journal homepage: www.sjfm.ir

Iranian Journal of Forensic Medicine

Iran J Forensic Med. 2025 Dec;31(3):184-193 Volume 31, Issue 3, Autumn 2025 Research Article

A Machine Learning Model for Investigating the Causes Influencing Student Dropout and Predicting Student Performance

Kiarash Zohori[1*] **PharmD, Marjan Fallah**[1] **PharmD**

[1] Department of Toxicology and Pharmacology, Faculty of Pharmacy, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran

***Correspondence to:** Kiarash Zohori, Email: kiarahzohori@yahoo.com

ARTICLE INFO

Article history:
Received: May 22, 2025
Accepted: September 15, 2025
Online Published: October 14, 2025

Keywords:
Machine learning
Metaheuristics
Student dropout
Decision-making

HIGHLIGHTS

1. Resources should be effectively allocated to support students to complete their education and prevent dropout.
2. With increasing age, dropout rates increase for both sexes, but women consistently have lower dropout rates across all age groups.

ABSTRACT

Introduction: Early identification of students at risk of dropping out is essential for providing timely support, improving retention rates, and promoting academic success. Machine learning offers a powerful approach to analyze patterns in student behavior, allowing universities to predict dropouts and implement preventive interventions.

Methods: A hybrid model combining Particle Swarm Optimization and Extreme Gradient Boosting algorithms was developed to classify at-risk students. Additional data analysis techniques were applied to examine the dataset and extract meaningful insights.

Results: The proposed model achieved an accuracy of 98.12%, outperforming alternative models. Academic variables were identified as the most influential factors in dropout decisions. Students who failed to pay tuition on time were at substantially higher risk of dropping out (87.05%) and had a lower probability of graduation (4.95%), compared to those who paid on time, who exhibited a dropout rate of 25.21% and a graduation rate of 55.13%. Dropout rates also increased with age for both genders, although women consistently showed lower dropout rates across all age groups.

Conclusion: Machine learning-based prediction of dropout risk enables institutions to target interventions more effectively, optimizing resource allocation and supporting students in completing their education.

How to cite: Zohori K, Fallah M. A Machine learning model for investigating the causes influencing student dropout and predicting student performance. Iran J Forensic Med. 2025;31(3):184-93.

[Downloaded from sjfm.ir on 2026-11-29]

Copyright © 2025 The Author(s); Published by Legal Medicine Research Center. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

www.sjfm.ir

مجله
پزشکی قانونی ایران

مرکز تحقیقات پزشکی قانونی

دوره ۳۱، شماره ۳، پاییز ۱۴۰۴، صفحات ۱۸۴ تا ۱۹۳

مقاله پژوهشی

یک مدل یادگیری ماشین برای بررسی علل موثر بر ترک تحصیل دانشجویان و پیش‌بینی عملکرد دانشجویان

کیارش ظهوری[1*] PharmD، مرجان فلاح[1] PharmD

[1] گروه سم‌شناسی و فارماکولوژی، دانشکده داروسازی، واحد آیت‌الل ه آملی، دانشگاه آزاد اسلامی، آمل، ایران

*نویسنده مسئول: کیارش ظهوری، پست الکترونیک: kiarahzohori@yahoo.com

اطلاعات مقاله

تاریخچه مقاله:
دریافت: ۱۴۰۴/۰۳/۰۱
پذیرش: ۱۴۰۴/۰۶/۲۴
انتشار برخط: ۱۴۰۴/۰۷/۲۲

واژگان کلیدی:
یادگیری ماشین
فراابتکاری
ترک‌تحصیل
تصمیم‌گیری

نکات ویژه

۱- بایستی منابع به‌طور موثر برای حمایت از دانشجویان برای تکمیل تحصیلات خود و جلوگیری از ترک تحصیل اختصاص یابد.
۲- با افزایش سن، نرخ ترک تحصیل در هر دو جنس افزایش می‌یابد، اما زنان به‌طور مداوم در تمام گروه‌های سنی نرخ ترک تحصیل کمتری دارند.

چکیده

مقدمه: شناسایی دانشجویانی که در معرض خطر ترک تحصیل هستند برای ارائه پشتیبانی به موقع، بهبود نرخ ماندگاری و تضمین موفقیت تحصیلی بسیار مهم است. یادگیری ماشینی را می‌توان برای تحلیل الگوهای رفتار دانش‌آموزان به کار برد و دانشگاه‌ها را قادر می‌سازد تا ترک تحصیل را پیش‌بینی کنند و اقدامات پیشگیرانه ای برای مداخله موثر انجام دهند.

روش بررسی: ترکیبی از الگوریتم بهینه‌سازی ازدحام ذرات و تقویت گرادیان شدید برای ایجاد یک مدل طبقه‌بندی‌کننده ارائه شد. همچنین از روش‌های تحلیل داده برای تحلیل و بررسی داده‌ها و به‌دست آوردن اطلاعات مفید استفاده شد.

