

مقایسه تکنیک های داده کاوی جهت شخصی سازی مطلوبتر در آموزش الکترونیک

هومن پاینده فر^۱، حسن سید رضی^۲، مسعود رهگذر^۳، احمد فراهی^۴

^۱دانشجوی کارشناسی ارشد، دانشکده کامپیوتر، دانشگاه پیام نور، تهران، ایران
hooman.payandehfar@gmail.com

^۲دانشیار دانشکده کامپیوتر، دانشگاه تهران، تهران، ایران

^۳استادیار دانشکده کامپیوتر، دانشگاه تهران، تهران، ایران

^۴استادیار دانشکده فناوری اطلاعات و ارتباطات، دانشگاه پیام نور، تهران، ایران

چکیده

ظهور و رشد بی نظیر اینترنت و وب در دهه های اخیر باعث شده است کاربردهای مختلف این تکنولوژی‌ها در زندگی روزمره بشری تاثیرات مثبت فراوانی داشته باشند. یکی از این موارد تاثیرگذار آموزش الکترونیک بوده است. آموزش الکترونیک می‌تواند به نوعی پدیدآورنده عدالت در آموزش باشد تا تمام افراد جامعه فارغ از موقعیت مکانی و زمانی به تحصیل علم پردازند. داده کاوی از شاخه‌هایی است که به وسیله شخصی سازی، می‌تواند در راستای افزایش کیفی سیستم‌های آموزش الکترونیک تأثیر بی‌بدیلی داشته باشد. هدف از داده کاوی استخراج دانش از مجموعه داده‌های بزرگ است و در این مقاله نیز به بررسی نحوه استفاده از داده کاوی در افزایش کیفیت سیستم‌های آموزش الکترونیک پرداخته شده است. در این مقاله، پیش‌گویی نتیجه تحصیلی کاربران سیستم آموزش الکترونیک را با استفاده از روش‌های یادگیری ماشین، انجام داده این که روش‌های درخت تصمیم، شبکه‌های عصبی، الگوریتم بیز و رگرسیون منطقی مورد استفاده ما واقع شده اند که روش رگرسیون منطقی با ۸۰.۹۶۶ درصد بیشترین دقت پیش‌گویی را در میان روش‌های استفاده شده داشته است. روش‌های درخت تصمیم و بیز ساده نیز دقت مشابه ۸۰.۱۱۳ درصد را داشته‌اند و در رده بعدی نیز شبکه عصبی با ۷۹.۲۶۱ درصد کمترین دقت را داشته است. نتایج بدست آمده می‌تواند با پیش‌بینی دقیق تر برای مازول پیش‌بینی نتیجه تحصیلی فرآگیران، کمک شایانی در افزایش کیفیت سیستم آموزش الکترونیک و شخصی سازی بهتر آن نماید.

کلمات کلیدی

آموزش الکترونیک، داده کاوی، پیش‌بینی، رگرسیون منطقی، درخت تصمیم، بیز ساده، شبکه عصبی.

بسیاری از رشته‌های تحت وب مطالب بدون توجه به تفاوت یادگیری دانشجویان به صورت یکنواخت آموزش داده می‌شود.

سیستم آموزشی هوشمند و سازگار در محیط‌هایی که نیاز به یادگیری شخصی غنی تری احساس می‌شود، راهگشا خواهد بود. سیستم آموزشی هوشمند و شخصی شده تلاش می‌کند بوسیله تهیی طرحی از ویژگی‌های شخصی، علائق و معلومات فرآگیران، شکلی از آموزش اختصاصی را به فرآگیران ارائه دهد. داده کاوی یا دانش استخراج خودکار الگوهای مفید از مجموعه داده‌های بزرگ،

۱- مقدمه

در سال‌های اخیر، نوآوری‌های مهمی در سیستم‌های آموزشی، در راستای معرفی تکنولوژی‌های نوینی همچون آموزش تحت وب رخ داده است [1]. این نوع از آموزش به کمک رایانه و کاملاً بدون وابستگی زمانی و مکانی خاص است [2]. این شیوه آموزشی اکنون بطور ویژه ای مورد توجه قرار گرفته است و در طی چند سال گذشته هزاران رشته تحت وب ثبت و گسترش یافته است. اما در

شناخته می شود. یک برگ از این درخت مقدار محتمل از یک مشخصه کامل رکوردهای تعریف شده را مشخص می کند. درخت های مذکور می توانند بعنوان مجموعه هایی از قوانین "if-then-else" نمایش داده شوند.

