مقایسه عملکرد روش های هوش مصنوعی در پیش‌بینی پیشرفت تحصیلی دانش‌آموزان

ملیحه نیک سیرت[1*]، استادیار، جواد طیبی[2]، دانشیار، محمدمهدی ایزدخواه[3]، استادیار

دانشکده مهندسی کامپیوتر و صنایع ـ دانشگاه صنعتی بیرجندـ بیرجندـ ایران ـ

چکیده: امروزه، روش های هوش مصنوعی می‌تواند از طریق کشف ارتباط بین پارامترهای 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 روش های ماشین بردار پشتیان، K-نزدیک‌ترین همسایه، درخت تصمیم و جنگل تصادفی به منظور پیش‌بینی پیشرفت تحصیلی به کار 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 با سایر مدل‌ها بهترین عملکرد را در سنجش پیشرفت تحصیلی دانش‌آموزان داشته است.

واژه‌های کلیدی: پیشرفت تحصیلی، هوش مصنوعی، ماشین بردار پشتیبان، K-نزدیک‌ترین همسایه، درخت‌های تصمیم، جنگل تصادفی، اعتباریابی.

* ملیحه نیک سیرت، **niksirat@birjandut.ac.ir**

پذیرفته شده در مجله محاسبات نرم

© پیش نویس مجله علمی محاسبات نرم، برای داوری در مجله

Comparing the performance of artificial intelligence methods in predicting students' academic progress

Malihe Niksirat [1*], Assistant Prof., Javad Tayyebi [2], Associate Prof., Mohammad Mahdi Izadkhah [3], Assistant Prof.

Dept. of Industrial and Computer Engineering, Birjand University of Technology, Birjand, Iran.

Abstract: Todays, artificial intelligence methods can extract the knowledge hidden in the educational datasets by discovering the relationship between different features. This knowledge can help educational systems in making better decisions and having more advanced plans to improve the academic performance of students. The aim of this study was to identify the factors affecting the academic progress of students and to use a technique that can predict the academic progress of students with the highest percentage of accuracy. Accordingly, artificial intelligence methods including Support Vector Machine(SVM), K-Nearest Neighbor(KNN), Decision Trees(DT) and Random Forest(RF) have been applied. Finally, the models was evaluated using Accuracy, Precision, F-measure, Sensitivity, Specificity and Classification error. The results showed that the Support vector machine had the best performance in predicting the academic progress of students.

Keywords: Academic achievement; Artificial Intelligence; Support vector machine; K-nearest neighbor; Decision trees; Random Forest; Validation.

*** Malihe Niksirat, niksirat@birjandut.ac.ir.**

پذیرفته شده در مجله محاسبات نرم

۱. مقدمه

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] و[2]).

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] و [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. ارائه راهکارهایی برای بهبود وضعیت تحصیلی این دانش‌آموزان است از اهداف پژوهش‌های اخیر است. تکنیک‌های هوش مصنوعی[1] می‌توانند پاسخ علمی برای این مسئله ارائه نمایند([6] و [7]). 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.

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

[1] Artificial intelligence

ناتک و زویلینگ[۲] در پژوهشی با هدف استخراج و مدیریت داده‌های آموزشی موسسات آموزش عالی آورده 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 و آنالیز داده‌ها مورد استفاده قرار می‌گیرد. درصد موفقیت دانشجویان یک موسسه نشان‌دهنده اعتبار آن موسسه و بسیار ارزشمند است[10].

شهیری و همکاران در پژوهشی با هدف کاربرد تکنیک‌های داده‌کاوی در پیش‌بینی عملکرد دانشجویان در کشور مالزی بیان می‌کند راهکارهای متفاوتی به منظور سنجش عملکرد تحصیلی دانشجویان وجود دارد که داده‌کاوی آموزشی یکی از محبوب-ترین و موثرترین روش‌های ارزیابی عملکرد دانشجویان است که جهت 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].

برنز[۳] و همکاران در مقاله‌ای با هدف شناسایی زود هنگام خطر در دانشجویان و پیش‌بینی و پیشگیری از افت تحصیلی با استفاده از داده‌های 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 به-روزرسانی 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].

هلال[۴] 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 غیر قانونی 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 معمولاً 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 عمران[۵] 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 پیش بینی پیشرفت تحصیلی در جدول زیر

