند اما در مورد این دسته از دانشجویان شواهد کافی (معدل کل پایین و معدل ترم قبل پایین) برای پیش‌بینی مشروطی ندارد.

ارزیابی مدل‌های حاصل از پژوهش

به‌طور معمول، پژوهشگر علاقه‌مند است بعد از ساخت و اجرای مدل، از کیفیت عملکرد آن مطلع شود. گره تحلیل^۱ در کلمنتاین این امکان را به پژوهشگر می‌دهد که عملکرد مدلی را که توسعه داده تحلیل نماید. در آن‌دسته از پژوهش‌های داده‌کاوی که داده‌ها قبل از مدل‌سازی، مورد بخش‌بندی^۲ قرار می‌گیرند این تحلیل‌ها بر روی مدل نهایی و بر مبنای داده‌های بخش اعتبار سنجی^۳ انجام می‌شود. توجه این رویکرد آن است که اگر مدل توسعه داده شده، مدل معتبری باشد باید بتواند پیش‌بینی قابل قبولی در مورد داده‌هایی که هیچ‌گونه استفاده‌ای از آن‌ها در مدل‌سازی و آزمون مدل به عمل نیامده است به عمل آورد.

اما برخلاف مدل‌های پیش‌بینی متغیرهای پیوسته از قبیل رگرسیون و مدل‌های ساختاری که معمولاً با شاخص‌های برازش از قبیل R^2 مورد ارزیابی قرار می‌گیرند در مدل‌های پیش‌بینی دودویی مبتنی بر شبکه عصبی، شاخص‌های برازش کاربرد ندارد. چراکه این‌گونه مدل‌ها از نوع غیرخطی بوده و بحث برازش خط معادله بر داده‌ها اساساً موضوعیت ندارد. در عوض، این‌گونه مدل‌ها با شاخص‌های دیگری ارزیابی می‌شوند که اصلی‌ترین آن‌ها عبارتند از: صحت^۱ و کارایی^۲.

جدول ۳. مقایسه شاخص‌های ارزیابی مدل‌های پژوهش در دانشگاه‌های مورد مطالعه

مورد مطالعه	شاخص صحت مدل		شاخص اطمینان (کارایی) مدل
	پیش‌بینی شرایط عادی	پیش‌بینی شرایط مشروطی	
دانشگاه ۱	۰٫۶۸	۰٫۶۶۱	٪۹۷٫۷۶
دانشگاه ۲	۰٫۷۶۹	۰٫۶۱۱	٪۹۹٫۳۵
دانشگاه ۳	۰٫۶۳۹	۰٫۶۰۶	٪۹۳٫۱۵

شاخص صحت مدل: این شاخص که به درصد بیان می‌شود بیانگر این است که فیلد هدف برای چه درصدی از رکوردها به درستی پیش‌بینی گردیده است. ملاحظه می‌شود که در هر سه دانشگاه صحت پیش‌بینی بالاتر از ۹۰ درصد و در دانشگاه‌های ۱ و ۲ بالاتر از ۹۵ درصد می‌باشد. شاخص کارایی مدل: قبل از پرداختن به اندازه عددی و تفسیر شاخص کارایی، توضیحی در خصوص مفهوم این شاخص، ضروری است. این توضیح در قالب یک مثال فرضی ارائه می‌شود: "چنانچه فرض شود برای یک رکورد که واقعاً دارای شرایط مشروطی (کد شناسه ۱) می‌باشد، مدل فرضی (الف) وضعیت این رکورد را با اندازه عددی (۰٫۷۵) و مدل فرضی (ب) با اندازه عددی (۰٫۹۸) پیش‌بینی نمایند.

با توجه به اینکه هر دوی این مقادیر عددی بالاتر از ۰٫۵ می‌باشند در جریان تبدیل نتایج پیش‌بینی این دو مدل، به کدهای دودویی، پیش‌بینی هر دوی این مدل‌ها به کد ۱ تبدیل می‌شود. و لذا پیش‌بینی هر دو مدل صحیح است. اما نکته در اینجاست که قابلیت اطمینان یا اعتماد مدل (ب) برای پیش‌بینی این رکورد، بسیار بالاتر از قابلیت اعتماد مدل (الف) می‌باشد. اما وضعیت زمانی پیچیده می‌شود که برای یک رکورد فرضی دیگر، مدل (الف) از قابلیت اعتماد بیشتری نسبت به مدل (ب) برخوردار باشد. به‌عنوان مثال: ۰٫۹۵ در مقابل ۰٫۶۶.

