



دانشگاه کاشان
University of Kashan

مجله محاسبات نرم

SOFT COMPUTING JOURNAL

تارنمای مجله: scj.kashanu.ac.ir



مقایسه عملکرد روش‌های هوش مصنوعی در پیش‌بینی پیشرفت تحصیلی دانش‌آموزان*

ملیحه نیک‌سیرت^{1*}، استادیار، جواد طیبی¹، دانشیار، محمدمهدی ایزدخواه¹، استادیار

¹ دانشکده مهندسی کامپیوتر و صنایع، دانشگاه صنعتی بیرجند، بیرجند، ایران.

چکیده

اطلاعات مقاله

تاریخچه مقاله:

دریافت 5 تیر ماه 1402

پذیرش 3 آذر ماه 1402

کلمات کلیدی:

پیشرفت تحصیلی

هوش مصنوعی

ماشین بردار پشتیبان

K-نزدیک‌ترین همسایه

درخت‌های تصمیم

جنگل تصادفی

اعتباریابی

امروزه، روش‌های هوش مصنوعی می‌تواند از طریق کشف ارتباط بین متغیرهای مختلف، دانش نهفته در مجموعه داده‌های آموزشی و تربیتی دانش‌آموزان را استخراج نماید. این دانش می‌تواند سیستم‌های آموزشی را در تصمیم‌گیری بهتر و داشتن طرح‌های پیشرفته‌تری در جهت بهبود عملکرد تحصیلی دانش‌آموزان کمک کند. شناسایی عوامل تاثیرگذار بر پیشرفت تحصیلی دانش‌آموزان و استفاده از روشی که با بیشترین درصد درستی بتواند پیشرفت تحصیلی دانش‌آموزان را پیش‌بینی کند، هدف پژوهش حاضر است. در این پژوهش روش‌های ماشین بردار پشتیبان، K-نزدیک‌ترین همسایه، درخت تصمیم و جنگل تصادفی به منظور پیش‌بینی پیشرفت تحصیلی بکار گرفته شده است. در نهایت اعتباریابی مدل‌ها با استفاده از معیارهای دقت، صحت، حساسیت، تشخیص‌پذیری، میزان خطای طبقه‌بندی و معیار ترکیبی مورد بررسی قرار گرفته است. یافته‌های پژوهش نشان می‌دهد که مدل ماشین بردار پشتیبان نسبت به سایر مدل‌ها بهترین عملکرد را در سنجش پیشرفت تحصیلی دانش‌آموزان داشته است.

© 1402 نویسنده‌گان. مقاله با دسترسی آزاد تحت مجوز CC-BY

1. مقدمه

علاوه بر این، افزایش سرعت کامپیوترها و پیشرفت علم هوش مصنوعی، کاربرد این حوزه را در علوم مختلف از جمله آموزش بسیار وسیع کرده است [3]. با این وجود، کاربرد روش‌های هوش مصنوعی جهت پیش‌بینی پیشرفت تحصیلی دانش‌آموزان در داخل کشور محدود بوده است. از طرف دیگر، طیف زیادی از عوامل تاثیرگذار بر پیشرفت یا افت تحصیلی، پیش‌بینی این مقوله را به امری پیچیده تبدیل کرده است [4]، [5].

با توجه به اهمیت بالای آموزش، مسئولین امر در پی تغییر مناسب و اصلاح ساختار برنامه‌های آموزشی هستند. یکی از اهداف جدی در بهبود مدیریت امور آموزش مدارس، تخمین عملکرد تحصیلی دانش‌آموزان در پایان سال تحصیلی با هدف شناسایی دانش‌آموزانی است که دچار افت تحصیلی شده‌اند.

پیشرفت تحصیلی دانش‌آموزان یکی از اهداف مهم نظام آموزش و پرورش است که در سال‌های اخیر توجه پژوهشگران این حوزه را به خود جلب کرده است. امروزه در سیستم‌های آموزش حجم عظیمی از داده‌های مختلف و گاه به ظاهر متناقض وجود دارد که می‌توان از آنها اطلاعات مفیدی در ارتقای عملکرد تحصیلی دانش‌آموزان استخراج کرد [1]، [2].

* نوع مقاله: پژوهشی

* نویسنده مسئول

پست(های) الکترونیک: niksirat@birjandut.ac.ir (نیک‌سیرت)

javadtayyebi@birjandut.ac.ir (طیبی)

izadkhah@birjandut.ac.ir (ایزدخواه)

از این پایگاه داده‌ها هدایت مدیران، محققان، برنامه‌ریزان و غیره برای اتخاذ تصمیم‌های راهبردی، توصیف وضعیت جاری و تهیه گزارش‌های مختلف می‌باشد [9].

ناتک و زویلینگ² در پژوهشی با هدف استخراج و مدیریت داده‌های آموزشی موسسات آموزش عالی ذکر کرده است که دانشگاه‌ها تمایل دارند بررسی کنند که دانشجویان آنها موفق خواهند بود و میزان موفقیت آنها را تخمین بزنند. ابزارهای هوش مصنوعی به عنوان یک روش جدید به منظور تجزیه و تحلیل داده‌ها مورد استفاده قرار می‌گیرد. درصد موفقیت دانشجویان یک موسسه بیانگر اعتبار آن موسسه و بسیار ارزشمند است [10].

شهیری و همکاران در پژوهشی با هدف کاربرد روش‌های داده-کاوی در پیش‌بینی عملکرد دانشجویان در کشور مالزی بیان می‌کند راهکارهای متفاوتی به منظور سنجش عملکرد تحصیلی دانشجویان وجود دارد که داده‌کاوی آموزشی یکی از موثرترین و محبوب‌ترین روش‌های ارزیابی عملکرد دانشجویان است که برای پیش‌بینی عملکرد دانشجویان و با هدف ارائه یک برنامه‌ریزی آموزشی هدفمند به مسئولین دانشگاه‌ها و موسسات آموزشی کمک می‌نماید تا بر عملکرد دانشجویان خود نظارت داشته باشند و فعالیت‌های یادگیری آنها را بهبود بخشند [11].

برنز³ و همکاران در مقاله‌ای با هدف شناسایی زود هنگام خطر در دانشجویان و پیش‌بینی و پیشگیری از افت تحصیلی با استفاده از داده‌های سیستم آموزشی و تکنیک‌های، تاکید کرده‌اند که برای پیشرفت تحصیلی دانشجویان و جلوگیری از افت یا ترک تحصیل آنها باید یک خلاقیت مناسب داشت که هم مقرون به صرفه باشد و هم به سرعت دانشجویان در معرض خطر را استخراج کند و برای بهبود عملکرد تحصیلی آنها برنامه ارائه نماید. برنز و همکاران سیستم تشخیص سریع طراحی کردند که اطلاعات صورت مرتب در پایان هر ترم تحصیلی به‌روزرسانی می‌شود و در تمام مقاطع تحصیلی دانشجو می‌تواند مفید باشد. این برنامه با یک تغییر کوچک در همه مراکز آموزشی در راستای نیازهای آن مجموعه قابل اجرا است [12].

ارائه راهکارهایی برای بهبود وضعیت تحصیلی این دانش‌آموزان است از اهداف پژوهش‌های اخیر است. تکنیک‌های هوش مصنوعی¹ می‌توانند پاسخ علمی برای این مساله ارائه نمایند [6]- [8]. استفاده از نتایج حاصل از این پژوهش می‌تواند مدیران مدارس را در جهت اصلاح فرآیندهای آموزشی و بهینه‌کردن آنها هدایت کند. از طرفی این مدل‌ها، مشاوره‌های هوشمند لازم را برای پیشگیری از افت تحصیلی دانش‌آموزان ارائه می‌نمایند. همان‌طور که در بیان شد فاکتورها و عوامل فراوانی در پیشرفت تحصیلی دانش‌آموزان موثر است. به منظور استخراج فاکتورهای موثر و پیش‌بینی پیشرفت تحصیلی دانش‌آموزان در این پژوهش از روش‌های هوش مصنوعی استفاده شده است.

