

آسیب شناسی افت تحصیلی دانش آموزان آموزان دبیرستانی شهر زاهدان با استفاده از داده کاوی

محمد رضا پور کیخایی^{۱*}، مجتبی اصغری^۲، منصوره شهرکی نادر^۳، مهشید صدیقی^۴

۱- هنرآموز رشته مهندسی کامپیوتو، هنرستان دکتر علی شریعتی، زاهدان، ایران ، reza.p93@hotmail.com

۲- کارشناس ارشد مهندسی کامپیوتو (نرمافزار)، مدیر نوسازی و تحول اداری شهرداری زاهدان، زاهدان، ایران asghari_258@yahoo.com

۳- کارشناس مهندسی کامپیوتو (نرمافزار)، شرکت گاز استان سیستان و بلوچستان، زاهدان، ایران

۴- کارشناس ارشد مهندسی کامپیوتو (نرمافزار) دانشگاه آزاد اسلامی ، واحد زاهدان ، زاهدان، ایران

چکیده

با توجه به اهمیت بالای آموزش و پرورش در توسعه اقتصادی، فرهنگی و اجتماعی جامعه به ویژه مقطع دبیرستان به لحاظ تاثیر عمده ای که در تربیت نیروی انسانی مورد نیاز جامعه دارد، همواره مورد توجه بوده است. در این راستا با توجه به این که نظام آموزش و پرورش هر ساله داده های آماری زیادی را در ارتباط با اطلاعات جمعیت شناختی دانش آموزان و نمرات درسی آنها جمع آوری و ذخیره می نماید لذا به منظور بهره برداری از دانش نهفته در این حجم اطلاعات و هدایت تصمیم گیری هوشمندانه مدیران آموزش و پرورش در بهبود هدایت تحصیلی دانش آموزان، این تحقیق به تحلیل اطلاعات دانش آموزان با استفاده از تکنیک های داده کاوی می پردازد. هدف از انجام این پژوهش شناسی افت تحصیلی دانش آموزان دبیرستانی شهر زاهدان می باشد. لذا در این تحقیق ابتدا مشخصه های تاثیرگذار بر افت تحصیلی دانش آموزان با بررسی مقالاتی که در این زمینه وجود دارد و مشورت با خبرگان شناسایی گردید. این مشخصه ها شامل: جنسیت، رشته تحصیلی، تحصیلات پدر، منزلت شغلی پدر، شغل مادر، معدل، عملکرد، معدل سال گذشته، استفاده از کلاس های فوق برنامه، پیشرفت تکنولوژی، ضربه روحی شدید، ارتباطات عاطفی و حوادث ناگهانی می باشد. داده های مربوط به ۱۱۲۶ دانش آموز از پایگاه داده های آموزش و پرورش جمع آوری شده است. سپس با استفاده از روش های گسسته سازی، پیش پردازش داده ها و سپس کلاسبندی داده ها انجام گرفت. در مرحله بعد مدل سازی داده ها به کمک درخت تصمیم صورت گرفت. برای این منظور مدل های درخت تصمیم CART، ۵C، QUEST و CHAID به عنوان پرکاربردترین مدل های کلاسبندی استفاده شد. برای مقایسه مدل های استفاده شده دقت کلاسبندی مدلها و ماتریس آشفتگی مورد استفاده قرار گرفت. نتایج حاصل از مدل CART حاکی از دقت بالاتر مدل فوق نسبت به سایر مدلها می باشد. یکی از نتایجی که از درخت تصمیم فوق می توان استخراج کرد رتبه بندی متغیرهای ورودی است که نتایج حاصل از این مدل نشان داد مشخصه تحصیلات پدر بیشترین تاثیر را بر عملکرد دانش آموز داشته است و بعد از آن مشخصه های پیشرفت تکنولوژی و معدل سال گذشته اهمیت بالایی داشته اند.

واژگان کلیدی: افت تحصیلی دانش آموزان، داده کاوی، درخت تصمیم CART، گسسته سازی.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



موسسه آموزش عالی بصیر

۱- مقدمه

کاربرد اصطلاح افت تحصیلی یا اتلاف در آموزش و پرورش، از زبان اقتصاددانان گرفته شده و آموزش و پرورش را به صنعتی تشبیه می کند که بخشی از سرمایه و مواد اولیه را که باید به محصول نهایی تبدیل می شد، تلف کرده و نتیجه مطلوب و مورد انتظار را به بار نیاورده است. البته کاربرد اصطلاح قصور و واماندگی در تحصیل مطلوب است. وقتی صحبت از افت تحصیلی می شود، منظور تکرار پایه تحصیلی در یک دوره و ترک تحصیل پیش از پایان دوره است. به عبارت دیگر افت تحصیلی شامل جنبه های مختلف شکست تحصیلی، چون غیبیت مطلق از مدرسه، ترک تحصیل قبل از موعده مقرر، تکرار پایه تحصیلی، نسبت میان سالهای تحصیلی دانش آموز و سالهای مقرر آموزش و کیفیت نازل تحصیلات می شود. منظور از افت تحصیلی کاهش عملکرد تحصیلی و درسی دانش آموز از سطح رضایتبخش به سطح نامطلوب است. افت تحصیلی یعنی دانش آموز در هنگام تحصیل بعد از یک دوره موفقیت تحصیلی یا متوسط، به تدریج ظرفیت یادگیری او کاهش یافته و یا تکرار پایه تحصیلی داشته باشد. به طورکلی نمرات او نسبت به ماه قبل یا سالهای قبل سیر نزولی محسوس از خود نشان می دهد. افت تحصیلی به عنوان شاخص اتلاف آموزشی، از نظر مفهوم عام به هر گونه ضعف در سوادآموزی و یا وجود بیسوادی گفته می شود. اما از نظر مفهوم خاص عبارت از محقق نشدن انتظارات آموزشی و شکست در اهداف آموزشی است و نا بسامانی در فرآیند یاددهی - یادگیری را در برمی گیرد و نهایتاً ناکامی های تحصیلی به شکل گریز از مدرسه، تجدیدی، مردودی و ترک تحصیلی به منصه ظهور می رسد. [1]

داده کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگو شناسی و یادگیری ماشین می باشد. این فرآیند پیچیده جهت شناسایی الگوها و مدل های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده می باشد؛ به طریقی که این الگوها و مدل ها برای انسان قابل درک باشند. داده کاوی به صورت یک محصول قابل خریداری نمی باشد؛ بلکه یک رشته علمی و فرآیندی است که باایستی به صورت یک پروژه پیاده سازی شود. [2]

