فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Fall 2022, 3 (3), 53-63
DOR: 20.1001.1.27832570.1401.3.3.5.7

Received: 16 July 2022
Accepted: 27 Aug 2022

مقاله پژوهشی

A Novel Semi-Supervised Approach to Improve the Performance of E-learning

Farhad Gharebaghi[1], Ali Amiri[2*]

1. Ph.D. Student, Department of Computer Engineering, Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran.
farhad.gharebaghi@iauz.ac.ir
2. Associate Professor, Computer Engineering Department, University of Zanjan, Zanjan, Iran.
(Corresponding Author) a_amiri@znu.ac.ir

Abstract:
Introduction: Academic success of students is one of the important goals in educational environments. In recent years, with the rapid proliferation of information technology in the field of education has led to a new model called "electronic learning". Nowadays, the application of this technology in education has made the education process go beyond time, geographical and political limitations. Providing effective training is an example benefit of e-learning. Considering an appropriate teaching method improves the performance of learners in the educational environment. In traditional education, there is a direct interaction between the learner and the teacher, which makes it possible for the teacher to adapt the teaching method to the learner's conditions. In electronic education, the absence of this direct interaction demands methods for the personalization of education. In this paper, using the data mining approach, the knowledge level of the learners is determined. This knowledge can be observed and collected by the system. Intelligent educational systems use a model of the learner that represents the level of literacy and skill of the learner in a specific field, and use it to analyze the learner's inputs to the system during educational interactions.
Method: In this paper, LP-MLTSVM semi-supervised learning method is used to improve the learning quality and the satisfaction of learners in e-learning. The LP-MLTSVM algorithm is a semi-supervised algorithm developed based on the support vector machine algorithm. The proposed model creates a multi-class classification. In the proposed model, the data of the electronic education center of Semin have been used. This center has about five thousand members, in which e-learning courses are conducted. In this method, all the training data are used in the construction of the model, with the difference that there is no need to label all the samples. In the proposed model, only 20% of the samples are labeled by an expert. 75% of the labeled data were used for training and 25% of them were used for testing the models.
Results: To evaluate the effectiveness of using the proposed system in the held course, the criteria of "academic success" and "academic satisfaction" have been used. In each of the groups, at the end of the course, a comprehensive test of 40 questions was presented to the learners for evaluation. The purpose of this test is to measure the academic success of the students in the mentioned groups. This work was done in order to check the effect of the proposed groups in the system. To explore the effect of these groups, the one-way analysis of variance was used and the amount of difference between the results of the groups was checked. In order to check the learners' satisfaction with the methods used in the first, second and third groups, at the end of each session, a questionnaire containing four questions was presented to the learners. These questions have 5 options, option 1 indicates lack of satisfaction and option 5 indicates complete satisfaction with the course. The results show the success of the proposed method.
Discussion: In the proposed method, the characteristics of the learners were investigated and suitable characteristics were created in order to predict the class variable which was the knowledge level of the learners. Then the research data was collected and solved using the proposed method. The results of the proposed method and the existing methods were compared with each other, and according to the evaluation criteria, the obtained results show that the proposed method is better. Finally, in order to evaluate, the presented model was used in a virtual course. The results of the course show the success and academic satisfaction of the students of the course with the proposed model.

Keywords: E-learning, Data Mining, Supervised Learning, Semi-Supervised Learning.

فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران

ارائه یک رویکرد شبه‌نظارتی برای بهبود عملکرد آموزش الکترونیکی

تاریخ دریافت: ۱۴۰۱/۰۴/۲۵
تاریخ پذیرش: ۱۴۰۱/۰۶/۰۵

دورهٔ سوم، پاییز ۱۴۰۱
شمارهٔ سوم، صص: ۵۳-۶۳

فرهاد قره باغی[۱]، علی امیری[۲*]

۱- دانشجوی دکتری، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران.
farhad.gharebaghi@iauz.ac.ir
۲- دانشیار، گروه مهندسی کامپیوتر، دانشگاه زنجان، زنجان، ایران. (نویسندهٔ مسئول) a_amiri@znu.ac.ir

چکیده: در عصر اطلاعات جهت‌گیری بسیاری از سامانه‌ها به سمت شخصی‌سازی اطلاعات برای کاربر است. یادگیری الکترونیکی نیز از این قاعده مستثنی نبوده و یکی از ملزومات آن وجود مکانیزمی شخصی‌شده برای کمک به یادگیری مؤثر یادگیرنده است. در سال‌های اخیر از روش‌های داده‌کاوی به‌طور گسترده در زمینه یادگیری الکترونیکی استفاده‌شده‌است. درواقع محققین مختلف با استفاده از روش‌های داده‌کاوی سعی در شناخت هرچه بیشتر یادگیرندگان و در نتیجه ایجاد یادگیری تطبیقی داشته‌اند. در تحقیقات مختلف از یادگیری نظارت‌شده برای شناخت یادگیرندگان استفاده‌شده‌است. با توجه به معایب الگوریتم های یادگیری نظارت‌شده در این تحقیق جهت شناسایی سطح دانش یادگیرندگان یک مدل داده‌کاوی شبه نظارتی پیشنهادشده‌است تا بتواند با شخصی‌سازی و هوشمندسازی محیط یادگیری الکترونیکی، عملکرد یادگیری و رضایت‌مندی یادگیرندگان را بهبود بخشد. در این مقاله با استفاده از رویکرد داده‌کاوی شبه نظارتی به شناسایی سطح دانش یادگیرندگان پرداخته‌شده‌است. به این منظور برای ایجاد مدل پیشنهادی از الگوریتم شبه نظارتی LP-MLTSVM استفاده‌شده‌است. در ساخت مدل پیشنهادی از داده‌های واقعی استفاده‌شده‌است. به منظور ارزیابی، در یک دوره مجازی از مدل پیشنهادشده استفاده‌شد. نتایج حاصل از دوره موفقیت و رضایت تحصیلی یادگیرندگان با مدل پیشنهادی را نشان‌می‌دهد.

واژه‌های کلیدی: یادگیری الکترونیکی، داده‌کاوی، یادگیری نظارت‌شده، یادگیری شبه نظارتی.

۱. مقدمه

موفقیت تحصیلی آموزش‌گیرندگان از اهداف مهم در محیط‌های آموزشی است. طی سال‌های اخیر، با رشد و توسعه سریع فنّاوری اطلاعات درعرصه‌های آموزش، الگوی جدیدی تحت عنوان «یادگیری الکترونیکی» پدیدآمده‌است[۱]. امروزه به‌کارگیری این فنّاوری در آموزش باعث‌شده که فرایند آموزش بتواند ورای محدودیت‌های زمانی، جغرافیایی و سیاسی جریان‌یابد[۲]. افزایش امکانات آموزشی برای یادگیرندگان در محیط‌های یادگیری حضوری، کم‌کردن هزینه آموزش، روزآمدتر کردن محتوای درسی، امکان ارائه مطالب درسی با رویکرد «بگاه[۱]»، فراهم‌آوردن آموزش اثربخش، نمونه‌هایی از مزایای یادگیری الکترونیکی است[۲]. لحاظ‌کردن یک شیوه مناسب آموزش باعث بهبود عملکرد فراگیران در محیط آموزشی می‌شود[۳].

