

# ارتقای کیفیت آموزش در سامانه‌های آموزش الکترونیکی با استفاده از داده‌کاوی آموزشی

بهرروز مقصودی<sup>۱</sup>، صادق سلیمانی<sup>۲</sup>، علی امیری<sup>۳</sup> و محسن افشارچی<sup>۴</sup>

چکیده: آموزش الکترونیکی، عبارت است از انجام فرایند آموزشی بر روی بستر ارتباطات الکترونیکی همچنین بهره‌گیری از فناوری شبکه برای طراحی، ارائه و توسعه آموزش می‌باشد که فراگیران، کارشناسان و تهیه‌کنندگان مطالب را در بر می‌گیرد. در این میان حجم وسیعی از اطلاعات مانند نحوه تعامل کاربر با سامانه‌های مدیریتی آموزش، دروس انتخابی دانشجو و نمرات دانشجویان ذخیره می‌گردد. این داده‌ها حاوی اطلاعات با ارزشی برای مطالعه و تحلیل رفتار دانشجویان و ارائه مشاوره به دانشجویان می‌باشد. به علت دور بودن استاد از دانشجو و نبود کارشناس گروه در سیستم‌های آموزش الکترونیکی نیاز به ارائه یک مشاوره مجازی و کارشناس گروه برخط ضروری می‌باشد که بتواند دانشجویان را در تصمیم‌گیری کمک کرده و در نهایت منجر به ارتقای کیفیت آموزش شود. هدف اصلی این پژوهش به دست آوردن تجربه‌هایی فراتر از تجربیات یک کارشناس گروه و مدیرگروه حقیقی با استفاده از داده‌کاوی و همچنین استفاده از این تجربیات در سیستم آموزش الکترونیکی جهت هدایت تحصیلی می‌باشد. در این پژوهش به کشف الگوهای نهفته در انتخاب واحد دانشجویان و پیش‌بینی نمرات آنان پرداخته شده است. همچنین تأثیر فعالیت، نحوه ورود، ساعت ورود، فصل و غیره در سامانه مدیریت آموزش الکترونیکی بررسی شده است.

کلمات کلیدی: آموزش الکترونیکی، داده‌کاوی آموزشی، قوانین انجمنی، پیش‌بینی

## ۱- مقدمه

آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی، فراگیری ماشینی و بازنمایی بصری داده می‌باشد. داده‌کاوی فرآیندی پیچیده جهت شناسایی الگوها، مدل‌های صحیح و بالقوه مفید در حجم وسیعی از داده است، به طوری که این الگوها و مدل‌ها برای انسان‌ها قابل درک باشند [۱]. داده‌کاوی به صورت یک محصول قابل خریداری نیست؛ بلکه یک رشته علمی و فرآیندی است که باید به صورت یک پروژه پیاده سازی شود. یکی از شاخه‌های جالب آن داده‌کاوی آموزشی (EDM) می‌باشد که امروزه در کشورهای درحال توسعه تحقیقات زیادی در این مورد در حال انجام است. از سال ۲۰۰۴ کنفرانس‌های مهمی از جمله AIED، UM، ITS. این کنفرانس‌ها فعالیت خود را در این زمینه گسترش دادند تا اینکه در سال ۲۰۰۸ اولین کنفرانس مستقل داده‌کاوی آموزشی در شهر مونترال کانادا برگزار گردید. درسال‌های اخیر، با توجه به مشکلات موجود در زمینه استخراج دانش

بنابر اعلام دانشگاه MIT دانش نوین داده‌کاوی (DM) یکی از ده دانش در حال توسعه‌ای است که دهه آینده را با انقلاب تکنولوژیکی مواجه می‌سازد. این تکنولوژی امروزه دارای کاربرد بسیار وسیعی در حوزه‌های مختلف است به گونه‌ای که حد مرزی برای کاربرد این دانش در نظر نگرفته و زمینه‌های کاری این دانش را از ذرات کف اقیانوس‌ها تا اعماق فضا می‌دانند. امروزه، بیشترین کاربرد داده‌کاوی دربانک‌ها، مراکز صنعتی و کارخانجات بزرگ، مراکز درمانی، بیمارستان‌ها، مراکز تحقیقاتی، بازاریابی هوشمند و بسیاری از موارد دیگر می‌باشد. داده‌کاوی پل ارتباطی میان علم

تاریخ دریافت مقاله ۹۰/۰۶/۳۱، تاریخ تصویب نهایی ۹۰/۱۱/۱۷

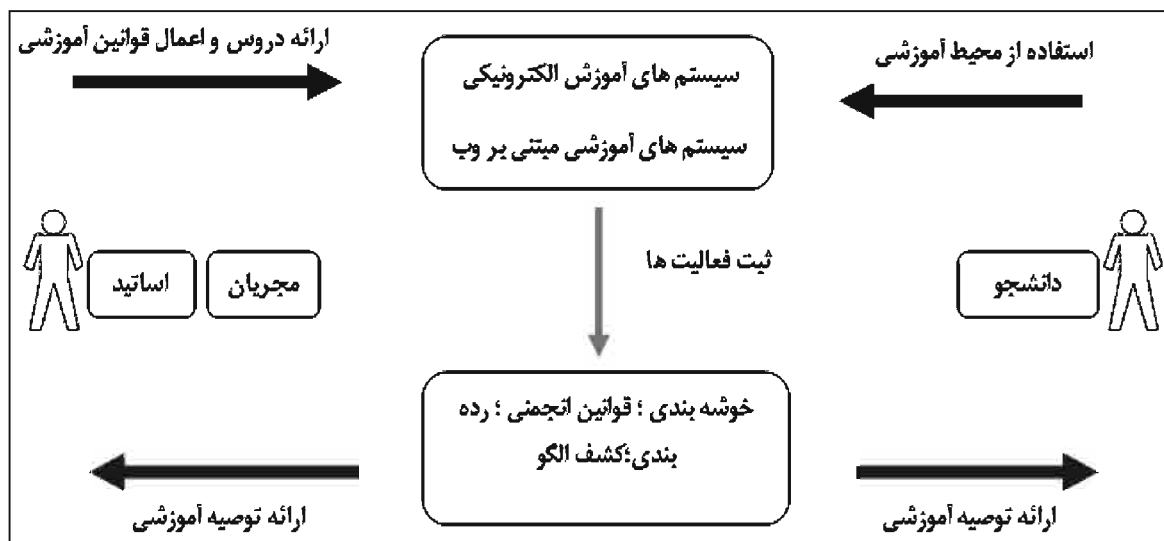
<sup>۱</sup> کارشناسی ارشد، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد زنجان (نویسنده مسئول)، پست الکترونیکی:

Behroozmagsodi@azu.ac.ir

<sup>۲</sup> مربی، دانشکده فنی مهندسی برق و کامپیوتر، دانشگاه کردستان

<sup>۳</sup> استادیار، دانشکده فنی برق و کامپیوتر، دانشگاه زنجان

<sup>۴</sup> استادیار، دانشکده فنی برق و کامپیوتر، دانشگاه زنجان



شکل ۱ ساختار داده کاوی آموزشی [۲]

سیستم‌های آموزشی هوشمند در برزیل، فعالیت خود را آغاز نمودند. تاریخچه این کارگروه در کنار مقالات ارائه شده در کارگاه‌های آموزشی و کنفرانس‌های برگزار شده، باعث تمایل پژوهشگران در استفاده از روش‌های داده‌کاوی در سیستم‌های آموزش و یادگیری شد. در وب از علم داده‌کاوی به عنوان استخراج دانش از پایگاه‌داده یاد می‌شود که امکان کشف دانش مفید و جدید را از میان حجم عظیم اطلاعات ذخیره شده فراهم می‌آورد. بنا به تحقیقات موجود روش‌های داده‌کاوی آموزشی اکثراً با روش‌های عادی داده‌کاوی متفاوت است؛ زیرا این حوزه نیازمند تشریح سلسله مراتب چند سطحی از داده‌های آموزشی و وابستگی ضمنی میان آن‌ها می‌باشد. رومرو و همکاران ساختار داده‌کاوی آموزشی را مانند شکل شماره ۱ نشان داده‌اند [۲]. الگوهای کشف شده با استفاده از داده‌کاوی به عنوان توصیه به دانشجو و به عنوان دانش موجود در داده به مجریان آموزشی اعلام می‌شود. در اکثر محیط‌های آموزش الکترونیکی اطلاعات دانشجو، رفتار و تعاملات دانشجو با سامانه آموزشی ذخیره و نگهداری می‌شود. در سیستم‌های سنتی از این اطلاعات صرفاً جهت بایگانی یا چاپ نمودار و گزارشات ساده مورد نیاز استفاده می‌شود. طرف دیگر قضیه، حجم زیاد و بی‌ساختار بودن این اطلاعات است

از پایگاه داده‌های عظیم و استفاده از آنها به منظور تصمیم‌گیری‌های مناسب، تکنیک‌های داده‌کاوی و استفاده از آنها در امر آموزش مورد توجه قرار گرفته است. این زمینه تحقیقاتی جدید، داده‌کاوی آموزشی نامیده می‌شود که به امر توسعه روش‌های کشف دانش از داده‌های محیط آموزشی خصوصاً دانشجویان می‌پردازد [۲]. در سایت این گروه، واژه داده‌کاوی آموزشی این گونه بیان شده است: داده‌کاوی آموزشی یک حوزه علمی نو ظهور است که به توسعه روش‌هایی برای کاوش و اکتشاف دانش در محیط‌های آموزشی می‌پردازد و از این روش‌ها برای درک بهتر دانشجویان و محیط آموزشی، که فرآیند یادگیری و آموزش در آن صورت می‌پذیرد، استفاده می‌شود [۳ و ۴]. همچنین در حوزه اصلی داده‌کاوی آموزشی، کاوش داده‌های پروفایل و فعالیت‌های دانشجویان معرفی شده است [۴ و ۵]. در حوزه اصلی دیگر داده‌کاوی، کاوش در داده‌های ثبت نام دانشجویان می‌باشد. کاربرد اصلی داده‌کاوی، در پیش‌بینی رفتار دانشجو و همچنین کشف قوانین نهفته جهت توصیه به دانشجو می‌باشد. همان طور که اشاره شد کارگروه تحقیقاتی داده‌کاوی آموزشی از سال ۲۰۰۴ و با برگزاری کارگاه آموزشی تحلیل تعاملات یادگیرنده آموزگار با هدف بهبود نتایج آموزشی در حاشیه هفتمین کنفرانس

تأثیری بر روند پیشرفت یا آفت تحصیلی داشته است. دانشجویان چه درس‌های را با هم انتخاب می‌کنند. شاید یکی از واضح‌ترین سؤالات اساتید این باشد که کدام قسمت، در سیستم آموزشی تأثیر بیشتری در یادگیری داشته است. دانشجویان موفق از چه الگویی در تعامل با سامانه استفاده کرده‌اند. به جرأت می‌توان گفت با استفاده از داده‌کاوی می‌توان از تعاملات دانشجویان با سامانه، الگو‌هایی را کشف و استخراج کرد که برای مدیران و دانشجویان کاربرد زیادی دارد.

## ۲- روش‌شناسی پژوهش

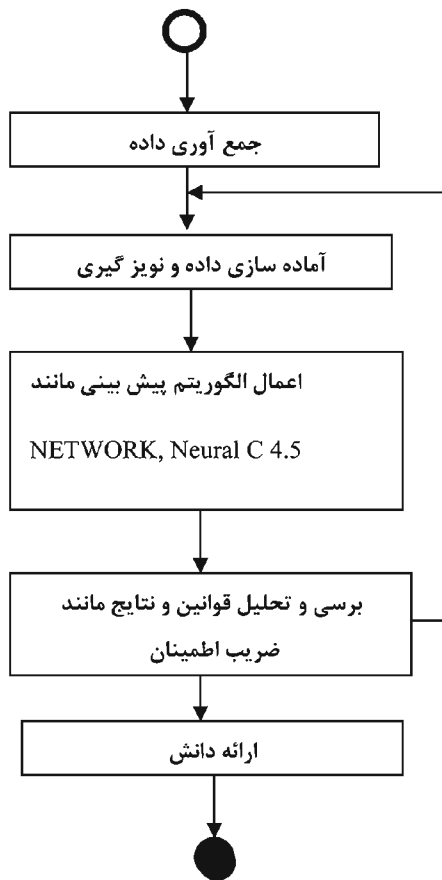
در این پژوهش سعی بر آنست با استفاده از داده‌های موجود، سیستمی ایجاد شود تا دانشجو در مسیر صحیح یادگیری هدایت شود. برای نیل به این هدف کارهای زیر انجام شده سپس به صورت توصیه در اختیار دانشجویان قرار می‌گیرد.