---

[۲] Natek and Zwilling

[۳] Berens

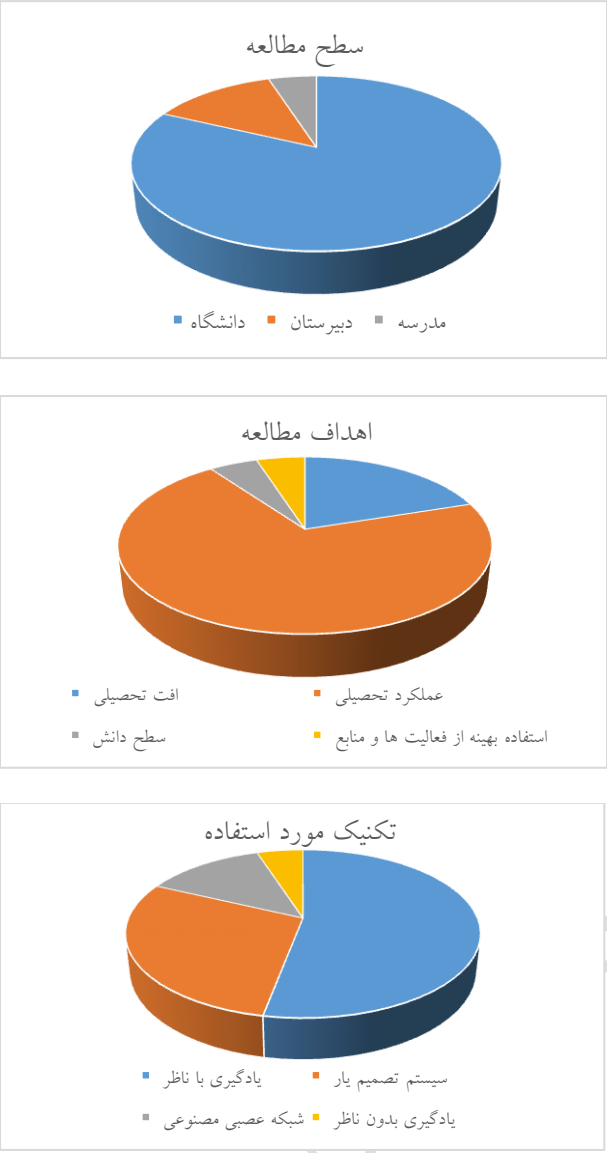
[۴] Helal

[۵] Emran

گردآوری و خلاصه شده است. در این جدول مولفه‌های (ستون-ها) زیر در نظر گرفته شده است.

- سطح دانش‌آموزان: هر مرجع، مجموعه داده‌های دانش‌آموزان 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 دانش-آموزان حاکی از عوامل پنهانی است که 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.
- تکنیک ها: تکنیک ها 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 ۱ به صورت گرافیکی آمار قابل توجهی در مورد تکنیک-ها، اهداف، سطح مطالعه و الگوریتم‌های در نظر گرفته شده در بررسی ادبیات را نشان می‌دهد. این نمودارها از جدول ۱ به منظور درک بهتر تأثیر مرور ادبیات ساخته شده‌اند.



شکل ۱: خلاصه سازی ادبیات موجود مساله تحقیق

جدول ۱: بررسی ادبیات موجود مسئله تحقیق

	سطح مطالعه	هدف	تکنیک	الگوریتم و روش حل
[16]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	درخت تصمیم[۶]، رگرسیون لجستیک[۷]، بیز ساده[۸]، پیش نمونه برداری اقلیت ترکیبی (SMOTE[۹])
[17]	دبیرستان	عملکرد تحصیلی	یادگیری با ناظر	K–نزدیکترین همسایه[۱۰]، 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[۱۱]
[20]	دانشگاه	افت تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی، درخت تصمیم، رگرسیون خطی[۱۲]، ماشین بردار پشتیبان[۱۳]
[21]	دانشگاه	افت تحصیلی	یادگیری با ناظر	الگوریتم ترکیبی دسته بندی، K–نزدیکترین همسایه، شبکه های RBN[۱۴]، ماشین بردار پشتیبان
[22]	دانشگاه	افت تحصیلی	یادگیری با ناظر	شبکه عصبی مصنوعی[۱۵]، 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[۱۶]، درخت تصمیم، رگرسیون لجستیک، بیز ساده، جنگل تصادفی، تحلیل بقا[۱۷]

[۶] Decision Tree

[۷] Logistic regression

[۸] Naïve Bayes

[۹] Synthetic Minority Over-Sampling

[۱۰] K-nearest neighbor

[۱۱] Radial Basis Networks

[۱۲] Linear regression

[۱۳] Support vector machine

[۱۴] Radial Basis Networks

[۱۵] Artificial neural network

پذیرفته شده در مجله محاسبات نرم

[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[۱۸]
[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
[31]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	شبکه های باور بیزی[۱۹]
[32]	دانشگاه	عملکرد تحصیلی	یادگیری با ناظر	تخصیص پنهان دریکله (LDA[۲۰])، 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[۲۱]، 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

[۱۶] AdaBoost

[۱۷] Survival Analysis

[18] Sequential Forward Selection

[19] Bayesian Belief Network

[20] *Latent Dirichlet Allocation*

[21] Gradient Boosted Tree

پذیرفته شده در مجله محاسبات نرم

[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[۲۲]، 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 دریکله، 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،
[50]، [44]، [54]، [47]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی[۲۳]
[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[۲۴]، فیلترینگ مشارکتی
[57]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	تجزیه ماتریس
[58]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	سیستم پیش بینی نمره[۲۵]

