(شناسایی عوامل تاثیرگذار بر نتیجه ارزیابی ساختمان های آموزشی اسکلتی با استفاده از تکنیک های داده کاوی)

جلال صفیاری، فوق لیسانس مهندسی کامپیوتر، موسسه غیر انتفاعی ماد سنندج
دکتر پرویز رشیدیان، استادیار علوم کامپیوتر، موسسه غیر انتفاعی ماد سنندج
دکتر امیر شیخ احمدی، استادیار نرم افزار کامپیوتر، دانشگاه آزاد اسلامی واحد سنندج

Archive of SID

چکیده

وزارت آموزش پرورش کشور ایران در حال حاضر دارای حدود 286123 ساختمان با مساحت 66764310 متر مربع با کاربری آموزشی،غیرآموزشی وجانبی می باشد. که از این میان تعداد 96786مورد از آنها با مساحت 48709430 متر مربع زیربنا، آموزشی می باشند و از نظر کیفی دارای سه ویژگی مستحکم، تخریبی و نیاز مند مقاوم سازی هستند. برای احداث مدارس جدید، مقاوم سازی وتجهیز یا توسعه مدارس موجود سالانه بودجه محدودی در نظر گرفته می شود. در این تحقیق سعی شده است عوامل تاثیرگذار بر نتیجه ارزیابی ساختمان های اسکلتی بررسی شود و عوامل تاثیر گذار بر نتیجه ارزیابی مدارس، شناسایی شود که برای این منظور از الگوریتم های داده کاوی استفاده شده و آنها را با هم مقایسه کرده ایم و بهترین الگوریتم و نتیجه را کشف کرده تا در تصمیمات آتی سعی شود این عوامل مد نظر قرار گیرد تا مدارسی با قدمت بالا و آسیب پذیری کمتر داشته باشیم.

کلیدواژه‌ها: نتیجه ارزیابی، اسکلت، ساختمان آموزشی، مقاوم سازی، تخریبی، داده کاوی.

1- مقدمه و بیان مسئله

به طور کلی استفاده همگانی از وب و شبکه جهانی به عنوان یک سیستم اطلاع رسانی جهانی مارا با حجم زیادی از داده ها مواجه می کند. این رشد انفجاری داده ها نیاز به یک سری تکنیک ها و ابزارهای جدید که توانایی پردازش هوشمندانه اطلاعات را دارا باشند نمایان می سازد. داده کاوی با پیدا کردن مجموعه ای از الگوهای جالب از دل داده های موجود در انباره داده ها می تواند چنین نیازی را برطرف کند. محققان در بسیاری از رشته ها نظیر پایگاه داده ها، یادگیری ماشین و آمار این موضوع را پیگیری کرده و تکنیک های مختلفی را برای داده کاوی معرفی کرده و آنها را طبقه بندی

www.SID.ir

می کنند. داده کاوی یکی از مهم ترین مراحل فرایند استخراج دانش در پایگاه داده به حساب می آید[1]. در این تحقیق هدف شناسایی عوامل تاثیر گذار بر نتیجه ارزیابی مدارس می باشد و ایجاد مدلی برای ساخت مدارس با بهترین نوع مصالح که دیر تر تخریب می شوند و پیش بینی اینکه یک ساختمان آموزشی ارزیابی نشده در کدام دسته می تواند قرار گیرد. بزرگترین محدودیت جاری نبود تعدادی فیلد مورد نیاز به تفکیک مدارس از جمله تعداد دانش آموز به روز موجود در مدارس، فاصله مدارس با گسل زلزله و ... می باشد که بدست آوردن آنها بسیار هزینه بردار و کاری طاقت فرسا است. اما این مجموعه داده های خام به خودی خود ارزشی ندارند، برای معنی بخشیدن به این داده ها باید آنها را تحلیل و تبدیل به اطلاعات یا بهتر از آنها دانش کرد[1]. با استفاده از تکنیک های داده کاوی سعی می شود الگو و فلوچارتی برای ارزیابی ساختمان های آموزشی کشف شود که ویژگی های یک ساختمان مستحکم را شناسایی کرده و پیشنهاد می دهد که ساختمان مستحکم باید دارای چه ویژگی هایی باشد. در مرحله دوم نیز ویژگی های یک ساختمان تخریبی و مرحله سوم ویژگی های یک ساختمان که نیازمند مقاوم سازی است را کشف می کند

