

عوامل موفقیت دانش آموزان رشته کامپیوتر هنرستان دکتر حسابی ابهر

رحمان عسگری^۱

چکیده

در این مقاله عوامل موفقیت دانش آموزان رشته کامپیوتر هنرستان دکتر حسابی ابهر مورد بررسی قرار گرفته است، تا بدانیم دانش آموزانی که با موفقیت فارغ التحصیل شده اند، قبل از ورود به این رشته دارای چه مشخصاتی بوده اند تا بتوانیم از روی اطلاعات بدست آمده از طریق داده کاوی بر روی این داده ها، به دانش آموزانی که میخواهند وارد این رشته شوند، جهت موفقیتشان در این رشته، مشاوره دهیم و بدانیم که آیا دانش آموزی که ثبت نام میکند، میتواند از این رشته با موفقیت فارغ التحصیل شود یا خیر.

کلمات کلیدی

داده کاوی، درخت تصمیم، خوشه بندی، الگوریتم K-means، الگوریتم J^{۴۸}.

Data Mining of success factors of students in the field of computer at Dr.Hesabi's technical school.

Rahman Asgari

Student MA IT ,Qazvin Islamic Azad University. Asgari.rahman@yahoo.com

Abstract

In this paper, the success factors of students in the field of computer at Dr.Hesabi's technical school has been studied in order to find out that students graduated successfully, what features they had before entering to this field. According to the findings of datamining on these inputs, we can counsel those students, what to study in this field successfully. In addition, we want to recognize that, can a student, who registers in this field, graduate successfully or not.

Key words

Data Mining, Decision tree, Clustering, K-means algorithm, J^{۴۸} algorithm

۱. مقدمه

در این مقاله سعی شده با داده کاوی بر روی فارغ التحصیلان رشته کامپیوتر، یک مدل مناسب جهت موفقیت دانش آموزانی که میخواهند وارد این رشته شوند، ارائه دهیم. تاکنون فعالیت هایی که در زمینه های مشابه صورت گرفته بیشتر بر روی مقطع ورود به دانشگاه بوده و در زمینه موضوع مقاله بیشتر به رشته های نظری پرداخته شده است. این مقاله با تکمیل پرسشنامه هایی توسط دانش آموزان فارغ التحصیل رشته کامپیوتر هنرستان دکتر حسابی ابهر (بین سالهای ۷۹ تا ۸۹) و با استفاده از نرم افزار Weka انجام گرفته است تا میزان و علل موفقیت و یا عدم موفقیت آن ها را مورد بررسی قرار دهیم.

۲. پیش پردازش داده ها

یکی از مراحل زمان بر و در عین حال بسیار مهم در کسب نتیجه مطلوب میباشد. در این مرحله نیاز است تا کیفیت داده های مورد بررسی را افزایش دهیم که شامل پاکسازی داده ها، تبدیل داده ها، یکپارچه سازی داده ها می باشد. [۱]

۱.۲. معرفی فیلدهای پرسشنامه

در این قسمت به معرفی فیلدهای پرسشنامه و توضیحی در رابطه با آنها میپردازیم، که فیلد هدف در این پروژه، فیلد مربوط به معدل دیپلم میباشد. (جدول ۱).