در آموزش سنتی تعامل مستقیمی بین یادگیرنده و آموزش‌دهنده وجوددارد که این امکان را فراهم‌می‌نماید که آموزش‌دهنده نحوه آموزش را با شرایط یادگیرنده تطبیق‌دهد. در آموزش الکترونیکی عدم وجود این تعامل مستقیم نیازمند روش‌هایی برای شخصی‌سازی و هوشمندسازی آموزش است[۵،۴].

در سال‌های اخیر از روش‌های داده‌کاوی به‌طور گسترده به منظور شخصی‌سازی و هوشمندسازی محیط یادگیری الکترونیکی استفاده‌شده‌است[۸،۷،۶]. درواقع محققین با استفاده از روش‌های داده‌کاوی سعی در شناخت هرچه بیشتر یادگیرندگان و در نتیجه ایجاد یادگیری تطبیقی داشته‌اند. داده‌کاوی فرایند تشخیص الگوهای معتبر، نو، مفید و نهایتاً قابل‌درک در داده‌ها است. کار روش‌های داده‌کاوی، انتخاب داده و ساخت مدلی است که واقعیت را نشان دهد. مدل مذکور الگوها و ارتباطات میان داده‌ها را شرح‌می‌دهد[۹]. طبقه‌بندی یکی از شناخته‌شده‌ترین وظایف داده‌کاوی است[۱۰]. از طبقه‌بندی به‌عنوان ابزاری جهت پیش‌بینی استفاده‌می‌شود. هدف طبقه‌بندی ابتدا این است که داده‌های آموزش را تحلیل و سپس مدل یا توصیف دقیقی برای هر کلاس با استفاده از مشخصه‌های در دسترس ارائه‌دهد. چنین توصیف کلاسی سپس درجهت طبقه‌بندی مجموعه داده‌های آزمون مورد استفاده قرارخواهدگرفت[۱۰].

تکنیک‌های یادگیری در داده‌کاوی به صورت‌های نظارت‌شده، نظارت نشده و شبه‌نظارتی تقسیم‌می‌شوند[۱۱]. در یادگیری نظارت‌شده، داده‌های آموزشی به کمک شخص خبره برچسب‌گذاری می‌شوند[۱۲]. خروجی این مدل ایجاد طبقه‌بند خواهدبود. درخت تصمیم و ماشین بردار پشتیبان نمونه‌هایی از این روش داده‌کاوی است[۱۴،۱۳]. یادگیری نظارت‌نشده به‌دنبال یافتن گروه‌هایی از داده‌های مشابه است بدون اینکه از قبل پیش‌بینی در مورد شباهت‌های موجود داشته باشد[۱۵]. از یادگیری نظارت‌نشده جهت خوشه‌بندی داده‌ها استفاده‌می‌شود. الگوریتم k میانگین نمونه مهمی از این روش داده‌کاوی است[۱۶]. یادگیری شبه‌نظارتی حالتی بین یادگیری نظارت‌شده و نظارت‌نشده است[۱۷]. باتوجه به حجم زیاد داده‌ها، برچسب‌گذاری کلیه داده‌های آموزشی مستلزم صرف زمان و هزینه بالایی خواهدبود[۱۸]. جهت صرفه‌جویی در انجام این کار از روش یادگیری شبه‌نظارتی استفاده‌می‌شود[۱۹]. در یادگیری شبه‌نظارتی از داده‌های برچسب‌دار و هم از داده‌های فاقد برچسب به صورت همزمان استفاده می‌شود تا دقت یادگیری بهبودیابد[۱۸].

در دهه‌های اخیر، یادگیری شبه‌نظارتی به عنوان یک مسیر جدید و اثرگذار در حوزه داده‌کاوی پدیدآمده‌است[۱۸]. این نوع یادگیری با بهره‌گیری از قابلیت‌های داده‌های برچسب‌دار در کنار حجم زیادی از داده‌های فاقد برچسب، فرایند آموزش مدل‌های داده‌کاوی را به میزان قابل‌توجهی بهبود بخشیده‌است[۱۷]. باتوجه به تازه‌بودن این روش در این مقاله جهت بهبود میزان یادگیری و رضایتمندی فراگیران در یادگیری الکترونیکی از روش یادگیری شبه‌نظارتی LP-MLTSVM استفاده‌شده‌است[۲۰].

در این مقاله با استفاده از رویکرد داده‌کاوی به شناسایی سطح دانش یادگیرندگان پرداخته‌شده‌است. امروزه در اکثر کاربردها و زمینه‌های مبتنی بر وب، «شخصی‌سازی» نقش مهمی ایفامی‌کند[۸]. شخصی‌سازی به فرایندی گفته‌می‌شود که در طی آن رفتار سامانه بر اساس دانشی که از کاربر دارد، تغییرمی‌یابد[۷]. این دانش می‌تواند توسط سامانه مشاهده و جمع‌آوری‌شود. سامانه‌های آموزشیار هوشمند مدلی از یادگیرنده را که نمایانگر سطح سواد و مهارت یادگیرنده در حوزه‌ی درسی است، به‌کارمی‌گیرند و از آن برای تحلیل ورودی‌های یادگیرنده به سامانه در طی تعاملات آموزشی استفاده‌می‌کنند.

به‌طور خلاصه اهداف این مقاله را می توان در موارد ذیل بیان‌نمود:

- استفاده از الگوریتم‌های داده‌کاوی شبه‌نظارتی به منظور ارتقاء دانش یادگیرندگان در محیط یادگیری الکترونیکی.
- شناسایی سطح دانش یادگیرندگان به منظور شخصی‌سازی و هوشمند سازی آموزش .
- روش ایجادشده در این تحقیق می‌تواند در محیط‌های یادگیری الکترونیکی به منظور تحقق یادگیری تطبیقی استفاده‌شود تا میزان رضایت و موفقیت تحصیلی یادگیرندگان افزایش‌یابد.

در روش پیشنهادی به بررسی مشخصه‌های یادگیرندگان پرداخته‌شده و از هشت مشخصه مناسب به منظور پیش‌بینی متغیر کلاس که همان سطح دانش یادگیرندگان است، استفاده‌شده‌است. پس از جمع‌آوری داده‌های تحقیق با استفاده از روش پیشنهادی به حل مسئله پرداخته‌شده‌است. نتایج حاصل از روش پیشنهادی و روش‌های موجود با یکدیگر مقایسه‌شده‌است که نتایج، نشان از بهتر بودن معیار دقت و انحراف معیار روش پیشنهادی دارد. در نهایت به منظور ارزیابی، مدل ارائه‌شده در

جدول ۱: خلاصه کارهای انجام‌شده

Ref. No	Author	Year	Techniques used	Highlighted contributions
[21]	Fok et al.	2018	Tensor flow deep learning	Evaluates the student's performance. The study involves the analysis of the influence of academic performance and their extra-curricular
[22]	Hamoud et al	2016	Tree, J48, Hoeffding	The performance of Portuguese students is studied
[23]	Huang et al	2011	SVM, MLP, RBF	Prediction of students' academic performance in an introductory engineering course
[24]	Raihana et al	2018	SVM	The study shows the importance of quality of life on academic performance.
[25]	Hasan et al	2018	Random Forest Tree algorithm	The study focuses in improving student's performance by early prediction
[26]	Maksud et al	2021	Artificial Neural Network(ANN), Naïve Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM)	Machine Learning Approaches to Digital Learning Performance Analysis
[27]	PanduRanga et al.,	2021	Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM) and k-Nearst Neighborhood (k-NN)	Student Classification Based on Cognitive Abilities and Predicting Learning Performances Using Machine Learning Models

دوره‌ای مجازی به‌کارگرفته‌شد. نتایج حاصل از دوره، موفقیت و رضایت تحصیلی یادگیرندگان با مدل پیشنهادی را نشان‌می‌دهد

۲. پیشینه پژوهش

در این بخش به برخی از روش‌هایی که با کار انجام‌شده در این مقاله بسیار مرتبط است پرداخته‌شده‌است.