روش های بسیار زیادی از جمله روش های آماری و تکنیک های داده کاوی، جهت شناسایی عوامل تاثیر گذار بر نتیجه ارزیابی استحکام ساختمان های آموزش پرورش وجود دارد. دراین تحقیق با توجه به اینکه ارزیابی ساختمان های اسکلتی و مصالح بنایی آیتم های مختلفی دارند، این ساختمان ها جداگانه مورد بررسی قرار گرفته و عوامل تاثیر گذار برهر یک از ساختمانها را شناسایی و الگوریتم ها جداگانه با هم مقایسه و ارزیابی می شوند. بهترین الگوریتم را که هم از دقت بالایی برخوردار بوده و هم سریع به نتیجه رسیده و همچنین نتایج خروجی به راحتی قابل تفسیر باشد به عنوان الگوریتم برتر، انتخاب شده و به نتایج حاصل از آن استناد خواهد شد.

2-روش تحقیق

روش های متعددی برای اجرای پروژه های داده کاوی وجود دارد که یکی از روش های قدرتمند در این زمینه متدولوژی کریسپ می باشد[2]. این مقاله نیز بر اساس این متدولوژی تنظیم شده است (شکل1). در ادامه به بررسی هر یک از این مراحل در جهت رسیدن به مدلی برای پیشگویی نتیجه ارزیابی ساختمان های آموزشی خواهیم پرداخت.

الف)شناخت سیستم

به کارگیری موفق داده کاوی مستلزم شناخت حوزه ای است که قرار است داده کاوی در آن به کار برده شود و علاوه بر آن شناخت کافی از روش ها و ابزارهای داده کاوی نیز لازم است. به طور کلی تیم داده کاوی بایستی دانش کافی در حوزه ای که قرار است بررسی شود داشته باشند.در گام نخست پژوهشگر با مطالعه و مشورت با کارشناسان ارزیابی ساختمان های اسکلتی و تعیین فاکتور های مهم در نتیجه ارزیابی ساختمان های اسکلتی سعی در شناخت کافی در حوزه مورد بررسی خواهد داشت.