[22] Extreme Learning Machine

[23] *collaborative filtering*

[24] Bayesian Knowledge Tracing

[25] Grade Prediction Advisor

[59]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	فیلترینگ مشارکتی
[60]	دانشگاه	عملکرد تحصیلی	سیستم تصمیم یار	الگوریتم بیشینه سازی امید[۲۶]، تجزیه ماتریس، تجزیه مقادیر منفرد[۲۷]
[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[۲۸] ، الگوریتم شیب اول دوقطبی[۲۹]، فیلترینگ مشارکتی، K–نزدیکترین همسایه، تجزیه ماتریس، الگوریتم شیب یک
[65]	دانشگاه	عملکرد تحصیلی	شبکه عصبی مصنوعی	تجزیه ماتریس
[66]، [59]، [60]، [61]، [63]	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
[74]، [66]، [68]	دانشگاه	سطح دانش	داده کاوی	داده کاوی
[76]	مدرسه	سطح دانش	یادگیری با ناظر	درخت تصمیم، درخت گرادیان تقویت شده
[78]، [70]	دانشگاه	افت تحصیلی	داده کاوی	داده کاوی

پذیرفته شده در مجله محاسبات نرم

[26] Expectation-Maximization

[27] Singular Value Decomposition

[28] Biased-Matrix Factorization

[29] Bipolar Slope One

یکی از نکات مهم، زمانی که مطالعات قبلی با تمرکز بر روش-های 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 اساسا 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 گرفته اند[80]. یک رویکرد جایگزین برای تحقیقات سنتی مبتنی بر نظرسنجی، یک استراتژی مبتنی بر داده است که از حجم وسیعی از داده‌ها که معمولاً در پایگاه‌های داده سازمانی موجود است، استفاده می‌کند و از تکنیک‌های یادگیری ماشین برای استخراج بینش 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.

۳. 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 مؤثر بر پیشرفت تحصیلی از روش تحلیل سیستماتیک با استفاده از نرم‌افزار مکس کیودا[30] 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 پرسش‌نامه با ۳۹ سؤال و ۴ مؤلفه شامل: عوامل فردی (سؤالات؛ ۱ تا ۱۵)، خانواده (سؤالات؛ ۱۶ تا ۲۷)، محیط و متغیرهای برون فردی (سؤالات؛ ۲۸ تا ۳۰)، عوامل درون مدرسه‌ای و سیستم آموزش‌وپرورش (سؤالات؛ ۳۱ تا ۳۹) است. نمره‌گذاری پرسش‌نامه بر اساس طیف لیکرت ۵ درجه‌ای (خیلی زیاد=۵؛ خیلی کم= ۱) می‌باشد. پایایی و روایی پرسشنامه توسط آزمون-های 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 ۴۱۴۵۲ نفر می‌باشد. حجم نمونه پژوهش بر اساس جدول مورگان ۳۸۴ 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 رپیدماینر[31] پیاده‌سازی شده است. همچنین جهت اعتباریابی مدل از شاخص‌های دقت، AUC، صحت، معیار ترکیبی، حساسیت، تشخیص‌پذیری و خطای طبقه‌بندی استفاده شده است[86].

۴. نتایج و یافته‌ها

[30] Maxqda

[31] RapidMiner

در این بخش ابتدا اطلاعات جمعیت‌شناختی نمونه مانند سن جنسیت و رشته‌ی 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 پشتیان و مدل K–نزدیک‌ترین همسایه به منظور سنجش پیشرفت تحصیلی دانش‌آموزان طراحی می‌شود.

۱،۴. اطلاعات جمعیت شناختی

نتایج آماری نشان داد که ۶۸/۸ درصد افراد شرکت‌کننده در تحقیق زن و ۳۱/۲ درصد از آن‌ها مرد هستند. همچنین میانگین سن افراد مورد پژوهش ۱۳/۷۴ با انحراف معیار ۱/۰۴ می‌باشد (جدول ۲).

همچنین بر اساس جدول ۳ می‌توان گفت اکثر افراد علاقه‌مند به رشته‌ی علوم تجربی هستند و فقط ۱۰/۴ درصد به رشته‌ی ریاضی و فیزیک علاقه دارند.

۲،۴. 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 مؤثر بر پیشرفت تحصیلی در جدول ۴ آمده است. نتایج نشان می دهد که بعد محیط و متغیرهای برون فردی دارای بیشترین میانگین (۳/۷۹) و بعد عوامل فردی دارای کمترین میانگین (۳/۴۷) هستند. سه شاخص تمرکز میانگین، مد و میانه برای این متغیرها به هم نزدیک بوده و از طرفی کشیدگی و چولگی نیز بین ۲- تا ۲+ می‌باشد که نشان‌دهنده تقارن نسبی داده‌ها است.

۴،۳. مدل K–نزدیک‌ترین همسایه

مدل K–نزدیک‌ترین همسایه که به‌اختصار KNN نامیده می‌شود برای تخمین خروجی داده جدید (تست) از K تا نزدیک‌ترین همسایهٔ نمونه تست 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 ۳ تا ۱۵ تغییر داده شده است و دقت الگوریتم بررسی شده است. مشاهدات نشان داد که بیشترین دقت زمانی به دست می‌آید که مقدار K برابر ۷ باشد. قابل ذکر است که برای مقادیر K بزرگتر از ۱۵، 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.

