

۱۴۵۶

پیش بینی موفقیت و عدم موفقیت فارغ التحصیلی دانشجویان رشته‌ی تحصیلی مکانیک دانشگاه شاهرود

محسن فرهادی^۱؛ سهیلا ملارستمی^۲؛ بهاره سادات میراسدی^۳

چکیده

در عصر یادگیری، یکی از چالش‌های اساسی در زمینه‌ی آموزش، ارتقا سطح کیفی آموزش و بررسی عوامل موثر بر آن می‌باشد. پیدا کردن الگوها و روابط میان پارامترهای مختلفی که به صورت پنهان، درون حجم داده‌های آموزشی قرار گرفته است، می‌تواند کمک شایانی به تصمیم‌گیرندگان در راستای بهبود فرایند آموزشی نماید. در این راستا، داده کاوی به عنوان ابزاری برای کشف دانش از پایگاه داده‌های آموزشی مورد استفاده قرار گرفته و از اطلاعات پنهانی که برای برنامه ریزی طولانی مدت و بهبود سطح آموزشی، می‌تواند به کارگرفته شود، پرده برداری می‌کند.

در این پروژه با به کارگیری تکنیک درخت تصمیم، دریک نمونه‌ی عملی، به شناسایی عوامل موثر در پیش بینی موفقیت یا عدم موفقیت دانشجویان رشته‌ی تحصیلی مکانیک در فارغ التحصیلی پرداخته ایم. امید است ارائه این پروژه، مسئولان آموزشی دانشگاه را قادر سازد تا با اجرای گام‌هایی مطمئن‌تر، نسبت به آینده عمل نمایند.

کلمات کلیدی

داده کاوی، پایگاه داده‌ی آموزشی، کشف دانش، درخت تصمیم.

کنفرانس داده کاوی ایران

۱- عضو هیات علمی دانشکده کامپیوتر و فناوری اطلاعات، mfarhadi@shahroodut.ac.ir

۲- دانشجوی کارشناسی مهندسی فناوری اطلاعات، mollarostami@yahoo.com

۳- دانشجوی کارشناسی مهندسی فناوری اطلاعات، bahare_mirasadi@yahoo.com

Predict success and failure of mechanical graduated students of shahrood university

Mohsen Farhadi; Soheila Mollarostami; Bahare Sadat Mirasadi

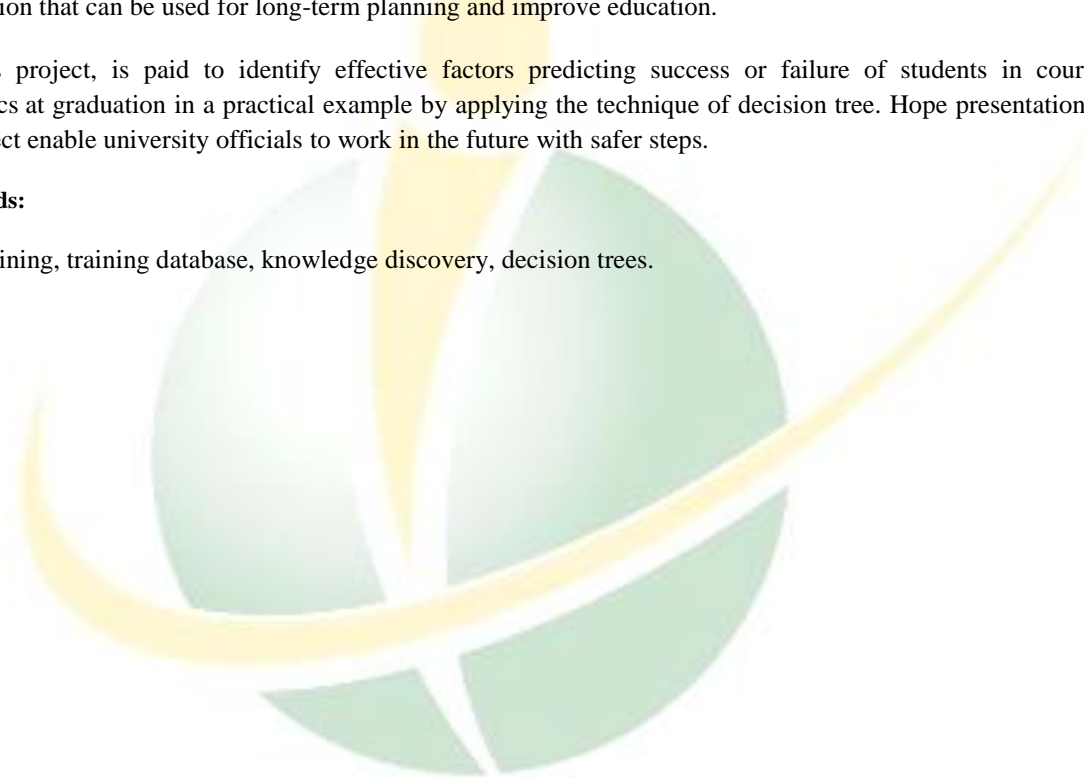
Abstract :

