A Data Mining Approach to Determine Better Methods for Learners' Assessment in E-learning Courses

Mohammad Rezapour[1], Mohammad Mehdi Sepehri[2*], Hassan Rezapour[3]

1. Master of ICT Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran
2. Associate Professor of Industrial Engineering, Faculty of Engineering & technology, Tarbiat Modares University, Tehran, Iran
3. Ph.D. student of Applied Mathematics, Network Flow, Qom University, Qom, Iran

Receive: 29/4/2012 Accept: 23/7/2013

Determining how to assess learners and contents of e-learning is essential, which is conducted by a professor or teaching assistant. As a result they can determine student assessment methods, such as holding an online test or periodic homework assignment. If the organizers know the effect of each activity in the quality of learning, then besides saving considerable time and resources to stakeholders courses, they can transfer the useful content, make a realistic assessment of students and will ultimately improve e-learning.

In this paper, first we use unsupervised techniques of data mining for clustering, describe the present status of learners, extract the hidden rules in e-learning data using rule mining, and discover the effective contents in desired results. Next, using supervised methods, we predict the results of courses. Using real data of an electronic course provided and with designing four different methods for data sampling and training system, predictions are performed and the methods are validated with an accuracy rate of 92.86%.

Our findings show that the methods of this study can help teachers for a better understanding of learners and the impact of training activities required such as describing the characteristics of learners based on the discovery of hidden patterns in their acquired scores. They can also determine the most effective learning activities and decide about the real measures of learners.

Keywords: Data mining, E-learning, Pattern discovery, Clustering, Prediction.

* Corresponding Author's E-mail: mehdi.sepehri@modares.ac.ir

Archive of SID

www.SID.ir

تعیین روش‌های برتر سنجش فراگیران در دوره‌های یادگیری‌الکترونیکی - رویکرد داده‌کاوی

محمد رضاپور[1]، محمدمهدی سپهری[2]*، حسن رضاپور[3]

۱- کارشناس‌ارشد مهندسی فناوری اطلاعات و ارتباطات، دانشکده فنی و مهندسی دانشگاه تربیت مدرس، تهران، ایران
۲- دانشیار بخش مهندسی صنایع، دانشکده فنی و مهندسی دانشگاه تربیت مدرس، تهران، ایران
۳- دانشجوی دکتری ریاضی‌کاربردی، گرایش تحقیق در عملیات، دانشگاه قم، قم، ایران

دریافت: ۹۱/۲/۱۰ پذیرش: ۹۲/۵/۱

چکیده

چگونگی سنجش فراگیران و تعیین محتویات مطالب از فعالیت‌های ضروری فرایندهای آموزش الکترونیکی می‌باشد. این فعالیت‌ها توسط استاد یا آموزشیار انجام می‌پذیرد و روش‌های ارزیابی دانشجویان، نظیر برگزاری آزمون‌های برخط یا واگذاری تکالیف مدت‌دار به آن‌ها تعیین می‌شوند. در صورتی که برگزارکنندگان بتوانند از میزان تأثیر هر فعالیت در کیفیت یادگیری فراگیران آگاهی پیدا کنند، ضمن صرفه‌جویی قابل توجه در وقت و منابع ذینفعان دوره‌ها، باعث انتقال مطالب مفیدتر و سنجش واقعی‌تر دانشجویان و درنهایت بهبود آموزش‌الکترونیکی خواهد شد.

در این مقاله نخست با روش‌های فاقد ناظر داده‌کاوی به خوشه‌بندی و توصیف وضعیت موجود فراگیران پرداخته و به‌کمک قاعده‌کاوی، قواعد نهفته در داده‌های آموزش الکترونیکی را استخراج و محتواهای مؤثر در نتیجه مطلوب فراگیران را کشف می‌کنیم، سپس با روش‌های باناظر به پیش‌بینی نتایج دوره‌ها می‌پردازیم. از این رو با بهره‌گیری از داده‌های واقعی فعالیت‌های یک درس الکترونیکی ارائه شده و با طراحی چهار روش مختلف برای نمونه‌برداری داده‌ها و آموزش سیستم با دو درخت DT و WJ48، پیش‌بینی‌ها اجرا شدند و روش‌ها با نرخ دقت ۹۲/۸۶ درصد اعتبارسنجی شدند. در اینجا نشان داده‌ایم که اسلوب‌های این مطالعه می‌توانند به استاد درس برای شناخت بهتر فراگیران و تأثیر فعالیت‌های آموزشی خواسته شده از آنان نظیر- توصیف ویژگی‌های فراگیران برپایه کشف الگوهای نهفته در نمره‌های اکتسابی آنان کمک نمایند. تعیین فعالیت‌های مؤثرتر یادگیری و تصمیم‌گیری درباره شاخص‌های سنجش واقعی‌تر فراگیران از مزیت‌های مهمی است که روش‌های فناورانه این مطالعه برای بهبود مدیریت دوره‌های یادگیری الکترونیکی فراهم می‌کنند.

کلیدواژه‌ها: داده‌کاوی، آموزش‌الکترونیکی، کشف‌الگو، خوشه‌بندی و پیش‌بینی.

---

* نویسندهٔ مسؤول مقاله: E-mail: mehdi.sepehri@modares.ac.ir

Archive of SID

www.SID.ir

۱ – مقدمه و بیان مسئله

با توسعه سریع فناوری اطلاعات به‌ویژه در عرصه اینترنت، آموزش الکترونیکی به‌عنوان کاربردی مهم در مقوله آموزش مطرح شده است. با استفاده از فناوری‌های شبکه‌ای در مواردی نظیر زمان و هزینه‌های فراگیران صرفه‌جویی شده و موجب استقبال از این نحوه آموزش شده است. با اِعمال داده‌کاوی (DM)[1] بر داده‌های آموزش الکترونیکی می‌توان به تحلیل روش‌های آموزشی، خصوصیات محتوا و میزان فراگیری پرداخت و به این کاربرد داده‌کاوی که بر کشف الگوها از داده‌ها و فایل‌های ثبتی حاصل از وب دلالت دارد، وب‌کاوی (WM)[2] هم گفته می‌شود [۱]. اجرای الگوریتم‌های داده‌کاوی بر داده‌های مربوط به نمره‌های فراگیران آموزش الکترونیکی می‌تواند به‌عنوان یک واسط هوشمندسازی موجب سنجش واقعی‌تر فراگیران، انتخاب یا اصلاح محتوای درسی مؤثرتر و نیز تعیین الگوهای برتر برگزاری آزمون‌ها در LMS شود. گسترش پژوهش بر داده‌های ورودی فرایند WUM[3] و تحلیل فایل‌های ثبتی به فهم رفتار کاربران و بهبود طراحی سایت کمک می‌کند [۲].

با این مقدم نقش داده‌کاوی در کشف روش‌های مؤثر اجرای آزمون‌ها، انتخاب محتوای کیفی‌تر و ارزشیابی واقعی‌تر فراگیران در دوره‌های آموزش الکترونیکی معلوم می‌شود. در این مقاله و در بخش ۲ ضرورت پژوهش بررسی شده و در بخش ۳ به مرور ادبیات موضوع و پیشینه پژوهش‌های این رویکرد پرداخته می‌شود؛ در بخش ٤ کلیات فرایند داده‌کاوی بر داده‌های آموزش الکترونیکی به‌همراه مراحل انجام پژوهش حاضر تشریح می‌شوند. در بخش ۵ روش‌های مختلف آزمایشات – شامل توصیفی و پیش‌بینانه– و یافته‌های این مطالعه آورده شده و در بخش ٦ به اعتبارسنجی روش‌های پیش‌بینی پرداخته شده است. در پایان نیز بخش ۷ برای نتایج و پیشنهاد برای تحقیقات افزون اختصاص پیدا کرده است.

