Tech. Edu. J. 14(4): 821-834, Autumn 2020

Technology of Education Journal
(TEJ)
Homepage: jte.sru.ac.ir

ORIGINAL RESEARCH PAPER

Predicting and analyzing the performance of students through data mining techniques to improve academic performance

M. Ghodoosi*[*,1], F. Mirsaeedi[1], H. Koosha[2]

[1] *Department of Industrial Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran*
[2] *Department of Industrial Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*

Received: 2 March 2019
Reviewed: 3 April 2019
Revised: 9 June 2019
Accepted: 19 June 2019

KEYWORDS:
Educational Data Mining
Unit Selection
Academic Performance
Logit Boost

* Corresponding author
M.ghodoosi@torbath.ac.ir
(+9851) 51240149

ABSTRACT

Background and Objectives: Nowadays, significant advancements in information technology and communication field in different societies are seen. Given that universities as a leading institution in the field of science have moved towards electronic processes in the management of education and educational environments, there are databases with a large amount of information. By analyzing this massive data of educational systems, methods can be provided to improve the educational status of students. Educational data mining has sought to discover the knowledge contained in the data of the educational system. One of the applications of educational data mining is to predict students' academic performance. Predicting students' academic performance and providing useful solutions is of particular importance in the success of educational systems and can help managers make right decisions to increase the efficiency of the educational system and better student performance. The purpose of this paper is to identify effective indicators on academic performance, predict students' academic status using data mining techniques, and finally present a new trend for modifying unit selection and educational strategies to increase the efficiency of the education system.

Methods: Steps of this research are determined according to CRISP model. In current research, databases containing 9 datasets of specialized courses in industrial engineering were used. The students had bachelor degrees. Indicators affecting students' performance have been identified based on previous research and expert opinions. Demographic data and academic records of undergraduate students were entered in database. After data preprocessing, 13 attributes were selected and different models were proposed to predict student's academic status in the next semester. Then, a comparison between the results of 4 different algorithms has been done.

Findings: All 13 attributes were identified to be effective according to information gain and gain ratio. These 13 attributes are as follow: GPA, total passed units, number of conditional terms, type of admission, marital status, gender, university admission year, living place, age, current semester, prerequisite course score, instructor of the course, repetition of the course. Among the 4 considered models, the Logit Boost algorithm is known as the best model for categorizing two class and multi-class according to the accuracy rate and ROC.

Conclusion: Because of acceptable performance of data mining algorithms, the use of these algorithms in predicting student performance is appropriate and the proposed model can be used as a support tool for decision making in educational systems. According to the obtained results and the opinion of academic experts, the unit selection process was redesigned. The proposed model can be used as a decision support tool in educational systems. The presented process uses the available data in educational systems and data mining science, provides useful knowledge to decision-makers to make the right and appropriate decision. Decision makers can make appropriate decisions by examining the predictions made by the data mining algorithm and obtaining useful information, in order to make the educational system more efficient.

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
58	3	4

مقاله پژوهشی

پیش‌بینی و تحلیل عملکرد دانشجویان به کمک تکنیک‌های داده‌کاوی به منظور بهبود عملکرد تحصیلی

محمد قدوسی*[۱]، فاطمه میرسعیدی[۱]، حمیدرضا کوشا[۲]

[۱] گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه تربت حیدریه، تربت حیدریه، ایران
[۲] گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

چکیده

پیشینه و اهداف: در حال حاضر پیشرفت‌های قابل توجهی در عرصه فناوری اطلاعات و ارتباطات در جوامع مختلف دیده می‌شود. با توجه به این پیشرفت‌ها، دانشگاه‌ها به عنوان یک نهاد پیشرو در عرصه علم، به سمت فرآیندهای الکترونیکی در مسیر مدیریت آموزش حرکت نموده‌اند و در محیط‌های آموزشی، پایگاه‌های اطلاعاتی با حجم اطلاعات زیاد وجود دارد. با تحلیل این داده‌های انبوه سیستم‌های آموزشی، می‌توان روش‌هایی را برای بهبود وضعیت آموزشی دانشجویان ارایه داد. داده‌کاوی آموزشی به دنبال کشف دانش موجود در داده‌های سیستم آموزشی بوده‌است. یکی از کاربردهای داده‌کاوی آموزشی، پیش‌بینی عملکرد تحصیلی دانشجویان است. پیش‌بینی عملکرد تحصیلی دانشجویان و ارائه راهکارهای مفید از اهمیت ویژه‌ای در موفقیت نظام‌های آموزشی برخوردار است و می‌تواند به تصمیم‌گیری درست مدیران، جهت افزایش بازدهی سیستم آموزشی و عملکرد بهتر دانشجویان، کمک شایانی کند. هدف مقاله حاضر، شناسایی شاخص‌های مؤثر بر عملکرد تحصیلی، پیش‌بینی وضعیت تحصیلی دانشجویان با استفاده از تکنیک‌های داده‌کاوی و در نهایت، ارائه روندی جدید برای اصلاح روش انتخاب واحد و راهکارهای آموزشی در جهت افزایش کارایی سیستم آموزش است.

روش‌ها: گام‌های این پژوهش بر اساس مدل Crisp تعیین شده است. در پژوهش حاضر، پایگاه‌داده‌ای شامل ۹ مجموعه داده از درس‌های تخصصی رشته مهندسی صنایع استفاده شدند. دوره تحصیلی دانشجویان در نظر گرفته شده کارشناسی بوده‌است. شاخص‌های تاثیرگذار بر عملکرد دانشجویان، بر اساس تحقیقات قبلی و نظر خبرگان شناسایی شده‌است. داده‌های جمعیت‌شناختی و سوابق تحصیلی دانشجویان مقطع کارشناسی رشته مهندسی صنایع وارد پایگاه داده شدند. پس از پیش‌پردازش داده‌ها، ۱۳ شاخص در نظر گرفته شد و با کمک الگوریتم‌های مختلف، مدل‌های مختلفی برای پیش‌بینی وضعیت تحصیلی دانشجویان در نیمسال بعدی ارائه گردید. مدل‌های شبکه بیزی، لوجیت بوست، پارت و درخت تصمیم به عنوان پرکاربردترین الگوریتم های داده‌کاوی آموزشی در این پژوهش مورد استفاده قرار گرفته و جهت بررسی عملکرد الگوریتم‌ها از دو شاخص صحت و سطح زیر نمودار عملکرد استفاده شد. ۹ پایگاه داده دروس در دو حالت دو و چند کلاسه در نظر گرفته‌شدند. در ادامه، مقایسه‌ای میان نتایج حاصل از ۴ الگوریتم مختلف صورت گرفته‌است.

یافته‌ها: با توجه به شاخص‌های بهره اطلاعات و نسبت بهره، تمامی ۱۳ شاخص در نظر گرفته شده، به عنوان شاخص‌های مؤثر شناسایی شدند. این شاخص‌ها عبارتند از: معدل، کل واحدهای گذرانده، تعداد ترم‌های مشروطی، نوع پذیرش، وضعیت تأهل، جنسیت، سال ورود به دانشگاه، سن، محل زندگی، ترم حاضر، نمره درس پیش‌نیاز، استاد درس، تکرارد در اخذ واحد. از بین ۴ مدل در نظر گرفته شده، بهترین مدل در دسته‌بندی و پیش‌بینی عملکرد آموزشی دانشجویان ، الگوریتم Logit Boost شناخته شد. این الگوریتم، در هر دو حالت دو و چندکلاسه براساس شاخص‌های درصد صحت و سطح زیر نمودار ROC عملکرد بهتری از خود نشان داده‌است.

نتیجه‌گیری: با توجه به عملکرد قابل قبول الگوریتم‌های داده‌کاوی، استفاده از این الگوریتم‌ها در پیش‌بینی عملکرد دانشجویان مناسب است و می‌توان مدل پیشنهادی را به عنوان یک ابزار پشتیبان تصمیم‌گیری در سیستم‌های آموزشی مورد استفاده قرار داد. در نهایت، با توجه به نتایج به‌دست آمده و نظرخواهی از خبرگان دانشگاهی، فرایند انتخاب واحد، بازطراحی گردید. فرایند ارایه شده با استفاده از داده‌های موجود در سیستم‌های آموزشی و علم داده‌کاوی، دانش مفیدی به تصمیم‌گیرندگان جهت تصمیم صحیح و مناسب ارایه می‌دهد. تصمیم‌گیرندگان می‌توانند با بررسی پیش‌بینی‌های انجام شده توسط الگوریتم داده‌کاوی و کسب اطلاعات مفید، تصمیمات مناسب اخذ نمایند، تا سیستم آموزشی بازدهی بیشتری داشته‌باشد.

تاریخ دریافت: ۱۱ اسفند ۱۳۹۷
تاریخ داوری: ۱۴ فروردین ۱۳۹۸
تاریخ اصلاح: ۱۹ خرداد ۱۳۹۸
تاریخ پذیرش: ۲۹ خرداد ۱۳۹۸

واژگان کلیدی:
داده‌کاوی آموزشی
انتخاب واحد
عملکرد تحصیلی

* نویسنده مسئول
M.ghodoosi@torbath.ac.ir
۰۵۱-۵۱۲۴۰۱۴۹

مقدّمه

مسأله بررسی عملکرد تحصیلی، یکی از مباحث مهم و مورد توجه محققین حوزه مدیریت آموزشی به شمار می‌رود [۱-۳]. یکی از موضوعات قابل توجه در عملکرد تحصیلی، پیش‌بینی درست عملکرد تحصیلی دانشجویان و اقدام به موقع و مشاوره به دانشجویان در معرض خطر افت تحصیلی است [۴].

