

شهرام باشکیان<sup>۱</sup>

Bashokian@Comp.iust.ac.ir<sup>۱</sup>

چکیده - داده‌کاوی آموزشی یکی از رویکردهای نوین در علم داده‌کاوی می‌باشد که به بررسی و استخراج اطلاعات از سیستم‌های آموزشی و بکارگیری نتایج جهت بهبود سیستم آموزش، اشاره دارد. در یک سیستم آموزشی عوامل بسیاری در آموزش و فرایند یادگیری دانشآموزان با دانشجویان تأثیر گذار است. در این تحقیق تأثیر عوامل تربیتی نظیر میزان ارتباط با دوستان در خارج از مدرسه، ارتباط با والدین، شرایط محیط زندگی و... بر عملکرد تحصیلی دانشآموزان مورد بررسی قرار گرفته است. استخراج ارتباط میان عوامل مختلف به کمک داده‌کاوی می‌تواند جهت ارائه راهکارهایی به منظور بهبود عملکرد دانشآموزان موثر باشد. کلید واژه- داده‌کاوی آموزشی، مدل رفتاری، خوشبندی، دسته‌بندی.

آموزشی می‌توان اطلاعات و دانش مفیدی را از آنها استخراج و در جهت بهبود فرآیند آموزش، افزایش بازدهی و کارایی دانشآموزان و دانشجویان، به کار گرفت. تحقیقات انجام گرفته پیشین در این زمینه، هریک حوزه متفاوتی از آموزش را تحت پوشش قرار داده است. در پژوهشی رومرو و همکارانش بررسی جامعی از تحقیقات این حوزه از سال ۱۹۹۵ تا ۲۰۰۵ را ارائه کرده‌اند [۱]. کاسترو تحلیل دقیق‌تری از مشکلات آموزش الکترونیکی، که به وسیله روش‌های داده‌کاوی قابل پاسخگویی هستند به عمل آورده است [۲]. بیکر و همکاران استفاده نادرست دانشآموزان از محیط آموزشی و رفتار غیر مسئولانه آنها را با استفاده از روش‌های رده‌بندی تشخیص داده‌اند [۳]. همچنین کتسینتس کارایی دانشآموزان را در محیط آموزشی در کنار ارتباط ویژگی‌های تأثیرگذار در بهبود آن را مورد بررسی قرار داده است [۴]. یودلسن دانشآموزان را بر حسب الگوی استفاده به دو گروه متمایل به خطاب و متمایل به درستی تقسیم‌بندی کرده و با استفاده از نتایج این دسته‌بندی، علل معلول رویداد اشتباہ در دانشآموزان مورد بررسی قرار گرفته است [۵]. هم‌لینن قوانین انجمنی را برای برطرف کردن مشکلات دانشآموزان در محیط آموزشی و ارائه مشاوره با آنها مورد استفاده قرار داده است [۶]. زین نیز از این قوانین برای هدایت فعالیت‌های دانشآموزان و پیشنهاد محتوای آموزشی استفاده کرده است [۷]. همچنین لو نیز قوانین استخراج شده برای بهینه‌سازی محیط آموزشی مجازی بر حسب مواردی که از منظر دانشآموز جالب است مورد بررسی قرار داده است [۸]. در این پژوهش مراحل مختلف فرآیند داده‌کاوی از جمله جمع‌آوری داده‌ها، آماده‌سازی و پیش‌پردازش داده‌ها را بررسی

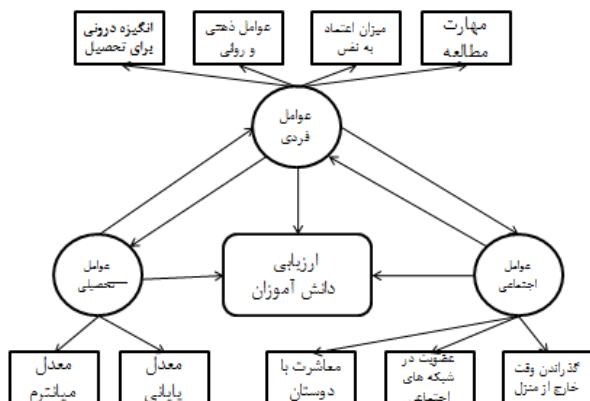
## ۱- مقدمه

سنجهش و ارزیابی جز لاینفک تعلیم و تربیت بشمار می‌رود که بدون استمرار دقیق آن، رسیدن به اهداف مورد نظر بصورت مطلوب، ناممکن خواهد بود. آموزش و پرورش فرآیندی است متشکل و به هم پیوسته که تمامی اجزای آن در ارتباط با یکدیگر و برای رسیدن به هدفی معین فعالیت می‌کنند. به طور کلی هدف آموزش و پرورش ایجاد تغییرات مطلوب در رفتار، نگرش‌ها، علایق، توانایی‌ها، مهارت‌ها، اعتقادات و به طور کلی همه ویژگی‌های فرد است. دستیابی به وضع مطلوب در هر نظام آموزشی، اعم از خرد یا کلان، مستلزم آن است که تحقیق و ارزیابی آموزشی در باره زمینه‌های مربوط انجام پذیرد و برنامه‌ریزی توسعه آموزشی به عمل آید و مدیریت آموزشی اعمال شود. از طرفی حجم بالای اطلاعات گاه در ظاهر متناقض در حوزه آموزشی به خصوص در مدارس منجر به تصمیم‌گیری‌های نادرست می‌گردد. به عنوان مثال بسیاری از مشکلات رفتاری و آموزشی دانشآموزان ریشه در عواملی دارد که عمدتاً از دید کادر آموزشی پنهان می‌ماند و در برخورد و ارائه راهکار مناسب برای معضل مورد بررسی، منجر به تصمیم‌های نادرست می‌گردد.