www.SID.ir
Archive of SID

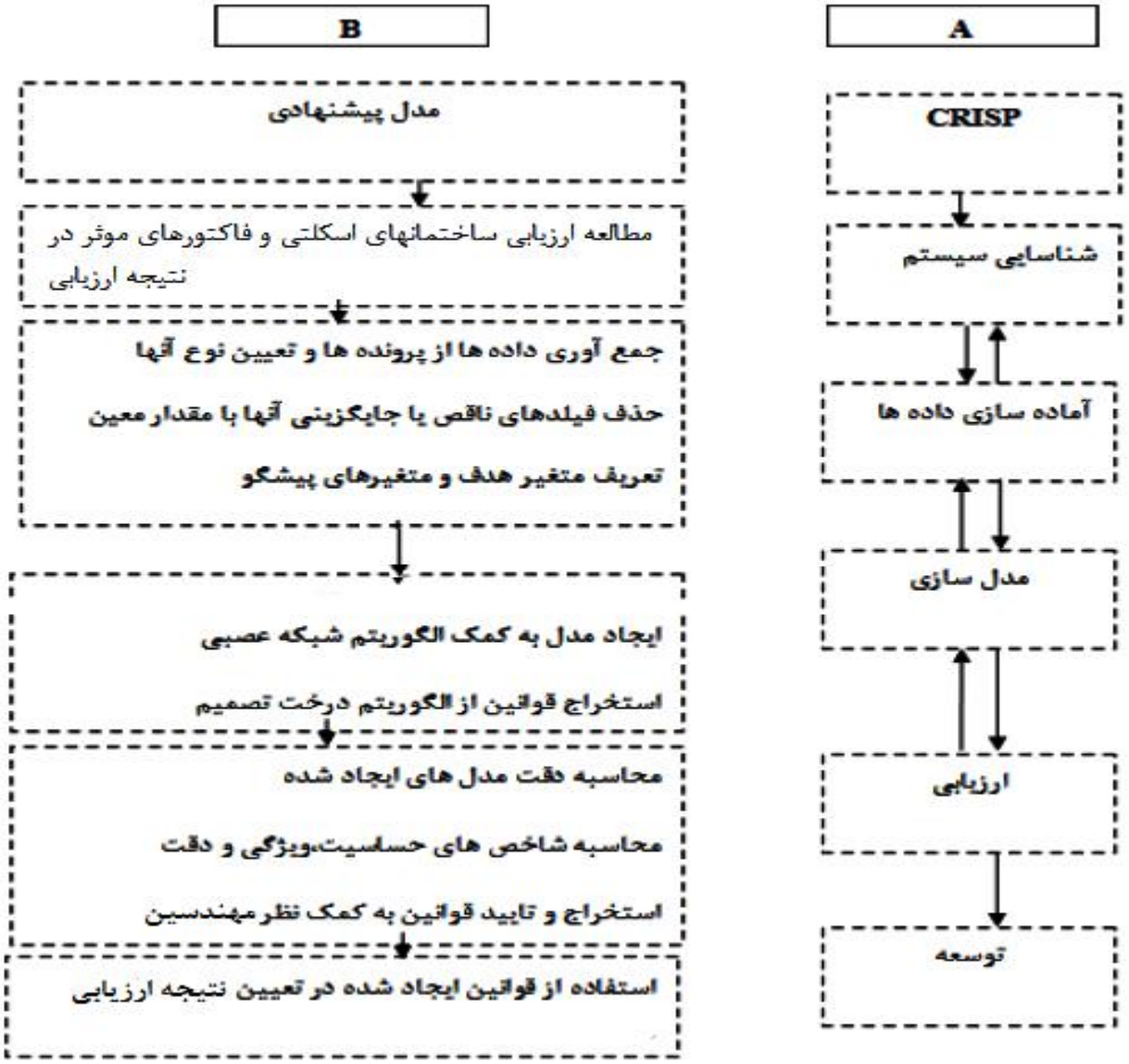
ب) مرحله آماده سازی داده ها

در این مرحله متغیر های موجود در مجموعه داده ها باید شناسایی شوند و با توجه به اهداف تعیین شده متغییرهای مهم و مرتبط انتخاب یا استخراج شوند. همچنین در این مرحله باید شناخت کلی از نوع و ماهیت داده ها ایجاد شود. مرحله جمع آوری و پیش پردازش داده ها مهمترین و زمان برترین مرحله در پروژه های داده کاوی است. دراین تحقیق از داده های ذخیره شده از بانک اطلاعاتی مدارس سازمان نوسازی مدارس کشور استفاده گردیده است. این جدول دارای تعداد130000 رکورد از مشخصات ساختمان ها بدست آمده و هر رکورد دارای 108 خصیصه است. این داده ها با درخواست کتبی از سازمان نوسازی مدارس قابل دریافت می باشد. در ادامه مهمترین این فیلد ها بررسی می شوند.

- **کدفضا**:یک کد 8 رقمی منحصر به فرد می باشد که هر مدرسه دارای یک کد فضا می باشد. البته این فیلد پس از اعمال الگوریتم های انتخاب ویژگی حذف گردید.
- **شناسه ساختمان**:یک کد 6 رقمی منحصر به فرد است که چون هر فضا ممکن است دارای چند ساختمان باشد این کد هم منحصر به فرد شده است تا شناسایی ساختمان ها نیز امکان داشته باشد.
- **سازه**:یک متغیر اسمی می باشد که شامل سه مقدار اسکلتی نیم اسکلتی و مصالح بنایی می باشد.