در این راستا با توجه به این که نظام آموزش و پرورش هر ساله داده های آماری زیادی را در ارتباط با اطلاعات جمعیت شناختی دانش آموزان و نمرات درسی آنها جمع آوری و ذخیره می نماید لذا به منظور بهره برداری از دانش نهفته در این حجم اطلاعات و هدایت تصمیم گیری هوشمندانه مدیران آموزش و پرورش در بهبود هدایت تحصیلی دانش آموزان، این تحقیق به تحلیل اطلاعات دانش آموزان با استفاده از تکنیک های داده کاوی می پردازد. جهت دست یابی به این هدف از تکنیک دسته بندی که یکی از تکنیک های داده کاوی می باشد، استفاده می شود. دسته بندی فرآیند یافتن مدلی است که با تشخیص دسته ها و یا مفاهیم داده میتواند دسته ناشناخته اشیاء دیگر را پیش‌بینی کند. دسته بندی یک تابع یادگیرنده است که یک قلم داده را به یکی از دسته های از قبل تعریف شده نگاشت میکند.

۲- پیشینه و ادبیات تحقیق

امروزه با گسترش سیستم های پایگاه داده و حجم بالای داده های ذخیره شده در این سیستم ها، نیاز به ابزاری است تا بتوان داده های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرار داد. داده کاوی یکی از مهمترین این روش هاست که به وسیله آن الگوهای مفید در داده ها با حداقل دخالت کاربران شناخته میشوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمان ها اتخاذ شوند. [3]

دسته بندی فرآیند یافتن مدلی است که با تشخیص دسته ها و یا مفاهیم داده می تواند دسته ناشناخته اشیاء دیگر را پیش بینی کند. دسته بندی یک تابع یادگیرنده است که یک قلم داده را به یکی از دسته های از قبل تعریف شده نگاشت می کند. داده های موجود به دو قسمت آموزش و آزمون تقسیم می شوند. داده های آموزش برای یادگیری قواعد توسط



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



موسسه آموزش عالی بصیر

سیستم استفاده می شوند و داده های آزمون برای بررسی دقت دسته بندی و جلوگیری از بیش برازش به کار می روند.
[3]

یکی از مهمترین انواع دسته بندیها شیوه درخت تصمیم است که با توجه به تفسیر پذیر بودن نتایج آن بسیار مورد استفاده قرار گرفته است. درخت تصمیمگیری یک ساختار درختی شبیه فلوچارت است که هر گره داخلی نسبتی بر روی هر ویژگی مشخص میکند و هر شاخه دستاورد تست را نشان میدهد. در این حالت هر گره داخلی نمایانگر یکی از متغیرهای ورودی است که از آن شاخهها با مقادیر ممکن برای مقادیر ورودی ایجاد میشوند. بالاترین گره در درخت، ریشه نام دارد. هر برگ نیز نمایشگر مقدار متغیر هدف به ازای متغیرهای ورودی است که در مسیر ریشه تا برگ مدنظر اشاره شده اند [11].

ساختار درخت تصمیم در یادگیری ماشین، یک مدل پیش بینی کننده می باشد که حقایق مشاهده شده در مورد یک پدیده را به استنتاج هایی در مورد مقدار هدف آن پدیده نقش می کند. تکنیک یادگیری ماشین برای استنتاج یک درخت تصمیم از داده ها، یادگیری درخت تصمیم نامیده می شود که یکی از رایج ترین روش های داده کاوی است. هر گره داخلی متناظر یک متغیر و هر کمان به یک فرزند، نمایانگر یک مقدار ممکن برای آن متغیر است. یک گره برگ، با داشتن مقادیر متغیرها که با مسیری از ریشه درخت تا آن گره برگ بازنمایی می شود، مقدار پیش بینی شده متغیر هدف را نشان می دهد. یک درخت تصمیم ساختاری را نشان می دهد که برگ ها نشان دهنده دسته بندی و شاخه ها ترکیبات فصلی صفاتی که منتج به این دسته بندی ها را بازنمایی می کنند. یادگیری یک درخت می تواند با تفکیک کردن یک مجموعه منبع به زیرمجموعه هایی برآسانس یک تست مقدار صفت انجام شود. این فرآیند به شکل بازگشتی در هر زیرمجموعه حاصل از تفکیک تکرار می شود. عمل بازگشت زمانی کامل می شود که تفکیک بیشتر سودمند نباشد یا بتوان یک دسته بندی را به همه نمونه های موجود در زیرمجموعه بحسب آمده اعمال کرد. [5]

در درخت تصمیمگیری یک سری سوال وجود دارد و با مشخص شدن پاسخ هر سوال یک سوال دیگر پرسیده می شود. این سوالات گرهها را تشکیل میدهند و به این صورت است که یک گره در بالا کشیده شده و برگهای آن در پایین میباشند. حال یک رکورد در گره ریشه وارد میشود و در این گره یک تست صورت میگیرد تا معلوم شود که این رکورد به کدام از گرههای فرزند شاخه پایینتر میرود. عموماً روش های مختلفی برای انتخاب این آزمون اولیه وجود دارد ولی هدف همه آنها یکی است یعنی انتخاب روشی که بهترین جداسازی را در کلاس های هدف انجام دهد. این فرآیند آنقدر ادامه پیدا می کند تا رکورد به گره برگ برسد. تمام رکوردهایی که به یک برگ از درخت می رست در یک کلاس قرار می گیرند. هچنین برای رسیدن از ریشه به برگ تنها یک راه وجود دارد و آن راه در واقع بیان قانونی است که برای دسته بندی رکوردهای ایجاد شده است. پس از آموزش درخت و ایجاد آن با داده های آزمایشی، اثربخشی درخت ایجاد شده باید اندازه گیری شود. برای این کار از یک مجموعه رکورد یا داده های آزمایشی استفاده میشود که متفاوت از داده های اولیه ایجاد کننده درخت میباشد. معیاری که اندازه گیری میشود عبارت است از «درصد داده هایی که درست دسته بندی می شوند و دسته پیش بینی شده با دسته واقعی آنها یکسان است». [6]