در سال‌های اخیر به منظور استفاده بهینه از ابیوه اطلاعات موجود در سازمان‌های آموزشی و استخراج دانش از این اطلاعات، راهبرد داده‌کاوی مورد استفاده قرار گرفته است. این رویکرد جدید که در اصطلاح داده‌کاوی آموزشی نامیده می‌شود به امر توسعه روش‌های کشف دانش از داده‌های محیط‌های آموزشی می‌پردازد. با بکارگیری تکنیک‌های داده‌کاوی برروی داده‌های

است. این تحقیق بر روی ۳۴۰ دانشآموز در مقطع متوسطه انجام گرفته است. در واقع روش اصلی جمع‌آوری داده‌ها در این تحقیق روش میدانی می‌باشد.

همانطور که اشاره شد داده‌ها در دو دسته اصلی داده‌های آموزشی و داده‌های تربیتی قرار دارند. داده‌های آموزشی از بانک‌های اطلاعاتی مدارس جمع‌آوری و در قالب فایل اکسل ذخیره شده‌اند. اطلاعاتی از قبیل نام و نام خانوادگی، شناسه منحصر‌بفرد، سن، مقطع تحصیلی و سوابق تحصیلی دانشآموزان شامل نمرات و نتایج امتحانات آنها از جمله اصلی‌ترین اطلاعات بdst آمده از این پایگاه‌داده‌ها می‌باشد. اطلاعات رفتاری و تربیتی دانشآموزان شامل اطلاعات بdst آمده از مطالعات رفتاری و جمع‌آوری نتایج پرسش‌نامه‌های توزیع شده در میان دانشآموزان است. این پرسش‌نامه‌ها سوالاتی از قبیل تاثیر عوامل اجتماعی، فردی و خانوادگی دانشآموزان را شامل می‌شوند که جمع‌آوری پاسخ‌های دانشآموزان پایگاه‌داده رفتاری-تربیتی دانشآموزان را شکل می‌دهد. همچنین در کنار دو پایگاه‌داده فوق اطلاعاتی نظری شرایط محیطی نیز گردآوری می‌گردد. در واقع به منظور ارائه روشی جامع و درست جهت کشف ارتباط میان تربیت و آموزش دانشآموزان، به بانک اطلاعاتی عظیمی شامل داده‌های آموزشی، فردی، خانوادگی و اجتماعی نیازمندیم. ارتباط داده‌های فوق در شکل ۲ بطور خلاصه نشان داده شده است. با بکارگیری درست روش‌های داده‌کاوی می‌توان تحلیل درستی از رفتارهای تربیتی و آموزشی دانشآموزان داشت. برای مثال می‌توان میزان تاثیر عضویت در شبکه‌های اجتماعی بر نتایج آموزشی دانشآموزان را مورد بررسی قرار داد. و یا ارتباط میان میزان گذراندن وقت با دوستان با سطح یادگیری و نمرات پایانی آنها را بدست آورد.

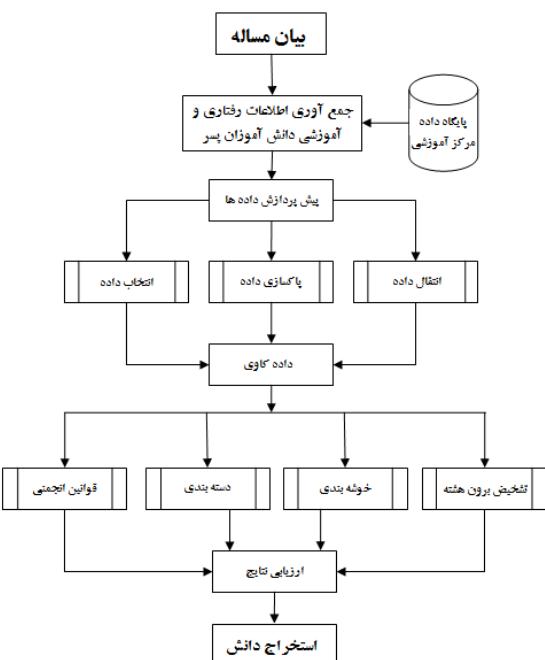


شکل ۲: مدل جمع‌آوری داده‌های تربیتی و آموزشی دانشآموزان

مجموعه داده‌های آموزشی جمع‌آوری شده انجام و الگوریتم‌های مختلف داده‌کاوی نظیر پیش‌بینی، رده‌بندی، خوشبندی و کشف قوانین انجمنی را بر روی داده‌ها اعمال کرده و در ادامه نشان خواهیم داد که چگونه از نتایج این الگوریتم‌ها در جهت شناخت رفتار و وضعیت دانشآموزان و تصمیم‌گیری برای تغییر برنامه‌های آموزشی در جهت افزایش بازدهی آنها و رفع مشکلات آموزشی استفاده خواهد شد.