برای مقدار K=7، ماتریس درهم‌ریختگی برای مدل KNN به صورت جدول ۵ 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 در هم‌ریختگی نشان می‌دهد که الگوریتم ۶۶ مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین ۱۰ مورد از رکوردهایی که در

کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم ۱۸ مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و ۵ 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.

۴،۴. 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، $O(mnlog_2n)$می‌باشد که در آن n تعداد رکوردها در مجموعه داده و m تعداد ویژگی ها است[82]. جدول ۶ ماتریس درهم‌ریختگی را برای مدل درخت تصمیم نشان می‌دهد.

ماتریس در هم‌ریختگی نشان می‌دهد که الگوریتم ۶۴ مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین ۱۵ مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم ۱۳ مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و ۷ مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است.

۵،۴. 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، $O(log_2n)$ می‌باشد که در آن n تعداد رکوردها در مجموعه داده است[81]. ماتریس درهم‌ریختگی برای مدل جنگل تصادفی در جدول ۷ نمایش داده شده است.

ماتریس در هم‌ریختگی نشان می‌دهد که الگوریتم ۶۵ مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین ۱۵ مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم ۱۳ مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و ۶ مورد از رکوردهایی که متعلق به کلاس B بوده‌اند را به اشتباه در کلاس A طبقه‌بندی کرده است.

از مقایسه ماتریس‌های در هم‌ریختگی 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 ار نتیجه می‌دهد.

۶،۴. مدل ماشین بردار پشتیبان

مدل ماشین بردار پشتیبان (SVM) یکی از مدل‌های توانمند و انعطاف پذیر 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] و [85]). بنابراین هدف از مدل SVM تقسیم بندی داده‌ها در کلاس ها 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 به صورت جدول ۸ به دست آمد.

ماتریس درهم‌ریختگی نشان می‌دهد که الگوریتم ۶۵ مورد از رکوردها که در واقع در کلاس A بوده‌اند را به درستی در کلاس A طبقه‌بندی کرده است. همچنین ۱۸ مورد از رکوردهایی که در کلاس B بوده‌اند را به درستی در کلاس B طبقه‌بندی کرده است. جالب توجه است که الگوریتم ۱۰ مورد از رکوردهایی که در کلاس A بوده‌اند را به اشتباه در کلاس B طبقه‌بندی کرده است و ۶ 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.

۷،۴. اعتباریابی مدل‌ها و مقایسه عملکرد روش‌های هوش مصنوعی

مجموعاً 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 دقت[32]، صحت[33]، معیار ترکیبی[34]، حساسیت[35]، تشخیص‌پذیری[36]، خطای طبقه‌بندی[37] و AUC ارزیابی و در ادامه گزارش شده است. به منظور ارزیابی دقیق تر هر یک از مدل‌های پیشنهادی، معیارهای اشاره شده برای تمام نمونه‌های استخراج شده محاسبه شده است. جدول ۲ میانگین هر یک از معیارها را به تفکیک مدل‌های پیشنهادی گزارش می‌کند.

نتایج جدول ۹ نشان می‌دهد که مدل SVM در تقریبا تمامی معیارها (به جز معیار AUC) عملکرد بهتری نسبت به سایر روش‌ها در طبقه بندی داده های 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

---

[32] Accuracy

[33] Precision

[34] F-measure

[35] Sensetivity

[36] Specificity

[37] Classification error

تفاوت زیادی با یکدیگر ندارند. با این وجود مدل جنگل تصادفی و SVM زمان اجرای مناسب تری 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.

۵. نتیجه

در این پژوهش از روش‌های K–نزدیک‌ترین همسایه، درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبان به منظور پیش-بینی 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 پیاده سازی این پژوهش بدست آورد. زمانیکه دانش‌آموزانی با پیش‌بینی افت تحصیلی در سال آتی شناسایی می‌شوند، می‌توان قبل از اینکه حتی این دانش آموزان از این وضعیت مطلع باشند، همزمان با مدیریت اضطراب ناشی از افت تحصیلی، با دادن مشاوره‌های تحصیلی هوشمند و مشوق های انگیزشی مناسب و همچنین انجام برنامه‌ریزی‌های تحصیلی کارآمد و سایر اقدامات پیشگیرانه از رسیدن دانش-آموزان به وضعیت بحرانی و افت آنها پیش‌گیری نمود.

می توان 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 دانش آموزان را افزایش دهد و هزینه ها را به ازای هر دانش آموز 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 واقع گردد.

سپاسگزاری

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 می داند از تمامی همکاران در اداره آموزش و پرورش استان خراسان جنوبی کمال تشکر و قدردانی را داشته باشد. همچنین نویسندگان مقاله مراتب قدردانی و امتنان خود را از داوران محترم اعلام می‌دارند. بی-شک نقطه نظرات ارزشمندشان در بهبود کیفیت مقاله نقش بسزایی داشته است.