این تحقیق تمام عوامل مؤثر بر پیشرفت تحصیلی دانش‌آموزان را استخراج کرده و بر اساس آنها با کاربرد روش‌های هوش مصنوعی مدلی ایجاد می‌کند که دقت مناسب و عملکرد بهبودیافته‌ای در پیش‌بینی پیشرفت تحصیلی دانش‌آموزان دارد. در این پژوهش از روش‌های ماشین بردار پشتیبان، K-نزدیکترین همسایه، جنگل تصادفی و درخت تصمیم استفاده شده است. همچنین اعتباریابی مدل‌ها با استفاده از معیارهای مختلف انجام شده است و عملکرد آنها با یکدیگر مقایسه شده است.

ساختار مقاله به شرح زیر تنظیم شده است. در بخش بعد ادبیات موضوع پژوهش مرور شده است. روش پژوهش در بخش سوم مقاله ارائه گردیده است. در بخش چهارم، روش ارائه شده و روش‌های هوش مصنوعی جهت پیش‌بینی پیشرفت تحصیلی در دانش‌آموزان مقطع متوسطه اول استان خراسان جنوبی بکار گرفته شده است و نتایج حاصل از روش‌های هوش مصنوعی با یکدیگر مقایسه شده است. در انتها جمع‌بندی مقاله ارائه شده است و پیشنهاداتی جهت پژوهش‌های آتی در این زمینه ارائه گردیده است.

2. بررسی ادبیات موضوع پژوهش

در هر سازمان برای امور آماری، بازرگانی، آموزشی، پرسنلی و غیره پایگاه داده‌هایی اخیراً خریداری و یا ایجاد شده است. هدف

² Natek and Zwillling

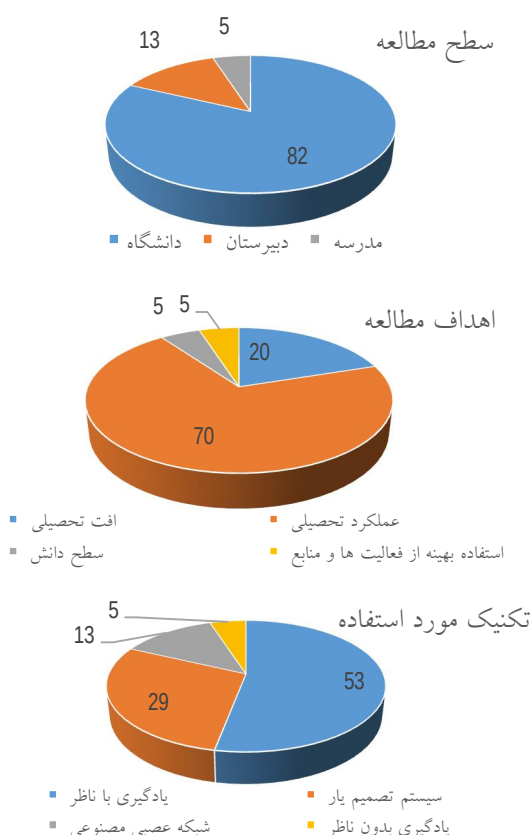
³ Berens

¹ Artificial intelligence

می‌تواند با توجه به سطح تحصیلی متفاوت باشد. اکثر موارد مورد مطالعه مربوط به سطح دانشگاه و سپس دبیرستان است.

- اهداف: اهداف شامل عملکرد تحصیلی، افت تحصیلی، داشتن و استفاده بهینه از منابع است.
- تکنیک‌ها: تکنیک‌ها شامل الگوریتم‌ها، روش‌ها و ابزارهای مختلفی است که داده‌ها را برای پیش‌بینی اهداف فوق تجزیه و تحلیل می‌کنند.
- الگوریتم‌ها و روش‌ها: الگوریتم‌های اصلی و روش‌های محاسباتی مورد استفاده در هر مورد در جدول به تفصیل آمده است. الگوریتم‌های دیگری که در این جدول نشان داده نشده‌اند نیز می‌توانند اعمال شوند.

شکل (1) به صورت گرافیکی آمار قابل توجهی در مورد روش‌ها، اهداف، سطح مطالعه و الگوریتم‌های در نظر گرفته شده در بررسی ادبیات را نشان می‌دهد. این نمودارها از جدول (1) به منظور درک بهتر تاثیر مرور ادبیات ساخته شده‌اند.



شکل (1): خلاصه‌سازی ادبیات موجود مساله تحقیق

هلال¹ در مقاله‌ای با هدف شناسایی عوامل تاثیرگذار در پیشرفت تحصیلی دانشجویان در مقطع کارشناس ارشد دانشگاه ایتالیا بیان می‌کند که دانشجویانی که بنا به دلایل اجتماعی و یا اقتصادی غیرقانونی هستند به احتمال زیاد افت تحصیلی دارند و همچنین دانشجویانی که با شرایط ویژه و یا بورسیه به دانشگاه وارد شده اند افت تحصیلی را به طور معمول تجربه می‌کنند. میزان حضور دانشجویانی که دسترسی کمتری به منابع درسی و تالار گفتگو دارند در دانشگاه بیشتر حضور است. همچنین دانشجویانی شاغل هستند و یا به صورت پاره‌وقت مشغول فعالیت می‌باشند در فعالیت‌های دانشگاه کمتر حضور دارند. شناسایی عوامل موثر بر پیشرفت تحصیلی دانشجویان به عنوان یک وظیفه اساسی و همراه با چالش زیاد است که بر عواملی چون جامعه و شرکت دانشجویان در دوره‌های مختلف موثر می‌باشد [13].

آل عمران² و همکاران در مقاله‌ای تحت عنوان فاکتورهای موثر بر عملکرد دانشجویان دبیرستانی می‌نویسد که برای استخراج عوامل موثر بر عملکرد دانشجویان و پیش‌بینی آن با استفاده از روش‌های داده‌کاوی شکاف تحقیقاتی وجود دارد. به راحتی نمی‌توان بهترین روش را برای پیش‌بینی عملکرد دانشجویان در تعیین کرد. علاوه بر این، شناسایی عوامل موثر در پیشرفت تحصیلی مورد غفلت قرار گرفته است و باید تحقیقات بیشتری انجام شود [15].

همچنین تعدادی از تحقیقات اخیر در زمینه کاربرد هوش مصنوعی در پیش‌بینی پیشرفت تحصیلی در جدول (1) گردآوری و خلاصه شده است. در این جدول مولفه‌های (ستون‌ها) زیر در نظر گرفته شده است.

- سطح دانش آموزان: هر مرجع، مجموعه داده‌های دانش - آموزان یک سطح خاص را تجزیه و تحلیل می‌کند. یک طبقه‌بندی وسیع از سطوح شامل مدرسه، دبیرستان و دانشگاه در نظر گرفته شده است. دانستن این اطلاعات می‌تواند مفید باشد زیرا مجموعه داده‌های مورد مطالعه از رفتار دانش‌آموزان حاکی از عوامل پنهانی است که

¹ Helal

² Emran

جدول (1): بررسی ادبیات موجود مساله تحقیق

مرجع	سطح مطالعه	هدف	تکنیک	الگوریتم و روش حل
[16]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	درخت تصمیم ¹ ، رگرسیون لجستیک ² ، بیز ساده ³ ، پیش نمونه برداری اقلیت ترکیبی (SMOTE ⁴)
[17]	دبیرستان	عملکرد تحصیلی	یادگیری با ناظر	K-نزدیکترین همسایه ⁵ ، بیز ساده، جنگل تصادفی
[18]	دانشگاه	افت تحصیلی	یادگیری با ناظر	درخت تصمیم
[19]	دانشگاه	افت تحصیلی	یادگیری با ناظر	شبکه های بیزی، درخت تصمیم، بیز ساده، شبکه های RBN ⁶
[20]	دانشگاه	افت تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، درخت تصمیم، رگرسیون خطی ⁷ ، ماشین بردار پشتیبان ⁸
[21]	دانشگاه	افت تحصیلی	یادگیری با ناظر	الگوریتم ترکیبی دسته بندی، K-نزدیکترین همسایه، شبکه های RBN ⁹ ، ماشین بردار پشتیبان
[22]	دانشگاه	افت تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی ¹⁰ ، شبکه های بیزی، درخت تصمیم،
[23]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، درخت تصمیم، رگرسیون لجستیک، بیز ساده
[24]	دانشگاه	افت تحصیلی	یادگیری با ناظر	آدابوست ¹¹ ، درخت تصمیم، رگرسیون لجستیک، بیز ساده، جنگل تصادفی، تحلیل بقا ¹²
[25]	دانشگاه	افت تحصیلی	یادگیری با ناظر	یادگیری عمیق، درخت تصمیم، K-نزدیکترین همسایه، مدل های خطی، بیز ساده
[26]	دانشگاه	افت تحصیلی	یادگیری با ناظر	ماشین بردار پشتیبان
[27]	دبیرستان	افت تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، درخت تصمیم، جستجوی مستقیم ترتیبی ¹³
[28]	مدرسه	افت تحصیلی	یادگیری با ناظر	جنگل تصادفی
[29]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، رگرسیون لجستیک، درخت تصمیم، بیز ساده، ماشین بردار پشتیبان
[30]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	پرسپترون چندلایه، رگرسیون خطی چندگانه، شبکه های RBN، ماشین بردار پشتیبان