## ۲- فرآیند کاوش

شکل ۱ مدل اولیه کشف دانش از پایگاه داده آموزشی را نشان می‌دهد. در این مدل مراحل مختلف داده‌کاوی شامل جمع‌آوری داده، پیش‌پردازش داده، اعمال تکنیک‌های داده‌کاوی و تحلیل نتایج به منظور کشف اطلاعات مفید از داده‌های آموزشی و رفتاری دانشآموزان انجام گرفته است.



شکل ۱: مدل اولیه کشف دانش از پایگاه‌داده آموزشی

جامعه آماری مورد بررسی در این تحقیق دانشآموزان پسر هترستان‌ها و دبیرستان‌های شهر تهران می‌باشند. داده‌های آموزشی نظری اطلاعات فردی و اطلاعات نمرات و وضعیت تحصیلی دانشآموزان از طریق بانک‌های اطلاعاتی مدارس جمع‌آوری گردیده است. همچنین داده‌های رفتاری دانشآموزان از طریق جمع‌آوری پرسش‌نامه میان دانشآموزان گردآوری شده

دانشآموز، شناسه منحصربفرد و سن دانشآموز از مجموعه خصیصه‌های مورد نظر حذف می‌گردد. در این مرحله ۲۵ مشخصه مورد نظر طبق جدول ۱ مورد بررسی قرار می‌گیرد.

### ۳- نتایج پژوهش

پس از جمع‌آوری و ذخیره نتایج در فایل اکسل، داده‌های به کمک ابزارهای داده‌کاوی Clementine و Knime مورد تجزیه و تحلیل قرار گرفته و تکنیک‌های کاوش قوانین انجمانی و دسته‌بندی به منظور کاوش اطلاعات و دانش مفید اجرا می‌گرددند. در واقع هدف از این بخش کشف دانش نهفته و ارتباط میان داده‌ها موجود می‌باشد. خروجی این بخش می‌تواند بصورت نمودارها و یا فایل‌های متون باشد که با بکارگیری آنها در یک سیستم ارزیابی و پیش‌نهاددهنده می‌توان ابزاری جهت ارزیابی تربیتی و آموزشی دانشآموزان راهه کرد.

در کاوش قوانین انجمانی بدبانی روابط جالب میان اقلام یک مجموعه از اقلام هستیم. یک مجموعه اقلام شامل بیش از یک قلم داده است، برای مثال مدل=متوسط یک نمونه از اقلام موجود در پایگاهداده است. هدف اصلی کاوش قوانین انجمانی پیدا کردن اقلامی است که در هر سطر از پایگاهداده باهم رخداده‌اند. برای مثال یک نمونه از قوانین انجمانی در پایگاهداده دانشآموزان بصورت زیر می‌تواند باشد:

۷۵٪ از دانشآموزانی که میانگین معدل متوسط دارند، عضو شبکه‌های اجتماعی اینترنوتی هستند.

با داشتن یک چین اطلاعاتی از رفتار دانشآموزان، می‌توان تصمیمات و راهبردهایی جهت بهبود سطح آموزشی دانشآموزان و یا امور تربیتی آنان اتخاذ کرد. از مهمترین الگوریتم‌های کاوش قوانین انجمانی می‌توان به الگوریتم Apriori اشاره کرد. خاصیت اولین بار توسط Agrawal در سال ۱۹۷۷ مورد استفاده قرار گرفت. بر اساس این خاصیت، یک مجموعه اقلام Kتایی فقط در صورتی متناسب است که همه زیر مجموعه‌های آن متناسب باشند. در کشف قوانین انجمانی دو معیار اندازه‌گیری به نام‌های پشتیبانی و درجه اطمینان جهت کاوش و کنترل نتایج و کاوش نتایج دلخواه بکار گرفته می‌شوند. مقدار پشتیبانی برای یک کاندیدا بصورت زیر محاسبه می‌شود:

$$\text{پشتیبانی} = \frac{\text{تعداد رکوردهای که شامل A و B هست}}{\text{تعداد کل رکوردها}}$$

برای مثال مقدار پشتیبانی ۶۴ درصد برای مجموعه اقلام مدل=خوب و برنامه‌ریزی=دارد) بدین معنا است که ۶۴ درصد