این روش توسط الگوریتم ID3 و جانشین وی C4.5 پیش برده می شود. C4.5 الگوریتم تعمیم یافته ای از ID3 است که مقادیر دست نیافتنی، دامنه مقادیر مشخصه پیوسته، هرس درخت های تصمیم و قاعده اشتراق را بدست می آورد.

۳-۲ روش شبکه عصبی [5]

عصب، سلول بیولوژیکی ویژه ای با توانایی پردازش اطلاعات می باشد. شبکه عصبی، یک گراف لایه ای است که خروجی یک گره، ورودی یک یا چند گره از لایه بعدی را تامین می کند.

ادراک چند لایه ای (MLP) یک پیش خورد پایه ای شبکه عصبی مصنوعی است که از انتشار الگوریتم برای آموزش استفاده می کند. یعنی اینکه در طول آموزش، اطلاعات در سراسر قسمتهای پیشین شبکه منتشر می شود و برای بروز کردن وزنهای اتصال مورد استفاده قرار می گیرد.

خروجی های اعصاب یک لایه، ورودی های لایه بعدی را فراهم می کنند. پس وقتی یک n تایی (x_1, \dots, x_n) بعنوان لایه ورودی یک شبکه به کار برده می شود، خروجی m تایی (c_1, \dots, c_m) حاصل می شود، که مقدار c_i ۱ است اگر n تایی ورودی به کلاس i تعلق داشته باشد و در غیر اینصورت صفر است.

۴-۲ روش رگرسیون منطقی [7]

مدل رگرسیون، تعمیم یافته اساس تئوری مدل رگرسیون خطی را نشان می دهد که شیوه رگرسیون خطی می تواند برای مدل هایی که متغیر پاسخ آنها گستته می باشد، استفاده شود. یک مدل عمومی از مدل خطی تعمیم یافته، رگرسیون منطقی می باشد. مدل رگرسیون منطقی احتمال وقایع اتفاق افتاده در یکتابع خطی از مجموع متغیرهای پیشگویی شده را نشان می دهد.

۳- داده کاوی در آموزش الکترونیک

آموزش از راه دور و یا یادگیری از تکنیک ها و متدهایی است که دستیابی به برنامه های آموزشی را برای دانشجویانی که به شکل زمانی و یا مکانی از مدرس بدور هستند فراهم می سازد. شاخه های فرعی متعددی برای آموزش از راه دور وجود دارد: آموزش مکاتبه ای مبتنی بر کاغذ، آموزش از طریق نوارهای ضبط شده، آموزش به کمک رایانه، آموزش چند رسانه ای، آموزش الکترونیکی و آموزش تحت وب.

می تواند برای دستیابی به مدل فرایند یادگیری و مدل یابی فرآیندان به کار رود و با جستجو و یافتن اطلاعات آموزشی سودمند بر مبنای آموزشی، در ارزیابی و بهبود سیستم آموزش نیز استفاده شود. هدف از ارائه این مقاله، اعمال روش های داده کاوی غیر گرافیکی به منظور پیش بینی رفتارهای آتی کاربران سیستم های آموزش الکترونیک بوده است. لذا پس از شرح این روش های داده کاوی برای اهداف پیش بینی، به مقایسه دقت این روش ها در پیش بینی نتیجه تحصیلی فرآیندان آموزش الکترونیک پرداخته شده است. به دلیل آنکه دقت پیش بینی در یک سیستم آموزشی هوشمند (ITS)، می تواند اهمیت زیادی در شخصی سازی بهتر آن ITS داشته باشد، بدین منظور نتایج بدست آمده از این مقاله می تواند کمک شایانی در رویکرد شخص سازی و ارائه سیستم توصیه گر مطلوبتر به عنوان کارهای آتی این تحقیق داشته باشد.

۲- داده کاوی و پیش بینی

داده کاوی، کاربرد تکنیک های یادگیری ماشین بر روی پایگاه داده برای کشف دانش و الگوهای نهفته در داده های پایگاه داده می باشد.

در ادامه، روش های غیر گرافیکی داده کاوی به منظور پیش بینی رفتارهای آتی فرآیندان شرح داده شده است.

