فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Spring 2021, 2 (1), 23-33
DOR: 20.1001.1.27832570.1399.1.2.4.0

Received: 2 Jan 2021
Accepted: 20 Feb 2021

Use Data Mining to Identify Factors Affecting Students' Academic Failure

Mahmood Najafi[1], Mehdi Afzali[2*], Mahmood Moradi[3]

1. Ms Student, Faculty of Electrical and Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran.
2. Assistant Professor, Faculty of Electrical and Computer Engineering, Islamic Azad University of Zanjan, Zanjan Iran. *Corresponding Author afzali@iauz.ac.ir
3. Assistant Professor, Department of Information Science and Dentistry, Razi University, Kermanshah, Iran.

Abstract

Knowledge extraction is one of the most significant problems of data mining. The principles raised in if-then format can be turned into real numbers in each section- as values which could be included in dataset. The suggested method in the present dissertation is application of decision tree algorithms, clustering and forum rules for extraction of final rules. In the suggested method, extraction of rules is defined as an optimization problem and objective was obtaining a rule of high confidence, generalization and understandability. The suggested algorithm for extraction of rules was obtained from and tested based on a dataset of educational failure of 256 art school students living in Zanjan. The results suggested that the j48 algorithm in decision tree and accuracy of 0.95 is the choice for the dataset of educational failure. Data clustering was done by K-Main algorithm with confidence coefficient of 0.95. After all, obtaining rules of high confidence coefficient was done based on forum rules from Apriori algorithm for the whole datasets. The results of present study could be used for inhibition of educational failure of students, improved quality of relationship of parents and authorities with students and enhancing the education they receive.

Keywords: Association rules, Clustering, Decision trees, Academic failure, Educational Data Mining

فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

کاربرد داده کاوی آموزشی جهت شناسایی عوامل مؤثر بر افت تحصیلی دانش‌آموزان

تاریخ دریافت: ۱۳۹۹/۱۰/۱۳
تاریخ پذیرش: ۱۳۹۹/۱۲/۰۲

سال دوم، بهار ۱۴۰۰
شمارهٔ اول، صص: ۲۳ – ۳۳

محمود نجفی[۱]، مهدی افضلی[۲*]، محمود مرادی[۳]

۱. کارشناسی ارشد، گروه مهندسی کامپیوتر، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران.
۲. استادیار، دانشکده مهندسی برق و کامپیوتر، واحد زنجان، دانشگاه آزاد اسلامی، زنجان، ایران.(نویسنده مسئول) afzali@iauz.ac.ir
۳. استادیار، گروه علم اطلاعات و دانش شناختی، دانشگاه رازی، کرمانشاه، ایران. mahmoudmoradi@razi.ac.ir

چکیده: استخراج دانش یکی از مهمترین مسائل داده‌کاوی می‌باشد. قوانینی که به‌صورت اگر – آنگاه مطرح‌می‌گردند این توانایی را دارند که به صورت اعداد حقیقی در هر جزء- به صورت تمام مقادیری که در مجموعه‌داده می‌تواند وجودداشته‌باشد، قرارگیرند. روش پیشنهادی در این مقاله استفاده از الگوریتم‌های درخت تصمیم و خوشه‌بندی و قوانین انجمنی برای استخراج قوانین می‌باشد. در روش پیشنهادی استخراج قوانین را به صورت یک مسئله بهینه‌سازی در آورده و هدف به‌دست‌آوردن قانونی با اطمینان بالا، عمومیت و قابلیت درک بالا می‌باشد. الگوریتم پیشنهادی برای استخراج قوانین ازمجموعه داده افت تحصیلی از اطلاعات فردی۲۵۶ دانش‌آموزان هنرستان-ها در زنجان جمع‌آوری و آزمایش‌شد. نتایج نشان‌داد که الگوریتم j48 در درخت تصمیم با دقت ۰,۹۵ برای مجموعه داده افت تحصیلی و خوشه‌بندی داده‌ها با الگوریتم K-Main با ضریب اطمینان ۰,۹۵ و به‌دست‌آوردن قوانین با ضریب اطمینان بالا با استفاده از قوانین انجمنی که از الگوریتم Apriori برای مجموعه داده‌ها اجرا و تولید شد. از نتایج به‌دست‌آمده از این تحقیق می‌توان برای پیشگیری از افت تحصیلی دانش‌آموزان و بهبود کیفیت ارتباط مسئولین و والدین با دانش‌آموزان و آموزش هرچه بهتر آنان استفاده‌کرد.

واژه‌های کلیدی: قوانین انجمنی، خوشه‌بندی، درخت تصمیم، افت تحصیلی، داده‌کاوی آموزشی.

۱. مقدمه

داده‌کاوی روشی در کشف دانش است که حجم بزرگی از داده را برای کشف دانش مفید و جذاب که درگذشته ناشناخته بوده جستجومی‌کند. با رشد سریع در اندازه و تعداد پایگاه داده‌ها، کاوش دانش، قواعد یا اطلاعات سطح بالا از داده‌ها به‌منظور پشتیبانی از تصمیم‌گیری‌ها و پیش‌بینی رفتارهای آتی، ضروری به نظرمی‌رسد. بنابراین نیاز به روش-های خودکار یا نیمه‌خودکار برای کشف دانش هست. کاوش قوانین یکی از وظایف مهم در داده‌کاوی، روند یافتن روابطی مابین خصیصه‌ها یا مابین مقادیر آن‌ها در یک پایگاه داده بزرگ است که در جهت امر تصمیم‌گیری کمک‌ساز باشند.

داده‌کاوی فرآیند استخراج اطلاعات(قوانین و الگوها) ضمنی، غیربدیهی، از قبل ناشناخته و بالقوه مفید از داده‌ها در پایگاه داده است[1]. تعاریف متنوعی از داده‌کاوی در مراجع مختلف ارائه‌شده‌است. اما تعریفی که در اکثر مراجع به اشتراک ذکرشده عبارت است از "استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده‌های بسیار بزرگ و پیچیده[۲]داده‌کاوی شامل مراحل مختلفی می‌باشد که این مراحل به صورت زیر تفکیک‌می‌شوند: تعیین اطلاعات گذشته، تمیزکردن داده و پردازش اولیه، ساخت انبارداده، انتخاب مجموعه داده‌های هدف، یافتن ویژگی‌های مورد استفاده و تعیین ویژگی‌های جدید، نمایش داده‌ها، انتخاب عملیات داده‌کاوی، استخراج الگو، ارزیابی و تحلیل الگوی به‌دست‌آمده و حذف الگوهای نامناسب و در آخر تفسیر نتایج داده‌ها و استنتاج از اطلاعات باارزش.[3]

فن‌های متنوعی در داده‌کاوی وجوددارند که الگوهای مختلفی را تولیدمی‌کنند. روش‌های کشف قوانین انجمنی[۱]، طبقه‌بندی داده‌ها[۲] و خوشه‌بندی[۳] از عمده‌ترین راهکارهایی محسوب‌می‌شوند که به تولید الگوهای خاص خود می‌پردازند[4].

طبقه‌بندی داده‌ها یکی از روش‌های مهم داده‌کاوی هست که برای نسبت دادن یک مورد، شیء یا یک رکورد به کلاس‌های از پیش تعریف‌شده مورد استفاده قرارمی‌گیرد. روش‌های مختلفی برای انجام طبقه‌بندی اطلاعات در حجم زیادی از داده‌ها پیشنهاد و استفاده‌شده-است. در این راستا روش‌های طبقه‌بندی متنوعی مانند، درخت تصمیم[۴]، شبکه‌های عصبی مصنوعی[۵]، ماشین بردار پشتیبان[۶]، نزدیک-ترین همسایه، روش‌های هیستوگرام و هسته ارائه شده‌اند.[5] به طور کلی در سال‌های اخیر استفاده از داده‌کاوی برای بهبود فرآیند برنامه-ریزی و انجام آموزش اهمیت زیادی پیدا کرده است[6].