داده‌های آموزشی جمع‌آوری شده از پایگاهداده‌های مدارس شامل نام دانشآموز، شناسه منحصربفرد، پایه تحصیلی، نیمسال تحصیلی، سن و معدل در قالب یک فایل اکسل با همان ویژگی‌ها ذخیره می‌شوند. در این تحقیق، ویژگی پایه تحصیلی شامل سه مقدار اول، دوم و سوم دبیرستان می‌باشد. ویژگی میانگین نمرات میان‌ترم هم در فاز پیش‌پردازش داده‌ها به چهار دسته خیلی خوب، خوب، متوسط و ضعیف تقسیم می‌شود. همچنین جهت نمایش بهتر نتایج عملیات داده‌کاوی، معدل دانشآموزان به چهار دسته عالی، خوب، متوسط و ضعیف دسته بندی شده است. مقدار معدل عالی شامل دانشآموزان با معدلی بیشتر از ۱۷، مقدار معدل خوب شامل دانشآموزان با معدلی بین ۱۵ تا ۱۷، مقدار معدل متوسط شامل دانشآموزان با معدلی بین ۱۵ تا ۱۵ و مقدار معدل ضعیف برای دانشآموزانی با معدل کمتر از ۱۲ در نظر گرفته شده است.

داده‌های تربیتی برای هر دانشآموز از پاسخ‌های وی به پرسشنامه تهیه شده بدست می‌آید. این پرسشنامه شامل سوالاتی در زمینه فعالیت‌های اجتماعی، رفتاری و تحصیلی دانشآموزان است. هر سوال در این پرسشنامه به عنوان اطلاعاتی برای یکی از عوامل به ظاهر موثر در زمینه تحصیلی دانشآموزان در نظر گرفته می‌شود. اطلاعات جمع‌آوری شده از پرسشنامه‌ها به فایل اکسل اطلاعات آموزشی دانشآموزان اضافه گردیده و این

فایل جهت پیش‌پردازش داده‌ها مورد بررسی قرار می‌گیرد.

برای اکتشاف داده‌ها و ساخت مدل اولیه، اغلب کار کردن با مجموعه‌های کاوش یافته منطقی تر به نظر می‌رسد در حالیکه جدول نمونه نهایی ممکن است شامل هزاران یا میلیون‌ها نمونه باشد. آماده‌سازی اندیشمندانه داده‌ها می‌تواند به میزان قابل توجهی اطلاعاتی که بوسیله داده‌کاوی اکتشاف می‌شوند را بهبود بخشد. پیش‌پردازش و آماده‌سازی داده‌ها مهمترین و زمان برترین مرحله در پیروزهای داده‌کاوی است. تقریباً ۹۰ تا ۷۵ درصد از زمان انجام یک پیروزه داده‌کاوی صرف این مرحله می‌شود و ۹۰ درصد موفقیت پیروزهای داده‌کاوی به آن بستگی دارد.

فرآیندهایی که در پیش‌پردازش انجام می‌شوند عبارت است از: تجمیع، نمونه‌گیری، کاوش بعد، انتخاب زیرمجموعه ویژگی‌ها، ایجاد ویژگی و تبدیلات داده. براساس نوع کاربردی که عمل داده‌کاوی باید روی آن انجام شود، تکنیک‌های مختلفی برای هریک از این اعمال مورد استفاده قرار می‌گیرد. براساس داده‌های آموزشی و تربیتی جمع‌آوری شده در مرحله قبل، ۱۰ مشخصه از ۱۳ مشخصه موجود انتخاب و تمامی رکوردهای آنها در عملیات داده‌کاوی مورد استفاده قرار می‌گیرد. پارامترهای نام

دسته‌بندی یا رده‌بندی یکی از مدل‌های پیش‌بینی داده‌ها می‌باشد. در مسائل دسته‌بندی هدف شناسایی ویژگی‌هایی است که گروهی را که هر مورد به آن تعلق دارد نشان می‌دهد. از این الگو هم برای فهم داده‌های موجود و هم پیش‌بینی نحوه رفتار داده‌های جدید استفاده کرد. در داده‌کاوی، مدل‌های دسته‌بندی با بررسی داده‌های دسته‌بندی شده قبلی ایجاد می‌شوند و یک الگوی پیش‌بینی کننده را بصور استقرایی می‌یابند. از مهمترین روش‌های دسته‌بندی اطلاعات روش درخت‌های تصمیم می‌باشد. درخت‌های تصمیم روشی برای نمایش یک سری قوانین هستند که منتهی به یک رده یا مقدار می‌شوند. برای مثال می‌خواهیم دانش‌آموzan را براساس معدل میانترم به دو دسته دانش‌آموzan خوب و ضعیف تقسیم کنیم. مولفه‌های اصلی یک درخت تصمیم عبارتند از: نود تصمیم، شاخه‌ها و برگ‌ها. در درخت تصمیم گیری برای نمایش اطلاعات یا دانش به صورت صحیح، یک روش بکارگیری قوانین است که از قوانین if-then به صورت زیر بهره گرفته می‌شود:

if Condition then Conclusion

اگر موقعیت در مقدم صحیح باشد این شرط نتیجه بخش است. زمانی که این قوانین استخراج شوند، دانش ضمنی و با ارزش موجود در داده‌ها به دانشی قابل استفاده تبدیل می‌گردد. درخت‌های تصمیم این قوانین را تولید می‌کنند. هر قانون با حرکت از سمت ریشه به سمت برگ بدست می‌آید. گره‌های میانی با قوار دادن And بین آنها مقدم و برگ نتیجه یا تالی را تشکیل می‌دهند. درک این قوانین برای انسان راحت‌تر از درک ساختار درختی است اما از آنجایی که ممکن است تعداد این قوانین زیاد شود از تکنیک‌های هرس استفاده می‌شود و دقت افزایش می‌یابد. برای مثال قانون زیر در پایگاهداده دانش‌آموzan از درخت تصمیم حاصل می‌شود که اگر معدل میانترم خوب و عضویت در شبکه اجتماعی کم باشد آنگاه معدل پایان‌ترم خوب خواهد بود.

