

بررسی دقیق روش‌های داده کاوی در پیش‌بینی رفتار فرایگیران آموزش الکترونیک

همون پاینده فر^۱، حسن سید رضی^۲، مسعود رهگذر^۳

^۱دانشجوی کارشناسی ارشد، دانشکده کامپیوتر، دانشگاه پیام نور
تهران، ایران

hooman.payandehfar@gmail.com

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

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

چکیده

ظهور و رشد بی نظیر اینترنت و وب در دهه های اخیر باعث شده است که کاربردهای مختلف این تکنولوژی‌ها در زندگی روزمره بشری تأثیرات 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 بپردازند. داده کاوی از شاخه هایی است که به وسیله شخصی سازی، می‌تواند در راستای افزایش کیفی سیستم‌های آموزش الکترونیک تاثیر 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 ۸۰.۹ درصد بیشترین دقیق پیش‌گویی را در میان روش‌های استفاده شده داشته است. روش‌های درخت تصمیم و بیز ساده نیز دقیق تر برای مازول پیش‌بینی نتیجه تحصیلی فرایگیران، کمک شایانی در افزایش کیفیت سیستم آموزش الکترونیک و شخصی سازی بهتر آن نماید.

کلمات کلیدی

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

رشته‌های تحت وب مطالب بدون توجه به تفاوت یادگیری

دانشجویان به صورت یکنواخت آموزش داده می‌شود سیستم آموزشی هوشمند و سازگار در محیط هایی که نیاز به یادگیری شخصی غنی تر احساس می‌شود، راهگشا خواهد بود. سیستم آموزشی هوشمند و شخصی شده تلاش می‌کند بوسیله تهیه طرحی از ویژگی‌های شخصی، علائق و معلومات فرایگیران، شکلی از آموزش اختصاصی را به فرایگیران ارائه دهد. داده کاوی یا دانش استخراج خودکار الگوهای مفید از مجموعه داده‌های بزرگ، می-

۱- مقدمه

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

چندین زیرمجموعه افزای می‌کند که توسط برگهای این درخت شناخته می‌شود. یک برگ از این درخت مقدار محتمل از یک مشخصه کامل رکوردهای تعریف شده را مشخص می‌کند. درخت های مذکور می‌توانند عنوان مجموعه هایی از قوانین "if-then-else" نمایش داده شوند.

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

۳-۲ روش شبکه عصبی [5]

عصب، سلول بیولوژیکی ویژه‌ای با توانایی پردازش اطلاعات می‌باشد شبکه‌ی عصبی، یک گراف لایه‌ای است که خروجی یک گره، ورودی یک یا چند گره از لایه بعدی را تأمین می‌کند. ادراک چند لایه‌ای (MLP) یک پیش خود پایه‌ای شبکه عصبی مصنوعی است که از انتشار الگوریتم برای آموزش استفاده می‌کند. یعنی اینکه در طول آموزش، اطلاعات در سراسر قسمتهای پیشین شبکه منتشر می‌شود و برای بروز کردن وزنهای اتصال مورد استفاده قرار می‌گیرد.

خروجی‌های اعصاب یک لایه، ورودی‌های لایه بعدی را فراهم می‌کنند. پس وقتی یک n تایی (x_1, \dots, x_n) عنوان لایه ورودی یک شبکه به کار برده می‌شود خروجی m تایی (c_1, \dots, c_m) حاصل می‌شود، که مقدار c_i , $i = 1, \dots, m$ است اگر x_j تایی ورودی به کلاس C_j 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 می‌دهد که شیوه رگرسیون خطی می‌تواند برای مدل هایی که متغیر پاسخ آنها گستینه می‌باشد، استفاده شود. یک مدل عمومی از مدل خطی تعمیم یافته، رگرسیون منطقی می‌باشد. مدل دگرسیون منطقی احتمال وقایع اتفاق افتاده در یک تابع خطی از مجموع متغیرهای پیشگویی شده را نشان می‌دهد.