Archive of SID

www.SID.ir



شکل1: گام های متدولوژی کریسپ(A) و چارچوب استفاده شده در این مطالعه(B)

- **ارزیابی ساختمان**: این فیلد یک متغییر اسمی است و به عنوان فیلد هدف در این مطالعه شناسایی شده است که پس از کارشناسی ساختمان های آموزش پرورش توسط کارشناسان سازمان نوسازی مدارس کشور که عمدتا مهندسین عمران و معماری بودند در سال 1389 تکمیل شده است که شامل 3 مقدار مستحکم، نیاز به مقاوم سازی و تخریبی می باشد.
- **آیین نامه زلزله**: این فیلد یک متغییر اسمی است که شامل 5 مقداربه ترتیب: قبل از ویرایش 1،ویرایش 1،ویرایش2،ویرایش3 و ویرایش 4 می باشد.
- **فرم بام**: این فیلد یک متغییر اسمی است که شامل دو مقدار تخت و شیروانی می باشد.

www.SID.ir

- **نوع فاضلاب**: این فیلد یک متغییر اسمی است که شامل دو مقدار چاه جذبی و فاضلاب شهری است.
- **شیب زمین**: این فیلد یک فیلد اسمی است که شامل چهار مقدار مسطح ، شیب کمتر از 8 ، شیب بین 8-15 و شیب بالای 15 می باشد.
- **سال ساخت**: یک متغییر اسمی است که از مقادیر 1294 تا 1396 را به خود اختصاص داده است.
- **نوع اسکلت**: یک متغییر اسمی است که شامل سه مقدار اسکلت فلزی،اسکلت بتونی و مصالح بنایی می باشد.
- **اقلیم**: یک متغییر اسمی است که شامل 4 مقدار سرد و کوهستانی،گرم و خشک،بارانی و کویر می باشد.
- **وزن سقف**: که شامل سه مقدار خوب ، بد و متوسط می باشد.
- **انسجام سقف**: که شامل سه مقدار خوب ، بد و متوسط می باشد.
- **خیز سقف**: که شامل سه مقدار خوب ، بد و متوسط می باشد.
- **نشست پی**: که شامل سه مقدار خوب ، بد و متوسط می باشد.
- **رطوبت دیوارها و نم باران**: که شامل سه مقدار خوب ، بد و متوسط می باشد.
- **درز انقطاع در پلان**: که شامل سه مقدار خوب ، بد و متوسط می باشد.

در این مرحله داده ها تشریح شده و با استفاده از روش های آماری دانش مقدماتی از داده ها کسب می گردد. این دانش مقدماتی در مراحل بعدی به تشریح و تفسیر نتایج مدل سازی داده کاوی کمک شایانی خواهد کرد.پس از پاک سازی داده ها باید زیرمجموعه ای مناسب از داده ها را انتخاب کرد به این معنی که اگر متغیرهای غیر مرتبطی در مجموعه داده ها وجود دارد حذف گردد. از طرفی اگر حجم داده ها بسیار زیاد است می توان از روش های نمونه گیری ساده و پیشرفته به منظور کاهش حجم داده ها استفاده نمود. در این تحقیق فیلدهایی که اطلاعاتی بوده و ارزش محاسباتی ندارند مانند آدرس مدرسه،کد فضای مدرسه،کد واحد سازمانی و غیره حذف شده و برای نمونه فقط مدارس استان کردستان مورد ارزیابی قرار گرفته است.

اکتشاف داده ها: برخی از روش های آماری مانند پراکندگی،همبستگی،مرکزیت و...به منظور کشف داده ها در این مرحله استفاده می شود. با توجه به اینکه بودجه نوسازی مدارس باید عادلانه در سطح کشور به استان ها تخصیص پیدا کند و در استان هم عادلانه بین شهرستانهای تابعه تقسیم شود. یک ارزیابی کلی از ساختمانهای استان به تفکیک شهرستان را در ادامه بررسی کرده ایم و دانشی کلی از

Archive of SID

www.SID.ir

اطلاعات را بدست آورده و مورد بررسی قرار می دهیم. دراین مرحله یک سری کارآماری برروی داده هاجهت درک بیشتراز داده ها انجام شده است که درادامه نتایج آماری راخواهیم دید(شکل2).