در این تحقیق الگوریتم‌های CART و C5.0 جهت تولید قوانین درخت‌های تصمیم بکار گرفته شده‌اند. دسته‌بندی داده‌ها با درختان تصمیم گیری یک فرآیند دو مرحله‌ای می‌باشد. در مرحله اول که به آن مرحله آموزش گفته می‌شود، مدلی براساس یک الگوریتم دسته‌بندی منطبق با داده‌های مربوط به مجموعه آموزشی ساخته می‌شود. مجموعه آموزشی بصورت تصادفی از پایگاهداده انتخاب می‌شود. در مرحله دوم، یادگیری از طریق یکتابع ( $y=f(x)$ ) انجام می‌شود که می‌تواند برچسب کلاس هر رکورد  $x$  از پایگاهداده را پیش‌بینی کند. مرحله یادگیری، خود

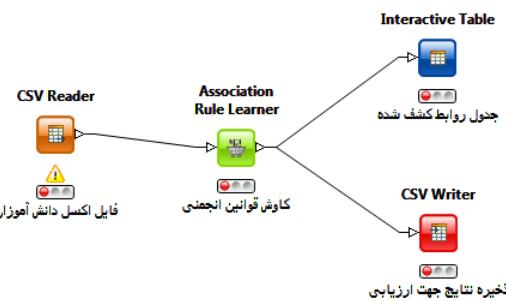
دانش‌آموzanی که دارای معدل خوب می‌باشند، جهت مطالعه دروس برنامه‌ریزی دارند.

درجه اطمینان معیاری برای اندازه‌گیری میزان دقیق قانون است که به صورت زیر محاسبه می‌شود:

$$\text{confidence} = P(B|A) = \frac{\text{تعداد دکورهای شامل A}}{\text{تعداد دکورهای شامل A} + \text{تعداد دکورهای غایب A}}$$

برای مثال درجه اطمینان  $100\%$  برای قانون (معدل میانترم خوب  $\leftarrow$  معدل کل خوب) بدین معنا است که تمامی دانش‌آموzanی که معدل میانترم خوبی داشته‌اند، معدل کل خوبی نیز کسب کرده‌اند. شکل ۳ مدل ساده‌ای از کشف قوانین اجمنی با استفاده از ابزار داده‌کاوی Knime را نشان می‌دهد.

شکل ۳: مدل کشف قوانین اجمنی به کمک ابزار Knime



طبق مدل ارائه شده در شکل ۳ قوانین اجمنی یافت شده از پایگاهداده دانش‌آموzan به کمک ابزار Knime استخراج گردیده است. با فرض حداقل پشتیبانی و درجه اطمینان  $100\%$ ، تعداد  $30700$  قانون معنادار از پایگاهداده دانش‌آموzan کشف گردیده است که از میان اطلاعات کشف شده تعداد اندکی از قوانین کاربردی و قابل استفاده جهت اتخاذ تصمیمات راهبردی در جهت بهبود کیفیت تحصیلی و تربیتی دانش‌آموzan است. برای مثال در  $4$  قانون کشف شده با مقادیر پشتیبانی و درجه اطمینان  $100\%$ ، قوانین حول  $3$  آیتم عضویت در شبکه‌های اجتماعی، داشتن انگیزه تحصیلی و اعتماد به نفس بالا خلاصه می‌شود. نتیجه حاصل از این کاوش بدین شرح است که  $60$  درصد دانش‌آموzanی که فاقد انگیزه تحصیلی هستند در شبکه‌های اجتماعی اینترنتی عضویت دارند. پس می‌توان تمهدیاتی که زمینه گسترش فرهنگ استفاده از شبکه‌های اجتماعی در میان این ۶۰ درصد از دانش‌آموzan اتخاذ کرد. جدول ۲ برخی از مهمترین قوانین اجمنی کشف شده به همراه تحلیل و تصمیمات پیشنهادی ممکن در جهت بهبود وضعیت تحصیلی و تربیتی دانش‌آموzan را نشان می‌دهد.

صورت از دست رفتم مقادیر یک پیش‌بینی، مقدار فاز قبلی پیش‌بینی را برای آن معیار در نظر می‌گیریم. در ادامه با توجه به تنظیمات اعمال شده فوق درخت‌های تصمیم تولید شده به ازای متغیرهای مختلف را نمایش و نتایج حاصل از آنها را مورد بررسی قرار خواهیم داد. می‌توان ادعا کرد که نتایج حاصل از این مرحله مفیدترین اطلاعات کشف شده از پایگاهداده دانش‌آموزان خواهد بود که در پیش‌بینی رفتار دانش‌آموزان آینده و اتخاذ تصمیمات آموزشی و تربیتی مورد استفاده قرار خواهد گرفت.