در چنین حالتی قابلیت اطمینان هر یک از مدل‌های فرضی الف و ب، می‌بایست با توجه کل رکوردها محاسبه شود و نه با توجه به یک یا تعداد معدودی از رکوردها. و این کاری است که الگوریتم شبکه عصبی انجام می‌دهد. این الگوریتم بعد از محاسبه فیلد هدف برای تمامی

رکوردها، فاصله مقادیر پیش‌بینی شده خود با وضعیت‌های واقعی را محاسبه نموده و بر مبنای آن شاخصی تحت عنوان ارزیابی عملکرد ارائه می‌نماید.

با توجه به مثال فرضی ارائه شده، حال می‌توان درک بهتری از شاخص کارآیی مدل پیش‌بینی مشروطی که در این پژوهش، توسعه داده شده، حاصل نمود.

اندازه عددی این شاخص در مدل نهایی (بر مبنای داده‌های بخش اعتبارسنجی، برای پیش‌بینی نیم سال عادی (با کد ۰)، برابر (۰٫۶۸) و برای پیش‌بینی نیم سال مشروط (کد ۱)، برابر (۰٫۶۶۱) می‌باشد.

ملاحظه می‌شود که شاخص عملکرد برای هر سه دانشگاه بالاتر از ۰٫۶ است که محدوده‌ای قابل قبول (ولی نه خیلی خوب) برای کارآیی مدل‌های حاصل از شبکه عصبی به حساب می‌آید.

فاصله‌ای که این شاخص تا عدد (۱) دارد مربوط به تأثیر وضعیت نیم سال از سایر متغیرهایی است که در این پژوهش در نظر گرفته نشده‌اند. لذا می‌توان این‌گونه نتیجه‌گیری کرد که اگر متغیرهای بیشتری در سیستم‌های آموزش دانشگاه‌ها ثبت شوند (مثلاً متغیرهای مربوط به طرح پایش سلامت دانشجویان) و از آن‌ها نیز در مدل‌سازی بهره برده شود می‌توان به مدل‌های پیش‌بینی با کارآیی بسیار بالاتری دست یافت.

زمان پیشنهادی برای اجرای ادواری مدل

به‌عنوان یک اصل علمی، بهترین زمان برای اجرای ادواری^۱ تقویمی هر مدل پیش‌بینی، برهه‌های زمانی‌ای است که اطلاعات مربوط به متغیرهای ورودی آن تا جای ممکن، "به روز" و "تثبیت" شده باشند. بر همین اساس، زمان پیشنهادی پژوهشگر برای اجرای ادواری مدل پیش‌بینی مشروطی نیم سال، بعد از حذف و اضافه^۲، در نیم سال مورد نظر می‌باشد. این زمان از جهات زیر بسیار مناسب است:

- در این زمان پیشنهادی، نوع و تعداد واحدهای درسی اخذ شده دانشجو در نیمسال مورد نظر، تثبیت شده است.
- معدل کل و معدل نیمسال قبل دانشجو تقریباً تثبیت شده است.
- دانشجو در هفته‌های ابتدایی نیمسال تحصیلی قرار داشته و فرصت کافی برای استفاده از نتایج مدل وجود دارد. منظور از استفاده از نتایج مدل، اجرای تدابیر لازم برای پیشگیری از مشروط شدن دانشجویان در معرض خطر مشروطی توسط دانشجو و دانشگاه می‌باشد.

1. Periodic

۲. واژه حذف و اضافه در برخی از دانشگاه‌ها به "حذف و اخذ" و در برخی از دانشگاه‌ها به "ترمیم واحدهای درسی" معروف است.