فوک و همکاران [۲۱] از موتور تحلیلی یادگیری عمیق برای ارزیابی عملکرد دانش‌آموز استفاده‌کرده‌اند. این مطالعه شامل تجزیه و تحلیل تأثیر عملکرد تحصیلی و فعالیت‌های فوق برنامه آن‌ها مانند خدمات و هنر است. دقت روش پیشنهادشده بین ۸۰ تا ۹۱ درصد بود. در این تحقیق در مجموع از ۲۰۰۰ مجموعه داده استفاده‌شده‌است. در این آزمایش ۷۵درصد داده برای آموزشی و ۲۵درصد داده‌ها برای آزمایشی در نظرگرفته‌شده‌است. حمود و همکاران [۲۲] عملکرد دانش آموزان پرتغالی را موردمطالعه قراردادند. جهت انجام کار از درخت تصمیم استفاده‌شده‌است. در این تحقیق نتایج آزمایش با الگوریتم‌های J48 Hoeffding,و C4.5 مقایسه‌شده‌است. نتایج نشان‌داد که الگوریتم J48 از سایر طبقه‌بندها بهتر عمل‌می‌کند و پیش‌بینی دقیقی از دانشجویانی که می‌توانند دوره‌های آموزش عالی را با موفقیت به پایان برسانند، ارائه‌می‌دهد. هوانگ و همکاران[۲۳] تحلیل مقایسه‌ای رگرسیون را با استفاده ازالگوریتم‌های خطی چندگانه، شبکه MLP، وماشین بردار پشتیبان (SVM) برای کشف بهترین مدل برای پیش‌بینی عملکرد تحصیلی دانش‌آموزان مورد مطالعه قراردادند. نتیجه مدل‌ها، اثربخشی آن‌ها در پیش بینی نمرات کارشناسی دانش آموزان در امتحان نهایی را نشان‌می‌دهد. ریحانا و همکاران [۲۴] پیشنهاددادند که با استفاده از پارامترهای کیفیت زندگی و عملکرد تحصیلی دانش آموزان می‌توان با استفاده از ماشین بردار پشتیبان، دانش‌آموزان را طبقه‌بندی‌نمود. در انجام این تحقیق از اطلاعات ۶۰ دانش آموز استفاده‌شده‌است. حسن و همکاران [۲۵] با استفاده از اطلاعات دانشجویان و فعالیت صورت گرفته آن‌ها پیشنهاددادند که می‌توان با استفاده درخت تصمیم عملکرد دانشجویان را پیش‌بینی نمود. در این تحقیق از اطلاعات ۲۲ دانشجو استفاده‌شده‌است. برای شبیه‌سازی تحقیق از نرم‌افزار WEKA استفاده‌شده‌است. نتایج این شبیه‌سازی نشان‌می‌دهد الگوریتم درخت جنگل تصادفی دقت بهتری را نسبت به سایر الگوریتم‌های درخت تصمیم دارد. مقصود و همکاران [۲۶] با شناسایی دانشجویان در معرض خطر تلاش‌کرده‌اند که عملکرد تحصیلی دانشجویان را اصلاح نمایند. آن‌ها با استفاده از استخراج ویژگی‌های مورد نظر تلاش‌کرده‌اند به کمک شناسایی این دانشجویان عملکرد تحصیلی ارتقا یابد. نتایج این تحقیق نشان‌می‌دهد که روش پیشنهادشده از دقت ۹۴/۸۲ درصد برخوردار است. باندورانگا و همکاران [۲۷] روشی را پیشنهاددادند که بتوان سطح یادگیری دانشجویان را طبقه‌بندی‌نمود. جهت انجام کار از معیارهای علاقه، سطح دانش، استدلال و توانایی‌های موضوعی ۳۱۳ نفر از دانشجویان که از سال ۲۰۱۷ تا ۲۰۱۹ در حال تحصیل بوده‌اند استفاده‌شده‌است. در انجام روش پیشنهادی از الگوریتم‌های شبکه‌های بیزین، درخت C4.5، K نزدیکترین همسایه و ماشین بردار پشتیبان استفاده‌شده‌است. نتایج تحقیق نشان‌می‌دهد که ماشین بردار پشتیبان و K نزدیکترین همسایه بیشترین دقت را در انجام کار از خود نشان‌داده‌اند. مطالعات انجام‌شده نشان‌می‌دهد درخت تصمیم و ماشین بردار پشتیبان از جمله بهترین روش‌ها برای پیش‌بینی دانش فراگیران در سیستم‌های یادگیری الکترونیکی است که در سال‌های اخیر از آن‌ها زیاد استفاده‌شده‌است. روش پیشنهادشده در این مقاله با این دو روش‌ها مقایسه‌شده‌است.

۳. مدل پیشنهادی

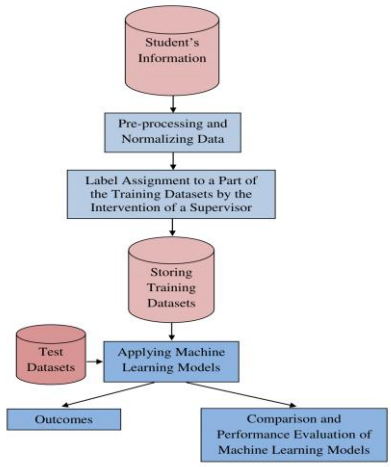
در مدل پیشنهادی ابتدا صفت‌هایی که دارای بیشترین رتبه هستند را انتخاب‌کردیم. جهت انجام این کار از ابزار موجود در نرم‌افزار WEKA کمک‌گرفتیم. پس از انجام این کار صفت‌های انتخاب‌شده را نرمال‌سازی نمودیم. با توجه به شبه‌نظارتی بودن مدل پیشنهادشده تنها بخش از نمونه‌های آموزشی را توسط شخص خبره برچسب‌گذاری نمودیم که این امر سبب صرفه‌جویی در مرحله برچسب‌گذاری داده‌ها می‌شود. با توجه به اینکه اکثر داده‌های آموزشی فاقد برچسب هستند جهت ساخت طبقه‌بند پیشنهادی از الگوریتم شبه‌نظارتی LP-MLTSVM استفاده‌گردید[۲۰]. الگوریتم LP-MLTSVM یک الگوریتم شبه‌نظارتی است که

بر پایه الگوریتم ماشین بردار پشتیبان توسعه‌یافته‌است. پس از ساخت طبقه‌بند، مدل پیشنهادی مورد ارزیابی قرارگرفت. جهت اندازه‌گیری میزان کارایی مدل پیشنهادی، مدل پیشنهادی را بر روی داده‌های آزمون اعمال‌نمودیم که نتایج حاصله موفقیت‌آمیز بودن این روش را نشان‌می-دهد. مراحل انجام مدل پیشنهادی در شکل (۱) بیان‌شده‌است.

