پیش بینی نمرات دانشجویان بر اساس زمان آزمون و فرصت مطالعاتی

محسن فرهادی[1]، مریم چوپانی[2]

چکیده

رشد روز افزون اطلاعات در سازمان‌ها و شرکت‌ها موجب افزایش پیچیدگی روند نگهداری، بازیابی و تحلیل اطلاعات شده است. داده کاوی یا کشف دانش در پایگاه داده‌ها، راهبرد جدیدی است در راستای کشف محتوی اطلاعات معتبر پنهان در پایگاه‌های داده.

مدیریت، بازیابی، نگهداری و تجزیه و تحلیل داده‌ها برای پیدا کردن الگوهای تازه و کاربردی در پایگاه داده‌های بزرگ از این طریق میسر می‌شود.

اغلب ارگانها به تصمیم گیری‌های استراتژیک و اتخاذ خط مشی‌های جدید برای عملکرد بهتر نیاز دارند. در واقع داده کاوی اطلاعاتی را در اختیار ما قرار می‌دهد، که برای گرفتن تصمیم هوشمندانه در انبوهی از اطلاعات به آن نیاز است.ابزار داده کاوی، داده را می گیرد و یک دید کلی از واقعیت را به شکل مدل می سازد، این مدل روابط موجود در داده‌ها را شرح می دهد.

داده کاوی در امر آموزش یکی از راهبردهای مطرح در راستای بهبود و توسعه اهداف آموزشی است. مطالعه و تحلیل اطلاعات در بخش آموزش عالی، عملکرد و کارایی روال‌های اجرایی را مورد سنجش قرار می‌دهد، در صورت لزوم امر پیش بینی را میسر می‌سازد و زمینه و ایده لازم برای اصلاح و بهبود بهره وری را فراهم می‌آورد.

یکی از موارد کاربردی در این زمینه، پیش بینی عملکرد دانشجویان بر اساس داده‌های موجود در پایگاه اطلاعاتی سیستم آموزشی، از قبیل زمان آزمون، فرصت مطالعاتی و .. می‌باشد. ما در این مقاله تأثیر جنبه‌های مختلف آزمون را در نمره‌ی اکتسابی دانشجویان در دروس مختلف را مورد بررسی قرار داده‌ایم.

واژه‌های کلیدی

داده‌کاوی، آموزش عالی، آزمون، پیش‌بینی نمرات

۱.عضو هیئت علمی دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود narfarhadi@yahoo.com

۲.فارغ التحصیل دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود maryamchoopany@yahoo.com

Students' scores prediction based on the exam date and the preparation time

Mohsen Farhadi[1]Maryam Choopany[2]

Abstract

Daily increasing amount of data in organizations and companies, the process of data storage, analysis and recovery is becoming more complex than ever. Data-mining or knowledge exploration in database is a new procedure in finding hidden valid knowledge and it also makes management, recovery, storage, and analysis of data to find new and practical patterns in database possible. Almost all companies and organizations need to make strategic decisions and determine new procedures to promote their actions and that is when data-mining comes in to provide us with required data to do so, among tons of data.

As a matter of fact, by means of data-mining, data is chosen and analyzed to give out a general perspective of facts, in the form of a model, explaining the relationships among the data. The study and analysis of data in higher education can evaluate the practicality of the procedures in use, make the necessary predictions possible, and can also provide us with ideas to reform the procedures and increase the efficiency.

One of the implements of data-mining is the prediction of students' performance at universities based on the existing data in the database of educational system such as the date of the exam, preparation time of the exam, etc. In this article, the relationship between different aspects of an exam and the students' scores has been studied.