از سوی دیگر، موضوعی که عصر حاضر را از دوران گذشته متمایز کرده، فناوری اطلاعات و ارتباطات است. میزان بهره‌مندی و کاربرد فناوری اطلاعات و ارتباطات با شکاف بین کشورها و افراد رابطه مستقیم دارد. بنابراین شاید بتوان گفت مهم‌ترین شاخص پیشرفت، میزان توسعه و کاربرد فناوری اطلاعات و ارتباطات در امر آموزش است [۵]. از آنجا که مؤسسات آموزش عالی و دانشگاه‌ها اصلی‌ترین نهادهای ادغام‌کننده فناوری رایانه در فرایندهای آموزشی هستند [۶] باید زیرساخت‌های لازم برای الکترونیکی شدن را فراهم کنند [۷].

با الکترونیکی شدن بسیاری از فرایندها، مؤسسات آموزشی پایگاه‌های داده‌ای متعددی در اختیار دارند که در بردارنده حجم بالایی از داده‌های مربوط به سوابق آموزشی، تحصیلی و ... است. داده‌کاوی آموزشی، روش‌هایی را برای تحلیل این داده‌های انبوه سیستم‌های آموزشی توسعه می‌دهد تا بتوان الگوهایی را برای بهبود وضعیت آموزشی و تصمیم‌گیری‌ها ارائه داد [۸]. نتایج حاصل از داده‌کاوی می‌تواند فرایند آموزش و یادگیری را برای طراحی یا بازطراحی محیط آموزشی به خوبی بهبود دهد [۹]. زیرا این نتایج می‌تواند سیستم‌های آموزش‌عالی را در تصمیم‌گیری بهتر و داشتن طرح پیشرفته‌تر در هدایت و مشاوره دانشجویان یاری کند [۱۰].

در مقاله بیکر [۱۱] پنج رویکرد در داده کاوی آموزشی ارائه شده‌است: پیش‌بینی، خوشه‌بندی، استخراج رابطه، کشف درون مدل‌ها و تقسیم داده‌ها برای قضاوت انسانی. در این زمینه، طبقه‌بندی و خوشه‌بندی فراگیران، جزء موضوعات فرعی اما مطرح در سالیان اخیر بوده‌است. بسیاری از پژوهشگران، اعضای یک جامعه آماری از فراگیران را به طبقه یا خوشه شامل: متوسط (اکثریت)، ضعیف و ممتاز تقسیم می‌کنند. پژوهشگران دیگری دو طیف بسیار ضعیف و بسیار قوی را نیز به سه طیف قبلی اضافه می‌کنند [۱۲].

یک کاربرد کلیدی عمده پیش‌بینی در داده‌کاوی آموزشی پیش‌بینی خروجی‌های آموزشی دانشجویان است. پژوهش‌ها در این حوزه در چند سطح مختلف انجام شده‌است: در سطح سیستم آموزش و پرورش، سطح دوره، یا در سطح درجه تحصیلی و غیره [۴]. دانش قابل کشف از طریق داده‌کاوی آموزشی نه تنها برای صاحبان سیستم (مدرسین و مسئولین آموزشی) بلکه برای کاربران سیستم (دانشجویان) نیز قابل استفاده است [۱۳]. بدین صورت که به دانشجویان کمک می‌کند تا در فرایند یادگیری از بازخوردهای داده‌کاوی آموزشی استفاده کرده و موفق عمل کنند. از سوی دیگر، مدرسین می‌توانند کارایی فرایند یادگیری را ارتقا دهند و مسئولین آموزشی منابع سازمانی اعم از مادی و انسانی را به نحو مناسب‌تری تخصیص دهند [۱۴]. به عنوان مثال، اگر در یک مدل پیش‌بینی، برای دانشجویی نمره پایین پیش‌بینی شود، آن دانشجو به طور بالقوه ضعیف شناخته می‌شود و با راهکارهایی از قبیل مسائل تکنیکی اضافه و بازنگری دروس قبلی می‌توان عملکرد وی را بهبود داد [۱۵]. شخصی‌سازی آموزش، بیشینه کردن کارایی سیستم آموزش و کاهش هزینه‌ها از فواید داده‌کاوی آموزشی است [۱۶].

برخی از پژوهشگران، با بررسی مقالات در حوزه داده کاوی آموزشی، مطالعات از نوع پیش‌بینی را به عنوان رایج‌ترین مطالعات در این حوزه شناسایی کردند. در حالی‌که در پژوهش‌های داخلی در حوزه پیش‌بینی اغلب از شیوه‌های کلاسیک آماری بهره می‌برند. البته با رایج شدن شیوه‌های نوین داده‌کاوی در حوزه‌های مهندسی و مدیریت، می‌توان امیدوار بود این روش‌ها در حوزه مدیریت آموزشی نیز رواج پیدا کند و شکاف ایجاد شده بین پژوهش‌های داخلی و خارجی از بین برود [۱۲]. از سویی دیگر، شناسایی فاکتورهای تأثیرگذار بر عملکرد دانشجویان در داده‌کاوی آموزشی یک ورودی حیاتی و بسیار مهم برای فهم و بهبود محیط آموزشی است [۱۷].

دسته‌بندی‌های مختلفی برای مقالات داده‌کاوی آموزشی ارائه شده است. برای مثال، پنا آیالا [۱۸] پژوهش‌های داده‌کاوی آموزشی را بررسی و تحلیل کرده و مقالات را براساس مدل، روش و الگوریتم مورداستفاده دسته‌بندی کرده‌است. شهیری، هوساین و رشید [۱۹] مقالاتی را که عملکرد دانشجویان را با استفاده از تکنیک داده‌کاوی پیش‌بینی کرده‌اند، بر اساس شاخص‌ها و روش‌ها بررسی و دسته‌بندی نمودند. رودریگز، زرات و ایسوتنی [۲۰] به بررسی مقالات در حوزه داده کاوی آموزشی در طول ۲۰ سال گذشته پرداخته و مقالات را براساس هدف اجرای داده‌کاوی آموزشی تقسیم کردند، شامل مدل‌های شناسایی الگوی رفتاری دانشجو، شناسایی فاکتورهای تأثیرگذار، مدل‌های بهبود عملکرد تحصیلی و افزایش همکاری و هماهنگی دانشجویان و اساتید. در مقاله حاضر از این دسته‌بندی استفاده شده و جدیدترین مقالات در این حوزه براساس این دسته‌بندی در هر زیربخش، توضیح داده‌شدند.

مدل‌های شناسایی الگوی رفتاری دانشجو

برخی از مقالات، که به دنبال شناسایی الگوی رفتاری دانشجو در طول تحصیل هستند مورد بررسی قرار می‌گیرند. این مقالات از نظر شاخص ارزیابی، دسته‌بندی شدند. برخی از مقالات تنها از شاخص درصد صحت برای ارزیابی الگوریتم‌ها استفاده کردند. از جمله، حیدری و یقینی [۲۱] با مقایسه تکنیک‌های مختلف داده‌کاوی، به شناسایی بهترین روش برای پیش‌بینی وضعیت تحصیلی دانشجویان داخل کشور پرداخته و به این نتیجه رسید که در خوشه‌بندی وضعیت تحصیلی، تکنیک نزدیک‌ترین همسایه و در پیش‌بینی وضعیت تحصیلی، شبکه عصبی کارآمدتر از سایر تکنیک‌ها می‌باشد. مقصودی و همکاران [۲۲] به کشف الگوهای نهفته در انتخاب واحد دانشجویان و پیش‌بینی نمرات

آنها در سامانه‌های آموزش الکترونیکی پرداخته‌اند. سن و اوکار [۱۷] به پیش‌بینی عملکرد دانشجویان مهندسی کامپیوتر پرداختند. سن و همکاران [۲۳] از الگوریتم‌های مختلف برای پیش‌بینی نمرات دانش‌آموزان استفاده کردند. نلتک و زوئیلینگ [۲۴] مطالعه‌ای روی اطلاعات دانشجویان در سال‌های ۲۰۱۲ تا ۲۰۱۳ انجام داده و نمرات را پیش‌بینی کردند. استرچ، کروز، سوآرس، موریا و آبرن [۲۵] موفقیت یا شکست و نمره دانشجو را در یک دوره با استفاده از متغیرهای اجتماعی از قبیل سن و جنسیت و ... پیش‌بینی می‌کنند. راچبوری و همکاران [۲۶] داده‌کاوی آموزشی را بر روی ۶۸۸۴ رکورد بین سال‌های ۲۰۰۴ تا ۲۰۱۰ انجام دادند. نق و همکاران [۲۷] عملکرد دانشجویان را در چند دانشگاه پیش‌بینی کرده‌اند. برای این کار حالات چهار کلاسه، سه و دو کلاسه را بررسی نموده و از الگوریتم‌های درخت تصمیم و شبکه بیضی با 10-fold Cross-validation استفاده کرده‌اند.