پذیرفته شده در مجله محاسبات نرم

جدول ۲: **شاخص‌های پراکندگی و مرکزی سن**

متغیر	میانگین	انحراف معیار	کمترین	بیشترین
سن	۱۳/۷۴	۱/۰۴	۱۲	۱۶

جدول ۳: توزیع فراوانی رشته مورد علاقه

رشته	تعداد	درصد
ریاضی و فیزیک	۳۹	۱۰/۴
علوم تجربی	۱۷۸	۴۷/۳
علوم انسانی	۶۶	۱۷/۶
سایر	۹۳	۲۴/۷
جمع کل	۳۷۶	۱۰۰
بدون پاسخ	۸	

جدول ۴: شاخص‌های توصیفی متغیرهای پرسشنامه

متغیر	میانگین	میانه	مد	انحراف معیار	چولگی	کشیدگی
عوامل فردی	۳/۴۷	۳/۵۳	۳/۴۰	۰/۵۲	−۰/۴۸	۰/۲۸
خانواده	۳/۷۸	۳/۸۳	۴/۱۷	۰/۴۶	−۰/۸۳	۰/۸۶
محیط و متغیرهای برون فردی	۳/۷۹	۳/۶۷	۴	۰/۶۷	−۰/۳۵	۰/۰۴
عوامل درون مدرسه ای و سیستم آموزش وپرورش	۳/۵۶	۳/۶۷	۳/۵۶	۰/۶۴	−۰/۵۷	۰/۰۱

جدول ۵: 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%	

پذیرفته شده در مجله محاسبات نرم

جدول ۶: 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%	

جدول ۷: 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%	

جدول ۸: 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%	

جدول ۹: ارزیابی و مقایسه عملکرد روش‌های یادگیری ماشین

معیار ارزیابی	KNN	درخت تصمیم	جنگل تصادفی	SVM
دقت (فاصله اطمینان ۹۵ 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
تشخیص‌پذیری	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

پذیرفته شده در مجله محاسبات نرم

مراجع

[1] Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, *25*(1), 28-47.

[2] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. Education Sciences, 11(9), 552.

[3] Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17, 1-21.

[4] Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science, 10(3), 637-655.

[5] Bai, X., Zhang, F., Li, J., Guo, T., Aziz, A., Jin, A., & Xia, F. (2021). Educational big data: Predictions, applications and challenges. Big Data Research, 26, 100270.

[6] Huang, C., Zhou, J., Chen, J., Yang, J., Clawson, K., & Peng, Y. (2023). A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction. Neural Computing and Applications, 35(16), 11517-11529.

[7] Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J., & Chamoso, P. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. Neurocomputing, 423, 713-720.

[8] Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. Computers & Electrical Engineering, 89, 106903.

[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., no. September, 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," pp. 41–50, 2009.

[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, no. 0123456789. Springer Netherlands, 2019.

[16] Kotsiantis, S.; Pierrakeas, C.; Pintelas, P. Predicting Students' Performance in Distance Learning using Machine Learning Techniques. Appl. Artif. Intell. 2004, 18, 411–426, doi:10.1080/08839510490442058.

[17] Navamani, J.; Kannammal, A. Predicting performance of schools by applying data mining techniques on public examination results. Res. J. Appl. Sci. Eng. Technol. 2015, 9, 262–271.

[18] Moseley, L.; Mead, D. Predicting who will drop out of nursing courses: A machine learning exercise. Nurse Educ. Today 2008, 28, 469–475, doi:10.1016/j.nedt.2007.07.012.

[19] Nandeshwar, A.; Menzies, T.; Nelson, A. Learning patterns of university student retention. Expert Syst. Appl. 2011, 38, 14984–14996, doi:10.1016/j.eswa.2011.05.048.

[20] Thammasiri, D.; Delen, D.; Meesad, P.; Kasap, N. A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. Expert Syst. Appl. 2014, 41, 321–330, doi:10.1016/j.eswa.2013.07.046.

from learning session data. Artif. Intell. Rev. 2018, 52, 1–27, doi:10.1007/s10462-018-9620-8.

[30] Huang, S.; Fang, N. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Comput. Educ. 2013, 61, 133–145, doi:10.1016/j.compedu.2012.08.015.

[31] Slim, A.; Heileman, G.L.; Kozlick, J.; Abdallah, C.T. Predicting student success based on prior performance. In Proceedings of the 2014 IEEE Symposiumon Computational Intelligence and DataMining (CIDM), Singapore, 16 April 2015; pp. 410–415, doi:10.1109/CIDM.2014.7008697.

[32] Zhao, C.; Yang, J.; Liang, J.; Li, C. Discover learning behavior patterns to predict certification. In Proceedings of the 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23 August 2016; pp. 69–73, doi:10.1109/ICCSE.2016.7581557.

[33] Chaudhury, P.; Mishra, S.; Tripathy, H.; Kishore, B. Enhancing the capabilities of student result prediction system. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, Udaipur, India, March 2016; doi:10.1145/2905055.2905150.

[34] Nespereira, C.; Elhariri, E.; El-Bendary, N.; Vilas, A.; Redondo, R. Machine learning based classification approach for predicting students performance in blended learning. Adv. Intell. Syst. Comput. 2016, 407, 47–56, doi:10.1007/978-3-319-26690-9_5.

[35] Sagar, M.; Gupta, A.; Kaushal, R. Performance prediction and behavioral analysis of student programming ability. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21 September 2016; pp. 1039–1045, doi:10.1109/ICACCI.2016.7732181.