- پیش‌بینی نمره دانشجویان.
- تحلیل سبد انتخاب واحد دانشجویان.
- تحلیل رفتار دانشجویان و ارائه قوانین.

پیش‌بینی نمره برای دانشجویان مبحث جالبی است. با استفاده از الگوریتم‌های داده‌کاوی می‌توان با تأثیر عوامل مؤثر از جمله میزان ورود به سایت، زمان‌های ورود به سایت، نحوه تعامل در سایت و غیره نمره دانشجو را در درس خاصی پیش‌بینی کرد. نکته قابل توجه در پیش‌بینی، ذکر این نکته می‌باشد که می‌توان نمره دانشجو، در دوره  $N+1$  را، در دوره  $N$  پیش‌بینی کرد و این یعنی پیشگیری قبل از اتفاق؛ به طور مثال در حین انتخاب واحد ترم جاری دانشجو با انتخاب هر درس، می‌تواند نمره آن درس را که هنوز به طور واقعی ثبت نشده است مشاهده کند و دانشجو به طور مثال می‌تواند دروسی را که پیش‌بینی شده است نمره کم می‌گیرد، حذف کند یا اینکه در صورت اخذ، مطالعه بیشتری داشته باشد. کاربرد جالب این پیش‌بینی برای مدیران آموزش و کارشناسان آموزشی است. به طور مثال برنامه‌ریزی ترم بعدی زمانی صورت می‌گیرد که ترم جاری هنوز تمام نشده است. برای کارشناسان گروه جالب است بدانند در ترم فعلی در یک درس خاص چند نفر مردود می‌شوند که برنامه ریزی خود را برآن اساس طراحی

که کارایی زیادی برای اساتید و دانشجویان ندارد؛ در حالی که با استفاده از داده‌کاوی آموزشی می‌توان با این مشکلات مقابله کرد. نکته قابل توجه این که با استفاده از داده‌کاوی آموزشی می‌توان الگو و قوانین موجود در سیستم را کشف کرد که این قوانین به طور واضح در سیستم مشخص نیستند. استفاده از این الگوهای کشف شده می‌تواند چاره ساز مشکلاتی باشد که امروزه دانشجویان درسیستم‌های آموزش الکترونیکی با آن مواجه هستند. نکته قابل توجه دیگر این که الگوهای استخراجی می‌توانند در ارتقای سیستم آموزش و حتی روش ارائه دروس تأثیرگذار باشند. امروزه با رشد وب، کاربرد فناوری وب مدار، آموزش و یادگیری در همه جا مورد توجه قرار گرفته است. این شیوه از آموزش به طور قابل ملاحظه‌ای اهمیت یافته و هزاران دوره درسی وب در چندین سال گذشته شروع به فعالیت کرده‌اند؛ به طوری که دانشجو و استاد هر دو به طور مجازی فعالیت می‌کنند و محدودیت مکانی وجود ندارد. همچنین در سال‌های اخیر تمایل مردم به علم‌آموزی و تحصیلات دانشگاهی، رندرو به رشد پذیرش در دانشگاه‌ها را به همراه داشته است. با این افزایش پذیرش دانشجو در رشته‌های مختلف به صورت الکترونیکی و نبود کارشناس گروه، مدیر گروه و غیره، لزوم هدایت صحیح دانشجویان و مشاوره بیش از پیش احساس می‌شود. نظر به اینکه دانشجویان در سیستم آموزش مجازی با یک سامانه آموزشی در ارتباط می‌باشند؛ اساتید، دانشجویان و مدیران از راه دور و از مکان‌های مختلف به این سامانه متصل می‌شوند و ارتباط واضحی در میان نیست، ممکن است دانشجویان دچار سردرگمی شوند. ورود به سامانه آموزشی و انجام تعاملات در برگیرنده اطلاعات زیادی است. امروزه هزینه‌های بسیار زیادی جهت پیاده‌سازی و اجرای سیستم‌های آموزش الکترونیکی صرف می‌شود. همچنین کارشناسان و مجریان برگزاری هر ساله برای پژوهش و بررسی آمار و تهیه نمودارهایی در روند پیشرفت، شکست دانشجو و علل ایجاد آن هزینه‌هایی را صرف می‌کنند. در سیستمی که دانشجو از راه دور مشغول به تحصیل است (DE) نیاز به مشاوره و راهنمایی یک امر ضروری است. همچنین برای مسئولین و اساتید جالب است بدانند دانشجویان از کدام قسمت‌های سایت بیشترین استفاده را کرده‌اند. قسمت‌های مختلف چه

الگوریتم های مختلف جهت پیش بینی و پیدا کردن قوانین انجمنی استفاده گردیده است. در پژوهش های مختلف، پیش بینی و استخراج قوانین انجمنی، بررسی شده است. در پژوهش رومرو و همکاران بررسی جامعی از پژوهش های این حوزه از سال ۱۹۹۵ تا ۲۰۰۵ ارائه شده است و خلاصه ای از دستاوردهای هر یک از پژوهش ها معرفی گردیده است. کاسترو تحلیل دقیق تری از مشکلات آموزش الکترونیکی، که به وسیله روش های داده کاوی قابل



شکل ۲ روند اجرایی پژوهش فعلی

پاسخگویی هستند، به عمل آورده است [۶]. علاوه بر این روش های رده بندی به مدل سازی داده ها با استفاده از مجموعه آزمون و ویژگی های آنها می پردازند. در آموزش الکترونیکی این روش ها معمولاً برای پیش بینی یک ویژگی مشخص، برحسب الگوهای پیشین مورد استفاده قرار می گیرد. چن روش های رده بندی را برای بررسی واکنش گروه های مختلف دانشجویان به راهبردهای مختلف

کنند و ظرفیت و گروه های مناسبی را ارائه دهند؛ همچنین وقتی در دوره N پیش بینی نمرات دانشجو در دوره N+1 صورت می گیرد. مسئولین امر می توانند در دوره N علل آفت را بررسی و راهکارهای خاصی را اعمال کنند.