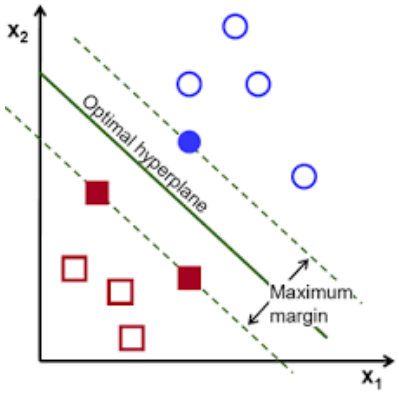
با توجه به اینکه روش پیشنهادشده در این مقاله با الگوریتم ماشین بردار پشتیبان[۱۴] و درخت تصمیم C4.5 [۲۸] مقایسه‌شده‌است در ادامه به تشریح این الگوریتم‌ها می‌پردازیم.

یکی از مدل‌های یادگیری ماشین نظارت‌شده الگوریتم ماشین بردار پشتیبان است که می‌تواند برای طبقه‌بندی مورداستفاده قرارگیرد. این روش از جمله روش‌های است که در سال‌های اخیر کارایی خوبی را نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان‌داده‌است. مبنای کار ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است که در تقسیم خطی داده‌ها سعی‌می‌کند خطی را انتخاب‌نماید که حاشیه اطمینان بیشتری داشته‌باشد. در این روش مسئله پیداکردن خط بهینه برای جداسازی داده‌هاست. شکل (۲) تصویری از طبقه‌بند ماشین بردار پشتیبان را نشان‌می‌دهد.

درخت تصمیم یکی از مشهورترین روش‌های نظارت‌شده جهت طبقه‌بندی در یادگیری ماشین است. درخت تصمیم با استفاده از تمامی نمونه‌های آموزشی و به کمک ساختار درختی، مدلی را جهت پیش‌بینی ایجادمی‌نماید. گره‌ها در درخت تصمیم نشان‌دهنده صفاتی هستند که به شاخه‌ها متصل‌بوده و منجر به تقسیم درخت شده و درخت را به گره‌های فرزند رهنمود می‌سازد. پیش‌بینی برچسب نمونه دیده نشده، با طی مسیر از گره ریشه و عبور از گره‌ها و شاخه‌ها تا رسیدن به یک گره برگ صورت‌می‌گیرد. الگوریتم‌های درخت تصمیم متعددی جهت طبقه-بندی داده‌های چندبرچسبی توسعه‌یافته‌است. نمونه‌ای از درخت‌های تعمیم‌یافته، درخت C4.5 است که به کمک تغییر در فرمول اینتروپی طبقه‌بندی را انجام‌می‌دهد.



شکل ۱: شمای مدل پیشنهادی



شکل ۲: طبقه‌بند ماشین بردار پشتیبان

۴. داده های تحقیق

در مدل پیشنهادی از داده‌های سامانه مرکز آموزش‌های الکترونیکی سمین استفاده‌شده‌است. این مرکز حدوداً دارای پنج هزار نفر عضو است. جهت ایجاد مدل پیشنهادی از ویژگی‌های جنسیت، شماره دانشجویی، کد دوره، کیفیت پاسخ به پرسش‌های پیش‌آزمون، زمان پاسخگویی به پرسش‌های پیش‌آزمون، تعداد پرسش‌های یادگیرندگان در دوره و مدت زمان مطالعه. داده‌های مربوط به این ویژگی‌ها نتایجی که در دوره‌های گذشته برگزارشده‌اند استخراج‌شده‌است. تعیین کلاس برچسب نمونه‌ها توسط شخص خبره صورت‌می‌پذیرد. باتوجه به حجم زیاد داده‌ها انجام این کار مستلزم صرف زمان و هزینه بالایی خواهد بود. جهت صرفه‌جویی در انجام این کار از روش یادگیری شبه‌نظارتی استفاده‌شده‌است. در این روش در ساخت مدل از تمامی داده‌های آموزشی استفاده‌می‌شود با این تفاوت که دیگر نیازی به برچسب گذاری کلیه نمونه‌ها نیست. در مدل پیشنهادی تنها ۲۰ درصد نمونه ها توسط شخص خبره برچسب گذاری شده است. از ۷۵ درصد داده‌های برچسب‌گذاری‌شده برای آموزش و از ۲۵ درصد آن‌ها جهت تست مدل‌ها استفاده‌شده‌است. باتوجه به نظارتی بودن سایر روش‌هایی که مدل پیشنهادی با آن‌ها مورد مقایسه قرارگرفته، این مدل‌ها تنها با این ۷۵ درصد نمونه های برچسب‌گذاری-شده آموزش‌داده‌شده‌اند. درحالی‌که در آموزش مدل پیشنهادی علاوه بر نمونه‌های برچسب‌گذاری از نمونه‌های فاقد برچسب نیز استفاده‌شده-است.

به منظور پیش بینی سطح دانش یادگیرندگان ، ویژگی‌های زیر در نظر گرفته‌شده‌است :

۱-کیفیت پاسخ به پرسش‌های پیش‌آزمون

یکی از مواردی که در این تحقیق مورد توجه قرارگرفته میزان آشنایی یادگیرندگان با محتویات دوره است. به این دلیل که برخی از یادگیرندگان با محتویات دوره آشنایی نسبتاً خوبی دارند و برخی نیز تا به حال با آن سروکار نداشته‌اند. در نتیجه با استفاده از پیش‌آزمون

می‌توان اطلاعات نسبی از دانش یادگیرندگان نسبت به محتویات دوره را سنجید.

یکی از مواردی که از پیش‌آزمون‌ها استخراج‌شده‌است کیفیت پاسخ یادگیرندگان به پرسش‌ها است. این مهم از طریق تعداد پاسخ‌های صحیح یادگیرندگان به پرسش‌ها حاصل‌می‌شود.

با توجه به تعداد پرسش‌های پیش‌آزمون به نرمال‌سازی این متغیر پرداخته‌شده‌است. به این صورت که تعداد پاسخ‌های صحیح یادگیرنده بر

جدول ۲: نرمال‌سازی کلاس داده‌ها

نمره	سطح دانش	کلاس
۰ -۱۰	بسیار پایین	۱
۱۰- ۲۰	پایین	۲
۲۰-۳۰	متوسط	۳
۳۰-۴۰	بالا	۴

تعداد کل پرسش‌های آزمون تقسیم‌شده‌است تا عددی در بازه صفر تا ۱ حاصل‌شود. به این ترتیب متغیر کیفیت پاسخ به پرسش‌های پیش‌آزمون

جدول۳: شرح ویژگی‌های مجموعه داده‌های آموزشی

شرح	نوع داده	ویژگی
خانم یا آقا	گسسته	جنسیت
عددی در بازه ۱ تا ۱۰۰	گسسته	کد دوره
برای هر دوره آموزشی کدی اختصاص داده شده است.	گسسته	کد دانشجویی
عددی در بازه صفر تا ۱	پیوسته	کیفیت پاسخ به پرسش‌های پیش آزمون
عددی در بازه صفر تا ۱	پیوسته	زمان پاسخگویی به پرسش‌های پیش‌آزمون
عددی در بازه صفر تا ۱	پیوسته	تعداد پرسش‌های یادگیرندگان در دوره
عددی در بازه صفر تا ۱ ۱ تا ۴	پیوسته	مدت زمان مطالعه
(بسیار پایین، پایین، متوسط، بالا)	گسسته	برچسب کلاس
۱ تا ۴ (بسیار پایین، پایین، متوسط، بالا)	گسسته	برچسب کلاس

ایجادمی‌شود.