۳- داده کاوی در آموزش الکترونیک

آموزش از راه دور و یا یادگیری از راه دور متشکل از تکنیک‌ها و متدهایی است که دستیابی به برنامه‌های آموزشی را برای دانشجویانی که به شکل زمانی و یا مکانی از مدرس بدور هستند فراهم می‌سازد. شاخه‌های فرعی متعددی برای آموزش از راه دور

تواند برای دستیابی به مدل فرایند یادگیری و مدل یابی فرآگیران 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 روش‌های داده کاوی غیر گرافیکی به منظور پیش‌بینی رفتارهای آتی کاربران سیستم‌های آموزش الکترونیک بوده است. لذا پس از شرح این روش‌های داده کاوی برای اهداف پیش‌بینی، به مقایسه دقت این روش‌ها در پیش‌بینی نتیجه تحصیلی فرآگیران آموزش الکترونیک پرداخته شده است. به دلیل آنکه دقت پیش‌بینی در یک سیستم آموزشی هوشمند (ITS)، می‌تواند اهمیت زیادی در شخصی‌سازی و ارائه سیستم توصیه گر مطلوبتر شایانی در رویکرد شخصی‌سازی و ارائه سیستم توصیه گر مطلوبتر به عنوان کارهای آتی این تحقیق داشته باشد.

۲- داده کاوی و پیش‌بینی

داده کاوی، کاربرد تکنیک‌های یادگیری ماشین بر روی پایگاه داده برای کشف دانش و الگوهای نهفته در داده‌های پایگاه داده می‌باشد.

در ادامه، روش‌های غیر گرافیکی داده کاوی به منظور پیش‌بینی رفتارهای آتی فرآگیران شرح داده شده است.

۱-۳ استنتاج بیزی [4]

یکی از مهمترین روش‌های اماری در داده کاوی روش استنتاج بیزی است. طبقه بندی بیزی ساده، روش آسان و موثری را برای دانش طبقه بندی فراهم می‌آورد.

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

۲-۲ روش مبتنی بر درخت تصمیم [6]

روش‌های مبتنی بر درخت تصمیم روش‌هایی عمومی و پ्रطریفارد برای استفاده در زمینه داده کاوی می‌باشند. رده بند درخت تصمیم از روش‌های سلسله مراتبی یا لایه‌ای برای طبقه بندی کردن استفاده می‌کند.

هر راس این درخت نمایشگر یک آزمایش یا تصمیم یکتا می‌باشد. بالهای خروجی یک راس، متناظر با همه احتمالهای حاصل از آزمایش روی راس می‌باشد. این پیشامدها مجموعه داده‌ها به

۴ - مجموعه داده

مجموعه داده مورد استفاده ما، 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 ۳۴، ۳۶ و ۳۱.۲ سال است. بیشینه و متوسط سن فرآگیران به ترتیب ۳۴، ۳۶ 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 و بیزگی های موجود در مجموعه داده، همراه با مقادیر آنها در جدول ۱ نشان داده شده است.

جدول ۱ ویژگی های استفاده شده در باره فرآیند

مقداری	ویژگی
مرد، زن	جنس
۳۴-۳۹	سن
مجرد، متأهل	خانواده
هیچ، یک یا دو، سه یا چهار	فرزند
خیر یا نیمه وقت، بله یا اضافه کار	کار
بله، خیر	سود کامپیوتري
بله، خیر	کار کامپیوتري
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

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

وجود دارد: آموزش مکاتبه ای مبتنی بر کاغذ، آموزش از طریق نواهای خبط شده، آموزش به کمک رایانه، آموزش چند رسانه ای، آموزش الکترونیکی و آموزش تحت وب.

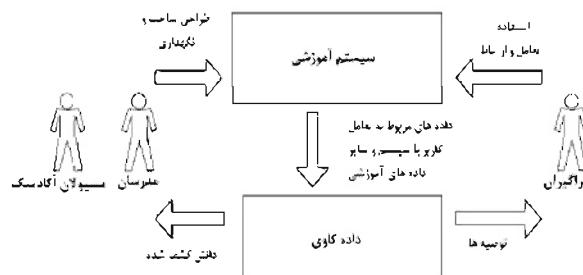
اخیراً مرسم ترین، آموزش تحت وب (آموزش الکترونیک) است که دانشجویان را قادر می سازد به راحتی از طریق اینترنت آموزش بگیرند. آموزش تحت وب، فرمی از آموزش راه دور است که اطلاعات از طریق اینترنت منتقل می شود.

سازمان هایی که پایگاه های آموزش از راه دور را 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 کاری 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.