۲ – ضرورت پژوهش

با وجود فواید متعددی که آموزش الکترونیکی به همراه داشته، هنوز ابعادی چون کنترل فراگیران و روش‌های آموزش برای بهبودسازی وجود داشته و درواقع ارتقای کیفیت آموزش فراگیران الکترونیکی، نیازمند روش‌ها و خدماتی فراتر از ساختار سنتی آموزش می‌باشد [۳].

Archive of SID

www.SID.ir

این پژوهش درواقع با نگاه نـقادانه به اعتبار کیفی دوره‌های آموزش الکترونیکی، از دو منظر «امکان مشورت فراگیران در آزمون‌های از راه دور» و نیز «میزان کیفیت محتوای درسی دوره‌ها»، به رفع نقیصه‌های احتمالی در برگزاری دوره‌ها کمک می‌نماید.

یک تفاوت اصلی بین محیط سنتی کلاس‌های رو در رو با آموزش مبتنی بر وب – که با عنوان «سیستم مدیریت آموزشی» (LMS)[٤] شناخته می‌شود– در داده‌های قابل حصول از هر کدام می‌باشد؛ کلاس‌های سنتی به‌طور صرف شامل اطلاعاتی درباره حضور دانشجو، اهداف آموزشی دوره و داده‌های حیطه انفرادی هستند در حالی که در LMS، اطلاعات قابل حصول بیشتری وجود دارند، زیرا تمامی عملکرد دانشجویان در این سیستم‌ها در فایل‌های ثبتی[٥] و پایگاه‌های داده ذخیره می‌شوند و به‌وسیله داده‌کاوی روی داده‌ها یا کشف دانش از پایگاه‌های داده (KDD)[٦] می‌توان به الگوهای مؤثر و موجود درآن‌ها پی بُرد [٤]. داده‌کاوی به‌عنوان رویکردی برای کشف دانش‌های ضمنی، ناشناخته، بالقوه و با ارزش معرفی می‌شود [٥، صص ۱-۲] که می‌توان از آن برای یک نوع از کشف دانش برای حل مسئله در زمینه خاص استفاده کرد [٦، ص ٥٨]. به‌علاوه وب‌کاوی به‌عنوان فرایند کاوش استفاده وب، معادل اِعمال روش‌های داده‌کاوی بر داده‌های فایل‌های ثبتی وب است که با آن می‌توان الگوهای رفتار فراگیران را احصا و آن‌ها را بهتر ارزیابی کرد [٧]. گام‌های اصلی به‌کارگیری فرایند وب‌کاوی در LMSها عبارتند از :«پیش‌پردازش»، «کشف الگو» و «تحلیل الگو» [٢].

۳ - پیشینه پژوهش

برخی خصوصیات و امکانات ویژه آموزش الکترونیکی از طریق LMS که نتایج بهتری نسبت به روش‌های سیستم آموزش سنتی و یا آموزش شنیداری یکطرفه نیز فراهم کرده‌اند، عبارتند از : ارائه مطالب آموزشی در قالب‌های متنوع چند رسانه‌ای به فراگیران، ارتباط پویا و دوطرفه با کاربر، آزمون بر خط و نیز تکالیف در زمان مشخص و امکان احصای مسیرهای آموزشی فراگیران با ردیابی و کاوش داده‌های حجیم از فایل‌های ثبتی و پایگاه داده مرتبط [٣]. ضرورت اجرای داده‌کاوی در حوزه کلان آموزش الکترونیکی در پژوهش‌های دیگر هم اشاره شده است [٨، صص ۱۳۵-۱٤٦]. در [٩، صص ۱۸۳-۲۲۱] نمونه‌هایی از مسائل آموزش الکترونیکی که

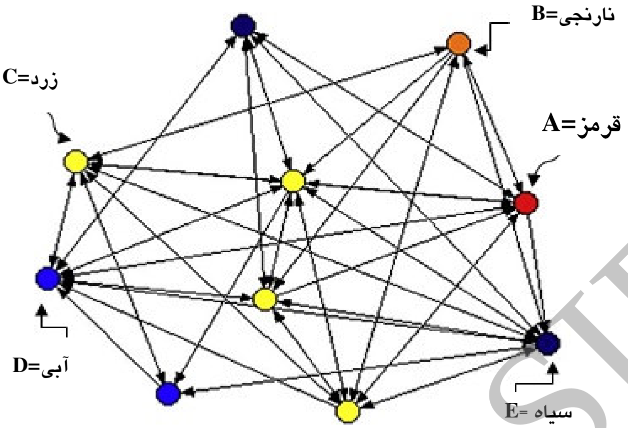
Archive of SID

www.SID.ir

می‌توان فناوری داده‌کاوی را برای حل آن‌ها اعمال کرد، ذکر شده‌اند: دسته‌بندی دانشجویان برحسب بازدهی یادگیریآنان و خوشه‌بندی براساس تشابه میزان و زمان استفاده از LMS، کشف رفتارهای یادگیری غیرمعمول و بهسازی فعل و انفعال LMS. در [۱۰، صص۱۲۰۱-۳٤۱۷] از راهکار داده‌کاوی دسته‌بندی[7] به نام ID3 برای ارزیابی کارایی دانشجویان آموزش الکترونیکی استفاده شده است. وب‌کاوی به‌عنوان شاخه‌ای از داده‌کاوی به کشف «الگوهای بالقوه مفید، معتبر و قبلاً ناشناخته» از فایل‌های ثبتی و داده‌های وبی می‌پردازد و در سیستم‌های مبتنی بر وب نتایج قابل توجهی به‌دست می‌دهد [۱]. زمینه‌های اصلی WM با به‌کارگیری DM بر فایل‌های ثبتی وب، محتویات وب و ساختارهای وبی حاصل شده که از همین تعریف، سه نوع تحلیل مختلف از WM منتج می‌شوند: کاوش استفاده‌ی وب (WUM)، کاوش محتوای وب (WCM)[8] و کاوش ساختار وب (WSM)[9]. داده‌های ورودی فرایند WUM، به‌طور عمده از فایل‌های ثبتی وب‌سرورها (فایل‌های ثبتی) تشکیل می‌شوند و می‌توانند فرمت‌های متنوعی داشته و حاوی آدرس IP کاربر، URL درخواست، نام کاربری در تصدیق هویت، فرد ایجادکننده درخواست (در صورت دسترس)، روش درخواست HTTP (GET یا POST)، زمان مراجعه و نتیجه درخواست (موفقیت، شکست، خطا و..)، بایت‌های انتقالی، URL ارجاعی (که یعنی URL صفحه‌ای که کاربر از آن آمده)، عامل کلاینت و غیره می‌باشند.