پایگاهداده اطلاعات دانش‌آموزان شامل ۱۰ خصوصیت اصلی است که مورد کاوش قرار گرفته است. در واقع هدف رسیدن به پیش‌بینی درستی از رفتار آموزشی و تربیتی دانش‌آموزان است بنابراین فاز تولید درخت‌های تصمیم را برای تمامی ویژگی‌های پایگاهداده اجرا گردیده است که از جمله مهمترین نتایج بدست آمده از اجرای این تکنیک به موارد زیر می‌توان اشاره کرد:

- در دسته‌ای از دانش‌آموزان اگر میزان گذراندن وقت در خارج از منزل متوسط و معدل میانترم هم متوسط باشد، به احتمال فراوان میزان گذراندن وقت با دوستان زیاد می‌باشد.

- در دسته‌ای از دانش‌آموزان اگر معدل میانترم خوب و میزان اعتماد به نفس زیاد باشد، به احتمال فراوان میزان مشکلات ذهنی کم است.

- در دسته‌ای از دانش‌آموزان اگر میزان پایه تحصیلی اول دبیرستان و میزان ارتباط با دوستان کم باشد، به احتمال فراوان معدل میانترم خوب می‌باشد.
- در دسته‌ای از دانش‌آموزان اگر میزان ارتباط با دوستان کم یا خیلی کم و معدل میانترم متوسط باشد، به احتمال فراوان معدل پایانی متوسط می‌باشد.

- در دسته‌ای از دانش‌آموزان اگر پایه اول دبیرستان و میزان ارتباط با دوستان کم و معدل میانترم متوسط باشد، به احتمال فراوان معدل پایانی متوسط می‌باشد.

- در دسته‌ای از دانش‌آموزان اگر پایه سوم دبیرستان و میزان ارتباط با دوستان کم و معدل میانترم متوسط باشد، به احتمال فراوان معدل پایانی خیلی خوب می‌باشد.

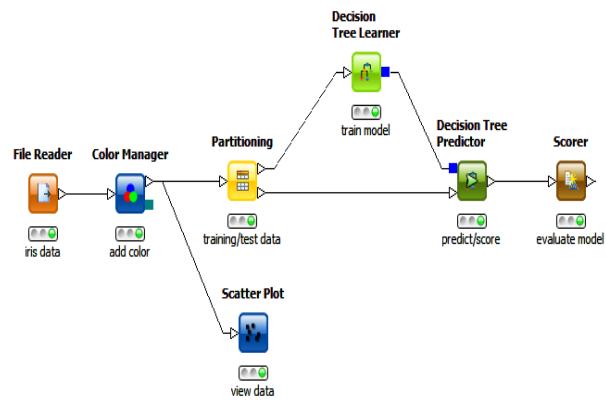
- در دسته‌ای از دانش‌آموزان اگر پایه اول دبیرستان و میزان اعتماد به نفس خیلی زیاد و معدل میانترم خوب باشد، به احتمال فراوان معدل پایانی خوب می‌باشد.

- در دسته‌ای از دانش‌آموزان اگر پایه اول دبیرستان و میزان اعتماد به نفس زیاد یا متوسط و معدل میانترم

طی دو گام اساسی رشد و هرس انجام می‌شود. در طول فرآیند آموزش، الگوریتم درخت تصمیم می‌بایست به صورت مکرر موثرترین روش، جهت تقسیم کردن مجموعه رکوردها به فرزندان را بباید. مرحله هرس، برای جلوگیری از برازش بیش از حد صورت می‌گیرد.

شکل ۴ مدلی از روش درخت‌های تصمیم‌گیری با ابزار KNIME را نشان می‌دهد. در این مدل ورودی فایل اکسل داده‌های دانش‌آموزان است.

شکل ۴: مدل درخت تصمیم‌گیری با ابزار Knime



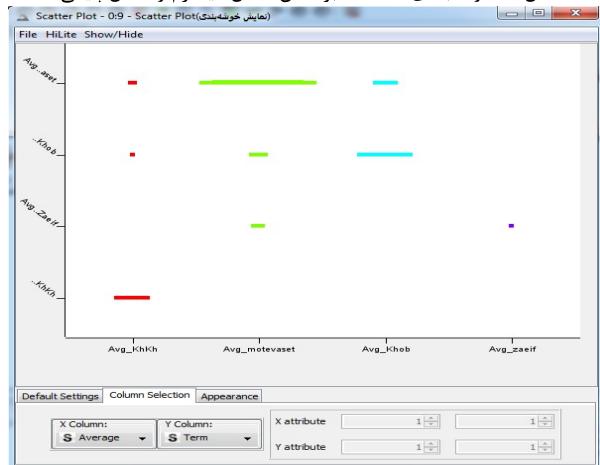
در ابزار Knime جهت ایجاد و استخراج اطلاعات به کمک درخت‌های تصمیم‌گیری از ماژول Partitioning جهت بخش بندي پایگاهداده استفاده شده است. در این ماژول امکان اعمال تنظیماتی جهت دسترسی به داده‌های فاز یادگیری وجود دارد. در این تحقیق ۲۰ درصد از داده‌ها به عنوان ورودی فاز یادگیری بصورت تصادفی انتخاب می‌شوند.