۱-۲ استنتاج بیزی [4]

یکی از مهمترین روش های آماری در داده کاوی روش استنتاج بیزی است. طبقه بندی بیزی ساده، روش آسان و موثری را برای داشت طبقه بندی فراهم می آورد.

مجموعه ورودی های a_j که $(j=1,2,\dots,n)$ به عنوان ورودی به ابزاری داده می شود و هر ورودی توسط این ابزار که قادر به اندازه گیری خصیصه هاست بررسی می شود. هر الگو به شکل یک بردار x_i در نظر گرفته می شود و الگوی a_j به طبقه w_i متعلق است، زمانی که احتمال تعلق ورودی a_j به طبقه w_i بیشترین باشد.

۲-۲ روش مبتنی بر درخت تصمیم [6]

روش های مبتنی بر درخت تصمیم روش هایی عمومی و پرطرفدار برای استفاده در زمینه داده کاوی می باشند. رده بند درخت تصمیم از روشهای سلسله مراتی یا لایه ای برای طبقه بندی کردن استفاده می کند.

هر راس این درخت نمایشگر یک آزمایش یا تصمیم یکتا می باشد. پالهای خروجی یک راس، متناظر با همه احتمالهای حاصل از آزمایش روی راس می باشد. این پیشامدها مجموعه داده ها را به چندین زیرمجموعه افزای می کند که توسط برگهای این درخت

مجموعه داده مورد استفاده ماء، اطلاعات ۳۵۲ فرآگیر در گیر در فعالیت های سیستم آموزش الکترونیک دانشگاه CMU در ایالت پنسیلوانیای آمریکا است. از این تعداد ۱۵۵ نفر در امتحان نهایی قبول و ۱۹۷ نفر نیز مردود یا غایب شده اند. در این مجموعه کمینه، بیشینه و متوسط سن فرآگیران به ترتیب ۴۶، ۲۴ و ۳۱.۲ سال است. فرآگیران باید در طول ترم حداقل سه تکلیف از ۴ تکلیف موجود را انجام دهند. در صورتی که فرآگیران این سه تکلیف را با موفقیت انجام داده باشند قادر خواهند بود در آزمون نهایی شرکت کنند.

ویژگی های موجود در مجموعه داده، همراه با مقادیر آنها در جدول ۱ نشان داده شده است.

مقادیر	ویژگی
مرد، زن	جنس
۲۴-۴۹	سن
مجرد، متاهل	خانواده
هیچ، یک یا دو، سه یا چهار	فرزند
خیر یا نیمه وقت، بله یا اضافه کار	کار
بله، خیر	سوانح کامپیوتری
بله، خیر	کار کامپیوتری
بله، خیر	تکلیف اول
مردود، غایب، خوب، عالی	آزمون اول
بله، خیر	تکلیف دوم
مردود، غایب، خوب، عالی	آزمون دوم
بله، خیر	تکلیف سوم
مردود، غایب، خوب، عالی	آزمون سوم
بله، خیر	تکلیف چهارم
مردود، غایب، خوب، عالی	آزمون چهارم
مردود یا غایب، قبول	آزمون نهایی

جدول ۱ ویژگی های استفاده شده درباره فرآگیر

توزیع مقادیر مختلف مربوط به هر کدام از ویژگی های نیز در شکل ۲ نمایش داده شده است.

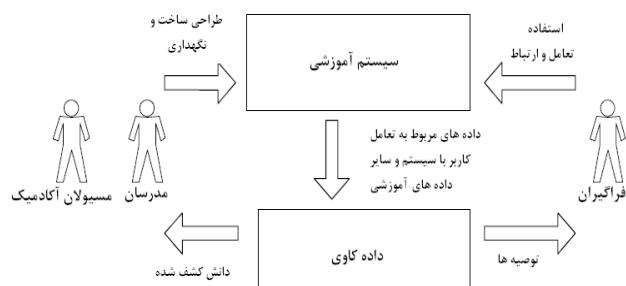
اخیراً مرسوم ترین، آموزش تحت وب (آموزش الکترونیک) است که دانشجویان را قادر می سازد به راحتی از طریق اینترنت آموزش بینند. آموزش تحت وب، فرمی از آموزش راه دور است که اطلاعات از طریق اینترنت منتقل می شود.

