

روش های داده کاوی برای سیستم مدیریت سطح علمی دانشجویان وموسسات آموزش عالی

۱محمدرضا افراش ، ۲دکتر محمد خدامرادی ، ۳مهرناز فروزان مفرد

۱کارشناس ارشد کامپیوتر - نرم افزار

۲عضو هیات علمی دانشکاه ازاد اسلامی اهواز ،استاد دانشگاه شهید چمران و دانشگاهای خوزستان

۳کارشناس ارشد کامپیوتر - نرم افزار

چکیده

موسسات آموزش عالی اغلب کنجکاووند که بدانند آیا دانشجویان طی دوران تحصیلشان به موفقیت دست می یابند یا نه ، قبل یا حین کلاس های دانشجویان ، موسسات سعی می کنند که درصد موفقیت آنها را تخمین بزنند، اما آیا تخمین زدن میزان موفقیت دانش جویانی که در این کلاس ها شرکت کرده اند امکان پذیر است ؟ آیا ویژگی خاصی از دانشجویان وجود دارد که بتواند با میزان موفقیت آنها در ارتباط باشد ؟ آیا اطلاعات مرتبطی از دانش جویان در اختیار موسسات آموزش عالی است که بتواند بر اساس آنها میزان موفقیت دانشجویان را پیش بینی کنند ، پاسخ به پرسش های فوق به طور کلی می تواند با استفاده از ابزار های داده کاوی بدست آید . متاسفانه الگوریتم داده کاوی بر روی مجموعه داده های در مقیاس بزرگ کارایی بهتری دارد ، در حالی که داده ها آموزشی که در اختیار موسسات آموزش عالی می باشد ، محدود است . در ذیل مجموعه داده های با مقیاس کوچک قرار می گیرد ، بنابراین ، تحقیق بر داده کاوی با مجموعه داده های در مقیاس کوچک تمرکز دارد و هدف آن پاسخ به پرسش های پژوهشی فوق با مقایسه ی دو داده کاوی با ابزار های متفاوت است . نتایج این تحقیق بسیار امیدوار کننده است و موسسات آموزش عالی را تشویق میکند که از ابزاره ای داده کاوی به عنوان بخش مهمی از سیستم مدیریت سطح علمی آموزشی استفاده کنند .

واژه های کلیدی

داده کاوی ، داده کاوی آموزشی ، داده کاوی بر روی مجموعه داده های در مقیاس کوچک ، سیستم مدیریت سطح علمی دانشجویان

معرفی :

در فرایند مدیریت علمی ، فن داده کاوی می تواند از مقدار زیادی داده ، اطلاعات پردازشی و معنا داری استخراج و کشف کند . امروزه در بخش های صنعتی توجه قابل ملاحظه ای به داده کاوی شده است و در جامعه به عنوان یک کل در نظر گرفته می شود . این فن رویکردی هست که در حال حاضر در تجزیه و تحلیل داده ها در کانون توجه می باشد و به عنوان یک ابزار تجزیه و تحلیل تازه ای در حال ظهور ، شناخته شده است . احتمالاً مشاهده کرده اید که در محل ها و موقعیت های زیادی از این روش برای رفع مشکل استفاده شده است ، مثل مشکلات مالی ، پزشکی ، بازاریابی ، بازار سهام ، ارتباطات از راه دور ، کشاورزی ، بهداشت و درمان و ارتباط با مشتری . با وجود اینکه اکثریت موسسات آموزش عالی اذعان کرده اند که میزان موفقیت دانشجویان برای اعتبار مسولان زیر ربط و مدیریت ارشد موسسه ی آموزش حایز اهمیت می باشد ، افراد مرتبط با نظام آموزشی توجه چندانی به نرم افزار های داده کاوی نکرده اند . بنابراین درک عمیق دلایل و پیش بینی درست موفقیت دانشجویان بسیار ارزشمند شده است . راه حل های olap و روش های آماری ابزارهای خوبی برای تجزیه و تحلیل به شمار می آیند اما داده کاوی رویکردی تازه است که به درک الگو های پنهان و پیش بینی داده ها کمک می کند ، و به طور خاص داده کاوی نظر محققان را به خود جلب کرده است چرا که بسیاری از ابزار های مستقل یا دستکناپ داده کاوی در بازار یافت می شود . از مشاهده ی ابزار های مختلف ، می توان نمونه های ذیل را به عنوان مثال در نظر گرفت protage , weka , spss, excel, Microsoft به عنوان سیستم جمع آوری اطلاعات و rapid miner و برخی از آنها (به عنوان مثال mscel mining tool) معمولاً در دسترس استادان موسسات آموزش عالی هستند و آنها می توانند از اطلاعات موجود در اکسل بهره ببرند . و این چرایی افزودن آن به تحقیق است . weka در میان دیگر desktop dm ها انتخاب شد . چرا که در مقایسه با ms excel از تجزیه و تحلیل داده کاوی به روش دیگری پشتیبانی می کند . علاوه بر این weka به عنوان ابزار تجزیه و تحلیل بسیار قانونمند در نظر گرفته شده است . که در آن با آموزش مدیریت داده ها در گیر است . اخیراً موسسات آموزش عالی مدیریت علمی را به منظور ایجاد یک محیط داده کاوی به عنوان ابزاری نون برای ساخت نظام مدیریت علمی ما شناخته شده است . رویکرد مدیریت علمی به فرایند داده کاوی در صورتی که همراه تکنولوژی کسب و