Keywords

Data-mining, higher education, exams, score prediction

۱. مقدمه

پیش‌بینی آینده در زمینه‌های مختلف همواره برای انسان جالب و جذاب بوده است. با اطمینان می‌توان گفت که پیش‌بینی آینده و روند تغییرات در همه‌ی حوزه‌ها از دغدغه‌های اصلی و همیشگی مدیران سطح بالا و میانی می‌باشد. اما همواره مشکلات فراوانی در برابر آن وجود داشته است که انجام پیش‌بینی‌های دقیق و قابل اعتماد را تقریباً غیر ممکن نموده است. وجود پارامترهای زیاد و در بسیاری موارد پنهان اینگونه موارد را به مسائلی بسیار پیچیده تبدیل نموده است که الگوریتم‌های غول‌پیکر ریاضی نیز از ارائه راهکاری مناسب برای ساخت یک مدل پیش‌بینی کارامد عاجز مانده‌اند. [۲]

امروزه در اکثر دانشگاه‌های ایران اسلامی بانک‌های اطلاعاتی وسیعی از ویژگی‌های دانشجویان موجود است که حجم بالایی از اطلاعات مربوط به سوابق آموزشی، تحصیلی و ... را شامل می‌شود. در سال‌های اخیر با توجه به مشکلات موجود در زمینه استخراج دانش از چنین پایگاه‌داده‌های عظیمی و استفاده از آنها به منظور تصمیم گیری‌های مناسب، تکنیک‌های داده کاوی و استفاده از آنها در امر آموزش مورد توجه قرار گرفته است. [۱] هدف پژوهش‌هایی این چنین نیز استخراج قوانین و الگوهایی قابل استنتاج و مرتبط از روی پایگاه داده دانشگاه به منظور بهبود کیفیت امور در آموزش عالی می‌باشد.

از جملهٔ تحقیقات مشابه انجام شده در این زمینه می‌توان مواردی همچون به‌کارگیری ابزارهای داده‌کاوی جهت پیش بینی موفقیت دانشجویان در فارغ التحصیلی توسط فرهادی و عطفایی به صورت ترمی [۱]، پیش‌بینی میزان موفقیت و عدم موفقیت دانشجویان رشته فناوری اطلاعات در فارغ‌التحصیلی، بر اساس معدل، توسط توحیدی مقدم و فرهادی [۲] که هر دو روش از ابزار درخت تصمیم استفاده کرده اند، را نام برد.

در یک نظرسنجی از داده کاوان در مورد اینکه در ۱۲ ماه اخیر از چه زبان برنامه نویسی برای تحلیل داده و داده کاوی استفاده کرده‌اید، زبان sql با ۳۲٪ در رتبه دوم قرارگرفت[۵]. در نظرسنجی سالیانه دیگر از داده‌کاوان در مورد موضوع کاری داده کاوی، در سال ۲۰۱۲ نیز، کاربرد داده کاوی در امر آموزش با اختصاص ۱۶٫۲٪ در رتبه پنجم قرار گرفته است [۵] که نشان دهنده آن است که بحث آموزش زمینه-ای است پرطرفدار و نیازمند تحقیق. بدین ترتیب این تحقیق با انتخاب زبان پرکاربرد SQL و موضوع آموزش عالی با بررسی تاثیر عوامل جانبی آزمون بر نمره اکتسابی آغاز شد، در ادامه اطلاعات گرد آوری شده و در پایگاه داده یکپارچه و اصلاح شد، سپس دسته هایی از اطلاعات برای تست و تحلیل به صورت تصادفی انتخاب شد و به منظور ایجاد یک الگو و سنجش آن تعیین شد، پس از آن نتایج تحلیل در حالت های مختلف بررسی شد، که نشان دهنده عوامل موثر در نتیجه مطلوب دانشجویان است.

۲. آماده سازی داده‌ها و پیش بینی نمرات

پس از انتخاب زبان اطلاعات مورد نظر از پایگاه داده دانشگاه دریافت شد، اطلاعات در مورد دانشجو، نمرات، گروه‌های درسی و درس‌ها در جداول جداگانه از پایگاه داده استخراج شد. در مرحله بعد با کمک نرم افزار SQL Server Management Studio ابتدا داده‌ها برای تحلیل آماده شدند. محاسباتی روی اطلاعات انجام شد و صفات مورد نیاز برای آنالیز برآورد شد و اطلاعات اضافی حذف شد. صفاتی همچون معدل متوالی دانشجو (معدل کل دانشجو از ترم اول تا ترم قبل از اخذ درس مورد پیش‌بینی) ، زمان امتحان (صبح یا ظهر بودن) برای شرکت در تحلیل، آماده شدند. صفاتی دیگر همچون معدل امتحان قبل در پایگاه داده (معیاری برای اندازه گیری درجه سختی امتحان قبلی دانشجو) ، فرصت مطالعاتی (فاصله زمانی امتحان تا امتحان قبلی) ، معدل نمره دهی استاد (میانگین نمرات ثبت شده در سیستم از یک استاد) و معدل درس(میانگین نمرات ثبت شده در سیستم از یک درس) محاسبه شدند، که تمامی صفات ورودی الگوریتم در بخش ضمایم در **Error! Unknown switch argument.** آمده است. زمانی که داده‌هابرای داده کاوی آماده شدند می‌توان وارد مرحله ی دوم از انجام داده کاوی شد.