شهرستان	بهسازی جزئی	درصد بهسازی جزئی از کل استان	تخریبی است	درصد تخریبی از کل استان	تغییر کاربری نظر کارشناسی	حداقل شرایط استحکام را دارد	درصد حداقل شرایط استحکام از کل استان	فاقد ارزش	نا مشخص	نیاز به مقاوم سازی دارد	درصد نیاز به مقاوم سازی از کل استان	مجموع تعداد ساختمانهای هر شهرستان
بانه	10	3.23	88	8.50	1	45	5.84	23	1	119	11.2	287
بیجار	62	20.00	61	5.89		54	7.00	12	1	184	17.4	374
دهگلان	14	4.52	82	7.92		58	7.52	16		53	5.02	223
دیواندره	8	2.58	141	13.62	1	57	7.39	66		23	2.18	296
سروآباد	3	0.97	73	7.05		30	3.89	21		46	4.36	173
سقز	159	51.29	99	9.57		160	20.75	53	2	102	9.67	575
سنندج	23	7.42	219	21.16	2	145	18.81	28	11	228	21.6	656
قروه	22	7.10	88	8.50	1	67	8.69	97		82	7.77	357
کامیاران	4	1.29	99	9.57		75	9.73	53	2	118	11.1	351
مریوان	5	1.61	85	8.21		80	10.38	57	3	100	9.48	330
جمع کل	**310**		**1035**		**5**	**771**		**426**	**20**	**1055**		**3622**

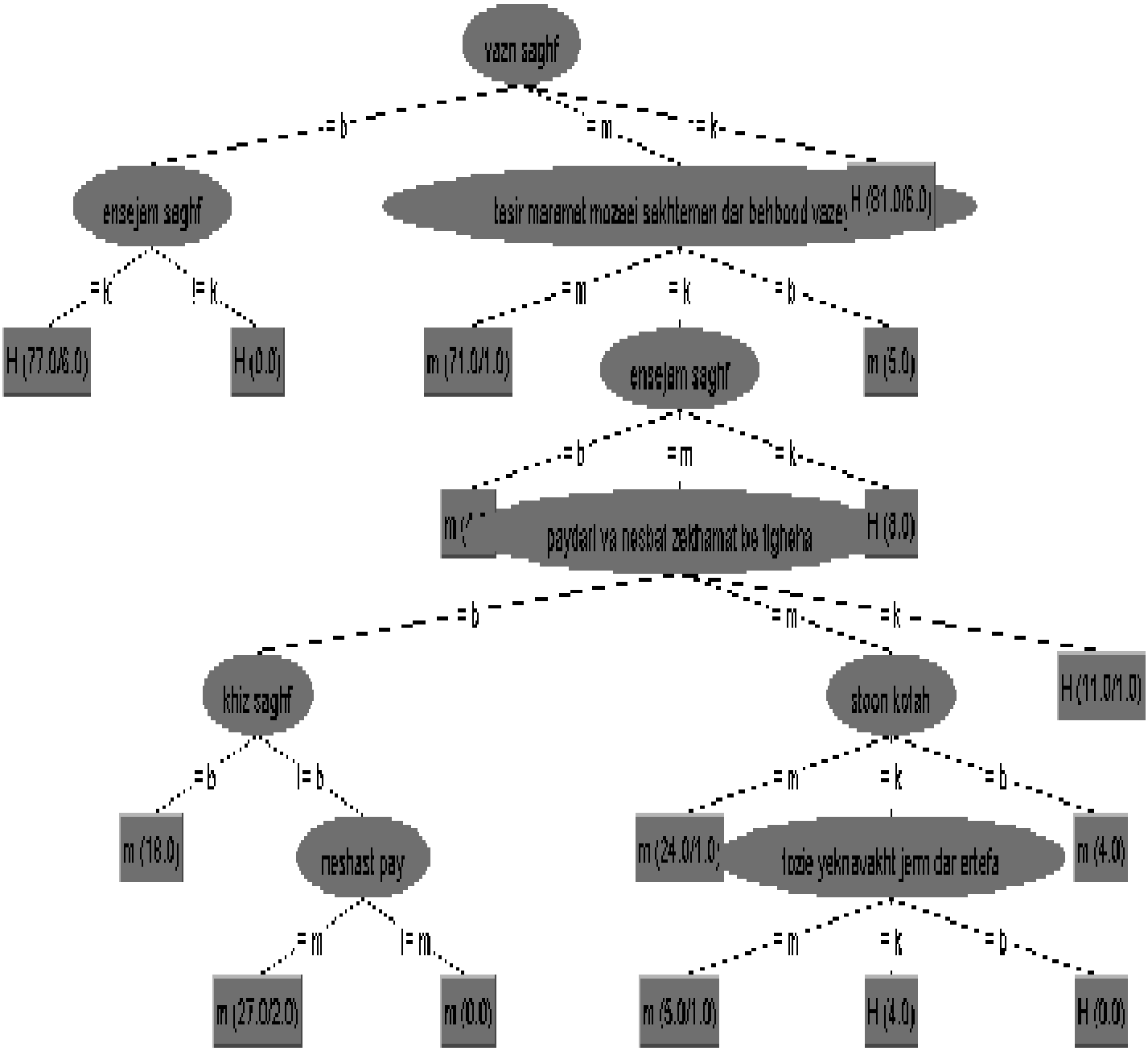
شکل2: نتایج آماری بر روی ساختمان های آموزشی استان کردستان