۲. داده‌کاوی آموزشی

از داده‌کاوی در زمینه آموزش وپرورش نیز استفاده‌می‌شود برای جستجو و یافتن داده‌های مورد نظر میان دانش‌آموزان و دانشجویان، ازآنجایی که فرآیند آموزش مجموعه‌ای است از آنچه که دانش‌آموز در مدرسه از طریق برنامه درسی تعیین‌شده می‌آموزد. یکی از مشکلات نظام آموزشی، افت‌تحصیلی قابل‌توجه دانش‌آموزان است که باعث خسارت‌های روحی و روانی به دانش‌آموز و خسارت‌های اقتصادی به جامعه می‌شود که این خسارت‌ها عموماً غیرقابل‌جبران هستند. شکست یا افت تحصیلی، عبارت است از وقوع ترک تحصیل زودرس و تکرار پایه تحصیلی در نظام آموزش و پرورش یک کشور؛ که یک نظام آموزشی کارآمد، کمترین افت تحصیلی و بالاترین بازدهی را دارد. به عبارت دیگر، افت تحصیلی شامل جنبه‌های مختلف شکست تحصیلی چون غیبت مطلق از مدرسه، ترک تحصیل دانش‌آموز و سال‌های مقرر آموزشی، تکرار پایه تحصیلی، کیفیت نازل تحصیلات دانش‌آموز در مقایسه با آنچه که باید باشد و کسب محفوظات به جای معلومات است[۷]

با توجه به اینکه عده قابل‌توجهی از دانش‌آموزان از افت شدید تحصیلی در سیستم آموزش و پرورش رنج‌می‌برند لذا جهت ازمیان‌بردن این معضل اساسی، اولین راه مطالعه و تحقیق در عوامل به وجود آورنده آن و اصلاح عوامل مشکل ساز است[8]. افت تحصیلی یعنی عدم موفقیت در تحصیل به معنای عدم دستیابی فراگیر، به مجموعه مهارتهایی است که انتظار می‌رود در نتیجه آموزش، آن را یاد بگیرند. به عبارتی، به معنای عملکرد دانشجویانی است که در امتحان شکست خورده‌اند یا مجبور شده‌اند مجدداً در درس امتحان دهند. جلوگیری از افت تحصیلی یک چالش جدی است، زیرا کسانی که از نظر تحصیلی شکست می‌خورند، در طول زندگی چالش‌های اجتماعی و اقتصادی چشم‌گیری را تجربه می‌کنند.

در سال‌های اخیر تحقیقات زیادی در زمینه افت تحصیلی دانش-آموزان در حوزه رشته‌های مدیریت تربیتی و روانشناسی تربیتی انجا شده است که هر کدام به تعدادی از شاخص‌های مؤثر در افت تحصیلی پرداخته‌اند.

ونتورا[۷] و رومرو برای اولین بار بحث داده‌کاوی آموزشی را مطرح‌کردند و فعالیت تخصصی آنان به طور رسمی از سال ۲۰۰۷ میلادی با برگزاری اولین کنفرانس جهانی داده‌کاوی آموزشی توسط انجمن بین‌المللی داده‌کاوی آموزشی آغازشد[9]. اولین اقدام در این زمینه را کمپل و ابلینگر با معرفی روش جدید "تجزیه وتحلیل آموزشی" و با استفاده از روش‌های آماری و به کارگیری تکنیک‌های داده‌کاوی انجام دادند. هدف آنها کمک به اعضای هیئت علمی و مشاوران آموزشی به منظور شناسایی و کمک به دانشجویان در معرض خطر بود. آنان به نوعی پایه‌گذار شاخه‌ای از داده‌کاوی بودند که در سال‌های بعد داده‌کاوی آموزشی یا به‌اختصار "EDM " نامیده‌شد.[10].

بیکر و یاسف[۸] ضمن ارائه تعریف رسمی برای داده‌کاوی آموزشی، آن را رشته‌ای در حال ظهور معرفی‌کردند که با تمرکز بر توسعه روش‌هایی به منظور تحقیق و بررسی مبتنی بر داده‌های برگرفته از فعالیت‌های آموزشی، برای درک و شناسایی بهتر رفتار دانشجو و تنظیم روش یادگیری مطلوب به‌کارمی‌رود. این رشته بر اهمیت تجزیه وتحلیل

داده‌های آموزشی برای توسعه مدلهای بهبود فعالیت‌های یادگیری و همچنین بهبود اثربخشی سازمانی، تأکیددارد[11].

بینکوفسکی[۹] در کنفرانس بین‌المللی جاسیک ساکای سال ۲۰۱۲ اظهار کرد، می‌توان از داده‌کاوی آموزشی به منزله "یادگیری تحلیلی" به کمک مجموعه‌ای از داده‌های حجیم آموزشی یادکرد. شکل ۱ نظریه بینکوفسکی را به تصویر کشیده‌است. داده‌کاوی آموزشی از روش‌ها و تکنیک‌های آماری، یادگیری ماشین و داده‌کاوی، به منظور تجزیه و تحلیل داده‌های جمع‌آوری‌شده برای توسعه و بهبود بهره‌می‌برد و بر ایجاد مخازن اطلاعات نظامند به منزله منابع داده در داده‌کاوی تمرکزمی‌کند.[12]

داده‌کاوی زمینه‌هایی را پوشش‌می‌دهد که به شرح زیر فهرست-شده-است:

۱. یافتن علل احتمالی تداوم یا انصراف تحصیل دانشجو؛
۲. ساخت مدل رفتار دانشجو به منظور تبیین تفاوت‌های رفتاری دانشجویان و پیش‌بینی مبتنی بر آن؛
۳. فراهم‌آوری محیط‌های فراگیر فردی و سیستم‌های پیشنهاددهنده؛
۴. فراهم‌آوری زمینه کار با حجم زیادی از داده‌های ذخیره‌شده توسط کاربران غیرمتخصص و کسب دانش عمیق درباره روش‌ها، فرایندها و الگوریتم‌های یادگیری[13]
۵. پشتیبانی مؤثرتر آموزشی از دانشجویان، به کمک نر افزارهای یادگیری فردی و یادگیری مشارکتی[14]
۶. بررسی و مطالعه تئوری‌های آموزشی؛
۷. بررسی، مطالعه و گسترش تئوری‌ها و نظریه‌های آموزشی برای یافتن شواهد تجربی به منظور ادراک عمیق از عوامل کلیدی مؤثر در یادگیری [11].

۳. پیشینه پژوهش

در پژوهشی، فراوانی الگوریتم‌های استفاده شده در مطالعات مختلف طی سال‌های ۲۰۰۳ تا ۲۰۱۲ مشخص‌شده‌است. براساس داده‌های این پژوهش از ۴۸ مطالعه انجام‌شده، در سیزده مورد از "شبکه‌های عصبی"، سیزده مورد "درخت تصمیم‌گیری"، هشت مورد "رگرسیون لجستیک"، از "ماشین‌بردار پشتیبان"، "بیزین" در سه مطالعه و از "جنگل تصادفی" در دو مورد و سایر الگوریتم‌ها ازجمله: "تحلیل تشخیص خطی"، "ژنتیک"، "نزدیک‌ترین همسایه" هرکدام یک مورد استفاده قرارگرفته‌اند.[۱۵]

پیشرفت روزافزون فن‌آوری، راه‌های جدیدی را برای گسترش و بهبود سیستم‌های آموزشی ایجادکرده‌است. تحقیقات انجام‌شده در سال‌های اخیر بیانگر این مطلب است که تصمیم‌گیرندگان عرصه آموزش عالی و مسئولین آموزش و پرورش به این باور رسیده‌اند که امروزه عملکرد تحصیلی دانش‌آموزان و دانشجویان تنها پیامد هوش فردی و یا امکانات سخت‌افزاری موجود در محیط نیست بلکه پیامد جنبه‌های مختلف آموزشی، تربیتی، اجتماعی و ... می‌باشد[۱۶].