کارشناس گروه وظیفه دارد دانشجویان را در حین انتخاب واحد، راهنمایی و دانشجو را جهت اخذ دروس مناسب با توجه به شرایط تحصیلی دانشجو راهنمایی کند. حال این سؤال مطرح است که کارشناس های گروه این مشاوره و راهنمایی را چطور انجام می دهند؟ در جواب این سؤال باید گفت کارشناسان گروه با کسب تجربه و آموزش هایی که از قبل دیده اند و با مشاهده انتخاب واحدهای دانشجویان در ترم های قبل، این تجربه ها را کسب نموده اند. سؤال مطرح شده این است چه کارشناس گروهی بهتر از پایگاه داده انتخاب واحد خود دانشجویان وجود دارد. شاید کارشناسان گروه تجربه ۲ یا ۳ ترم داشته باشند؛ ولی در یک پایگاه داده، انتخاب واحد و اطلاعات زیاد دیگری از دانشجویان در چندین ترم تجربه های بسیار نهفته است. با داده کاوی می توان این تجربه ها را استخراج و از آن جهت مشاوره و راهنمایی استفاده کرد. با استفاده از تحلیل سبد انتخاب واحد دانشجویان می توان تجربه های زیر را کسب کرد:

- چه درس هایی بهتر است با هم انتخاب شود.
- چه درس هایی در چه ترم هایی انتخاب شود، بهتر است.
- چه درس هایی بهتر است با هم برداشته نشوند.
- چه درس هایی در انتخاب واحد دانشجو باعث افت تحصیلی می شوند.
- با توجه به شرایط دانشجو چه درس هایی انتخاب شود، بهتر است.

با وجود تحلیل سبد انتخاب واحد و ارائه این تجربه ها به دانشجو در حین انتخاب واحد، از طریق سامانه دیگر نیاز به هیچ گونه کارشناس گروه و نیروی انسانی نیست.

دانشجویان آموزش الکترونیکی با مراجعه به سامانه آموزشی، رفتار خاصی را انجام می دهند. با مشاهده رفتار دانشجویان در یک بازه زمانی خاص می توان رفتار و الگوی رفتاری آن ها را به دست آورد و جهت مشاوره یا ارتقای کیفیت آموزشی استفاده کرد. در شکل شماره ۲ روند اجرای پژوهش فعلی نمایش داده شده است. در این میان از

هملینن قوانین انجمنی را برای برطرف کردن مشکلات دانشجویان در محیط آموزشی و ارائه مشاوره به آنها مورد استفاده قرار داده است [۱۳].

زین نیز از این قوانین برای هدایت فعالیت‌های دانشجویان و پیشنهاد محتوای آموزشی استفاده کرده است [۱۴]. هونگ در پژوهشی به طور خاص مناسب‌ترین محتوای آموزشی برای ارائه به دانشجو را مشخص می‌کند [۱۵].

همچنین لو نیز قوانین استخراج شده برای بهینه سازی محیط آموزش مجازی برحسب مواردی که از منظر دانشجو جالب است مورد بررسی قرار داده است [۱۶].

مرکلو روشی را برای تعیین ویژگی‌های اصلی متمایز کننده دانشجویان از منظر کارایی ارائه کرده است [۱۷]. در تمام مقالات و پژوهش‌هایی که در زمینه داده‌کاوی صورت پذیرفته است، جمع‌آوری داده و پیش پردازش آن جزء اصلی‌ترین مراحل می‌باشد و بیشترین زمان و هزینه را به خود اختصاص می‌دهد. ماده اولیه به کار رفته در داده-کاوی، داده است؛ از این رو سنگ بنای عملیات داده‌کاوی خوب، به کارگیری و دسترسی به داده‌های اولیه خوب و مناسب است، که از آن به آماده‌سازی یا پیش‌پردازش داده‌ها یاد می‌شود [۱۸]. فقدان داده با کیفیت، برابر با فقدان کیفیت در نتایج کاوش است و ورودی بد، خروجی بد به دنبال دارد. از آنجا که هدف اصلی سامانه‌های آموزشی، آموزش است و برای کاربردهای تحلیلی و داده‌کاوی ایجاد نگشته‌اند.

آموزشی مورد بررسی قرار داده است [۷]. بهروز مینایی بیدگلی به پیش بینی کارایی دانشجویان و نمرات پایانی آنها با استفاده از ترکیبی از روش‌های رده‌بندی پرداخته است [۸]. بیکر و همکاران استفاده نادرست دانشجویان از محیط آموزشی و رفتار غیر مسئولانه آنها را با استفاده از روش‌های رده‌بندی تشخیص داده‌اند [۹]. همچنین کتسینتیس کارایی دانشجویان را در محیط آموزشی در کنار ارتباط ویژگی‌های تأثیرگذار در بهبود آن مورد بررسی قرار داده است [۱۰].

یودلسن دانشجویان را بر حسب الگوی استفاده به دو گروه متمایل به خطا و متمایل به درستی تقسیم بندی کرده و با استفاده از نتایج این دسته‌بندی، علل معمول رویداد اشتباه در دانشجویان مورد بررسی قرار گرفته است [۱۱]. کار دیگری که در این زمینه انجام شده است، استخراج قوانین انجمنی می‌باشد. قوانین انجمنی برای استخراج ارتباط میان ویژگی‌های مختلف در پایگاه داده به کار می‌رود. نتایج این الگوریتم‌ها معمولاً به صورت مجموعه‌ای از قوانین  $X \rightarrow Y$  ارائه می‌شود که در آن  $X, Y$  مجموعه‌ای از ویژگی‌ها هستند. وبیلزهل قوانین انجمنی را برای ساخت عامل‌های توصیه دهنده به دانشجویان مورد استفاده قرار داده است. این عوامل فعالیت‌های آموزشی مختلف را مناسب با الگوی استفاده دانشجویان به آنها توصیه کرده و میان‌برهایی برای حذف منابع آموزشی غیر ضروری به آنها پیشنهاد می‌کنند [۱۲].