ج) مدل سازی

روش های داده کاوی متنوعی برای مدلسازی وجود دارد. در این مرحله با استفاده از تکنیک های داده کاوی به ارائه مدل پرداخته شد. مدل سازی با استفاده از نرم افزار weka3.6.9 انجام شد و در این مطالعه از روشهای پیش گویانه در داده کاوی استفاده شده است در این مرحله الگوریتم های درخت تصمیم، شامل الگوریتم های(J48, NBTree, BFTree, LMT, RandomTree, UserClassifier) و شبکه عصبی شامل لگوریتم های (Logistic, RBFNetwork, SimpleLogistic, SMO, MultilayerPerceptron) با بکارگیری متغیرهای ورودی و تعیین متغیر هدف ایجاد شدند. برای ساخت مدل های درخت تصمیم و شبکه عصبی فیلدهای وزن سقف، انسجام سقف، خیز سقف، درز انقطاع درپلان، نشست پی، اتصال بین دیوار و نما، بازشوها و... به عنوان متغیرهای پیشگو و فیلد نتیجه ارزیابی نیز به عنوان متغیر هدف تعیین گردید. مدل شبکه عصبی دارای یک لایه ورودی، تعدادی لایه پنهان و یک لایه خروجی است. هرگره ورودی را دریافت کرده، آن را پردازش نموده و خروجی را تولید می کند. تعیین اینکه آیا هر ورودی به گره خروجی خواهد رسید یا خیر به وزن آن ورودی بستگی دارد[3]. یک درخت تصمیم ترکیب تعدادی استلزام منطقی(قانون اگر -آنگاه)است . درخت-های تصمیم تنها نمایشی از فرایند تصمیم گیری نیستند بلکه می توان از آنها در حل مسائل طبقه بندی نیز استفاده کرد .معمولا مجموعه قوانین استخراج شده از درخت تصمیم، مهمترین

www.SID.ir
Archive of SID

اطلاعاتی است که از آنها بدست می آید[4]. در (شکل3) درخت تصمیم که توسط الگوریتم j48 graft ایجاد شده است نشان داده شده است.