دسته‌ی دیگری از مقالات، علاوه بر نرخ صحت، سایر شاخص‌ها را نیز مورد بررسی قرار دادند؛ مانند کابکچیوا و همکاران [۲۸] و کابکچیوا [۲۹] که در پژوهش‌های خود عملکرد ۱۰۳۳۰ دانشجو را در بخش آموزش بلغارستان مورد بررسی قرار دادند. آن‌ها از الگوریتم‌های دسته‌بندی مختلف برای ۵ کلاس و با 10-fold Cross-validation استفاده کرده‌اند. اسکوئی و اکبری [۳۰] ۵۰۰ دانش‌آموز دبیرستانی و ۶۰۰ دانشجو در مقطع لیسانس را از نظر عملکرد با الگوریتم‌های مختلف در دو کلاس قبول یا رد مورد بررسی قرار دادند. یهوآلا [۳۱] پذیرش یا رد ۱۱۸۷۳ دانشجو را با الگوریتم‌های درخت تصمیم و بیزی ساده و 10-fold Cross-validation مورد بررسی قرار داده‌است. کار و همکاران [۳۲] پنج الگوریتم را در داده‌کاوی آموزشی در هند با هم مقایسه کردند. پندی و تارونا [۳۳] رویکردی ترکیبی برای پیش‌بینی عملکرد دانشجویان در هند ارائه داده و با الگوریتم‌های دسته‌بندی مقایسه کردند که الگوریتم ترکیبی بهتر عمل کرده‌است.

برخی دیگر از مقالات نیز از شاخص صحت استفاده نکردند و از سایر شاخص‌ها مانند مساحت منحنی مشخصه عملکرد، دقت و ... استفاده کردند. ابوسعا [۳۴] براساس فاکتورهای اجتماعی و شخصی، داده‌کاوی آموزشی را انجام داده‌است. یحیی [۳۵] از الگوریتم PSO برای دسته‌بندی در داده‌کاوی آموزشی استفاده کرده و با سایر الگوریتم‌ها مقایسه نموده است. سنتانا و همکاران [۳۶] اثربخشی الگوریتم‌های داده‌کاوی در دو پایگاه داده دانشگاهی را در برزیل بررسی کردند. ماشین بردار پشتیبان بسیار بهتر از سایر الگوریتم‌ها عمل کرده‌است. فرناندس و همکاران [۳۷] موفقیت یا عدم موفقیت دانش‌آموزان برزیلی را پیش‌بینی کردند. ونلی و همکاران [۳۸] نیز برای داده‌کاوی آموزشی از الگوریتم ژنتیک استفاده کرده‌اند.

مدل‌های شناسایی فاکتورهای تأثیرگذار بر عملکرد

مقالاتی که به شناسایی فاکتورهای اصلی تأثیرگذار بر عملکرد تحصیلی دانشجو می‌پردازند در این زیربخش بررسی می‌شوند. تاری، مینایی، فراهی و نیکنام پیرزاده [۳۹] عوامل مؤثر در میزان مشروطی دانشجویان دانشگاه پیام نور استان قم را بررسی کرده‌اند. احمدی و همکاران [۴۰] فاکتورهای تأثیرگذار بر رویگردانی دانشجویان شهریه‌پرداز را با استفاده از تکنیک‌های داده‌کاوی شناسایی و تحلیل کرده و از الگوریتم‌های دسته‌بندی برای پیش‌بینی انصراف استفاده کردند. رحمتی و همکاران [۴۱] به شناسایی عوامل تأثیرگذار بر مشروطی دانشجویان دانشگاه شهید باهنر پرداخته‌اند. امین بیدختی و همکاران [۱۲] با استفاده از شبکه عصبی به پیش‌بینی آسیب‌پذیری تحصیلی دانشجویان مقطع کارشناسی رشته‌های مهندسی در کوتاه مدت (نیمسال تحصیلی) اقدام کرده‌اند. در این پژوهش، معدل نیمسال قبل، معدل کل، زوج یا فرد بودن نیمسال، نوع واحدهای اخذشده و مبادرت به فعالیت‌های فوق برنامه در نیمسال، از مؤثرترین عوامل در پیش‌بینی وضعیت تحصیلی دانشجویان شناخته‌شده‌است. رستمی و همکاران [۵] با استفاده خوشه‌بندی فازی به شناسایی روابط عوامل تأثیرگذار بر یادگیری پرداختند و در ادامه با استفاده از تکنیک‌های داده‌کاوی نمرات دانشجویان را پیش‌بینی کردند.

مدل‌های بهبود عملکرد دانشجو

هدف برخی مقالات، بهبود عملکرد دانشجو به وسیله برجسته کردن مهارت‌ها یا دغدغه‌های دانشجو در طول فرایند یادگیری است. مانند، دی پیر و رابو [۴۲] ابتدا ابعاد مختلف سبک یادگیری را در قالب مدل فلدر- سیلورمن سنجیده‌اند و سپس، با استفاده از الگوریتم k-میانگین به گروه‌بندی یادگیرندگان براساس سبک یادگیری پرداخته‌اند. این مدل در دوره‌ای واقعی مورد ارزیابی قرار گرفته و نتایج مطلوبی از جمله موفقیت تحصیلی بالاتر و جذابیت بیشتر به همراه داشته‌است. رومرو و ونتورا [۸] مناسب بودن کمّی و کیفی و اطلاعات شبکه اجتماعی را در مورد استفاده انجمنی در برابر الگوریتم‌های دسته‌بندی و خوشه‌بندی کلاسیک برای شکست یا موفقیت دانشجویان در یک دوره بررسی کرده‌اند. شوکر و همکاران [۴۳] تأثیر داده‌کاوی را بر اثربخشی آموزش آنلاین بررسی کردند. بولدو و آگگان [۴۴] به داده‌کاوی آموزشی بر روی پایگاه‌های داده در استانبول ترکیه پرداختند. آنها از قوانین انجمنی استفاده نمودند و خاطر نشان کردند با استفاده از نتایج تحقیق می‌توان مهارت‌های دانشجویان را شناسایی و عملکرد آنان را بهبود بخشید. عبدالله، هراوان، احمد و دریس [۴۵] رویکردی برای پشتیبانی از قوانین انجمنی در داده‌کاوی آموزشی ارائه دادند. اعتقاد این پژوهشگران بر این است که با استفاده از نتایج داده‌کاوی آموزشی می‌توان استانداردها و مدیریت آموزشی فعلی را بررسی کرد و ارتقا بخشید. همسا و همکاران [۴۶] از الگوریتم ژنتیک فازی و درخت تصمیم برای پیش‌بینی نمره ریاضی دانشجویان لیسانس و ارشد گروه کامپیوتر استفاده کردند و معتقدند روش‌های سیستماتیک می‌توانند به دلیل پیش‌بینی زودهنگام، عملکرد دانشجویان در امتحان پایانی را بهبود بخشند.

همکاری و هماهنگی بین استاد و دانشجو

تحقیقاتی به جهت همکاری و هماهنگی بین استاد و دانشجو در طول اجرای فعالیت‌های آموزشی انجام شده‌است؛ مانند، گیتوئی [۴۷] الگوریتمی هوشمند برای مشاوره انتخاب واحد دانشجو برمبنای تحلیل چارت دروس و پیش‌بینی نمرات طراحی و تحلیل کرده‌اند. هوآنگ و فنگ [۱۵] عملکرد دوره را براساس عملکرد دانشجویان در دوره‌های پیش‌نیاز و امتحانات میان ترم پیش‌بینی کرده‌اند. الگامل [۴۸] براساس فاکتورهای مختلف از جمله پیشینه ریاضی دانشجو، مهارت حل مسأله و ... به پیش‌بینی عملکرد تحصیلی با استفاده از قوانین انجمنی پرداختند. در نتیجه‌گیری این مقاله آمده‌است که می‌توان با استفاده از نتایج داده‌کاوی آموزشی، فرایند برنامه‌ریزی تحصیلی را تغییر داد و تأثیرات این تغییر را ارزیابی نمود. آسیف، مرکرون، علی و حیدر [۴] عملکرد دانشجویان دوره کارشناسی را بررسی کردند. ابتدا دستاوردهای دانشجو را در پایان برنامه چهار ساله پیش‌بینی کردند؛ سپس پیشرفت‌های معمول را مطالعه نمودند و آنها را با نتایج پیش‌بینی ترکیب کردند. با استفاده از درخت تصمیم و الگوریتم k-means به مدل‌سازی پرداختند و با استفاده از این مدل می‌توان به دانشجویان با عملکرد بالا مشاوره و فرصت داد و به دانشجویان با عملکرد پایین به موقع هشدار داد و پشتیبانی کرد.

خلاصه‌ای از مقالات در جدول ۱ (در انتهای مقاله)، آمده و شاخص‌های تأثیرگذار در جدول آورده شده است. شاخص‌های مذکور در جدول، مطابق شاخص‌های استفاده شده در این مقاله است. پژوهش پیش رو را می‌توان در هر چهار دسته جای داد. در دسته اول است؛ زیرا با پیش‌بینی نمرات دانشجویان در هر نیمسال می‌توان الگوی رفتاری دانشجو در دروس تخصصی مهندسی صنایع را شناسایی کرد. در دسته دوم است؛ زیرا ابتدا با بررسی مقالات و با استفاده از نظر خبرگان، شاخص‌های تأثیرگذار بر نمرات دانشجویان استخراج می‌شوند و سپس، با استفاده از ابزار انتخاب ویژگی شاخص‌ها، موارد تأثیرگذار شناخته‌شدند و هیچ یک از مقالات تمام این شاخص‌ها را در کنار هم استفاده نکردند. در دسته سوم هم قرار می‌گیرد؛ زیرا می‌توان مهارت‌ها یا ضعف‌های دانشجو را در هر یک از دروس تشخیص داد و برای بهبود عملکرد دانشجویان راه‌حل ارائه نمود. در دسته چهارم قرار دارد؛ زیرا دانشجوها، استادان و مسئولین آموزشی در ابتدای نیمسال می‌توانند با همفکری یکدیگر با توجه به پیش‌بینی‌های صورت‌گرفته، برنامه‌ریزی بهتری برای طول ترم تحصیلی ارائه دهند.