با کمک نرم افزارSQL Server BusinessIntelligence Development Studio از Microsoft decision trees برای پیش بینی استفاده شد. در مجموع ۶۵۵۲۵ داده از پایگاه داده دانشگاه صنعتی شاهرود به فرایند تحلیل داده شد. ۹۰٪ از داده‌ها برای آموزش و تحلیل به صورت تصادفی انتخاب شد و۱۰٪ دیگر برای تست الگوی به دست آمده اختصاص داده شد.

۲-۱ درخت تصمیم

درخت تصمیم یکی از رایج‌ترین تکنیک‌های داده‌کاوی است. معمول‌ترین زمینه‌ی کاری درخت تصمیم، دسته‌بندی می‌باشد. الگوریتم درخت تصمیم مایکروسافت که توسط سرویس‌های خدمات تحلیلی ارائه شده است، تکنیک‌های دسته‌بندی و رگرسیون را پشتیبانی می‌نماید و برای مدل‌های پیشگویانه بسیار خوب عمل می‌کند. با استفاده از این الگوریتم مشخصه‌های گسسته و پیوسته قابل پیش‌بینی

هستند. مقصود اصلی در درخت تصمیم، تقسیم داده‌ها به صورت بازگشتی به زیرمجموعه‌هایی است؛ به گونه‌ای که هر زیرمجموعه در برگیرنده وضعیت همگنی از متغیر هدف می‌باشد. زمانیکه پردازش بازگشتی کامل شد، درخت تصمیم شکل گرفته است [۴].

۲-۱-۱ ویژگی‌ها و مزایای درخت تصمیم مایکروسافت

- از مزایای استفاده از درخت تصمیم نسبت به سایر الگوریتم‌های داده‌کاوی این است که مدل آن سریع ساخته و آسان‌تر تفسیر می‌شود [۴].
- پیش‌بینی‌ها بر اساس درخت تصمیم مؤثرتر است [۴].
- روش‌های مختلفی برای رشد یک درخت وجود دارد و می‌توان از فرمول‌های متنوعی برای تعیین چگونگی تقسیم درخت استفاده نمود [۴].
- بر پایه تنظیم پارامتری می‌باشد یعنی اینکه رشد درخت بر اساس تعداد متغیرهای ورودی و با تنظیم پارامترها قابل کنترل است [۴].
- یکی از ویژگی‌های منحصر به فرد درخت تصمیم مایکروسافت این است که می‌تواند برای آنالیز وابستگی‌ها نیز به کار گرفته شود [۴].

پس از اجرای الگوریتم درخت تصمیم ماکروسافت و شبکه وابستگی‌ها به دست آمد. برای مثال یکی از برگ‌های درخت تصمیم برای دانشجویانی که معدل آنها بزرگتر ۱۲،۰۶۷ و کوچکتر از ۱۴،۲۶۷ باشد به صورت تصویر ۱ بدست آمد که ۶۰۵۶ مورد در این شرط صدق می‌کرد، به این معنی که پیش بینی می‌شود مواردی که در این بازه معدل کل قرار دارند نمره درس آنها با فرمول بدست آمده برای grade در تصویر ۱ پیش‌بینی می‌شود.