سازمان هایی که پایگاه های آموزش از راه دور را راه اندازی می کنند، بطور خودکار حجم عظیمی از اطلاعات را توسط خدمت گزارهای (سرورهای) وب تولید و در قسمت ثبت و قایع (log) خدمت گزارها جمع آوری می کنند. محیط های آموزشی تحت وب قادرند بیشتر رفتارهای یادگیری فرآگیران را ثبت و حجم عظیمی از پروفایل ها را فراهم کنند. اخیراً، علاقه زیادی به آنالیز خودکار اطلاعات تعاملات فرآگیر با محیط آموزش تحت وب توسط داده کاوی بوجود آمده است [8].

عملکرد داده کاوی در سیستم آموزش الکترونیک، یک سیکل تعاملی طرح فرضیه، آزمودن و اصلاح است (شکل ۱). دانش کاوش شده باید وارد چرخه سیستم و هدایت شده، و بطور کلی باعث تسهیل و افزایش سطح یادگیری شود.

همانطور که در شکل ۱ می بینیم آموزش دهندهان و مسئولان آکادمیک، مسئول طراحی، برنامه ریزی و حفظ سیستم آموزشی هستند که فرآگیران از آن استفاده می کنند و با آن تعامل دارند. با استفاده از تمام داده های موجود درباره دوره های آموزشی، فرآگیران و تعاملات، تکنیک های متفاوتی از کاوش اطلاعات به کار گرفته می شود تا اطلاعات مفیدی برای پیشرفت و بهبود فرآیند آموزش الکترونیکی بدست آید.

دانش بدست آمده نه تنها می تواند بوسیله تعلیم گر جهت بالابردن کیفیت آموزشی سیستم و برنامه ریزی برای آینده، استفاده شود بلکه می تواند در غالب سیستم آموزشی هوشمند و سازگار برای بهبود فرآیند یادگیری فرآگیران، استفاده گردد [9].

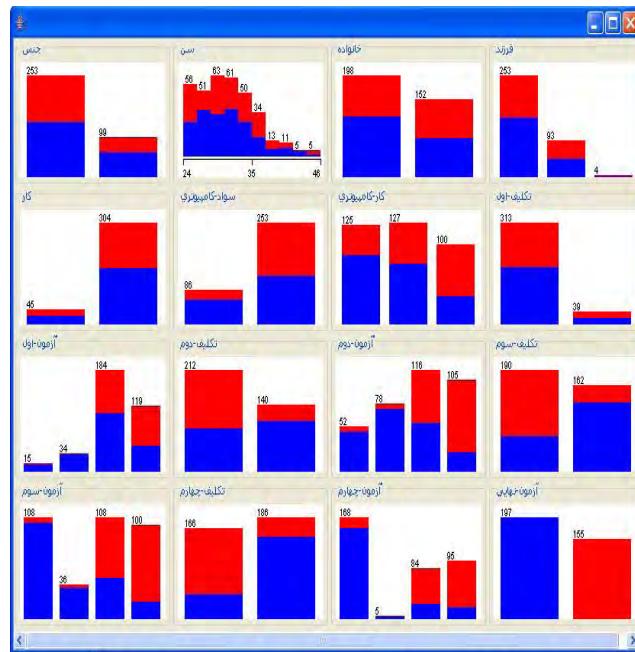


شکل ۱ چرخه استفاده از داده کاوی در سیستم آموزش الکترونیک

۴- مجموعه داده

۱-۶ روش ارزیابی متقاطع

در روش ارزیابی متقاطع برای تعیین میزان کیفیت یک طبقه بنده کننده، مقداری از داده آموزشی به عنوان مجموعه تست استفاده می شود. به عنوان مثال ۹۰٪ مجموعه داده به عنوان داده آموزشی استفاده می شود و ۱۰٪ بقیه برای تست روش استفاده می شود. تعداد نمونه هایی که به طور صحیح طبقه بنده می شوند به عنوان دقت آن روش لحاظ می گردد. این فرایند ۱۰ بار تکرار می شود و هر مرتبه با نگاه داشتن ۱۰٪ داده آموزشی اجرا می شود. هنگامی که این فرایند تکمیل می شود، میانگین مقادیر دقت مراحل به عنوان میزان دقت نهایی آن روش لحاظ می شود. اگر میزان مقدار داده تست نگاه داشته شده ۱۰ درصد کل داده آموزشی باشد به این شیوه، ارزیابی متقاطع ده-دسته گفته می شود. می توان تعداد دسته ها را به غیر از ده نیز انتخاب نمود. به عنوان مثال اگر ۵٪ از داده به عنوان تست نگاه داشته شود، آنگاه ارزیابی متقاطع ۲۰-دسته را خواهیم داشت.