درخت تصمیمگیری یکی از ابزارهای قوی و متداول برای دسته بندی و پیش بینی میباشد. درخت تصمیمگیری بر خلاف شبکه های عصبی به تولید قانون میر达زد. یعنی درخت تصمیمگیری پیش بینی خود را در قالب یک سری قوانین توضیح میدهد. در حالیکه در شبکه های عصبی تنها پیش بینی بیان میشود و چگونگی آن در خود شبکه پنهان باقی می ماند. همچنین در درخت تصمیمگیری برخلاف شبکه های عصبی لزومی ندارد که داده ها لزوماً بصورت عددی باشند.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



موسسه آموزش عالی بصیر

سیری^{*} (۲۰۱۵) در تحقیقی به پیش بینی ترک تحصیل دانشجویان در دانشگاه با استفاده از شبکه های عصبی مصنوعی پرداخته است. در این پروژه روند انتقال دانشجو به دانشگاه وهمچنین کاهش نارسانی دانشجویی و بهبود فرآیندهای آموزشی با کمک تکنیک های داده کاوی، به ویژه شبکه های عصبی مصنوعی بررسی شده است. نتایج نشان داد که شبکه های عصبی مصنوعی به خوبی توانسته است دلایل ترک تحصیل و میزان اهمیت آنها در ترک تحصیل دانشجویان را پیش بینی نماید. [9]

اوزک[†] و همکاران (۲۰۱۴) در تحقیقی علت ترک تحصیل دانشجویان با استفاده از روش های داده کاوی در یک برنامه آموزش آنلاین را بررسی کرده اند. داده های جمع آوری مربوط به ۱۸۹ دانش آموز و شامل ۱۰ متغیر، که شامل: جنس، سن، سطح آموزشی، تجربه قبلی آنلاین، شغل، خودکارآمدی، آمادگی، دانش قبلی، منبع کنترل و وضعیت ترک تحصیل علی می باشد. نتایج نشان داد که تجربه قبلی آنلاین، خودکارآمدی و آمادگی یادگیری آنلاین به عنوان مهم ترین عوامل در پیش بینی ترک تحصیل می باشد. [10]

اندری و صالحی (۱۳۹۰) در یک تحقیق با استفاده از تکنیک های قوانین همبستگی و درخت تصمیم اطلاعات ۵۹۹ دانش آموز مقطع راهنمایی را که از طریق پرسشنامه جمع آوری شده بود را تحلیل کرده اند و عوامل موثر در بروز افت تحصیلی را معرفی نموده اند. به عنوان مثال اگر عزت نفس دانش آموز بالا، جنسیت او پسر و وضعیت اقتصادی آن دانش آموز ضعیف باشد، آنگاه برای او افت تحصیلی وجود دارد. [12]

قلندری (۱۳۹۱) رابطه‌ی بین ویژگی‌های شخصیتی و عملکرد تحصیلی دانش آموزان سال سوم دبیرستان شهر زنجان را مورد بررسی قرار داد و به این نتیجه دست یافت که بین ویژگی‌های شخصیت (مسؤولیت پذیری، انعطاف پذیری و برونگرایی) و عملکرد تحصیلی دانش آموزان رابطه‌ی مثبت معناداری وجود دارد و ویژگی‌های شخصیت می تواند درصد عملکرد تحصیلی را تبیین نماید. [13]

هدایتی و قاسمیان (۱۳۹۲) با تحلیل اطلاعات پیش از ۲۵۹ نفر از دانشجویان و با بهره گیری از روش های داده کاوی، عدم انتخاب رشته براساس علاقه‌ی فردی را به عنوان مهمترین عامل در قرار گرفتن دانشجویان در خوش‌یاری شناسایی کرده اند. [13]

دادفر و همکاران (۱۳۹۲) به بررسی عوامل خانوادگی موثر بر افت تحصیلی دانش آموزان دختر و پسر سال اول دیلویستان شهرستان دره شهر پرداخته اند. بر اساس نتایج به دست آمده بین عوامل خانوادگی با افت تحصیلی فرزندان در دو گروه دختر و پسر رابطه معناداری وجود دارد. عوامل خانوادگی بررسی شده شامل پایگاه اجتماعی -اقتصادی، سطح تحصیلات در خانواده، داشتن امکانات و نابه سامانی و مشکلات در خانواده می باشد که با افت تحصیلی رابطه معنادار دارند. [13]



* Siri
† Ozekes



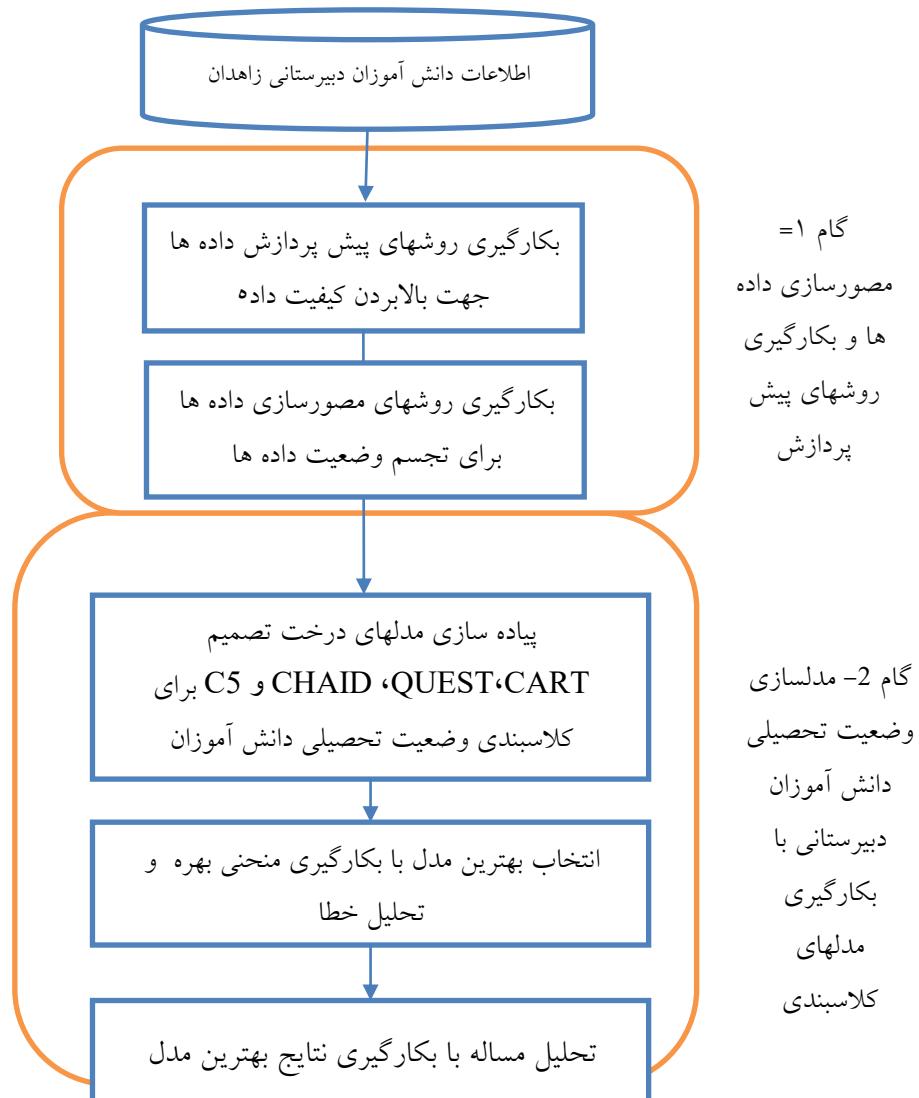
مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