یافته‌ها: مدل ارائه شده دقت ۹۸/۱۲ درصد نشان داد که دقت بالاتری نسبت به دیگر مدل‌ها دارد و همچنین نشان داده شد داده‌های آکادمیک متغیرهای موثرتری در انصراف دانشجویان از دانشگاه هستند. دانشجویانی که شهریه خود را به موقع پرداخت نکرده‌اند، به طور قابل توجهی بیشتر در معرض خطر ترک تحصیل (۸۷/۰۵ درصد) و کمتر در معرض فارغ‌التحصیلی (۴/۹۵ درصد) هستند، در مقایسه با دانشجویانی که شهریه خود را به موقع پرداخت می‌کنند که نرخ ترک تحصیل کمتری (۲۵/۲۱ درصد) و نرخ فارغ‌التحصیلی بالاتری (۵۵/۱۳ درصد) دارند. همچنین، با افزایش سن، نرخ ترک تحصیل در هر دو جنس افزایش می‌یابد، اگرچه زنان در تمام گروه‌های سنی همواره نرخ ترک تحصیل کمتری نشان می‌دهند.

نتیجه‌گیری: با به‌کارگیری یادگیری ماشینی برای پیش‌بینی خطرات ترک تحصیل، مؤسسات می‌توانند منابع خود را به‌طور موثرتری برای حمایت از دانشجویان در به‌پایان رساندن تحصیلاتشان تخصیص دهند.

مقدمه

ترک دانشگاه و تحصیل موانع جدی بر سر راه رشد اقتصادی و رفاه اجتماعی است [۱]. دانشجویانی که دانشگاه را بدون تکمیل مدرک تحصیلی خود ترک می‌کنند، با چشم‌انداز شغلی محدودی روبه‌رو می‌شوند، اغلب به طور قابل توجهی کمتر از فارغ‌التحصیلان درآمد دارند که ثبات اقتصادی طولانی مدت آنها را کاهش می‌دهد [۲]. این بی‌ثباتی مالی می‌تواند استرس بیشتری را بر خانواده‌ها وارد کند، به ویژه در مواردی که والدین در آموزش فرزندان خود سرمایه‌گذاری کرده‌اند. شکست تحصیلی در سطح دانشگاه نیز شکاف بین تحصیلکرده و بی‌سواد را افزایش می‌دهد و نابرابری را تشدید می‌کند و مانع تحرک اجتماعی می‌شود [۳]. مؤسسات از نرخ بالای ترک تحصیل رنج می‌برند، زیرا درآمد حاصل از شهریه را از دست می‌دهند و ممکن است شهرت خود را تخریب کنند و جذب دانشجویان آینده را سخت‌تر کنند. در مقیاس بزرگ‌تر، جوامع از کمک‌های بالقوه این افراد غافل می‌شوند که منجر به

[Downloaded from sjfm.ir on 2026-11-29]

نیروی کار کمتر ماهر و نوآوری کندتر می‌شود. همچنین افرادی که ترک تحصیل می‌کنند، بیشتر با بیکاری مواجه می‌شوند، به خدمات اجتماعی متکی هستند و از پیامدهای بهداشتی ضعیف‌تر رنج می‌برند [۴]. اثر موج‌دار به جوامع گسترش می‌یابد؛ جایی که نابرابری‌های آموزشی می‌تواند به ناآرامی اجتماعی، نرخ بالاتر جرم و جنایت و کاهش مشارکت مدنی کمک کند. پرداختن به نرخ ترک تحصیل از دانشگاه مستلزم سرمایه‌گذاری در خدمات حمایتی دانشجو، کمک‌های مالی، و مشاوره شغلی و همچنین سیاست‌هایی است که آموزش عالی را در دسترس‌تر و مرتبط‌تر با بازار کار امروزی می‌کند [۵]. ناکامی در مقابله با این موضوع، توسعه اقتصادی و پیشرفت اجتماعی را در درازمدت تضعیف می‌کند. در آمریکا، نرخ ترک تحصیل نشان می‌دهد که ۳۹ درصد از دانشجویانی که برای اولین بار و تمام وقت به دنبال مدرک لیسانس هستند، برنامه مدرک خود را در عرض ۸ سال کامل نمی‌کنند [۶]. با استفاده از تکنولوژی‌های جدیدی همچون هوش مصنوعی، می‌توان به بررسی عوامل موثر بر ترک تحصیل و پیش‌بینی آن برای جلوگیری و یا کاهش خسارت‌های ناشی از آن پرداخت.

گونزالس و همکاران با استفاده از روش‌های تصمیم‌گیری چندمعیاره فازی به بررسی متغیرهایی همچون رضایت از انتخاب رشته، راهبردهای خودتنظیمی و مشارکت در داخل دانشگاه پرداختند تا بتوانند علت تصمیم دانشجویان برای ترک دانشگاه را مورد مطالعه قرار دهند[۷]. کورتاپ و لایکگارد با استفاده از روش‌های چندمتغیره آماری به بررسی چگونگی ادراک دانشجویان علوم انسانی از محیط مطالعه، ملاحظات ترک تحصیل و متغیرهای زمینه و تاثیر آنها در ترک تحصیل پرداختند [۸]. این مطالعه نشان داد که ادراک دانش‌آموزان از محیط مطالعه‌شان بین ۱۵/۸ تا ۳۶/۹ درصد از ترک تحصیل را توضیح می‌دهد، در حالی که ملاحظات ترک تحصیل و پارامترهای پیش‌زمینه تنها بین صفر تا ۹/۱ درصد و بین ۷/۹ تا ۲۱/۴ درصد ترک تحصیل را توضیح می‌دهند [۸]. پرز و همکاران تحلیل پیش‌بینی دانش‌آموزان ترک تحصیل در کلمبیا را مورد بحث قرار دادند. متغیرهایی که بر ترک تحصیل دانش‌آموزان در کلمبیا تأثیر می‌گذارد، جمعیت‌شناسی و سوابق کارنامه دانش‌آموزان است. این متغیرها برای پیش‌بینی دانش‌آموزان ترک تحصیل استفاده می‌شوند و متغیرهای حاصل به طور قابل توجهی دانش‌آموزان ترک تحصیل را تحت تاثیر قرار می‌دهند [۹]. چن و همکاران همچنین در مورد پیش‌بینی ترک تحصیل در ایالات متحده تحقیق کردند. در مطالعه چن، متغیرهای مورد استفاده برای پیش‌بینی ترک تحصیل عبارت بودند از اطلاعات دبیرستان، جمعیت‌شناسی، ثبت نام در دانشگاه و اطلاعات در هر ترم. بر اساس نتایج تحلیل، متغیرهای انتخاب شده به طور قابل توجهی نرخ ترک تحصیل دانشجویان را پیش‌بینی می‌کنند [۱۰].