اندری و همکاران[۱۷] با استفاده از تکنیک‌های قوانین همبستگی و درخت تصمیم، اطلاعات ۵۹۹ دانش‌آموز مقطع راهنمایی را که از طریق پرسشنامه جمع‌آوری شده‌بود را تحلیل-کرده‌اند و عوامل مؤثر در بروز افت تحصیلی را معرفی‌نموده‌اند. به عنوان مثال اگر عزت نفس دانش‌آموز بالا، جنسیت او پسر و وضعیت اقتصادی آن دانش‌آموز ضعیف باشد، آنگاه برای او افت تحصیلی وجوددارد.

اسکندریان و همکاران[۱۸] با استفاده از شبکه‌های بیزین عملکرد تحصیلی ۱۹۹ نفر از دانش‌آموزان دوره متوسطه را ارزیابی‌کرده‌اند و در نهایت بر اساس بهبود به‌دست‌آمده در مدل تولید شده، شبکه‌های بیزین را به عنوان ابزاری مناسب جهت ارزیابی عملکرد تحصیلی دانش-آموزان به منظور غلبه بر افت تحصیلی آن‌ها معرفی‌نموده‌اند.

قلندری[۱۹] رابطه بین ویژگی‌های شخصیتی و عملکرد تحصیلی دانش‌آموزان سال سوم دبیرستان شهر زنجان را مورد بررسی قرارداد و به این نتیجه دست‌یافت که بین ویژگی‌های شخصیت (مسئولیت-پذیری، انعطاف‌پذیری و برونگرایی) و عملکرد تحصیلی دانش‌آموزان رابطه مثبت معناداری وجود دارد و ویژگی‌های شخصیت می‌تواند ٪۵۹ عملکرد تحصیلی را تبیین‌نماید.

هدایتی و همکاران[۲۰] با تحلیل اطلاعات بیش از ۲۵۹ نفر از دانشجویان و با بهره‌گیری از روش‌های داده‌کاوی، عدم انتخاب رشته بر اساس علاقه‌ی فردی را به عنوان مهمترین عامل در قرارگرفتن دانشجویان در خوشه‌ی ضعیف شناسایی‌کرده‌اند.

هدایتی و جوانمرد[۲۱] با استفاده از روش‌های داده‌کاوی، از جمله درخت تصمیم‌گیری CART و الگوریتم خوشه‌بندی K-Means و تحلیل داده‌های جمع‌آوری‌شده از پرسشنامه‌هایی که توسط تمامی دانش‌آموزان سال اول دبیرستان دخترانه مهدیه شهر یزد تکمیل‌شده-بود، به بررسی آن دسته از عواملی پرداخته‌اند که بر روی افت تحصیلی این دانش‌آموزان مؤثر هستند.

پرویز اسدی ورمله و همکارنش[۲۲] با استفاده از از ترکیب و مدل-سازی دو الگوریتم قوانین همبستگی و درخت تصمیم پژوهشی بر روی ۴۸۱۰ نفر دانش‌آموز سال اول دبیرستان صورت‌دادند و قوانین موجود بین این فیلدها استخراج شده‌است. داده‌های استفاده‌شده در این تحقیق از طریق پایگاه داده نمرات دانش‌آموزان و پرونده‌های مشاوره‌ای و تربیتی آن‌ها جمع‌آوری شده‌است. در ارائه این تحلیل استفاده-شده-است.

اسماعیل‌پور و همکارانش[۲۳] به بررسی افت تحصیلی دانش-آموزان به هنگام تبدیل‌شدن مدرسه از حالت سنتی به هوشمند با استفاده از روش الگوریتم ژنتیک در مناطق کم‌برخوردار پرداخته‌اند.

ایرجی و همکارانش[۲۴] در پژوهشی با به‌کارگیری داده‌کاوی سعی در آسیب‌شناسی افت تحصیلی هنرجویان هنرستان‌های فنی و حرفه‌ای در رشته‌های مختلف بین سال‌های ۸۰ تا ۸۳ نمودند. ایجاد پایگاه اطلاعاتی داده جهت فرایند داده‌کاوی، پیاده‌سازی راه حل پروژه داده-کاوی، تنظیم منابع داده مورد استفاده، اجرای الگوریتم‌های کاوش با استفاده از پایگاه داده مایکروسافت SQL Server2005 انجام‌شده است. در این پژوهش تلاش‌می‌شود با توجه به مشخصه‌های ورودی هنرجویان مجموعه آماری نمونه، روند نمرات در طول تحصیل و ارتباط آن با مشخصه‌های خروجی، دلایل افت تحصیلی مورد تحلیل و بررسی قرارگیرد و با کشف الگوهای مناسب رشد، نمایه‌ای برای پیش‌بینی هنرجویان موفق ایجادگردد.

یقینی و همکارانش[۲۵] با استفاده از تکنیک‌های داده‌کاوی وضعیت تحصیلی آتی دانشجویان شامل معدل نیمسال آینده، معدل کل در زمان فارغ‌التحصیلی، و وضعیت فارغ‌التحصیلی پیش‌بینی‌شده-است. برای ساخت مدل‌های مورد نظر از تکنیک‌های مختلفی نظیر شبکه‌های عصبی، درخت‌های تصمیم و SVM استفاده‌شده‌است. این مدل‌ها برای داده‌های سیستم آموزشی دانشگاه علم و صنعت ایران پیاده‌سازی شده‌اند. عملکرد هریک از مدل‌ها، مورد بررسی قرارگرفته و نتایج به‌دست‌آمده با یکدیگر مقایسه‌شده‌اند. اعتبارسنجی انجام‌شده بر روی مدل‌ها اثبات می‌کند که نتایج به‌دست‌آمده دقیق و قابل‌اعتماد بوده‌اند.

ابدو و همکارنش در سال ۲۰۰۷ در پژوهشی با موضوع " تأثیر رفتار معلم و نوع برخورد او در کلاس با وضعیت تحصیلی " که به کمک قوانین انجمنی در داده‌کاوی به این تنیجه رسیدندکه ۷۱٪ دانش‌آموزان علت پیشرفت تحصیلی خود را " همدلی معلم " می‌دانند. آن‌ها گفته‌اند ما در کلاس با این معلم راحت بودیم و این راحتی باعث-می‌شد ما مطالب را راحت‌تر یادبگیریم. یک معلم مهربان و صمیمی می‌تواند حتی خلاء عاطفی و روانی دانش‌آموزان را که با مشکلات خانواده ارتباط دارد پرکند. تشویق زبانی معلم و پاداش‌دادن او انرژی-های روانی فرد را فعال‌کرده و روز به روز پیشرفت تحصیلی و تعادل رفتاری دانش‌آموز بهتر خواهدشد.

۴. اهداف پژوهش و راه کار پیشنهادی

در زمینه بررسی افت تحصیلی دانش‌آموزان پژوهش‌های قابل‌توجهی صورت‌گرفته و روش‌های مختلفی نظیر خوشه‌بندی، شبکه‌های بیزین، درخت تصمیم به‌طور جداگانه به‌کار گرفته‌شده‌است. در پژوهش حاضر ما عوامل مؤثر بر افت تحصیلی دانش‌آموزان هنرستان پسرانه شهر زنجان با استفاده از ترکیب الگوریتم‌های درخت تصمیم، خوشه‌بندی و قوانین انجمنی بررسی‌کرده‌ایم.