[36] Verhun, V.; Batyuk, A.; Voityshyn, V. Learning Analysis as a Tool for Predicting Student Performance. In Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 11 September 2018; Volume 2, pp. 76–79, doi:10.1109/STC-CSIT.2018.8526741.

[37] Backenköhler, M.; Wolf, V. Student performance prediction and optimal course selection: An MDP

[21] Dewan, M.; Lin, F.;Wen, D.; Kinshuk. Predicting dropout-prone students in e-learning education system. In Proceedings of the 2015 IEEE 12th Intl Conference on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conference on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conference on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, China, 14 August 2016; pp. 1735–1740, doi:10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.315.

[22] Tan, M.; Shao, P. Prediction of student dropout in E-learning program through the use of machine learning method. Int. J. Emerg. Technol. Learn. 2015, 10, 11–17, doi:10.3991/ijet.v10i1.4189.

[23] Sultana, S.; Khan, S.; Abbas, M. Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. Int. J. Electr. Eng. Educ. 2017, 54, 105–118, doi:10.1177/0020720916688484.

[24] Chen, Y.; Johri, A.; Rangwala, H. Running out of STEM: A comparative study across STEM majors of college students At-Risk of dropping out early. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, Australia, March 2018; pp. 270–279.

[25] Nagy, M.; Molontay, R. Predicting Dropout in Higher Education Based on Secondary School Performance. In Proceedings of the 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), Las Palmas de Gran Canaria, Spain, 21 June 2018; pp. 000389–000394, doi:10.1109/INES.2018.8523888.

[26] Serra, A.; Perchinunno, P.; Bilancia, M. Predicting student dropouts in higher education using supervised classification algorithms. Lect. Notes Comput. Sci. 2018, 10962 LNCS, 18–33, doi:10.1007/978-3-319-95168-3_2.

[27] Gray, C.; Perkins, D. Utilizing early engagement and machine learning to predict student outcomes. Comput. Educ. 2019, 131, 22–32, doi:10.1016/j.compedu.2018.12.006.

[28] Chung, J.; Lee, S. Dropout early warning systems for high school students using machine learning. Child. Youth Serv. Rev. 2019, 96, 346–353, doi:10.1016/j.childyouth.2018.11.030.

[29] Hussain, M.; Zhu,W.; Zhang,W.; Abidi, S.; Ali, S. Using machine learning to predict student difficulties

[46] Sorour, S.; Goda, K.; Mine, T. Estimation of Student Performance by Considering Consecutive Lessons. In Proceedings of the 4th International Congress on Advanced Applied Informatics, Okayama, Japan, 12 June 2016, pp. 121–126, doi:10.1109/IIAI-AAI.2015.170.

[47] Guo, B.; Zhang, R.; Xu, G.; Shi, C.; Yang, L. Predicting Students Performance in Educational Data Mining. In Proceedings of the 2015 International Symposium on Educational Technology (ISET),Wuhan, China, 24 March 2016; pp. 125–128, doi:10.1109/ISET.2015.33.

[48] Rana, S.; Garg, R. Prediction of students performance of an institute using ClassificationViaClustering and ClassificationViaRegression. Adv. Intell. Syst. Comput. 2017, 508, 333–343, doi:10.1007/978-981-10-2750-5_35.

[49] Anand, V.K.; Abdul Rahiman, S.K.; Ben George, E.; Huda, A.S. Recursive clustering technique for students' performance evaluation in programming courses. In Proceedings of the 2018 Majan International Conference (MIC), Muscat, Oman, 19 March 2018; pp. 1–5, doi:10.1109/MINTC.2018.8363153.

[50] Bydžovská, H. Student performance prediction using collaborative filtering methods. Lect. Notes Comput. Sci. 2015, 9112, 550–553. doi:10.1007/978-3-319-19773-9_59.

[51] Bydžovská, H. Are collaborative filtering methods suitable for student performance prediction? Lect. Notes Comput. Sci. 2015, 9273, 425–430, doi:10.1007/978-3-319-23485-4_42.

[52] Park, Y. Predicting personalized student performance in computing-related majors via collaborative filtering. In Proceedings of the 19th Annual SIG Conference on Information Technology Education, Fort Lauderdale, FL, USA, 3 October 2018; p. 151, doi:10.1145/3241815.3241875.

[53] Liou, C.H. Personalized article recommendation based on student's rating mechanism in an online discussion forum. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 60–65, doi:10.1109/HICSS.2016.16.

[54] Elbadrawy, A.; Karypis, G. Domain-aware grade prediction and top-n course recommendation. In Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15 approach. Lect. Notes Comput. Sci. 2018, 10729 LNCS, 40–47, doi:10.1007/978-3-319-74781-1_3.

[38] Hsieh, Y.Z.; Su, M.C.; Jeng, Y.L. The jacobian matrix-based learning machine in student. Lect. Notes Comput. Sci. 2017, 10676 LNCS, 469–474, doi:10.1007/978-3-319-71084-6_55.