۳- روش ارائه شده

در این تحقیق با توجه به اینکه محقق قصد دستکاری متغیرها را ندارد، روش تحقیق مبنی بر گستته سازی و کلاس بندی است. در بخش مطالعات استنادی، برای شناخت چهار چوب مساله و با هدف تدوین مبانی نظری و مفاهیم اساسی موضوع پژوهش، از روش کتابخانه ای استفاده شده است. در این روش یافته های سایر پژوهش ها از طریق مطالعه منابع داخلی و خارجی در دسترس، جمع آوری و مورد بررسی قرار گرفته است. ابزار اصلی گردآوری اطلاعات بانک اطلاعات آموزش و پرورش است. جامعه آماری این تحقیق دانش آموزان دبیرستان شهر زاهدان می باشد. نمونه انتخاب شده شامل ۱۱۲۶ دانش آموز در شهر زاهدان بوده که اطلاعات مورد نیاز برای هر دانش آموز شامل: جنسیت، رشته، تحصیلات پدر، منزلت شغلی پدر، شغل مادر، معدل، عملکرد، معدل سال گذشته، استفاده از کلاس های فوق برنامه، پیشرفت تکنولوژی، ضربه روحی شدید، ارتباطات عاطفی و حوادث ناگهانی می باشند. شکل ۱ مدل ارائه شده در این تحقیق به همراه گام های پیاده سازی آن را نشان می دهد.



شکل ۱: مدل ارائه شده در این تحقیق

دومین کنفرانس ملی پژوهش های نوین در مهندسی برق و کامپیوتر



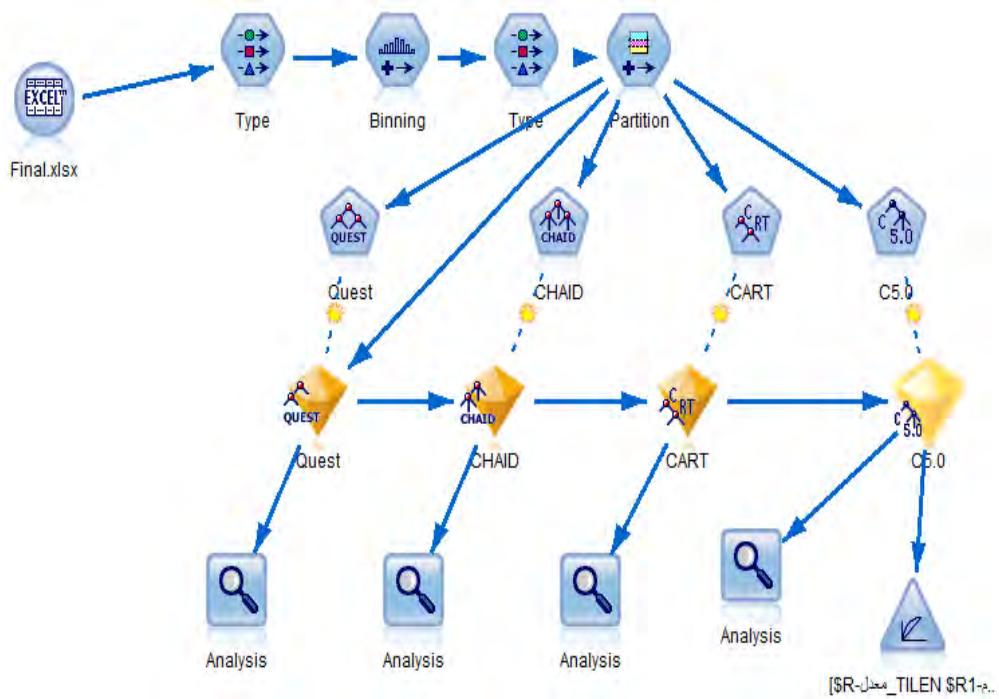
تهران - شهریور ۱۳۹۶

موسسه آموزش عالی بصیر

۴- نتایج تجزیه و تحلیل داده ها

۴-۱- پیاده سازی مدل های کلاسیبندی

در این گام با بکارگیری اطلاعات دانش آموزان (جنسیت، رشته، تحصیلات پدر، منزلت شغلی پدر، شغل مادر، معدل سال قبل، استفاده از کلاسها فوچ برنامه، پیشرفت تکنولوژی، ضربه روحی شدید، ارتباطات عاطفی، حوادث ناگهانی) به عنوان مشخصه ها ورودی، کلاس عملکردی دانش آموز که مبتنی بر معدل او بدست آمده است (ضعیف، متوسط و قوی) با بکارگیری مدل های مختلف درخت تصمیم شبیه سازی می گردد. مشخصه های ورودی در غالب بردارهای p_1, p_2, \dots, p_n به همراه کلاس عملکردی دانش آموز مشخص است و هدف، بدست آوردن یک مدل است که بطور خودکار به هر یک از دانش آموزان یک برچسب C_i بگمارد. بنابراین سیستم کلاسیبند الگو تابعی مانند $\{C_1, C_2, \dots, C_m\}$ است که بر روی هر الگو تعریف می شود و برای مجموعه های از خصلتهای ایجاد شده کلاس مربوطه را پیش‌بینی می کند. با توجه به وجود مشخصه های کیفی به عنوان متغیرهای ورودی استفاده از مدل های درخت تصمیم موثرتر می باشد. برای این منظور مدل های درخت تصمیم QUEST, C5, CART و CHAID به عنوان پرکاربردترین مدل های کلاسیبندی مورد استفاده قرار می گیرد. برای ارزیابی و مقایسه مدل های درخت تصمیم مورد استفاده ابتدا داده ها را به دو مجموعه مجازی آموزش و آزمون افزای می کنیم. ۸۰ درصد نمونه ها را به عنوان داده آموزش و ۲۰ درصد آن را به عنوان داده آزمون استفاده مینماییم. شکل ۲ نحوه پیاده سازی مدل های درخت تصمیم در نرم افزار داده کاوی IBM SPSS MODELER(ver.14) را نشان می دهد.