دانشگاه‌ها به طور فزاینده‌ای از الگوریتم‌های یادگیری ماشین برای تحلیل مجموعه داده‌های بزرگ از جمله عملکرد تحصیلی، حضور و غیاب، وضعیت مالی، مشارکت در فعالیت‌های دانشگاه و شاخص‌های سلامت روان می‌توانند استفاده کنند تا دانشجویان در معرض خطر ترک تحصیل را شناسایی کنند [۱۱]. این مدل‌ها می‌توانند الگوهای پنهان و علائم هشدار اولیه را که ممکن است فوراً برای مشاوران یا کارکنان دانشگاهی آشکار نشوند، آشکار کنند و به مؤسسات اجازه می‌دهند تا فعالانه مداخله کنند. با تطبیق خدمات پشتیبانی (مانند تدریس خصوصی، کمک مالی، یا مشاوره سلامت روان) براساس پیش‌بینی‌های یادگیری ماشین، دانشگاه‌ها می‌توانند نرخ ماندگاری را بهبود بخشند [۱۲].

احمد و شهزادی با استفاده از الگوریتم پرسپترون چندلایه به پیش‌بینی و تحلیل عملکرد دانشجویان پرداختند و مدلی با دقت ۹۳/۲ درصد ارائه دادند [۱۳]. برنس و همکاران با

[Downloaded from sjfm.ir on 2026-11-29]

استفاده از الگوریتم‌های یادگیری ماشین همچون جنگل تصادفی به بررسی و شناسایی دانشجویان متمایل به انصراف از تحصیل پرداختند و مدلی با دقت ۸۹ درصد ارائه دادند [۱۴]. دیاز و همکاران با مقایسه چندین الگوریتم یادگیری ماشین به این نتیجه رسیدند که الگوریتم ماشین بردار پشتیبان دقت بالاتری برای پیش‌بینی دانشجویان متمایل به انصراف درمورد مطالعاتی‌شان دارد که دقتی در حدود ۹۴ درصد نشان داد [۱۵]. هرناندز و همکاران با استفاده از الگوریتم شبکه عصبی به عملکرد دانشجویان را مدل کردند و مدل آنها دقت ۹۴/۵ درصد را نشان داد [۱۶]. العوفی و همکاران بر روی پیش‌بینی میزان انصراف از دانشجویان کار کردند و با الگوریتم‌های یادگیری عمیق مدلی ساختند که دقتی در حدود ۹۶ درصد داشت [۱۷]. در این مقاله بر خلاف پژوهش‌های قبلی عوامل موثر بر انصراف از تحصیل دانشجویان با استفاده از روش ترکیبی یادگیری ماشین و الگوریتم فراابتکاری مورد بررسی قرار خواهد گرفت. علاوه بر آن، بر اساس تحقیقات قبلی، پژوهشگران از تکنیک‌های انتخاب و یا اهمیت ویژگی در مطالعات خود بهره برده‌اند، اما در این مقاله از هردوی این روش‌ها برای به دست آوردن یک مدل کارآمد و ارائه راهکارهای مدیریتی مناسب‌تر استفاده خواهد شد. در نهایت از ادغام نوآورانه الگوریتم‌های یادگیری ماشین و فراابتکاری برای تنظیم پارامترهای مدل استفاده می‌شود تا مدلی با دقت بالا و ارائه الگوریتم کاربردی به دست بیاید. نتایج این مسئله در راستای فهم درست عوامل موثر بر انصراف از تحصیل دانشجویان است که می‌تواند باعث کاهش مضرات این پدیده شود.

روش بررسی

این یک مطالعه توسعه‌ای است که با استفاده از نرم‌افزار پایتون به تحلیل داده‌های آکادمیک، اقتصادی-اجتماعی و جمعیتی مربوط به دانشجویان می‌پردازد تا بتواند دانشجویان در معرض انصراف از تحصیل را شناسایی کند و اطلاعات مفیدی را برای کاهش تعداد انصرافی‌ها در اختیار تصمیم‌گیران قرار دهد.

فراابتکاری

الگوریتم بهینه‌سازی ازدحام ذرات ، یک روش بهینه‌سازی است که در دسته الگوریتم‌های فراابتکاری قرار می‌گیرد و از رفتار گروهی پرندگان و ماهی‌ها الهام گرفته شده است [۱۸]. این الگوریتم که به دلیل کارآیی و توانایی خود در حل مسائل بهینه‌سازی شناخته می‌شود، یک سیستم اجتماعی را شبیه‌سازی می‌کند که در آن ذرات به عنوان نماینده افراد با یکدیگر همکاری می‌کنند تا بهترین راه‌حل را در یک فضای جستجو بیابند. توانایی حل مسئله این الگوریتم بر اساس مجموعه‌ای از اصول اساسی است که عملکرد آن را هدایت می‌کند [۱۹].