یکی از فواید ابزارها و روش‌های تحلیل داده‌ها، تسهیل این کار از راه ارزیابی میزان تأثیر محتوا در کیفیت یادگیری فراگیران است. نمونه‌ای از اینگونه تحلیل‌ها بر داده‌ها با عنوان شبکه‌های تحلیل شخص[10] در [۱۱، صص ۵۸۸-۵۹۹] ارائه شده است که در آن هر گره نشانگر نمره پایانی دانشجو بوده و در شکل‌های ۱ و ۲ آمده است. در این شبکه میزان فعالیت‌های دانشجویان در صفحات گفتگو را همراه با نتایج نمرات نهایی ایشان مشاهده می‌شود. در این تحلیل دو دسته دانشجو از یک دوره دانشگاه UBC[11] انتخاب شده‌اند. شبکه دسته نخست شامل دانشجویان دارای فعالیت کمتر می‌باشد (به فلش‌های خروجی از هر گره توجه شود)؛ براساس شکل ۱ بیشتر این افراد، نمرات C (متوسط) تا E (مردود) اخذ کرده‌اند.

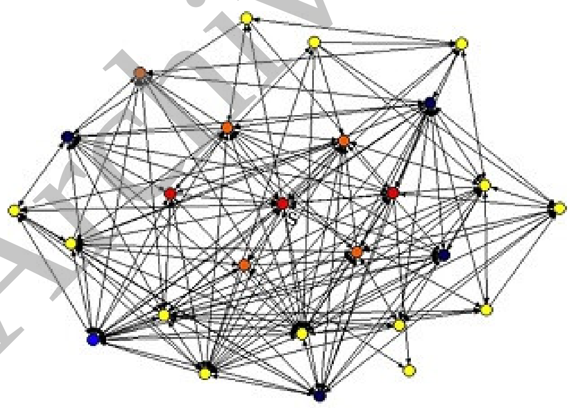
www.SID.ir



مأخذ : (Macfadyen, 2010).

شکل ۱ شبکه تحلیل شخص مربوط به دانشجویان کم‌تحرّک

اما شکل ۲، شبکه آن دسته از دانشجویان فعالی را نمایش می‌دهد که ٨٤ درصد آن‌ها نمره A کسب کرده‌اند:



مأخذ : (Macfadyen, 2010)

شکل ۲ شبکه تحلیل شخص از دانشجوبان فعال، که در آن هر گره، نشانگر نمره پایانی دانشجوست: قرمز=A، نارنجی=B، زرد=C، سیاه/آبی=D یا مردود.

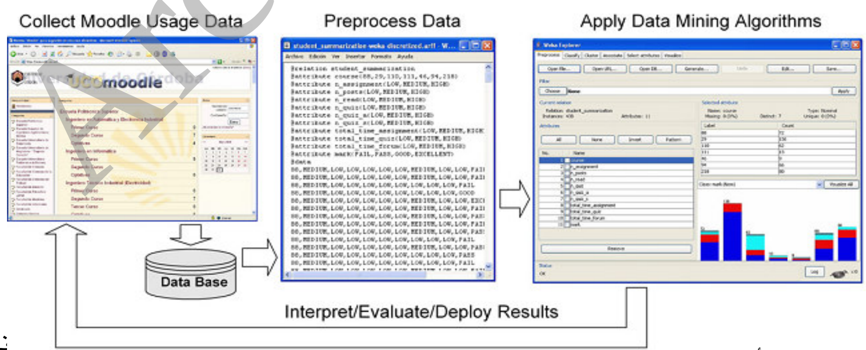
Archive of SID

www.SID.ir

دستیابی به برآوردهای بالا در سیستم‌های آموزشی پیشین، به‌مراتب مشکل و یا مستلزم هزینه‌های افزون و بررسی‌های انسانی بوده است که البته از این تحلیل‌ها می‌توان به‌عنوان پشتیبان آن سیستم‌ها نیز بهره‌برداری کرد. در [٤]، نتایجی از داده‌کاوی بر فعالیت‌های محیط آموزشی Moodle-VLE با استفاده از ابزار وِکا[۱۲] ارائه شده است.

۴ - روش‌شناسی و پیش‌پردازش داده‌ها

روش‌های سنتی تحلیل داده‌های آموزش‌الکترونیکی، فرضیه‌محور بوده است [۱۲]، امّا در روش‌های وب‌کاوی با رویکردی کشف‌گرا، مدل‌های تحلیلی ایجاد می‌کنند که به بررسی و استخراج الگوهای عملکرد دانشجویان می‌پردازد. این روش‌ها به‌صورت چرخه تکراری عمل کرده و دانش مکشوفه در مجموع به‌عنوان حلقه‌ای در تسهیل و ارتقای کیفیت آموزشی و تصمیم‌سازی در این راستا، ایفای نقش می‌کند. این فرایند در شکل ۳ نشان داده شده و شامل چهار مرحله - مشابه مراحل عمومی فرایند داده‌کاوی- می‌باشد [٤]: (الف)- تجمیع داده‌ها: استفاده سیستم مدیریت‌آموزشی به‌وسیله دانشجویان و تجمیع داده‌های فعالیتهای آنان درون پایگاه داده. (ب)- پیش‌پردازش داده‌ها: داده‌ها تمیز شده و به فرمت مناسب برای کاوش تبدیل می‌شوند. (ج)- اِعمال الگوریتم‌های داده‌کاوی: الگوریتم‌های داده‌کاوی اجرا شده تا مدل‌های کشف‌شده را ایجاد و دانش حاصل را خلاصه‌سازی کنند. (د)- تفسیر، ارزیابی و گسترش نتایج: نتایج و مدل‌های حاصل شده تفسیر شده و به‌وسیله گردانندگان سیستم و اساتید مورد استفاده قرار می‌گیرند تا با اخذ بازخورد فعالیت دانشجویان و روش‌های مؤثر دوره‌ها، کیفیت آموزش را ارتقا بخشند.



نه

ماخذ:(Romero,

شکل ۳ فرایند داده‌کاوی در آموزش

Archive of SID

www.SID.ir

داده اولیه در نظر گرفته‌ایم. نخست وضعیت فراگیری دانشجویان این درس را با استفاده از روش خوشه‌بندی[۱۳] توصیف می‌کنیم؛ سپس با تکنیک‌هایی از نوع یادگیری «با ناظر»[۱٤] [۱۳]، پس از آنکه سیستم با اطلاعات موجود در متغیر هدف آموزش دید، پیش‌بینی[۱۵] صورت می‌گیرد. برای پالایش و به اصطلاح تمیزکردن داده‌ها[۱٦]، نمرات ثبت‌نشده یا مقادیر تهی[۱۷] را با مقدار صفر جایگزین کرده[۱۸] و ستون‌های تمام تهی مربوط به کوئیزهایی را که اصلاً برگزارنشده‌اند، حذف می‌کنیم تا آماده‌ی تغذیه به ابزار داده‌کاو رپیدماینِر[۱۹] شوند؛ انتخاب این ابزار براساس در دسترس بودن و ارائه رایگان آن در اینترنت صورت گرفته است. بخشی از نمونه داده‌های پالایش شده را در جدول ۱ آورده‌ایم که در آن Q نشانگر نمرات آزمون بر خط (کوئیز آن‌لاین) و Sn بیانگر میزان فعالیت دانشجو در محتوای جلسه شماره n می‌باشد.

در پیش‌پردازش داده‌ها، نخست براساس ستون امتیاز نهایی[۲۰] از جدول ۱، سه برچسب در نظر می‌گیریم (تا بعد سیستم را آموزش دهیم): برچسب A (بزرگ‌تر یا مساوی با ۸۰)، برچسب B (بزرگ‌تر یا مساوی با ۵۰ و کمتر از ۸۰) و برچسب C (کمتر از ۵۰).