۲-زمان پاسخگویی به پرسش‌های پیش‌آزمون

یکی دیگر از مواردی که از پیش‌آزمون‌ها استخراج‌شده‌است، مدت زمان پاسخگویی یادگیرنده به پرسش‌هاست. این متغیر نیز می‌تواند معیار نسبتاً خوبی از میزان آشنایی یادگیرندگان بر مطالب دوره باشد.

بر روی این متغیر نیز عمل نرمال‌سازی انجام شده‌است تا اعدادی بین صفر تا ۱ حاصل شود. به این ترتیب که تمامی زمان‌های یادگیرندگان بر بزرگترین زمان موجود در بین آن‌ها تقسیم‌شده تا اعداد مورد نظر حاصل‌شوند.

۳-تعداد پرسش‌های یادگیرندگان در دوره

یکی از برتری‌های یادگیری الکترونیکی نسبت به یادگیری حضوری این است که یادگیرندگان بدون نگرانی می‌توانند پرسش خود را مطرح‌کنند. به همین دلیل یکی از معیارهایی که می‌توان بر اساس آن میزان دانش یادگیرندگان از مطالب دوره را مدل کرد، تعداد پرسش‌های آن‌ها در طول دوره می‌باشد. به این صورت که یادگیرندگانی که تعداد پرسش بیشتری را مطرح‌می‌کنند معمولاً از یادگیرندگان دیگر آشنایی کمتری نسبت به مطالب دوره دارند. بر این اساس می‌توان این متغیر را به عنوان معیاری در این تحقیق درنظرگرفت. این متغیر نیز با استفاده از تقسیم مقادیر موجود بر بیشترین تعداد پرسش بین یادگیرندگان به اعدادی در بازه ۰ تا ۱ تبدیل‌شده‌است.

۴-مدت زمان مطالعه

یکی دیگر از متغیرهای متمایزکننده یادگیرندگان از یکدیگر مدت زمان مطالعه آن‌ها در جلسات مختلف می‌باشد. درواقع این مهم از مدت زمان سپری‌کردن آن‌ها در سامانه یادگیری الکترونیکی حاصل می‌شود. معمولاً یادگیرندگانی که آشنایی کمتری با محتوای دوره دارند مدت زمان بیشتری را در سامانه سپری‌می‌کنند تا مقصود مطالب را درک‌کنند. بنابراین این متغیر نیز می‌تواند تمایزی بین یادگیرندگان مختلف با سطوح دانش متفاوت ایجادکند. این متغیر نیز با استفاده از تقسیم مقادیر موجود بر بیشترین مدت زمان سپری‌شده توسط یادگیرندگان به اعدادی در بازه ۰ تا ۱ تبدیل‌شده‌است.

۵-سطح دانش یادگیرندگان

این متغیر که همان هدف مدل می‌باشد از طریق پس‌آزمونی که در انتهای دوره برگزارمی‌شود حاصل‌شده‌است. مقادیر مربوط به آن از نمرات یادگیرندگان در دوره‌های گذشته استخراج‌شده‌است و به عنوان متغیر هدف برای مدل قرارگرفته‌است.

نرمال‌سازی این مشخصه به این صورت بوده‌است که یادگیرندگان بر اساس نمراتی که دریافت‌کرده‌اند در چهار سطح دانش بسیار پایین، پایین، متوسط و بالا قرارگرفته‌اند. جدول (۲) نرمال‌سازی مربوط به این متغیر در چهار سطح مذکور را نشان‌می‌دهد:

جدول (۳) ویژگی‌های مجموعه داده را با جزئیات نشان‌می‌دهد. هر ویژگی با مقادیر اسمی، پیوسته یا گسسته توصیف‌می‌شود.

داده‌های جمع‌آوری‌شده برای آموزش مدل پیشنهادی، مربوط به ۵۰۰۰ نفر از یادگیرندگانی است که در گذشته دوره مورد نظر را به‌پایان رسانده‌اند و اطلاعات آن‌ها از سامانه مدیریت یادگیری احصاشده‌است.

۵. ارزیابی مدل پیشنهادی

به منظور ارزیابی مدل پیشنهادی و به کارگیری آن در دوره‌ای جدید، ۶۹ نفر از یادگیرندگان درنظرگرفته‌شده‌اند.

این افراد به‌طور تصادفی به سه گروه ۲۳ نفره تقسیم‌شده‌اند که مشخصات هر یک از گروه‌ها عبارتند از:

گروه اول: یادگیرندگانی که میزان دانش آن‌ها با استفاده از درخت تصمیم آموزش داده‌شده‌است.

گروه دوم: یادگیرندگانی که میزان دانش آن‌ها با استفاده از ماشین بردار پشتیبان آموزش داده‌شده‌است.

گروه سوم: یادگیرندگانی که میزان دانش آن‌ها با استفاده از الگوریتم پیشنهادی LP-MLTSVM آموزش داده‌شده‌است.

مدل آموزش دیده‌شده با روش پیشنهادی در دوره‌ای جدید استفاده-شده‌است و به پیش‌بینی میزان دانش یادگیرندگان از محتوای دوره پرداخته‌است. به این ترتیب با اطلاع از میزان آشنایی یادگیرندگان با مطالب، درس‌هایی با سطوح متفاوت به آن‌ها تخصیص داده‌می‌شود. داده‌های مربوط به مشخصه‌ها به منظور پیش‌بینی میزان دانش یادگیرندگان در دوره جدید به این صورت جمع‌آوری شده‌است که دو مشخصه کیفیت و زمان پاسخگویی به پرسش‌های پیش‌آزمون پس از برگزاری پیش‌آزمون یادگیرندگان حاصل‌شده‌اند. دو مشخصه تعداد پرسش‌های مطرح‌شده و مدت زمان مطالعه از سه جلسه ابتدایی برگزار-شده که به‌صورت عمومی به مباحث اولیه پرداخته‌شده، جمع‌آوری‌شده-است. به این ترتیب این دو مشخصه نیز در حد خوبی یادگیرندگان را از یکدیگر از نظر آشنایی با مطالب و میزان دانش متمایزمی‌کند. در نهایت با تشخیص سطح دانش یادگیرندگان از محتویات دوره، یادگیرندگانی با دانش بسیار پایین و پایین، درس‌های سطح پایین، یادگیرندگان با سطح دانش متوسط درس‌های با سطح متوسط و یادگیرندگان با سطح دانش بالا، درس‌های سطح بالا را در طول دوره دریافت‌می‌کنند.

برای ارزیابی میزان تأثیر استفاده از سامانه پیشنهادی در دوره برگزارشده، از معیارهای «موفقیت تحصیلی» و «رضایت تحصیلی» استفاده‌شده‌است که در ادامه به هر یک از آن‌ها پرداخته‌می‌شود.

۶. معیارهای ارزیابی مدل پیشنهادی

در این مدل پیشنهادی ما یک مسئله چهار کلاسه بسیار پایین ، پایین ، متوسط و بالا را بیان‌می‌کنیم. جهت ارزیابی مدل پیشنهادی بر اساس ماتریس درهم‌ریختگی جدول (۴) از معیارهای صحت، حساسیت، دقت و معیار اف استفاده‌شده‌است.