جدول (۱) معرفی فیلدهای پرسشنامه

عنوان فیلد	توضیح	عنوان فیلد	توضیح
Father-t	تحصیلات پدر	Mathematic No	نمره درس ریاضی اول متوسطه
Mother -t	تحصیلات مادر	Physic No	نمره درس فیزیک اول متوسطه
Revenue	میزان درآمد ماهیانه خانواده	Interest	میزان علاقه به این رشته
Worked in honarestan	شاغل در دوران هنرستان	Computer at home	داشتن کامپیوتر قبل از ورود به این رشته
Avg in ۱ school	معدل سال اول متوسطه	High - level	علاقه به تحصیل در دانشگاه
member of your family	میزان توصیه آشنایان	Council	میزان مشاوره با دبیر مشاور
Market	تاثیر گذاری بازار کار جهت انتخاب این رشته	Live	دوران تحصیل نزد چه کسی زندگی میکرده
Application	میزان تاثیر گذاری کاربردی بودن کامپیوتر جهت انتخاب این رشته	Your friend in classmate	همکلاسی بودن دوستان صمیمی
Lived Father	پدر در قید حیات است	Sincere Friend	محل تحصیل دوستان صمیمی
No of your sis&bro	تعداد خاهر و برادر	Daily worked by computer	ساعت کارکرد روزانه با کامپیوتر
Child-family	چندمین فرزند خانواده	Career and art	نمره درس حرفه فن
Financial-problem	داشتن مشکل مالی	Which action with computer	کاری که با کامپیوتر انجام میداده
Worked in ۱ school	شاغل در اول دبیرستان	Your parent in your choice	حمایت والدین از انتخاب این رشته
The teachers in your chioce	تاثیرگذاری معلمان قبل دوران هنرستان	The member of your family that studied	تحصیل کردن یکی از آشنایان در این رشته
Supplementary	انتخاب این رشته بخاطر گرایشهای مختلف	The efficacy in deiction	میزان تاثیر گذاری آن فرد تحصیل کرده در این رشته
Good name of the computer	انتخاب این رشته بخاطر خوش نامی	Haphazard	انتخاب اتفاقی این رشته
Avg of Diplom	معدل دیپلم		

۳. دسته بندی

دسته بندی اختصاص دادن داده ها به گروه هایی از پیش تعریف شده میباشد که هدف آن یادگیری قواعد یا ساختن مدلی جهت پیش بینی وضعیت آینده میباشد، برخی از روشهای دسته بندی عبارتند از:

- شبکه های بیز
- نزدیک ترین همسایه
- شبکه های عصبی
- درخت تصمیم
- رگرسیون [۲]

۱.۳. درخت تصمیم

یکی از روشهای دسته بندی ساخت درخت تصمیم می باشد، که این روش یکی از پرکاربردترین روش های دسته بندی میباشد.

از جمله ویژگی های درخت تصمیم:

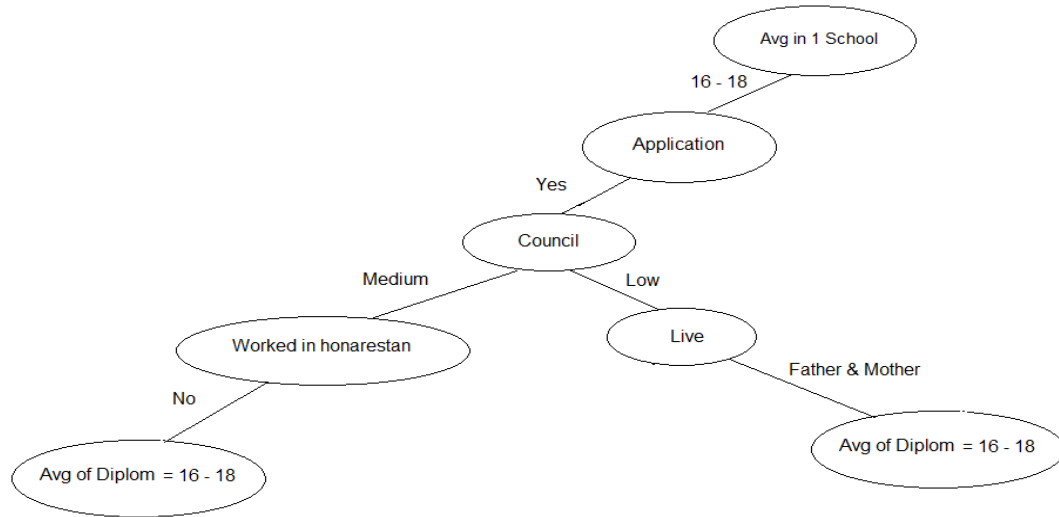
- استفاده ساده
- راحت تر بودن فهم نتایج
- عدم نیاز به حذف داده ها
- ایجاد قواعد استنتاجی

قابل استفاده برای داده های عددی و غیر عددی [۳]

۱,۱,۳. ایجاد مدل بر اساس درخت تصمیم

در این مرحله بر اساس داده های وارد شده به نرم افزار Weka، مدل درخت تصمیم بر اساس الگوریتم J48 ساخته شده است. شکل(۱).