جدول ۱ نمونه داده‌های اولیه

Surname	Gender	S۱	Q۱	S۲	S۳	S۴	S۵	S۶	S۷	S۸	S۹	Final Grade	Label
نام	جنسیت	جلسه ۱	کوئیز ۱	جلسه ۲	جلسه ۳	جلسه ۴	جلسه ۵	جلسه ۶	جلسه ۷	جلسه ۸	جلسه ۹	نمره نهایی	برچسب
Student1	M	-	-	۱	-	-	-	۱	-	-	۱	۹/۳۸	C
Student2	M	-	-	-	-	-	-	-	-	-	-	-	A
Student3	F	۱۳	۱۰	۲۰	۸	۷	۷	٦	٦	۷	٦	۱۰۰	A
Student4	F	-	-	-	-	-	-	-	-	-	-	-	A
Student5	M	۱۳	۱۰	۲۰	۸	۷	۷	٦	٦	۷	٦	۱۰۰	A
Student6	F	۱۳	-	۲۰	۸	۱	-	٦	٦	۷	٦	۹۱/۷۸	A
Student7	M	۱۳	۸	۲۰	۸	۷	۷	٦	٦	۷	٦	۹۱/۷۸	A
Student8	M	۱۳	۱۰	۲۰	۸	۷	-	٦	٦	۷	٦	۱۰۰	A
Student9	M	۱۳	۱۰	۲۰	۸	۷	۷	٦	٦	۷	٦	۱۰۰	A
Student10	M	۱۳	-	۲۰	۸	۱	۳	٤	۱	-	-	۷۷/٦۱	B
Student11	M	۱۳	۱۰	۲۰	۸	۷	۷	٦	٦	۷	٦	۱۰۰	A
…	…	…	…	…	…	…	…	…	…	…	…	…	…
برچسب C: نمرات کمتر از ۵۰				برچسب B: نمرات بزرگ‌تر یا مساوی با ۵۰ و کمتر از ۸۰						برچسب A: نمرات بزرگ‌تر یا مساوی با ۸۰			

Archive of SID

www.SID.ir

۵- پردازش‌ها و یافته‌ها

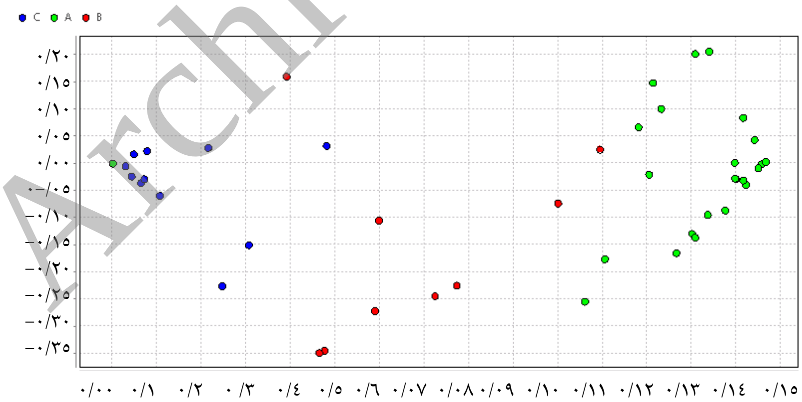
۵-۱- روش‌های توصیفی

تحلیل‌های «فاقد ناظر»[۲۱] (بدون آموزش سیستم با متغیر هدف) را بر داده‌ها اجرا می‌کنیم.

۵-۱-۱- توصیف با خوشه‌بندی

یکی از الگوریتم‌های رایج خوشه‌بندی، عملگر K-Means است که در زمینه‌های پژوهشی دیگر در داخل کشور، مثل خدمات بانکی نیز استفاده شده است [۱۴، ص ۷۹]. در این پژوهش نخست فراگیران را براساس شباهت‌های آن‌ها به‌وسیله عملگر K-Means خوشه‌بندی می‌کنیم. نتیجه گرافیکی خوشه‌بندی سیستم در شکل ٤ آمده است؛ همان‌گونه که ملاحظه می‌شود، تجانس درون خوشه‌ها و عدم تجانس بین خوشه‌ها، نشان از اعتبار خوشه‌بندی دارد. به‌علاوه با مقایسه نتایج خوشه‌بندی سیستم در این شکل (خوشه‌بندی وضعیت نهایی دانشجویان، با توجه به نمرات آنان در فعالیت‌ها) با خوشه‌بندی دستی خودمان (برچسب‌دهی براساس رؤیت نمرات نهایی حاصل شده)، موضوع جالبی ملاحظه می‌شود: به‌جز موارد نادر (مثل قرارگرفتن یک مورد سبز در خوشه آبی‌ها و ی‌ا دو تا قرمز در سبزها)، انطباق خوشه‌ها برقرار بوده و به تعبیری، خوشه‌بندی سیستم «تقریباً همه‌جا»[۲۲] با وضعیت واقعی منطبق است.

Archive of SID



شکل ۴ مدل گرافیکی خوشه‌بندی

www.SID.ir

۵-۱-۲- قاعده‌کاوی[۲۳]

در این قسمت، قواعد نهفته در داده‌های برچسب زده شده را به‌وسیله اجرای عملگرهای قاعده‌کاو استخراج می‌کنیم؛ نخست با عملگر یادگیرنده Basic Rule استنباط می‌کنیم که جدول ۲ را خواهیم داشت. براساس قواعد مندرج در این جدول، در صورت شرکت داشتن در فعالیت‌های S۸ و S٦ و نیز شرکت نداشتن در فعالیت S۱، کلاس A به‌دست می‌آید. برچسب‌های نشان داده شده در انتهای قواعد شرطی این جدول (و نیز به‌طور مشابه در جدول‌های بعدی)، به این صورت تفسیر می‌شوند که اگر هر یک از نتایج سه‌گانه موجود در مؤلفه‌های داخل پرانتز محتمل‌تر بوده باشند، برچسب مربوط به آن مؤلفه به‌عنوان نتیجه نهایی انتخاب می‌شود؛ پس اگر نتیجه قاعده به‌صورت (nA, nB, nC) استنباط شده باشد، به معنای این است که به‌ترتیب، برچسب‌های A، B و C دارای nA، nB و nC مصداق بوده‌اند و به‌طور مثال در صورتی که nA از بقیه بزرگ‌تر باشد، برچسب A نشان داده می‌شود.

جدول ۲ قواعد استخراج شده با عملگر اول (Basic Rule)

Rule Model	شرح قواعد
If S۸ >= ۰ then A (A=٤۲, B=۱, C=۱)	در قاعده ۱، ٤۲ تفر از دانشجویان فعال در محتوای جلسه هشتم، به برچسب A راه یافته‌اند؛
If S۱ <= ۰ then A (A=٥, B=۰, C=۲)	در قاعده دوم، ۵ نفر از غیرفعالان جلسه اول به برچسب A رسیده‌اند.
If S٦ >= ۰ then A (A=٥, B=٤, C=۰)	قاعده سوم بیانگر تأثیر محتوای جلسه ششم در برچسب‌های A و سپس B است.
If S۲ <= ۰ then C (A=۰, B=۰, C=٥)	براساس قاعده چهارم عدم فعالیت در محتوای جلسه دوم، منجر به برچسب C شده است.
If Q۱ <= ۰ then C (A=۰, B=۱, C=۲) else B (A=۲, B=۳, C=۱)	قاعده آخر می‌گوید که عدم فعالیت در کوئیز منجر به برچسب C شده، گرچه فعال بودن در آن نیز برچسب B را به‌دنبال داشته است.
Number of Rules : ٥	

اما عملگر W-JRip که از مجموعه عملگرهای وِکا است، توسط فردی به نام کوهن به‌عنوان نسخه ارتقا یافته الگوریتم یادگیری به نام IREP[۲٤] پیشنهاد شد [۱۵، صص ۱۱۵-۱۲۳].