در این پژوهش، از الگوریتم‌های داده‌کاوی برای پیش‌بینی نمرات دانشجویان مهندسی صنایع در دروس تخصصی استفاده شده‌است. علاوه بر شاخص‌های تأثیرگذار مذکور در سایر مقالات، دو شاخص استاد درس و نمره درس پیش‌نیاز، براساس نظر خبرگان به لیست شاخص‌ها اضافه شده و تمامی شاخص‌ها بر پیش‌بینی، مؤثر شناسایی شدند. براساس نتایج، کارایی و دقت الگوریتم‌ها در پیش‌بینی نمرات دانشجویان، به اثبات رسیده و مدلی برای بهبود فرایند انتخاب واحد دانشجویان مبتنی بر تکنیک‌های داده‌کاوی در این پژوهش ارائه شده‌است.

در این مقاله، پس از مقدمه، در بخش روش تحقیق، گام‌های پژوهش توضیح داده‌شده و در ادامه، نتایج و نتیجه‌گیری به تفصیل آمده‌است.

روش تحقیق

در این پژوهش، گام‌های تحقیق، مطابق با روش CRISP-DM تعیین شده‌است. مدل مرجع CRISP-DM برای داده‌کاوی یک نمای کلی از چرخه زندگی یک پروژه داده‌کاوی را نشان می‌دهد و شامل فازهای پروژه است: [۴۹]

- *شناخت سیستم:* اولین گام، بر اهداف و نیازمندی‌های پروژه از نمای کسب و کار تمرکز دارد. سپس این دانش را به تعریف یک مسأله داده‌کاوی تبدیل می‌کند و یک نقشه مقدماتی از نقشه طراحی می‌کند تا اهداف محقق شود.
- *شناخت داده‌ها:* این مرحله با یک مجموعه داده شروع می‌شود و با عملیات‌ها ادامه می‌یابد. به منظور آشنا شدن با داده‌ها، مسایل کیفیت داده را تعریف کرد و یک دید کلی نسبت به داده‌ها کشف می‌کند یا زیرمجموعه‌های جالبی را که مفروضات را برای اطلاعات پنهان شکل می‌دهد، پیدا می‌کند. یک رابطه بسیار نزدیک بین درک داده و درک کسب و کار وجود دارد. فرموله کردن مسأله داده‌کاوی و نقشه پروژه به حداقل مقدار درک از داده‌های در دسترس نیاز دارد.
- *آماده‌سازی داده‌ها:* مرحله آماده‌سازی داده‌ها، شامل همه فعالیت‌هایی است که پایگاه داده نهایی را می‌سازد. فعالیت‌هایی مانند مطرح کردن، ثبت کردن، انتخاب مشخصه، پالایش داده، ایجاد مشخصه‌های جدید و تغییر داده برای ابزارهای مدل‌سازی.
- مدل‌سازی: روش‌های مختلف مدل‌سازی انتخاب و به کار بسته می‌شود. پارامترهای آن‌ها اندازه‌گیری می‌شود تا مقادیر، بهینه شود. به طور معمول، چند روش برای یک نوع مسأله داده‌کاوی وجود دارد. بعضی روش‌ها به قالب‌های خاصی از داده نیاز دارد.
- ارزیابی: قبل از ادامه کار تا گسترش نهایی مدل، بسیار مهم است که مدل ارزیابی شود و گام‌های طی شده تا ساخت مدل بازبینی شود تا مطمئن شویم اهداف کسب و کار محقق می‌شود.
- *توسعه:* ساخت مدل معمولاً پایان پروژه نیست. اغلب نیاز است، دانش به دست آمده در مسیری که مشتری بتواند از آن استفاده کند، سازماندهی و ارائه شود.

شکل ۱، مراحل انجام پژوهش را نشان می‌دهد. در ادامه توضیح هر یک از مراحل به تفصیل آمده‌است که براساس روش CRISP، گام اول (تعیین اهداف پروژه) را می‌توان به عنوان شناخت سیستم در نظر گرفت؛ زیرا در این مرحله با هدفی که از انجام پروژه داریم آشنا می‌شویم. گام دوم (استخراج شاخص‌های تأثیرگذار) را می‌توان به عنوان شناخت داده‌ها در نظر گرفت؛ زیرا داده ابتدایی براساس این گام‌ها جمع‌آوری می‌شود. مرحله تدوین پایگاه داده، نیز دقیقا همان

کارهایی را انجام می‌دهد که در آماده‌سازی داده‌ها بدان اشاره شد. استفاده از ابزارهای مختلف داده‌کاوی مطابق مرحله مدل‌سازی است. در بررسی صحت پیش‌بینی به طور هم‌زمان هم مدل، ارزیابی می‌شود، هم می‌توان برای گسترش مدل تصمیم گرفت.

تعیین اهداف و دامنه پروژه

در این مقاله، داده‌کاوی آموزشی برای دانشجویان مهندسی صنایع دانشگاه تربت حیدریه انجام می‌شود تا بتوان با استفاده از این تکنیک، در هنگام انتخاب واحد، بازخورد اطلاعاتی به دانشجو و اساتید برای افزایش کارایی و پیشرفت تحصیلی ارائه داد.

استخراج شاخص‌های تأثیرگذار

انتخاب شاخص‌های مورد استفاده در پیش‌بینی از اهمیت ویژه‌ای برخوردار است. با مرور پژوهش‌های پیشین می‌توان متغیرهای کارآمد را شناسایی و گزینش کرد [۱۲]. بنابراین، با بررسی مقالات داده‌کاوی آموزشی، شاخص‌های تأثیرگذار در داده‌کاوی آموزشی استخراج شده‌است. علاوه بر شاخص‌های مذکور در مقالات، شاخص استاد درس، تکرار در اخذ واحد درسی و نمرات دروس پیش‌نیاز نیز براساس نظر خبرگان به شاخص‌ها اضافه شد. لیست کامل شاخص‌ها به همراه مشخصات، در جدول ۲ آمده‌است.

استفاده از الگوریتم‌های داده‌کاوی

پس از بررسی مقالات ارائه شده در این زمینه و با استفاده از نرم‌افزار وکا (WEKA) پس از پیش‌پردازش داده‌ها (تخمین مقادیر از دست رفته و ...) با استفاده از انتخاب مشخصه از روش‌های نسبت بهره و بهره اطلاعات، تمامی شاخص‌ها انتخاب شدند. مزایای این روش‌ها، آسانی، سادگی، استقلال از مدل پیش‌بینی و آسانی فهم خروجی است [۵۰]. بنابراین، می‌توان نتیجه گرفت شاخص‌های مذکور، در پیش‌بینی نمرات دانشجویان تأثیرگذار هستند. سپس، از الگوریتم‌های مختلف دسته‌بندی برای پیش‌بینی نمرات استفاده شده که معرفی هر یک در ادامه مطالب آمده‌است.

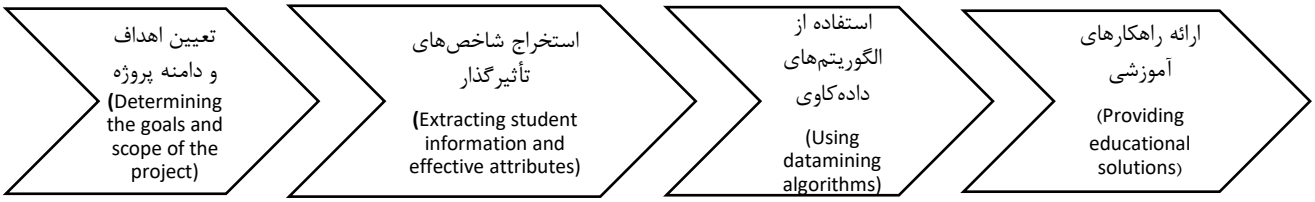
- *شبکه بیزی:* این الگوریتم، مدل گرافیکی است که می‌تواند توزیع احتمال شرطی را توصیف کند. دسته‌بندهای بیزی، الگوریتم‌های دسته‌بندی متداولی هستند که با توجه به سادگی، کارایی محاسباتی عملکرد خوبی در مسایل دنیای واقعی دارند. مزیت دیگر این مدل‌ها در این است که سرعت آموزش و ارزیابی بالایی دارند و از درصد صحت قابل توجهی برخوردارند [۵۱].
- *لوجیت بوست:* این الگوریتم در چند سال اخیر گسترش یافته‌است. Boosting ارائه شده تا چند الگوریتم را ترکیب می‌کند و عمکرد پیش‌بینی را بهبود می‌بخشد [۵۲].
- *پارت (PART):* از این الگوریتم برای کشف قوانین مختلف و شناسایی دانش و الگو استفاده می‌شود [۵۳].
- *درخت تصمیم J48:* درخت‌های تصمیم ابزارهای قوی و متداولی برای دسته‌بندی هستند. مزیت درختان تصمیم در این است که قواعدی را ایجاد می‌کنند که به راحتی قابل فهم و تفسیر است. نیاز به آماده‌سازی پیچیده داده‌ها ندارند و برای متغیرهای عددی و بازه‌ای عملکرد مناسب دارند [۵۱].