اهداف مشخص تحقیق، شامل اهداف علمی (کلی و جزئی) و اهداف کاربردی این پژوهش عبارتند از:

- شناسایی عوامل مؤثر بر افت تحصیلی دانش‌آموزان
- شناسایی و بررسی ویژگی‌های دانش‌آموزان دارای افت تحصیلی

این پژوهش تلاشی هست تا با جمع‌آوری اطلاعات آموزشی و فردی دانش‌آموزان دبیرستان پسرانه شهر زنجان به تحلیل عوامل مرتبط با پدیده‌ی افت تحصیلی بپردازیم. پس از جمع‌آوری داده‌ها، آماده‌سازی و پیش پردازش داده‌ها، تکنیک‌های خوشه‌بندی و قوانین انجمنی بر روی این داده ها اعمال شدند. در نهایت، عوامل موثر بر افت تحصیلی وشناسی وقوع افت تحصیلی پس از ارزیابی‌های مختلف استخراج گردیدند. در این پژوهش از تکنیک‌های خوشه‌بندی و قوانین انجمنی برای پیشگیری از افت تحصیلی دانش‌آموزان و بهبود کیفیت ارتباط مسئولین و والدین با دانش‌آموزان استفاده‌شده‌است، که داه‌ها مورد نیاز با استفاده از پرسشنامه از دانش‌آموزان شاغل در هنرستان‌های شهر زنجان گردآوری شده‌است.

در این مطالعه برای جمع‌آوری اطلاعات از پرسش‌نامه استفاده‌شده که بین ۲۵۶ نفر از دانش آموزان هنرستان‌های شهر زنجان توزیع شده-است. تعداد ۱۳۷ نفر از دانش آموزان در مقطع سوم دبیرستان و ۱۱۹ نفر در مقطع دوم دبیرستان می‌باشند. رشته‌های تحصیلی دانش‌آموزان و تعداد آن‌ها در جدول ۱ نشان‌داده‌شده‌است.

جدول ۱: رشته‌های تحصیلی و تعداد دانش‌آموزان

رشته تحصیلی	تعداد	رشته تحصیلی	تعداد
نقشه	۹	تربیت بدنی	۳۲
هنر	۱۱	مکا نیک خودرو	۲۰
مدیریت	۲۰	ساخت وتولید	۳۸
کامپیوتر	۵۴	تاسیسات	۴
صنایع	۸	حسابداری	۴۸
گرافیگ	۱۲		
معداد کل	**۲۵۶**		

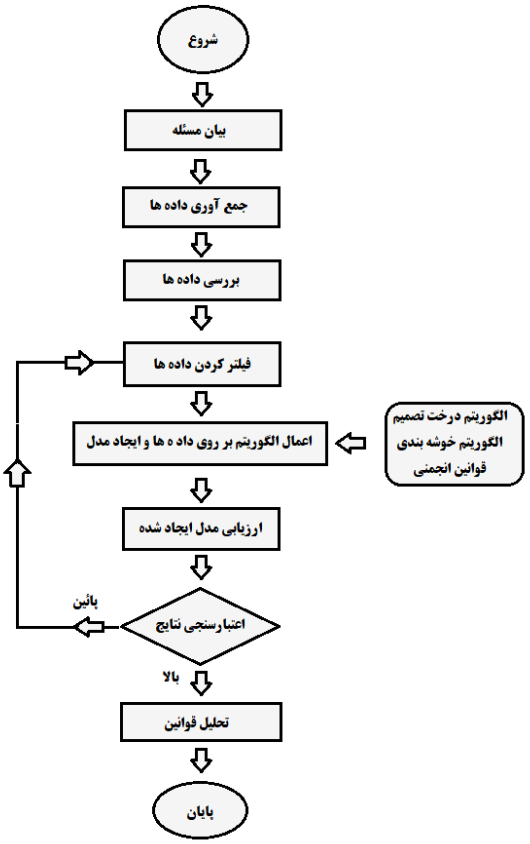
روایی داده‌های جمع‌آوری‌شده از پرسش‌نامه‌ها با نظر متخصصان تأییدگردید. سپس داده‌ها که شامل ۲۰۰ رکورد اطلاعاتی به همراه ۲۰ فیلد بودند، در یک فایل صفحه گسترده اکسل Excel قرارگرفتند.

۱،۴. روش پیشنهادی

داده‌ها با استفاده از ابزارهای پیش‌پردازش داده نرم‌افزار، کارهای فیلترکردن حذف ویژگی‌های اضافی و در صورت نیاز تبدیل ویژگی اسمی به عددی، تعداد ۲۸ ویژگی برای ادامه کاوش انتخاب‌شدند.

داده‌های انتخابی در این تحقیق که از طریق پرسشنامه جمع‌آوری شده‌بود به صورت فایل اکسل به فرمت CSV تبدیل و سپس به نرم-افزار Weka منتقل‌گردید. با استفاده از ابزارهای پیش‌پردازش داده این نرم‌افزار، کارهای فیلترکردن، حذف ویژگی‌های اضافی و در صورت نیاز تبدیل ویژگی اسمی به عددی انجام‌گردید. با توجه به ماهیت تحقیق که بررسی ویژگی افت تحصیلی دانش‌آموزان می‌باشد، تصمیم گرفته‌شد تا برروی خوشه‌بندی متمرکزشویم.

بعد از مشخص نمودن ورودی نرم‌افزار از آنجا که باید اعتبار روش ارائه‌شده در هر پژوهشی مورد سنجش قرارگیرد، برای سنجش اعتبار و صحت مدل، با استفاده از ویژگی‌های نرم‌افزار Weka داده‌های به دو مجموعه داده آموزشی (۹۰٪) و آزمایشی(۱۰٪) استفاده شده‌است. میزان اعتبار با نتایج داده‌های جدید آزمون‌می‌شود و داده‌های آزمایشی به‌منزله ناظر به الگوریتم وارد شده و نتایج میزان صحت آن را ارزیابی می‌کند معیار اعتبار و صحت مدل، بسته به صحت طبقه‌بندی یا تفکیک داده‌های آزمایشی است. در این پژوهش، مجموع داده‌ها با میزان اعتبار سنجی ده مرتبه تکرار برای کلیه الگوریتم‌ها سنجیده-می‌شود.



شکل ۲: فلوچارت روش انجام پژوهش

قبل اعمال الگوریتم‌های داده‌کاوی می‌توان فیلدها یا خصیصه‌هایی که نیاز پردازش یا حذف دارند را با استفاده از فیلتر در نرم‌افزار weka این پردازش‌ها را انجام‌داد که باعث بهبود الگوریتم مورد استفاده می-گردد. در این پژوهش، برخی فیلدها با استفاده از فیلتر Remove حذف‌شده است. همچنین یکی دیگر از فیلترهای موجود، Discretize است که با استفاده از آن می‌توان مقادیر یک صفت پیوسته را به تعداد دلخواه بازه گسسته تبدیل‌کرد(مانند فیلد سن، معدل و یا میزان درآمد خانواده)

۲,۴. معیارهای ارزیابی مدل

یکی از مهمترین قسمت‌هایی که در دسته‌بندی باید به آن توجه نمود، ارزیابی مدل دسته‌بندی است. ارزیابی یک مدل دسته‌بندی می‌تواند براساس نمونه‌های آموزشی و آزمایش صورت‌گیرد. برای ارزیابی باید برچسبی که مدل افت تحصیلی نسبت داده شده‌است مقایسه‌شود. وقوع حالات مختلف برای دسته‌ها باتوجه به مجموعه داده‌های ورودی برای دسته‌بندی با مقادیر FN, TN, FP, TP برای دو دسته مثبت و منفی در جدول شماره ۲ نشان داده شده‌است.

جدول ۲: ماتریس آشفتگی برای یک مسئله دسته‌بندی دو دسته‌ای

نوع رکورد	رکوردهای تخمینی (Predicated Records)		
رکوردهای واقعی (Actual Records)	نوع دسته	دسته -	دسته +
	دسته -	TN	FP
	دسته +	FN	TP

TN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز آن را به‌درستی منفی تشخیص‌داده‌است.