[39] Han, M.; Tong, M.; Chen, M.; Liu, J.; Liu, C. Application of Ensemble Algorithm in Students' Performance Prediction. In Proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 16 November 2017; pp. 735–740, doi:10.1109/IIAI-AAI.2017.73.

[40] Shanthini, A.; Vinodhini, G.; Chandrasekaran, R. Predicting students' academic performance in the University usingmeta decision tree classifiers. J. Comput. Sci. 2018, 14, 654–662, doi:10.3844/jcssp.2018.654.662.

[41] Ma, C.; Yao, B.; Ge, F.; Pan, Y.; Guo, Y. Improving prediction of student performance based on multiple feature selection approaches. In Proceedings of the ICEBT 2017, Toronto, ON, Canada, 2017; pp. 36–41, doi:10.1145/3141151.3141160.

[42] Tekin, A. Early prediction of students' grade point averages at graduation: A data mining approach [Ö˘grencinin mezuniyet notunun erken tahmini: Bir veri madencili˘ gi yakla¸sidotlessmidotless]. Egit. Arastirmalari Eurasian J. Educ. Res. 2014, 207–226, doi:10.14689/ejer.2014.54.12.

[43] Pushpa, S.; Manjunath, T.; Mrunal, T.; Singh, A.; Suhas, C. Class result prediction using machine learning. In Proceedings of the 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bengaluru, India, 19 August 2018; pp. 1208–1212, doi:10.1109/SmartTechCon.2017.8358559.

[44] Howard, E.; Meehan, M.; Parnell, A. Contrasting prediction methods for early warning systems at undergraduate level. Internet High. Educ. 2018, 37, 66–75, doi:10.1016/j.iheduc.2018.02.001.

[45] Villagrá-Arnedo, C.; Gallego-Duran, F.; Compan-Rosique, P.; Llorens-Largo, F.; Molina-Carmona, R. Predicting academic performance from Behavioural and learning data. Int. J. Des. Nat. Ecodyn. 2016, 11, 239–249, doi:10.2495/DNE-V11-N3-239-249.

[63] Jembere, E.; Rawatlal, R.; Pillay, A. Matrix Factorisation for Predicting Student Performance. In Proceedings of the 2017 7thWorld Engineering Education Forum (WEEF), Kuala Lumpur, Malaysia, 16 November 2018; pp. 513–518, doi:10.1109/WEEF.2017.8467150.

[64] Sweeney, M.; Lester, J.; Rangwala, H. Next-term student grade prediction. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 1 November 2015; pp. 970–975, doi:10.1109/BigData.2015.7363847.

[65] Adán-Coello, J.; Tobar, C. Using collaborative filtering algorithms for predicting student performance. Lect. Notes Comput. Sci. 2016, 9831 LNCS, 206–218, doi:10.1007/978-3-319-44159-7_15.

[66] Rechkoski, L.; Ajanovski, V.; Mihova, M. Evaluation of grade prediction using model-based collaborative filtering methods. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17 April 2018; pp. 1096–1103, doi:10.1109/EDUCON.2018.8363352.

[67] Adewale Amoo, M.; Olumuyiwa, A.; Lateef, U. Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach. Int. J. Sci. Technol. Educ. Res. 2018, 9, 1–8, doi:10.5897/IJSTER2017.0415.

[68] Gedeon, T.; Turner, H. Explaining student grades predicted by a neural network. In Proceedings of 1993 International Conference on Neural Networks, Nagoya, Japan, 25 October 1993; Volume 1, pp. 609–612.

[69] Arsad, P.M.; Buniyamin, N.; Manan, J.A. A neural network students' performance prediction model (NNSPPM). In Proceedings of the 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Kuala Lumpur, Malaysia, 27 November 2013; pp. 1–5, doi:10.1109/ICSIMA.2013.6717966.

[70] Iyanda, A.; D. Ninan, O.; Ajayi, A.; G. Anyabolu, O. Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models. Int. J. Mod. Educ. Comput. Sci. 2018, 10, 1–9, doi:10.5815/ijmecs.2018.06.01.

[71] Dharmasaroja, P.; Kingkaew, N. Application of artificial neural networks for prediction of learning performances. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-September 2016, pp. 183–190, doi:10.1145/2959100.2959133.

[55] Pero, v.; Horváth, T. Comparison of collaborative-filtering techniques for small-scale student performance prediction task. Lect. Notes Electr. Eng. 2015, 313, 111–116, doi:10.1007/978-3-319-06773-5_16.

[56] Song, Y.; Jin, Y.; Zheng, X.; Han, H.; Zhong, Y.; Zhao, X. PSFK: A student performance prediction scheme for first-encounter knowledge in ITS. Lect. Notes Comput. Sci. 2015, 9403, 639–650, doi:10.1007/978-3-319-25159-2_58.

[57] Xu, K.; Liu, R.; Sun, Y.; Zou, K.; Huang, Y.; Zhang, X. Improve the prediction of student performance with hint's assistance based on an efficient non-negative factorization. IEICE Trans. Inf. Syst. 2017, E100D, 768–775, doi:10.1587/transinf.2016DAP0026.