بحث و نتیجه‌گیری

پژوهش حاضر اساساً با این تفکر، طرح‌ریزی و اجرا گردید که داده‌های موجود در سیستم‌های اطلاعاتی دانشگاه‌ها، حاوی دانش نهفته‌ای هستند که با بهره‌گیری از اصول علم داده‌کاوی می‌توان آن را استخراج و تبیین نمود. و از آن‌جا که این پژوهش در فضای موضوعی فراگیران ضعیف دنبال می‌شد، تلاش گردید با توسعه مدل‌های مبتنی بر داده‌کاوی به دانش و مکانیزمی برای شناسایی به‌موقع دانشجویان آسیب‌پذیر در دانشگاه‌های مورد مطالعه دست یافته شود.

در این پژوهش، واژه "آسیب‌پذیری تحصیلی" به‌عنوان یک اصطلاح نوآورانه برای نگاه به ضعف تحصیلی دانشجویان ابداع و معرفی گردید. و بر این فرض کلی استوار است که چنان‌چه دانشجویان آسیب‌پذیر، به‌موقع، مورد شناسایی قرار گیرند امکان مراقبت و پیشگیری از بحران تحصیلی آن‌ها وجود دارد.

پژوهش حاضر نشان داد که متغیرها، داده‌ها و اطلاعات موجود در سیستم‌های اطلاعاتی دانشگاه‌های مورد مطالعه تا حد بسیار مفیدی در پیش‌بینی مشروطی دانشجو در یک نیمسال تحصیلی قابل استفاده‌اند. این پاسخ بدین معناست که می‌توان به‌جای آن‌که به سیستم‌های اطلاعات مدیریت دانشگاه‌ها فقط در سطح اطلاعاتی و کارکردی نگریست به‌عنوان بستری برای کشف و استخراج دانش (حداقل دانش پیش‌بینی وضعیت نیمسال) نگاه کرد. چنین رویکردی نیاز پژوهشگران آتی به شیوه‌های پیمایشی و ابزارهای پرسشنامه‌ای را به شدت می‌کاهد. در واقع این پژوهش هم‌زمان با پژوهش‌های نوین داده‌کاوی آموزشی این پیام را تقویت می‌نماید که در سال‌های آتی، پژوهش‌های پیمایشی مبتنی بر پرسشنامه تنها در مواقعی توجیه می‌یابند که داده‌های مورد نیاز آن‌ها در سیستم‌های اطلاعاتی موجود نباشند.

مدل‌های این پژوهش نشان دادند که اهمیت نسبی هر متغیر پیش‌بین از دانشگاهی به دانشگاه دیگر متفاوت است. این بدان معناست که در داده‌کاوی‌های آموزشی نمی‌توان به دنبال مدل جامعی بود که پیش‌بینی قابل قبولی برای همه دانشگاه‌ها (حتی برای سه دانشگاه که همگی از نوع فنی مهندسی هستند) انجام دهد. چنین نتیجه‌ای با تمامی پژوهش‌های مشابه پیشین از این منظر هم‌راستاست که هر دانشگاه، نیازمند مدل اختصاصی خودش می‌باشد. بنابراین هیچ‌گونه قصد یا ادعای تعمیم نتایج در این پژوهش وجود ندارد.

از آنجا که الگوریتم‌های شبکه عصبی به‌طور کلی الگوریتم‌های بسیار دقیق و کارآمدی هستند هر گونه ضعف در شاخص‌های صحت و کارایی مدل‌های حاصله را می‌توان ناشی از نامناسب بودن یا ناکافی بودن داده‌ها و متغیرها دانست. یافته‌های این پژوهش نشان دادند که صحت مدل‌های پژوهش در حد متعارف مدل‌های حاصل از شبکه عصبی هستند لکن کارایی مدل‌ها قدری پایین (گرچه قابل قبول) است. و چنین پیامدی قابل انتظار است. زیرا از انبوه متغیرهایی که می‌