TP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز آن را درست تشخیص‌داده است.

FP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی دسته آن‌ها را به‌اشتباه مثبت تشخیص‌داده‌است.

FN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی دسته آن‌ها را به‌اشتباه منفی تشخیص داده‌است.

با توجه به پارامترهای مطرح‌شده در رابطه‌های زیر معیارهای ارزیابی مختلفی ارائه‌شده‌است که ازجمله آن‌ها می‌توان به معیار درستی، دقت، فراخوانی و معیار F-measure اشاره‌کرد.

مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بند معیار Accuracy می‌باشد. این معیار دقت کل یک دسته‌بند را محاسبه می-کند. این معیار نشان‌دهنده این موضوع است که چند درصد از کل مجموعه داده‌ها به‌درستی دسته‌بندی شده‌است. رابطه شماره ۱ نحوه محاسبه معیار درستی را نشان‌می‌دهد[26]

رابطه شماره ۱

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

دو مقدار TN, TP مهم‌ترین مقادیری هستند که باید بیشینه شوند تا کارایی دسته‌بندی به حداکثر برسد. با توجه به این نکته که در مسائل دسته‌بندی ممکن است بین تعداد دسته‌های مختلف توازنی برقرار نباشد، ممکن است یک دسته دارای نمونه‌های خیلی بیشتر از دسته دیگر باشد. در نتیجه مدل نهایی به سمت دسته با بیش‌ترین نمونه سوق پیدامی‌کند. بنابراین دسته دارای تعداد نمونه کم عملاً تأثیر

چندانی در بهبود یا عدم بهبود کارایی نخواهدداشت. می‌توان نتیجه‌گرفت که معیار Accuracy در مجموعه داده‌هایی که دارای دسته نامتعادل با تعداد مختلف نمونه، معیار مناسبی نمی‌باشد.

Precision: معیار precision درصدی را نشان‌می‌دهد که از میان تمامی دسته‌ها که توسط دسته‌بند به آن دسته نسبت داده شده‌اند. به عبارتی دقت دسته i را با توجه به کل مواردی نشان‌می‌دهد که برچسب i برای نمونه مورد بررسی توسط دسته‌بند پیشنهاد شده‌است. نحوه محاسبه این معیار در رابطه شماره ۲ نشان داده شده‌است. اندیس i در این پارامترها به این مفهوم است که پارامترها باید برای هر دسته i در این پارامترها باید برای هر دسته محاسبه شوند[26]

رابطه شماره ۲

$$Precision_i = \frac{TP_i}{TP_i+FP_i}$$

Recall: دقت دسته‌بندی دسته i را با توجه به کل نمونه‌های با برچسب i نشان‌می‌دهد. نحوه محاسبه این معیار در رابطه شماره ۳ نشان داده شده است.

رابطه شماره ۳

$$Recall_i = \frac{TP_i}{TP_i+FN_i}$$

نکته قابل‌توجه این است که معیار Recall کارایی دسته‌بند را با توجه به تعداد رخداد دسته i نشان‌می‌دهد در حالیکه Precision اساساً مبتنی بر دقت پیش‌بینی دسته می‌باشد و بیاینگر آن است که به چه میزان می‌توانیم به خروجی اعتمادکنیم.

F-Measure: این معیار از ترکیب معیارهای Precision و Recall به‌دست‌می‌آید و در مواردی استفاده‌می‌شود که نتوان اهمیت ویژه‌ای را برای هر یک از دو معیار Recall و Precision نسبت به یکدیگر قائل‌شد. رابطه شماره ۴ نحوه محاسبه این معیار را نشان می‌دهد:

$$F-measure = \frac{2*Precision_i*Recall_i}{Precision_i+Recall_i}$$

۵. اعمال الگوریتم درخت تصمیم

الگوریتم درخت تصمیم، یک الگوریتم تولید مدل کلاس‌بندی و رگرسیون است که توسط خدمات تحلیلی به منظور استفاده در مدل‌های پیش‌بینی در مورد مشخصه‌های گسسته و پیوسته فراهم‌شده‌است. به منظور مشخصه‌های گسسته، این الگوریتم پیش‌بینی‌هایی را بر مبنای روابط بین ستون‌های ورودی در یک مجموعه داده انجام‌می‌دهد. از مقادیر و وضعیت‌های این ستون‌ها برای پیش‌بینی وضعیت یک ستون که به عنوان ستون قابل‌پیش‌بینی انتخاب شده‌است، استفاده‌می‌نماید. در خصوص مشخصه‌های پیوسته الگوریتم از روش رگرسیون خطی برای تعیین اینکه درخت تصمیم‌گیری در کجا تقسیم‌می‌شود، استفاده‌می‌نماید. اگر بیش از یک ستون قابل‌پیش‌بینی داشته‌باشیم و یا اگر داده ورودی شامل یک جدول تو در تو باشد در آن صورت یک مجموعه قابل پیش‌بینی وجوددارد و الگوریتم به ازای هر یک از ستون‌های قابل پیش‌بینی یک درخت تصمیم‌گیری می‌سازد[27]

درخت تصمیم‌گیری C4.5 که توسط weka استفاده‌می‌گردد تحت عنوان J48 شناخته‌می‌شود که درخت تصمیم‌گیری دودیی ایجاد می‌کند. C4.5 درخت‌های تصمیم‌گیری را از مجموعه داده‌های آموزشی می‌سازد که مانند ID3 از مفهوم آنتروپی اطلاعات استفاده‌می‌کند. مهمترین مزیت J48 این است که تعداد نقاط تصمیم‌گیری را محدود می‌کند. در ادامه این تحقیق الگوریتم J48 را برروی داده‌های مورد مطالعه اعمال نموده که در شکل ۳ نتیجه حاصل از اجرای این الگوریتم را نشان‌می‌دهد.

```
Correctly Classified Instances         243               95.2941 %
Incorrectly Classified Instances        12                4.7059 %
Kappa statistic                          0.9046
Mean absolute error                      0.0898
Root mean squared error                  0.2093
Relative absolute error                 18.1901 %
Root relative squared error             42.133  %
Total Number of Instances              255
Ignored Class Unknown Instances                 1

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.958     0.053      0.958     0.958     0.958      0.936    NO
                 0.947     0.042      0.947     0.947     0.947      0.936    YES
Weighted Avg.    0.953     0.048      0.953     0.953     0.953      0.936

=== Confusion Matrix ===

   a   b   <-- classified as
 136   6 |   a = NO
   6 107 |   b = YES
```

شکل ۳: خروجی حاصل از اجرای الگوریتم J48

بعد از اجرای الگوریتم درخت تصمیم نتاج زیر به‌دست‌آمد: مقدار TP=136، مقدار FN=6، مقدار FP=6، مقدار TN =107 می‌باشد. به عبارت دیگر نتیجه ماتریس آشفتگی حاکی از آن است که از تعداد ۱۴۲ نفر دانش‌آموز که در دسته a قرار دارند، ۱۳۶ نفر به‌درستی در همین دسته دسته‌بندی شده‌اند و ۶ نفر به صورت نادرست در دسته b دسته‌بندی شده‌است. همچنین از ۱۱۳ نفر دانش‌آموز که در دسته b دسته قرار دارند ۱۰۷ نفر به‌درستی در دسته b دسته‌بندی شده‌اند اما ۶ نفر به صورت نادرست در دسته a دسته‌بندی شده ست. نتایج به‌دست آمده از اجرای الگوریتم j48 برای معدل امسال در مجموعه داده نشان‌دهنده این است که کسانی که معدل آن‌ها در c قرار گیرد افت تحصیلی دارند.