جدول ۴ : جدول درهم‌ریختگی

Classifier		True Value			
	Class	Fast (F)	Average (A)	Slow (S)	Very slow (V)
Predicted Values	Fast (F)	F-F	F-A	F-S	F-V
	Average (A)	A-F	A-A	A-S	A-V
	Slow (S)	S-F	S-A	S-S	S-V
	Very slow (V)	V-F	V-A	V-S	V-V

معیار صحت بیانگر نمونه هایی است که درست تشخیص-داده‌شده‌اند. این معیار با استفاده از قطر ماتریس درهم‌ریختگی به‌صورت زیر محاسبه-می‌شود:

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Nagative}{\sum Total\ Population}$$

حساسیت به معنی نسبتی از موارد مثبت است که آزمایش آن‌ها را به‌درستی به‌عنوان نمونه مثبت تشخیص‌داده‌است. این پارامتر به صورت زیر محاسبه می‌شود:

$$Recall\ or\ Sensitivity = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Nagative}$$

از معیار دقت برای بیان نسبت پاسخ‌های درست در هر دسته استفاده می‌شود. این معیار به‌صورت زیر محاسبه‌می‌شود:

$$Precision = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Positive}$$

معیار اف، معیار مهم دیگری است که از آن برای ارزیابی عملکرد طبقه‌بندها مورد استفاده قرارمی‌گیرد. این معیاراز ترکیب دو معیار حساسیت دقت حاصل‌می‌شود. معیار اف به صورت زیر تعریف‌می‌شود:

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

جهت ارزیابی مدل پیشنهادی، مدل پیشنهادی براساس معیارهای بیان شده فوق با الگوریتم‌های نظارتی، ماشین بردار پشتیبان و درخت تصمیم مقایسه‌شده‌است. نتایج شبیه سازی در جداول (۵) ، (۶) ، (۷) و(۸) بیان‌شده‌است.

برای ارزیابی، مدل پیشنهادی در یک دوره مجازی به کارگرفته‌شد. به این منظور جهت ارزیابی معیارهای موفقیت و رضایت تحصیلی یادگیرندگان استفاده‌شد.

جدول ۵: نتایج عملکرد الگوریتم شبه نظارتی LP-MLTSVM

Class	*Accuracy*	*Recall*	*Precision*	$F-Measure$
Very-Slow (V)	0.9315	0.9512	0.9981	0.9734
Slow (L)	0.9315	0.9506	0.93257	0.9414
Average (A)	0.9315	0.7033	0.9265	0.7996
Fast (H)	0.9315	0.9455	0.9057	0.9251
Avg	0.9315	0.8876	0.9407	0.9098

جدول ۶: نتایج عملکرد الگوریتم غیرنظارتی ماشین بردار پشتیبان

Class	*Accuracy*	*Recall*	*Precision*	$F-Measure$
Very-Slow (V)	0.9102	0.9321	0.9321	0.9321
Slow (L)	0.9102	0.9264	0.9264	0.9264
Average (A)	0.9102	0.6557	0.8642	0.7456
Fast (H)	0.9102	0.8608	0.9852	0.9188
Avg	0.9102	0.8437	0.9269	0.8807

جدول ۷: نتایج عملکرد الگوریتم غیرنظارتی درخت تصمیم

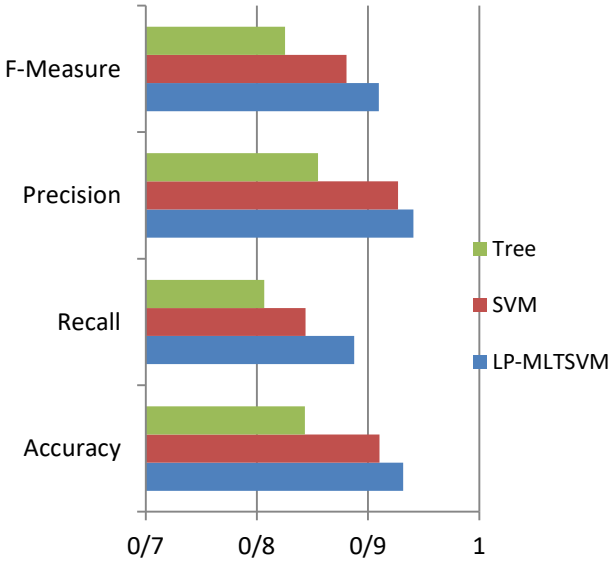
Class	*Accuracy*	*Recall*	*Precision*	$F-Measure$
Very-Slow (V)	0.8432	0.9324	0.9028	0.9173
Slow (L)	0.8432	0.8779	0.8264	0.8513
Average (A)	0.8432	0.8066	0.8357	0.8208
Fast (H)	0.8432	0.6092	0.8558	0.7117
Avg	0.8432	0.8065	0.8551	0.8252

جدول ۸: مقایسه عملکرد مدل‌ها

Class	*Accuracy*	*Recall*	*Precision*	$F-Measure$
Semi-supervised	0.9315	0.8876	0.9407	0.9098
Svm	0.9102	0.8437	0.9269	0.8807
Decision tree	0.8432	0.8065	0.8551	0.8252

۱،۶. موفقیت تحصیلی

در هر یک از گروه‌ها در پایان دوره، آزمون جامع ۴۰ پرسشی برای ارزیابی به یادگیرندگان ارائه‌شد. هدف از این آزمون سنجش میزان موفقیت تحصیلی یادگیرندگان در گروه‌های مذکور می‌باشد. میانگین و انحراف استاندارد نمرات گروه‌های مختلف در جدول(۹) آمده‌است. این کار به این منظور انجام شد که تأثیر گروه‌های پیشنهادی در سامانه بررسی شود. برای کاوش تأثیر این گروه‌ها از تحلیل واریانس یک‌طرفه [۲۹] استفاده و میزان تفاوت بین نتایج گروه‌ها در جدول(۱۰) بررسی شده است.



شکل ۳: مقایسه عملکرد مدل‌ها

جدول ۹: میانگین و انحراف استاندارد نمرات گروه‌ها

	گروه اول	گروه دوم	گروه سوم
تعداد یادگیرندگان	۲۳	۲۳	۲۳
میانگین نمرات	۲۲/۴	۲۴/۵	۲۵/۸
انحراف استاندارد	۵/۲	۷/۶	۴/۸

جدول(۱۰) نتایج تحلیل واریانس آزمون نهایی را نشان‌می‌دهد. با توجه به نتایج، نمرات یادگیرندگان در گروه اول در مقایسه با گروه دوم تفاوت معناداری ندارد؛ اما بین نمره‌های گروه‌های اول با سوم تفاوت معناداری مشاهده‌می‌شود. همچنین بین نتایج گروه‌های دوم و سوم نیز تفاوت معناداری مشاهده‌نمی‌شود. بنابراین یادگیرندگانی که در گروه سوم از روش پیشنهادی بهره‌برده‌اند بیشترین موفقیت تحصیلی را نسبت به دو گروه دیگر داشته‌اند.

جدول۱۰: نتایج تحلیل واریانس نمره‌های نهایی

گروه‌های مقایسه شده	میزان آماره‌ی F	مقدار P در سطح معناداری ۰/۰۵	وجود تفاوت معنی‌دار
گروه اول و دوم	۲/۸	۴/۱۷	خیر
گروه اول و سوم	۵/۳	۴/۱۷	بلی
گروه دوم و سوم	۱/۴	۴/۱۷	خیر

۲,۶. رضایت تحصیلی

به منظور بررسی رضایت یادگیرندگان از روش‌های به کار برده‌شده در گروه اول، دوم و سوم در پایان هر جلسه، پرسشنامه‌ای که حاوی چهار پرسش بود، به یادگیرندگان ارائه‌شد. این پرسش‌ها دارای ۵ گزینه است که گزینه ۱ نشان‌دهنده عدم‌رضایت و عدد ۵ رضایت کامل از دوره است. هر یک از این پرسش‌ها هدفی را دنبال‌می‌کنند و به دنبال اندازه‌گیری تأثیر هر یک از روش‌ها است که در جدول (۱۱) نشان داده شده‌است.