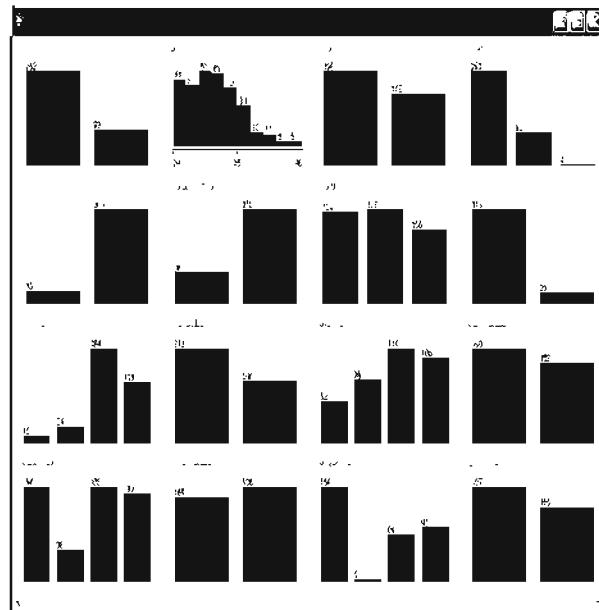
همانطور که در شکل ۱ می بینیم، آموزش دهنگان و مستولان آکادمیک، مسئول طراحی، برنامه ریزی و حفظ سیستم آموزشی هستند که فرآگیران از آن استفاده می کنند و با آن تعامل دارند با استفاده از تمام داده های موجود درباره دوره های آموزشی، فرآگیران و تعاملات، تکنیک های متفاوتی از کاوش اطلاعات به کار گرفته می شود تا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

دانش بدست آمده نه تنها می‌تواند بوسیله تعلیم گرجهت بالابردن کیفیت آموزشی سیستم و برنامه ریزی بهتر برای آینده، استفاده شود بلکه می‌تواند در غالب سیستم آموزشی هوشمند و سازگار برای پیمود فرآیند یادگیری فرآگیران، استفاده گردد [9].



شکل ۱ چرخه استفاده از داده کاوی در سیستم آموزش الکترونیک

در روش ارزیابی متقاطع برای تعیین میزان کیفیت یک طبقه بندی کننده، مقداری از داده آموزشی به عنوان مجموعه تست استفاده می‌شود. به عنوان مثال ۹۰٪ مجموعه داده به عنوان داده آموزشی استفاده می‌شود و ۱۰٪ بقیه برای تست روش استفاده می‌شود. تعداد نمونه‌هایی که به طور صحیح طبقه بندی می‌شوند به عنوان دقت آن روش لاحظ می‌گردد. این فرایند ۱۰ بار تکرار می‌شود و هر مرتبه با نگاه داشتن ۱۰٪ داده آموزشی اجرا می‌شود. هنگامی که این فرایند تکمیل می‌شود، میانگین مقادیر دقت مراحل به عنوان میزان دقت نهایی آن روش لاحظ می‌شود. اگر میزان مقدار داده تست نگاه داشته شده ۱۰ درصد کل داده آموزشی باشد به این شیوه، ارزیابی متقاطع ده دسته گفته می‌شود. می‌توان تعداد دسته‌ها 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 ویژگی‌های مختلف

۶-۲ دقت و یادآوری
پس از آموزش و تست داده، بعد از اینکه تعیین شد هر نمونه به کدام کلاس تعلق دارد، می‌توان مقادیر دقت و یادآوری هر طبقه بندی کننده را محاسبه نمود. این مقادیر برای هر کلاس نشان می‌دهد که طبقه بندی کننده در مورد هر کلاس چگونه عمل کرده است. در واقع دقت و یادآوری بیشتر در بازیابی اطلاعات برای ارزیابی دقت الگوریتم‌های مختلف استفاده می‌شود اما این مقادیر برای روش‌های طبقه بندی نیز می‌تواند استفاده شود.

اگر a تعداد نمونه که به کلاس صحیح خود طبقه بندی 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 می‌شوند:

$$\text{یادآوری} = \frac{a}{b}$$

$$\text{دقت} = \frac{a}{c}$$

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 می‌آید:

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

در این مقاله، برای انجام تجربیات مربوط به پیش‌بینی نتیجه تحصیلی (نتیجه آزمون نهایی) فرآگیران در این سیستم آموزش الکترونیک، از ۴ روش مختلف شبکه‌های 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 می‌بایست از بین ابزارهای داده کاوی موجود، محیط مناسبی را برای پیشبرد تحقیقمان انتخاب می‌کردیم. از میان این ابزارها، آنهای که منبع باز (Open Source) بودند را در اولویت قرار دادیم. لذا در نهایت به ۳ ابزار داده کاوی Weka [10]، Orange [11] و Yale [12] رسیدیم. Weka و Orange از الگوریتم‌های بیشتری نسبت به Weka پشتیبانی می‌کردند. ما در نهایت ابزار داده کاوی Weka را به دلیل قابلیت پشتیبانی آن از مجموعه داده‌های بزرگ، برگزیدیم.