شکل ۲: مدلسازی درخت تصمیم با بکارگیری نرم افزار IBM SPSS MODELER(Ver.18)

برای مقایسه مدل های استفاده شده دقت کلاسیبندی مدلها و ماتریس آشفتگی مورد استفاده قرار می گیرد. جدول ۱ ماتریس آشفتگی مربوط به مدل QUEST را نشان می دهد. همانطور که مشخص است دقت تعمیم مدل که مربوط به داده های آزمون است ۸۴.۱۶ درصد بدست آمده است که نشان دهنده قدرت بالای مدل فوق می باشد. همچنین ماتریس آشفتگی نشان دهنده دقت بالای مدل در پیش بینی هر یک از کلاس های عملکردی دانش آموزان است.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



'Partition'	1_Training		2_Testing	
Correct	756	83.54%	186	84.16%
Wrong	149	16.46%	35	15.84%
Total	905		221	

'Partition' = 1_Training	1	2	3
1	222	71	6
2	18	257	44
3	1	9	277
'Partition' = 2_Testing	1	2	3
1	51	13	1
2	2	49	10
3	1	8	86

جدول ۱: نتایج حاصل از مدل درخت تصمیم QUEST

جدول ۲ نتایج مدل CHAID را نشان می دهد. نتایج حاصل از مدل CHAID نشان دهنده عملکرد ضعیف مدل فوق می باشد. ماتریس آشتفتگی نشان می دهد که مدل CHAID نتوانسته است کلاس عملکردی ۱ (دانش آموزان ضعیف) را پیش بینی کند و در این کلاس خطای ۱۰۰ درصد دارد و بنابراین مدل مناسبی نمی باشد.

'Partition'	1_Training		2_Testing	
Correct	519	57.35%	136	61.54%
Wrong	386	42.65%	85	38.46%
Total	905		221	
'Partition' = 1_Training			2	3
1			35	264
2			238	81
3			6	281
'Partition' = 2_Testing			2	3
1			6	59
2			47	14
3			6	89

جدول ۲: نتایج حاصل از مدل درخت تصمیم CHAID

جدول ۳ نتایج مدل CART را نشان می دهد. نتایج حاصل از مدل CART حاکی از دقت بالاتر مدل فوق نسبت به مدل های QUEST و CHAID می باشد و به تفکیک کلاس های عملکردی دانش آموزان نیز مدل CART نتوانسته عملکرد مناسبی داشته باشد.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



'Partition'	1_Training	2_Testing		
Correct	775	85.64%	188	85.07%
Wrong	130	14.36%	33	14.93%
Total	905		221	

'Partition' = 1_Training	1	2	3
1	241	52	6
2	20	269	30
3	2	20	265
'Partition' = 2_Testing	1	2	3
1	53	11	1
2	3	50	8
3	1	9	85

جدول ۳: نتایج حاصل از مدل درخت تصمیم CART

جدول ۴ نتایج مدل C5 را نشان می دهد نتایج حاصل از مدل C5 نشان دهنده عملکرد ضعیف مدل فوق می باشد. ماتریس آشفتگی نشان می دهد که مدل C5 نتوانسته است کلاس عملکردی ۱ (دانش آموzan ضعیف) را پیش بینی کند و در این کلاس خطای ۱۰۰ درصد دارد و بنابراین مدل مناسبی نمی باشد. نتایج حاکی از عملکرد مناسب تر مدل C5 نسبت به مدل‌های دیگر است.

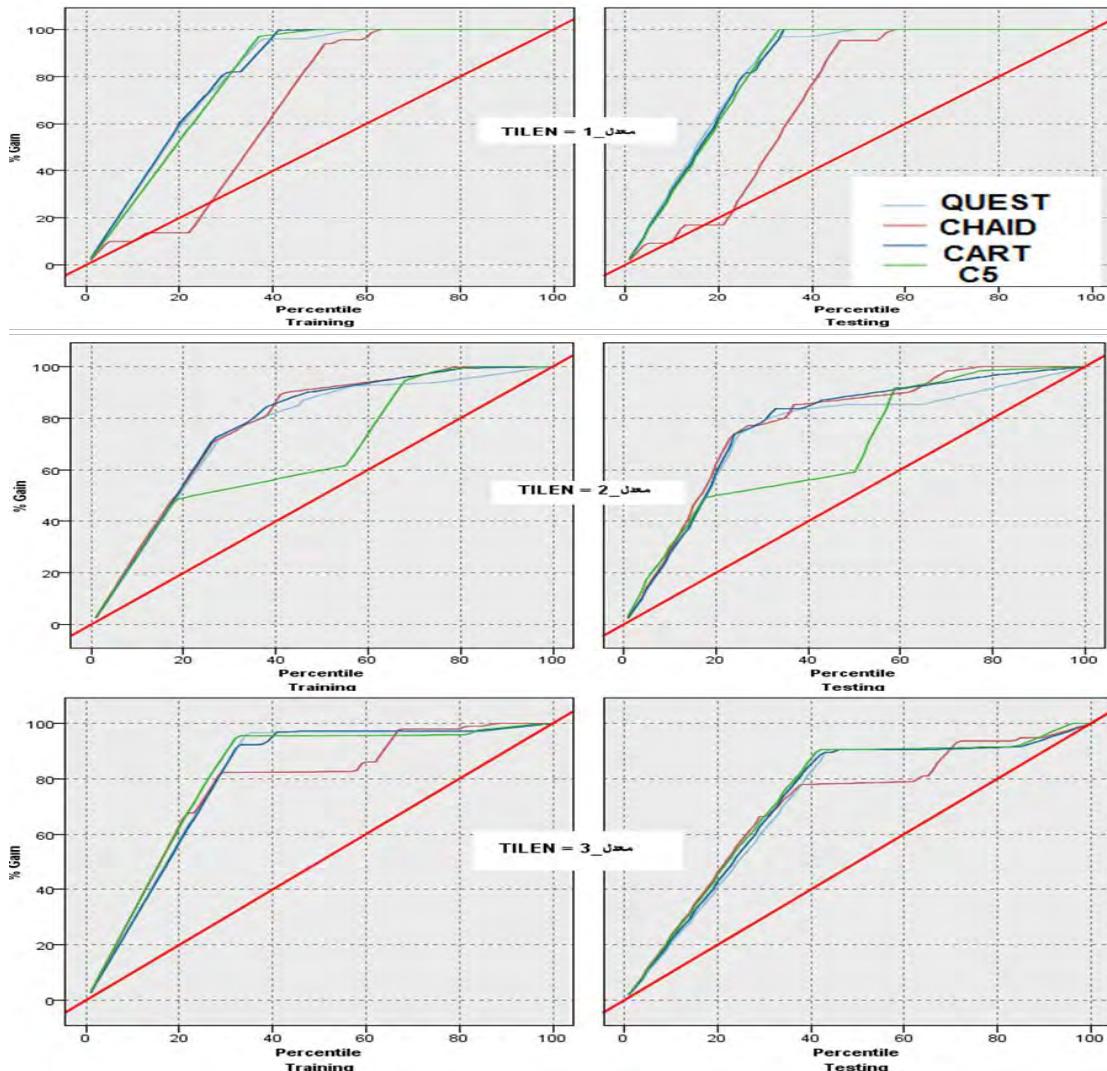
'Partition'	1_Training	2_Testing	
Correct	471	52.04%	122
Wrong	434	47.96%	99
Total	905		221
'Partition' = 1_Training		2	3
1		290	9
2		197	122
3		13	274
'Partition' = 2_Testing		2	3
1		65	0
2		36	25
3		9	86