جدول1

1. Gpa معدل متوالی دانشجو از ترم اول تا ترم قبل از اخذ درس مورد پیش‌بینی
2. Ostadavg معدل نمره دهی استاد
3. Crsnameavg معدل درس مورد پیش‌بینی در کل پایگاه داده
4. Grade نمره مورد پیش بینی یا نمره مورد انتظار

```
Existing Cases: 6056
Missing Cases: 0
GRADE = 13.362+0.961*(GPA-14.521)+0.665*(Ostadavg-13.927)+0.649*(Crsnameavg-13.367)
```

تصویر ۱

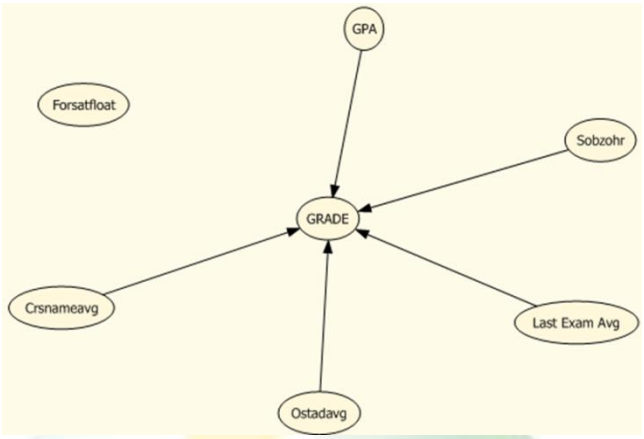
۲-۲ شبکه وابستگی‌ها

شبکه وابستگی‌ها معرف قدرت مشخصه‌های ورودی و پیش‌بینی کننده در پیش‌بینی متغیر هدف می‌باشد. هر نود در شبکه وابستگی‌ها بیانگر یک مشخصه و هر بردار نشانگر وجود یک رابطه بین دو نود می‌باشد. هر بردار دارای یک جهت از سمت مشخصه‌ی ورودی به سمت مشخصه‌ی قابل پیش‌بینی می‌باشد.

استفاده از این نمودار زمانی مفید است که تعداد زیادی مشخصه‌ی قابل پیش‌بینی وجود داشته باشد. این نمودار در تحلیل‌های اکتشافی داده‌ها بسیار سودمند است [۴].

۳. تحلیل پیش‌بینی‌ها

پس از بدست آمدن درخت تصمیم از روی ۹۰٪ از داده‌ها که به صورت تصادفی انتخاب شدند، پیش‌بینی با ۱۰٪ داده‌های باقی‌مانده انجام شدتا دقت عمل درخت تصمیم بررسی شود، در نتیجه **Error! Unknown switch argument.** باscore قابل قبول 1.44 بدست آمد که نشان می‌دهد پیش بینی درخت تصمیم به خوبی انجام شده است ، سپس شبکه وابستگی‌ها به صورت Error! Unknown switch argument. حاصل شد.



نمودار۱

ترتیب صفات موثر بر نمره نهایی به صورت جدول2 بدست آمد که همانطور که آمده است، معدل نمره دهی استاد اولین و موثرترین صفت درتعیین نمره بدست آمده است.

جدول2

نمره مورد پیش بینی یا نمره مورد انتظار Grade
1. Ostadavg معدل نمره دهی استاد
2. Gpa معدل متوالی دانشجو از ترم اول تا ترم قبل از اخذ درس مورد پیش‌بینی
3. Crsnameavg معدل درس مورد پیش‌بینی در کل پایگاه داده
4. Sobzohr صبح یا ظهر بودن زمان امتحان
5. Last exam avg معدل امتحان قبل
6. Forsat float یا فرصت مطالعاتی فاصله زمانی امتحان تا امتحان قبل