¹ Decision Tree

² Logistic regression

³ Naïve Bayes

⁴ Synthetic Minority Over-Sampling

⁵ K-nearest neighbor

⁶ Radial Basis Networks

⁷ Linear regression

⁸ Support vector machine

⁹ Radial Basis Networks

¹⁰ Artificial neural network

¹¹ AdaBoost

¹² Survival Analysis

¹³ Sequential Forward Selection

ادامه جدول (1): بررسی ادبیات موجود مساله تحقیق

مرجع	سطح مطالعه	هدف	تکنیک	الگوریتم و روش حل
[31]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه های باور بیزی ¹
[32]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	تخصیص پنهان دریگله (LDA ²)، رگرسیون لجستیک، ماشین بردار پشتیبان
[33]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	پیش نمونه برداری اقلیت ترکیبی
[34]	دبیرستان	عملکرد تحصیلی	یادگیری با ناظر	جنگل تصادفی، ماشین بردار پشتیبان
[35]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	درخت گرادیان تقویت شده ³ ، مدل های خطی، جنگل تصادفی
[36]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	درخت تصمیم
[37]	دبیرستان	عملکرد تحصیلی	یادگیری با ناظر	فرآیند تصمیم مارکو
[38]	دبیرستان	عملکرد تحصیلی	یادگیری با ناظر	یادگیری ماشین ژاکوبین ماتریس-محور، پرسپترون چندلایه، شبکه های RBN، ماشین بردار پشتیبان
[39]	مدرسه	عملکرد تحصیلی	یادگیری با ناظر	آدابوست، شبکه عصبی مصنوعی، درخت تصمیم، جنگل تصادفی
[40]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	آدابوست، درخت تصمیم، ماشین بردار پشتیبان
[41]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	بیز ساده، K-نزدیکترین همسایه، رگرسیون لجستیک، ماشین بردار پشتیبان
[42]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، ماشین یادگیری حداکثر ⁴ ، بیز ساده، جنگل تصادفی، ماشین بردار پشتیبان
[43]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	جنگل تصادفی، ماشین بردار پشتیبان
[44]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی
[45]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	ماشین بردار پشتیبان
[46]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	تخصیص پنهان دریگله، ماشین بردار پشتیبان
[47]	دبیرستان	عملکرد تحصیلی	یادگیری بدون ناظر	شبکه عصبی مصنوعی، پرسپترون چندلایه، بیز ساده
[48]	دانشگاه	عملکرد تحصیلی	یادگیری بدون ناظر	یادگیری با ناظر، یادگیری بدون ناظر
[49]	دانشگاه	عملکرد تحصیلی	یادگیری بدون ناظر	رگرسیون لجستیک
[44]، [47]، [50]، [54]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی ⁵

¹ Bayesian Belief Network² Latent Dirichlet Allocation³ Gradient Boosted Tree⁴ Extreme Learning Machine⁵ Collaborative filtering

ادامه جدول (1): بررسی ادبیات موجود مساله تحقیق

مرجع	سطح مطالعه	هدف	تکنیک	الگوریتم و روش حل
[51]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی، ماشین بردار پشتیبان
[53]	دانشگاه	استفاده بهینه از فعالیت ها و منابع	سیستم تصمیم یار	فیلترینگ مشارکتی، سیستم تصمیم یار
[56]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	جستجوی دانش بیزی ¹ ، فیلترینگ مشارکتی
[57]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	تجزیه ماتریس
[58]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	سیستم پیش بینی نمره ²
[59]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی
[60]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	الگوریتم بیشینه سازی امید ³ ، تجزیه ماتریس، تجزیه مقادیر منفرد ⁴
[60]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	تجزیه ماتریس
[61]	دانشگاه	استفاده بهینه از فعالیت ها و منابع	سیستم تصمیم یار	فیلترینگ مشارکتی
[62]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی
[63]	دانشگاه	افت تحصیلی	شبکه عصبی مصنوعی	فیلترینگ مشارکتی، تجزیه مقادیر منفرد
[64]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	تجزیه ماتریس جهت دار ⁵ ، الگوریتم شیب اول دوقطبی ⁶ ، فیلترینگ مشارکتی، K-نزدیکترین همسایه، تجزیه ماتریس، الگوریتم شیب یک
[65]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	تجزیه ماتریس
[59]، [60]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	شبکه عصبی مصنوعی
[61]، [63]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	شبکه عصبی مصنوعی
[66]، [60]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	شبکه عصبی مصنوعی
[70]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	شبکه عصبی مصنوعی، رگرسیون لجستیک
[73]	مدرسه	عملکرد تحصیلی	شبکه عصبی مصنوعی	شبکه عصبی مصنوعی، ماشین یادگیری حداکثر
[66]، [68]	دانشگاه	سطح دانش	داده کاوی	داده کاوی
[74]	دانشگاه	سطح دانش	داده کاوی	داده کاوی
[76]	مدرسه	سطح دانش	یادگیری با ناظر	درخت تصمیم، درخت گرادیان تقویت شده
[70]، [78]	دانشگاه	افت تحصیلی	داده کاوی	داده کاوی

¹ Bayesian Knowledge Tracing

² Grade Prediction Advisor

³ Expectation-Maximization

⁴ Singular Value Decomposition

⁵ Biased-Matrix Factorization

⁶ Bipolar Slope One

پرسش‌نامه با 39 سوال و 4 مولفه شامل: عوامل فردی (سوالات؛ 1 تا 15)، خانواده (سوالات؛ 16 تا 27)، محیط و متغیرهای برون فردی (سوالات؛ 28 تا 30)، عوامل درون مدرسه‌ای و سیستم آموزش و پرورش (سوالات؛ 31 تا 39) است. نمره‌گذاری پرسش‌نامه بر اساس طیف لیکرت 5 درجه‌ای (خیلی زیاد=5؛ خیلی کم=1) می‌باشد. پایایی و روایی پرسش‌نامه توسط آزمون‌های آماری بررسی و اثبات شده است.

جامعه آماری پژوهش دانش‌آموزان مقطع متوسطه اول استان خراسان جنوبی است که بر اساس آمار دریافت شده از سازمان آموزش و پرورش برابر 41452 نفر می‌باشد. حجم نمونه پژوهش بر اساس جدول مورگان 384 نفر می‌باشد. به منظور ارزیابی دقیق تر مدل‌های هوش مصنوعی در این پژوهش، سه نمونه تصادفی استخراج و تجزیه و تحلیل شده است.

به منظور طراحی مدل سنجش پیشرفت تحصیلی، از روش‌های هوش مصنوعی شامل، ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و K-نزدیک‌ترین همسایه استفاده شده است. این مدل‌ها با استفاده از نرم‌افزار رپیدماینر² پیاده‌سازی شده است. همچنین جهت اعتباریابی مدل از شاخص‌های دقت، AUC، صحت، معیار ترکیبی، حساسیت، تشخیص‌پذیری و خطای طبقه‌بندی استفاده شده است [86].

4. نتایج و یافته‌ها

در این بخش ابتدا اطلاعات جمعیت‌شناختی نمونه مانند سن جنسیت و رشته تحصیلی مورد علاقه ارائه می‌شوند و در ادامه شاخص‌های توصیفی هر یک از متغیرها بررسی خواهند شد. سپس با استفاده از روش‌های هوش مصنوعی مدل‌های مختلفی شامل مدل درخت تصمیم، مدل جنگل تصادفی، مدل ماشین بردار پشتیبان و مدل K-نزدیک‌ترین همسایه به منظور سنجش پیشرفت تحصیلی دانش‌آموزان طراحی می‌شود.