شکل ۲ توزیع مقادیر مربوط به ویژگی های مختلف

۶-۲ دقت و یادآوری

پس از آموزش و تست داده، بعد از اینکه تعیین شد هر نمونه به کدام کلاس تعلق دارد، می توان مقادیر دقت و یادآوری هر طبقه بنده کننده را محاسبه نمود. این مقادیر برای هر کلاس نشان می دهد که طبقه بنده کننده در مورد هر کلاس چگونه عمل کرده است. در واقع دقت و یادآوری بیشتر در بازیابی اطلاعات برای ارزیابی دقت الگوریتم های مختلف استفاده می شود اما این مقادیر برای روش های طبقه بنده نیز می تواند استفاده شود. اگر a تعداد نمونه که به کلاس صحیح خود طبقه بنده شده باشد، b تعداد نمونه متعلق به یک کلاس و c نیز تعداد نمونه اختصاص داده شده به یک کلاس باشد، آنگاه مقادیر دقت و ارزیابی به صورت زیر محاسبه می شوند:

$$\text{یادآوری} = \frac{a}{a+b}$$

$$\text{دقت} = \frac{a}{c}$$

به عنوان مثال اگر مجموعه داده ما حاوی ۱۰۰۰ نمونه باشد، آنگاه اگر یک روش ۸۰۰ نمونه به این کلاس اختصاص دهد که ۶۰۰ نمونه از آن واقعاً به این کلاس تعلق داشته باشد آنگاه دقت این روش ۷۵٪ و یادآوری آن ۶۰٪ خواهد بود.

F-Measure ۳-۶

F-Measure مقیاسی است که از تلفیق دو معیار دقت و یادآوری بدست می آید:

$$F = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

در این مقاله، برای انجام تجربیات مربوط به پیش بینی نتیجه تحصیلی (نتیجه آزمون نهایی) فرآگیران در این سیستم آموزش الکترونیک، از ۴ روش مختلف شبکه های عصبی، درخت تصمیم، رگرسیون منطقی و الگوریتم بیز استفاده کرده ایم که در ادامه فرایند مورد استفاده و نتایج بدست آمده را توضیح خواهیم داد.

۵- انتخاب ابزار داده کاوی

ما در ابتدا می بایست از بین ابزارهای داده کاوی موجود، محیط مناسبی را برای پیشبرد تحقیقمان انتخاب می کردیم. از میان این ابزارها، آنهایی که منبع باز (Open Source) بودند را در اولویت قرار دادیم. لذا در نهایت به ۳ ابزار داده کاوی Orange [10] Weka [11] و Yale [12] رسیدیم. Weka و Yale از الگوریتم های بیشتری نسبت به Orange پشتیبانی می کردند. ما در نهایت ابزار داده کاوی Weka را به دلیل قابلیت پشتیبانی آن از مجموعه داده های بزرگ، برگزیدیم.

۶- ارزیابی

انتخاب فاکتورهای موثر برای ارزیابی روش های یادگیری ماشین اعمال شده، یکی از مسایل مهم در این مقاله بوده است. در ادامه به معرفی فاکتورهایی که در ارزیابی روش های مختلف، مورد استفاده واقع شده اند، می پردازیم.

۰.۸۱۴	F-Measure
۱۲۹	تعداد نمونه صحیح اختصاص داده شده از کلاس قبول
۱۵۳	تعداد نمونه صحیح اختصاص داده شده از کلاس مردود یا غایب
۸۰.۱۱۳	دقت نهایی روش (درصد)

جدول ۳ نتایج بدست آمده از اعمال تکنیک بیز ساده

MLP که نوعی شبکه عصبی است نیز برای پیش بینی نتیجه تحصیلی استفاده شده است. تعداد نمونه های صحیح اختصاص داده شده برابر با ۲۷۹ عدد است که منتج به دقت ۷۹.۲۶۱ درصد شده است. نتایج مربوط به این روش در جدول ۴ نشان داده شده است