شکل (۱) مدل به دست آمده بر اساس درخت تصمیم



۲,۱,۳. قوانین تولید شده از درخت تصمیم

- فارغ التحصیلانی که معدل اول دبیرستانشان بین ۱۶-۱۸ باشد و رشته کامپیوتر را به خاطر کاربرد فراوان آن انتخاب کرده باشند و با دبیر مشاوره خود درباره انتخاب رشته کامپیوتر به مقدار متوسط مشاوره کرده باشند و در دوران تحصیل در هنرستان شاغل نباشند، معدل دیپلم آنها بین ۱۶-۱۸ خواهد بود.
- فارغ التحصیلانی که معدل اول دبیرستانشان بین ۱۶-۱۸ باشد و رشته کامپیوتر را به خاطر کاربرد فراوان آن انتخاب کرده باشند و با دبیر مشاوره خود درباره انتخاب رشته کامپیوتر به مقدار کمی مشاوره کرده باشند و نزد پدر و مادر زندگی کنند، معدل دیپلم آن بین ۱۶-۱۸ خواهد بود.

۳,۱,۳. ارزیابی درصد پاسخ صحیح مدل

- با انجام عمل ساخت مدل درخت تصمیم با استفاده از الگوریتم J48، توانستیم به پاسخ صحیح ۸۳ درصدی برسیم، که درصد خوبی برای این مسئله است.

۴. خوشه بندی

- خوشه بندی، گروه بندی نمونه های مشابه میباشد، به این معنا که بر اساس میزان شباهت نمونه ها، آنها را در گروه های مختلف قرار میدهم.

۱,۴. تحلیل خوشه ای

- ابزاری است برای کشف ساختار نمونه ها بدون فرضیات روشهای آماری.

۲,۴. کاربردهای خوشه بندی

- تجزیه تحلیل شباهت ها یا عدم شباهت ها.
- کاهش حجم یا بعد.

۳,۴. ایجاد مدل بر اساس خوشه بندی

- در این مرحله بر اساس داده های وارد شده به نرم افزار Weka، مدل خوشه بندی بر اساس الگوریتم K-Means ساخته شده است. شکل(۲).

شکل (۲) مدل به دست آمده بر اساس خوشه بندی

Attribute	Cluster#							
	Full Data (269)	0 (32)	1 (96)	2 (39)	3 (27)	4 (30)	5 (27)	6 (18)
father-t	rudiment	Diplom	rudiment	Diplom	Diplom	Diplom	rudiment	rudiment
mother -t	rudiment	rudiment	rudiment	Diplom	Diplom	Diplom	rudiment	lowbrow
No of your sis&bro	2.0	3.0	2.0	2.0	1.0	3.0	4 and high	4 and high
child-family	1.0	3.0	1.0	2.0	1.0	3.0	3.0	4 and lesser
revenue	medium	medium	medium	medium	medium	medium	medium	medium
financial-problem	No	No	No	No	No	No	Yes	No
Live	Father and mother	Father and mother	Father and mother	Father and mother	Father and mother	Father and mother	Father and mother	Father and mother
Worked in 1 school	No	No	No	No	No	No	No	No
Worked in Honar	No	No	No	No	No	No	No	No
Lived Father	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg in 1 school	16-18	16-18	16-18	14-16	12-14	14-16	16-18	14-16
career and art	18-20	16-18	18-20	18-20	18-20	16-18	16-18	16-18
mathematic No.	16-18	16-18	16-18	16-18	14-16	12-14	14-16	14-16
physic No	14-16	18-20	16-18	14-16	12-14	14-16	14-16	16-18
Avg of Diplom	14-16	16-18	16-18	14-16	12-14	14-16	14-16	14-16
your friend in classmate	No	No	Yes	Yes	No	No	No	No
Sincere Friend	Fani	Fani	Fani	Fani	Fani	Fani	nazari	Fani
Interest	Much	Much	Much	Much	Much	Much	medium	Much
computer at home	Yes	No	Yes	No	No	Yes	No	Yes
daily worked by computer	1-2	1-2	1-2	1-2	1-2	2-3	1-2	3-4
Which action with computer	Internet	Graphic	Internet	Game	Internet	Game	Game	Game
Council	No	No	No	No	little	little	medium	medium
your parent in your choice	Much	medium	Much	Much	Much	Much	medium	little
The member of your family	No	No	No	Much	Much	medium	medium	No
The member of your family that studied	No	No	No	No	No	No	No	No
the efficacy in deciction	No	No	No	Much	No	medium	No	little
marketing	medium	medium	No	Much	No	medium	Much	medium
haphazard	No	No	No	No	No	No	No	No
the teachers in your chioce	No	No	No	medium	No	little	little	No
high - level	Much	Much	Much	Much	Much	Much	Much	little
supplementary	No	No	No	No	No	Yes	No	No
Application	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Good name of the computer	No	Yes	No	No	No	Yes	No	No