1.1.4. اطلاعات جمعیت شناختی

نتایج آماری نشان داد که 68/8 درصد افراد شرکت‌کننده در

یکی از نکات مهم، زمانی که مطالعات قبلی با تمرکز بر روش‌های سنجش پیشرفت تحصیلی ارزیابی می‌شود، این است که به نظر می‌رسد علم داده و تکنیک‌های پیچیده‌تر تجزیه و تحلیل داده‌ها هنوز به طور کامل مورد استفاده قرار نگرفته‌اند. علیرغم علاقه فزاینده به این موضوع و اهمیت غیرقابل انکار آن، بر اساس دستاوردهای محقق از بررسی ادبیات موجود، محققان اساساً از طیف گسترده‌ای از روش‌های سنتی برای سنجش پیشرفت تحصیلی استفاده کرده‌اند. تحقیقات در مورد پیشرفت تحصیلی به طور سنتی مبتنی بر نظرسنجی بوده است (به عنوان مثال، بررسی یک گروه دانش آموز و پیگیری آنها برای یک دوره مشخص برای تعیین موفقیت آنها). مطالعات مبتنی بر نظرسنجی به دلیل عدم قابلیت تعمیم‌پذیری آنها برای سایر زمینه‌ها و دشواری و هزینه اجرا در مقیاس بزرگ مورد انتقاد قرار گرفته‌اند [80]. یک رویکرد جایگزین برای تحقیقات سنتی مبتنی بر نظرسنجی، یک استراتژی مبتنی بر داده است که از حجم وسیعی از داده‌ها که به طور معمول در پایگاه‌های داده سازمانی موجود است، استفاده می‌کند و از تکنیک‌های یادگیری ماشین برای استخراج بیش از داده‌ها استفاده می‌کند، چیزی که مطالعه حاضر قصد دارد دنبال نماید. اگرچه ادبیات موجود برتری رویکردهای مبتنی بر داده و تکنیک‌های هوش مصنوعی را بر روش‌های مبتنی بر نظرسنجی نشان می‌دهد، استفاده از هوش مصنوعی در زمینه آموزش هنوز در مراحل اولیه است.

3. روش پژوهش

این تحقیق به طراحی و اعتباریابی مدل سنجش پیشرفت تحصیلی بر پایه روش‌های هوش مصنوعی می‌پردازد. قلمرو مکانی این تحقیق مدارس مقطع متوسطه اول استان خراسان جنوبی می‌باشد. محقق به منظور استخراج عوامل موثر بر پیشرفت تحصیلی از روش تحلیل سیستماتیک با استفاده از نرم‌افزار مکس کیودا¹ بهره برده است. جهت تعیین ارتباط هر یک از عوامل استخراج شده بر پیشرفت تحصیلی، بر اساس نتایج حاصل از تحلیل سیستماتیک، پرسش‌نامه محقق ساخته طراحی شد. این

² RapidMiner

¹ Maxqda

3.4. مدل K-نزدیک‌ترین همسایه

مدل K-نزدیک‌ترین همسایه که به اختصار KNN نامیده می‌شود برای تخمین خروجی داده جدید (آزمون) از K تا نزدیک‌ترین همسایه نمونه آزمون در داده‌های آموزش کمک می‌گیرد. این مدل در سه مرحله برچسب داده‌های جدید را تخمین می‌زند:

- در مرحله اول مدل KNN فاصله نمونه آزمایشی را با تمامی نمونه‌های آموزشی محاسبه می‌کند.
- سپس در مرحله دوم براساس فاصله به دست آمده در مرحله اول، K تا از نزدیک‌ترین همسایه‌ها را از داده‌های آموزشی به داده‌های آزمایشی پیدا می‌کند.
- در مرحله آخر بر اساس رای‌گیری از بین K نزدیک‌ترین همسایه، کلاسی که بیشترین همسایه را از داده آموزشی به نمونه آزمایشی دارد انتخاب شده و نمونه جدید به آن کلاس تخصیص می‌یابد.

یکی از پارامترهای مهم در مدل KNN مقدار K است که باید به صورت آزمون و خطا مشخص شود. برای این منظور مقادیر متفاوتی برای K در نظر گرفته و مدل را بر روی مجموعه داده اجرا می‌کنیم در نهایت بهترین مقدار K را در نظر گرفته و برای مقایسه این مدل با سایر مدل‌ها ملاک عمل قرار می‌دهیم. برای این منظور مقدار K از 3 تا 15 تغییر داده شده است و دقت الگوریتم بررسی شده است. مشاهدات نشان داد که بیشترین دقت زمانی به دست می‌آید که مقدار K برابر 7 باشد. قابل ذکر است که برای مقادیر K بزرگتر از 15، هزینه محاسباتی الگوریتم به شدت افزایش پیدا می‌کند. با این وجود، با افزایش بیشتر مقدار K دقت الگوریتم بهبود پیدا نمی‌کند. بنابراین مقدار $K=7$ به عنوان مقدار بهینه انتخاب شده است.

تحقیق زن و 31/2 درصد از آنها مرد هستند. همچنین میانگین سن افراد مورد پژوهش 13/74 با انحراف معیار 1/04 می‌باشد (جدول (2)).

جدول (2): شاخص‌های پراکندگی و مرکزی سن

متغیر	میانگین	انحراف معیار	کمترین	بیشترین
سن	13/74	1/04	12	16

همچنین بر اساس جدول (3) می‌توان گفت اکثر افراد علاقه‌مند به رشته‌ی علوم تجربی هستند و فقط 10/4 درصد به رشته ریاضی و فیزیک علاقه دارند.

جدول (3): توزیع فراوانی رشته مورد علاقه

رشته	تعداد	درصد
ریاضی و فیزیک	39	10/4
علوم تجربی	178	47/3
علوم انسانی	66	17/6
سایر	93	24/7
جمع کل	376	100
بدون پاسخ	8	

2.4. تحلیل توصیفی متغیرهای پژوهش

در این بخش به ارائه و بحث در مورد شاخص‌های توصیفی هر یک از متغیرهای پژوهش پرداخته شده است. شاخص‌های تمرکز و پراکندگی برای عوامل موثر بر پیشرفت تحصیلی در جدول (4) آمده است. نتایج نشان می‌دهد که بعد محیط و متغیرهای برون فردی دارای بیشترین میانگین (3/79) و بعد عوامل فردی دارای کمترین میانگین (3/47) هستند. سه شاخص تمرکز میانگین، مد و میانه برای این متغیرها به هم نزدیک بوده و از طرفی کشیدگی و چولگی نیز بین 2- تا 2+ می‌باشد که نشان‌دهنده تقارن نسبی داده‌ها است.

جدول (4): شاخص‌های توصیفی متغیرهای پرسشنامه

متغیر	میانگین	میانه	مد	انحراف معیار	چولگی	کشیدگی
عوامل فردی	3/47	3/53	3/40	0/52	-0/48	0/28
خانواده	3/78	3/83	4/17	0/46	-0/83	0/86
محیط و متغیرهای برون فردی	3/79	3/67	4	0/67	-0/35	0/04
عوامل درون مدرسه ای و سیستم آموزش و پرورش	3/56	3/67	3/56	0/64	-0/57	0/01

ماتریس درهم‌ریختگی را برای مدل درخت تصمیم نشان می‌دهد.

جدول (6): ماتریس درهم‌ریختگی مدل درخت تصمیم			
	True A	True B	Class precision
Pred. A	64	13	83.12%
Pred. B	7	15	68.18%
Class recall	90.14%	35.57%	

ماتریس در هم‌ریختگی نشان می‌دهد که الگوریتم 64 مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین 15 مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم 13 مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و 7 مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است.

5.4. مدل جنگل تصادفی

مدل جنگل تصادفی مجموعه‌ای از درختان تصمیم‌گیری را با استفاده از داده‌ها، آموزش می‌دهد. پس از تولید درخت‌های تصمیم و آموزش آنها، نتیجه پیش‌بینی هر درخت به صورت جداگانه مشخص می‌شود. کارایی الگوریتم جنگل تصادفی در مورد مجموعه داده‌های بزرگ به اثبات رسیده است. همچنین پیچیدگی زمانی الگوریتم جنگل تصادفی، $O(\log_2 n)$ می‌باشد که در آن n تعداد رکوردها در مجموعه داده است [81]. ماتریس درهم‌ریختگی برای مدل جنگل تصادفی در جدول (7) نمایش داده شده است.