کار شود ، منجر به برخی از همکاری های جدید و ارزش افزوده به راحل های داده کاوی می شود . این تحقیق بر روی سیستم مدیریت علمی موسسات آموزش عالی به عنوان بخشی از آموزش و کار تحقیقاتی تمرکز داشته است .متاسفانه بیشتر تکنیک های داده کاوی بر روی نمونه های با مقیاس بزرگ کارایی بهتری دارد . و به دنبال آن آندونی (Andonie) در سال ۲۰۱۰ گزارش داد که برخی از تکنیک های داده کاوی ، ممکن است مانند شبکه های عصبی قادر به انجام وظیفه ی یادگیری نباشند .(طی مراحل یاد گیری) . چرا که مجموعه داده های با مقیاس کوچک قادر نیستند اطلاعات کافی را برای پر کردن خلاء بین نمونه های بسیار کوچک فراهم کنند . در نتیجه چندین نفر از محققان نتیجه گیری کردند که مجموعه داده های کوچک دامنه تکنیک های داده کاوی را محدود می کنند . دست کم از نظر موسسات آموزش عالی ، در زندگی واقعی موقعیت های آموزشی بسیاری است که در آن مجموعه داده های با مقیاس کوچک امری عادی محسوب می شود . بنابراین داده های که دانشجویان در کلاس های درس جمع اوری می کنند ، یک مثال خوب از این وضعیت است . حتی اگر دانشجویان زیادی در این کلاس ها شرکت کنند ، داده ها به عنوان مجموعه داده ای با مقیاس کوچک تلقی می شوند . آندونی حتی گفت که هیچ راه حل بهینه ی جهانی برای مشکل مجموعه داده های با مقیاس کوچک وجود ندارد . بنابراین چه فرصت هایی در اختیار موسسات آموزش عالی است اگر بخواهند با استفاده از داده کاوی نوین ، میزان موفقیت دانشجویان را در مواد درسی مختلف بدست آورند . داده های دانشجویان در اختیار موسسات آموزش عالی محدود می باشد و ذیل مجموعه داده های در مقیاس کوچک قرار می گیرند . (جدول شماره ۱) ، روش شرح داده شده در این تحقیق بسیار جالب می باشد و به خوبی تحت حمایت ابزار های داده کاوی است . هدیف این تحقیق تاکید بر روی مجموعه داده کاوی های در مقیاس کوچک دانشجویان می باشد که با استفاده از روش های داده کاوی تجزیه و تحلیل می شوند بنابراین تلاش برای پاسخ دادن به پرسش های گوناگون موسسات آموزش عالی ، با استفاده از اطلاعات مواد درسی دانشجویان و ابزار های داده کاوی برای تجزیه و تحلیل (که معمولا در اختیار موسسات آموزش عالی می باشد) صورت می گیرد . برای موسسات آموزش عالی مهم ترین نتیجه ، پیش بینی میزان معتبر موفقیت دانشجویان و آزمایش کاربر می باشد . Bukwifs and willaim در سال ۲۰۰۰ اظهار کردند که فن آوری ، تمایز میان انواع علوم را (علوم ناشناخته نسبت به علوم شناخته شده) را مشخص می کند . تحقیق ایشان از مدل مفهومی به عنوان داده کاوی در مجموعه داده های دانشجویان پشتیبانی می کند . برخی از کلید های تاثیر گذار و نمره نهایی پیش بینی شده در یک درس انتخابی را کشف می کند . علاوه بر این همان طور که محققان اظهار کردند مجموعه داده های در مقیاس کوچک موسسات آموزش عالی می تواند برای تجزیه و تحلیل تکنیک های داده کاوی مورد استفاده قرار می گیرد . در این موارد می توان از الگوریتم های خاص برای فایق آمدن بر داده های که اغلب چنین نیستند استفاده کرد چرا که بیشتر بر روی داده هایی در مقیاس بزرگ کارایی دارد .

سال تحصیلی	دانشجو	جنسیت	سال تولد	شاغل	وضعیت جسمانی (ورزش)	ثبت نام	نوع تحصیل	نمره فعالیت ها	نمره امتحان	نمره نهایی	دسته بندی کلی
2010-2011	1	زن	1988	خیر	خیر	اولین بار	تمام وقت	48	46	92	10
2010-2011	2	مرد	1990	خیر	خیر	اولین بار	تمام وقت	38	33	71	7
2010-2011	3	زن	1990	خیر	خیر	اولین بار	تمام وقت	39	30	69	7
2010-2011	4	زن	1990	خیر	خیر	اولین بار	تمام وقت	47	35	82	8
2010-2011	5	زن	1989	خیر	خیر	اولین بار	تمام وقت	39	36	75	7
2010-2011	6	مرد	1990	خیر	خیر	اولین بار	تمام وقت	38	30	68	7
2010-2011	7	زن	1990	خیر	بله	مجدد	تمام وقت	39	36	75	7
...
2012-2013	106	زن	1990	خیر	بله	مجدد	نیم وقت	44	30	74	8

جدول ۱ - نمونه ای از مجموعه داده های دانشجویان (صفات گرفته شده عبارتند از : سال تحصیل ، نام دانشجو # ، جنسیت ، سال تولد ، شاغل ، وضعیت جسمانی * حضور در رشته ی ورزشی خاص ، ثبت نام * برای اولین بار یا مجدد ، و ...) .

۲- روش و اصول :

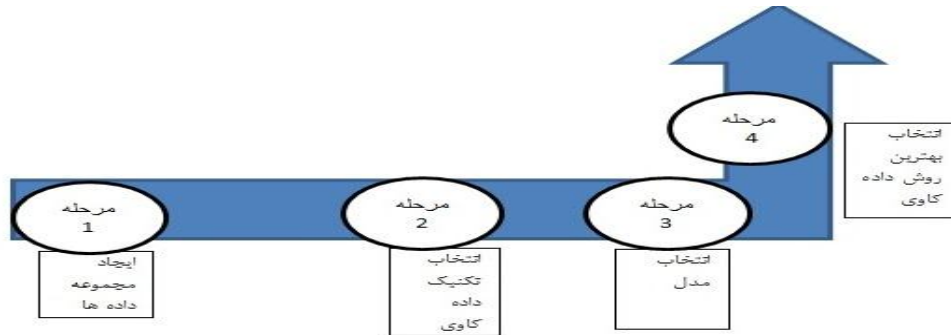
برای ساخت الگوی داده کاوی میزان موفقیت دانشجویان از ابزار ms excel استفاده می شود . چندین محقق مراحل مختلفی برای ساخت الگوی داده کاوی که عملا بر روی فرایند تحول تمرکز دارد ، پیشنهاد کرده اند . turban از الگوی آماده سازی داده ها با ادغام داده ها ، پاک کردن داده ها ، تبدیل داده ها و کاهش مراحل داده ، پشتیبانی کرد . berry and linoff الگوی داده کاوی جامع را به عنوان یک فرایند تکراری حاوی چند مرحله تعریف کرده اند . مدل مورد نظر برای ساخت الگوی داده کاوی میزان موفقیت دانشجویان مرتبط با مدارس بین المللی برای مطالعات تجاری و اجتماعی مورد استفاده قرار گرفت : مرحله ی ساخت شناسایی داده های مورد نیاز است که در آن موسسات آموزش عالی می بایست میزان موفقیت دانشجویان را مربوط به موارد خاصی پیش بینی کنند . اکتساب داده ها بر روی دروس انفورماتیک برنامه های کارشناسی ارشد ، اقتصاد در جامعه نوین در سالهای تحصیلی ۲۰۱۰-۲۰۱۱ (۴۲ دانشجوی) ، ۲۰۱۱-۲۰۱۲ (۳۲ دانشجوی) و ۲۰۱۲-۲۰۱۳ (۳۲ دانشجوی) متمرکز بود . داده ها از نرم افزار وب novis وارد می شوند - سیستم اطلاعات آموزشی - افزودن module به ms excel به همراه داده کاوی و سرور ms sql که بر روی یک لپتاپ نصب شده بود . اعتبار ، کشف ، تمییز دادن ، پس و پیش کردن داده ها جزو فعالیت های بعدی محسوب می شود که در آن ستون ها غیر مرتبط حذف شده اند (به عنوان مثال گروه ، گفته ها ، فعالیت های انفرادی ، شماره ی ثبت ، (عنوان و محتوای ستون) انجام شده است . و ترتیب ستون ها بر اساس (ترتیب منطقی برای پیش بینی موفقیت ، آخرین