در این پژوهش، تقسیم‌بندی کل داده‌ها به مجموعه داده‌های آموزشی و آزمایشی به روش اعتبارسنجی k-باره(K-fold) انجام شده‌است. در این روش، داده‌ها به k دسته تقسیم می‌شوند و k-1 دسته برای آموزش و دسته kام برای آزمایش استفاده می‌شود و تا زمانی‌که همه دسته‌ها به عنوان دسته آزمایش به‌کار روند؛ این روند ادامه می‌یابد. برآورد دقت الگوریتم‌ها در این روش از سایر روش‌ها بسیار بیشتر است [۵۴]. در این مقاله، مقدار k، برابر ۱۰ در نظر گرفته‌شده‌است. الگوریتم‌های مختلف براساس شاخص‌های صحت و مساحت منحنی مشخصه عملکرد با یکدیگر مقایسه شدند. در ادامه تعریف هر یک آمده‌است و در جدول ۳، مقالاتی که از این شاخص‌ها برای ارزیابی استفاده کردند مشخص شده‌است.

- *صحت:* یکی از معیارهای ارزیابی مدل‌های دسته‌بندی است که مقدار آن برابر درصد مشاهداتی است که توسط روش مورد استفاده، به درستی دسته بندی شده است [۵۵].
- *نمودار مشخصه عملکرد (ROC):* مساحت زیر منحنی ROC یک شاخص ترکیبی است که نشان می‌دهد مدل با چه احتمالی موقعیت مثبت را نسبت به موقعیت منفی انتخاب می‌کند. بیشترین حد این شاخص ۱ و کمترین آن ۰/۵ است [۵۶]. این شاخص، کارایی الگوریتم را نشان می‌دهد.

ارائه راهکارهای آموزشی

با استفاده از بازخوردهایی که از پایگاه داده به هنگام پیش انتخاب واحد دانشجویان به مدیر گروه و دانشجویان ارائه می‌شود؛ می‌توان راهکارهایی برای بهبود وضعیت تحصیلی دانشجویان و خدمات آموزشی ارائه داد. برای این کار، در دروسی که نسبت دانشجویان با عملکرد ضعیف زیاد هستند می‌توان با استفاده از الگوریتم رگرسیون لجستیک، شاخص‌های مؤثر را شناسایی و براساس آن‌ها راهکار ارائه داد. رگرسیون لجستیک ابزاری برای توصیف روابط میان یک متغیر اسمی و مجموعه‌ای از متغیرهای پیش‌بینی‌کننده (شاخص‌های مؤثر) است. این تکنیک زمانی که متغیر برچسب محدود به دو یا چند مقدار اسمی باشد؛ قابل استفاده است. رگرسیون لجستیک فرمولی را بدست می‌آورد که احتمال رخ دادن یک اتفاق را به عنوان تابعی از متغیرهای مستقل پیش‌بینی می‌کند [۵۷-۵۸].



شکل۱: مراحل انجام پژوهش
Fig. 1: Steps of research

جدول ۲: شاخص‌ها
Table 2: Attributes

ردیف (Sort)	شاخص‌ها (Attribute)		مقیاس (Scale)		وضعیت‌های قبل اطلاق (Applicable situations)
1	معدل	GPA	پیوسته	Continuous	[0,20]
2	کل واحدهای گذرانده	Total passed units	گسسته	Discrete	x≥0
3	تعداد ترم‌های مشروطی	Number of conditional terms	گسسته	Discrete	x≥0
4	نوع پذیرش	Type of admission	گسسته	Discrete	4 3 2 1
5	وضعیت تأهل	Marital status	دودویی	Binary	متاهل Married مجرد Single
6	جنسیت	Gender	دودویی	Binary	مرد Male زن Female
7	سال ورود به دانشگاه	University admission year	گسسته	Discrete	92 91 90 89
8	سن	Age	گسسته	Discrete	x≥18
9	محل زندگی	Living place	دودویی	Binary	2 1
10	ترم حاضر	Current semester	گسسته	Discrete	1≤x≤11
11	نمره درس پیشنیاز	Prerequisite course score	پیوسته	Continuous	[0,20]
12	استاد درس	instructor of the course	گسسته	Discrete	کدگذاری Codding (x≥1)
13	تکرار در اخذ واحد	Repeating the course	گسسته	Discrete	x≥1

جدول ۳: بررسی پیشینه پژوهش شاخص ارزیابی
Table 3: The literature review of evaluation index

	شاخص (Criteria)	منبع (Reference)
صحت	Accuracy	[51] [28] [27] [17] [23] [32] [24] [31] [26] [60] [21] [40] [47][22]
سطح زیر منحنی مشخصه عملکرد	Area under ROC	[32] [31] [37] [12]

نتایج و بحث

براساس نظر پژوهشگران و مقالات داده‌کاوی آموزشی بهتر است نمرات براساس بازه پیش‌بینی شوند [۵۹]. پس از آماده‌سازی داده‌ها، دو مجموعه داده برای هر یک از دروس تخصصی براساس متغیر پیش‌بینی تدوین شد که شامل موارد زیر است:

- دوکلاسه: برچسب متغیر پیش‌بینی به‌صورت a و b که به ترتیب شامل بازه‌های [۰-۱۰) و [۱۰-۲۰] در نظر گرفته شده‌است.
- چندکلاسه: برچسب متغیر پیش بینی به صورت a و b و c که به ترتیب شامل بازه‌های [۰-۱۰)، [۱۰-۱۷) و [۱۷-۲۰] در نظر گرفته شده‌است.

پایگاه‌های داده برای هر درس با الگوریتم‌های مختلف طبقه‌بندی آزمایش شدند. برای مقایسه حالت‌های دو و چندکلاسه، درصد صحت الگوریتم‌های مختلف بین دروس میانگین گرفته و نتایج مطابق شکل ۲ است.

همان‌طور که در شکل ۲ نشان داده شده، حالت دوکلاسه از صحت بیشتری برخوردار است. ولی در صورت نیاز سیستم به بررسی دقیق‌تر نمرات، حالت چندکلاسه کارایی بیشتری دارد؛ زیرا اطلاعات دقیق‌تری از عملکرد دانشجو ارائه می‌کند. حال برای بررسی دقیق‌تر و مقایسه بین الگوریتم‌ها، از دو شاخص نرخ صحت و منحنی ROC در جدول ۳ استفاده شده‌است.

همان‌طور که در جدول ۳ مشـــاهده می‌شـــود؛ در تمامی دروس و در حالت دو و چندکلاسه الگوریتم Logit Boost در هر دو شاخص بهتر از سـایر الگوریتم‌ها عمل کرده‌اسـت. با توجه به مقادیر جدول ۴، توانایی الگوریتم‌های داده‌کاوی در ســیســـتم آموزش دانشـــگاه به اثبات رسـیده‌اسـت. مقالاتی که در حوزه داده‌کاوی آموزشـی کار شـده‌اسـت؛ کارایی و صحت پیش‌بینی الگوریتم‌های داده‌کاوی را تصدیق می‌کنند. برای نمونـه امین بیـدختی و همکـاران (۱۳۹۶) [۱۲] و راچبوری و همکاران (۲۰۱۸) [۲۶] به این نتیجه رســـیدند که می‌توان از نتایج داده‌کاوی آموزشی در جهت پیش‌بینی عملکرد تحصیلی دانشجویان و کمک به بهبود عملکرد آنان استفاده کرد. بنابراین می‌توان روند انتخاب واحد را بهبود داد و از نتایج داده‌کاوی آموزشـی در این فرایند به جهت افزایش کارایی آموزش و پیشـرفت تحصـیلی دانشجویان بهره برد. روند انتخاب واحد پیشـنهادی و سـنتی در شـکل ۳ آمده‌اسـت. همان‌طور که در این شـــکل نشـــان داده‌شـــده، مدل‌های داده‌کاوی، پایگاه‌های داده دانشجویان و نظرات خبرگان در روند انتخاب واحد لحاظ شده‌است.

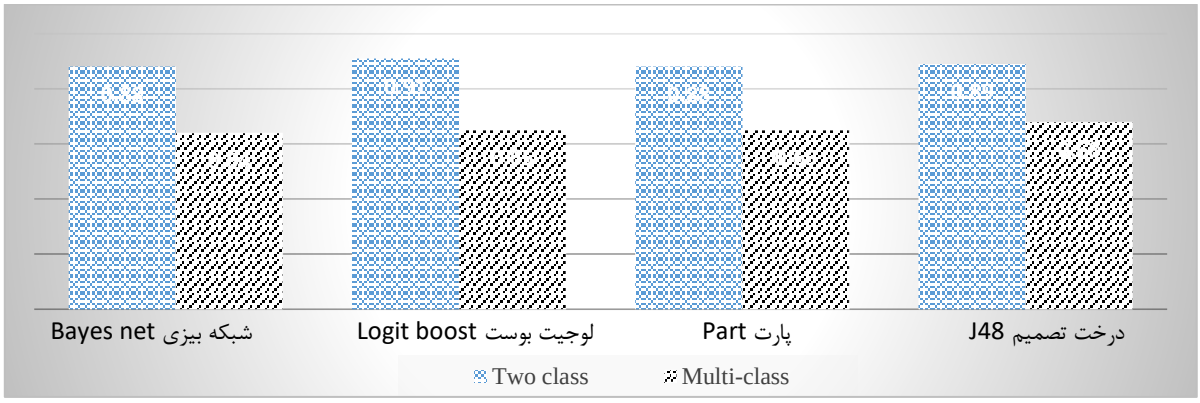
در مرحله داده‌کاوی آموزشـــی، اطلاعات دانشـــجویان (کســـانی که پیش‌ثبت‌نام انجام دادند) وارد پایگاه داده دروس مربوطه می‌شـــود. ســـپس با اســـتفاده از الگوریتم Logit Boost عملکرد دانشـــجویان پیش‌بینی می‌شود و نتایج به دانشجو و مدیرگروه برای تصمیم‌گیری و ارتقاء عملکرد دانشجویان، اطلاع‌رسانی خواهدشد. مدیر گروه، عملکرد دانشجویان در دروس تخصصی را بررسی و نسبت دانشجویان با عمکرد ضعیف را محاسبه می‌کند. در دروسی که نسبت زیادی از دانشجویان از عملکرد ضعیفی برخوردارند، با استفاده از الگوریتم رگرسیون لجستیک شاخص مؤثر، شناسایی و براساس آن راهکارهای آموزشی ارائه می‌شود. برای مثال، اگر شـــاخص نمره پیش‌نیاز، بر عملکرد دانشـــجویان تأثیر دارد؛ از اســـتاد درس مربوطه تقاضـــا شـــود تا قســـمت‌های مهم درس پیش‌نیاز در ترم آتی مرور شـود. یا اگر شـاخص اسـتاد درس از اهمیت بالایی برخوردار است از استادان با تجربه و نمره ارزشیابی بالا در گروه برای درس مورد نظر استفاده شود.