جدول (7): ماتریس درهم‌ریختگی مدل جنگل تصادفی			
	True A	True B	Class precision
Pred. A	65	13	83.33%
Pred. B	6	15	71.43%
Class recall	91.55%	53.57%	

ماتریس در هم‌ریختگی نشان می‌دهد که الگوریتم 65 مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس

برای مقدار $K=7$ ، ماتریس درهم‌ریختگی برای مدل KNN به صورت جدول (5) به دست آمد. علاوه بر این، پیچیدگی زمانی الگوریتم جنگل تصادفی، $O(dn^2)$ می‌باشد که در آن n تعداد رکوردها در مجموعه داده و d بعد داده است [81]، [83]. ماتریس درهم‌ریختگی نشان می‌دهد که الگوریتم 66 مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین 10 مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم 18 مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و 5 مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است. علاوه بر این از این ماتریس در ادامه برای محاسبه معیارهای ارزیابی الگوریتم و مقایسه عملکرد این مدل با سایر مدل‌ها استفاده می‌کنیم.

جدول (5): ماتریس درهم‌ریختگی الگوریتم KNN			
	True A	True B	Class precision
Pred. A	66	18	78.57%
Pred. B	5	10	66.67%
Class recall	92.96%	35.71%	

4.4. مدل درخت تصمیم

درخت تصمیم، درختی مانند مجموعه‌ای از گره‌ها است که برای ایجاد یک تصمیم در مورد وابستگی مقادیر به یک کلاس در نظر گرفته شده است. در ساختار درخت تصمیم، نتایج پیش‌بینی در قالب تعدادی از قواعد تشریح می‌شود. هر مسیر از ریشه تا یک برگ، یک قانون را توضیح می‌دهد و در نهایت برچسب برگ با کلاسی که بیشترین مقدار رکورد در آن تعلق گرفته مشخص می‌شود. از آنجا که ساختار درخت تصمیم یک فلوچارت ساده است، یکی از سریع‌ترین روش‌ها برای شناسایی متغیرهای تأثیرگذار و روابط بین دو یا چند متغیر محسوب می‌شود. قابل ذکر است که پیچیدگی زمانی الگوریتم جنگل تصادفی، $O(mn \log_2 n)$ می‌باشد که در آن n تعداد رکوردها در مجموعه داده و m تعداد ویژگی‌ها است [82]. جدول (6)

جدول (8): ماتریس درهم‌ریختگی مدل SVM

	True A	True B	Class precision
Pred. A	65	10	86.67%
Pred. B	6	18	75.00%
Class recall	91.55%	64.29%	

ماتریس درهم‌ریختگی نشان می‌دهد که الگوریتم 65 مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین 18 مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم 10 مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و 6 مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است. علاوه بر این از این ماتریس در ادامه برای اعتباریابی الگوریتم و مقایسه عملکرد این مدل با سایر مدل‌ها استفاده می‌کنیم.

7.4. اعتباریابی مدل‌ها و مقایسه عملکرد روش‌های هوش

مصنوعی

در مجموع در این پژوهش از روش‌های K-نزدیک‌ترین همسایه، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی به منظور پیش‌بینی پیشرفت تحصیلی در دانش‌آموزان مقطع متوسطه اول استان خراسان جنوبی استفاده شده است. به منظور بررسی عملکرد این روش‌ها، معیارهای دقت¹، صحت²، معیار ترکیبی³، حساسیت⁴، تشخیص‌پذیری⁵، خطای طبقه‌بندی⁶ و AUC ارزیابی و در ادامه گزارش شده است. به منظور ارزیابی دقیق‌تر هر یک از مدل‌های پیشنهادی، معیارهای اشاره شده برای تمام نمونه‌های استخراج شده محاسبه شده است. جدول (9) میانگین هر یک از معیارها را به تفکیک مدل‌های پیشنهادی گزارش می‌کند.

¹ Accuracy

² Precision

³ F-measure

⁴ Sensitivity

⁵ Specificity

⁶ Classification error

A طبقه‌بندی کرده است. همچنین 15 مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم 13 مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و 6 مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است.

از مقایسه ماتریس‌های درهم‌ریختگی برای دو مدل جنگل تصادفی و مدل درخت تصمیم مشاهده می‌شود که مدل جنگل تصادفی از دقت مناسب‌تری برخوردار است. دلیل این امر این است که الگوریتم جنگل تصادفی یک روش طبقه‌بندی است که داده‌ها را به زیرمجموعه‌هایی تقسیم کرده و در هر زیرمجموعه از یک درخت تصمیم استفاده می‌کند. در نهایت با هدف بهبود دقت پیش‌بینی، میانگین نتایج درخت تصمیم در نظر گرفته می‌شود. بنابراین جنگل تصادفی به جای استفاده از یک درخت تصمیم، پیش‌بینی را از تمام درخت‌ها و براساس رای اکثریت انجام می‌دهد. بنابراین استفاده از تعداد بیشتر درختان در جنگل دقت بالاتری را نتیجه می‌دهد.

6.4. مدل ماشین بردار پشتیبان

مدل ماشین بردار پشتیبان (SVM) یکی از مدل‌های توانمند و انعطاف‌پذیر هوش مصنوعی است که هم برای رگرسیون و همچنین طبقه‌بندی مورد استفاده قرار می‌گیرند. اخیراً، استفاده از این مدل به دلیل قابلیت آنها در برخورد با چندین متغیر پیوسته و قطعی بسیار مطلوب بوده است.

اساساً یک مدل SVM، کلاس‌های متفاوت را در یک ابر صفحه در فضای چندبعدی نمایش می‌دهد. این ابر صفحه با هدف کاهش میزان خطا به صورت تکراری توسط SVM تولید می‌شود [84]، [85]. بنابراین هدف از مدل SVM تقسیم‌بندی داده‌ها در کلاس‌ها و یافتن یک ابر صفحه با حداکثر حاشیه می‌باشد. قابل توجه است که پیچیدگی زمانی الگوریتم SVM، $O(n^3)$ می‌باشد که در آن n تعداد رکوردها در مجموعه داده است [72]. بعد از پیاده‌سازی مدل SVM بر روی مجموعه داده پیشرفت تحصیلی دانش‌آموزان ماتریس درهم‌ریختگی جدول (8) به دست آمد.

جدول (9): ارزیابی و مقایسه عملکرد روش‌های یادگیری ماشین

معیار ارزیابی	KNN	درخت تصمیم	جنگل تصادفی	SVM
دقت (فاصله اطمینان 95 درصد)	77.27 (0.765,0.797)	78.9 (0.764,0.822)	81.32 (0.802,0.825)	82.24 (0.551,0.820)
AUC	0.632	0.873	0.849	0.843
صحت	65.95	69.22	72.13	74.96
معیار ترکیبی	48.23	62.43	63.41	71.09
حساسیت	36.34	53.02	54.32	63.99
تشخیص پذیری	92.57	89.87	92.08	92.17
خطای طبقه‌بندی	23.23	20.20	19.19	16.16
زمان اجرای الگوریتم (ثانیه)	3.13	3.28	3.10	3.10

نتایج جدول (9) نشان می‌دهد که مدل SVM در تقریباً تمامی معیارها (به جز معیار AUC) عملکرد بهتری نسبت به سایر روش‌ها در طبقه‌بندی داده‌های پیشرفت تحصیلی داشته است. علاوه بر این، مدل جنگل تصادفی، مدل درخت تصادفی و مدل KNN به ترتیب در رتبه‌های بعدی قرار دارد. همچنین زمان اجرای الگوریتم‌ها در این جدول گزارش شده است. همان‌طور که مشاهده می‌شود، الگوریتم‌های مختلف از نظر زمان اجرا تفاوت زیادی با یکدیگر ندارند. با این وجود مدل جنگل تصادفی و SVM زمان اجرای مناسب‌تری داشته‌اند. همچنین از نظر میزان خطا در طبقه‌بندی، بیشترین میزان خطا مربوط به مدل K-نزدیک‌ترین همسایه و کمترین میزان خطا مربوط به مدل ماشین بردار پشتیبان می‌باشد. بنابراین از دیدگاه میزان خطا هم مدل ماشین بردار پشتیبان عملکرد بهتری نسبت به سایر مدل‌ها دارد.