میزان	مورد
۰.۷۶۶	دقت کلاس قبول
۰.۸۱۳	دقت کلاس مردود یا غایب
۰.۷۶۱	یادآوری کلاس قبول
۰.۸۱۷	یادآوری کلاس مردود یا غایب
۰.۷۶۴	F-Measure کلاس قبول
۰.۸۱۵	F-Measure کلاس مردود یا غایب
۱۱۸	تعداد نمونه صحیح اختصاص داده شده از کلاس قبول
۱۶۱	تعداد نمونه صحیح اختصاص داده شده از کلاس مردود یا غایب
۷۹.۲۶۱	دقت نهایی روش (درصد)

جدول ۴ نتایج بدست آمده از اعمال تکنیک شبکه عصبی

روش آخر طبقه بندی مورد نظر، روش رگرسیون منطقی می باشد. این روش در فرایند انجام شده موفق به طبقه بندی صحیح ۲۸۵ مورد یعنی ۸۰.۹۶۶ درصد از مجموعه داده شده است که بهترین نتیجه بدست آمده است. جزئیات مربوط به نتایج بدست آمده از این روش در جدول ۵ نشان داده شده است.

میزان	مورد
۰.۷۷۵	دقت کلاس قبول
۰.۸۳۹	دقت کلاس مردود یا غایب
۰.۸	یادآوری کلاس قبول
۰.۸۱۷	یادآوری کلاس مردود یا غایب
۰.۷۸۷	F-Measure کلاس قبول
۰.۸۲۸	F-Measure کلاس مردود یا غایب
۱۲۴	تعداد نمونه صحیح اختصاص داده شده از کلاس قبول
۱۶۱	تعداد نمونه صحیح اختصاص داده شده از کلاس مردود یا غایب
۸۰.۹۶۶	دقت نهایی روش (درصد)

جدول ۵ نتایج بدست آمده از اعمال تکنیک رگرسیون منطقی

F-Measure برای مثال فوق ۰.۶۶ است که برای ارزیابی دقته یک طبقه بندی می تواند استفاده شود.

۷ - نتایج بدست آمده

در این مقاله تعدادی از روش های مختلف یادگیری ماشین برای پیش بینی نتیجه تحصیلی فرآگیران بر اساس نتیجه نهایی امتحانات آنها استفاده شده است. روش های مورد نظر شامل درخت تصمیم، شبکه های عصبی، الگوریتم بیز و رگرسیون منطقی می باشند. برای درخت تصمیم، روش C4.5 را انتخاب نموده ایم و J48 که پیاده سازی از این الگوریتم است نیز استفاده شده است. همانگونه که بیان شد برای ارزیابی روش های مختلف از شیوه ارزیابی متقطع-ده دسته استفاده می شود. به این ترتیب که عمل آموزش و طبقه بندی، ۱۰ دفعه انجام می شود و در هر دفعه ۵ درصد مجموعه داده به عنوان تست کنار گذاشته شده و به این منظور استفاده می شود. از ۳۵۲ نمونه موجود در مجموعه داده تعداد ۲۸۲ عدد یعنی ۸۰.۱۱۳ درصد، کلاس مربوطه به شکل صحیح طبقه بندی شده اند. سایر اطلاعات بدست آمده از پیاده سازی این روش در جدول ۲ آمده است.

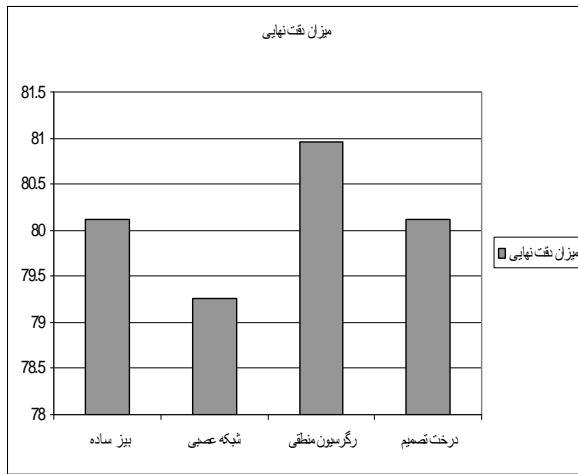
میزان	مورد
۰.۷۵۴	دقت کلاس قبول
۰.۸۴۳	دقت کلاس مردود یا غایب
۰.۸۱۳	یادآوری کلاس قبول
۰.۷۹۲	یادآوری کلاس مردود یا غایب
۰.۷۸۳	F-Measure کلاس قبول
۰.۸۱۷	F-Measure کلاس مردود یا غایب
۱۲۶	تعداد نمونه صحیح اختصاص داده شده از کلاس قبول
۱۵۶	تعداد نمونه صحیح اختصاص داده شده از کلاس مردود یا غایب
۸۰.۱۱۳	دقت نهایی روش (درصد)