نتایج اجرای الگوریتم j48 با اعتبار ۱۰ در جدول ۳ بیان شده‌است. براساس این نتایج از تعدادکل نمونه که ۲۵۶ نمونه بود ۲۵۵ در دسته‌بندی شرکت داده شده‌است و یکی از نمونه در دسته‌بندی شرکت داده نشده است. که از ۲۵۵ مورد ۲۴۳ نمونه صحیح دسته‌بندی شده‌است. و دوازده نمونه نادرست دسته‌بندی شده و ضریب اطمینان برای مدل ایجاد شده توسط الگوریتم J48 برابر با ۹۵,۷۰ است. ضریب دسته‌بندی نادرست ۴,۳۰ می‌باشد.

با هدف راستی‌آزمایی و تعیین دقت، الگوریتم J48 با برخی از الگوریتم‌های دیگر درخت تصمیم که در دیگر کارهای مربوط به افت

تحصیلی استفاده شده است ازجمله: ADTree و DecisionStump داده‌های موجود اجرا گردید و نتایج به‌دست آمده باهم مقایسه‌شد.

با بررسی نتایج و با توجه به آینکه هرچه قدرضریب اطمینان بالاتر باشد، مدل ایجادشده مدل بهتری خواهد،بود؛ مشخص‌گردید که j48 نسبت به دو الگوریتم دیگر از دقت بیشتری برخوردار بوده و مدل الگوریتم طول بهتری است.

جدول ۳: مقایسه الگوریتم‌ها

گوریتم	j4	ADTre	DecisionStum
خ دقت	۹۵,۷	۹۴,۹	۹۴,۹

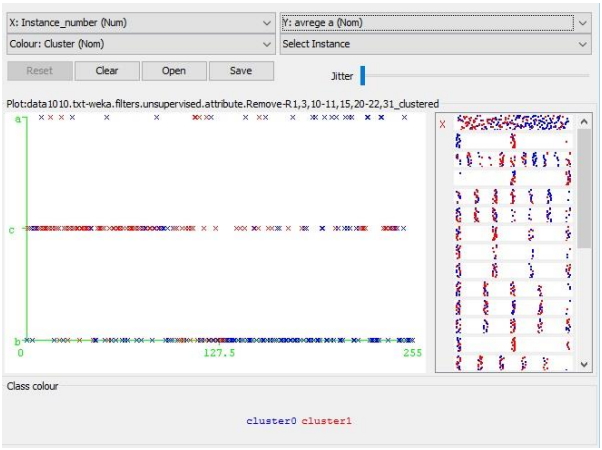
۶. اعمال الگوریتم خوشه‌بندی K-MEANS

روش‌ها و الگوریتم‌های متعددی برای تبدیل اشیاء به گروه‌های همشکل یا مشابه وجوددارد. الگوریتم K-Means یکی از ساده‌ترین و محبوب‌ترین الگوریتم‌هایی است که در داده‌کاوی، به‌خصوص در حوزه "یادگیری نظارت نشده[۱۰]" به‌کارمی‌رود. معمولا در حالت چندمتغیره، باید از ویژگی‌های مختلف اشیا به منظور طبقه‌بندی و خوشه‌کردن آن‌ها استفاده‌کرد. به این ترتیب با داده‌های چند بعدی سروکارداریم که معمولا به هر بعد از آن، ویژگی یا خصوصیت گفته‌می‌شود. با توجه به این موضوع، استفاده از توابع فاصله مختلف در این‌جا مطرح‌می‌شود. ممکن است بعضی از ویژگی‌های اشیا کمی و بعضی دیگر کیفی باشند. به هر حال آنچه اهمیت دارد روشی برای اندازه‌گیری میزان شباهت یا عدم شباهت بین اشیاء است که باید در روش‌های خوشه‌بندی لحاظ شود.

الگوریتم K-Means از گروه روش‌های خوشه‌بندی تفکیکی[۱۱] محسوب می‌شود و درجه پیچیدگی محاسباتی آن برابر با O(ndk+1) است، به شرطی که n تعداد اشیاء، d بعد ویژگی‌ها و k تعداد خوشه‌ها باشد. همچنین پیچیدگی زمانی برای این الگوریتم برابر با O(nkdi) است، که البته منظور از i تعداد تکرارهای الگوریتم برای رسیدن به جواب بهینه است.

۱,۶. مصورسازی نتایج

روش دیگر برای کسب اطلاعات در مورد هرخوشه، مصورسازی است. انتخاب‌های مختلف برای هرکدام از سه بعد نمودار حاصل (محور X، محورY، رنگ) نمودارهای مختلفی را نتیجه‌می‌دهد که می‌توان از آن‌ها اطلاعات مورد نظر را به‌دست‌آورد. بد از اجرای خوشه‌بندی نتایج قابل-استفاده‌ای به‌دست‌آمد. به عنوان مثال: در شکل ۴ محور X ها نماینده شماره خوشه، محور Y ها نماینده شماره نمونه در بانک اطلاعاتی، و رنگ‌ها نماینده جنسیت هستند (قرمز: پسران، آبی: دختران). همانطورکه مشاهده‌می‌شود خوشه۰ بیشتر توسط دختران و خوشه ۱ توسط پسران احاطه شده‌است.



شکل ۴: مصورسازی نتیجه حاصل از خوش‌بندی

۲,۶. تفسیر خوشه‌های ایجادشده از الگوریتم K-means

بعد از نتیجه اجرای الگوریتم k-means می‌توان خروجی را به دست آورد که در شکل ۵ قابل‌مشاهده است.

```
Cluster centroids:
                                Cluster#
Attribute               Full Data       0         1
                          (256)     (145)     (111)
=============================================================
SEX                             M         F         M
Field of study           COMPUTER  COMPUTER sagttolid
academic year                   C         C         B
father s level of education     C         C         C
mother s level of education     D         D         C
job status of father            C         C         C
job status of mother            D         D         D
 average income                 A         A         A
Social position                 B         B         B
 number of family members       C         C         C
tudent groups and activities    C         C         C
number of students in class     A         A         B
amount of interest in study     C         B         C
Student employment             NO        NO        NO
avrege a                        c         b         c
Intend to continue studying     A         A         A
Hoping to study with employment C         B         C
value of education              C         A         C
academic position of friends    C         C         C
average study                   B         B         A
Incentives                   Both      Both      Both
average use of the mass media   E         C         E
 goal of                      JOB    FUTURE       JOB
Academic drop                  NO        NO       YES


Time taken to build model (full training data) : 0.14 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       145 ( 57%)
1       111 ( 43%)
```

شکل ۵: نتایج حاصل از اجرای الگوریتم خوشه‌بندی

خوشه (۰): در این خوشه ۱۴۵ نفر (۵۳٪) قراردارند، که جنسیت اکثریت دانش‌آموزان این خوشه دختر می‌باشد. تحصیلات پدر در سطح راهنمایی و مادر در سطح دیپلم و شغل پدر خانواده این دانش‌آموزان آزاد و شغل مادر خانه‌دار است. میانگین سن پدران بین ۴۰ تا ۴۴ سال و میانگین سن مادران بین ۳۵ تا ۳۹ سال می‌باشد. میانگین درآمد

خانواده و موقعیت اجتماعی خانواده‌های این خوشه در سطح متوسط و تعداد متوسط افراد خانواده نیز چهار نفر است. سرپرست دانش‌آموزانی که در این خوشه قراردارند برعهده پدر خانواده می باشد. میزان مشارکت دانش‌آموزان این خوشه در گروه‌های دانش‌آموزی متوسط و تعداد دانش‌آموزان کلاس این خوشه بین ۱۵ تا ۲۰ نفر است. این دانش‌آموزان علاقه زیادی به تحصیل دارند. عدم اشتغال، قصد ادامه تحصیل، امید زیاد به اشتغال درآینده با ادامه تحصیل از ویژگی‌های آن‌هاست. میانگین مطالعه روزانه دانش‌اموزان این شاخه، بین یک تا دو ساعت و مشوق تحصیلی آن‌ها والدینشان هستند. روزانه بین و تا سه ساعت از رسانه‌های جمعی (رادیو، تلویزیون و ...) و یا اینترنت استفاده-می‌کنند. با والدین ارتباط صمیمی دارند. دانش‌آموزان این خوشه دارای افت تحصیلی نیستند.