توانند در سرنوشت دانشجو در یک نیمسال تأثیرگذار باشند در این پژوهش فقط از متغیرهایی بهره برده شد که اطلاعات آن‌ها در سیستم‌های اطلاعات دانشگاه‌ها وجود داشته است. با این تفسیر می‌توان به دانشگاه‌ها توصیه نمود که با ثبت داده‌هایی از قبیل نتایج طرح پایش سلامت دانشجویان در سیستم‌های دانشگاهی، پتانسیل پیش‌بینی سیستم‌های خود را افزایش دهند. پژوهش حاضر از منظر توجه به فعالیت‌های فوق برنامه با پژوهش عازفی و همکاران (۱۳۹۱) مشابه و تا حدی قابل مقایسه است. لکن از جهت کارکرد متغیرها متفاوت می‌باشد. چرا که در پژوهش حاضر از مشارکت دانشجویان در فعالیت‌های فوق برنامه به‌عنوان یک متغیر پیش‌بین در وضعیت تحصیلی استفاده گردیده است در حالی که در پژوهش اشاره شده، از دیدگاه دانشجویان به ارزیابی فعالیت‌های فوق برنامه پرداخته شده است.

منابع

- ابوالقاسمی، م. و میرالی رستمی، ا. (۱۳۹۲). عوامل مؤثر بر پیشرفت تحصیلی دانشجویان دانشکده‌های فنی و مهندسی دانشگاه تهران به منظور ارائه مدلی برای پیش‌بینی پیشرفت تحصیلی آنها. *فصلنامه آموزش مهندسی ایران*، ۵۸، ۸۴-۶۷.
- احمدوند، م. ع. (۱۳۸۰). مقایسه سبک‌های مطالعه و یادگیری دو گروه قوی و ضعیف فراگیران. *مجله علوم انسانی دانشگاه سیستان و بلوچستان*، ۲۰-۷.
- چنگیزی آشتیانی، س.، شمسی، م. و محمد بیگی، ا. (۱۳۸۸). فراوانی افت تحصیلی و برخی از عوامل مؤثر بر آن از دیدگاه دانشجویان علوم پزشکی اراک. *مجله علمی پژوهشی دانشگاه علوم پزشکی اراک*، ۱۲(۴): ۳۳-۲۳.
- دسترنج، م.، بلوکی، ص. و موذن، م. (۱۳۹۲). بررسی عوامل مؤثر بر افت تحصیلی دانشجویان دانشگاه پیام نور بستک در سال ۱۳۸۹. *فصلنامه تخصصی علوم اجتماعی دانشگاه آزاد اسلامی واحد شوشتر*، ۷(۲۰): ۲۵۸-۲۴۱.
- دلارام، م.، آئین، ف. و فروزنده، ن. (۱۳۹۱). عوامل مؤثر بر مشروط شدن دانشجویان دانشگاه علوم پزشکی شهرکرد. *مجله پزشکی هرمزگان*، ۲(۱۶۳): ۱۶۳-۱۷۲.
- رحمتی، ع. لسانی، م. و خلیل زاده، ر. (۱۳۹۱). عوامل مرتبط با مشروطی دانشجویان شهید باهنر کرمان در سال ۱۳۸۸-۸۹ و ارائه مدل تحلیلی آن. پایان‌نامه کارشناسی ارشد. *دانشگاه شهید باهنر کرمان، دانشکده ادبیات و علوم انسانی*.
- رودباری، م.، احمدی، آ. و عبادی فرد آذر، ف. (۱۳۸۹). تعیین عوامل مؤثر بر پیشرفت تحصیلی دانشجویان دانشگاه علوم پزشکی ایران در سال تحصیلی ۸۸-۸۹. *فصلنامه طب و تزکیه*، ۱۹(۳)، ۴۸-۳۷.
- زارعی، ج.، عزیزی، ا. و کاظمی، ا. (۱۳۹۳). بررسی عوامل مؤثر بر کاهش میل به ادامه تحصیل در دانشجویان مقطع کارشناسی دانشگاه جندی شاپور اهواز. *مجله توسعه آموزش در علوم پزشکی*، ۷(۱۵): ۵۹-۴۹.
- شهرابی، جمال (۱۳۹۲). *داده کاوی ۲*. چاپ دوم. انتشارات جهاد دانشگاهی، واحد دانشگاه صنعتی امیرکبیر. تهران. ایران.
- شهرابی، جمال و زارع، ابوالفضل (۱۳۹۲). *داده کاوی با کلمنتاین*. چاپ اول. انتشارات جهاد دانشگاهی، واحد دانشگاه صنعتی امیرکبیر. تهران. ایران.
- صباغیان، زهرا و پور کاظمی، محمد حسین. (۱۳۸۷). بررسی و مقایسه موفقیت تحصیلی دانشجویانی که از توالی رشته در دوره کارشناسی ارشد برخوردارند با دانشجویانی که فاقد توالی رشته اند. *دو فصلنامه مدیریت و برنامه‌ریزی در نظام‌های آموزشی*، ۲۱(۲)، ۲۱-۷.
- عازفی، محبوبه، قهرمانی، محمد، ابوالقاسمی، محمود. و خرسندی یامچی، اکبر. (۱۳۹۲). اثربخشی فعالیت‌های فوق برنامه دانشگاه شهید بهشتی. *دو فصلنامه مدیریت و برنامه‌ریزی در نظام‌های آموزشی*، ۶(۱۰)، ۷۳-۴۸.