id	time	userid	ip	course	module	cmid	action	url
1	1190371124	2	127.0.0.1	1	user	0	update	view.php?id=2&course=1
2	1190371258	2	127.0.0.1	1	course	0	view	view.php?id=1
3	1190371419	2	127.0.0.1	1	user	0	view	view.php?id=2&course=1
4	1190371477	2	127.0.0.1	1	user	0	update	view.php?id=2&course=1
5	1190371477	2	127.0.0.1	1	user	0	view	view.php?id=2&course=1
6	1190371483	2	127.0.0.1	1	user	0	logout	view.php?id=2&course=1
7	1190371503	2	127.0.0.1	1	user	0	login	view.php?id=2&course=1
8	1190371504	2	127.0.0.1	1	course	0	view	view.php?id=1
9	1190371793	2	127.0.0.1	1	course	0	view	view.php?id=1
10	1190371817	2	127.0.0.1	1	user	0	logout	view.php?id=2&course=1
11	1190371832	2	127.0.0.1	1	user	0	login	view.php?id=2&course=1
12	1190371833	2	127.0.0.1	1	course	0	view	view.php?id=1
13	1190373026	2	127.0.0.1	1	course	0	view	view.php?id=1
14	1190373090	2	127.0.0.1	1	course	0	view	view.php?id=1
15	1190373343	2	127.0.0.1	1	course	0	view	view.php?id=1
16	1190289170	2	127.0.0.1	1	course	0	new	view.php?id=2

شکل ۳ نمایی از اطلاعات موجود در پایگاه داده مودل

جدول ۱ جدول رفتار دانشجو

نام صفت	شرح صفت
USER ID	شماره کاربری کاربر
C.VIWE	تعداد مشاهده درس
C.POST	تعداد پست مطلب
C.FROUM	تعداد مشاهده فروم ها
C.UPLOAD	تعداد ارسال تمرین
GRADE	نمره دانشجو

### ۳- نتایج و بحث

در این قسمت به ترتیب نتایج حاصل از پژوهش ارائه شده است.

**قوانین انجمنی:** در واقع کاری که ما در این تحقیق به آن پرداخته‌ایم، این است که به تحلیل انتخاب واحدهای مشابه که توسط دانشجویان دیگر انجام شده است بپردازیم. هر انتخاب واحد خود ابتدا به منزله یک سبد خرید و هر درس به مثابه یک محصول یا کالا می‌باشد. دروسی را که بیشتر با هم اخذ شده‌اند و به عبارتی در بیشتر مواقع در یک سبد خرید قرار گرفته‌اند شناسایی شده و در انتخاب واحد به دانشجویان جدید پیشنهاد می‌گردد. با این کار عملاً ما نقش یک کارشناس گروه را که راهنمایی لازم جهت اخذ و یا عدم اخذ یک درس را ارائه می‌نماید، ایفا کرده‌ایم. به عنوان مثال در مقاله فعلی از طریق الگوریتم‌های سبد خرید به دانشجو پیشنهاد گرفتن یک درس خاص به صورت زیر داده می‌شود:

(جنسیت مونث باشد و پایگاه‌داده و برنامه سازی ۲ در انتخاب واحد موجود بود) IF  
 ("شما بهتر است درس مدار منطقی را نیز انتخاب کنید") PRINT  
 (سیستم عامل و برنامه سازی ۱ در انتخاب واحد موجود بود و نیمسال دوم) IF  
 ("شما بهتر است گرافیک را نیز انتخاب کنید") PRINT

نکته قابل تأمل این است که قوانینی که از طریق داده‌کاوی پیدا و به دانشجو پیشنهاد داده می‌شود، خیلی وسیعتر و بهتر از پیشنهاد کارشناس گروه مجازی است؛ چرا که با استفاده از داده‌کاوی قوانینی نیز استخراج می‌شوند که به صورت صریح در سیستم وجود ندارد و حتی شاید کارشناسان گروه با این قوانین ناآشنا باشند.

جمع‌آوری و آماده‌سازی داده در آنها بسیار دشوار می‌باشد. به طور مثال مودل (MOODLE) که جزء معروف‌ترین سیستم مدیریتی یادگیری می‌باشد، خلاصه‌ای از اطلاعات دانشجویان مانند ساعت ورود، تعداد صفحات مورد بازدید، عملیات انجام داده در هر صفحه، تالارهای گفتگو و غیره را نمایش می‌دهد. در شکل شماره ۳ نمایی از پایگاه داده سیستم مدیریت مودل که در آن اطلاعات فعالیت‌های دانشجو ذخیره گشته است، قابل مشاهده می‌باشد. در پژوهش کنونی از داده‌های ذخیره شده از بانک اطلاعاتی سیستم مدیریت آموزش مجازی مودل دانشگاه علم و صنعت استفاده شده است. در بسیاری از پروژه‌های گرد-آوری اطلاعات نحوه استفاده کاربران از سیستم و پروفایل کاری آن‌ها نیازمند فرآیند طاقت‌فرسایی است [۱۹]. به طور مثال در داده‌کاوی آموزشی مجبور به بررسی اطلاعات وب سرورها و استخراج اطلاعات مربوط به دسترسی منابع هستیم؛ اما امروزه بسیاری از سیستم‌های مدیریتی کار را آسانتر نموده‌اند. سیستم مدیریت آموزش مودل یک مکانیزم ثبت وقایع و فعالیت‌های کاربر برای استفاده در سیستم گزارش‌گیری دارد. کلیه اطلاعات و فعالیت‌های کاربرد در بانک اطلاعاتی رابطه‌ای نگهداری می‌شود. در جدول شماره ۱ لیست رفتارهای خاصی که برای کشف قوانین انجمنی از رفتار دانشجو لازم است گردآوری شده است. با استفاده از این جدول در فصل بعد رفتار دانشجو در سامانه مورد تحلیل قرار خواهد گرفت و قوانین انجمنی نهفته در سیستم کشف می‌شود. سایر جداول نیز برای پیش بینی نمره به طریق مشابه استخراج گردید. قابل ذکر است اطلاعات دروس انتخابی دانشجو و انتخاب واحد در سیستم آموزش مودل وجود ندارد. دانشگاه علم و صنعت به دلایل خاصی حاضر به ارائه قسمتی از بانک اطلاعاتی که انتخاب واحد دانشجویان در آن وجود دارد نشد. برای رفع مشکل پایگاه‌داده انتخاب واحد دانشگاه آزاد واحد سنج دریافت و جایگزین آن قسمت گردید.