Archive of SID
www.SID.ir

به‌وسیله این عملگر، قواعد برای هر کلاس واقع در مجموعه آموزش دهنده ایجاد می‌شوند و دانش استخراج شده را در قالب قوانین شرطی IF-THEN نمایش می‌دهد [۱٦]. حال این عملگر را جایگزین عملگر قبلی می‌کنیم که قواعد جدول ۳ به‌دست خواهد آمد:

جدول ۳ قواعد استخراج شده با عملگر دوم (W-JRip)

JRIP Rules
(S۲ <= ۱٤) => label=B (۱۳/۰)
(S۱ <= ٥) => label=A (٥٥/۰)
=> label=C (٦/۰)
Number of Rules: ۳

بنا بر قواعد اخیر هم اگر بارم فعالیت دانشجو در محتوای جلسه دوم (S۱) از ۱٤ بیشتر نباشد، برچسب B حاصل شده، در حالی که دو برچسب دیگر و به‌خصوص A در صورتی ظاهر می‌شوند که فعالیت وی در محتوای جلسه اول (S۱) بارمی بیشتر از ٥ کسب ننماید.

۵-۲- روش‌های پیش‌بینانه[۲۵]

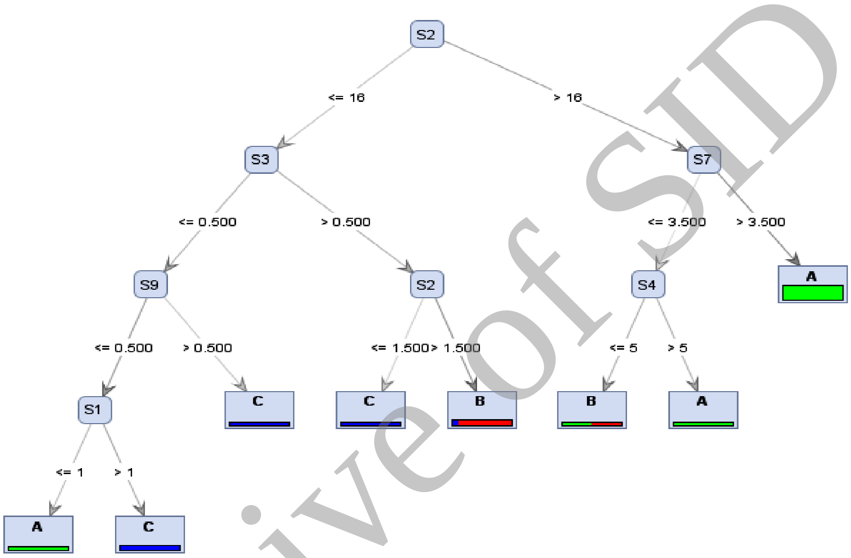
در این قسمت از الگوریتم‌های یادگیرنده استفاده می‌کنیم. توجه داریم که تعداد زیادی از این‌گونه الگوریتم‌ها موجودند و اینکه سیستم را با کدام یک آموزش دهیم، بنا به دلایلی است که در ادامه شرح می‌دهیم. نتایجی که با اجرای «درخت تصمیم» بر داده‌هایی که نمرات نهایی آن‌ها از قبل مشخص هستند، به واقعیت بسیار نزدیک بودند که این موضوع در بخش ششم اعتبارسنجی شده است. همچنین خصوصیات درخت WJ٤٨ را در قسمت ٥-۲-۳ توضیح می‌دهیم. اکنون علاوه بر تکنیک‌های توصیفی قبلی با سه روش پیش‌بینانه مختلف نیز تحلیل می‌نماییم. کل دانشجویان این درس ۷٤ نفر بوده‌اند که توزیع نمره‌های نهایی‌آن به‌صورت C (۱۱)، A(٥٤)، B(۹) می‌باشد.

Archive of SID

www.SID.ir

۵-۲-۱- آموزش سیستم با انتخاب محدوده‌ای از داده‌ها

نخست برای آموزش الگوهای نهفته در داده‌ها به سیستم[۲٦]، ٦۰ مورد اول را انتخاب می‌کنیم[۲۷] که درخت تصمیم[۲۸] شکل ۵ را خواهیم داشت:



شکل ۵ درخت تصمیم مربوط به ٦۰ نمونه نخست

همان طور که ملاحظه می‌شود بنا بر درخت اخیر، اگر نمره‌های فعالیت‌های S۲ و S۷ به‌ترتیب بیشتر از ١٦ و ۳/۵ باشند، قرار گرفتن در کلاس A تضمین شده است. قوانین متناظر با درخت اخیر در جدول ٤ آمده است که یکی از معیارهای قابل تأمل، نقش معکوس فعالیت S۱ نسبت به کلاس A می‌باشد.

Archive of SID

www.SID.ir

جدول ۴ قوانین متناظر با درخت تصمیم شکل ۵

```
If (S۲ <= ۱٦) then
 {
  if   (S۳ <= ۰/٥۰۰) then
     {
       if  (S۹ <= ۰/٥۰۰) then
         if    (S۲    <=    ۱)    then:    A    (A=۳,    B=۰,    C=۰)
           else (S۱ > ۱) then: C (A=۰, B=۰, C=٤)
       if (S۹ >  ۰/٥۰۰) then: C (A=۰, B=۰, C=۳)
     }
   Else (S۳ > ۰/٥۰۰)
       if (S۲ <=۱/٥۰۰) then : C (A=۰, B=۰, C=۲)
        else    (S۲    >۱/٥۰۰)    then:    B    (A=۰,    B=۷,    C=۱)
 }
Else (S۲ > ۱٦)
 {
   if  (S۷ <=  ۳/٥۰۰) then
     {
       if    (S٤    <=    ٥)    then:    B    (A=۱,    B=۱,    C=۰)
        else    (S٤    >    ٥)    then:    A    (A=۲,    B=۰,    C=۰)
     }
    else (S۷ > ۳/٥۰۰) then: A (A=۳٦, B=۰, C=۰)
 }
```

در این روش بعد از آموزش سیستم به‌وسیله درخت فوق، نتیجه پیش‌بینی سیستم روی مابقی موارد (مورد ٦۱ تا ۷٤ که وضعیت نهایی آن‌ها به سیستم آموزش داده نشده) به‌صورت جدول ٥ خواهد بود:

جدول ٥ نتایج پیش‌بینی با روش اول (آموزش سیستم با درخت تصمیم شکل ٥)