مدلی برای پیش‌بینی آسیب‌پذیری تحصیلی مقطع کارشناسی مبتنی بر شبکه عصبی ۱۰۱

علیزاده، سمیه و ملک محمدی، سمیرا (۱۳۹۳). داده کاوی و کشف دانش گام به گام با نرم افزار کلمنتاین. چاپ سوم. انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی. تهران. ایران.

غفاری، ا.، کارشکی، ح. و رضائی، م. (۱۳۸۹). مقایسه میزان و عوامل مؤثر بر افت تحصیلی دانشجویان ایرانی و افغانی دانشگاه فردوسی مشهد و دانشگاه هرات. پایان نامه کارشناسی ارشد. دانشگاه فردوسی مشهد. دانشکده علوم تربیتی و روانشناسی. کانتاردزیک، ام. داده کاوی. ترجمه علیخانزاده، امیر (۱۳۹۲). چاپ سوم. انتشارات علوم رایانه. بابل. ایران.

لشکرگیر، سعادت جو، لطیف. (۱۳۹۲). پیش بینی عملکرد دانش آموزان با استفاده از تکنیک داده کاوی و ارائه راهکار مناسب برای بهبود آن. پایان نامه کارشناسی ارشد. دانشگاه آزاد اسلامی. واحد یزد.

مقیم، سید محمد، خنیفر، حسین، زرنندی، نفیسه، جوکار، احمد، بیان معمار، سید احمد. و فرجی ارمکی، اکبر. (۱۳۸۹). مطالعه و بررسی رابطه بین عوامل آموزشی و افت تحصیلی در میان دانش آموزان پسر آموزشگاه‌های راهنمایی. دو فصلنامه مدیریت و برنامه ریزی در نظام‌های آموزشی، ۳(۴)، ۱۶۰-۱۴۴