در زمینه بهینه‌سازی مسائل، الگوریتم بهینه‌سازی ازدحام ذرات با قرار دادن تصادفی گروهی از ذرات در فضای راه‌حل آغاز می‌شود. موقعیت هر ذره در این فضا نمایانگر یک راه‌حل ممکن است و مسیر حرکت آن تحت تأثیر دو عامل کلیدی قرار دارد: تجربه شخصی آن و عملکرد بهترین ذره در اطراف آن. ذرات با تنظیم مداوم موقعیت خود سعی می‌کنند به بهینه جهانی نزدیک شوند و همزمان تعادلی میان اکتشاف و بهره‌برداری برقرار کنند [۱۹]. این روش همکاری‌محور به الگوریتم اجازه می‌دهد فضای جستجو را به‌صورت غیرمتمرکز به‌طور کارآمد کاوش و بهره‌برداری کند. محاسبه حرکت ذرات از طریق تغییر سرعت با استفاده از رابطه (۱) انجام می‌شود.

$$\vec{V}_{ij}(h+1) = \varphi_1 e_1 \left(\vec{z}_{ij}(h) - \vec{x}_{ij}(h)\right) + \varphi_2 e_2 \left(\vec{z}_{gi}(h) - \vec{y}_{ij}(h)\right) + W\vec{V}_{ij}(h) \quad (۱)$$

در این معادله، $\vec{V}_{ij}(h)$ نشان‌دهنده سرعت ذره i در تکرار h است، در حالی که $\vec{V}_{ij}(h)$ مکان ذره i را در همان تکرار

[Downloaded from sjfm.ir on 2026-11-29]

مشخص می‌کند. مقادیر e_1 و e_2 عداد تصادفی در بازه (۰، ۱) هستند و φ_1 و φ_2 ثابت‌هایی هستند. همچنین، $\vec{Z}_{ij}(h)$ بهترین مکان یک ذره را نشان می‌دهد و $\vec{Z}_{gj}(h)$ نشان‌دهنده بهترین مکان کل جمعیت ذرات است. مقدار W نیز ثابت است و مکان جدید ذره از طریق رابطه (۲) محاسبه می‌شود.

$$\vec{y}_{ij}(h+1) = \vec{y}_{ij}(h) + \vec{V}_{ij}(h+1) \qquad (۲)$$

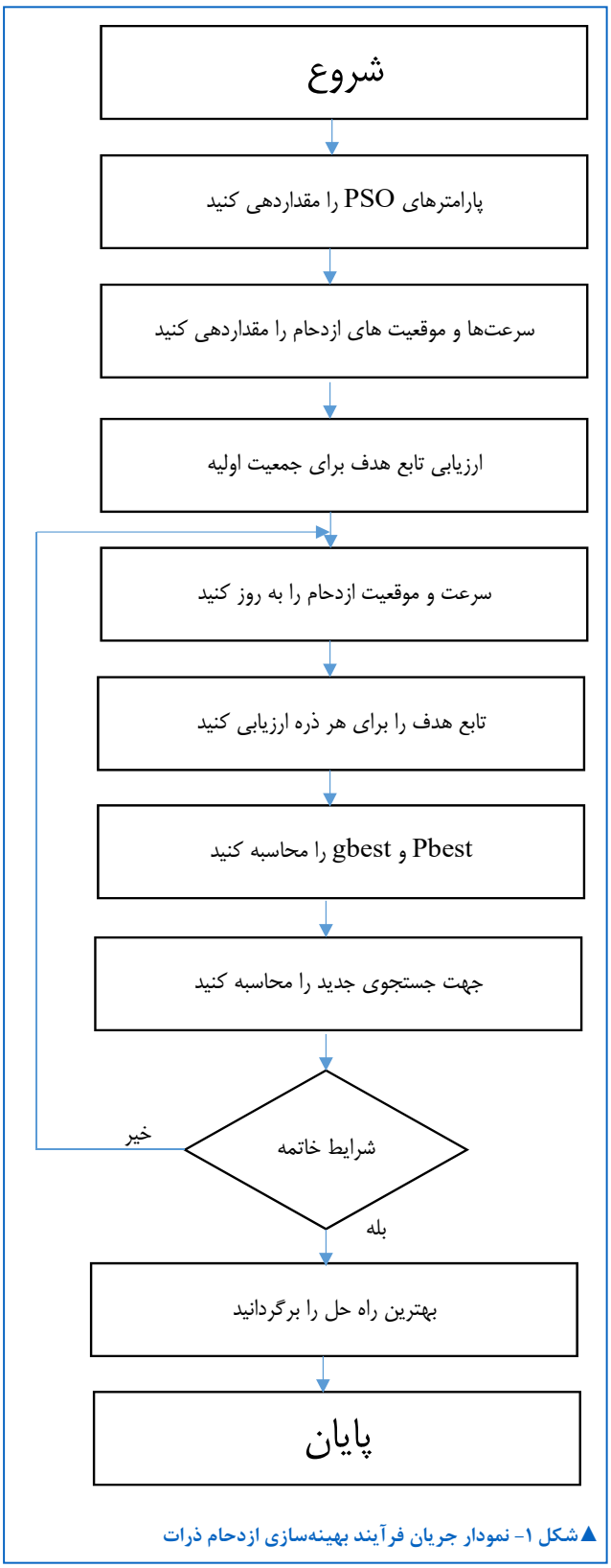
در این مقاله، از الگوریتم بهینه‌سازی ازدحام ذرات برای بهینه‌سازی فراپارامترهای مدل یادگیری ماشین به منظور بهبود دقت عملکرد استفاده شده است. نمودار جریان فرآیند بهینه‌سازی ازدحام ذرات در شکل ۱ نمایش داده شده است.