جدول ۴: نتایج حاصل از مدل درخت تصمیم C5



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



شکل ۳: منحنی بهره برای مدل‌های مورد استفاده

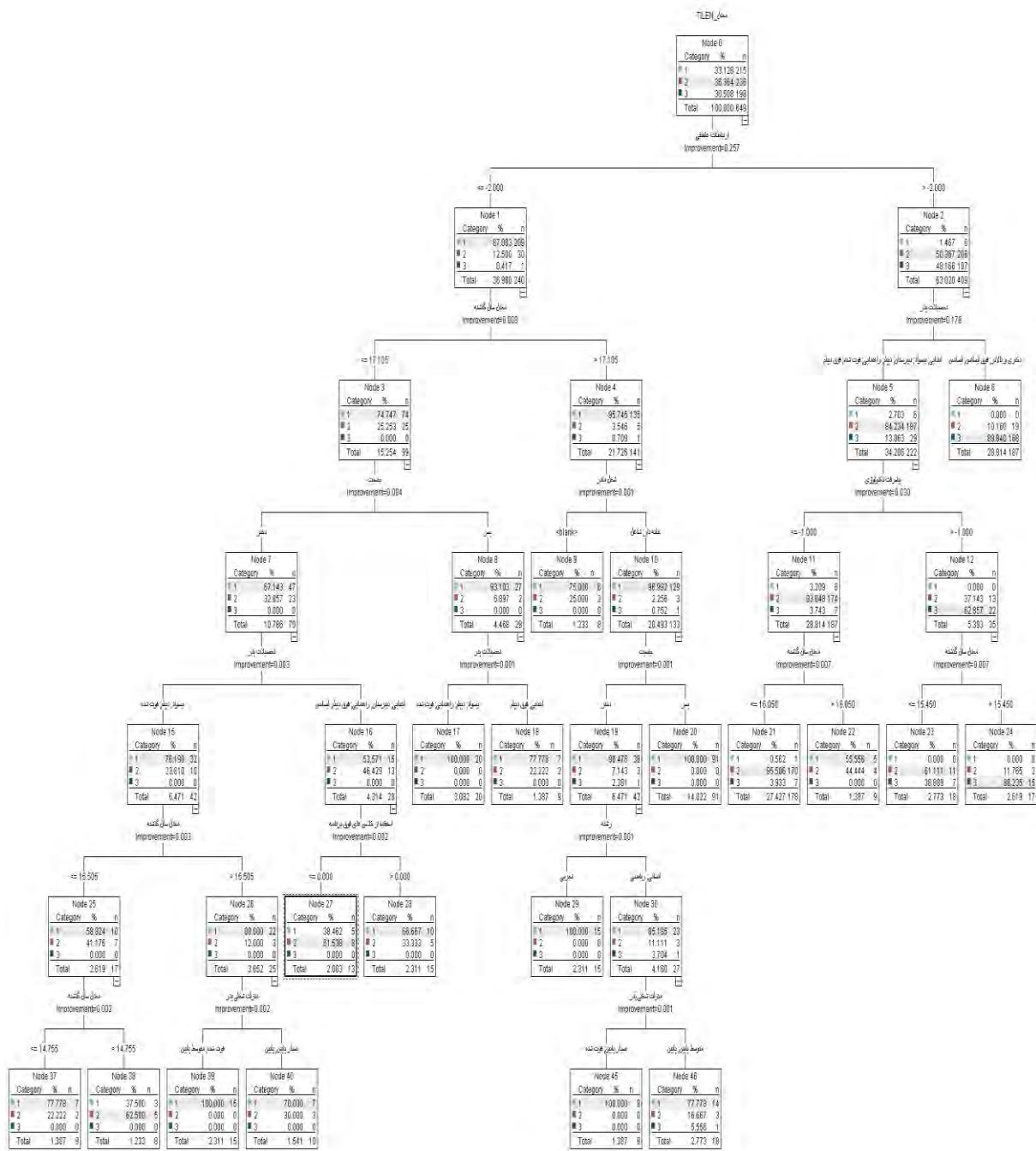
برای ارزیابی بهتر مدل‌های فوق می‌توان از منحنی بهره^{*} استفاده کرد. منحنی فوق برای هر یک از کلاس‌های عملکردی و داده‌های آموزش و آزمون بطور جداگانه رسم می‌شود و هرچه منحنی بهره مدلی بالاتر باشد مدل فوق مدل بهتری است. آنطور که در شکل ۳ مشخص است منحنی بهره مدل‌های CART و CHAID بالاتر قرار گرفته و بنابراین مدل‌های فوق بهتر از دو مدل دیگر است. با توجه به عملکرد بهتر CART از این مدل برای تحلیل‌های تخصصی استفاده می‌شود.