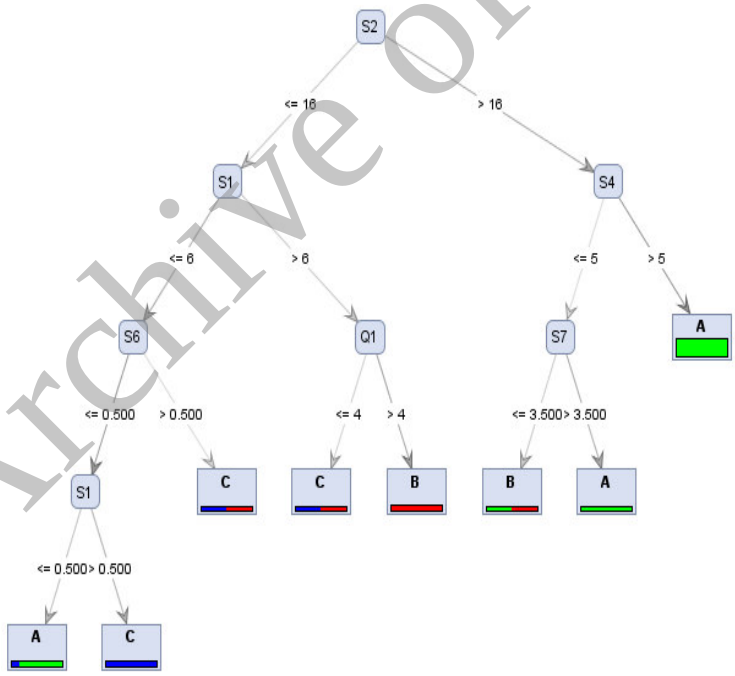
Label	Nominal	C (۱), A (۱۲), B (۱)
Prediction (Label)	Nominal	C (۲), A (۱۲), B (۰)
Confidence (C)	Real	[۰/۰۰۰ ; ۱/۰۰۰]
Confidence (A)	Real	[۰/۰۰۰; ۱/۰۰۰]
Confidence (B)	Real	[۰/۰۰۰; ۰/۰۰۰]

Archive of SID
www.SID.ir

در جدول ٥ نتایج پیش‌بینی شده (ردیف دوم) تا حد زیادی نزدیک به واقعیت (ردیف اول) است؛ چگونگی محاسبه میزان «نرخ دقت»[۲۹] این روش را در جدول ۷ خواهیم دید. اما اعتبار این نتایج تا چه حد بوده و می‌توان با آن‌ها تصمیم‌سازی نمود؟ برای کسب نتایج منطقی‌تر، تحلیل‌های داده‌کاوی را با روش‌های دیگری نیز بررسی می‌کنیم.

۵-۲-۲- آموزش سیستم با نمونه‌برداری داده‌ها

اکنون در مرحله انتخاب موارد آموزش‌دهنده سیستم، به جای انتخاب محدوده‌ای موارد با روش ASS[۳۰] نمونه‌های را انتخاب و سیستم را آموزش می‌دهیم. درخت حاصل در مرحله آموزش به صورت شکل ٦ می‌باشد و ملاحظه می‌شود که در این درخت، نقش کوئیز ۱ بعد از فعالیت‌های S۲ و S۱ تعیین‌کننده است :



شکل ۶ درخت تصمیم مربوط به موارد نمونه‌برداری با ASS

Archive of SID

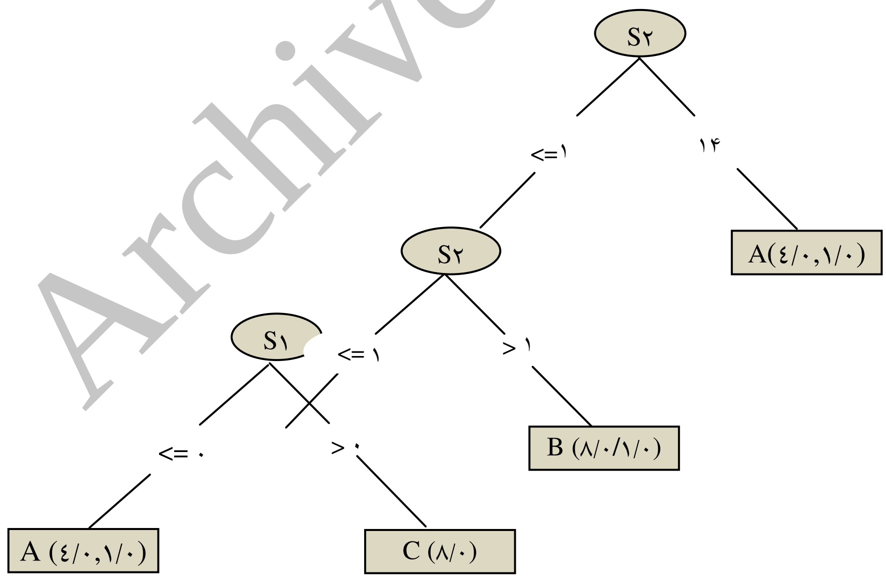
www.SID.ir

۵-۲-۳- آموزش سیستم با درخت تصمیم وکا

برای آنکه بتوان پیش‌بینی واقع‌بینانه‌تری داشت، باید به یکی از معیارهای منطقی‌بودن روش توجه داشت که لزوم تأیید نظر خبرگان فن درباره درختهای تصمیم مرحله‌ی آموزش سیستم می‌باشد. پس از بکارگیری درخت قابل قبول از نتایج می‌توان مرحله‌ی پیش‌بینی تصمیم‌سازی نمود. اینک درخت تصمیم دیگری را در دو روش قبل جایگزین می‌نمائیم. بدین منظور، درخت WJ٤٨– از شاخه‌ی وکا– را جایگزین DT می‌نمائیم.

درخت WJ٤٨ از الگوریتم معروفی به نام کوئین‌لَن[۳۱] گرفته شده که الگوریتمی بهینه‌سازی شده از الگوریتم پایه‌ای ID۳ و بر اساس راهکار استنتاج بالا به پایین از درختهای تصمیم (TDIDT)[۳۲] می‌باشد [۱۷؛ ۱۸، صص ۸۳۷-۸۴۰]. درخت WJ٤٨ از این واقعیت استفاده می‌کند که می‌توان از طریق تقسیم مجموعه داده‌ها به زیرمجموعه‌های کوچکتر، هر خصوصیت داده‌ها را به عنوان یک «تصمیم» در نظر گرفت؛ به این ترتیب، WJ٤٨ یک « درخت تصمیم» از روی داده‌ها ایجاد می‌کند که برگهای این درخت، قوانین موجود در داده‌ها را نمایش می‌دهند [۱۹، صص ۱۱۵-۱۲۲؛ ۲۰].