۶- ارزیابی

انتخاب فاکتورهای موثر برای ارزیابی روش‌های یادگیری ماشین اعمال شده، یکی از مسائل مهم در این مقاله بوده است. در ادامه به معرفی فاکتورهایی که در ارزیابی روش‌های مختلف، مورد استفاده واقع شده‌اند، می‌پردازیم.

۶-۱ روش ارزیابی متقاطع

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
۰.۸۱۴	F-Measure کلاس مردود یا غایب
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)

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

جدول ۴ نتایج بدست آمده از اعمال تکنیک شبکه عصبی

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)

روش آخر طبقه‌بندی مورد نظر، روش رگرسیون منطقی می‌باشد. این روش در فرایند انجام شده موفق به طبقه‌بندی صحیح ۲۸۵ عدد یعنی ۸۰.۹۶۴ درصد از مجموعه داده شده است که بهترین نتیجه بدست آمده است. جزیيات مربوط به نتایج بدست آمده از این روش در جدول ۵ نشان داده شده است.

F-Measure برای مثال فوق ۶۵٪ است که برای ارزیابی دقت یک طبقه‌بندی می‌تواند استفاده شود.

۷- نتایج بدست آمده

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

جدول ۲ نتایج بدست آمده از اعمال تکنیک C4.5

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)

روش بیز ساده روش دیگری است که برای طبقه‌بندی فرآگیران و پیش‌بینی نتیجه تحصیلی کاربران استفاده شده است. این روش نیز ۲۸۲ عدد از نمونه‌ها را به شکل صحیح طبقه‌بندی کرده است و دقت نهایی این روش همانند روش C4.5 ۸۰.۱۱۳ درصد می‌باشد. نتایج این روش در جدول ۳ تشریح شده است.

جدول ۳ نتایج بدست آمده از اعمال تکنیک بیز ساده

جدول ۸ ماتریس آشتفتگی 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 آشتفتگی برای روش رگرسیون منطقی
Confusion Matrix for Logistic Regression

a b <- classified as

161 36 | a = مردود_غایب

31 124 | b = قبول

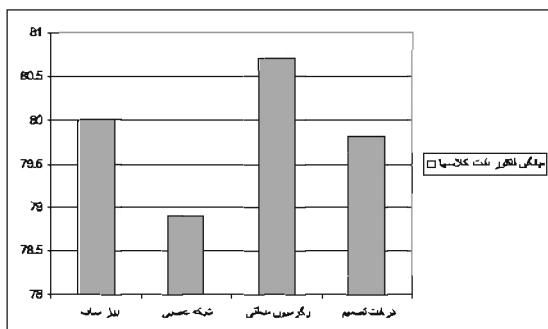
جدول ۱۰ ماتریس آشتفتگی برای روش درخت تصمیم
Confusion Matrix for Decision Tree

a b <- classified as

156 41 | a = مردود_غایب

29 126 | b = قبول

در نمودارهای شکل ۳ و ۴ و ۵، میانگین مقادیر فاکتورهای دقت و یادآوری و F-measure کلاسهای برای چهار روش، اورده شده است. در تمام این سه نمودار ملاحظه می‌گردد که روش رگرسیون منطقی، بالاترین مقادیر و روش شبکه عصبی، کمترین مقادیر را داشته است.



شکل ۳ میانگین فاکتور دقت کلاسهای در ۴ روش

جدول ۵ نتایج بدست آمده از اعمال تکنیک رگرسیون

منطقی

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

نتایج بدست آمده از اعمال ۴ روش مختلف یادگیری ماشین در

جدول ۶ نشان داده شده است.

جدول ۶ مقایسه روش های مختلف

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

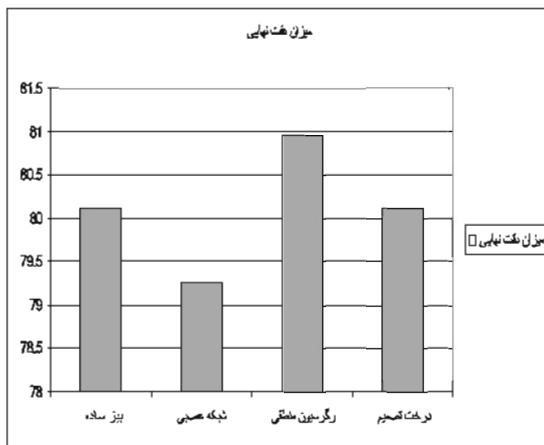
در جداول ۷ الی ۱۰ ملاحظه می‌شود که روش رگرسیون منطقی، قادر به پیش‌بینی کلاس صحیح تعداد نمونه‌های بیشتری نسبت به سایر روش‌ها است.