In age of learning, improving education and its identify effective factors is one of the main challenges in the field of education. Finding Patterns and their relationships between the various parameters which are hidden inside a large number of training data, can help decision-makers in order to improve the educational process. In this way data mining is used for knowledge discovery from database training and is unveiled from hidden information that can be used for long-term planning and improve education.

In this project, is paid to identify effective factors predicting success or failure of students in courses mechanics at graduation in a practical example by applying the technique of decision tree. Hope presentation of the project enable university officials to work in the future with safer steps.

Key words:

data mining, training database, knowledge discovery, decision trees.



کنفرانس داده کاوی ایران

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌ها ی ذخیره شده در این سیستم‌ها، به ابزاری نیاز است تا بتوان این داده‌ها را پردازش کرد و اطلاعات حاصل از آن را در اختیار کاربران قرار داد. معمولاً کاربران پس از طرح فرضیه‌ای، براساس گزارشات مشاهده شده به اثبات یا رد آن می‌پردازند، درحالی که امروزه به روش‌هایی نیاز داریم که به اصطلاح به کشف دانش (Knowledge Discovery) بپردازند، یعنی روش‌هایی که با کمترین دخالت کاربر و به صورت خودکار، الگوها و رابطه‌های منطقی را بیان نمایند. یکی از روش‌های بسیار مهمی که با آن می‌توان الگوهای مفیدی را در میان داده‌ها تشخیص داد، داده‌کاوی است. این روش که با حداقل دخالت کاربران همراه است، اطلاعاتی را در اختیار آن‌ها و تحلیل‌گران قرار می‌دهد تا براساس آن‌ها، تصمیمات مهم و حیاتی را در سازمان خود اتخاذ نمایند. باید توجه داشت که هر چه حجم داده‌ها بیشتر و روابط میان آن‌ها پیچیده‌تر باشد، دسترسی به اطلاعات نهفته در میان داده‌ها مشکل‌تر می‌شود و نقش داده‌کاوی به عنوان یکی از روش‌های کشف دانش، آشکارتر می‌گردد.

داده‌کاوی آموزشی نیز در کشور با توجه خاصی همراه بوده است و تلاش افراد براین بوده تا شکاف دانشی تا حدودی در این حوزه از میان برداشته شود. افزایش تعداد مقالات در کنفرانس‌های داده‌کاوی برگزار شده در سال‌های اخیر، گواهی بر این ادعاست.

در این مقاله، هدف اصلی را بررسی و شناخت عوامل موفقیت و عدم موفقیت دانشجویان در فارغ التحصیلی قرار می‌دهیم. به همین منظور، با اعمال مراحل داده‌کاوی با استفاده از الگوریتم درخت تصمیم بر انباره داده‌ی مرتبط با اطلاعات آموزشی دانشجویان رشته‌ی تحصیلی مکانیک، به شناخت مدل موردنظر می‌پردازیم.