ستون به عنوان ویژگی های پیش بینی نمره ی نهایی (تنظیم شده است . برای پیش بینی شفاف تر صورت های مختلف به صورت (بله و خیر) تبدیل شدند . برای جلوگیری از مشکلات داده ها ، نام های دانش آموزان به اعداد تبدیل می شوند . علاوه بر این برای افزایش ارزش پیش بینی اطلاعات ، متغیر های مشتق شده اضافه شد (ستون های اضافی) : پژوهش سالانه بر اساس جنس (بر گرفته از نام دانش جویان) و نوع پژوهش (با ادغام دانش جویان تمام وقت و پاره وقت) می باشد . ساخت الگوی مجموعه داده ها مرحله ی اساسی جهت ساخت الگوی داده کاوی به همراه مجموعه داده های دانش آموزان ذیل می باشد (جدول شماره ۱) سال تحصیلی (۲۰۱۰-۲۰۱۱ و ۲۰۱۱-۲۰۱۲ و ۲۰۱۲-۲۰۱۳) جنسیت (مونث ، مذکر) سال تولد دانشجویان (برای مثال ۱۹۹۸) ، شاغل (بله و خیر) وضعیت جسمانی برای مثال رشته ای ورزشی و غیره (بله و خیر) ثبت نام (برای اولین بار یا مجدد) نوع تحصیل (تمام وقت یا پاره وقت) نمرات فعالیت ها (۵۰-۰) نمرات امتحان (۵۰-۰) ، نمره ی نهایی (۵۰-۰) و نمره ی قابل پیش بینی . مجموعه ی داده ها به دو گروه تقسیم می شوند . اولین گروه مجموعه ی داده ها به عنوان دادهای آموزشی که از کلاس های انفورماتیک در طول سال های ۲۰۱۰، ۲۰۱۱، ۲۰۱۳ گرفته شده است ، تعیین می شود . دومین گروه مجموعه داده ها به عنوان داده های قابل پیش بینی از کلاس های انفورماتیک طی سالهای ۲۰۱۲-۲۰۱۳ ، نمره واقعی و نهایی برای آزمایش نتایج در دسترس بود (۳۲ ردیف از نمره واقعی و نهایی داده ها) . شرح مجموعه داده ها از تعداد اندک نمونه ها صورت می گیرد اما برآورد شد که محتوای آنها مرتبط با فرایند یادگیری ماشینی نیست . در مرحله ی بعدی فن آوری ، فن آوری داده کاوی انتخاب شد که در آن شرکت ماکروسافت به طور کلی سه فرصت از سطح داده کاوی برای تجزیه و تحلیل ارائه داد . سطح پایه شامل ابزار جدول اکسل می باشد که حاوی ویژگی های تجزیه و تحلیل داده کاوی می باشد . سطح متوسط شامل داده کاوی ms excel که ویژگی هایی اضافه می کند ، سطح تخصصی شامل سرور ms sqt ، دارای قابلیت های داده کاوی می باشد . همه ی سطوح داده کاوی با استفاده از الگوریتم هایی از سرور ms sql داده کاوی ، با روابط کاربری متفاوت ، تکنیک های مختلف و تعدادی از پارامتر ها ی برای مدیریت فرایند داده کاوی استفاده می کنند . به منظور اهداف تحقیق سطح اساسی انتخاب شده بعد از این مرحله که در آن تکنیک الگو سازی داده کاوی انتخاب شده ابزار های جدول ms برای ارائه تکنیک های داده کاوی مورد استفاده قرار گرفتند . تجزیه و تحلیل نکات کلیدی و تاثیر گذار ، شناسایی زیر مجموعه ها ، مثال ها پیش بینی و ترجمه کردن استنهاها ، تحلیل سناریو ، پیش بینی محاسبه و تحلیل سبد خرید . بعضی از تجزیه و تحلیل ها برای مجموعه داده کاوی های دانشجویان به دلیل محدود بودن داده یا محتوا مناسب نیست . برای مثال پیش بینی ، تحلیل سناریو و سبد خرید . از میان روش های دیگر قابل استفاده ترین روش کلید های تاثیر گذار و استفاده از مثال هاست که به صورت پویا برای تحقیق های زیر مورد استفاده قرار می گیرد . با توجه به مدل آموزش ، تکنیک استفاده از نمونه ها استفاده شد (برای مثال انتخاب ستون های حاوی نمونه) . بعد از انتخاب ترکیب های مختلف از ستون های محققان با حذف تعدادی از دانش آموزان (که اصلا مرتبط نبودن) و نمرات امتحان (مستقیما به نمره ی نهایی اشاره می کند) ، مرتبط ترین ویژگی ها را کشف کردند . پژوهش به طور پویا از برنامه های ماکروسافت اکسل برای ساخت الگوی داده کاوی و مجموعه داده پژوهی مربوط استفاده می کنند . سر انجام بهترین الگو انتخاب شد و نتایج ارزیابی شدند . با مقایسه کردن نتایج حاصل از پیش بینی برای دانش جویان در طول سال ۲۰۱۲-۲۰۱۳ بر اساس پیش بینی خصوصیات کلاس : نمره نهایی ، محققان با نمره واقعی دانشجویان یکسان ، می توانند بهترین پارامترها را برای تکنیک نمونه یابی داده کاوی استفاده کنند . نمونه یابی داده کاوی نمرات نهایی قابل پیش بینی ، نمرات پایانی دانشجویان محیا بود . تا زمانی که چند فرض واقعی درست بدست آید می توان این مدل را بها اعتماد به نفس به کار برد . فرض اول نشان می دهد که نتایج گذشته برای پیش بینی آینده مثر ثمر تلقی می شوند (چنانچه نسل دانشجویان تغییر کند ، خصوصیات و رفتارشان نسبت به گذشته تغییر می کند که منتج به الگوی مختلف دانشجویی می شود . فرض دیگر این است که داده ها در دسترس باشند . محققان بر این باورند که اطلاعات همیشه در اختیار موسسه آموزش عالی خواهد بود . آخرین فرض این است که داده ها حاوی آن چیزی هستند که محققان درصد پیش بینی آن هستند - پژوهش مدل داده کاوی شامل داده های مربوط به تجزیه و تحلیل داده ها می باشد .

۲.۲ - ساختن مدل داده کاوی دانشجویان با weka

داده ها با استفاده از ابزار داده کاوی weka ساخته شدند (شکل ۱) . به طور کلی اولین گام در کار با ابزار تجزیه تحلیل داده کاوی ایجاد مجموعه داده ها می باشند(مرحله ۱) . در این روش همانطور که در بالا شرح داده شد اطلاعات ذخیره می شوند به جز ستون آخر که به عنوان نمره ی نهایی محسوب می شود . این ویژگی با استفاده از شاخص های زیر به یک عدد ترتیبی تبدیل شد : با ارزیابی ۸ و ۱۰ ، مقداری بالا انتخاب شد و با ارزیابی ۷ و ۸ مقدار متوسط اختصاص یافت و برای ارزیابی کم تر از ۶ مقداری پایین اختصاص یافت (در مورد شاخص تصمیم گیری m5p خصوصیات کلاس همان عددی باقی می ماند) در مرحله ی بعد فن آوری های داده کاوی انتخاب شدند ، ابزار weka ابزاری که بتواند این مرحله را تجزیه و تحلیل کند در نظر گرفته شد (مرحله ی ۲) . به روش خاصی که از weka استفاده می کرد بکار گرفته شده (مرحله ۳) . در این روش سه روش شاخص تصمیم گیری گوناگون ارزیابی شد . j48 , m5p , reptree ، این روش ها به عنوان بهترین و متناسب ترین روش برای تجزیه و تحلیل داده هایی که در آخرین مرحله انجام شده است ، انتخاب شده اند (مرحله ۴) (در صورتی که در m5p تبدیل هیچ مقادیری که در ارتباط با خصوصیات کلاس باشد نیاز نبود و مقادیر نمرات نهایی به عنوان عدد باقی ماند) . در پایان در آخرین مرحله (مرحله ۵) ، داده ها با هر یک از الگوها تجزیه و تحلیل شدند و بهترین آنها انتخاب شد . به عبارت دیگر مدلی را که خصوصیات کلاس را با میزان درستی بالا پیش بینی کرد ، انتخاب کردند ، (هر یک از مدلها بر روی برنامه آموزشی ارزیابی شد و درستی