* Gain Curve

مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶

موسسه آموزش عالی بصریر



CART شکل ۴: درخت تصمیم





مهندسی برق و کامپیووتر

تهران - شهریور ۱۳۹۶



شکل ۴: قواعد بدست آمده از درخت تصمیم

قاعده ۱ : نشان می دهد که اگر ارتباطات عاطفی ۲- باشد و معدل سال گذشته بیشتر از ۱۷ و دانش آموز دختر باشد و تحصیلات پدر بی سواد یا دبیلم و یا فوت شده باشد و م معدل سال گذشته از 14.75 بزرگتر باشد آنگاه دانش آموز در سال آتی عملکرد متوسطی خواهد داشت. آنطور که مشخص است ارتباطات عاطفی عاملی بسیار تعیین کننده در عملکرد دانش آموز می باشد.

یکی از نتایجی که از درخت تصمیم فوق می‌توان استخراج کرد رتبه بندی متغیرهای ورودی است بر مبنای اهمیتی که در تفکیک کلاس‌ها دارند، می‌باشد. شکل ۵ رتبه بندی فوق را نشان می‌دهد.



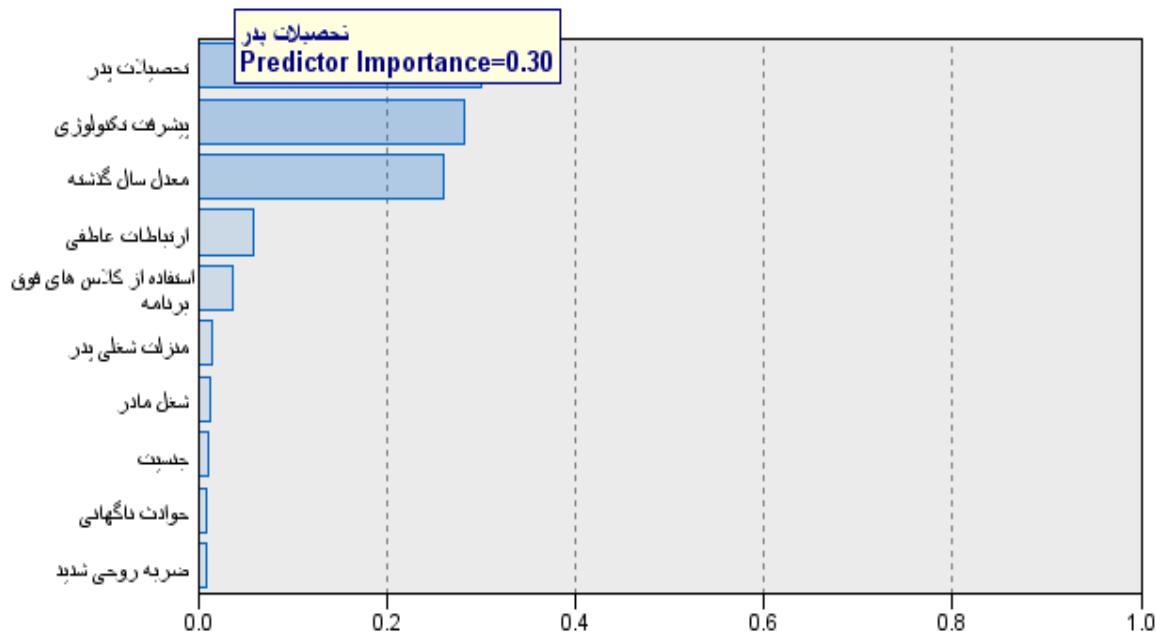
مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



موسسه آموزش عالی بصیر

Predictor Importance

Target: Mعدل_TILEN 

شکل ۵: رتبه بندی مشخصه ها بر مبنای مدل CART

همانطور که مشخص است مشخصه تحصیلات پدر بیشترین تاثیر را بر عملکرد دانش آموز داشته است و بعد از آن مشخصه های پیشرفت تکنولوژی و معدل سال گذشته اهمیت بالایی داشته اند. از طرفی، استفاده از وسائل الکترونیکی نظیر موبایل، لب تاب، تبلت، ماهواره، کنسول های بازی و سایر مواردی که در دسته کلاس های فوق برنامه و ارتباطات عاطفی قرار می گیرند در جهت بهبود کیفیت آموزشی ارائه شده به دانش آموزان می تواند در بهبود عملکرد آنها تاثیر بسزایی داشته باشد.

۵- بحث و نتیجه گیری

با توجه به اهمیت بالای آموزش و پرورش در توسعه اقتصادی، فرهنگی و اجتماعی جامعه به ویژه مقطع دبیرستان به لحاظ تاثیر عمده ای که در تربیت نیروی انسانی مورد نیاز جامعه دارد، همواره مورد توجه بوده است. چالشهای موجود در این مقطع باعث شده که مسئولین آموزش و پرورش در پی تغییر و اصلاح ساختار برنامه های آموزشی این مقطع برآیند. برای دستیابی به این هدف نیاز به تحلیل و توصیف داده های آماری موجود از اهمیت بالایی برخوردار است. در این راستا با توجه به این که نظام آموزش و پرورش هر ساله داده های آماری زیادی را در ارتباط با اطلاعات جمعیت شناختی دانش آموزان و نمرات درسی آنها جمع آوری و ذخیره می نماید لذا به منظور بهره برداری از دانش نهفته در این حجم اطلاعات و هدایت تصمیم گیری هوشمندانه مدیران آموزش و پرورش در بهبود هدایت تحصیلی دانش آموزان، این تحقیق به تحلیل اطلاعات دانش آموزان با استفاده از تکنیک های داده کاوی می پردازد. جهت دست یابی به این هدف از تکنیک دسته بندی که یکی از تکنیک های داده کاوی می باشد، استفاده می شود. داده های مورد استفاده در این تحقیق از پایگاه داده های آموزش و پرورش جمع آوری شده است. سپس با استفاده از روش های گسسته سازی، پیش پردازش داده ها صورت می گیرد و کلاس بندی داده ها انجام می گیرد. در مرحله بعد مدل سازی داده ها به کمک درخت تصمیم صورت می گیرد.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



برای این منظور مدل های درخت تصمیم CHAID، CART، C₅، QUEST و کلاسیبندی مورد استفاده قرار می گیرد. جهت ارزیابی مدل های فوق از منحنی بهره استفاده شده است. نتایج نشان داد که منحنی بهره مدل های CART و CHAID بالاتر قرار گرفته و بنابراین مدل های فوق بهتر از دو مدل دیگر است. با توجه به عملکرد بهتر CART از این مدل برای تحلیل های تخصصی استفاده شده است. یکی از نتایجی که از درخت تصمیم فوق می توان استخراج کرد رتبه بندی متغیرهای ورودی است بر مبنای اهمیتی که در تفکیک کلاس ها دارند، می باشد. همانطور که مشخص است مشخصه تحصیلات پدر بیشترین تاثیر را بر عملکرد دانش آموز داشته است و بعد از آن مشخصه های پیشرفت تکنولوژی و معدل سال گذشته اهمیت بالایی داشته اند.