با توجه به عملکرد دانشجویان و بررسـی شـاخص‌های مؤثر بر عملکرد، می‌تـوان راهـکارهایی ارائـه داد که از مراجـعه به خـبرگان دانشـگاهی استخراج شده و شامل موارد زیر است:

- بررسی افزایش زمان کلاس ۳ ساعته به ۴ ساعت در سقف قانونی و ارائه به مدیر آموزش؛
- تصمیم جهت برگزاری کلاس حل تمرین برای درس؛
- استفاده از استادان با تجربه در گروه آموزشی؛
- استفاده از استادان با نمره ارزشیابی بالاتر؛
- گذاشـتن دوره‌های مکمل نرم‌افزار و آموزش مفاهیم توسـط انجمن علمی جهت افزایش یادگیری؛
- برگزاری بازدید از صنایع مرتبط با درس؛
- پیشنهاد مرور بخش‌های مهم دروس پیش‌نیاز در صورت نیاز؛
- بررسی عملکرد مدرس دروس پیش‌نیاز؛
- بررسی مجدد نحوه تدریس، سرفصل طرح درس و بازنگری در طرح درس؛
- تخصـیص زمان مناسـب به دروس بحرانی و اسـتفاده از روزهای آزاد دانشجویان جهت تمرکز بیشتر.

از این مدل می‌توان به عنوان یک ابزار پشـــتیبانی در ســـیســـتم‌های آموزش دانشگاه استفاده کرد. مزایای مدل پیشنهادی و راهکارهای ارائه شده در این مقاله عبارتند از:

- جلوگیری از افت تحصیلی دانشجویان؛
- شناسایی شاخص‌های مؤثر بر عملکرد تحصیلی؛
- مشاوره به‌موقع به دانشجویان؛
- شناسایی نقاط قوت و ضعف در عملکرد تحصیلی؛
- تخصیص مناسب زمان و هزینه جهت کارایی بیشتر؛
- کاهش هزینه‌های فرایندی سیستم دانشگاه؛
- افزایش مزیت رقابتی در بین سایر گروه‌ها و دانشگاه‌ها؛
- نزدیک شدن به سطح استاندارد آکادمیک؛
- افزایش رضایت‌مندی دانشجویان؛
- تقویت پایه علمی و افزایش موفقیت در آزمون ورودی کارشـناسـی ارشد؛
- افزایش سطح اطلاعات و آمادگی بیشتر جهت ورود به بازار کار.



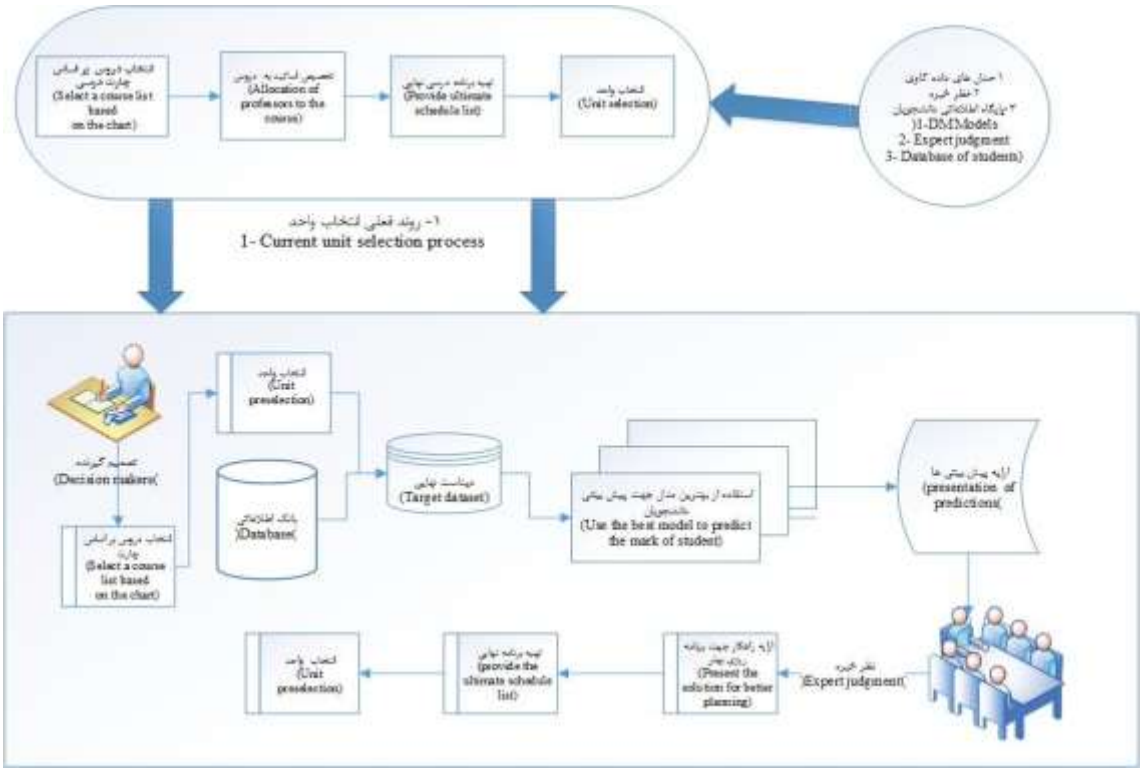
شکل ۲: میانگین نرخ صحت

Fig. 2: Average of accuracy

جدول ۴: مقایسه الگوریتم‌ها در حالت دو و چندکلاسه

Table 4: Comparison of algorithms in two class and multi-class

درس (Course)		الگوریتم‌ها (algorithms)		دوکلاسه (Two class)		چندکلاسه (Multi-class)	
				صحت (Accuracy)	سطح زیر منحنی مشخصه عملکرد (ROC)	صحت (Accuracy)	سطح زیر منحنی مشخصه عملکرد (ROC)
اقتصادمهندسی	Engineering economics	شبکه بیزی	Bayes Net	0.89	0.721	0.71	0.599
		لوجیت بوست	Logit Boost	0.89	0.759	0.74	0.666
		پارت	PART	0.88	0.622	0.66	0.576
		درخت تصمیم	J48	0.88	0.572	0.71	0.532
کنترل موجودی	Inventory control	شبکه بیزی	Bayes Net	0.83	0.675	0.56	0.575
		لوجیت بوست	Logit Boost	0.85	0.735	0.67	0.655
		پارت	PART	0.8	0.613	0.62	0.643
		درخت تصمیم	J48	0.83	0.538	0.66	0.653
طرح‌رزی واحدهای صنعتی	Facility layout	شبکه بیزی	Bayes Net	0.99	0.5	0.61	0.557
		لوجیت بوست	Logit Boost	0.99	0.5	0.67	0.703
		پارت	PART	0.99	0.5	0.65	0.663
		درخت تصمیم	J48	0.99	0.5	0.63	0.631
تحقیق در عملیات ۱	Operations research 1	شبکه بیزی	Bayes Net	0.95	0.5	0.76	0.445
		لوجیت بوست	Logit Boost	0.95	0.603	0.8	0.709
		پارت	PART	0.9	0.5	0.72	0.675
		درخت تصمیم	J48	0.95	0.5	0.76	0.549
تحقیق در عملیات ۲	Operations research 2	شبکه بیزی	Bayes Net	0.76	0.579	0.47	0.557
		لوجیت بوست	Logit Boost	0.84	0.796	0.62	0.697
		پارت	PART	0.79	0.674	0.51	0.592
		درخت تصمیم	J48	0.8	0.538	0.57	0.603
احتمال	Probability	شبکه بیزی	Bayes Net	0.81	0.493	0.64	0.531
		لوجیت بوست	Logit Boost	0.89	0.74	0.72	0.616
		پارت	PART	0.83	0.576	0.65	0.611
		درخت تصمیم	J48	0.84	0.548	0.67	0.582
کنترل پروژه	Project control	شبکه بیزی	Bayes Net	0.98	0.145	0.72	0.658
		لوجیت بوست	Logit Boost	0.98	0.336	0.74	0.703
		پارت	PART	0.97	0.135	0.64	0.647
		درخت تصمیم	J48	0.98	0.145	0.59	0.612
کنترل کیفیت	Quality control	شبکه بیزی	Bayes Net	0.87	0.836	0.57	0.61
		لوجیت بوست	Logit Boost	0.91	0.841	0.77	0.711
		پارت	PART	0.9	0.72	0.71	0.595
		درخت تصمیم	J48	0.9	0.793	0.74	0.588
آمار	Statistics	شبکه بیزی	Bayes Net	0.84	0.523	0.74	0.511
		لوجیت بوست	Logit Boost	0.85	0.691	0.77	0.63
		پارت	PART	0.84	0.579	0.7	0.594
		درخت تصمیم	J48	0.82	0.502	0.75	0.562