آنها دوباره آزمایش شد (. از آنجا که اختلاف میان مقدار آموزش پیش بینی شده و داده های واقعی پایین تر بود این مدل به عنوان پیش بینی کننده ی بهتری برای تجزیه و تحلیل داده های واقعی در نظر گرفته شد .



شکل (۱) - فرایند داده کاوی با استفاده از نرم افزار weka در چهار مرحله انجام می شود .

۳- نتایج

۳.۱- نکات کلیدی و تاثیر گذار در نمره ی نهایی دانشجویان

در ابتدای تحقیق محققان به دنبال پیدا کردن داده ها هستند که بیش ترین تاثیر را برای پیش بینی نمره ی نهایی دارا می باشد. تجزیه و تحلیل نکات کلیدی و تاثیر گذار ، ابزاری درست و مفید برای این کار محسوب می شود . (جدول شماره ۲) بیشترین تاثیر نسبی برای متغیرهای مختلف را نشان می دهد . برای مثال آخرین سال تحصیل 2012-2013 که در آن جای نمره ی نهایی خالی است (محققین در صدد پیش بینی بودند) ، نشان دهنده ی ۱۰۰ تاثیر نسبی موافق با ستون های خالی است . نمرات نهایی بالا در ردیف نمران نهایی بالا و نمرات پایین فعالیت کلاسی یا نمرات پایین امتحانی در ردیف نمرات پایانی پایین قرار می گیرند . نتایج نشان می دهد که باید برخی از این ستون ها از پیش بینی نهایی حذف شوند یا دسته کم برای تولید نتایج مربوط ، ترکیب ستون های مختلف را امتحان کنند .

۳.۲- نمونه یابی - پیش بینی نمره نهایی دانشجویان در سال تحصیلی ۲۰۱۲-۲۰۱۳

هدف گام بعدی تحقیق ، پیش بینی نمره ی نهایی داده های از دست رفته ۲۰۱۲-۲۰۱۳ با استفاده از نمونه یابی می باشد . داده کاوی بر اساس ماشین یادگیری ، از نمونه های قبل در سال های تحصیلی 2011-12 استفاده می کند . برای نتایج مرتبط دانشجویان ، نمرات امتحان و نمره ی پایانی که از تجزیه و تحلیل حذف شده اند . گزارش الگوی موجود در جدول شماره ۳ گوناگونی فاکتور های تاثیر گذار را نشان می دهد . نتایج تصمیم محقق را برای حذف نکات کلمات کلیدی و تاثیر گذار از تجزیه و تحلیل توجیه می کند . پیش بینی نمره ی نهایی دانشجویان در سال تحصیلی ۲۰۱۳-۲۰۱۲ بخش پایانی تجزیه و تحلیل داده کاوی با استفاده از روش نمونه یابی داده کاوی است . نتایج ۳۲ دانشجو در جدول شماره ۴ نشان داده شده است . پیش بینی بر اساس داده های علمی پنهان در اطلاعات دانشجویان می باشد . الگوریتم های متفاوتی برای بهره بردن از اطلاعات سال های تحصیلی ۲۰۱۱-۲۰۱۰ و ۲۰۱۲-۲۰۱۱ استفاده شده است که به منظور الگوی پیش بینی داده کاوی صورت می گیرد (اطلاعات آموزشی) و هدف آن پیش بینی نمره ی نهایی دانشجویان (missing student) سال های ۲۰۱۲-۱۳ است . همان طور که ملاحظه می کنید بیشتر نمرات پایانی دانشجویان مثبت می باشد (بالای ۶) چرا که بیشتر نتایج پیشین نیز مثبت بودند .

۳.۳- تغییر داده ها (تحلیل داده ها)

تفسیر به عنوان مرحله ی نهایی اما حیاتی داده کاوی در نظر گرفته می شود . به منظور ارزیابی ارتباط داده کاوی پیش بینی نمره ی نهایی دانشجویان، محقق ، نمره نهایی واقعی دانش جویان را با مقدار پیش بینی شده مقایسه می کند و نتایج زیر را محاسبه و دربارہ ی آن ها بحث شد .

Column	value	favors	Relative impact
Study year	2101-2013		
Final points _100_	92	10	100
Activities points _50_	39.42	7	100
Final points _100_	77	8	100
_Final points _100_	85	9	100
_Exam points _50_	9	2	100
_Final points _100_	53	2	100
_Exam points _50_	13	2	100
_Final points _100_	8	4	100
_Exam points _50_	41	4	100

جدول ۲ - گزارش کلیدی مؤثر برای نمره ی نهایی (جدول شامل مقادیر زیر می باشد ، سال تحصیل ، امکانات ، عوامل مؤثر مرتبط و نمره ی نهایی فعالیت ها)

در فهرست ۳۲ دانشجویان جدول شماره ۵ تنها ۲۰ دانشجو به آزمون نزدیک شده اند (نمره ی پایانی واقعی در سومین ستون) . به کار بردن داده کاوی الگوهای پیش بینی نمرات پایانی همه ی دانش جویان را نزدیک می کند ، که شامل ۱۲ دانشجو(۳۷.۵ درصد) بدون نمره ی پایانی واقعی می باشد . با

گذاشتن عدد صفر، پیش بینی برای ۷ نمره ی پایانی (۲۱.۸ درصد) درست از آب درآمد. با گذاشتن عدد ۱ پیش بینی برای ۱۱ نمره ی پایانی (۴۳.۳ درصد) درست از آب درآمد. پیش بینی مثبت نمره ی نهایی برای ۱۳ دانشجو از ۳۲ دانش جو (۴۰.۶ درصد) موفقیت آمیز بوده و سرانجام اگر ۱۲ دانشجو (بدون نمره نهایی واقعی از داده ها حذف می شدند، پیش بینی نمره ی نهایی مثبت برای ۱۳ نفر از ۲۰ دانشجو موفقیت آمیز بود (۶۵ درصد) که این خود نشان دهنده ی پتانسیل روش داده کاوی برای مجموعه داده های دانش آموزان در مقیاس کوچک که هنوز حایز اهمیت می باشد، است. هر چند که اگر مجموعه داده ها در مقیاس بزرگتری باشند و یا حتی داده های مرتبط بیشتری از دانش جویان در دست رس باشد، نتایج بهتری حاصل می شود، این وقتی است که پارامترها و ستون ها به دقت در تجزیه و تحلیل داد ها ترکیب شوند.