- Blass, E., Jasman, A. & Shelley, S. (2010). Visioning 2035: The Future of the Higher Education Sector in the UK. *Futures*, 42, 445–453.
- Chen, S. and Voyles, D. (2013). HESI Admididion Assessment Scores: Predicting Student Ssuccess. *Journal of Professional Nursing*, 9(25): 32–37
- Dorothyjean, C. (2012). Potential for Significant Reductions In Dropout Rates: Analysis of An Entire ۳rd Grade State Cohort. *Economics of Education Review*, 31, 644–662.
- Giambona, F., Erasmo, V., & Vassiliadis, E. (2011). Educational Systems Efficiency in European Union Countries. *Studies in Educational Evaluation*, 37, 108–12.
- Gil-Galván, R., & Gil-Galván, F. J. (2013). How to Use Professional and Life Projects to Guide University Students towards Optimal Professional Development. *Procedia - Social and Behavioral Sciences*, 93, 1901 – 1905.
- Gqweta, N. (2012). A perspective of final year diagnostic radiography students. *Radiography*, 18, 212-217.
- Hamaideh, S. H., & Hamdan-Mansour, A. M. (2013). Psychological, Cognitive and Personal Variables that Predict College Academic Achievement among Health Sciences Students. *Nurse Education Today*, 34, 703-708.
- Hardinger, K. L., Schauner, S., Graham, M., & Garavalia, L. (2013). Admission Predictors of Academic Dismissal for Provisional and Traditionally Admitted Students. *Currents in Pharmacy Teaching and Learning*, 5, 33–38.
- Islam Shovon, H. and Haque, M. (2012). An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree. *International Journal of Advanced Computer Science and Applications*, 3(8): 145- 149
- Juklová, K. (2012). Analysis of University Education In Terms of the Level of Cognitive Study Goals – From the Perspective of Future Teachers. *Procedia - Social and Behavioral Sciences*, 69, 1610 – 1615.
- Knauss, P. J. & Wilsson, P. (2013). Predicting Early Academic Success: HESI Admissions Assessment Exam. *Journal of Professional Nursing*, 29, 28-31.
- Laurens, C., Krist, D. W., Erwin, O., & Ides, N. (2012). Efficiency and Equity in Private and Public Education: A Nonparametric Comparison. *European Journal of Operational Research*, 202, 563–573.
- Mikolaj, H., Steven, R. (2013). Divergent Historical Experiences and Inequality in Academic Achievement: The Case of Poland. *The Journal of Socio-Economics*, 42, 1–12.

- Nasir, N. A., Rasid, N. S., Ahmad, N. & Noor Shah, M. S. (2013). Bicluster Analysis as an Effective Tool to Measure Students Overall Performance. *Procedia - Social and Behavioral Sciences*, 90, 593 – 598.
- Reichmann, M. (2012). Future-Oriented Higher Education: Which Key Competencies Should Be Fostered Through University Teaching And Learning? *Futures*, 44, 127–135.
- Rodgers, S., Stenhouse, R., McCreddie, M., & Small, P. (2013). Recruitment, Selection and Retention of Nursing and Midwifery Students in Scottish Universities. *Nurse Education Today*, 33, 1301–131.
- Rosander, P., Bäckström, M., & Stenberg, G. (2011). Personality Traits and General Intelligence as Predictors of Academic Performance: A Structural Equation Modeling Approach. *Learning and Individual Differences*, 21, 590–596.
- Saklofske, D. H., Austin, E. J., Mastoras, S. M., Beaton, L., & Osborne, S. E. (2012). Relationships of Personality, Affect, Emotional Intelligence and Coping with Student Stress and Academic Success: Different Patterns of Association for Stress and Success. *Learning and Individual Differences*, 22, 251–257.
- Stukalina, Y. (2013). Management of the Educational Environment: The Context in Which Strategic Decisions Are Made. *Procedia - Social and Behavioral Sciences*, 99, 1054 – 1062.
- U.S. Department Of Education, for Each And Every Child (2013). *A Strategy for Education Equity and Excellence*, Washington, D.C.
- Vesela, D., & Klimova, K. (2013). Supporting Creative Industries with Innovative University Study Programs. *Procedia - Social and Behavioral Sciences*, 81, 152 – 156
- Warwick, P. (2014). The International Business of Higher Education. A Managerial Perspective on the Internationalization of UK Universities. *The International Journal of Management Education*, 12, 91-103
- Weber, A. S. (2011). The Role of Education in Knowledge Economies in Developing Countries. *Procedia Social and Behavioral Sciences*, 15, 2589–2594
- Willcockson, I. U., Johnson, C. W., Hersh, W. & Bernstam, E. V. (2009). Predictors of Student Success in Graduate Biomedical Informatics Training: Introductory Course and Program Success. *Journal of the American Medical Informatics Association*, 16, 837–846
- Zimmermann, J., Heinimann, H.R. & Bachmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. *Journal of Educational Data Mining*, 7(3), 151- 176
- Zoghbi, C., Fabiana, A. R., & Enlinson, M. (2013). Education Production Efficiency: Evidence from Brazilian Universities. *Economic Modeling*, 31, 94–103.