نویسی پیشرفته را نیز اخذ نماید. این قانون در همه مواقع صحیح می‌باشد.

IF (ALGORITHMDESIGN=YES AND ECONOMICS = YES) Then OPERATINGSYSTEM = YES CONF:(0.98)

معنا و مفهوم این قانون این است که اگر دانشجویی در برگه انتخاب واحد خود دروس طراحی الگوریتم و اقتصاد مهندسی را اخذ کرده است، بهتر است که درس سیستم عامل را نیز اخذ نماید. این قانون در ۰/۹۸ مواقع صحیح می‌باشد.

IF (ADVANCEDPROGRAMMING =YES AND LOGIC=YES AND LANGUAGE= YES) Then DATASTRUCTURE = YES CONF:(0.85)

معنا و مفهوم این قانون این است که اگر دانشجویی در برگه انتخاب واحد خود دروس برنامه نویسی پیشرفته و مدار و زبان تخصصی را اخذ کرده است، بهتر است که درس ساختمان داده را نیز اخذ نماید. این قانون در ۰/۸۵ مواقع صحیح می‌باشد.

IF (COMPUTERARCHITECTURE =YES AND COMPUTENETWORK = YES ) Then ALGORITHMDESIGN = YES CONF:(0.90)

معنا و مفهوم این قانون این است که اگر دانشجویی در برگه انتخاب واحد خود دروس معماری کامپیوتر و شبکه‌های کامپیوتری را اخذ کرده است، بهتر است که درس طراحی الگوریتم را نیز اخذ نماید. این قانون در ۰/۹۰ مواقع صحیح می‌باشد.

IF (DIFFERENTIALEQUATIONS =YES AND DISCRETEMATHEMATICS = YES) THEN DATASTRUCTURE = YES CONF:(0.80)

معنا و مفهوم این قانون این است که اگر دانشجویی در برگه انتخاب واحد خود دروس معادلات دیفرانسیل و ریاضیات گسسته را اخذ کرده است، بهتر است که درس ساختمان داده را نیز اخذ نماید. این قانون در ۰/۸۰ مواقع صحیح می‌باشد.

**درخت تصمیم:** اساساً درخت‌های تصمیم مثل ID3 ، C4.5 و C5 به این منظور ایجاد شدند که از روی یک سری داده آماری، درختی ایجاد کنند که تا برای داده‌های آماری مشابه بتوان از آن درخت برای کلاسه‌بندی استفاده کرد. برای مثال یک سری داده اولیه انتخاب می‌کنیم و بر

در سیستم‌های آموزش الکترونیکی می‌توان به محض انتخاب واحد توسط دانشجو با توجه به قوانین کشف شده در سیستم، دانشجو را در انتخاب واحد کمک کرد و به او پیشنهاد اخذ یا حذف درس خاصی داد. در سیستم‌های آموزش الکترونیکی به علت نبود کارشناس گروه مستقیم و نبود هر گونه مشاوره‌ای می‌توان دانشجویان را به صورت زیر در امور آموزش راهنمایی کرد:

- پیدا کردن روابط و قوانین موجود در سیستم
- ارائه برخط توصیه‌ها به دانشجو
- اعمال قوانین کشف شده در انتخاب واحد.
- ایجاد مشاوره مجازی

در این مقاله قوانینی استخراج گردیده که با قاطعیت می‌توان گفت حتی به ذهن کارشناسان گروه خطور نکرده و یا شاید تا به حال به آن توجه نشده است. پس از فرآیند جمع‌آوری، پاک‌سازی و منسجم کردن داده‌ها از منابع مختلف باید قوانین انجمنی و الگوهای ضمنی و قوانینی که در داده‌ها پنهان هستند را استخراج کنیم. همان طور که می‌دانیم یافتن قوانین انجمنی در حالت کلی شامل یافتن ارتباطهایی میان صفات یا مقدار صفات در یک پایگاه داده است که به صورت زیر بیان می‌گردند:

IF <برخی شرطها برقرار بودند> THEN

<حمق‌دیری را برای سایر صفات پیش‌بینی کن >

به قسمتی که در کنار IF می‌آید، مقدم و به قسمت دیگر تالی گفته می‌شود. معیارهایی که در یافتن قوانین انجمنی استفاده می‌شوند، به طور معمول عبارتند از: Support Count : تعداد رکودهایی که تمام شرطهای قانون را برآورده می‌سازند.

Comprehensibility : تعداد صفتهای موجود در قانون [۲۰]. پس از اعمال الگوریتم ۴۰۰ قانون کشف و در میان آن‌ها ۱۰ قانون به عنوان قوانین اصلی انتخاب شد. در زیر لیستی از قوانین همراه با تشریح آن آمده است.

IF (TASTRUCTURE=YES AND LOGIC=YES AND STATISTICS = YES) Then ADVANCEDPROGRAMMING = YES CONF: (1)

معنا و مفهوم این قانون این است که اگر دانشجویی در برگه انتخاب واحد خود دروس ساختمان داده و مدار و آمار و احتمالات را اخذ کرده است، بهتر است که درس برنامه

خود را عوض و از مدل‌های آموزشی دیگر یا منابع دیگر جهت آموزش استفاده نمایند. نحوه کار و مشخصات این درخت تصمیم در جدول شماره ۲ آمده است. ایجاد این نوع درخت در نرم افزار وکا صورت گرفته است. فیلد هدف یا به عبارت دیگر فیلدی که پیش بینی آن انجام می‌شود، نمره Database می‌باشد. تعداد دفعات اجرای این الگوریتم روی داده‌های آموزشی ۱۰ بار می‌باشد.