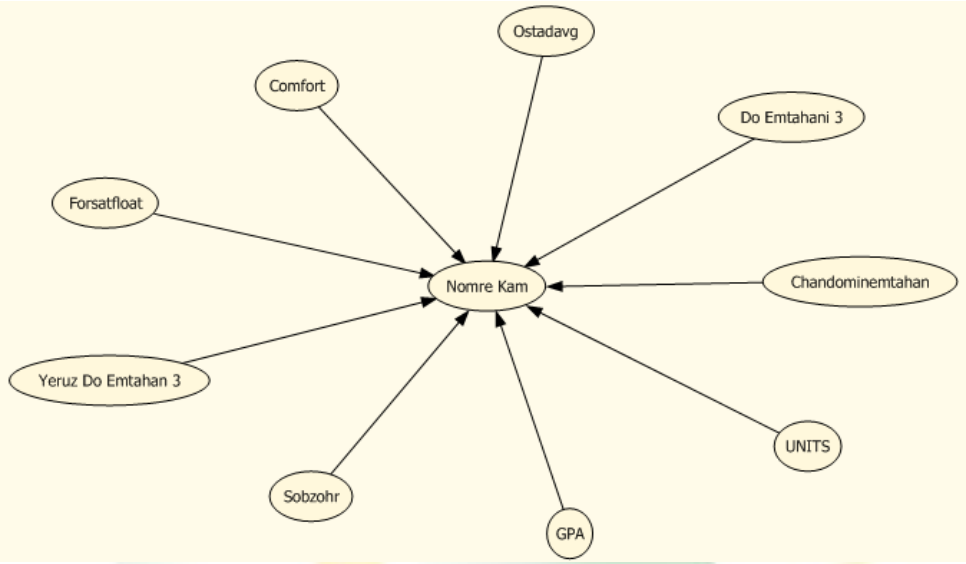
باقی صفات ورودی که در **Error! Unknown switch argument.** آمده‌اند اما در شبکه وابستگی **Error! Unknown switch argument.** پدیدار نشدند حاکی از این مطلب است که الگوریتم این صفات را در پیش‌بینی بی‌تاثیر دانسته است.
در طی اجرای پروژه بر آن بودیم تا رابطه نمره و دو امتحانی بودن را بیابیم اما به این هدف نرسیدیم لذا موضوع پیش بینی را به کسری نمره یعنی تفاضل نمره از ۲۰ تغییر دادیم تا ببینیم چه عواملی موجب نتیجه نامطلوب دانشجویان می‌شود.

۴. اجرای دوباره آنالیز با کسری نمره

همانظور که پیش‌تر گفته شد در این بخش متغیری به نام کسری نمره تعریف شد تا تاثیر عناصری چون دو امتحان در یک روز بر نمره سنجیده شود. با حذف معدل امتحان قبلی و معدل درس (crsnameavg , last exam avg) و اکتفا به دسته بندی چهار سطحی

انجام شده برای همین صفات (در متغیرcomfort last examو comfort) الگوریتم اجرا شد. ورودی‌های الگوریتم در این بخش در **Error! Unknown switch argument.** آمده است و دو صفت حذف شده، با علامت ضربدر در جلوی آنها، مشخص شده اند.

نتیجه این تغییر، یعنی حذف دو متغیر پیوسته‌ی avg و اکتفا به بازه‌بندی در۵ سطح، کاهش تعداد سطوح و تعداد برگها به معنی دسته بندی بهتر وافزایش score را به همراه داشت و مشاهده شد که عوامل تاثیر گزار تغییر کرد.



نمودار۲

ترتیب تاثیرات در dependency network بر روی کسری نمره (nomrekam) درجدول ۳ آمده است