یادگیری ماشین

یادگیری نظارت‌شده یکی از سه حوزه اصلی یادگیری ماشین است که از دو بخش اصلی تشکیل شده است: ۱) رگرسیون و ۲) طبقه‌بندی.

الگوریتم تقویت گرادیان شدید یک روش یادگیری ماشینی قدرتمند و محبوب است که در دسته یادگیری گروهی قرار دارد. این الگوریتم تقویتی با ترکیب چند یادگیرنده ضعیف، معمولاً درخت‌های تصمیم‌گیری، یک مدل پیش‌بینی قوی ایجاد می‌کند. از ویژگی‌های برجسته این الگوریتم، توانایی آن در اجرای موثر وظایف رگرسیون و طبقه‌بندی است. همچنین، به دلیل دقت بالای پیش‌بینی و استحکام، محبوبیت زیادی پیدا کرده است [۲۰].

در اصل، الگوریتم تقویت گرادیان شدید به صورت متوالی مجموعه‌ای از درخت‌های تصمیم را ایجاد می‌کند، به طوری که هدف هر درخت اصلاح خطاهای درخت‌های قبلی است. این الگوریتم یک تابع هدف منظم‌شده را به حداقل می‌رساند که از دو بخش اصلی تشکیل شده است: تابع ضرر که تفاوت بین مقادیر پیش‌بینی‌شده و واقعی را



▲شکل ۱- نمودار جریان فرآیند بهینه‌سازی ازدحام ذرات

[Downloaded from sjfm.ir on 2026-11-29]

اندازه‌گیری می‌کند و عبارت منظم‌سازی که از پیچیدگی بیش از حد مدل و بیش برازش داده‌های آموزشی جلوگیری می‌کند [۲۱]. تابع هدف الگوریتم تقویت گرادیان شدید به صورت رابطه (۳) قابل بیان است.

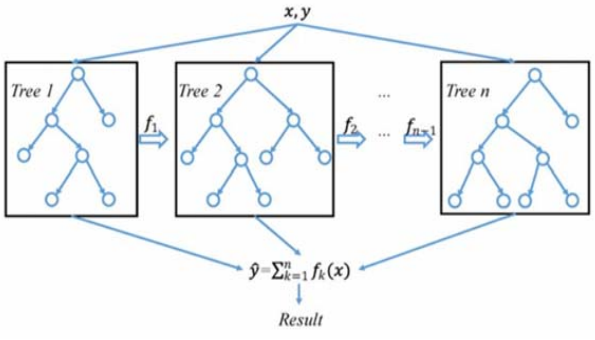
$$Objective(\Theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (۳)$$

در این معادله، L تابع زیان است، y_i برچسب واقعی مشاهده i-ام، $\hat{y}_i$ مقدار پیش‌بینی شده، K تعداد درخت‌ها، Ω(f_s) عبارت تنظیم‌کننده برای درخت k-ام، و Θ مجموعه پارامترهایی است که باید بهینه شوند.

نوآوری اصلی الگوریتم تقویت گرادیان شدید در به‌کارگیری تقویت گرادیان است، جایی که هر درخت جدید با گرادیان منفی تابع زیان، نسبت به پیش‌بینی مدل فعلی، همگام می‌شود. این رویکرد به الگوریتم اجازه می‌دهد تا بر تصحیح خطاهای مدل موجود تمرکز کرده و در نهایت به یک مدل دقیق‌تر و با تعمیم بهتر برسد [۲۲]. ساختار کلی الگوریتم تقویت گرادیان شدید در شکل ۲ به تصویر کشیده شده است.

معیارهای ارزیابی

برای ارزیابی یک مدل یادگیری ماشین می‌توان از معیارهای گوناگونی استفاده کرد (رابطه ۴-۷).



▲شکل ۲- ساختار الگوریتم تقویت گرادیان شدید(۳۲)

$$Precision = \frac{TP}{TP + FP} \quad (۴)$$

$$Recall = \frac{TP}{TP + FN} \quad (۵)$$

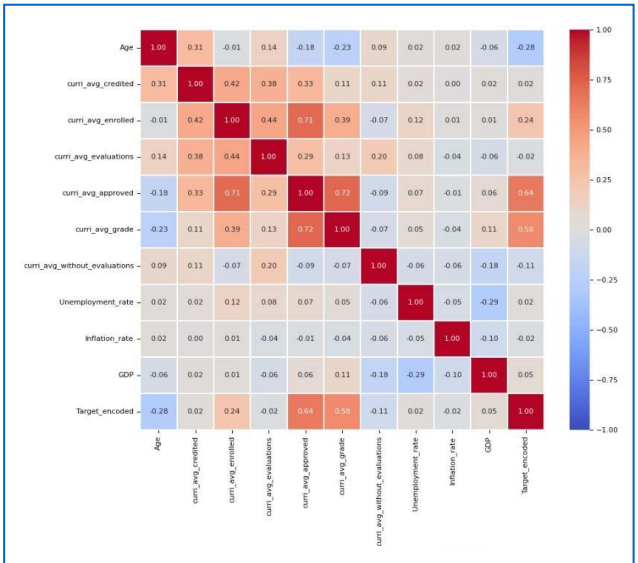
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۶)$$