5. نتیجه‌گیری

در این پژوهش از روش‌های K-نزدیک‌ترین همسایه، درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبان به منظور پیش‌بینی پیشرفت تحصیلی در دانش‌آموزان مقطع متوسطه اول استان خراسان جنوبی استفاده شده است. نتایج اعتباریابی مدل‌های هوش مصنوعی نشان داد که مدل ماشین بردار پشتیبان، عملکرد بهتری از نظر دقت، صحت، حساسیت، تشخیص پذیری، معیار

ترکیبی و خطای طبقه‌بندی نسبت به سایر مدل‌ها دارد. نتایج سودمندی را می‌توان از پیاده‌سازی این پژوهش به دست آورد. زمانی که دانش‌آموزانی با پیش‌بینی افت تحصیلی در سال آتی شناسایی می‌شوند، می‌توان قبل از اینکه حتی این دانش‌آموزان از این وضعیت مطلع باشند، همزمان با مدیریت اضطراب ناشی از افت تحصیلی، با دادن مشاوره‌های تحصیلی هوشمند و مشوق‌های انگیزشی مناسب و همچنین انجام برنامه‌ریزی‌های تحصیلی کارآمد و سایر اقدامات پیشگیرانه از رسیدن دانش‌آموزان به وضعیت بحرانی و افت آنها پیش‌گیری نمود. می‌توان مدل ساخته شده را به عنوان یک سیستم پشتیبان تصمیم‌گیری در مدیریت آموزشی مورد بهره‌برداری قرار داد. سیستم آموزشی از این طریق می‌تواند اثر بخشی سیستم خود را افزایش دهد، نرخ مردودی و افت دانش‌آموزان را کاهش داده، نرخ فراغت از تحصیل را ارتقاء، موفقیت دانش‌آموزان را افزایش دهد و هزینه‌ها را به ازای هر دانش‌آموز کاهش دهد.

موارد ذیل به عنوان جهت‌گیری‌های آتی در تحقیقات آینده مدنظر می‌باشد: پیشنهاد می‌شود تحقیقات مشابهی در جوامع آماری متفاوت و نیز در سایر مناطق کشور ایران انجام و نتایج مقایسه گردد. همچنین پیشنهاد می‌شود در تحقیقات جدید، از ابزارهای اندازه‌گیری بروز و مصاحبه‌های تشخیصی برای بررسی عوامل مؤثر بر پیشرفت تحصیلی دانش‌آموزان استفاده شود. علاوه بر این، مدل‌های مجزا و مبتنی بر جنسیت در

پیش‌بینی پیشرفت تحصیلی دانش‌آموزان می‌تواند مفید باشد. اداره آموزش و پرورش استان خراسان جنوبی کمال تشکر و قدردانی را داشته باشد.

سپاسگزاری

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی ندارند.

این مقاله برگرفته از طرح پژوهشی مرتبط و مورد حمایت اداره کل آموزش و پرورش استان خراسان جنوبی می‌باشد. بدین وسیله نویسنده مقاله بر خود لازم می‌داند از تمامی همکاران در

مراجع

- [1] X. Chen, D. Zou, H. Xie, G. Cheng, and C. Liu, "Two decades of artificial intelligence in education," *Educ. Technol. Soc.*, vol. 25, no. 1, pp. 28-47, 2022.
- [2] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, 2021.
- [3] E. Alyahyan and D. Dustegor, "Predicting academic success in higher education: Literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, pp. 1-21, 2020.
- [4] S. Hussain and M.Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," *Ann. Data Sci.*, vol. 10, no. 3, pp. 637-655, 2023.
- [5] X. Bai, F. Zhang, J. Li, T. Guo, A. Aziz, A. Jin, and F. Xia, "Educational big data: Predictions, applications and challenges," *Big Data Res.*, vol. 26, p. 100270, 2021.
- [6] C. Huang, J. Zhou, J. Chen, J. Yang, K. Clawson, and Y. Peng, "A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction," *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11517-11529, 2023.
- [7] A. Rivas, A. Gonzalez-Briones, G. Hernandez, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, vol. 423, pp. 713-720, 2021.
- [8] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Comput. Electr. Eng.*, vol. 89, p. 106903, 2021.
- [9] R. Campagni, D. Merlini, R. Sprugnoli, and M.C. Verri, "Data mining models for student careers," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5508-5521, 2015.
- [10] S. Natek and M. Zwilling, "Student data mining solution-knowledge management system related to higher education institutions," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6400-6407, 2014.
- [11] A.M. Shahiri, W. Husain, and N.A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414-422, 2015.
- [12] J. Berens, S. Oster, K. Schneider, and J. Burghoff, "Early detection of students at risk - Predicting student dropouts using administrative student data and machine learning methods," *Schumpeter Sch. Bus. Econ.*, pp. 0-32, 2018.
- [13] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D.J. Murray, "Identifying key factors of student academic performance by subgroup discovery," *Int. J. Data Sci. Anal.*, vol. 7, no. 3, pp. 227-245, 2019.
- [14] G.W. Dekker, M. Pechenizkiy, and J.M. Vleeshouwers, "Predicting students drop out: A case study," *Proc. 2nd Int. Conf. Educ. Data Mining, (EDM), 2009, Cordoba, Spain, 2009*, pp. 41-50.
- [15] A. Abu, S. Mostafa, A. Emran, and K. Shaalan, *Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques*, Springer Netherlands, 2019.
- [16] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," *Appl. Artif. Intell.*, vol. 18, pp. 411-426, 2004, doi: 10.1080/08839510490442058.
- [17] J. Navamani and A. Kannammal, "Predicting performance of schools by applying data mining

- techniques on public examination results,” *Res. J. Appl. Sci. Eng. Technol.*, vol. 9, pp. 262-271, 2015.
- [18] L. Moseley and D. Mead, “Predicting who will drop out of nursing courses: A machine learning exercise,” *Nurse Educ. Today*, vol. 28, pp. 469-475, 2008, doi: 10.1016/j.nedt.2007.07.012.
- [19] A. Nandeshwar, T. Menzies, and A. Nelson, “Learning patterns of university student retention,” *Expert Syst. Appl.*, vol. 38, pp. 14984-14996, 2011, doi: 10.1016/j.eswa.2011.05.048.
- [20] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, “A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition,” *Expert Syst. Appl.*, vol. 41, pp. 321-330, 2014, doi: 10.1016/j.eswa.2013.07.046.
- [21] M. Dewan, F. Lin, D. Wen, and Kinshuk, “Predicting dropout-prone students in e-learning education system,” in *Proc. 2015 IEEE 12th Int. Conf. Ubiquitous Intell. Comput. (UIC) and Related Conferences*, Beijing, China, 2016, pp. 1735-1740, doi: 10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.315.
- [22] M. Tan and P. Shao, “Prediction of student dropout in E-learning program through the use of machine learning method,” *Int. J. Emerg. Technol. Learn.*, vol. 10, pp. 11-17, 2015, doi:10.3991/ijet.v10i1.4189.
- [23] S. Sultana, S. Khan, and M. Abbas, “Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts,” *Int. J. Electr. Eng. Educ.*, vol. 54, pp. 105-118, 2017, doi: 10.1177/0020720916688484.
- [24] Y. Chen, A. Johri, and H. Rangwala, “Running out of STEM: A comparative study across STEM majors of college students At-Risk of dropping out early,” in *Proc. 8th Int. Conf. Learn. Analytics Knowl.*, Sydney, Australia, 2018, pp. 270-279.
- [25] M. Nagy and R. Molontay, “Predicting dropout in higher education based on secondary school performance,” in *Proc. 2018 IEEE 22nd Int. Conf. Intell. Eng. Syst. (INES)*, Las Palmas de Gran Canaria, Spain, 2018, pp. 000389-000394, doi: 10.1109/INES.2018.8523888.
- [26] A. Serra, P. Perchinunno, and M. Bilancia, “Predicting student dropouts in higher education using supervised classification algorithms,” *Lect. Notes Comput. Sci.*, vol. 10962, pp. 18-33, 2018, doi: 10.1007/978-3-319-95168-3_2.
- [27] C. Gray and D. Perkins, “Utilizing early engagement and machine learning to predict student outcomes,” *Comput. Educ.*, vol. 131, pp. 22-32, 2019, doi: 10.1016/j.compedu.2018.12.006.
- [28] J. Chung and S. Lee, “Dropout early warning systems for high school students using machine learning,” *Child. Youth Serv. Rev.*, vol. 96, pp. 346-353, 2019, doi: 10.1016/j.chilyouth.2018.11.030.
- [29] M. Hussain, W. Zhu, W. Zhang, S. Abidi, and S. Ali, “Using machine learning to predict student difficulties from learning session data,” *Artif. Intell. Rev.*, vol. 52, pp. 1-27, 2018, doi: 10.1007/s10462-018-9620-8.
- [30] S. Huang and N. Fang, “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models,” *Comput. Educ.*, vol. 61, pp. 133-145, 2013, doi: 10.1016/j.compedu.2012.08.015.
- [31] A. Slim, G.L. Heileman, J. Kozlick, and C.T. Abdallah, “Predicting student success based on prior performance,” in *Proc. 2014 IEEE Symp. Comput. Intell. Data Min. (CIDM)*, Singapore, 2015, pp. 410-415, doi: 10.1109/CIDM.2014.7008697.
- [32] C. Zhao, J. Yang, J. Liang, and C. Li, “Discover learning behavior patterns to predict certification,” in *Proc. 2016 11th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Nagoya, Japan, 2016, pp. 69-73, doi: 10.1109/ICCSE.2016.7581557.
- [33] P. Chaudhury, S. Mishra, H. Tripathy, and B. Kishore, “Enhancing the capabilities of student result prediction system,” in *Proc. 2nd Int. Conf. Inf. Commun. Technol. Competitive Strategies*, Udaipur, India, 2016, doi: 10.1145/2905055.2905150.
- [34] C. Nespereira, E. Elhariri, N. El-Bendary, A. Vilas, and R. Redondo, “Machine learning based classification approach for predicting students performance in blended learning,” *Adv. Intell. Syst. Comput.*, vol. 407, pp. 47-56, 2016, doi: 10.1007/978-3-319-26690-9_5.
- [35] M. Sagar, A. Gupta, and R. Kaushal, “Performance prediction and behavioral analysis of student programming ability,” in *Proc. 2016 Int. Conf. Adv. Comput., Commun. Inf. (ICACCI)*, Jaipur, India,