همچنین از ماژول Dscition Tree Learner جهت فاز یادگیری درخت تصمیم استفاده شده است. در تنظیمات ای ماژول روش هرس کردن درخت، روش هرس از قبل انتخاب گردیده است. روش هایی که در آنها، رشد درخت قبل از اینکه به نقطه ای برسد که کاملاً مثال‌های آموزشی را دسته بندی کند متوقف می‌شوند. نمونه هایی از معیارهای متوقف کردن رشد درخت عبارتند از: از تست chi-squared برای آزمایش معناداری تقسیم بیشتر مثال‌های آموزشی استفاده می‌کنیم، و یا معیار MDL را بکار گرفته و در صورتی که رشد بیشتر باعث کاهش این معیار شود، اجازه رشد درخت را می‌دهیم. که در این تحقیق معیار MDL بکار گرفته شده است.

همچنین استراتژی مقابله با مقادیر از دست رفته را استراتژی LastPrediction در نظر می‌گیریم. به این ترتیب که در

درخت‌های تصمیم و کاوش قوانین انجمنی بود که اطلاعاتی در مورد تاثیر ویژگی‌های مختلف در سطح رفتاری و آموزشی دانش‌آموزان در اختیار قرار داد. برای مثال در مورد پیش‌بینی معدل پایانی دانش‌آموزان، مواردی همچون پایه تحصیلی، میزان ارتباط با دوستان، معدل میانترم از اهمیت بالایی برخوردار بود. از نتایج بدست آمده در این پژوهش می‌توان به منظور اتخاذ تصمیمات و ارائه مشاوره‌های مفید به والدین و دانش‌آموزان و کشف موارد مهم در کیفیت آموزشی دانش‌آموزان بهره برد.

شکل ۵: خوشه‌بندی اطلاعات براساس معدل میانترم و معدل پایانی



## ۵- مراجع

- [1] Romero C. and Ventura S., "Educational Data Mining: A Survey from 1995 to 2005", *Expert Systems with Applications*, 2007, pp.135-146.
- [2] Castro F., Vellido A., Nebot A. and Mugica F., "Applying Data Mining Techniques to Learning Problems", in *Studies in Computational Intelligence*, Berlin and Heidelberg: Springer-Verlag, 2007, pp.183-221.
- [3] Baker, R., & Yacef, K., "The State of Educational Data mining in 2009: A Review Future Visions", *Journal of Educational Data Mining*, 2009.
- [4] Kotsiantis S.B., Paliouras 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.
- [5] 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.
- [6] 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.
- [7] Zaiane O., "Building a Recommender Agent for E-Learning Systems", in Proceedings of the International Conference on Computers in Education, 2009, pp.55.
- [8] 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.

خوب باشد، به احتمال فراوان معدل پایانی متوسط

می‌باشد.

همانطور که مشاهده می‌شود با بررسی نتایج بدست آمده از تحلیل درخت‌های تصمیم مختلف و با در نظر گرفتن ویژگی هدف مشخص، می‌توان اطلاعات بسیار مفیدی از پایگاهداده دانش‌آموزان و رفتارهای تربیتی و آموزشی آنها استخراج کرد. هریک از قوانین بدست آمده به نحوی گویای برخی از مشکلات و نقاط قوت دانش‌آموزان از نظر روابط اجتماعی، عاطفی و تحصیلی است که با ارائه راهکارهای مناسب می‌توان آنها را کنترل و سطح تحصیلی دانش‌آموزان را بهبود بخشید. همچنین از اطلاعات بدست آمده در این بخش می‌توان جهت پیش‌بینی رفتار دانش‌آموزان در ترم‌ها و دوره‌های بعدی استفاده کرد.

خوشه‌بندی یعنی یافتن گروه‌هایی از اشیا بطوریکه اشیا موجود در یک گروه تا حد امکان شبیه به هم باشند و با اشیا موجود در سایر گروه‌ها متفاوت باشند. با استفاده از این تکنیک می‌توان دانش‌آموزان را براساس ویژگی‌های مختلفی به عنوان مثال میزان مشارکت در کلاس، گروه‌بندی کرد و با هر گروه بصورت مناسبی رفتار کرد. در این تحقیق از الگوریتم K-Means برای خوشه‌بندی دانش‌آموزان استفاده شده است. در ابزار Knime به منظور انجام عملیات خوشه‌بندی نیاز به مازول‌های مدیریت رنگ، الگوریتم K-Means و نمایش خوشه‌ها می‌باشد. در این تحقیق به علت پراکندگی کم داده‌ها و جامعه آماری نسبتاً کم نمی‌توان از خوشه‌بندی به عنوان منبع مفیدی جهت کشف داشت جدید بهره برد. در واقع نتایج بدست آمده از خوشه‌بندی داده‌ها به نحوی در مراحل کاوش قوانین انجمنی و تولید درخت‌های تصمیم پوشش داده شده است. برای مثال شکل ۵ خوشه‌بندی اطلاعات را براساس دو ویژگی معدل میانترم و معدل پایانی نشان می‌دهد. همانگونه که مشاهده می‌شود می‌توان از این عملیات دانش‌آموزان را براساس معدل پایانی و میانترم در چهار خوشه دسته‌بندی کرد که بیشترین فراوانی زمانی است که معدل پایانی و میانترم متوسط باشد.