جدول شماره ۲ اطلاعات اجرای الگوریتم

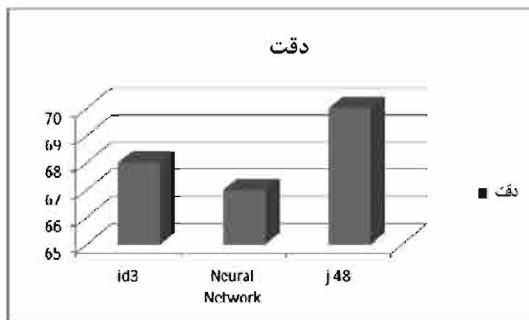
Scheme	weka.classifiers.trees
Instances	263
Attributes	Gender , age, math, strucher program1, program2 , database
Test mode	10-fold cross-validation

میزان دقت در پیش‌بینی و صحت الگوریتم بر روی داده‌های تست به صورت جدول شماره ۳ می‌باشد. داده‌هایی که برای تست مدل در نظر گرفته می‌شوند، داده‌هایی هستند که کل جامعه آماری را پوشش می‌دهند. به طور تقریب دقت مدل درخت تصمیم ساخته شده ۷۰٪ می‌باشد. یعنی از هر ۱۰۰ نمره مربوط به دانشجو ۷۰ نمره درست پیش‌بینی می‌شوند.

جدول شماره ۳ دقت کلاس بندی

صحت / خطا	تعداد نمونه	دقت
Correctly Classified Instances	184	69.962 %
Incorrectly Classified Instances	79	30.038 %

همچنین در نمودار شماره ۱ مقایسه‌ای بین سایر روش‌های کلاس‌بندی انجام گرفته است. با توجه به نمودار، روش درخت تصمیم J48 از همه مناسب تر می‌باشد.



نمودار شماره ۱ دقت سایر روش‌ها

اساس یک الگوریتم مشخص درخت را ایجاد می‌کنیم. سپس با داده‌های تست درخت ایجاد شده را ارزیابی می‌کنیم، تا ببینیم چقدر کارا است. این درخت، ساختار خیلی سریع و کارایی برای تصمیم‌گیری روی کلاس یک داده جدید خواهد بود. در این بخش با به کارگیری الگوریتم که در فصل‌های پیشین معرفی گردید، مدلی از نوع درخت تصمیم ساخته خواهد شد [۲۱].

برای اعمال درخت تصمیم سه عمل اساسی انجام می‌شود :

۱- ایجاد مجموعه آموزشی

۲- ایجاد مجموعه تست

۳- اعمال الگوریتم C.45

وکا ۱۰ بار جداگانه و هر بار با استفاده از مجموعه آموزشی و مجموعه تست جداگانه‌ای برای طراحی مدل اقدام می‌کند و در نهایت بهترین مدل را به عنوان درخت تصمیم ارائه می‌دهد. مدل ساخته شده در این پژوهش از دقت مناسبی برخوردار است. با مدل ساخته شده می‌توان در سامانه‌های آموزش الکترونیکی در حین انتخاب واحد و با توجه به نمرات قبلی وی پیش‌بینی مناسبی از نمرات دروس انتخابی ترم جاری آن‌ها انجام داد. از پیش‌بینی صورت گرفته افراد مختلف به صورت زیر استفاده می‌کنند.

**دانشجویان:** در حین انتخاب واحد و قبل از سر رسید فصل امتحانات می‌توانند نمره دروس خود را مشاهده کنند و در صورت ضعیف بودن در آن درس امور پیش‌گیرانه را انجام دهند و نقطه قوت این مقاله نیز همین می‌باشد.

**مدیران و مجریان:** مدیران و مجریان قبل از شروع برنامه‌ریزی هر درس می‌توانند نتیجه کاری دانشجویان را پیش‌بینی کنند و در درس‌هایی که احساس می‌شود اکثر دانشجویان نمره آن‌ها کم می‌شود، عملیات کنترلی برای پیشگیری از آفت تحصیلی انجام دهند. هر ساله خیلی از مدیران دانشگاهی سهم چشمگیری از پژوهش‌های خود را بر بررسی آفت تحصیلی و عوامل مرتبط با آن می‌کنند.

**استادان:** استادان و مجریان آموزشی می‌توانند قبل از شروع کلاس نتایج و نمرات و دانشجویان را پیش‌بینی کنند و با توجه به نمرات سیاست‌های آموزش خاصی را اعمال نمایند؛ به طور مثال هنگامی که میانگین نمرات پیش‌بینی شده برای دانشجویان کم باشد، می‌توانند سیستم آموزشی



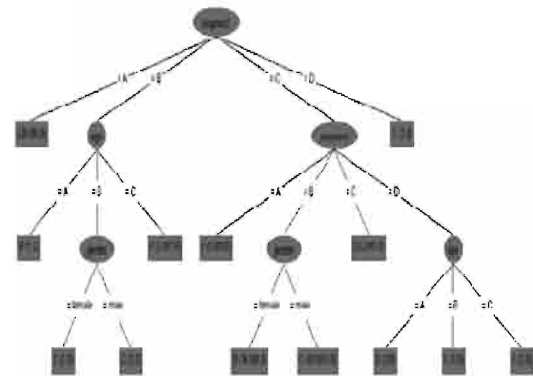
مشکل اساسی در این پژوهش و سایر پژوهش‌های انجام گرفته مشابه، داده و در دسترس نبودن داده می‌باشد. کارگروه داده‌کاوی آموزشی در سال ۲۰۰۸ با همکاری مرکز آموزش یادگیری پترزبورگ مستقر در دانشگاه کارنگی ملون به گردآوری مجموعه داده‌هایی در قالب استاندارد و مناسب برای استفاده در روش‌های داده‌کاوی آموزشی پرداخته است؛ که با در نظر گرفتن تمامی ملاحظات مربوط به حفظ مالکیت معنوی به صورت رایگان و از طریق سایت این مرکز در اختیار پژوهشگران این حوزه قرار گرفته است. انتشار عمومی این مجموعه داده در این پژوهش با در نظر گرفتن حقوق معنوی دانشگاه علم و صنعت می‌تواند جهت ارتقای سامانه آموزشی و همچنین کمک به این رشته نو پا در کشور باشد. به عنوان گام بعدی در راستای ادامه فعالیت‌ها می‌توان ابزاری برای بررسی و تحلیل و ارائه توصیه‌هایی در سطوح مختلف کاربری و مدیریتی در سیستم مدیریت آموزشی و یادگیری ایجاد نمود. این امکانات را می‌توان در قالب ماژول‌هایی متن باز در این سیستم‌ها اضافه شود.