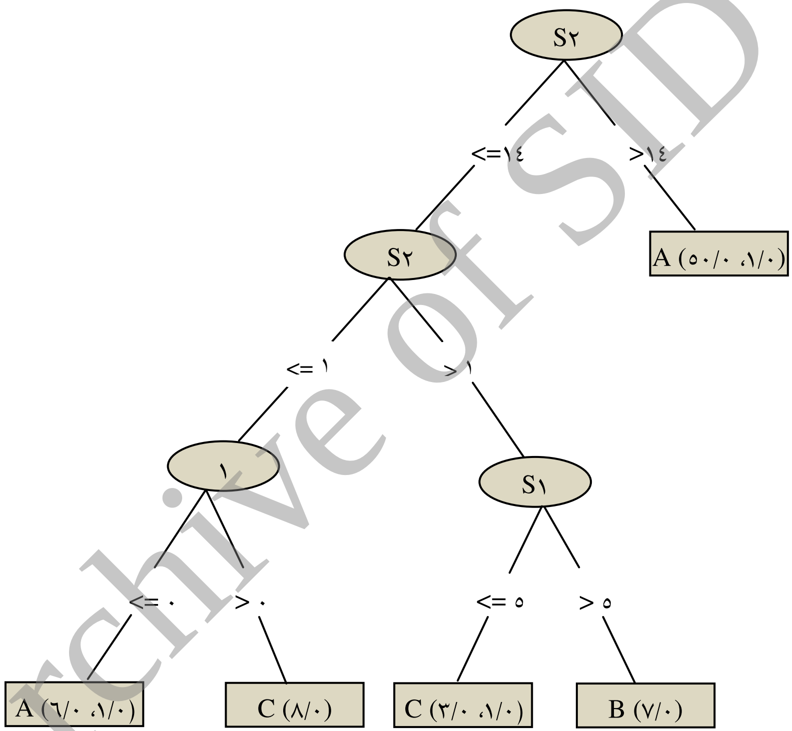
در این قسمت، روش سوم را با بازسازی روش نخست قبلی را توسط درخت جدید، طراحی می‌نمائیم؛ الگوهای آموزش دیده توسط درخت WJ٤٨ در شکل ۷ آمده‌اند:



شکل ۷ درخت WJ48 روی ٦۰ نمونه‌ی نخست

Archive of SID
www.SID.ir

در فاز پایانی، روش دوم انتخاب نمونه‌های مرحله آموزش سیستم– که به صورت ASS انجام داده‌بودیم– را با درخت WJ٤۸ اجرا می‌نمائیم که روش چهارم شکل می‌گیرد. در قوانین استنباط شده از درخت جدید نیز نقش نمره فعالیت۲ S به‌عنوان شاخصی در منتج شدن کلاس A بوده و متعاقب آن، شرکت نداشتن در فعالیت S۱ به چنین نتیجه‌ای کمک کرده است (شکل ۸).



شکل ۸ درخت WJ٤۸ روی موارد نمونه‌برداری شده با ASS

Archive of SID

www.SID.ir

قوانین معادل با درخت فوق در جدول ٦ نمایانده شده‌اند:

جدول ۶ قوانین حاصل از اجرای درخت WJ٤٨ پس از نمونه‌برداری ASS

```
If (S٢ <= ١٤) then
{
  If  (S٢ <= ١) then
   { If   (S١ <= ٠) then: A (٦/٠)
     If   (S١ > ٠) then: C (٨/٠)
   }
  Else If  (S٢ > 1) then
   { If  (S١ <= ٥) then: C (٣/٠)
     If  (S١ > ٥) then: B (٧/٠)
   }
}
If (S٢ > ١٤) then: A (٥٠/٠)
-----------------------------------
Number of Leaves :    ٥
Size of the tree :    ٩
```

۶- اعتبارسنجی روش‌ها

در بخش ۵-۲، چهار روش مختلف را برای پیش‌بینی طراحی نمودیم و در این بخش دقت آنها را اعتبارسنجی می‌کنیم. ابتدا در قسمت ۵-۲-۱، روش اول آموزش سیستم با نمرات ٦٠ فراگیر ابتدای لیست صورت گرفت و سیستم با درخت تصمیم آموزش دید که نتایج پیش‌بینی در جدول (۵) نمایش داده شدند. نرخ دقت روش ذکر شده، با استفاده از عملگر Classification Performance برابر با ٩٢/٨٦ درصد سنجیده می‌شود (جدول ۷).

Archive of SID

www.SID.ir

جدول ۷ میزان دقت پیش‌بینی به روش نخست (آموزش سیستم با درخت تصمیم شکل ۵)

Accuracy: ۹۲/۸۶%				
	true C	true A	true B	Class precision
pred. C	1	۰	۱	۵۰/۰۰%
pred. A	0	۱۲	۰	۱۰۰/۰۰%
pred. B	0	0	۰	۰/۰۰%
class recall	۱۰۰/۰۰%	۱۰۰/۰۰%	۰/۰۰%	

در روش دوم (در قسمت ۵-۲-۲)، روش نمونه‌برداری را تغییر دادیم و میزان دقت نهائی نتایج پیش‌بینی سیستم - روی همان ١٤ موردی که در روش قبل مورد پیش‌بینی قرار گرفتند- مشابه روش پیشین (٩٢٫٨٦٪) محاسبه می‌گردد (جدول ۸).

جدول ۸ دقت پیش‌بینی با روش دوم (نمونه‌برداری با الگوریتم ASS)

Accuracy: ۹۲/۸۶%				
	true C	true A	true B	Class precision
pred. C	۰	۰	۰	۰/۰۰%
pred. A	۱	۱۲	۰	۹۱/۳۱%
pred. B	۰	۰	۱	۱۰۰/۰۰%
class recall	۰/۰۰%	۱۰۰/۰۰%	۱۰۰/۰۰%	

در روش‌های سوم و چهارم که درخت یادگیری را از مجموعه‌ی وکا جایگزین درخت تصمیم دو روش قبل نمودیم، باز هم نرخ دقت پیش‌بینی بر روی ١٤ مورد باقیمانده برابر با همان روش‌ها و معادل ٩٢/٨٦ درصد به‌دست‌آمد. بنابراین به‌طور صرف توجه به دقت فنی یک روش نمی‌تواند تعیین‌کننده باشد، بلکه علاوه بر آن مستلزم تعیین صحت قوانین استنباط شده با نظر متخصصان موضوع نیز می‌باشد.

Archive of SID

www.SID.ir

۷ - نتیجه‌گیری

در این مقاله با اجرای الگوریتم‌های مختلفی از داده‌کاوی، داده‌های آموزش الکترونیکی تحلیل شد. نخست با روش‌های توصیفی، فراگیران خوشه‌بندی شده و ملاحظه شد که یک کاربرد تحلیل‌های توصیفی داده‌کاوی، گروه‌بندی دانشجویان براساس همخوانی و تشابه آنان است که می‌توان از آن رسته‌بندی متناسب فراگیران براساس توانمندی علمی آن‌ها را استنتاج کرد. سپس با طراحی چهار ترکیب مختلف برای روش‌های پیش‌بینانه تا میزان دقت۸۶/۹۲ درصد قرارگرفتن فراگیران در دسته‌های مختلف پیش‌بینی شد. با این تحلیل‌ها، الگوهای قابل استفاده و جدیدی نیز کشف شدند:

۱-بنابر نتیجه مشترک الگوریتم‌ها، فعالیت دانشجو در محتوای جلسه دوم (S2) دارای نقش مؤثر برای قرارگرفتن در کلاس A نمره نهایی بوده و از سویی، فعالیت دانشجو در محتوای جلسه اول (S1) از نقش منفی برخوردار است. همچنین در میان این چهار روش، تنها در یکی از درخت‌ها، نقش کوئیز ۲ مشاهده شد که آن هم بعد از فعالیت‌های S2 و S1 تعیین‌کننده بود.