Column	value	favors	Relative impact
Type of study	Part time	10	13
Activities points_50	47,000-43,539	10	13
Registration	Repeat	2	100
Study year	11-2010	4	17
Type of study	Part time	6	33
Gender	Male	6	21
Study year	12-2011	7	43
Activities points_50	38,902-33,000	7	36
Status_sport	بله	7	31
Employment	بله	7	18
Activities points_50	41,221-38,902	8	25
Year of birth	1,983,240-1,967,148	8	21
Activities points_50	47,000-43,539	9	36
Activities points_50	47,000-43,539	9	24
Type of study	Part time	9	10
Employment	No	9	10
Activities points_50	43,539-41,221	9	10

جدول ۳ - تجزیه و تحلیل نمونه ها برای نمره ی نهایی دانشجویان (جدول شامل مقداری از خصوصیات گرفته شده برای تجزیه و تحلیل است مثل ثبت نام (نخستین بار یا مجدد) نمرات فعالیتها ها و غیره

۳.۴ - عوامل موثر در ویژگی های نمرات پایانی دانشجویان با استفاده از ابزار weka

طراحی و آرایه دروس جدید به دانشجویان یکی از وظایف مهم تصمیم گیرندگان آموزش عالی می باشد به خصوص در محیط های پویا و متغیر، که در آن مجموعه داده ها با توجه به دانشجویان محدود می باشد و حاوی اطلاعات قابل توجهی نیست. در این مورد محققان از weka به عنوان ابزار تجزیه و تحلیل داده بدون هیچ مشکلی استفاده کردند. به علاوه محققان در مطالعات خود در طول سال از این ابزار به عنوان یکی از ابزار های پیش بینی موفقیت دانشجویان بهره بردند که در آن ویژگی های نمرات پایانی می تواند به عنوان عامل موثر پیش بینی در کنار دیگر مشخصه ها مورد استفاده قرار گیرد: ویژگی های مثل: نوع تحصیل (تمام وقت یا پاره وقت)، سن، وضعیت اشتغال، رشته تحصیل دانشجویان. این یافته، فقط از چندین ویژگی برای تجزیه و تحلیل استفاده می کند که در آن نمره ی پایانی در کل فرایند پیش بینی نقش کلیدی را بازی می کند و از نظر طراحی برای مفهوم نظام آموزش عالی بسیار جالب می باشد. زمینه لازم را برای پیشنهاد و برنامه بودجه بندی آموزش عالی رو فراهم می کند.

۳.۵ مدل RepTree

مرحله ی آموزش نشان دهنده ی ۹۷.۰۵۸۸ درصد می باشد که در آن ۶۶ مثال به درستی طبقه بندی شده اند و ۲.۹۴۱۲ درصد (۲ مثال) نادرست طبقه بندی شده اند. (شکل ۲). مرحله ی آزمایش نشان داد که ۱۰۰ درصد مثال ها (۲۰ از ۲۰ مثال) به درستی طبقه بندی شده اند.

۳.۶ مدل J48

مرحله ی آموزش نشان داد که ۹۸.۵۲۹۴ درصد (۹۷ مثال) به درستی طبقه بندی شده بودند و ۱.۴۷۰۶ درصد نادرست طبقه بندی شده است.

۳.۷ مدل M5P

مدل M5P نشان دهنده ی ظریب همبستگی ۰.۹۳۵۸ و خطای نسبی مطلق ۳۲.۲۸۸۴ درصد می باشد. ارزیابی نتایج مربوط به جدول شماره ۶ نشان می دهد که ۱۸ مورد از ۲۰ مورد با مدل J48 به درستی طبقه بندی شده اند و از آنجا که از مدل RepTree استفاده شده است همه ی موارد به درستی طبقه بندی شده اند (۲۰ مثال از ۲۰ مثال) بنابراین در هنگام استفاده از داده های با مقیاس کوچک همراه با ساختار الگوریتم داده کاوی (شاخص های تصمیم گیری) می توان به این نکته پی برد که بیش از کل میزان پیش بینی درصد بالایی دارد (بالتر از ۸۰ درصد)

سال تحصیلی	میانگین	حداکثر	سال تولد	شاغل	وضعیت جسمانی (تورم)	دسته نام	نوع تحصیل	نمره حداقل ها	نمره امتحان	نمره تئوری	میانگین نمره کلی
2012-13	75	سرتی	1990	نه	نه	انودن عالی	تمام وقت	38	22	60	7
2012-13	76	مرد	1990	نه	نه	انودن عالی	تمام وقت	38			7
2012-13	77	مرد	1992	نه	نه	انودن عالی	تمام وقت	44			8
2012-13	78	مرد	1992	نه	نه	انودن عالی	تمام وقت	38	30	68	7
2012-13	79	مرد	1990	نه	نه	انودن عالی	تمام وقت	43			8
2012-13	80	مرد	1983	نه	نه	انودن عالی	تمام وقت	43	19	62	8
2012-13	81	سرتی	1991	نه	نه	انودن عالی	تمام وقت	43	49	92	7
2012-13	82	سرتی	1973	نه	نه	انودن عالی	تمام وقت	44	42	86	8
2012-13	83	سرتی	1988	نه	نه	انودن عالی	تمام وقت	42			7
2012-13	84	سرتی	1989	نه	نه	انودن عالی	تمام وقت	38			7
2012-13	85	سرتی	1990	نه	نه	انودن عالی	تمام وقت	38			7
2012-13	86	سرتی	1976	نه	نه	انودن عالی	تمام وقت	44	36	80	8
2012-13	87	سرتی	1987	نه	نه	انودن عالی	تمام وقت	44	32	76	8
2012-13	88	سرتی	1992	نه	نه	انودن عالی	تمام وقت	38	39	77	7
2012-13	89	سرتی	1992	نه	نه	انودن عالی	تمام وقت	43	24	67	7
2012-13	90	سرتی	1992	نه	نه	انودن عالی	تمام وقت	44			9
2012-13	91	مرد	1992	نه	نه	انودن عالی	تمام وقت	38	30	68	7
2012-13	92	مرد	1992	نه	نه	انودن عالی	تمام وقت	38	8	44	7
2012-13	93	مرد	1992	نه	نه	انودن عالی	تمام وقت	38	24	62	7
2012-13	94	سرتی	1992	نه	نه	انودن عالی	تمام وقت	38			7
2012-13	95	مرد	1991	نه	نه	انودن عالی	تمام وقت	43	36	79	8
2012-13	96	مرد	1992	نه	نه	انودن عالی	تمام وقت	38	32	70	8
2012-13	97	سرتی	1990	نه	نه	انودن عالی	تمام وقت	41			7
2012-13	98	سرتی	1991	نه	نه	انودن عالی	تمام وقت	38			7
2012-13	99	سرتی	1968	نه	نه	انودن عالی	تمام وقت	44	38	82	7
2012-13	100	سرتی	1998	نه	نه	انودن عالی	تمام وقت	44	46	90	8
2012-13	101	مرد	1989	نه	نه	انودن عالی	تمام وقت	38	17	55	7
2012-13	102	مرد	1990	نه	نه	انودن عالی	تمام وقت				8
2012-13	103	سرتی	1988	نه	نه	انودن عالی	تمام وقت	38	25	63	8
2012-13	104	سرتی	1999	نه	نه	انودن عالی	تمام وقت	44			9
2012-13	105	سرتی	1988	نه	نه	انودن عالی	تمام وقت	44			9
2012-13	106	سرتی	1990	نه	نه	انودن عالی	تمام وقت	44	30	74	9