۶- کارهای آتی

۱. با توجه به پیشرفت تکنولوژی و جمع آوری داده های زیاد در پایگاه داده های آموزش و پرورش، پیشنهاد می گردد تا محققان آتی جهت آسیب شناسی افت تحصیلی دانش آموزان، متغیرهای بیشتری را مورد آزمایش قرار دهند.
۲. پژوهش حاضر بر روی دانش آموزان دیبرستانی انجام گرفته است. از آنجایی که شرایط حاکم پیرامون سایر سازمان ها از قبیل دانشگاه ها و مدارس و ... نیز نسبتا مشابه آن است؛ لازم است تحقیقاتی مشابه این تحقیق بر روی آن ها صورت گیرد.
۳. با توجه به نتایج حاصل از این تحقیق، عواملی چون تحصیلات پدر و پیشرفت تکنولوژی و معدل سال گذشته بیشترین تاثیر را در عملکرد دانش آموزان داشته است، لذا پیشنهاد می گردد در تحقیقات آتی این موارد با حساسیت بیشتری بررسی گرددند.
۴. در این تحقیق جهت بررسی افت تحصیلی دانش آموزان از روش درخت تصمیم گیری استفاده شده است. پیشنهاد می گردد سایر روش های داده کاوی از قبیل شبکه های عصبی مصنوعی، شبکه های بیزین، الگوریتم های خوش بندی و ... نیز بررسی شده و نتیجه آنها با نتایج این تحقیق مقایسه گردد.
۵. محدود شدن جامعه آماری به شهرستان زاهدان از محدودیت های این پژوهش بود، لذا پیشنهاد می شود که این متغیرها در سایر نقاط استان سیستان و بلوچستان و حتی کشور ایران برای تعمیم دهی نتایج بررسی شود.
۶. با توجه به اینکه اطلاعات مورد استفاده در این تحقیق دارای توزیع یکسانی بر اساس تعداد دانش آموزان دختر و پسر نبوده؛ امید است با در دست داشتن اطلاعات کاملی در حوزه های آموزش و پرورش، بتوان نتایج دقیق تری در جهت آسیب شناسی افت تحصیلی دانش آموزان به دست آورد.



مهندسی برق و کامپیوتر

تهران - شهریور ۱۳۹۶



موسسه آموزش عالی بصیر

منابع و مأخذ

- [۱] علیزاده گرجی، خورشید، ۱۳۸۹، رابطه ای افت تحصیلی با افت ریاضی مدرسه ای، یازدهمین کنفرانس آموزش ریاضی ایران.
- [۲] شرابیانی، اشکان، ۱۳۹۰، طراحی یک سیستم هوشمند برای پیش بینی بیماری قلبی با استفاده از تکنیک های داده کاوی، پایان نامه کارشناسی ارشد مهندسی صنایع، دانشگاه صنعتی شریف.
- [۳] علیزاده، سمیه؛ غضنفری، مهدی و تیمورپور، بابک، ۱۳۹۲، داده کاوی و کشف دانش، چاپ سوم، تهران، مرکز انتشارات دانشگاه علم و صنعت ایران.
- [۴] صنیعی آباده، محمد؛ محمودی، سینا و طاهرپور، محدثه، ۱۳۹۱، داده کاوی کاربردی، چاپ اول، تهران، نیاز دانش.
- [۵] اسماعیلی، مهدی، ۱۳۹۳، مفاهیم و تکنیک های داده کاوی، چاپ سوم، دانشگاه صنعتی امیرکبیر، ایران، تهران.
- [۶] شهرابی، جمال، ۱۳۹۰، داده کاوی، چاپ دوم، تهران، جهاد دانشگاهی (دانشگاه صنعتی امیرکبیر).
- [۷] اندری، سحر و صالحی، پرستو، ۱۳۹۰، تحلیل عوامل مؤثر بر افت تحصیلی دانش آموزان مقطع راهنمایی با استفاده از داده کاوی، پنجمین کنفرانس داده کاوی ایران، دانشگاه صنعتی امیرکبیر، ایران، تهران.
- [۸] قلندری، حسین، ۱۳۹۱، رابطه بین ویژگیهای شخصیتی و عملکرد تحصیلی دانش آموزان، نخستین همایش ملی شخصیت و زندگی نوین، ایران، سنترج.
- [۹] هدایتی، بهرام و جوانمرد، مهدی، ۱۳۹۲، بررسی عوامل موثر بر افت تحصیلی دانش آموزان سال اول دبیرستان با استفاده از تکنیک های داده کاوی، مهندسی کامپیوتر و توسعه پایدار با محوریت شبکه های کامپیوتری، مدل سازی و امنیت سیستم ها، موسسه آموزش عالی خاوران، مشهد.
- [۱۰] دادر، رضا و همکاران، ۱۳۹۲، بررسی عوامل خانوادگی موثر بر افت تحصیلی دانش آموزان دختر و پسر سال اول دبیرستان شهرستان دره شهر در سال تحصیلی ۹۰-۹۱، مجله علمی پژوهشی دانشگاه علوم پزشکی، ایلام دوره بیست و یکم، شماره ششم.

- [۱۱] Han, J., Kamber, M., 2006, Chapter 1: Introduction, Data mining concepts and techniques, 2nd edition, Morgan Kaufman.
- [۱۲] Siri, A., 2015, Predicting Students' Dropout at University Using Artificial Neural Networks. Italian Journal of Sociology of Education, 7(2), pp. 225-247.
- [۱۳] Ozekes, s., Yukselturk, E. and Türel. Y.K., 2014, Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program, European Journal of Open, Distance and e-Learning, v 17, pp.118-133.