[58] Sheehan, M.; Park, Y. pGPA: A personalized grade prediction tool to aid student success. In Proceedings of the sixth ACM conference on Recommender systems, Dublin, Ireland, 3 September 2012, pp. 309–310, doi:10.1145/2365952.2366027.

[59] Lorenzen, S.; Pham, N.; Alstrup, S. On predicting student performance using low-rank matrix factorization techniques. In Proceedings of the 9th European Conference on E-Learning—ECEL 2010 (ECEL2010), Porto, Portugal, 5 November 2010; pp. 326–334.

[60] Houbraken, M.; Sun, C.; Smirnov, E.; Driessens, K. Discovering hidden course requirements and student competences from grade data. In Proceedings of the UMAP '17: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, July 2017; pp. 147–152, doi:10.1145/3099023.3099034.

[61] Gómez-Pulido, J.; Cortés-Toro, E.; Durán-Domínguez, A.; Crawford, B.; Soto, R. Novel and Classic Metaheuristics for Tunning a Recommender System for Predicting Student Performance in Online Campus. Lect. Notes Comput. Sci. 2018, 11314 LNCS, 125–133, doi:10.1007/978-3-030-03493-1_14.

[62] Chavarriaga, O.; Florian-Gaviria, B.; Solarte, O. A recommender system for students based on social knowledge and assessment data of competences. Lect. Notes Comput. Sci. 2014, 8719 LNCS, 56–69, doi:10.1007/978-3-319-11200-8_5.

FSKD), Changsha, China, 13 August 2016; pp. 745–751, doi:10.1109/FSKD.2016.7603268.

[72] Abdiansah, A., & Wardoyo, R. (2015). Time complexity analysis of support vector machines (SVM) in LibSVM. Int. J. Comput. Appl, 128(3), 28-34.

[73] Musso, M.; Kyndt, E.; Cascallar, E.; Dochy, F. Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. Frontline Learn. Res. 2013, 1, 42–71. doi:10.14786/flr.v1i1.13.

[74] Mala Sari Rochman, E.; Rachmad, A.; Damayanti, F. Predicting the Final result of Student National Test with Extreme Learning Machine. Pancar. Pendidik. 2018, 7, doi:10.25037/pancaran.v7i1.159.

[75] Villegas-Ch, W.; Lujan-Mora, S.; Buenano-Fernandez, D.; Roman-Canizares, M. Analysis of web-based learning systems by data mining. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Guayas, Ecuador, 16 October 2018; pp. 1–5, doi:10.1109/ETCM.2017.8247553.

[76] Karagiannis, I.; Satratzemi, M. An adaptive mechanism for Moodle based on automatic detection of learning styles. Educ. Inf. Technol. 2018, 23, 1331–1357, doi:10.1007/s10639-017-9663-5.

[77] Masci, C.; Johnes, G.; Agasisti, T. Student and school performance across countries: A machine learning approach. Eur. J. Oper. Res. 2018, 269, 1072–1085, doi:10.1016/j.ejor.2018.02.031.

[78] Johnson, W. Data mining and machine learning in education with focus in undergraduate cs student success. In Proceedings of the 2018 ACM Conference on International Computing Education Research, Espoo, Finland, August 2018; pp. 270–271, doi:10.1145/3230977.3231012.

[79] Liu, D.; Richards, D.; Froissard, C.; Atif, A. Validating the effectiveness of the moodle engagement analytics plugin to predict student academic performance. In Proceedings of the 21st Americas Conference on Information Systems (AMCIS 2015), Fajardo, Puerto Rico, 13 August 2015.

[80] Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. *Journal of Business Research*, *90*, 186-195.

[81] Bao, J., Wei, S., Lv, J., & Zhang, W. (2020, March). Optimized faster-RCNN in real-time facial expression classification. In IOP Conference Series: Materials Science and Engineering (Vol. 790, No. 1, p. 012148). IOP Publishing.

[82] Sani, H. M., Lei, C., & Neagu, D. (2018). Computational complexity analysis of decision tree algorithms. In *Artificial Intelligence XXXV: 38th SGAI International Conference on Artificial Intelligence, AI 2018, Cambridge, UK, December 11–13, 2018, Proceedings 38* (pp. 191-197). Springer International Publishing.

[83] Alizadeh, H., Minaei-Bidgoli, B., & Amirgholipour, S. K. (2009). A New Method for Improving the Performance of K Nearest Neighbor using Clustering Technique. *J. Convergence Inf. Technol.*, *4*(2), 84-92.

[84] AlyanNezhadi, M. M., Hosseini, M., Qazanfari, H., & Kamandi, A. (2022). Content-based image retrieval using support vector machine and texture difference histogram features. Soft Computing Journal, 11(1), 10-21.

[85] Ebtia, M., Hoseini, S. M., & Khochiani, R. (2022). Credit rating of bank customers using a new ensemble method based on support vector machine: a case study of Pasargad bank. Soft Computing Journal, 10(2), 2-15.

[86] Khosravi, A., Abdulmaleki, H., & Fayazi, M. (2021). Predicting the academic status of admitted applicants based on educational and admission data using data mining techniques. Soft Computing Journal, 9(2), 94-113.