شکل ۳: مقایسه انتخاب واحد پیشنهادی با روش فعلی

Fig. 3: Comparison of the proposed unit selection with the current method

نتیجه‌گیری

در این مقاله، به پیش‌بینی عملکرد تحصیلی دانشجویان پرداختیم. جامعه آماری مورد مطالعه، داتشجویان کارشناسی مهندسی صنایع دانشگاه تربت حیدریه هستند. ابتدا شاخص‌های مؤثر بر عملکرد تحصیلی دانشجویان از مقالات مرتبط استخراج و چند شاخص جدید نیز براساس نظر خبرگان به لیست شاخص‌ها اضافه شد. با استفاده از اطلاعات دموگرافیک و تحصیلی دانشجویان، پایگاه داده برای هر یک از دروس تخصصی مهندسی صنایع تدوین گردید. سه نوع پایگاه داده بر حسب متغیر برچسب تشکیل شد. یک پایگاه دوکلاسه و یک پایگاه سه کلاسه، که با توجه به نرخ صحت پیش‌بینی پایگاه دوکلاسه از صحت بالاتری برخوردار بود. پس از بررسی شاخص‌های نرخ صحت و منحنی مشخصه عملکرد، الگوریتم Logit Boost از عملکرد بهتری برخوردار بود. بنابراین توانایی و کارایی این الگوریتم در پیش‌بینی عملکرد تحصیلی دانشجویان به اثبات رسید. با استفاده از نظر خبرگان، روند جدیدی برای انتخاب واحد دانشجویان با استفاده از داده‌کاوی آموزشی ارائه شد. در این روند، پس از انجام پیش‌انتخاب واحد دانشجویان، اطلاعات وارد پایگاه داده مربوطه می‌شود و عملکرد دانشجویان برای ترم آتی پیش‌بینی می‌شود.

تحقیقاتی همسو با پژوهش پیش رو انجام شده که توانایی و کارایی دانش داده‌کاوی را در پیش‌بینی و بهبود عملکرد تحصیلی دانشجویان به اثبات رسانده‌اند. مقالاتی که در مقدمه مقاله به آن‌ها اشاره شد گواه بر این قضیه است که داده‌کاوی آموزشی می‌تواند سبب ارتقا و پیشرفت تحصیلی دانشجویان شود. در این خصوص چهار رویکرد وجود دارد شامل شناسایی الگوی رفتاری دانشجو، شناسایی عوامل و شاخص‌های تأثیرگذار بر عملکرد تحصیلی، بهبود عملکرد تحصیلی و ایجاد همکاری و هماهنگی بین استاد و دانشجو که تمامی این موارد با داده‌کاوی آموزشی و نتایج مستخرج شده از آن امکان‌پذیر است.

در راستای اجرای روش پیشنهادی، ممکن است محدودیت‌هایی نیز وجود داشته باشد؛ از جمله:

- *جمع‌آوری اطلاعات:* برای تشکیل پایگاه داده نیاز به تعداد زیادی داده وجود دارد که این امر زمان‌بر است. برخی از این داده‌ها در سیستم آموزش موجود و برخی دیگر باید جمع‌آوری شوند و به درستی برای تشکیل یک پایگاه داده آماده برای داده‌کاوی کنار هم قرار بگیرند.
- *وجود داده‌های پرت:* با توجه به حجم بالای داده‌ها در سیستم آموزش ممکن است شرایط خاص یک دانشجو باعث به وجود آمدن داده پرت در سیستم شود. در صورتی که داده‌های پرت یا از دست رفته زیادی در پایگاه داده وجود داشته‌باشد منتج به نتایج اشتباه و گمراه‌کننده خواهدشد؛ که این امر با پیش‌پردازش صحیح و اصولی تا حدی قابل کنترل است.
- *محدودیت اجرای راهکارها:* در برخی از موارد راهکارهایی جهت بهبود شرایط و عملکرد تحصیلی دانشجویان ارائه می‌شود که با توجه به شرایط دانشگاه و استادان ممکن است در برهه‌زمانی خاصی، امکان‌پذیر نباشد؛ مانند افزایش ساعت تدریس، تغییر استاد درس و مواردی از این قبیل.

برای تحقیقات آتی در این زمینه می‌توان از نتایج ارزشیابی اساتید در روند داده‌کاوی استفاده نمود و شاخص‌های علمی و اخلاقی اساتید را در پایگاه داده در نظر گرفت.

مشارکت نویسندگان

این مقاله با مشارکت و همکاری نویسندگان نگارش شده است و تمامی نویسندگان نقش یکسانی در نگارش مقاله داشته‌اند. ارایه ایده اولیه، جمع‌آوری داده‌ها، اجرای مدل‌های داده‌کاوی، تحلیل و آنالیزنتایج، نگارش مقاله و نظارت بر مسیر پژوهش از بخش‌های نگارش مقاله بوده است.

تشکر و قدردانی

این مقاله مستخرج از نتایج طرح تحقیقاتی اجرا شده به شماره قرارداد ۱۳ از محل اعتبارات پژوهشی دانشگاه تربت‌حیدریه است که به این وسیله تشکر و قدردانی می‌گردد.

تعارض منافع

این پژوهش با حمایت مالی دانشگاه تربت‌حیدریه در قالب طرح پژوهشی انجام شده است.

منابع و مآخذ

[1] Pokay P, Blumenfeld PC. Predicting achievement early and late in the semester: the role of motivation and use of learning strategies. *Journal of Educational Psychology*. 1990; 82(1): 41-50.

[2] Grudnitski G. A forecast of achievement from student profile data. *Journal of Accounting Education. 1997;* 15(4): 549-558.

[3] Ransdell S. Predicting college success: the importance of ability and non-cognitive variables. *International Journal of Educational Research. 2001;* 35 (4): 357-364.

[4] Asif R, Merceron A, Ali SA, Haider NG. Analyzing undergraduate students' performance using educational data mining. Computers & Education. 2017; 113: 177-194.

[5] Rostami M, Ayat S, Saghari F, Yaghoobi F. [Applying fuzzy clustering to assess and anticipate students' educational progress in learning environments]. *Technology of Education Journal.* 2015; 10(1): 23-36. Persian.

[6] Akour I. *Factors influencing faculty computer literacy and use in Jordan: A multivariate analysis. D.B.A. dissertation.* US: Louisiana Tech University; 2010.

[7] Sarboland K. [Providing an e-learning model on teachers' satisfaction of learning in Ardabil Islamic Azad University]. *Technology of Education Journal.* 2019; 13(3): 603-614. Persian.

[8] Romero C, Ventura S. Data mining in education. *Wiley Interdiscilinary Review: Data Mining Knowledge Discovery*. 2013; 3(1): 12-27.

[9] Romero C, Ventura S. Educational data mining: A Survey from 1995 to 2005. *Expert Systems with Applications.* 2007; 33: 135-146.

[10] Yang M. *Data Mining Techniques Applied to Texas Woman's University's Enrollment data – What Can the Data Tell us?* [master's thesis]. US: Texaz Woman's University; 2006.

[11] Baker R. Data mining for Education. *International Encyclopedia of Education.* 2010; 7(3): 112-118.

[12] Aminbeidokhti A, Fathian Boroojeni M, Nameni A. [A Neural Network Based Model for Predicting Educational Vulnerability of Undergraduate Students]. *Journal of Management and Planning In Educational Systems*. 2017; 10(18): 81-102. Persian.

[13] Ranjan J, Malik K. Effective educational process: a datamining approach. *Journal of Information and Knowledge Management Systems.* 2007; 37(4): 502-515.

[14] Romero C, Ventura S, Garcia E. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education.* 2008; 51: 368-384.

[15] Huang S, Fang N. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*. 2013; 61: 133-145.

[16] Zhang Y, Oussena S, Clark T, Kim H. *Use data mining to improve student retention in higher education – A case study*. Paper presented at ICEIS. Funchal, Portugal; 2010.

[17] Sen B, Ucar E. Evaluating the achievements of computer engineering department of distance education students with data mining methods. *Procedia Technology*. 2012; 1: 262-267.

[18] Pena-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*. 2014; 41: 1426-1432.

[19] Shahiri AM, Husain W, Rashid NA. A Review on Predicting Student's Performance using Data Mining Techniques. The Third Information Systems International Conference. *Procedia Computer Science.* 2015; 72: 414-422.

[20] Rodrigues MW, Zarate LE, Isotani S. Educational Data Mining: A review of evaluation process in the e-learning, *Telematics and Informatics.* 2018; 35(6): 1701-1717.

[21] Heydari S, Yaghini M. [Classification and prediction of students' educational status using data mining techniques]. *Higher Educaion Letter.* 2011; 12: 107-124. Persian.

[22] Maghsoudi B, Sulaimany S, Amiri A, Afsharchi M. Teaching [Quality Improvement of electronic learning systems using educational data mining]. *Technology of Education Journal.* 2013; 6(4): 277-286. Persian.

[35] Yahya AA. Swarm intelligence-based approach for educational data classification. *Journal of King Saud University – Computer and Information Sciences.* 2018; 31: 35-51.

[36] Santana MA, Costa EB, Fonseca B, Rego J, Araujo FF. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior.* 2017; 73: 247-256.

[37] Fernandes EP, Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. V. (2018). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research.* 2018; 94(C):335-343.

[38] Wanli X, Rui G, Eva P, Sean G. Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*. 2015; 47: 168-181.