جدول ۴ - پیش بینی دسته بندی نهایی برای سال تحصیلی ۲۰۱۳-۲۰۱۲ (جدول شامل صفات نشان داده شده در ستون ها می باشد)

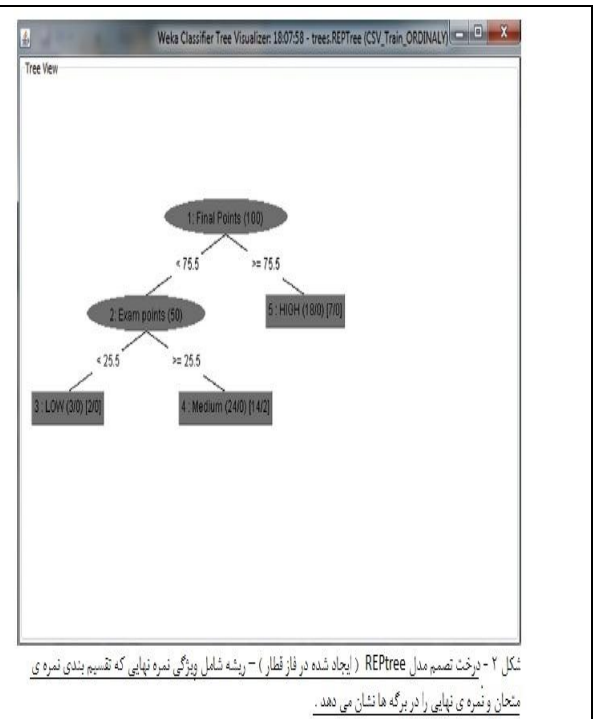
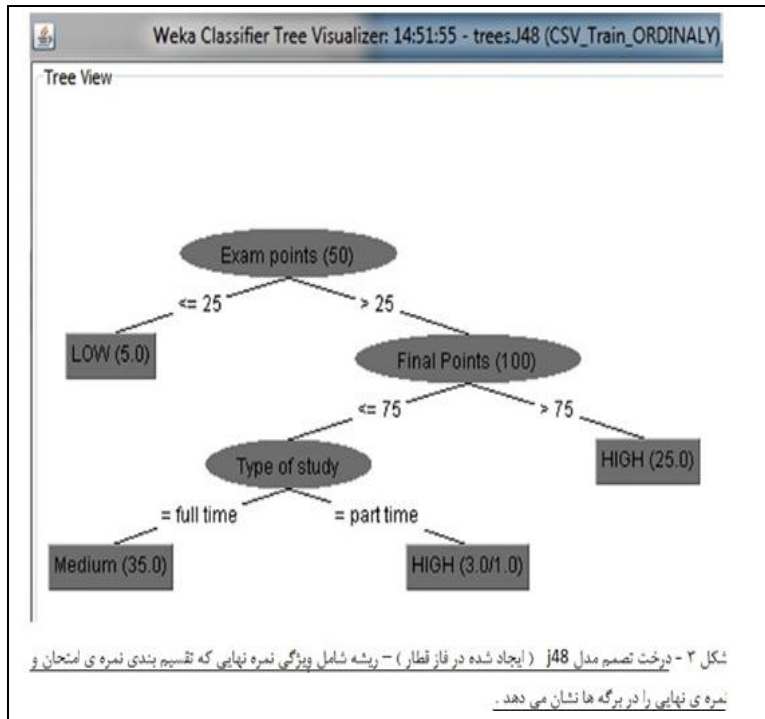
سال تحصیلی	دانشجو	دسته بندی نهایی	مقدار پیش بینی
2012-2013	75	4	7
2012-2013	76		7
2012-2013	77	7	7
2012-2013	78		7
2012-2013	79	7	8
2012-2013	80	3	8
2012-2013	81	110	7
2012-2013	82	9	8
2012-2013	83		7
2012-2013	84		7
2012-2013	85		7
2012-2013	86	8	8
2012-2013	87	8	8
2012-2013	88	8	8
2012-2013	89	4	9
2012-2013	90		9
2012-2013	91	7	7
2012-2013	92	1	7
2012-2013	93	4	7
2012-2013	94		7
2012-2013	95	8	8
2012-2013	96	7	8
2012-2013	97		7
2012-2013	98		7
2012-2013	99	8	8
2012-2013	100	9	8
2012-2013	101	3	8
2012-2013	102		7
2012-2013	103	4	7
2012-2013	104		8
2012-2013	105		9
2012-2013	106	7	9

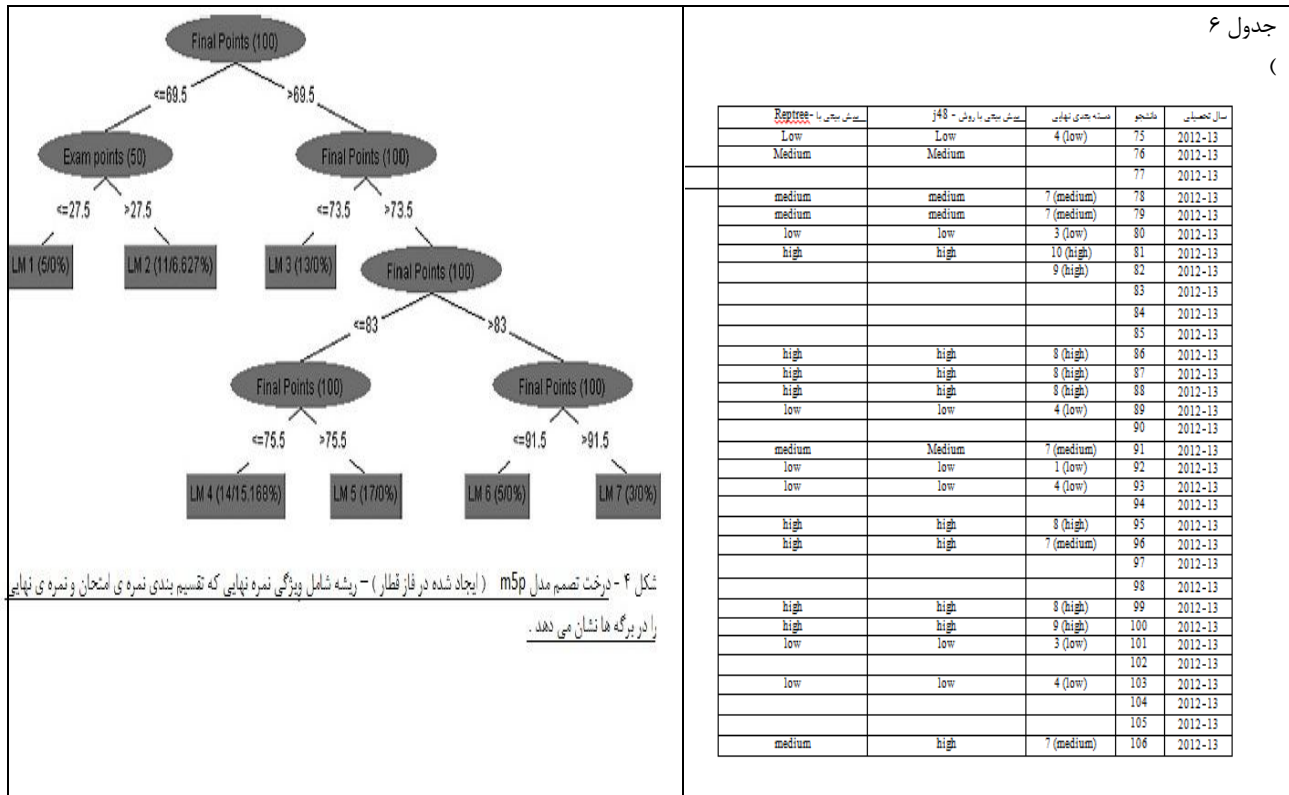
جدول ۵ - مقایسه ی نمره ی نهایی پیش بینی شده ی واقعی