فنی و حرفه ای ، علاقه مندی فراوان به رشته کامپیوتر ، داشتن کامپیوتر در خانه قبل از ورود به رشته کامپیوتر ، مشغول بودن با کامپیوتر در اینترنت روزانه به مدت ۲ ساعت ، عدم نقش دبیر مشاوره در انتخاب رشته کامپیوتر ، حمایت فراوان توسط والدین در انتخاب رشته کامپیوتر ، عدم توصیه سایر اعضای خانواده و آشنایان جهت انتخاب رشته کامپیوتر ، عدم تحصیل یکی از اعضای خانواده در رشته کامپیوتر ، عدم تأثیر سایر اعضای خانواده در تصمیم گیری جهت انتخاب رشته کامپیوتر ، عدم انتخاب رشته کامپیوتر جهت بازار کار آن ، اتفاقی نبودن انتخاب رشته کامپیوتر ، عدم تأثیر مدرسان دوران قبل از ورود به رشته کامپیوتر در علاقه مندی به این رشته ، داشتن تمایل زیاد جهت ادامه تحصیل به درجات بالاتر ، عدم انتخاب رشته کامپیوتر بخاطر داشتن گرایش های مختلف ، انتخاب رشته کامپیوتر بخاطر کاربرد فراوان آن ، مهم نبودن خوشنامی رشته کامپیوتر.

• فارغ التحصیلانی که با مشخصات زیر هستند، در یک خوشه قرار میگیرند و به عنوان دانش آموزان نا موفق، با معدل دیپلم ۱۴-۱۲ شناخته میشوند :

میزان تحصیلات پدر و مادر دیپلم ، تعداد خواهر و برادر ۱ نفر ، اولین فرزند خانواده ، با درآمد ماهیانه متوسط ، نداشتن مشکل مالی ، زندگی نزد پدر و مادر ، شاغل نبودن در اول دبیرستان و دوران هنرستان ، پدر در قید حیات ، دارای معدل اول دبیرستان ۱۴ - ۱۲ ، نمره درس حرفه و فن ۲۰-۱۸ ، نمره درس ریاضی ۱۶-۱۴ ، نمره درس فیزیک ۱۴-۱۲ ، با دوستانش همکلاسی نباشد ، تحصیل دوستان صمیمی در شاخه فنی و حرفه ای ، علاقه مندی زیاد به رشته کامپیوتر ، نداشتن کامپیوتر در خانه قبل از ورود به رشته کامپیوتر ، مشغول بودن با کامپیوتر در اینترنت روزانه به مدت ۲ ساعت (در کافی نت و یا با دوستان) ، نقش کم دبیر مشاوره در انتخاب رشته کامپیوتر ، حمایت فراوان توسط والدین در انتخاب رشته کامپیوتر ، توصیه زیاد سایر اعضای خانواده و آشنایان جهت انتخاب این رشته ، عدم تحصیل یکی از اعضای خانواده در رشته کامپیوتر ، عدم تأثیر سایر اعضای خانواده در تصمیم گیری جهت انتخاب این رشته، بی اهمیت بودن بازار کار رشته کامپیوتر ، اتفاقی نبودن انتخاب این رشته ، تأثیر متوسط مدرسان دوران قبل از ورود به رشته کامپیوتر در علاقه مندی به این رشته ، داشتن تمایل زیاد جهت ادامه تحصیل به درجات بالاتر ، عدم انتخاب رشته کامپیوتر بخاطر داشتن گرایش های مختلف آن ، انتخاب رشته کامپیوتر بخاطر کاربرد فراوان آن ، مهم نبودن خوشنامی رشته کامپیوتر.