### پی‌نوشت

- 1 Data Mining
- 2 Education Data Mining
- 3 Distance Education

### مراجع

- [1] Han J. and Kamber M., *Data Mining concepts and techniques*, Morgan Kaufmann publishers, 2006.
- [2] Romero C. and Ventura S., *Educational data mining: A survey from 1995 to 2005*, Expert Systems with Applications, 2007, pp.135-146.
- [3] Baker R. and Yacef K., *The State of Educational Data Mining in 2009: A Review and Future Visions*, Journal of Educational Data Mining, Vol.1, No.1, 2010, pp.3-17.
- [4] Romero C., Ventura S. and Garcia E., *Data Mining in Course Management Systems: MOODLE Case Study and Tutorial*, Computers & Education, Vol.51, No.1, 2008, pp.368-384.
- [5] Suneetha K.R. and Krishnamoorthi R., *Identifying User Behavior by Analyzing Web Server Access Log File*, International Journal of Computer Science and Network Security, Vol.9, 2009, p.6.



شکل شماره ۴ نمایی از درخت تصمیم

به جز الگوریتم C4.5 الگوریتم‌های کلاس‌بندی دیگری مانند Neural Network نیز اعمال گردید که در ادامه به مقایسه دقت این مدل‌ها می‌پردازیم. در شکل شماره ۴ قسمتی از درخت تصمیم ایجاد شده به علت بزرگ بودن درخت آمده است.

### ۴- نتیجه‌گیری

این پژوهش به ارائه راهبردی برای به کارگیری روش‌های داده‌کاوی در سامانه‌های آموزش و یادگیری می‌پردازد که به طور خاص به تحلیل اطلاعات مربوط به استفاده دانشجویان از محیط آموزش مجازی و ارتباط عوامل مرتبط با هم در سامانه آموزش مجازی اختصاص دارد. در این پژوهش از اطلاعات واقعی دانشجویان مرکز آموزش الکترونیکی دانشگاه علم و صنعت و دانشگاه آزاد اسلامی واحد سنندج استفاده شده است. ابتدا انبار داده با توجه به نیاز پژوهش استخراج گردید. این انبار داده بستر مناسبی برای بررسی رفتار دانشجویان است. پس از گردآوری و آماده‌سازی داده با استفاده از روش‌های داده‌کاوی قوانین و دانش موجود در انبار، داده پیدا و به عنوان توصیه به دانشجویان و اساتید پیشنهاد گردید. همچنین با استفاده از مدل درخت تصمیم به پیش‌بینی نمره دانشجویان در درس پایگاه داده پرداخته شد، که مدل ساخته شده از دقت مناسبی برخوردار است.

- [6] Castro F., Vellido A., Nebot A. and Mugica F., *Applying Data Mining Techniques toe-Learning Problems*, in *Studies in Computational Intelligence*, Berlin and Heidelberg: Springer-Verlag, **2007**, pp.183-221.
- [7] Chen G., Liu C., Ou K. and Liu B., *Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology*, *Journal of Educational Computing Research*, Vol.23, No.3, **2000**, pp.305-332.
- [8] Minaei-Bidgoli B. and Punch B., *Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System*, *Genetic and Evolutionary Computation*, Vol.2, **2003**, pp.2252-2263.
- [9] Baker R.S., Corbet A.T. and Koedinger K., *Detecting Student Misuse of Intelligent Tutoring Systems*, in *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, Maceio, Brazil, **2004**, pp.531-540.
- [10] Kotsiantis S.B., Pierrakeas C.J. and Pintelas P.E., *Predicting Students' Performance in Distance Learning Using Machine Learning Techniques*, *Applied Artificial Intelligence*, Vol.18, No.5, **2004**, pp.411-426.
- [11] Yudelson M.V. and et al., *Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS*, in *Proceedings of the AAAI Workshop on Educational Data Mining*, Boston, MA, USA, **2006**, pp.1-8.
- [12] Cocea M. and Weibelzahl S., *Can Log Files Analysis Estimate Learners' Level of Motivation?*, in *Proceedings of the 14th Workshop on Adaptation and User Modeling in Interactive Systems*, Hildesheim, Germany, **2006**, pp.32-35.
- [13] Hamalainen W. and Vinni M., *Comparison of machine learning methods for intelligent tutoring systems*, in *Proceedings of the 8th international conference in intelligent tutoring systems*, Taiwan, **2006**, pp.525-534.
- [14] Zaiane O., *Building a recommender agent for e-learning systems*, in *Proceedings of the International Conference on Computers in Education*, **2002**, pp.55.
- [15] Hwang G.J., Hsiao C.L. and Tseng C.R., *A computer-assisted approach to diagnosing student learning problems in science*, *Journal of Information Science and Engineering*, Vol.19, **2003**, pp.229-248.
- [16] Lu J., *Personalized e-learning material recommender system*, in *Proceedings of the International conference on information technology for application*, Utah, USA, **2004**, pp.374-379.
- [17] Markellou P., Mousourouli I., Spiros S. and Tsakalidis A., *Using semantic web mining technologies for personalized e-learning experiences*, in *Proceedings of the web-based education conference*, Grindelwald, Switzerland, **2005**, pp.461-826.
- [18] Ramli A.A., *Web usage mining using apriori algorithm: UUM learning care portal case*, in *Proceedings of the International Conference on Knowledge Management*, Malaysia, **2005**, pp.1-19.
- [19] Minaei - Bidgoli B., Tan P. and Punch W., *Mining interesting contrast rules for a web-based educational system*, in *Proceedings of the international conference on machine learning applications*, Louisville, KY, **2004**, pp.1-8.
- [20] Romero C., Ventura S. and Bra P.D., *Knowledge discovery with genetic programming for providing feedback to courseware author*, *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, Vol.14, No.5, **2004**, pp.425-464.
- [21] Yu P., Own C. and Lin L., *On learning behavior analysis of web based interactive environment*, in *Proceedings of the Workshop on Implementing Curricular Change in Engineering Education*, Oslo, Norway, **2001**, pp.1-10.