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (۷)$$

مثبت واقعی (TP) زمانی رخ می‌دهد که مدل به درستی یک نتیجه مثبت را پیش‌بینی کند. منفی واقعی (TN) نیز زمانی است که مدل با دقت یک نتیجه منفی را پیش‌بینی کند. مثبت کاذب (FP) هنگامی اتفاق می‌افتد که مدل به اشتباه یک نتیجه مثبت را پیش‌بینی کند، و منفی کاذب (FN) زمانی رخ می‌دهد که مدل به اشتباه یک نتیجه منفی را پیش‌بینی کند.

یافته‌ها

در این قسمت، تحلیل‌هایی بر روی داده‌ها برای به دست آوردن اطلاعاتی از متغیرها و رابطه آنها ارائه شده و پس از



▲شکل ۲- تحلیل همبستگی متغیرها

[Downloaded from sjfm.ir on 2026-11-29]

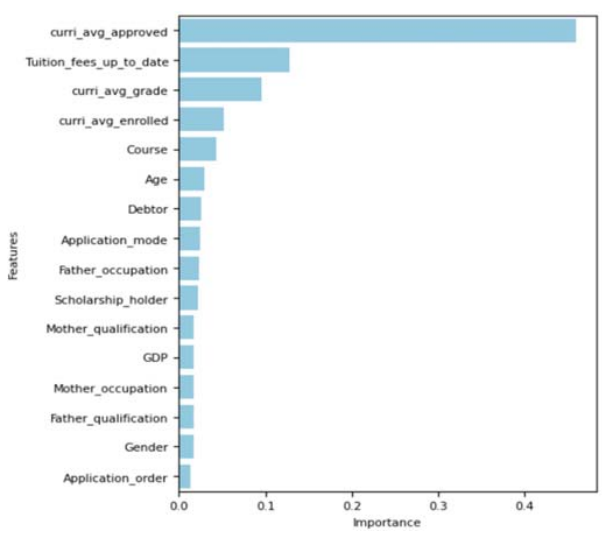
آن، یک مدل پیش‌بینی‌کننده ساخته می‌شود.
در شکل ۳ میزان همبستگی متغیرهای مستقل را با متغیر وابسته نشان داده شده است که اطلاعات مفیدی را برای ساخت مدل فراهم می‌کند.
پس از بررسی رابطه بین برخی از متغیرها و انصراف از تحصیل دانشجویان، مدل پیش‌بینی بر اساس متغیرهای مستقل ساخته می‌شود. برای این هدف، پس از استفاده از برخی تکنیک‌های آماده‌سازی داده‌ها بر روی مجموعه داده، داده‌ها به مجموعه آموزشی و مجموعه آزمایشی تقسیم شد که به ترتیب ۸۰ و ۲۰ درصد داده‌ها به آن تخصیص داده شد.
برخی از الگوریتم‌های طبقه‌بندی برای ساخت مدل یادگیری ماشینی استفاده می‌شوند و نتایج در جدول ۱ ذکر شده است.
بر اساس جدول ۱، ترکیب الگوریتم طبقه‌بندی‌کننده تقویت گرادیان شدید و الگوریتم بهینه‌سازی ازدحام ذرات عملکرد بهتری را نسبت به دیگران نشان می‌دهد. بنابراین برای ساخت مدل به عنوان الگوریتم اصلی انتخاب می‌شود.
در جدول ۲ پارامترهای مهم الگوریتم‌های بهینه‌سازی ازدحام ذرات و یادگیری ماشین قید شده‌اند.
در مرحله آخر، تحلیلی برای شناخت اهمیت ویژگی و تأثیر هر متغیر بر متغیر هدف پس از آموزش مدل انجام می‌شود. نتایج در شکل ۴ نشان داده شده است.
مهم‌ترین ویژگی‌ها عبارتند از تعداد واحدهای برداشته شده توسط دانشجو، وضعیت پرداخت شهریه توسط دانشجو و معدل دانشجو و پس از آن تعداد واحدهای درسی که دانشجو در آن شرکت دارد، نوع دوره‌های انتخابی و سن هستند. در میان این ویژگی‌های برتر، چهار داده آکادمیک، یکی داده اقتصادی-اجتماعی و یکی داده جمعیتی است.
شناسایی دانشجویانی که در معرض خطر ترک تحصیل هستند از طریق یادگیری ماشینی، برای دانشگاه‌ها برای اجرای مداخلات هدفمند و بهبود نرخ ماندگاری بسیار مهم است. فارغ‌التحصیلان عموماً به عملکرد تحصیلی بالاتری دست می‌یابند که در تعداد بیشتری از واحدهای درسی مصوب و نمرات بالاتر نسبت به کسانی که ترک

▼ جدول ۱- مقایسه دقت مدل‌های ایجاد شده توسط

الگوریتم	دقت مدل (درصد)
ماشین برداری پشتیبان	۸۹/۶
درخت تصمیم	۸۷/۳
رگرسیون لجستیک	۹۰/۲
جنگل تصادفی	۹۲/۳
تقویت گرادیان شدید	۹۵/۹
تقویت گرادیان شدید + بهینه‌سازی ازدحام ذرات	۹۸/۱۲