جدول ۲ نتایج بدست آمده از اعمال تکنیک C4.5

روش بیز ساده روش دیگری است که برای طبقه بندی فرآگیران و پیش بینی نتیجه تحصیلی کاربران استفاده شده است. این روش نیز ۲۸۲ عدد از نمونه ها را به شکل صحیح طبقه بندی کرده است و دقته نهایی این روش همانند روش C4.5 ۸۰.۱۱۳ درصد می باشد. نتایج این روش در جدول ۳ تشریح شده است.

میزان	مورد
۰.۷۴۶	دقت کلاس قبول
۰.۸۵۵	دقت کلاس مردود یا غایب
۰.۸۳۲	یادآوری کلاس قبول
۰.۷۷۷	یادآوری کلاس مردود یا غایب
۰.۷۸۷	F-Measure کلاس قبول

میان روش های استفاده شده، داشته است. روش های درخت تصمیم و بیز ساده نیز دقیق مشابه ۸۰.۱۱۳ درصد را داشته اند و در رده بعدی نیز شبکه عصبی با ۷۹.۲۶۱ درصد کمترین دقیق را داشته است.



شکل ۳ نمودار مقایسه دقت نهایی ۴ روش

۷- نتیجه گیری

در این مقاله، به نحوه استفاده از روش های داده کاوی در سیستم های آموزش الکترونیک پرداخته شده است. هدف از استفاده روش های داده کاوی، غنی کردن، توسعه و شخصی کردن این محیط هاست. سیستم آموزشی هوشمند و شخصی شده تلاش می کند بوسیله تهیه طرحی از ویژگی های شخصی، علایق و معلومات فرآگیران، شکلی از آموزش اختصاصی را به فرآگیران ارائه دهد. داده کاوی یا دانش استخراج خودکار الگوهای مفید از مجموعه داده های بزرگ، می تواند برای دستیابی به مدل فرایند یادگیری و مدل یابی فرآگیران به کار رود و با جستجو و یافتن اطلاعات آموزشی سودمند بر مبنای استناد آموزشی، در ارزیابی و بهبود سیستم آموزشی نیز استفاده شود.

در این مقاله، به پیش بینی رفتار فرآگیران آموزش الکترونیک پرداخته شده است. اهمیت تخمین دقیق نتیجه تحصیلی آتی فرآگیران برای تأمین یک سیستم آموزشی متناسب با توانایی فرآگیری و در نتیجه یاری وی در این فرایند برای موفقیت بیشتر، ضروری است. اگر مدرس و یا سیستم آموزشی بتواند نتیجه تحصیلی فرآگیر را در هر مقطع زمانی از دوره تحصیلی پیش بینی کند، قادر خواهد بود که برنامه آموزشی متناسب با فرآگیر را به وی ارایه نماید و از افت تحصیلی فرآگیران ضعیف جلوگیری نماید و یا شکوفایی بیشتر فرآگیران مستعد را فراهم سازد و این همان مفهوم شخصی سازی در آموزش الکترونیک است. برای این امر از ۴ روش مختلف یادگیری ماشین شامل رگرسیون منطقی، درخت تصمیم،

نتایج بدست آمده از اعمال ۴ روش مختلف یادگیری ماشین در جدول ۶ نشان داده شده است.

طبقه بندی کننده	میزان دقت (درصد)
درخت تصمیم	۸۰.۱۱۳
بیز ساده	۸۰.۱۱۳
شبکه عصبی	۷۹.۲۶۱
رگرسیون منطقی	۸۰.۹۶۶

جدول ۶ مقایسه روش های مختلف

در جداول ۷ الی ۱۰ ملاحظه می شود که روش رگرسیون منطقی، قادر به پیش بینی کلاس صحیح تعداد نمونه های بیشتری نسبت به سایر روش ها است.