## ۴- نتیجه‌گیری

در این پژوهش مدل داده‌کاوی پیشنهادی شامل عملیات خوشه‌بندی، کاوش قوانین انجمنی و تولید درخت‌های تصمیم‌گیری بروی پایگاهداده اطلاعات دانش‌آموزان متشكل از اطلاعات ۳۴۰ دانش‌آموز اجرا و نتایج حاصل از آن مورد بررسی و تحلیل قرار گرفت. بیشترین اطلاعات بدست آمده از کاوش

جدول ۱ : فیلدهای اطلاعاتی پیش‌پردازش شده

مجموعه مقادیر		توضیح	نام فیلد	ردیف	مشخصات فیلدها
[1,2,3]	اول، دوم، سوم دبیرستان	مقطع تحصیلی	Degree	۱	
[0-20]	معدل دانش آموزان از ۰ تا ۲۰	میانگین نمرات میان‌ترم	Term	۲	
[0-20]	معدل دانش آموزان از ۰ تا ۲۰	معدل کل	Average	۳	
[1-5]	خیلی کم، کم، متوسط، زیاد، خیلی زیاد	معاشرت با دوستان	Friends	۴	
[1-5]	خیلی کم، کم، متوسط، زیاد، خیلی زیاد	گذراندن وقت خارج از منزل	Out_home	۵	
[1,2]	بله، خیر	عضویت در شبکه‌های اجتماعی	Social_Net	۶	
[1-4]	خیلی خوب، خوب، متوسط، بد	مهارت‌های مطالعه	Reading	۷	
[1-5]	خیلی کم، کم، متوسط، زیاد، خیلی زیاد	اعتماد به نفس	Self	۸	
[1-5]	خیلی کم، کم، متوسط، زیاد، خیلی زیاد	عوامل ذهنی و روانی	Mine	۹	
[1,2]	دارد، ندارد	انگیزش درونی برای تحصیل	Goal	۱۰	

جدول ۲ : برخی از مهم‌ترین قوانین انجمنی کشف شده

	قانون انجمنی	مشخصات	تحلیل	پیشنهاد راهبردی
۱	$\text{Avg\_Motevaset} \rightarrow \text{Goal\_YES}$	Sup=0.5 Conf=0.8	۵۰ درصد دانش آموزانی که معدل متوسطی دارند انگیزه‌ای برای تحصیل دارند.	می‌توان جلسات مشاوره‌ای برای دانش آموزان جهت تعیین و شناخت اهداف تحصیل برگزار کرد.
۲	$\text{SocNet\_Yes} \rightarrow \text{Avg\_Motevaset}$	Sup=0.5 Conf=0.7	۵۰ درصد دانش آموزانی عضو شبکه‌های اجتماعی هستند معدل متوسطی دارند.	می‌تواند جلسات مشاوره‌ای جهت شناساندن اثرات استفاده نادرست از شبکه‌های اجتماعی برگزار کرد.
۳	$\text{AvgTerm\_motevaset} \rightarrow \text{Avg\_Motevaset}$	Sup=0.5 Conf=0.8	۵۰ درصد دانش آموزانی که معدل میان‌ترم متوسطی دارند معدل پایانی متوسطی دارند.	برگزاری جلساتی جهت آموزش مهارت‌های مطالعه و کلاس‌های تقویتی قبل از امتحانات پایانی
۴	$\text{SocNet\_Yes}, \text{AvgTerm\_Motevaset} \rightarrow \text{Avg\_Motevaset}$	Sup=0.4 Conf=0.9	۴۰ درصد از دانش آموزانی که عضو شبکه‌های اجتماعی هستند و معدل میان‌ترم متوسطی دارند، معدل پایانی متوسطی دارند	بررسی و برگزاری جلسات معرفی تأثیرات شبکه‌های اجتماعی بر روی دانش آموزان برای والدین
۵	$\text{Read\_Khob} \rightarrow \text{Goal\_YES}$	Sup=0.4 Conf=0.7	۴۰ درصد از دانش آموزانی که مهارت مطالعه خوبی دارند، انگیزه‌ای برای تحصیل دارند.	ترغیب دانش آموزان و افزایش انگیزش برای تحصیل و آموزش مهارت‌های مطالعه
۶	$\text{Goal\_YES} \rightarrow \text{Paye\_Aval}$	Sup=0.3 Conf=0.4	۳۰ درصد از دانش آموزانی که دارای انگیزه تحصیلی هستند در پایه اول دبیرستان تحصیل می‌کنند.	برگزاری جلسات مشاوره با دانش آموزان پایه‌های دوم و سوم دبیرستان



This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.