▼ جدول ۲- پارامترهای مهم الگوریتم‌های بهینه‌سازی ازدحام ذرات و تقویت گرادیان شدید

بهینه‌سازی ازدحام ذرات		تقویت گرادیان شدید	
اسم پارامتر	مقدار	اسم پارامتر	مقدار
تعداد ذرات	۵۰	نرخ یادگیری	۰/۱
C_1	۱	حداکثر عمق	۴
2_1	۱/۵	حداقل وزن	۶
تعداد تکرار	۱۰۰	gamma	۰



▲ شکل ۳- تحلیل همبستگی متغیرها

[Downloaded from sjfm.ir on 2026-11-29]

تحصیل می‌کنند منعکس می‌شود. دانشجویانی که شهریه پرداخت نشده دارند به طور قابل توجهی بیشتر احتمال دارد که از تحصیل خارج شوند (۸۷/۰۵ درصد) و احتمال بسیار کمتری برای فارغ‌التحصیلی (۴/۹۵ درصد) نسبت به کسانی که به موقع پرداخت می‌کنند، دارند که نرخ ترک تحصیل کمتر (۲۵/۲۱ درصد) و نرخ فارغ‌التحصیلی بالاتر (۵۵/۱۳ درصد) دارند. نرخ ترک تحصیل با افزایش سن برای هر دو جنس افزایش می‌یابد، اگرچه زنان به طور مداوم نرخ ترک تحصیل را در تمام گروه‌های سنی کمتر نشان می‌دهند. علاوه بر این، نرخ ترک تحصیل در دوره‌ها بسیار متفاوت است، از ۱۵/۴ تا ۶۶/۷ درصد، به طوری که از هر ۱۷ دوره، ۵ درس دارای نرخ بالای ۵۰ درصد هستند. با استفاده از یادگیری ماشینی برای پیش‌بینی خطرات ترک تحصیل، موسسات بهتر می‌توانند منابع را برای حمایت از دانشجویان در تکمیل تحصیلات خود تخصیص دهند.

بحث

مدل ترکیبی این پژوهش با دقت ۹۸/۱۲ درصد نشان داد که نسبت به مطالعات مشابه عملکرد برتری دارد، به‌ویژه در مقایسه با کار Jain و همکاران [۲۴] که با چارچوب PSO-SMOTE به دقت ۸۶ درصد و AUC=۰/۹۵۹ دست یافته بودند و یا پژوهش Khatun و همکاران [۲۵] که با XGBoost و روش‌های توضیح‌پذیری (SHAP/LIME) دقت ۹۴/۴ درصد گزارش کردند. یافته‌های ما اهمیت داده‌های آکادمیک به‌ویژه وضعیت پرداخت شهریه و معدل را در پیش‌بینی ترک‌تحصیل تأیید کرد، در حالی‌که سایر مطالعات بیشتر بر متغیرهای جمعیت‌شناختی یا روش‌های تبیین‌پذیری تمرکز داشتند. برتری این مدل در افزایش صحت پیش‌بینی می‌تواند به دانشگاه‌ها کمک کند تا مداخلات مدیریتی دقیق‌تری برای کاهش نرخ ترک‌تحصیل طراحی کنند. با این وجود، محدودیت‌هایی مانند استفاده از داده‌های یک جامعه خاص و فقدان متغیرهای روان‌شناختی یا انگیزشی وجود دارد که تعمیم‌پذیری نتایج را کاهش می‌دهد. در نتیجه، توصیه می‌شود پژوهش‌های آتی علاوه بر بهره‌گیری از داده‌های متنوع‌تر، از رویکردهای توضیح‌پذیرتر نیز برای تقویت اعتبار یافته‌ها استفاده کنند.

نتیجه‌گیری

یک مدل ترکیبی از الگوریتم بهینه‌سازی ازدحام ذرات و تقویت گرادیان شدید با دقت ۹۸/۱۲ درصد ارائه شده برای شناسایی دانشجویانی که در معرض خطر ترک تحصیل هستند. همچنین با استفاده از روش‌های تحلیل داده و تکنیک‌های آماری نشان داده شد که داده‌های آکادمیک بیشترین تأثیر را در ترک تحصیل دارند. دانشجویانی که شهریه خود را به‌موقع پرداخت نمی‌کنند، بیشتر در معرض ترک تحصیل (۸۷/۰۵ درصد) و کمتر در معرض فارغ‌التحصیلی (۴/۹۵ درصد) هستند، در حالی که پرداخت به‌موقع شهریه منجر به کاهش نرخ ترک تحصیل (۲۵/۲۱ درصد) و افزایش احتمال فارغ‌التحصیلی (۵۵/۱۳ درصد) می‌شود. همچنین با افزایش سن، نرخ ترک تحصیل در هر دو جنس افزایش می‌یابد، اما زنان به‌طور مداوم در تمام گروه‌های سنی نرخ ترک تحصیل کمتری دارند. برای پژوهش‌های بعدی، می‌توان از روش‌های یادگیری ماشین و عمیق دیگر برای ارائه مدل با دقت‌های بیشتر استفاده کرد. ضمن اینکه می‌توان از انواع دیگر از داده‌ها برای تحلیل این موضوع استفاده کرد.