خوشه (۱): در این خوشه ۱۱۱نفر (۴۷٪) قراردارند که جنسیت اکثریت دانش‌آموزان این خوشه پسر می‌باشد. تحصیلات والدین در سطح راهنمایی و شغل پدر خانواده این دانش‌آموزان آزاد و شغل مادر خانه-دار است. میانگین سن پدران ۴۵ تا ۴۹ سال و میانگین مادران این خوشه ۳۵ تا ۳۹ می‌باشد. میانگین درآمد خانواده و موقعیت اجتماعی خانواده‌های این خوشه در سطح خوب و تعداد متوسط افراد خانواده نیز در حال حاضر چهار نفر است. سرپرستی اکثر دانش‌آموزان این خوشه نیز برعهده پدر خانواده می باشد. میزان مشارکت دانش‌آموزان این خوشه در گروه‌های دانش‌آموزی متوسط و تعداد دانش‌آموزان کلاس این خوشه بین ۲۱ تا ۲۵ نفر است. این دانش‌آموزان علاقه زیادی به تحصیل دارند. عدم اشتغال، قصد ادامه تحصیل، امید زیاد به اشتغال درآینده با ادامه تحصیل از ویژگی‌های آن‌ها است. میانگین مطالعه روزانه دانش‌اموزان این شاخه، کمتر از یک ساعت و مشوق تحصیلی آن‌ها والدینشان هستند. روزانه بیش از چهار سه ساعت از رسانه‌های جمعی (رادیو، تلویزیون و ...) و یا اینترنت استفاده‌می‌کنند. والدین این دانش‌آموزان ارتباط صمیمی با هم دارند. دانش‌آموزان این خوشه دارای افت تحصیلی هستند.

۷. اعمال الگوریتم Apriori

این الگوریتم کار را با مقدارپشتیبانی معادل صددرصد برای تمامی data item ها اغازمی‌کند. سپس در هر مرحله از تکرار میزان 5% از آن را کم‌می‌کند تا اینکه حداقل ده قانون با میزان اطمینان برابر ۰,۹ ایجاد شوند و یا اینکه میزان پشتیبانی به ۱۰٪ برسد. البته کلیه این اعداد و ارقام به صورت پیش‌فرض تعیین‌شده‌اند و شما می‌توانید آن‌ها را تغییردهید.

```
priori
======

inimum support: 0.3 (77 instances)
inimum metric <confidence>: 0.9
umber of cycles performed: 14

enerated sets of large itemsets:

ize of set of large itemsets L(1): 27

ize of set of large itemsets L(2): 45

ize of set of large itemsets L(3): 11

est rules found:

1. number of students in class=B 89 ==> SEX=M 88    conf:(0.99)
2. SEX=F 91 ==> number of students in class=A 89    conf:(0.98)
3. Student employment=NO avrege a=b 89 ==> Academic drop=NO 87    conf:(0.98)
4. SEX=F Student employment=NO 85 ==> number of students in class=A 83    conf:(0.98)
5. Student employment=NO Academic drop=YES 83 ==> avrege a=c 80    conf:(0.96)
6. avrege a=b 112 ==> Academic drop=NO 107    conf:(0.96)
7. Academic drop=YES 113 ==> avrege a=c 107    conf:(0.95)
8. avrege a=c 113 ==> Academic drop=YES 107    conf:(0.95)
9. Student employment=NO avrege a=c 85 ==> Academic drop=YES 80    conf:(0.94)
0. SEX=M avrege a=c 83 ==> Academic drop=YES 78    conf:(0.94)
```

تصویر شماره ۶: نتیجه اجرای الگوریتم Apriori

این الگوریتم کارخود را با مقدار پشتیبانی معادل صددرصد برای تمامی data itemها آغازمی‌کند. سپس در هر مرحله از تکرار میزان 5% ازآن را کم‌می‌کند تا اینکه حداقل ده قانون با میزان اطمینان برابر ۰,۹ ایجادشوند و یا اینکه میزان پشتیبانی به ۱۰٪ برسد. مسئله کشف قوانین به‌عنوان یک مسئله بهینه‌سازی با استفاده از الگوریتم Apriori با یک کینگ ساده حل‌کنیم و از این الگوریتم برای کشف قوانین از مجموعه داده‌ی که جمع‌آوری‌شده استفاده و نتایج تحلیل گردید. با توجه به قوانین کشف‌شده در الگوریتم Apriori می‌توان مهمترین قوانین کشف‌شده در جدول ۴ بیان‌کرد.

جدول ۴: بهترین قوانین حاصل از الگوریتم Apriori

ضریب اطمینان	قانون	نیجه
گر دانش آموز شاغل نباشد و معدل بین ۱۴-۱۶ باشد	انش آموز افت تحصیلی ندارد	۰,۹
اگر دانش آموز شاغل نباشد و افت تحصیلی داشته باشد	عدل دانش آموز برابر با زیر ۱۳	۰,۹
گر معدل دانش آموز بین ۱۴-۱۶ باشد	انش آموز افت تحصیلی ندارد	۰,۹
گر دانش آموز افت تحصیلی داشته باشد	عدل دانش آموز برابر با زیر ۱۳	۰,۹
گر دانش آموز شاغل نباشد و معدل برابر با زیر ۱۳ باشد	انش آموز افت تحصیلی دارد	۰,۹
اگر دانش آموز پسر باشد و معدل برابر با زیر ۱۳ باشد	انش آموز افت تحصیلی دارد	۰,۹

نتایج آزمایش‌ها نشان‌می‌دهد این الگوریتم در تولید قوانین بسیار توانمند عمل‌نموده و طبقه‌بندی مناسبی ارائه‌می‌نماید.

۸. نتیجه‌گیری

به‌طورکلی در این پژوهش با تحلیل مجموعه داده‌ها و ویژگی‌های انتخابی علل افت تحصیلی با بهره‌گیری از الگوریتم درخت تصمیم j48

و خوشه‌بندی K_Means و الگوریتم قوانین انجمنی Apriori بررسی و مؤثرترین قوانین مرتبط با افت تحصیلی و خوشه ضعیف استخراج و پیش‌بینی افت تحصیلی استخراج‌شد. نتایج داده‌کاوی نشان‌می‌دهد شاغل بودن دانش‌آموز و معدل پایین، بیشترین نشانه در افت تحصیلی دانش‌آموزان می‌باشد.

با بررسی ویژگی‌های دوخوشه به‌دست‌آمده از الگوریتم خوشه‌بندی K-Means مشخص‌می‌شود که افت تحصیلی بیشتر در بین دانش‌آموزان پسر در مقطع تحصیلی دوم دیده می‌شود. بیشتر این دانش‌آموزان شاغل بوده و سن پدرانشان بین ۴۵ تا ۴۹ است که نسبت به سن پدران خوشه دیگر تقریباً بالا است(یعنی تفاوت سنی زیاد بین پدر با فرزند). تعداد افراد کلاس این خوشه نسبت به خوشه‌ای که افت تحصیلی ندارند تقریبا زیاد است. همچنین دانش‌آموزانی که دارای افت تحصیلی هستند میانگین مطالعه درسی روزانه کمتر از یک ساعت داشته و روزانه بیش از چهار ساعت از رسانه‌های جمعی و اینترنت استفاده‌می‌کنند. همچنین علاقه زیادی به تحصیل ندارند. تنها هدفشان از ادامه تحصیل، داشتن شغل در آینده است.