۲- روش انجام تحقیق

بطور کلی، داده‌کاوی (که گاهی اوقات اکتشاف اطلاعات یا دانش نامیده می‌شود) عبارتست از فرآیندی که از چشم اندازهای مختلف به تحلیل داده‌ها می‌پردازد و جمع‌بندی آن‌ها را در قالب اطلاعات مفیدی ارائه می‌کند. مراحل اجرای داده‌کاوی شامل:

- جمع‌آوری داده‌ها، بررسی و تحلیل آن‌ها و در نهایت، انتخاب داده‌های مناسب
- یکپارچه سازی داده‌ها (پاک‌سازی و...)
- ساخت مدل
- تحلیل و بهره برداری از مدل

نرم افزار داده‌کاوی یکی از ابزارهای تحلیل اطلاعات است. این نرم افزار به کاربران امکان می‌دهد، اطلاعات را از ابعاد و زوایای بسیار متفاوت تحلیل و طبقه بندی کنند و روابطی را که در آن‌ها شناسائی نموده اند به اجمال بیان نمایند. در این پروژه نرم افزار "microsoft sql server 2008" مورد استفاده قرار گرفته است.

۲-۱- جمع‌آوری داده‌ها^۴

اطلاعات دانشجویان رشته‌ی تحصیلی مکانیک از سال تحصیلی ۱۳۷۸ تا ۱۳۹۱ جمع‌آوری شده است. این اطلاعات شامل شماره دانشجویی، نظام ورودی، دوره (روزانه، شبانه)، ترم ورود، جنسیت، معدل کل، وضعیت دانشجو (فارغ التحصیل، انتقالی، مهمان، اخراج و...)، شماره درس، نمره، کد درس (اصلی، پایه، عمومی، تخصصی، اختیاری)، تاریخ و روز و ساعت امتحان و... می‌باشد.

۲-۲- ایجاد جداول فارغ التحصیلی

جدول (۱) فیلدهای اطلاعاتی انتخاب شده

ردیف	نام فیلد	توضیح فیلد	مجموعه مقادیر
1	ID	شماره دانشجویی	
2	Sysno	دوره (روزانه ، شبانه)	(1,2)
3	Sex	جنسیت (مرد، زن)	(0,1)
4	Gpa	معدل کل	
5	Trmavg	معدل ترم	
6	Stdstatus	وضعیت دانشجو (مشغول به تحصیل، فارغ التحصیل، انصراف، اخراج)	[105,110,112,10]
7	Trmno	شماره ترم جاری	
8	Entrtrm	شماره ترم ورودی	
9	Trmunt	تعداد واحد اخذ شده در هر ترم	
10	Generalnum	تعداد دروس عمومی	
11	Basenum	تعداد دروس پایه	
12	Mainnum	تعداد دروس اصلی	
13	Generalavg	معدل دروس عمومی	
14	Baseavg	معدل دروس پایه	
15	Mainavg	معدل دروس اصلی	
16	Agototalpassunit	تعداد واحد گذرانده تا ترم قبل	

هدف این مقاله، ایجاد مدلی به منظور پیش بینی معدل کل فارغ التحصیلی دانشجویان و همچنین شناخت عوامل مؤثر بر آن است. از این رو، جدول هایی را با ترکیب اطلاعات تحصیلی چهار ترم اول دانشجو (ترم یک و دو، ترم یک و دو و سه، ترم دو و سه، ترم سه و چهار، ترم دو و سه و چهار، ترم یک و دو و سه و چهار) ایجاد می کنیم. گام بعدی، انتخاب بهترین مدل ایجاد شده از مجموعه های ترمی و استفاده از آن برای پیش بینی است.