- 2016, pp. 1039-1045, doi: 10.1109/ICACCI.2016.7732181.
- [36] V. Verhun, A. Batyuk, and V. Voityshyn, "Learning analysis as a tool for predicting student performance," in Proc. IEEE 13th Int. Sci. Tech. Conf. Comput. Sci. Inf. Technol. (CSIT), 2018, vol. 2, pp. 76-79, doi: 10.1109/STC-CSIT.2018.8526741.
- [37] M. Backenkohler and V. Wolf, "Student performance prediction and optimal course selection: An MDP approach," Lect. Notes Comput. Sci., vol. 10729, pp. 40-47, 2018, doi: 10.1007/978-3-319-74781-1_3.
- [38] Y.Z. Hsieh, M.C. Su, and Y.L. Jeng, "The jacobian matrix-based learning machine in student," Lect. Notes Comput. Sci., vol. 10676, pp. 469-474, 2017, doi: 10.1007/978-3-319-71084-6_55.
- [39] M. Han, M. Tong, M. Chen, J. Liu, and C. Liu, "Application of ensemble algorithm in students' performance prediction," in Proc. 6th IIAI Int. Congr. Adv. Appl. Inf. (IIAI-AAI), Hamamatsu, Japan, 2017, pp. 735-740, doi: 10.1109/IIAI-AAI.2017.73.
- [40] A. Shanthini, G. Vinodhini, and R. Chandrasekaran, "Predicting students' academic performance in the university using meta decision tree classifiers," J. Comput. Sci., vol. 14, pp. 654-662, 2018, doi: 10.3844/jcssp.2018.654.662.
- [41] C. Ma, B. Yao, F. Ge, Y. Pan, and Y. Guo, "Improving prediction of student performance based on multiple feature selection approaches," in Proc. ICEBT 2017, Toronto, ON, Canada, 2017, pp. 36-41, doi: 10.1145/3141151.3141160.
- [42] A. Tekin, "Early prediction of students' grade point averages at graduation: A data mining approach [Ogrencinin mezuniyet notunun erken tahmini: Bir veri madenciligi yaklasimi]," Eurasian J. Educ. Res., no. 54, pp. 207-226, 2014, doi: 10.14689/ejer.2014.54.12.
- [43] S. Pushpa, T. Manjunath, T. Mrunal, A. Singh, and C. Suhas, "Class result prediction using machine learning," in Proc. 2017 Int. Conf. Smart Technol. Smart Nation (SmartTechCon), Bengaluru, India, 2018, pp. 1208-1212, doi: 10.1109/SmartTechCon.2017.8358559.
- [44] E. Howard, M. Meehan, and A. Parnell, "Contrasting prediction methods for early warning systems at undergraduate level," Internet High. Educ., vol. 37, pp. 66-75, 2018, doi: 10.1016/j.iheduc.2018.02.001.
- [45] C. Villagra-Arnedo, F. Gallego-Duran, P. Compan-Rosique, F. Llorens-Largo, and R. Molina-Carmona, "Predicting academic performance from Behavioural and learning data," Int. J. Des. Nat. Ecodyn., vol. 11, no. 3, pp. 239-249, 2016, doi: 10.2495/DNE-V11-N3-239-249.
- [46] S. Sorour, K. Goda, and T. Mine, "Estimation of Student Performance by Considering Consecutive Lessons," in Proc. 4th Int. Congr. Adv. Appl. Informat., Okayama, Japan, 2016, pp. 121-126, doi: 10.1109/IIAI-AAI.2015.170.
- [47] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, "Predicting Students Performance in Educational Data Mining," in Proc. 2015 Int. Symp. Educ. Technol. (ISET), Wuhan, China, 2016, pp. 125-128, doi: 10.1109/ISET.2015.33.
- [48] S. Rana and R. Garg, "Prediction of students performance of an institute using ClassificationViaClustering and ClassificationViaRegression," Adv. Intell. Syst. Comput., vol. 508, pp. 333-343, 2017, doi: 10.1007/978-981-10-2750-5_35.
- [49] V.K. Anand, S.K. Abdul Rahiman, E. Ben George, and A.S. Huda, "Recursive clustering technique for students' performance evaluation in programming courses," in Proc. 2018 Majan Int. Conf. (MIC), Muscat, Oman, 2018, pp. 1-5, doi: 10.1109/MINTC.2018.8363153.
- [50] H. Bydzovska, "Student performance prediction using collaborative filtering methods," Lect. Notes Comput. Sci., vol. 9112, pp. 550-553, 2015, doi: 10.1007/978-3-319-19773-9_59.
- [51] H. Bydzovska, "Are collaborative filtering methods suitable for student performance prediction?" Lect. Notes Comput. Sci., vol. 9273, pp. 425-430, 2015, doi: 10.1007/978-3-319-23485-4_42.
- [52] Y. Park, "Predicting personalized student performance in computing-related majors via collaborative filtering," in Proc. 19th Annu. SIG Conf. Inf. Technol. Educ., Fort Lauderdale, FL, USA, 2018, p. 151, doi: 10.1145/3241815.3241875.
- [53] C.H. Liou, "Personalized article recommendation based on student's rating mechanism in an online discussion forum," in Proc. 2016 49th Hawaii Int. Conf. Syst. Sci. (HICSS), Koloa, HI, USA, 2016, pp. 60-65, doi: 10.1109/HICSS.2016.16.