تأییدیه اخلاقی: پژوهش حاضر قسمتی از یک طرح تحقیقاتی با همین عنوان است.
تضاد منافع: هیچ‌گونه تعارض منافعی از سوی نویسنده اعلام نشده است.
سهم نویسندگان: تمام نویسندگان سهم یکسانی (هر یک ۵۰ درصد) در مقاله داشتند. به شکلی که نویسنده اول طراحی اولیه مطالعه، نظارت بر نگارش نهایی مقاله و نویسنده دوم نیز جمع‌آوری و تحلیل داده‌ها و نگارش اولیه مقاله را برعهده داشت.
منابع مالی: این پژوهش مورد حمایت مالی قرار نگرفته است.

[Downloaded from sjfm.ir on 2026-11-29]

Resources

1. Hilal M, Khabbache H, Ali DA. Dropping out of school: A psychosocial approach. Advances in Medicine, Psychology, and Public Health. 2024;1(1):26-36.
2. Brand JE. Overcoming the odds: The benefits of completing college for unlikely graduates: Russell Sage Foundation; 2023. doi: 10.1353/book.112674.
3. Powell F, Scanlon M, Leahy P, Jenkinson H, Byrne O. The Making of a Left-Behind Class: Educational Stratification, Meritocracy and Widening Participation: Policy Press; 2024. doi: 10.56687/9781447367970.
4. Contini D, Cugnata F, Scagni A. Social selection in higher education. Enrolment, dropout and timely degree attainment in Italy. High Educ (Dordr). 2018;75(5):785-808. doi: 10.1007/s10734-017-0170-9.
5. Tovar E. The role of faculty, counselors, and support programs on Latino/a community college students' success and intent to persist. Community College Review. 2015;43(1):46-71. doi: 10.1177/0091552114553788.
6. Hanson M. College Dropout Rates; 2024. Available from: https://educationdata.org/college-dropout-rates.
7. Galve-González C, Bernardo AB, Castro-López A. Understanding the dynamics of college transitions between courses: Uncertainty associated with the decision to drop out studies among first and second year students. Eur J Psychol Educ. 2024;39(2):959-78. doi: 10.1007/s10212-023-00732-2.
8. Qvortrup A, Lykkegaard E. The Malleability of Higher Education Study Environment Factors and Their Influence on Humanities Student Dropout—Validating an Instrument. Educ Sci (Basel). 2024;14(8):904. doi: 10.3390/educsci14080904.
9. Pérez B, Castellanos C, Correal D, editors. Predicting student drop-out rates using data mining techniques: A case study. IEEE Colombian Conference on Applications in Computational Intelligence; 2018: Springer. doi: 10.1007/978-3-030-03023-0_10.
10. Chen Y, Johri A, Rangwala H, editors. Running out of stem: a comparative study across stem majors of college students at-risk of dropping out early. Proceedings of the 8th international conference on learning analytics and knowledge; 2018. doi: 10.1145/3170358.3170410.
11. Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn Environ. 2022;9(1):11. doi: 10.1186/s40561-022-00192-z.
12. Maldonado S, Miranda J, Olaya D, Vásquez J, Verbeke W. Redefining profit metrics for boosting student retention in higher education. Decis Support Syst. 2021;143:113493. doi: 10.1016/j.dss.2021.113493.
13. Ahmad Z, Shahzadi E. Prediction of Students' Academic Performance Using Artificial Neural Network. Bulletin of Education and Research. 2018;40(3):157-64.
14. Berens J, Schneider K, Gortz S, Oster S, Burghoff J. Early Detection of Students at Risk--Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. Journal of Educational Data Mining. 2019;11(3):1-41. doi: 10.2139/ssrn.3275433.
15. Díaz I, Bernardo AB, Esteban M, Rodríguez-Muñiz LJ, editors. Variables influencing university dropout: A machine learning-based study. International Conference on EUropean Transnational Education; 2020: Springer. doi: 10.1007/978-3-030-57799-5_10.
16. Rodríguez-Hernández CF, Musso M, Kyndt E, Cascallar E. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. Computers and Education: Artificial Intelligence. 2021;2:100018. doi: 10.1016/j.caeai.2021.100018.
17. El Aouifi H, El Hajji M, Es-Saady Y. A hybrid approach for early-identification of at-risk dropout students using LSTM-DNN networks. Educ Inf Technol (Dordr). 2024;29(14):18839-57. doi: 10.1007/s10639-024-12588-0.
18. Amani MA, Sarkodie SA. Mitigating spread of contamination in meat supply chain management using deep learning. Sci Rep. 2022;12(1):5037. doi: 10.1038/s41598-025-87201-6.
19. Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. Soft comput. 2018;22(2):387-408. doi: 10.1007/s00500-016-2474-6.
20. Li J, An X, Li Q, Wang C, Yu H, Zhou X, et al. Application of XGBoost algorithm in the

[Downloaded from sjfm.ir on 2026-11-29]

optimization of pollutant concentration. Atmos Res. 2022;276:106238. doi: 10.1016/j.atmosres.2022.106238.
21. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments. 2023;31(6):3360-79. doi: 10.1080/10494820.2021.1928235.
22. Jabeur SB, Mefteh-Wali S, Viviani JL. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Oper Res. 2024;334(1):679-99. doi: 10.1007/s10479-021-04187-w.
23. Wang Y, Pan Z, Zheng J, Qian L, Li M. A hybrid ensemble method for pulsar candidate classification. Astrophys Space Sci. 2019;364:139. doi: 10.1007/s10509-019-3602-4.

[Downloaded from sjfm.ir on 2026-11-29]