بحث

این تحقیق با استفاده از داده کاوی نوین که در اختیار موسسات آموزش عالی می باشد ، امکان پیش بینی میزان موفقیت دانشجویان ثبت نامی در کلاس های مختلف آموزشی را میسر می کند . تحقیق به طور شفاف نشان می دهد که ابزار های داده کاوی دسکتاپ از نظر قابلیت استفاده و سهولت استفاده به وضعیت مطلوبی رسیده اند و بدون سرمایه گذاری گسترده نتایج کاربردی را فراهم می کند . نتایج تحقیق عملکرد دانشجویان ، در اختیار تصمیم گیرندگان (همانند استادان) آموزش عالی می باشد ، و ویژگی های خاص دانشجویان (به عبارت دیگر اطلاعات دانشجویان) را به طور سری انتقال می دهد که می تواند به خوبی با میزان موفقیت دانشجویان همراه شود . علی رغم اینکه این حقیقت آشکار که الگوریتم داده کاوی بر روی داده ها با مقیاس بزرگ کارایی بهتری دارد ، متمرکز بر روی دانشجویان در مقیاس کوچک موسسات آموزش عالی است و در ذیل مجموعه داده های در مقیاس کوچک قرار می گیرد . نتایج نشان می دهد که مجموعه داده های دانشجویان در مقیاس کوچک در تجزیه و تحلیل خاص داده کاوی ، استفاده از ابزار های داده کاوی را محدود نکرد . علاوه بر این ابزار weka که به عنوان ابزار تجزیه و تحلیل مقایسه ای مورد استفاده قرار گرفت نشان می دهد که با استفاده از الگوی شاخص تصمیم گیری ، درصد بالایی از پیش بینی درست (به خصوص با مدل RepTree) حاصل می شود (درستی هم در حین آزمایش تایید شد) . به علاوه شاخص m5p کارایی خوبی نسبت به دیگر شاخص های تصمیم گیر نداشت ، چرا که به فرضیه های بیشتری نیاز داشت . در واقع RepTree به عنوان شاخص تصمیم گیری سریع دانشجویان در نظر گرفته شد که یکی شاخص تصمیم گیری می سازد و از اطلاعات بدست آمده به عنوان تقسیم معیار استفاده می کند و با استفاده از کاهش خطاها آن را بازسازی می کند . این مدل فقط برای ویژگی های عددی ارزش دارد . کاهش مقادیر و ارتباط با مدل c4.5's که از موارد تجربی استفاده می کند می باشد . با توجه به نکته ی فوق الذکر ، شاید این طور تفسیر شود که چون ویژگی های پیش بینی شده به مقادیری معمولی تبدیل شد ، مدل reptree در مقایسه با مدل j48 حساسیت کمتری نسبت به کاهش مقادیر دارد بنابراین درستی پیش بینی بر روی

مجموعه آزمون ها بهتر جواب داد ، اگر چه بر روی مجموعه آموزشی در مقایسه با j48 (۹۷ درصد تا ۷۰ تا ۹۸ درصد) برای میزان پیش بینی ، کارایی نسبتا کمتری دارد . از این رو ، j48 الگوریتم C4.5 را اندکی اصلاح می کند که یک شاخص تصمیم گیری طبقه بندی شده برای مجموعه اطلاعات داده شده با پارتیشن بندی برگشت پذیر ، تولید کند . الگوریتم تمام آزمون ها ی که بتواند مجموعه داده ها را تقسیم کند و آزمونی را انتخاب کند که حاوی با ارزش ترین اطلاعات گرفته شده باشد . با توجه به اطلاعات ارزیابی شده ، درستی j48 کمتر از reptime بود که می تواند بدیهی شمرده شود ، چرا که در مقایسه با reptime ، حساس تر تلقی می شود .