۲-این پژوهش نشان می‌دهد فعالیت دانشجویان در محتوای درسی مهم دوره‌ها، بیش از کسب نمرات بهتر در «آزمون‌های از راه دور» به موفقیت نهایی آن‌ها کمک کرده است؛ بنابراین دستاورد جدید این پژوهش برای مدیریت این دوره‌ها، ضرورت رویکرد به‌روزرسانی محتواهای درسی مؤثر و الزام شرکت دادن دانشجویان در آن‌ها می‌باشد.

۳-درضمن روش‌های مختلف پیش‌بینی برحسب عوامل زیر تنظیم شدند:

الف- چگونگی نمونه‌برداری داده‌ها در مرحله آموزش

ب – الگوریتم مورد استفاده در ایجاد درخت تصمیم

ج – گستره داده‌های تحت آزمون پیش‌بینی

که نرخ دقت یکسانی به‌دست آمد، بنابراین نتیجه گرفته شد که به‌طور الزام توسیع تعداد داده‌ها موجب بالارفتن ضریب دقت در روش‌ها نشده و استفاده از نظر خبرگان حوزه آموزش در اعتبار قوانین استخراج شده برای آموزش سیستم می‌تواند قبل از اعمال نهایی آن‌ها در کیفیت پیش‌بینی مؤثر باشد. در خصوص تحقیقات آینده نیز پیشنهاد می‌شود تا داده‌های ثبت شده وب‌سرور با انواع دیگری از فایل‌های ثبتی – مثل فایل‌های ثبتی پروکسی یا فایل‌های لاگ سرویس‌گیرنده- تکمیل شوند.

Archive of SID

www.SID.ir

٤- این تحلیل‌های فناورانه می‌تواند در پیش‌بینی روند رشد تحصیلی فراگیران و دسته‌بندی آن‌ها براساس خوشه‌های مبتنی بر معیارهای فعالیت تحصیلی آنان کمک نمایند و با تعیین آن دسته از عواملی که در تحقق اهداف آموزشی تأثیر بالایی دارند، سیاستگزاری ارائه خدمات و مدیریت برگزاری دوره‌های آموزش الکترونیکی را بهبود بخشند.

۸ - پی‌نوشت‌ها

1. Data Mining
2. Web Mining
3. Web Usage Mining
4. Learning Management System
5. Log files
6. Knowledge Discovery and Data Mining
7. Classification
8. Web Content Mining
9. Web Structure Mining
10. Ego
11. University of British Columbia, 2009
12. Weka
13. Clustering
14. Supervised
15. Prediction
16. Data cleaning
17. missing values

۱۸. تعیین مقدار جایگزین، توسط عملگری به نام «Missing value replenishment» انجام می‌شود.

19. Rapid Miner
20. Final num
21. Unsupervised
22. Almost every where
23. Rule mining
24. Incremental Reduced Error Pruning
25. Predictive

۲٦. به این گام، مرحله «آموزش سیستم» (System Training) گفته می‌شود.

۲۷. انتخاب محدوده‌ای از داده‌ها را می‌توان به کمک عملگری به نام «Example Range Filter» انجام داد.

28. Decision tree
29. Accuracy rate

Archive of SID

www.SID.ir

30. Absolute Stratified Sampling
31. Quinlan algorithm (C4.5)
32. Top-Down Induction of Decision Trees

۹ - منابع

[1] Mustapasa O. & et al., “Implementation of semantic web mining on E-learning”, Science Direct. Istanbul-Turkey, Department of Software Engineering, University of Bahçesehir, January 2010.

[2] Psaromiligkos Y. & et al., “Mining log data for the analysis of learners’ behavior in web-based learning management systems”, Springer-Verlag. Piraeus, Greece. Technological Educational Institute of Piraeus, January 2009.

[3] Chuang H., Shen C., “A study on the relationship between learning path, learning style, and e-learning performance”, IEEE Explore. Taiwan, Department of Information Management, National Yunlin University of Science and Technology, July 2008.

[4] Romero C., Ventura S., Garcıa E., “Data mining in course management systems: Moodle case study and tutorial”, Science Direct. Cordoba-Spain, Department of Computer Sciences and Numerical Analisys, University of Cordoba, 2008.

[5] Mohammad Rezapour, M. Khavanin-zadeh & M.M. Sepehri, “Implementation of predictive data mining techniques for identifying the risk factors of early avf failure in hemodialysis patients”, Computational and Mathematical Methods in Medicine, Vol. 2013, Article ID 830745, 8 pages, 2013. http://dx.doi.org/10.1155/2013/830745.

[6] Hassanzadeh, A., Ghanbari, M. H., & Elahi, S., “Classification of mobile banking users by data mining approach: Comparison between artificial neural networks and naïve bayes techniques”, *journal of Management Research in Iran* (Modares Human Sciences); 16(2 (75)); pp. 57-71, Summer 2012 .

[7] Sun J., Xie Y., “A recommender system based on web data mining for personalized e-learning”, IEEE Xplore. Xiamen, China. Department of Computer Science & Technology, Xiamen University of Technology, XMUT, 2009.

Archive of SID

www.SID.ir

[8] Romero C., Sebastian V., "Educational data mining: A survey from 1995 to 2005", *Expert Systems with Applications, 33*, No. 1, Pp. 135-146, 2007.

[9] Castro F., Alfredo V., Àngela N., Francisco M., "Applying data mining techniques to e-learning problems", In Evolution of Teaching and Learning Paradigms in Intelligent Environment; Springer Berlin Heidelberg, pp. 183-221, 2007.

[10] Baradwaj B. K., Saurabh P., "Mining educational data to analyze students' performance", arXiv preprint arXiv: 1201-3417, 2012.

[11] Mac Fadyen Leah P., Dawson S., "Mining lms data to develop an ‘‘early warning system”, for educators: A proof of concept”, *Science Direct, Computers & Education,* 54: pp. 588–599, 2010.

[12] Gaudioso E., Talavera L., “Data mining to support tutoring in virtual learning communities: Experiences and challenges”, In C. Romero & S. Ventura (Eds.), Data Mining in E-learning. Southampton, UK: Wit Press, 2006.

[13] Nakhaeizadeh G., “Tutorial data mining & its application in industry & commerce”, Part 2. 2008.

[14] Shahin, A., Salehzadeh, R., & Ghandehari, M., "Integration of clustering, AHP and KANO approaches to describe SAMAN bank services – with a case study in QOM”, *Jouranal of Management Research in Iran* ; 16(1 (74)); pp. 73-91, spring 2012 .

[15] Cohen W., “Fast effective rule induction”; Proceedings of the 12th International Conference on Machine Learning; ML95, pp. 115–123, 1995.

[16] Witten I. H., Frank M., Hall M. A., “Data mining: Practical machine learning tool and technique with Java implementation”, Morgan Kaufmann, San Francisco, Calif, USA, 3rd Edition, 2011.

[17] Quinlan J., “C4.5: Programs for machine learning”, Morgan Kaufmann, San Francisco, Calif, USA, 1993.

[18] Ilczuk G. & et al., “Rough set techniques for medical diagnosis systems”, IEEE,

Archive of SID

www.SID.ir

Computers in Cardiology, 32, pp. 837–840, 2005.

[19] Cohen P., Jensen D., "Over fitting explained", Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics, pp. 115–122, 1997.

[20] Serapião A. B. S., Bannwart A. C., "Knowledge discovery for classification of three-phase vertical flow patterns of heavy oil from pressure drop and flow rate data", *Journal of Petroleum Engineering*; Vol. 2013, 2012.

Archive of SID

www.SID.ir