۳- کار با محیط microsoft visual studio

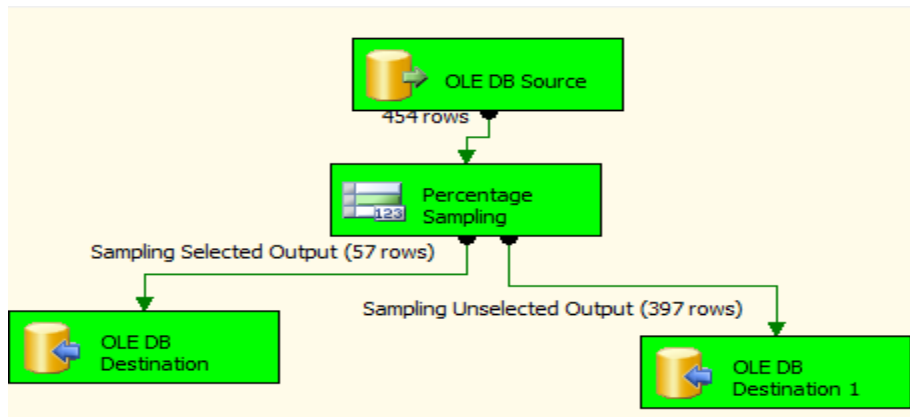
کنفرانس داده کاوی ایران

۳-۱- ساخت مدل

در این مرحله، جدول مربوط به دانشجویان فارغ التحصیل در محیط "SQL Server Business Intelligence Development Studio"، مورد بررسی قرار می گیرد. به این منظور داده های مرتبط با هر مجموعه ی ترمی، به دو قسمت ^۵test و ^۶train تقسیم بندی می شود.

^۵ -آزمایشی

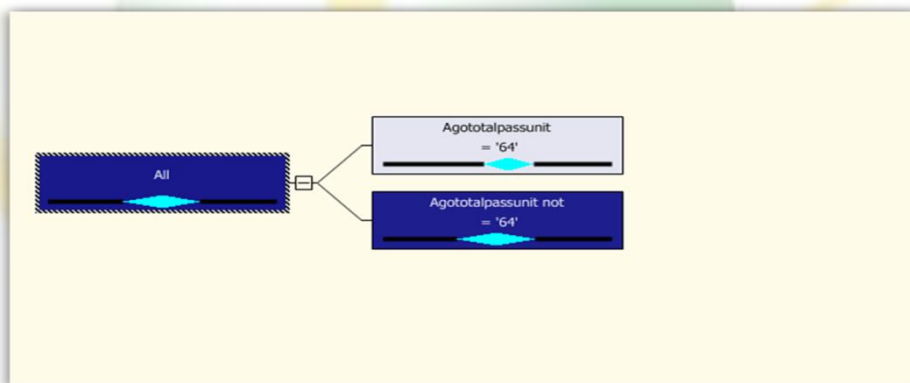
^۶ -آموزشی



شکل(۱) داده های تقسیم بندی شده به دو مجموعه train و test

۳-۲-درخت تصمیم^۷

درخت تصمیم یکی از ابزارهای قوی و متداول برای پیش بینی و مدل سازی است. در شکل زیر، درخت تصمیم برای مجموعه ی ترمی "یک و دو و سه و چهار" نمایش داده شده است.



شکل(۲)درخت تصمیم

$$Gpa=13.919+0.833*(fourtrmavg-13.620)+0.128*(fourbasenum-6.004) \quad (۱)$$

باتوجه به فرمول (۱) به دست آمده از درخت تصمیم فوق، می توان معدل دروس پایه ومعدل دروس اصلی را به عنوان بیشترین عوامل مؤثر در معدل کل دانشجو در مجموعه ترم های "یک و دو و سه و چهار" معرفی کرد.