5- نتیجه گیری و آینده ی کاری

این تحقیق به پرسش های تحقیقاتی پاسخ می دهد و از نتیجه های امید وار کننده حمایت می کند . به این معنی که داده کاوی پتانسیل تبدیل شدن به بخش نظام مدیریت سطح علمی موسسات آموزش عالی را دارا می باشد . بر اساس فرایند مدل (nonoka) از یک چارچوب مبتنی بر دانش، (nonoka toyama & hirata) ، این نکته بسیار حایز اهمیت می باشد که سازمان ها تصمیمات را بر اساس تجزیه و تحلیل داده ها اتخاذ کنند ، نه اینکه متکی بر کارکنان باشند . به عبارت دیگر بعضا کشف علوم پنهان در داده های سازمانها و تبدیل آن ها به علوم شفاف برای ارتقاء فرایند تصمیم گیری بسیار حایز اهمیت می باشد . با توجه به مبحث تحقیق کیفیت فرایند آموزش بسیار مهم تلقی می شود . نتایج این تحقیق بر این موضوع که داده کاوی به طور کلی محدود به مجموعه داده های با مقیاس بزرگ نیست ، صحه می گذارد ، با این همه اکثر محققان در تحقیق های پیشین این یافته را پیش بینی کرده بودند . استفاده خاص از تکنیک های داده کاوی برای ساختار مجموعه داد ها در مقیاس کوچک می تواند نتایج مثری در پی داشته باشد . داده کاوی برای مجموعه داده های دانشجویان در مقیاس کوچک مثالی از تاثیر استفاده تکنولوژی داده کاوی در ارتقاء سطح علمی نظام مدیریتی موسسات آموزش عالی می باشد ، در گذشته تلاش های فراوان برای پیش بینی موفقیت تحصیلی دانشجویان صورت گرفت که موفقیت های گوناگونی هم به همراه داشت . این تحقیق بر روی مجموعه داده های دانشجویان متمرکز بود و نتایج با مقایسه دو تکنولوژی داده کاوی متفاوت (مثل Microsoft and weka) بدست می آمد که هر دو نتیجه ، پیش بینی درستی را نشان می دهد و باتوجه به قدرت این نتیجه گیری ، می توان آنها را ابزاری برای ارتقاء سطح علمی نظام مدیریتی موسسات آموزش عالی به شمار آورد . یک نتیجه گیری مهم دیگر این است که چندین الگوریتم شاخص تصمیم گیری که در داده کاوی مورد استفاده قرار می گیرد ، در واقع برای کار بر روی مجموعه داده های در مقیاس کوچک بسیار علمی (کاربردی) می باشند . این الگوریتم ها امتیاز پیش بینی بالایی دارند . ابزار weka به عنوان منبع آزاد و باز تجزیه و تحلیل داده کاوی در نظر گرفته می شود و ابزار امیدوار کننده ای برای ارزیابی و تجزیه و تحلیل مجموعه داده ای در مقیاس کوچک است و نیز برای تجزیه و تحلیل و تفسیر داده ها به منظور ارتقاء سطح علمی مدیریت سازمان ها مفید است . مثلا یکی از مواردی که در این تحقیق آمده است ، نتایج تحقیق می تواند به مدیریت موسسات آموزش عالی کمک کند تا در مورد مشکلات موفقیت دانشجویان بحث و گفتگو کنند . بعضی از این مشکلات به سادگی قابل حل نیستند چرا که مدل های آماری معمولا برای کشف الگو ها و علوم پنهان به اندازه ی کافی عمیق نیستند ، بیشتر برای فرایند آموزشی و سازمان ها کارایی دارند . با دانستن میزان موفقیت دانشجویان در مدیریت موسسات آموزش عالی می توان به طور موثری با برنامه ریزی استراتژی و سیاست موسسه به مدت طولانی حتی قبل از روند ثبت نام دانش جویان کمک کرد . کمک و همکاری این مطالعه در فرصت های جدید

برای مدیریت موسسات آموزش عالی است و هدف آن پیش بینی اولیه موفقیت دانشجویان و آماده سازی اقدامات برای تشویق دانشجویان به منظور افزایش کارایی آنها است، البته اگر پیش بینی ها نیازهای موسسات آموزش عالی را برآورده نکنند. در سطح نظری، این تحقیق از مجموعه داد های در مقیاس کوچک دفاع می کند و توسط داده کاوی کنونی دسکتاپ و فن آوری پشتیبانی می شود. در سطح پیاده سازی داده های تحقیق و نتایج اثبات کرده است که پیش بینی انجام شده با هر دو روش در تحقیق، بسیار موفقیت آمیز بوده است. علاوه بر این یافته های ما یک راهنمایی عملی برای تصمیم گیرندگان موسسات آموزش عالی و استادان تهیه می کند تا آنها به توانند مشکلات در ارتباط با موفقیت دانشجویان را شناسایی کنند که گاهی در اطلاعات استادان درباره ی دانشجویانشان نهفته است. سر انجام این مدل، این امکان را فراهم می کند که فوراً بعد از ثبت نام یا هر دوره ی دیگری از سال تحصیلی، برای ارتقاء میزان موفقیت دانشجویان دست به کار شوند. همانطور که انتظار داشتیم، این پژوهش می تواند با مقایسه بین المللی از مجموعه داده های دانشجویان گرفته شده از کشورهای مختلف با استفاده از فن آوری های مختلف داده کاوی گسترش یابد. مطالعه ی مقایسه ای، از میزان موفقیت دانش جویان برای اندازه های مختلف از مجموعه داده های مربوط به دانشجویان می تواند تا حد زیادی پتانسیل تحقیق را گسترش دهد و دیدگاه جدید چگونگی استفاده از تکنولوژی داده کاوی را تشکیل دهد و این روش ها (الگوریتم ها) به عنوان یک ابزار مهم برای توسعه سیستم های مدیریت موسسات آموزش عالی است. محققان امید وارند که این نتایج مدیریت موسسات آموزش عالی را تشویق کند که با بکار بردن ابزار های داده کاوی در فعالیت های روزمره میزان موفقیت دانش جویان را ارتقاء دهند.

منابع

- Andonie, R. (2010). Extreme data mining: Inference from small datasets. *International Journal of Computers, Communications & Control*, V(3).
- Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future vision, *Journal of Educational Data Mining* (Vol. 1, Issue 1, pp.3–17).
- Becerra-Fernandez, I., Gonzales, A., & Sabherwal, R. (2004). *Knowledge management, challenges, solutions, and technologies*. Prentice Hall: Pearson.
- Berry, M., & Linoff, G. (2000). *Mastering data mining. The art and science of customer relationship management*. Wiley.
- Jashapara, A. (2011). *Knowledge management, an integrated approach* (2nd ed.). Prentice Hall.
- Marjetić, D., & Lesjak, D. (2012). Financing of higher education and the role and dilemmas of tariff groups. *International Journal of Management in Education*, 6(1/2), 56–72.
- Natek, S., & Lesjak, D. (2013). Improving knowledge management by integrating HEI process and data models. *The Journal of Computer Information Systems*, 53(4), 81–86.
- Negnevitsky, M. (2011). *Artificial intelligence, a guide to intelligent systems* (3rd ed.). Pearson Education Limited.
- Nonaka, I., Toyama, R., & Hirata, T. (2008). *Managing flow, a process theory of knowledge-based firm*. Palgrave Macmillan.
- Osei-Bryson, K.-M. (2010). Towards supporting expert evaluation of clustering results using a data mining process model. *Information Sciences*, 180(3), 414–431.
- Rojko, K., Lesjak, D., & Vehovar, V. (2011). Information communication technology spending in (2008-) economic crisis. *Industrial Management and Data Systems*, 111(3), 391–409.
- Romero, C., & Ventura, S. (2006). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
- Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support System*, 46(1), 287–299.
- Trunk Širca, N., Babnik, K., & Breznik, K. (2013). Towards organisational performance: Understanding human resource management climate. *Industrial Management and Data Systems*, 113(3), 367–384.
- Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32, 1761–1768.
- Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligent systems* (9th ed.). Pearson.
- Wan, S., & Lei, T. C. (2009). A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan. *Knowledge-Based System*, 22(8), 580–588.
- Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management and Data Systems*, 108(5).
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., et al. (2008). Top

10 algorithms in data mining. Knowledge Information Systems, 14, 1–37.

Yuan, J. L., & Fine, T. (1998). Neural-network design for small training sets of high dimension. IEEE Transactions on Neural Networks, 9.

Zanakis, S. H., & Fernandez, I. B. (2005). Competitiveness of nations: A knowledge discovery examination. European Journal of Operational Research, 166(1),185–211.

Zhuang, Z. Y., Churilov, L., Burstein, F., & Sikaris, K. (2009). Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. European Journal of Operational Research, 195, 662–675.