۵.۴. نتیجه خوشه بندی

در اینجا نتیجه این خوشه ها بصورت مقایسه ای در یک جدول آورده شده است. جدول شماره ۲.

مهمترین مشخصات	فارغ التحصیلان موفق	فارغ التحصیلان ناموفق
تحصیلات والدین	زیر دیپلم	دیپلم
تعداد برادر و خواهر	۲	تک فرزند
مشکل مالی	خیر	خیر
شاغل بودن در دوران متوسطه	خیر	خیر
معدل اول دبیرستان	۱۶ - ۱۸	۱۲ - ۱۴
نمره درس ریاضی اول دبیرستان	۱۶ - ۱۸	۱۴ - ۱۶
نمره درس فیزیک اول دبیرستان	۱۶ - ۱۸	۱۲ - ۱۴
همکلاسی با دوستان صمیمی	بله	خیر
میزان علاقه به رشته کامپیوتر	زیاد	زیاد
داشتن کامپیوتر قبل از ورود به این رشته	بله	خیر
میزان مشاوره با دبیر مشاور جهت انتخاب این رشته	عدم مشاوره	کم
میزان حمایت والدین	زیاد	زیاد
میزان توصیه سایر اعضای خانواده و آشنایان	عدم توصیه	زیاد

۵. نتیجه گیری

داده کاوی این امکان را میسر کرد تا در مورد داده های تحصیلی دانش آموزان با تکنیک هایی که استفاده کردیم به نتایج زیر برسیم که:

- سطح تحصیلات والدین اکثر فارغ التحصیلان موفق، زیر دیپلم بوده، تعداد برادر و خواهر ۲، معدل اول دبیرستانشان ۱۸ - ۱۶ بوده ، نمره درس ریاضی اول دبیرستانشان ۱۸ - ۱۶ بوده، نمره درس فیزیک اول دبیرستانشان ۱۸ - ۱۶ بوده، همکلاسی بودن با دوستان صمیمی، به میزان زیادی به رشته کامپیوتر علاقه دارند ، قبل از ورود به این رشته کامپیوتر داشته اند و بازار کار رشته کامپیوتر در انتخاب این رشته برایشان مهم نبوده است.
 - سطح تحصیلات والدین اکثر فارغ التحصیلان نا موفق، دیپلم بوده ، معدل اول دبیرستانشان ۱۴ - ۱۲ بوده ، نمره درس ریاضی اول دبیرستانشان ۱۶ - ۱۴ بوده ، همکلاس نبودن با دوستان صمیمی، نمره درس فیزیک اول دبیرستانشان ۱۴ - ۱۲ بوده، به میزان فراوان به رشته کامپیوتر علاقه دارند ، قبل از ورود به این رشته کامپیوتر نداشته اند و بازار کار رشته کامپیوتر در انتخاب این رشته برایشان مهم نبوده است.
- از عوامل مهم این سیستم، میزان تحصیلات والدین، تعداد خواهر و برادر، معدل سال اول دبیرستان، نمرات دروس ریاضی و فیزیک، همکلاس بودن با دوستان صمیمی، داشتن کامپیوتر قبل از ورود به این رشته و میزان توصیه سایر اعضای خانواده جهت انتخاب این رشته میباشد.

۶. فهرست منابع

- [۱] مهدی غضنفری، سمیه عزیزاده، بابک تیمورپور، داده کاوی و کشف دانش، ۱۳۸۷، انتشارات دانشگاه علم و صنعت ایران.
- [۲] محسن صفاریان، اعظم خیر آبادی، پیش بینی رتبه داوطلبان کنکور سراسری دانشگاهها با استفاده از داده کاوی، چهارمین کنفرانس داده کاوی ایران، دانشگاه صنعتی شریف، ۱۰ آذر ۱۳۸۹، تهران.
- [۳] مرجان عظیمی، تحلیل خرید و فروش در مشاور املاک با داده کاوی، دانشگاه آزاد کرج، بهمن ۱۳۸۹.

کنفرانس داده کاوی ایران