جدول ۱۱: صورت و هدف پرسش‌های رضایت تحصیلی

شماره پرسش	صورت پرسش	هدف پرسش
۱	تا چه میزان مطالب ارائه شده در این جلسه را آموختید؟	سطح مناسب درس‌ها
۲	آیا سطح دشواری درس‌های ارائه شده برای شما مناسب بود؟	دقت در شناسایی سطح دانش
۳	چه میزان از نحوه ارائه مفاهیم راضی هستید؟	سطح مناسب درس‌ها
۴	آیا تمایل دارید برای بار دیگری از این سامانه استفاده کنید؟	مطلوبیت و جذابیت دوره

پاسخ‌های یادگیرندگان به پرسش‌های پرسشنامه رضایت تحصیلی در پایان هر جلسه جمع‌آوری و میانگین نمرات هر گروه برای هر پرسش محاسبه شده‌است که این موارد درجدول(۱۲) آمده‌است. این مقادیر برای گروه‌ها، به کمک روش آماری تحلیل واریانس یک‌طرفه مورد بررسی قرارگرفته‌است که جدول (۱۳) نتایج آن را نشان می‌دهد.

همان‌طور که در جدول (۱۲) مشهود است میانگین نمرات پرسش‌های مطرح‌شده برای گروه سوم که از مدل پیشنهادی استفاده‌کرده‌اند در تمامی آن‌ها بیشتر است. اما میزان بیشتر بودن آن‌ها و همچنین انحراف استاندارد نمرات، تفاوت را بین گروه‌های مختلف مشخص‌می‌کند. به همین منظور برای بررسی تفاوت‌های معنی‌دار بین گروه‌های مختلف در اهداف متفاوت از آنالیز واریانس یک‌طرفه استفاده شده‌است. به این ترتیب مقادیر موجود در جدول (۱۳) این تفاوت‌ها را بررسی‌می‌کند.

جدول۱۲: پاسخ یادگیرندگان به پرسشنامه رضایت تحصیلی

شماره پرسش	معیار	گروه اول	گروه دوم	گروه سوم
۱	میانگین نمرات	۲/۸	۳/۳	۳/۷
	انحراف استاندارد	۰/۵۵	۰/۵	۰/۶۵
۲	میانگین نمرات	۲/۴	۲/۸	۳/۲
	انحراف استاندارد	۰/۶	۰/۷	۰/۵۵
۳	میانگین نمرات	۲/۸	۳/۲	۳/۸
	انحراف استاندارد	۰/۶	۰/۷	۰/۶۵
۴	میانگین نمرات	۲/۴	۲/۶	۲/۸
	انحراف استاندارد	۰/۷	۰/۵۵	۰/۵۵

همان‌طور که در جدول (۱۳) مشاهده‌می‌شود، پرسش نخست مربوط به سطح مناسب درس‌ها می‌باشد. در این زمینه مدل پیشنهادی با گروه اول و همچنین با گروه دوم مقایسه‌شد. نتایج نشان‌می‌دهند که یادگیرندگان گروه سوم که از سامانه پیشنهادی استفاده‌می‌کردند به طور معناداری رضایت بیشتری نسبت به گروه‌های اول و دوم دارند. این مسئله نشان از بهتر بودن دقت سامانه پیشنهادی در شناخت سطح دانش یادگیرندگان می‌باشد. این موضوع همچنین در پرسش‌های دوم و سوم نیز به همین ترتیب به اثبات می‌رسد. در پرسش چهارم پرسشی کلی در مورد مطلوبیت دوره مطرح‌می‌شود که هدف از آن سنجش میزان تأثیر سامانه پیشنهادی بر جذب یادگیرندگان می‌باشد. همانطور که نتایج نشان‌می‌دهند گروه دوم نتوانسته جذابیت قابل‌ملاحظه‌ای نسبت به گروه نخست که از هیچ یک از سامانه‌های استفاده‌نمی‌کند ایجادکند، اما گروه سوم که از سامانه پیشنهادی استفاده‌می‌کند موفق شده‌است مطلوبیت بیشتری نسبت به گروه اول برای یادگیرندگان فراهم‌آورد.

با توجه به نتایج مشاهده‌می‌شود که رضایت تحصیلی در گروه سوم در سطح مطلوبی قراردارد و استفاده از این روش، باعث بروز اختلاف معناداری در رضایت تحصیلی یادگیرندگان در مقایسه با دو گروه دیگر شده‌است.

جدول۱۳: نتایج مقایسه گروه‌ها بر مبنای پرسش‌های پرسشنامه رضایت تحصیلی

پرسش	هدف مورد بررسی	گروه‌های مقایسه شده	میزان آماره ی F	مقدار P در سطح معنادار ی ۰/۰۵	وجود تفاوت معنی‌دار
۱	سطح مناسب درس‌ها	دوم و سوم	۱۵/۱۶	۴/۱۷	بلی
	سطح مناسب درس‌ها	اول و سوم	۲۷/۶۵	۴/۱۷	بلی
۲	دقت در شناسایی سطح دانش	دوم و سوم	۸/۱۲	۴/۱۷	بلی
۳	سطح مناسب درس‌ها	دوم و سوم	۶/۴۷	۴/۱۷	بلی
۴	مطلوبیت و جذابیت دوره	اول و دوم	۳/۳۳	۴/۱۷	خیر
	مطلوبیت و جذابیت دوره	اول و سوم	۶/۵	۴/۱۷	بلی

به منظور ارزیابی مدل پیشنهادی، مدل در دوره‌ای واقعی به‌کارگرفته‌شد و نتایج آن‌ها با گروه‌هایی با آموزش متفاوت مقایسه‌شد. نتایج حاصل نشان از موفقیت و رضایت تحصیلی بیشتر یادگیرندگانی دارد که مدل پیشنهادی در آموزش آن‌ها به‌کاررفته‌است.

۷. نتیجه‌گیری

یادگیرندگان به شیوه‌های متفاوتی یادمی‌گیرند و یاددهندگان نیز به شیوه‌های متفاوتی آموزش‌می‌دهند، سامانه‌های یادگیری الکترونیکی کنونی در قیاس با آموزش و یادگیری سنتی دارای نقایصی جدّی هستند. به عنوان مثال می‌توان به فقدان تعامل مؤثر بین یادگیرنده و یاددهنده، عدم‌ارائه محتوا و بازخورد به صورت تطبیقی و شخصی‌شده اشاره‌کرد. بنابراین از حوزه‌های تحقیقاتی مهم در یادگیری الکترونیکی افزودن هوشمندی و تعامل به این سامانه‌ها و همچنین شخصی‌سازی محیط یادگیری است. تحقیق حاضر نیز درواقع به این جنبه‌های محیط یادگیری پرداخته و با استفاده از شناسایی سطح دانش یادگیرندگان سعی در شخصی‌سازی درس‌های ارائه‌شده به آن‌ها داشته‌است.