شکل3: نتیجه اعمال الگوریتم بر روی ساختمان های اسکلتی آموزشی

در این تحقیق برای بهره بردن از الگوریتم های وکا لازم است که داده های موجود، به روشهای مختلف تبدیل شوند . چون بعضی از الگوریتم ها با داده های عددی غیر فعال می شدند و لازم شد که داده ها به کاراکتر تبدیل شوند تا با الگوریتم مورد نظر سازگار شوند. این پیاده سازی ها با به کارگیری

www.SID.ir

الگوریتم های معمول در وکا و همچنین اعمال روش های مختلف انتخاب ویژگی، انجام شده است. جدول 2 نتیجه ترکیب الگوریتم انتخاب ویژگی به روش GeneticSearch با 11 الگوریتم انتخابی را نشان می دهد.

جدول2 نتیجه ترکیب روش های انتخاب ویژگی و الگوریتم های دسته بندی

الگوریتم	مدل سازی بدون انتخاب ویژگی		مدل سازی با انتخاب ویژگی			
					نحوه انتخاب ویژگی	
	مدت زمان ساخت مدل	دقت مدل	مدت زمان ساخت مدل	دقت مدل	method	Attribute evaluator
trees.J48	0.02	89.13	0.02	89.31	GeneticSearch	CfsSubsetEval
trees.NBTree	1.72	93.58	0.88	92.73	GeneticSearch	CfsSubsetEval
trees.BFTree	0.19	88.03	0.13	88.03	GeneticSearch	CfsSubsetEval
trees.LMT	6	94.01	5.58	95.72	GeneticSearch	CfsSubsetEval
trees.RandomTree	0	86.32	0.1	88.03	GeneticSearch	CfsSubsetEval
trees.UserClassifier	6.61	55.55	11.49	55.55	GeneticSearch	CfsSubsetEval
functions.MultilayerPerceptron	51.33	96.15	34.58	96.58	GeneticSearch	CfsSubsetEval
functions.SMO	0.08	97.00	0.16	97.00	GeneticSearch	CfsSubsetEval
functions.SimpleLogistic	2.34	94.01	1.72	95.72	GeneticSearch	CfsSubsetEval
functions.RBFNetwork	0.06	94.44	0.06	96.58	GeneticSearch	CfsSubsetEval
functions.Logistic	0.08	97.43	0.05	96.15	GeneticSearch	CfsSubsetEval

Archive of SID

د) ارزیابی

در این مرحله پس از ایجاد مدل بایستی به ارزیابی مدل ایجاد شده پرداخت .برای بررسی صحت مدل داده ها به دو دسته آموزش(80درصد)و آزمون(20درصد)تقسیم شدند. داده های بخش آموزش مدل را می سازند و داده های بخش آزمون مدل ایجاد شده را مورد ارزیابی قرار می دهند.جهت ارزیابی مدل ها می توان از شاخص های حساسیت، ویژگی، دقت، ارزش اخباری مثبت و ارزش اخباری منفی استفاده کرد .جهت محاسبه این شاخص ها از ماتریس تداخلی ایجاد شده در محیط نرم

www.SID.ir

افزار استفاده شد .در جدول2 شاخص های مقایسه دقت الگوریتم شبکه عصبی و الگوریتم های درخت تصمیم ارائه شده است.

ه) توسعه

ساخت مدل معمولا پایان پروژه نیست حتی اگر هدف مدل افزایش دانش است، دانش بدست آمده نیاز به سازماندهی دارد و نمایش به طریقی که کاربر بتواند از آن استفاده کند. دانش کشف شده باید سازماندهی شده و به شکل قابل ارائه برای دیگران درآید. ما در این مرحله بعد از ایجاد گزارشات لازم سعی کردیم توضیح دهیم که بر اساس مدلهای ایجاد شده تاثیرگذارترین فاکتورها در تعیین نتیجه ارزیابی ساختمان های اسکلتی آموزشی کدامند. از الگوریتم های مورد استفاده دراین مطالعه، الگوریتم Logistic دارای بالاترین میزان دقت (97.43 درصد) بود. در مرحله ارزیابی نظرات مهندسان سازمان نوسازی مدارس نیز در مورد قوانین ایجادشده اعمال می گردد. به این ترتیب که قوانین بدست آمده به مهندسین نوسازی مدارس ارائه شده و قوانینی که از نظرآنها معتبربودند به عنوان قوانین نهایی ارائه گردیدند. برخی از مهم ترین موارد این قوانین در ادامه ارائه شده است. بنابراین طبق نظر مهندسین ساختمان می توان گفت که فاکتورهای سال ساخت ساختمان، نوع اسکلت، نوع مصالح، وزن سقف، خیز سقف، نوع اقلیم آب و هوایی و ... بیشترین تاثیر را در مشخص کردن نتیجه ارزیابی دارا هستند و این در حالی است که بر اساس مقایسه های انجام شده بر اساس اولویت بندی متغیرها توسط الگوریتم های مورد بررسی نیز این متغیرها جزو فاکتورهای اول قرار گرفته اند، که نشان از اهمیت این متغیرها در تعیین نتیجه ارزیابی ساختمان های آموزشی را دارند برخی از مهم ترین قوانینی که از نظر مهندسین سازمان نوسازی مدارس در این بررسی مورد تایید قرار گرفتند عبارتند از:

1. از 105 مورد که انسجام سقف= متوسط و نتیجه ارزیابی= مقاوم سازی است، 98 موردآن وزن سقف=متوسط است. conf:(0.93)

2. از 103 مورد که انسجام سقف= متوسط و سیستم سازه ای و کیفیت پله= متوسط ، 95 موردآن وزن سقف=متوسط است. conf:(0.92)

3. از103 مورد که پایداری و نسبت ضخامت به تیغه ها=متوسط و نتیجه ارزیابی= مقاوم سازی است،95 مورد آن وزن سقف= متوسط است. conf:(0.92)

4. از 102 مورد که انسجام سقف= متوسط و پایداری و نسبت ضخامت به تیغه ها = متوسط، 94 موردآن وزن سقف=متوسط است. conf:(0.92)

5. از 105 مورد که انسجام سقف= متوسط و اتصال بین دیوار ونما = متوسط، 96 موردآن وزن سقف=متوسط است. conf:(0.91)

Archive of SID
www.SID.ir

نتیجه گیری

در این تحقیق بااستفاده ازالگوریتم شبکه عصبی و درخت تصمیم به ارائه مدل و استخراج قوانین آن در راستای پیشگویی احتمال نتیجه ارزیابی ساختمان پرداخته شد. بهترین نتایج از الگوریتم Logistic بدست آمد که دقت آن (97.43 درصد) بود. بیشترین فاکتورهای تاثیرگذار بر نتیجه ارزیابی ساختمان های اسکلتی سال ساخت، نوع مصالح، وزن سقف،خیز سقف، انسجام سقف، سیستم سازه ای و کیفیت پله، نوع فاضلاب و اقلیم آب و هوایی بودند.

سپاسگزاری

این مقاله برگرفته از پایان نامه کارشناسی ارشد اینجانب می باشد، بدین وسیله ازاساتید راهنما، مشاور و همچنین کارکنان محترم نوسازی مدارس استان کردستان کمال تشکر و قدردانی را دارم.

مراجع

1.Subbalakshmi G, Road Y. [Decision Support in Heart Disease Prediction System using Naive Bayes]. Indian Journal of Computer Science and Engineering 2011; 2(2): 183-195.

2. Jianxin C, Yanwei X, Guangcheng X, Jianqiang Y, Dongbin Z. [A Comparison of Four Data Mining Models: Bayes, Neural Network, SVM and decision Trees in Identifying Syndromes in Coronary Heart Disease]. Springer- Verlag Berlin Heidelberg 2007:1274–1279.

3. IBM SPSS Modeler CRISP-DM Guide. Available from:www.spss.ch/upload/1107356429_CrispDM1 among Cardiovascular Risk Factors]. Available from: www. civilica. Com/Paper- IDMC03-IDMC03_ 065.

4 **.** Margaret R, Kraft C, Desouza A. [Data Mining in Healthcare Information Systems. Case Study of a Veterans' Administration Spinal Cord Injury Population].Proceedings of the 36th Hawaii International Conference on System Sciences 2003

Archive of SID

www.SID.ir