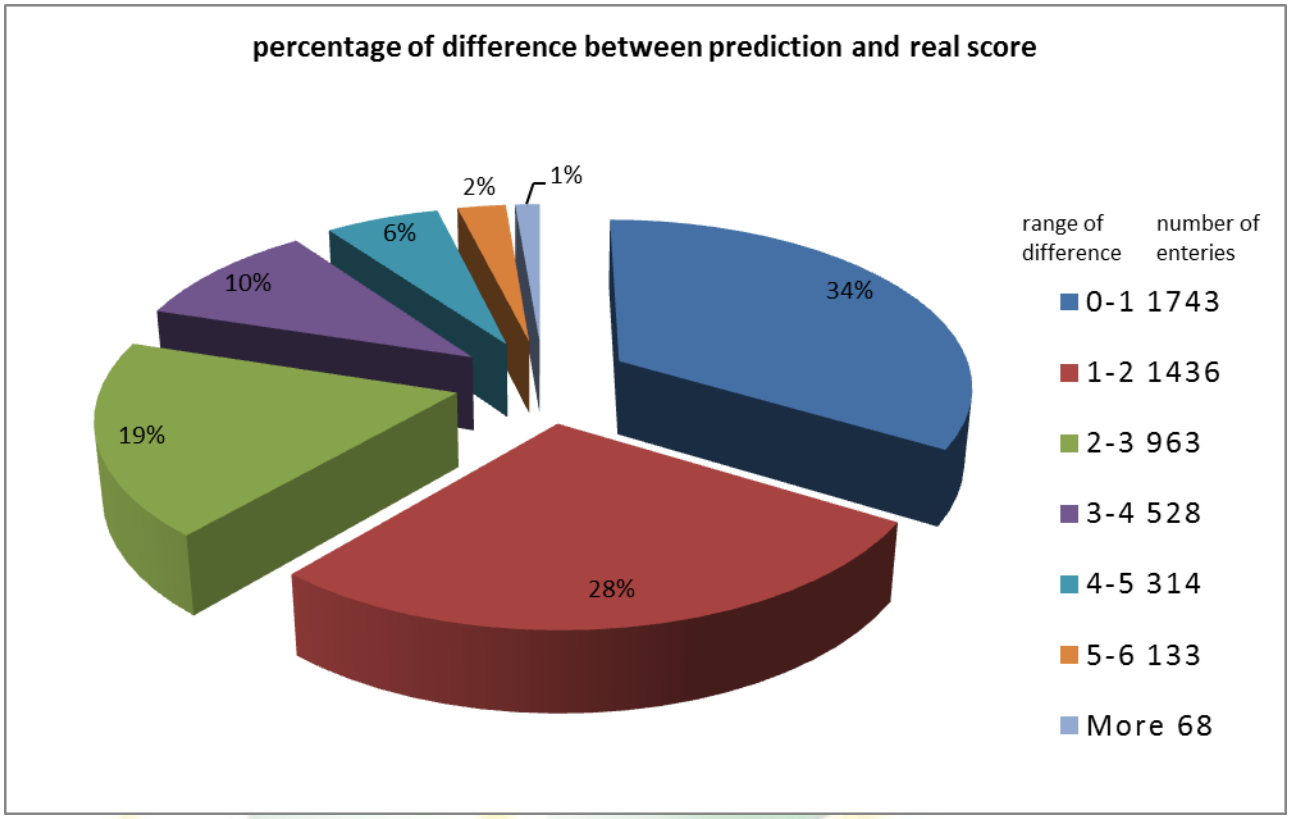
جدول ۳

1. Ostadavg معدل نمرات ثبت شده توسط هر استاد
2. Gpa معدل متوالی دانشجو
3. Comfort درجه راحتی یک درس
4. Units تعداد واحدهای یک درس
5. Chandominemtahan تعداد امتحانهایی که یک دانشجو تا یک روز داشته است
6. Yeruz do emtahan دانشجو در یک روز دو امتحان در دو ساعت متفاوت دارد
7. Forsatfloat فرصت مطالعاتی یا فاصله زمانی با امتحان قبلی
8. Do emtahani نشان می‌دهد دانشجو در یک روز، امتحانی دیگر، در ساعت یکسان با درس مورد پیش بینی دارد
9. Sobzohr صبح یا ظهر بودن زمان امتحان

۵. گزارش گیری و سنجش نتایج

پس از آنالیز گزارش گرفته شد و بدست آمدن الگوها ، برای ۱۰٪ از داده ها (۵۱۸۵ داده) که از قبل برای تست جدا شده بودند پیش بینی دوباره انجام شد.لازم به ذکر است که این داده ها از ابتدا جدا شده بودند و در جریان بدست آمدن الگو نقشی نداشتند، پس از آن نمرات پیش‌بینی شده با نمرات واقعی در مقابل هم قرار گرفتند و تفاوت آنها محاسبه شد.برای مثال درتصویر ۲ خطای پیش بینی و

تعداد آن در بازه های متفاوت آمده است که چه تعداد از ۵۱۸۵ پیش‌بینی، خطایی بین صفر تا یک داشته اند وچند درصد از کل را تشکیل می‌دهند. همانطور که مشخص است بیشترین نسبت مربوط به خطاها با تفاضل کمتر از یک است.