یکی از مهم‌ترین بخش‌های داده‌کاوی و به‌ویژه دسته‌بندی انتخاب مشخصه‌های مجموعه داده می‌باشد تا با استفاده از آن بتوان پیش‌بینی دقیق‌تری از متغیر کلاس مسئله انجام‌داد. در این تحقیق چهار مشخصه از اطلاعات یادگیرندگان به منظور پیش‌بینی میزان سطح دانش آن‌ها درنظرگرفته‌شد. نتایج حاصل از به‌کارگیری مدل پیشنهادی، موفقیت و رضایت تحصیلی بیشتر یادگیرندگان را نسبت به حالت‌های دیگر نشان‌می‌دهد که این خود گواهی بر درستی مشخصه‌های انتخابی مدل می‌باشد.

در این مقاله به شناسایی سطح دانش یادگیرندگان به منظور شخصی‌سازی آموزش پرداخته‌شد. در روش پیشنهادشده به بررسی مشخصه‌های یادگیرندگان پرداخته‌شد و چهار مشخصه مناسب به منظور پیش‌بینی متغیر کلاس که همان سطح دانش یادگیرندگان بود ، ایجادشد. سپس داده‌های تحقیق جمع‌آوری‌شده و با استفاده از روش پیشنهادی به حل آن پرداخته‌شد. نتایج حاصل از روش پیشنهادی و روش موجود در ادبیات تحقیق با یکدیگر مقایسه‌شد، که با توجه به معیارهای ارزیابی، نتایج به‌دست‌آمده نشان از بهتر بودن روش پیشنهادی دارد. در نهایت به منظور ارزیابی، مدل ارائه‌شده در دوره‌ای مجازی به‌کارگرفته‌شد. نتایج حاصل از دوره موفقیت و رضایت تحصیلی یادگیرندگان دوره با مدل پیشنهادی را نشان‌می‌دهد.

منابع

[1] N. Abdallah and O. Abdallah, "INVESTIGATING FACTORS AFFECTING STUDENTS' SATISFACTION WITH E-LEARNING: AN EMPIRICAL CASE STUDY," *J. Educ. Online*, vol. 19, no. 1, 2022, doi: 10.9743/jeo.2022.19.1.3.

[2] A. Z. Al Rawashdeh, E. Y. Mohammed, A. R. Al Arab, M. Alara, and B. Al-Rawashdeh, "Advantages and disadvantages of using E-learning in university education: Analyzing students' perspectives," *Electron. J. e-Learning*, vol. 19, no. 2, 2021, doi: 10.34190/ejel.19.3.2168.

[3] M. Hooda, "Learning Analytics Lens: Improving Quality of Higher Education," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, 2020, doi: 10.30534/ijeter/2020/24852020.

[4] A. Mustafa, "The personalization of e-learning systems with the contrast of strategic knowledge and learner's learning preferences: an investigatory analysis," *Appl. Comput. Informatics*, vol. 17, no. 1, 2021, doi: 10.1016/j.aci.2018.08.001.

[5] F. Essalmi, L. J. Ben Ayed, M. Jemni, Kinshuk, and S. Graf, "A fully personalization strategy of E-learning scenarios," *Comput. Human Behav.*, vol. 26, no. 4, 2010, doi: 10.1016/j.chb.2009.12.010.

[6] S. Rajper, N. A. Shaikh, Z. A. Shaikh, and G. A. Mallah, "Automatic Detection of Learning Styles on Learning Management Systems using Data Mining Technique," *Indian J. Sci. Technol.*, vol. 9, no. 15, 2016, doi: 10.17485/ijst/2016/v9i15/85959.

[7] W. Dwi Yuniarti, E. Winarko, and A. Musdholifah, “Data mining for student assessment in e-leaming: A survey,” 2020, doi: 10.1109/ICIC50835.2020.9288533.
[8] R. Krishnan *et al.*, “Smart Analysis of Learners Performance Using Learning Analytics for Improving Academic Progression: A Case Study Model,” *Sustain.*, vol. 14, no. 6, 2022, doi: 10.3390/su14063378.
[9] [9] M. Maggioni, “What is ... Data Mining,” *Bull. Am. Math. Soc.*, vol. 59, no. 4, 2012.
[10] S. Neelamegam and E. Ramaraj, “Classification algorithm in Data mining : An Overview,” *Int. J. P2P Netw. Trends Technol.*, vol. 4, no. 8, 2013.
[11] S. H. Liao, P. H. Chu, and P. Y. Hsiao, “Data mining techniques and applications - A decade review from 2000 to 2011,” *Expert Systems with Applications*, vol. 39, no. 12. 2012, doi: 10.1016/j.eswa.2012.02.063.
[12] V. Cherkassky and F. Mulier, *from Data: Concepts, Theory and Methods*. Wiley-IEEE Press, 2007.
[13] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, “Analysis of Various Decision Tree Algorithms for Classification in Data Mining,” *Int. J. Comput. Appl.*, vol. 163, no. 8, 2017, doi: 10.5120/ijca2017913660.
[14] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.
[15] M. E. Celebi and K. Aydin, *Unsupervised learning algorithms*. Springer, 2016.
[16] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, vol. 1, pp. 281–296.
[17] X. Goldberg, “Introduction to semi-supervised learning,” *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 6, pp. 1–116, 2009, doi: 10.2200/S00196ED1V01Y200906AIM006.
[18] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Mach. Learn.*, 2020, doi: 10.1007/s10994-019-05855-6.
[19] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning. Adaptive Computation and Machine Learning*. 2010.
[20] F. Gharebaghi and A. Amiri, “LP-MLTSVM: Laplacian Multi-Label Twin Support Vector Machine for Semi-Supervised Classification,” *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2021.3139929.
[21] W. W. T. Fok *et al.*, “Prediction model for students’ future development by deep learning and tensorflow artificial intelligence engine,” 2018, doi: 10.1109/INFOMAN.2018.8392818.
[22] A. K. Hamoud, “Selection of Best Decision Tree Algorithm for Prediction and Classification of Students’ Action,” *Am. Int. J. Res. Sci. Technol. Eng. Math.*, vol. 16, no. 1, 2016.
[23] S. Huang and N. Fang, “Work in progress - Prediction of students’ academic performance in an introductory engineering course,” 2011, doi: 10.1109/FIE.2011.6142729.
[24] Z. Raihana and A. M. Farah Nabilah, “Classification of students based on quality of life and academic performance by using support vector machine,” *J. Acad. Univ. Teknol. MARA Negeri Sembilan*, vol. 6, no. 1, 2018.
[25] R. Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarker, “Student Academic Performance Prediction by using Decision Tree Algorithm,” 2018, doi: 10.1109/ICCOINS.2018.8510600.
[26] M. Ahamad and N. Ahmad, “Machine Learning Approaches to Digital Learning Performance Analysis,”
Int. J. Comput. Digit. Syst., vol. 10, no. 1, 2021, doi: 10.12785/IJCDS/100187.
[27] T. P. R. Vital, K. Sangeeta, and K. K. Kumar, “Student classification based on cognitive abilities and predicting learning performances using machine learning models,” *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, 2021, doi: 10.12785/ijcds/100107.
[28] T. Korting, “C4. 5 algorithm and Multivariate Decision Trees,” *Image Process. Div. Natl. Inst. Sp.* , no. Section 2, 2010.
[29] QualtricsXM, “What is ANOVA (Analysis Of Variance) and what can I use it for?,” *Exp. Manag.*, 2020.