[39] Tari M, Minai B, Farahi A, Niknam Pirzadeh M. *Prediction of educational performance using CART algorithm, neural network and predictive accuracy increase using combination model.* Paper presented in the 3rd Iranian Conference on Electrical and Electronics Engineering (ICEEE2011), Gonabad; 2011. Persian.

[40] Ahmadi A, Karimzadgan D, Khairati Kazeroon T. [Data mining of Students Withdrawal at University of Tehran, Focusing on Fee Paid Students (to prevent customer churn)]. *Journal of Information Technology Managementd.* 2015; 7(2): 217-238. Persian.

[41] Rahmati A, Lesani M, Khalilzadeh R. *The related factors to the students of Shahid Bahonar Kerman in 2009-2010 and its analytical model* [master's theses] Kerman: Shahid Bahonar University; 2012. Persian.

[42] Deypir M, Raboo A. [Using Educational Data Mining for Grouping Learners in an E-Learning Environment for Customizing Learning Program]. *Journal of Management and Planning in Educational Systems*. 2018; 11(1): 83-108. Persian

[43] Shukor NA, Tasir Z, Meijden HV. *An examination of online learning effectiveness using data mining.* Papere presented in the Global Conference on Business & Social Science. Kuala Lumpur; 2014.

[44] Buldu A, Ucgun K. Data mining application on students' data. *Procedia Social and Behavioral Sciences*. 2010; 2: 5251-5259.

[45] Abdullah Z, Herawan T, Ahmad N, Deris MM. Mining significant association rules from educational data using critical relative support approach. *Procedia - Social and Behavioral Sciences*. 2011; 28: 97-101.

[46] Hamsa H, Indiradevi S, Kizhakkethottam JJ. Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Global Colloquium in recent Advancement and Effectual Researches in Engineering, Science

[23] Sen B, Ucar E, Delen D. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*. 2012; 39: 9468-9476.

[24] Natek S, Zwilling M. Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications*. 2014; 41(14): 6400-6407.

[25] Strecht P, Cruz L, Soares C, Merdes-Moreria J, Abren R. *A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance*. Paper presented at the 8th International Conference on Educational Data Mining. Madrid, Spain; 2015.

[26] Rachburee N, Punlumjeak W, Rugtanom S, Jaithavil D, Pracha M. A Prediction of Engineering Students Performance from Core Engineering Course using Classification. *Computer Science and Engineering*. 2018; 6(7): 43-48.

[27] Nghe TN, Janecek P, Haddawy P. *A Comparative Analysis of Techniques for Predicting Academic Performance*. Paper presented in the 37th ASEE /IEEE Frontiers in Education Conference; 2007.

[28] Kabakchieva D, Stefanova K, Kismov VS. *Analyzing University Data for Determining Student Profiles and Predicting Performance*. Paper presented in the 4th International Conference on Educational Data Mining. The Netherlands; 2011.

[29] Kabakchieva D. Predicting Student Performance by Using Data Mining Methods for classification. *Cybernetics and Information Technologies. 2013;* 13(1): 61-72.

[30] Oskouei RJ, Askari M. Predicting Academic Performance with Applying Data Mining Techniques (Generalizing the results of two Different Case Studies). *Computer Engineering and Applications Journal.* 2014; 79-88.

[31] Yehuala MA. Application of Data Mining Techniques for Student Success and Failure Prediction (The Case of Debre Markos University). *International Journal of Scientific & Technology Research*. 2015; 4(4): 91-94.

[32] Kaur P, Singh M, Josan GS. Classification and prediction based data mining algorithms to predict slow learners in education sector. In 3th International Conference on Recent Trends in Computing 2015(ICRTC-2015). *Procedia Computer Science*. 2015; 57: 500-508.

[33] Pandey M, Taruna S. Towards the integration of multipleclassifier pertaining to the Student'sperformance prediction. *Perspectives in Science*. 2016; 8: 364-366.

[34] Abu Saa A. Educational Data Mining & Students' Performance Prediction. *International Journal of Advanced Computer Science and Applications. 2016;* 5(7): 212-220.

[58] Asif R, Merceron A, Pathan M. Predicting student academic performance at degree level: A case study. *Intelligent Systems and Applications*. 2015; 1: 49-61.

معرفی نویسندگان

AUTHOR(S) BIOSKETHES

محمد قدوسی کارشناسی ارشد مهندسی صنایع و فارغ‌التحصیل دانشگاه صنعتی شریف بوده و در حال حاضر روی روش‌های داده کاوی و بهینه سازی سیستم‌های اموزشی به کمک روش‌های داده‌کاوی به عنوان زمینه تخصصی پژوهش می‌کنند.

Ghodoosi, M. Instructor, Industrial Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran

M.ghodoosi@torbath.ac.ir

فاطمه میرسعیدی فارغ‌التحصیل کارشناسی ارشد مهندسی صنایع از دانشگاه صنعتی سجاد هستند و روی کاربرد روش‌های داده‌کاوی در سیستم‌های آموزشی و مدیریت استراتژیک سازمان‌ها تحقیقات خود را انجام می‌دهند.

Mirsaeedi, F. Lecturer, Industrial Engineering, University of Torbat Hydarieh, Torbat Heydarieh, Iran

Fatemeh.mirsaeedi@yahoo.com

حمیدرضا کوشا فارغ‌التحصیل دکتری مهندسی صنایع از دانشگاه تربیت مدرس هستند. ایشان مقاطع کارشناسی و ارشد خود را در دانشگاه صنعتی شریف در رشته مهندسی صنایع گذارانده‌اند و روی روش‌ها و تکنیک‌های داده‌کاوی و کاربرد آن‌ها، مشغول به پژوهش هستند.

Koosha, M. Assistant Professor, Industrial Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

koosha@um.ac.ir

and Technology (RAEREST 2016). *Procedia Technology*. 2016; 25: 326-332.

[47] Gitue A. *The intelligent algorithm is to advise on the selection of student units based on the analysis of the chart of courses, the prediction of the grades and the pattern chosen by the previous students* [master's thesis]. Sanandaj: University of Kordestan; 2014. Persian.

[48] ElGamal AF. An Educational Data Mining Model for Predicting Student Performance in Programming Course. *International Journal of Computer Applications.* 2013; 70(17): 22-28.

[49] Wirth R, Hipp J. *CRISP-DM: Towards a Standard Process Model for Data Mining*. Paper presented in the 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining. Manchester, UK; 2000.

[50] Bagherzadeh F, Ramezankhani A, Azizi F, Hadaegh F, Khalili D. A tutorial on variable selection for clinical prediction models: Feature selection methods in data-mining could improve the results. *Journal of Clinical Epidemiology.* 2015; 71: 76-85.

[51] Cai YD, Chou KC. Using LogitBoost classifier to predict protein structural classes. *Journal Theoretical Biology*. 2006; 238(1): 172-176.

[52] Setayesh M, Fatahi Nafchi H, Abaspoor S, Roostayi M. [Providing a New Approach to Issuing Audit Report Using Data Mining (Case Study: Companies Listed in Tehran Stock Exchange)]. *Audit Science.* 2014; 19 (57): 5-26. Persian.

[53] Toloui Ashlaghi A, Nikoumaram H, Maghdoori Sharbiani F. Classification of applicants for credit facilities of banks using Support Vector Machine. *Management Researches*. 2010; 84: 1-19. Persian.

[54] Han J, Kamber M, Pei J. *Data Mining: Concepts and Techniques*. US: Morgan Kaufman; 2012.

[55] Alimohammadi AM. Abbasimehr MH, Javaheri A. Prediction of Stock Return Using Financial Ratios: A Decision Tree Approach. *Journal of Financial Management Strategy.* 2016; 3(4): 125-146.

[56] Hand D, Mannila H, Smyth P. *Principles of Data Mining*. Cambridge: MIT Press; 2001.

[57] Giudici P. *Applied data mining: statistical methods for business and industry*. New York: John Wiley & Sons; 2003.

Citation (Vancoure): Ghodoosi M, Mirsaeedi F, Koosha H. [Predicting and analyzing the performance of students through data mining techniques to improve academic performance]. *Tech. Edu. J.* 2020; 14(4): 821-834

http://dx.doi.org/10.22061/jte.2019.4902.2134

COPYRIGHTS

©2020 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.

جدول ۱: شاخص‌های در نظر گرفته شده در پژوهش‌ها

Table 1: Attribute considered in researches

منبع (Reference)	شاخص‌ها (Attributes)													
	معدل (GPA)	کل واحدهای گذرانده (Total passed units)	تعداد ترم‌های مشروطی (Number of conditional terms)	نوع پذیرش (Type of admission)	وضعیت تأهل (Marital status)	جنسیت (Gender)	سال ورود به دانشگاه (University admission year)	سن (Age)	محل زندگی (Living place)	ترم حاضر (Current semester)	نمره درس پیشنیاز (Prerequisite course score)	استاد درس (instructor of the course)	تکرار در اخذ واحد (Repeat the course	سایر (Other)
[21]	*	*	*	*		*	*	*						*
[39]	*	*		*			*	*	*	*				*
[22]						*		*						*
[47]														*
[40]	*	*	*		*	*	*	*	*					*
[5]		*			*	*		*	*					*
[12]	*	*	*											*
[27]						*		*						*
[28]	*					*		*	*	*				*
[17]						*		*						*
[23]						*								*
[29]	*					*		*	*	*				*
[24]						*		*					❋	*
[32]								*						*
[31]						*		*						*
[25]				*	*	*		*						*
[46]	*													*
[34]	*					*		*	*					*
[36]						*		*	*	*				*
[26]														*
[37]						*		*	*					*
Current article	*	*	*	*	*	*	*	*	*	*	*	*	*	