نتایج حاصل از اجرای الگوریتم Apriori در استخراج قوانین نسبت به دیگر روش‌های که برای به‌دست آوردن افت تحصیلی به‌کار گرفته‌می‌شود دارای ضریب اطمینان مناسبی بوده و می‌توان از این روش در زمینه آموزش، کاربردی‌شده و می‌توان در آموزش و پرورش و همچنین دانشگاه‌های مختلف برای تحلیل افت تحصیلی ا مورد استفاده قرار گیرد. ترکیب این سه الگوریتم می‌تواند توسط روانشناسان و مشاوران به عنوان یک نرم‌افزار کمکی استفاده‌شود.

مراجع

[1] Piatetsky-Shapiro G., Frawley W. J.,(1991) *"Knowledge Discovery in Databases",* AAAI/MIT Press, 13(3), pp. 57-70.

[۲] ابراهیم‌خانی، سمیه؛ افضلی، مهدی؛ شکوهی، علی(۳۹۰). "پیش‌بینی و بررسی عوامل تصادفات جاده‌ای با استفاده از الگوریتم‌های داده‌کاوی". *فصلنامه دانش انتظامی زنجان*. س ۱، ش ۱، (زمستان): ۱۱۱-۱۲۷.

[3] Fayyad U. M., Piatetsky-Shapiro G., Smyth P., (1996)*"From Data Mining to Knowledge Discovery: An Overview"*, In Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, pp. 1-34.

[4] Li, Xu; Zaian, Osar R.; Li, Zhanhuai (2006) *Advanced Data Mining and Applications*. Berlin: Springer.

[5] Williams, Graha J.; Simoff, Simeon J. (2006*). Data Mining: Theory, Methodology, Techniques, and Applications*. Berlin: Springer.

[6] Bousbia, N., & Belamri, I. (2014). Which Contribution Does EDM Provide to Computer-Based Learning Environments?. In *Educational data mining* (pp. 3-28). Springer International Publishing.

[۷] دورکین، پل(۱۳۷۶). *مشکلات یادگیری و رفتاری دانش آموزان*، ترجمه پرویز صالحی، تهران: نشر مرکز

[8] Suneetha. K.R and Krishnamoorthi. R, 2009 . Identifying User Behavior by Analyzing Web Server Access Log File, International Journal of Computer Science and Network Security Vol. 9 No. 4 pp. 327-332

[9] Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1): 135–146.

[10] Campbell, J.P., Oblinger, D.G. (2007). *Academic Analytics*. Available in: http://net.educase.edu/ir/library/pdf/PUB6101.pdf.

[11] Baker, R. & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1 (1): 3-17.

[12] Bienkowski, M., Feng M. & Means B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, Available in: http://www.cra.org/ccc/files/docs/learning-analytics-ed.pdf.

[13] García, E., Romero, C., Ventura, S. & Castro, C. (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education*, 14 (2): 77-88.

[14] Pechenizkiy, M., Calders, T., Vasilyeva, E. & Bra, P. D. (2008). Mining the student assessment data: Lessons drawn from a small scale case study. *in 1st International Educational Data Mining Conference (EDM2008)*. Montreal Canad, June 20-21 2008.

[۱۵] احمدی، سیدعلی اکبر؛ کریم زادگان، داوود؛ خیراتی کازرونی، تورج. (۱۳۹۴). "داده‌کاوی دانشجویان انصرافی دانشگاه تهران با تمرکز بر حفظ دانشجویان شهریه‌پرداز (جلوگیری از روی-گردانی مشتری)". دو فصلنامه مدیریت فناوری اطلاعات, دوره ۷, شماره ۲(تابستان): ۲۱۷- ۲۳۸.

[۱۶] تمنائی فر، م.ر؛ سلامی محمدآبادی، ف؛ دشتبان زاده، س.(۱۳۹۰) "ارتباط بین پنج رگه شخصیتی و عملکرد تحصیلی دانشجویان"، نخستین همایش ملی شخصیت و زندگی نوین، ایران، سنندج.

[۱۷] اندری، سحر؛ صالحی، پرستو(۱۳۹۰)،. "تحلیل عوامل مؤثر بر افت تحصیلی دانش‌آموزان مقطع راهنمایی با استفاده از داده‌کاوی"، پنجمین کنفرانس داده کاوی ایران، دانشگاه صنعتی امیرکبیر، ایران، تهران.

[۱۸] اسکندریان، پریناز؛ خان‌تیموری، علیرضا؛ فاتحی مرج، علیرضا؛ شنیده، فاطمه. (۱۳۹۰). "بررسی کاهش عملکرد تحصیلی دانش-آموزان جهت بهبود و ارتقای انگیزه تحصیلی بر اساس شبکه‌های بیزین"، پنجمین کنفرانس داده‌کاوی ایران، دانشگاه صنعتی امیرکبیر، ایران، تهران،

[۱۹] قلندری، ح؛ (۱۳۹۱). "رابطه بین ویژگی های شخصیتی و عملکرد تحصیلی دانش آموزان"، نخستین همایش ملی شخصیت و زندگی نوین، ایران، سنندج،

[۲۰] هدایتی، بهرام؛ قاسمیان، محمد حسن. (۱۳۹۲). "بررسی ویژگی های فردی موثر بر عملکرد تحصیلی دانشجویان با استفاده از تکنیکهای داده کاوی". پنجمین کنفرانس علمی فناوری اطلاعات و دانش. شیراز: دانشگاه شیراز.

[۲۱] هدایتی، بهرام؛ جوانمرد، مهدی، (۱۳۹۳). "بررسی عوامل موثر با افت تحصیلی دانش‌آموزان سال اول دبیرستان با استفاده از تکنیک‌های داده‌کاوی، همایش مهندسی کامپیوتر و توسعه پایدار با محوریت شبکه‌های کامپیوتری، مدلسازی و امنیت سیستم ها، مشهد

[۲۲] اسدی ورمله، پرویز؛ احمدی، هادی؛ حسنی پیرمحمدی، حشمت‌الله. (۱۳۹۳). "بررسی علل افت تحصیلی دانش‌آموزان سال اول دبیرستان با استفاده از تکنیک‌های داده‌کاوی"، دومین کنفرانس بین المللی دستاوردهای نوین در علوم مهندسی و پایه، اردبیل

[۲۳] اسماعیل‌پور، منصور؛ شکور، زرینا، نادری‌فر، وحید(۱۳۸۷). "مدرسه هوشمند، افت و پیشرفت تحصیلی، آموزش الکترونیکی ، الگوریتم ژنتیک، داده کاوی"، پنجمین کنفرانس ملی و دومین کنفرانس بین المللی یادگیری و آموزش الکترونی، تهران.

[۲۴] ایرجی، اعظم؛ مینایی، بهروز؛ شکورنیاز ونوس (۱۳۹۲). "بکارگیری فن آوری داده کاوی به منظور آسیب شناسی افت تحصیلی هنرجویان هنرستانی و استخراج نمایه ساز توصیفی در ارائه تمایز دانش آموزان ضعیف و ممتاز " تهران، دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران.

[۲۵] یقینی مسعود؛ اکبری، امین؛ شریفی، سیدمحمدمهدی(۱۳۸۷). "پیش بینی وضعیت تحصیلی دانشجویان با استفاده از تکنیک‌های داده کاوی"، دومین کنفرانس داده کاوی ایران، تهران

[26] Yeh, H.-C, Yang M.-L, Lee L. C.(2007). An Empirical Study of Credit Scoring Model for Credit Card in. Innovative Computing, Information and Control

[27] Chan, C & Lewis, B 2002, 'A basic primer on data mining', *Information Systems Management*, vol. 19, no. 4, pp. 56-60.

پی‌نوشت

1. Association Rule Mining
2. Classification,
3. Clustering
4. Decision Tree
5. Artificial Neural Network
6. Support Vector Machine
7. Romero & Ventura
8. Baker, R. & Yacef
9. Bienkowski
10. Unsupervised Learning
11. Partitioning Clustering