- [54] A. Elbadrawy and G. Karypis, "Domain-aware grade prediction and top-n course recommendation," in Proc. 10th ACM Conf. Recommender Syst., Boston, MA, USA, 2016, pp. 183-190, doi: 10.1145/2959100.2959133.
- [55] V. Pero and T. Horvath, "Comparison of collaborative-filtering techniques for small-scale student performance prediction task," Lect. Notes Electr. Eng., vol. 313, pp. 111-116, 2015, doi: 10.1007/978-3-319-06773-5_16.
- [56] Y. Song, Y. Jin, X. Zheng, H. Han, Y. Zhong, and X. Zhao, "PSFK: A student performance prediction scheme for first-encounter knowledge in ITS," Lect. Notes Comput. Sci., vol. 9403, pp. 639-650, 2015, doi: 10.1007/978-3-319-25159-2_58.
- [57] K. Xu, R. Liu, Y. Sun, K. Zou, Y. Huang, and X. Zhang, "Improve the prediction of student performance with hint's assistance based on an efficient non-negative factorization," IEICE Trans. Inf. Syst., vol. E100D, no. 4, pp. 768-775, 2017, doi: 10.1587/transinf.2016DAP0026.
- [58] M. Sheehan and Y. Park, "pGPA: A personalized grade prediction tool to aid student success," in Proc. 6th ACM Conf. Recommender Syst., Dublin, Ireland, 2012, pp. 309-310, doi: 10.1145/2365952.2366027.
- [59] S. Lorenzen, N. Pham, and S. Alstrup, "On predicting student performance using low-rank matrix factorization techniques," in Proc. 9th Eur. Conf. e-Learning (ECEL 2010), Porto, Portugal, 2010, pp. 326-334.
- [60] M. Houbraken, C. Sun, E. Smirnov, and K. Driessens, "Discovering hidden course requirements and student competences from grade data," in Proc. UMAP'17: Adjunct Publ. 25th Conf. User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 2017, pp. 147-152, doi: 10.1145/3099023.3099034.
- [61] J. Gomez-Pulido, E. Cortes-Toro, A. Duran-Dominguez, B. Crawford, and R. Soto, "Novel and Classic Metaheuristics for Tuning a Recommender System for Predicting Student Performance in Online Campus," Lect. Notes Comput. Sci., vol. 11314, pp. 125-133, 2018, doi: 10.1007/978-3-030-03493-1_14.
- [62] O. Chavarriaga, B. Florian-Gaviria, and O. Solarte, "A recommender system for students based on social knowledge and assessment data of competences," Lect. Notes Comput. Sci., vol. 8719, pp. 56-69, 2014, doi: 10.1007/978-3-319-11200-8_5.
- [63] E. Jembere, R. Rawatlal, and A. Pillay, "Matrix Factorisation for Predicting Student Performance," in Proc. 2017 7th World Eng. Educ. Forum (WEEF), Kuala Lumpur, Malaysia, 2018, pp. 513-518, doi: 10.1109/WEEF.2017.8467150.
- [64] M. Sweeney, J. Lester, and H. Rangwala, "Next-term student grade prediction," in Proc. 2015 IEEE Int. Conf. Big Data, Santa Clara, CA, USA, 2015, pp. 970-975, doi: 10.1109/BigData.2015.7363847.
- [65] J. Adan-Coello and C. Tobar, "Using collaborative filtering algorithms for predicting student performance," Lect. Notes Comput. Sci., vol. 9831, pp. 206-218, 2016, doi: 10.1007/978-3-319-44159-7_15.
- [66] L. Rechkoski, V. Ajanovski, and M. Mihova, "Evaluation of grade prediction using model-based collaborative filtering methods," in Proc. 2018 IEEE Global Eng. Educ. Conf. (EDUCON), Santa Cruz de Tenerife, Spain, 2018, pp. 1096-1103, doi: 10.1109/EDUCON.2018.8363352.
- [67] M. Adewale Amoo, A. Olumuyiwa, and U. Lateef, "Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach," Int. J. Sci. Technol. Educ. Res., vol. 9, pp. 1-8, 2018, doi: 10.5897/IJSTER2017.0415.
- [68] T. Gedeon and H. Turner, "Explaining student grades predicted by a neural network," in Proc. 1993 Int. Conf. Neural Netw., Nagoya, Japan, 1993, vol. 1, pp. 609-612.
- [69] P.M. Arsad, N. Buniyamin, and J.A. Manan, "A neural network students' performance prediction model (NNSPPM)," in Proc. 2013 IEEE Int. Conf. Smart Instrum., Meas. Anal. (ICSIMA), Kuala Lumpur, Malaysia, 2013, pp. 1-5, doi: 10.1109/ICSIMA.2013.6717966.
- [70] A. Iyanda, O.D. Ninan, A. Ajayi, and O.G. Anyabolu, "Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models," Int. J. Mod. Educ. Comput. Sci., vol. 10, pp. 1-9, 2018, doi: 10.5815/ijmecs.2018.06.01.
- [71] P. Dharmasaroja and N. Kingkaew, "Application of artificial neural networks for prediction of learning performances," in Proc. 2016 12th Int. Conf. Nat.

- Comput., Fuzzy Syst. Knowl. Discov. (ICNC-FSKD), Changsha, China, 2016, pp. 745-751, doi: 10.1109/FSKD.2016.7603268.
- [72] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (SVM) in LibSVM," *Int. J. Comput. Appl.*, vol. 128, no. 3, pp. 28-34, 2015.
- [73] M. Musso, E. Kyndt, E. Cascallar, and F. Dochy, "Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks," *Frontline Learn. Res.*, vol. 1, pp. 42-71, 2013, doi: 10.14786/flr.v1i1.13.
- [74] E. Mala Sari Rochman, A. Rachmad, and F. Damayanti, "Predicting the Final result of Student National Test with Extreme Learning Machine," *Pancar. Pendidik.*, vol. 7, no. 1, 2018, doi: 10.25037/pancaran.v7i1.159.
- [75] W. Villegas-Ch, S. Lujan-Mora, D. Buenano-Fernandez, and M. Roman-Canizares, "Analysis of web-based learning systems by data mining," in *Proc. 2017 IEEE 2nd Ecuador Tech. Chapters Meeting (ETCM)*, Guayas, Ecuador, 2018, pp. 1-5, doi: 10.1109/ETCM.2017.8247553.
- [76] I. Karagiannis and M. Satratzemi, "An adaptive mechanism for Moodle based on automatic detection of learning styles," *Educ. Inf. Technol.*, vol. 23, pp. 1331-1357, 2018, doi: 10.1007/s10639-017-9663-5.
- [77] C. Masci, G. Johnes, and T. Agasisti, "Student and school performance across countries: A machine learning approach," *Eur. J. Oper. Res.*, vol. 269, pp. 1072-1085, 2018, doi: 10.1016/j.ejor.2018.02.031.
- [78] W. Johnson, "Data mining and machine learning in education with focus in undergraduate cs student success," in *Proc. 2018 ACM Conf. Int. Comput. Educ. Res., Espoo, Finland*, 2018, pp. 270-271, doi: 10.1145/3230977.3231012.
- [79] D. Liu, D. Richards, C. Froissard, and A. Atif, "Validating the effectiveness of the moodle engagement analytics plugin to predict student academic performance," in *Proc. 21st Americas Conf. Inf. Syst. (AMCIS 2015)*, Fajardo, Puerto Rico, 2015.
- [80] D. Delen and H. M. Zolbanin, "The analytics paradigm in business research," *J. Bus. Res.*, vol. 90, pp. 186-195, 2018.
- [81] J. Bao, S. Wei, J. Lv, and W. Zhang, "Optimized faster-RCNN in real-time facial expression classification," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 790, no. 1, p. 012148, Mar. 2020.
- [82] H.M. Sani, C. Lei, and D. Neagu, "Computational complexity analysis of decision tree algorithms," in *Artificial Intelligence XXXV: Proc. 38th SGAI Int. Conf. Artif. Intell. (AI 2018)*, Cambridge, UK, 2018, pp. 191-197.
- [83] H. Alizadeh, B. Minaei-Bidgoli, and S.K. Amirgholipour, "A New Method for Improving the Performance of K Nearest Neighbor using Clustering Technique," *J. Convergence Inf. Technol.*, vol. 4, no. 2, pp. 84-92, 2009.
- [84] M.M. AlyanNezhadi, M. Hosseini, H. Qazanfari, and A. Kamandi, "Content-based image retrieval using support vector machine and texture difference histogram features," *Soft Comput. J.*, vol. 11, no. 1, pp. 10-21, 2022, doi: 10.22052/scj.2022.246175.1053 [In Persian].
- [85] M. Ebtia, S. M. Hoseini, and R. Khochiani, "Credit rating of bank customers using a new ensemble method based on support vector machine: a case study of Pasargad bank," *Soft Comput. J.*, vol. 10, no. 2, pp. 2-15, 2022, doi: 10.22052/scj.2022.243227.1016 [In Persian].
- [86] A. Khosravi, H. Abdulmaleki, and M. Fayazi, "Predicting the academic status of admitted applicants based on educational and admission data using data mining techniques," *Soft Comput. J.*, vol. 9, no. 2, pp. 94-113, 2021, doi: 10.22052/scj.2021.242837.0 [In Persian].