۳-۳- نمودار پراکندگی^۸

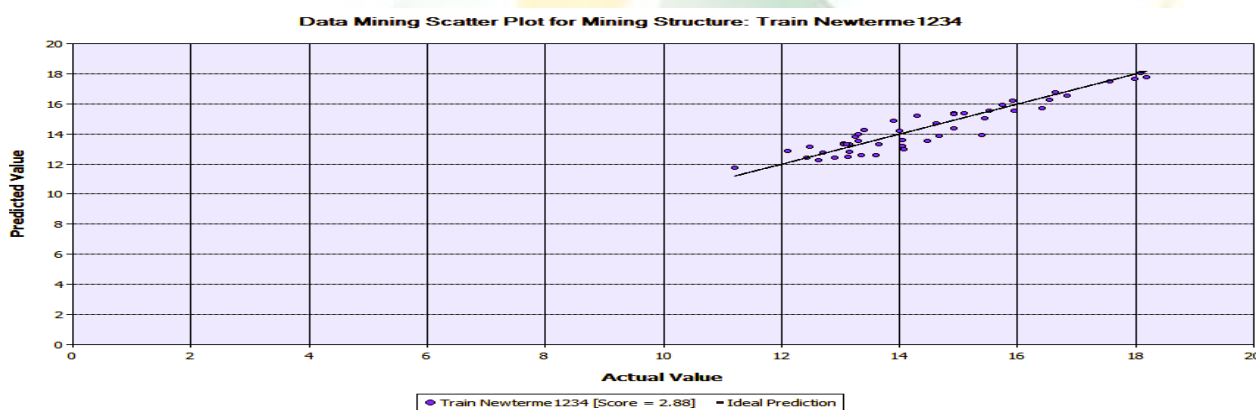
همانطور که از اسم این نمودار پیداست، وظیفه ی این نمودار نمایش پراکندگی داده هاست. پراکندگی داده ها با استفاده از دو مجموعه ی train و test ارزیابی می شود. برای انتخاب بهترین نمودار کفایست به مقدار SCORE توجه شود.

^۷ decision tree-
^۸ Scatter plot-

جدول (۲) مقادیر SCORE به دست آمده در مجموعه‌های ترمی

شماره ترم	مقدار SCORE
۱ و ۲	۱,۶۸
۳ و ۲	۲,۰۹
۳ و ۲	۱,۶۹
۴ و ۳	۱,۶۴
۴ و ۳ و ۲	۱,۴۷
۴ و ۳ و ۲ و ۱	۲,۸۸

در جدول فوق، مقدار SCORE برای هر یک از مجموعه ترم‌ها بیان شده است. بهترین نمودار پراکندگی، بیشترین مقدار SCORE را خواهد داشت. با کمی بررسی میتوان به این نتیجه دست یافت که، بهترین نمودار پراکندگی، مرتبط با مجموعه ترم‌هایی است که بیشترین تأثیر را بر روی پیش بینی خواهد داشت. این نمودار شکل منسجم تری داشته و پراکندگی داده‌ها در آن کمتر خواهد بود. نمودار زیر، مربوط به پراکندگی داده‌ها در ترم‌های "یک و دو و سه و چهار" است.

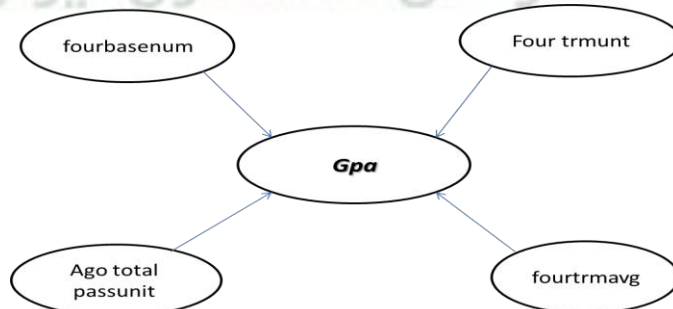


شکل (۳) پراکندگی مجموعه ترم‌های "یک و دو و سه و چهار"

محور افقی بیانگر مقدار معدل واقعی و محور عمودی بیانگر مقدار معدل پیش بینی شده است.

۳-۴- نمودار وابستگی^۹

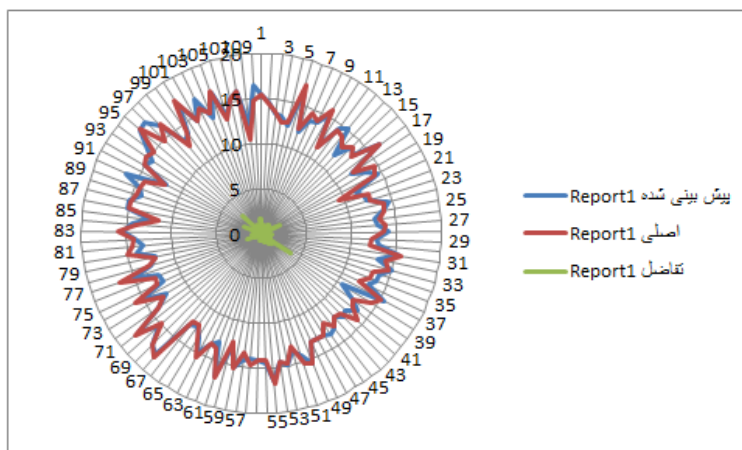
نمودار وابستگی و یا به عبارتی شبکه وابستگی، بیانگر عناصری است که بیشترین تأثیر را در پیش بینی خواهند داشت.