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

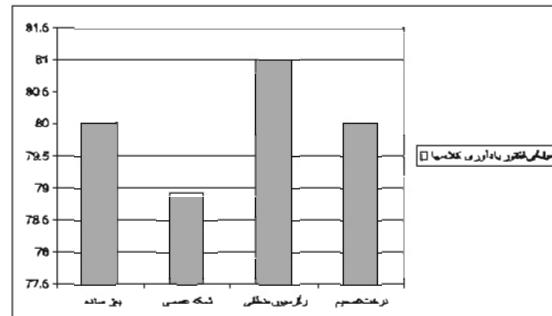
a b <- classified as

153 44 | a = مردود_غایب

26 129 | b = قبول



شکل ۶ نمودار مقایسه دقت نهایی ۴ روش

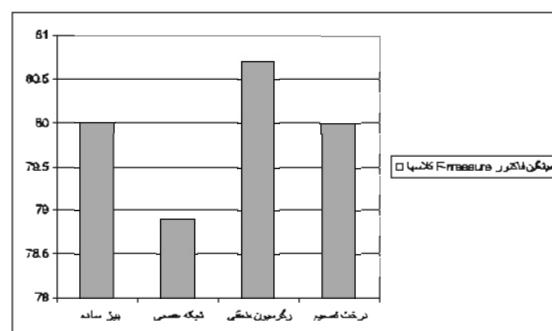


شکل ۴ میانگین فاکتور یادآوری کلاسها در ۴ روش

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 داشن 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 تحلیلی آنی فرآگیران برای تامین یک سیستم آموزشی مناسب با توانایی فرآگیری و در نتیجه یاری وی در این فرایند برای موقفيت بیشتر، ضروری است. اگر مدرس و یا سیستم آموزشی بتواند نتیجه تحلیلی فرآگیر را در هر مقطع زمانی از دوره تحصیلی پیش بینی کند، قادر خواهد بود که برنامه آموزشی مناسب 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 مашین شامل رگرسیون منطقی، درخت تصمیم، بیز ساده و شبکه عصبی استفاده شده است که در مقایسه این ۴ روش، رگرسیون منطقی با ۸۰.۹۶٪ 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 ۱۱۳٪ درصد را داشته اند



شکل ۵ میانگین فاکتور F-measure کلاسها در ۴ روش

در نمودار شکل ۶ مقایسه دقت نهایی ۴ روش آورده شده است. همانطور که در نمودار شکل ۶ ملاحظه می شود، در مقایسه این ۴ روش، رگرسیون منطقی به عنوان یک روش کاملاً کلاسیک برخلاف انتظار، با ۸۰.۹۶٪ درصد، بیشترین دقت در پیش بینی را در میان روش های استفاده شده، داشته است. روش های درخت تصمیم و بیز ساده نیز دقت مشابه ۱۱۳٪ درصد را داشته اند و در ده بعدی نیز شبکه عصبی با ۷۹.۲۶٪ درصد کمترین دقت را داشته است.

- [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.waiko.ac.nz/ml/weka/>
- [11] Orange, University of Ljubljana, Slovenia, <http://www.ailab.si/orange>
- [12] Yale, Rapid-I, <http://rapid-i.com/>

و در رده بعدی نیز 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 پیش‌بینی نتیجه تحصیلی فرآینران استفاده گردد، به شخصی سازی مطلوبتری منتج خواهد گردید. اما بدینهی است که باید در راستای افزایش دقت نتایج بدست آمده از اعمال روش‌های داده‌کاوی در پیش‌بینی، به عنوان کارهای آتی این فعالیت تلاش نمود. هر چند که دقت بدست آمده ۸۰.۹۵٪ درصد دقت مناسبی برای یک روش داده‌کاوی است، اما در صورتی که بخواهیم از یک مازول پیش‌بینی 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 پیش‌بینی نتایج تحصیلی کاربران در راستای شخصی سازی پروفایل آموزشی فرآینران می‌باشد.

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 2 nd 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.