جدول ۷ ماتریس آشفتگی برای روش بیز ساده
==== Confusion Matrix for Naïve Bayesian ===

a b <-- classified as
153 44 | a = مردود_غایب
26 129 | b = قبول

جدول ۸ ماتریس آشفتگی برای روش شبکه عصبی
==== Confusion Matrix for Multi Layer Perceptron ===

a b <-- classified as
161 36 | a = مردود_غایب
37 118 | b = قبول

جدول ۹ ماتریس آشفتگی برای روش رگرسیون منطقی
==== Confusion Matrix for Logistic Regression ===

a b <-- classified as
161 36 | a = مردود_غایب
31 124 | b = قبول

جدول ۱۰ ماتریس آشفتگی برای روش درخت تصمیم
==== Confusion Matrix for Decision Tree ===

a b <-- classified as
156 41 | a = مردود_غایب
29 126 | b = قبول

همانطور که در نمودار شکل ۳ ملاحظه می شود، در مقایسه این ۴ روش، رگرسیون منطقی به عنوان یک روش کاملاً کلاسیک برخلاف انتظار، با ۸۰.۹۶۶ درصد، بیشترین دقیق را در

مراجع

- [1] Ha, S., Bae, S., & Park, S, *Web mining for distance education*, IEEE international conference on management of innovation and technology, 2000.
- [2] Brusilovsky, P., & Peylo, C, *Adaptive and intelligent web-based educational systems*, International Journal of Artificial Intelligence in Education, 13, 156–169, 2003.
- [3] Han, J.;Kamber, M.; and Tung, A., *Spatial Clustering Methods in Data Mining: A Survey In Miller, H. and Han, J., eds.,Geographic Data Mining and Knowledge Discovery*, Taylor and Francis., 2001.
- [4] Kontkanen, P.,Myllymaki, P., and Tirri, H., *Predictive data mining with finite mixtures*, Proceeding 2nd international Conference “Knowledge Discovery and Data Mining (KDD,96), 1996.
- [5] Weiss, S. M. and Kulikowski C. A., *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, Morgan Koufman, 1991.
- [6] Mitchell, Tom M., *Machine Learning*, McGraw-Hill, 1997.
- [7] Kantardzic M., *Data Mining: Concepts, Models, Methods, and Algorithms*, Wiley publishing, 2003.
- [8] Muehlenbrock, M, *Automatic action analysis in an interactive learning environment*, 2005.
- [9] Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J. *Web usage mining project for improving web-based learning sites*. Web mining workshop, Cataluna, 2005.
- [10] Weka, University of Waiko, New Zealand, <http://www.cs.waikato.ac.nz/ml/weka/>
- [11] Orange, University of Ljubljana, Slovenia, <http://www.ailab.si/orange>
- [12] Yale, Rapid-I, <http://rapid-i.com/>

بیز ساده و شبکه عصبی استفاده شده است که در مقایسه این ۴ روش، رگرسیون منطقی با ۸۰.۹۶۶ درصد، بیشترین دقت در پیش بینی را در میان روش های استفاده شده داشته است. همچنین روش های درخت تصمیم و بیز ساده نیز دقت مشابه ۸۰.۱۱۳ درصد را داشته اند و در ردی بعدی شبکه عصبی با ۷۹.۲۶۱ درصد کمترین دقت را داشته است. لذا اگر که در این سیستم آموزش الکترونیک، از تکنیک رگرسیون منطقی به منظور پیش بینی نتیجه تحصیلی فراغیران استفاده گردد، به شخصی سازی مطلوبتری منتج خواهد گردید.

اما بدیهی است که باید در راستای افزایش دقت نتایج بدست آمده از اعمال روش های داده کاوی در پیش بینی، به عنوان کارهای آتی این فعالیت تلاش نمود. هر چند که دقت بدست آمده ۸۰.۹۶۶ درصد دقت مناسبی برای یک روش داده کاوی است، اما در صورتی که بخواهیم از یک مازول پیش بینی کننده نتیجه تحصیلی در یک سیستم واقعی آموزش الکترونیک بهره جوییم، قطعاً این دقت کافی و کارآمد نخواهد بود و باید در افزایش این دقت تلاش نمود.

همچنین یک گام آرمانی برای مرحله بعدی این تحقیق، استفاده نتایج بدست آمده در یک سیستم آموزش الکترونیک واقعی و تلاش برای استفاده از نتایج مازول پیش بینی نتایج تحصیلی کاربران در راستای شخصی سازی پروفایل آموزشی فراغیران می باشد.