شکل (۴) وابستگی مجموعه ترم‌های "یک و دو و سه و چهار"

۴- بررسی میزان خطای مدل درخت تصمیم

در این قسمت از مقاله به ارزیابی پیش‌بینی‌های صورت گرفته می‌پردازیم؛ بدین منظور، برای انجام عملیات گزارش‌گیری و پیش‌بینی معدل کل، مشخصات ۱۰۳ دانشجوی به عنوان ورودی در نظر گرفته می‌شود. نمودار (۱)، نشان‌دهنده‌ی میزان اختلاف معدل واقعی و معدل پیش‌بینی شده است.



نمودار (۱) اختلاف معدل کل با معدل پیش‌بینی شده

با توجه به نتایج بدست آمده از نمودار بالا، اختلاف بین معدل واقعی و معدل پیش‌بینی شده در ۸۴٫۲٪ دانشجویان، کمتر از یک و میانگین درصد خطای نسبی ۴٫۲ است.

۵- نتیجه

کاربردهای داده کاوی در آموزش عالی به تازگی مورد توجه قرار گرفته است. دانشگاه‌ها با به کارگیری داده‌های تاریخی و استخراج دانش معنادار از آن، می‌توانند الگوهای خاص و با ارزش در مورد دانشجویان موفق، یک برنامه یا روش تدریس موفق بیابند.

این پژوهش تلاشی برای پیاده‌سازی مدل‌های داده کاوی پیش‌بینی‌کننده، همانند درخت تصمیم، به منظور بررسی عوامل مؤثر بر موفقیت و عدم موفقیت در فارغ‌التحصیلی دانشجویان رشته‌ی تحصیلی مکانیک دانشگاه شاهرود، بر اساس گذشته‌ی تحصیلی آنان بوده است. بر اساس نتایج حاصل از ساخت مدل‌های پیش‌بینی‌کننده، فاکتورهای تعداد دروس پایه و معدل کل در طول چهار ترم اول تحصیلی، به عنوان تأثیرگذارترین عوامل بر معدل کل دانشجویان در فارغ‌التحصیلی شناخته شده است.

قابل ذکر است، نتایج این مقاله نقطه‌ی پایانی برای تحقیق در این حوزه نیست و می‌توان با در نظر گرفتن فاکتورهای دیگر، به یافته‌های جدیدی دست یافت.

- [۱] مینایی، بهروز؛ میرافضل، هانیه سادات؛ هانی، سیدحسن؛ "شناسایی عوامل موثر بر افت تحصیلی دانشجویان با استفاده از قوانین انجمنی و تحلیل خوشه بندی" ششمین کنفرانس داده کاوی ایران، دانشگاه صنعتی امیرکبیر، ۱۳۹۱.
- [۲] ابراهیمی، ملیحه؛ رحیم اف، حامد؛ "کشف دانش از عملکرد دانشجویان با استفاده از قوانین انجمنی جهت تعیین وابستگی دروس رشته‌ی کامپیوتر و فناوری اطلاعات" ششمین کنفرانس داده کاوی ایران، دانشگاه امیرکبیر، ۱۳۹۱.
- [۳] مرادخانی، فرزاد؛ محمدی، مسعود؛ گوهرزاد، مریم؛ " بررسی تاثیر استفاده از اینترنت بر یادگیری دانشجویان با استفاده از فنون داده کاوی " ششمین کنفرانس داده کاوی ایران، دانشگاه امیرکبیر، ۱۳۹۱.
- [۴] شهرابی، جمال؛ شکورنیا، ونوس؛ داده کاوی در **sql server**، انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر، چاپ اول، ۱۳۸۸.



کنفرانس داده کاوی ایران