تصویر ۲ خطای پیش بینی و تعداد آن در بازه های متفاوت

۶. نتیجه‌گیری

با توجه به تاثیر قابل توجه اساتید در نمره دانشجویان دریافت می‌شود که نحوه تدریس و نمره دهی استاد به طور مستقیم بر دو عنصر نمره‌ی دانشجو و کسری نمره از بیست موثر است، لذا توجه به این موضوع از جهات مختلف، موضوعی است که باید بیش از اینها مورد بررسی قرار گیرد. پیشنهاد می‌گردد برای هر استاد معیاری به نام معدل استاد ایجاد و نهادینه شود تا مقایسه و بررسی عملکرد اساتید تسهیل شود چراکه این معیار اولین ملاک موثر چه درکسب نمره و چه درنتیجه منفی دانشجویان در آزمونمی‌باشد. فرصت مطالعاتی، داشتن دو امتحان در یک روز و یک ساعت عوامل دیگر در نتیجه منفی دانشجو در امتحانات بود که ناشی از چینش نامناسب برنامه امتحانی است، لذا برای بازدهی بیشتر و برنامه ریزی صحیح امتحانات بایستی تلاش کرد تا این امور با بهره گیری از هوش مصنوعی مکانیزه شود.

۷. ضمایم

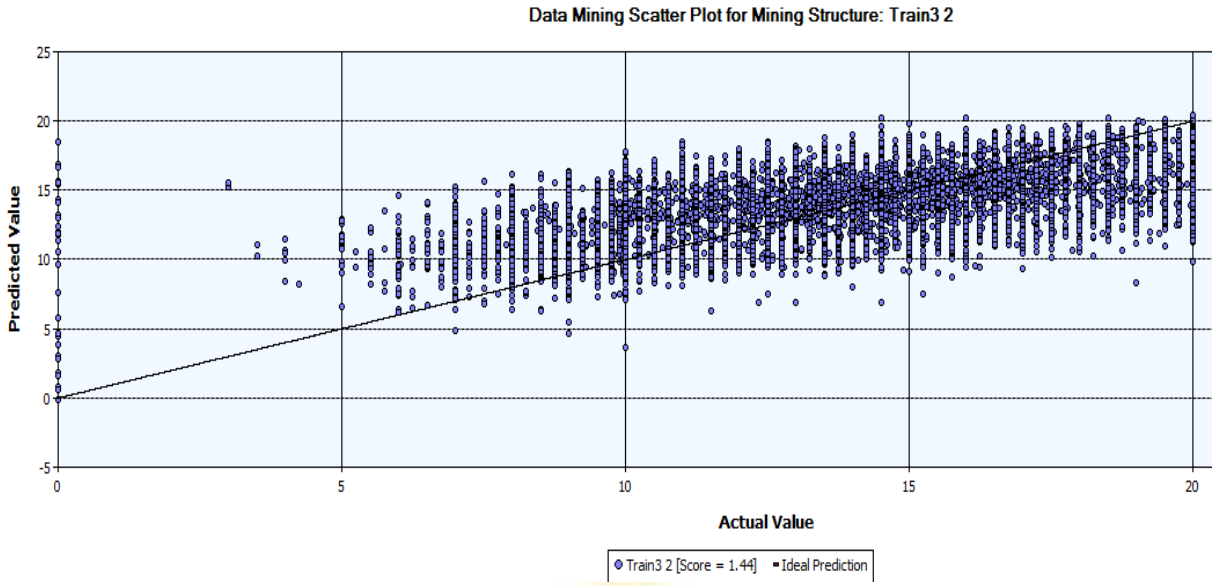
Mining model structure:

	Content Type	Data Type
Chandominemtahan	Ordered	Long
Comfort	Ordered	Long
Crsnameavg	Continuous	Double
Do Emtahani 3	Discrete	Text
Forsatfloat	Continuous	Double
GPA	Continuous	Double
GRADE	Continuous	Double
Key Id Trm Crs	Key	Text
Last Exam Avg	Continuous	Double
Last Exams Comfort	Discretized	Long
Ostadavg	Continuous	Double
Sobzohr	Discrete	Text
Unit	Discrete	Text
Yeruz Do Emtahan 3	Discrete	Text

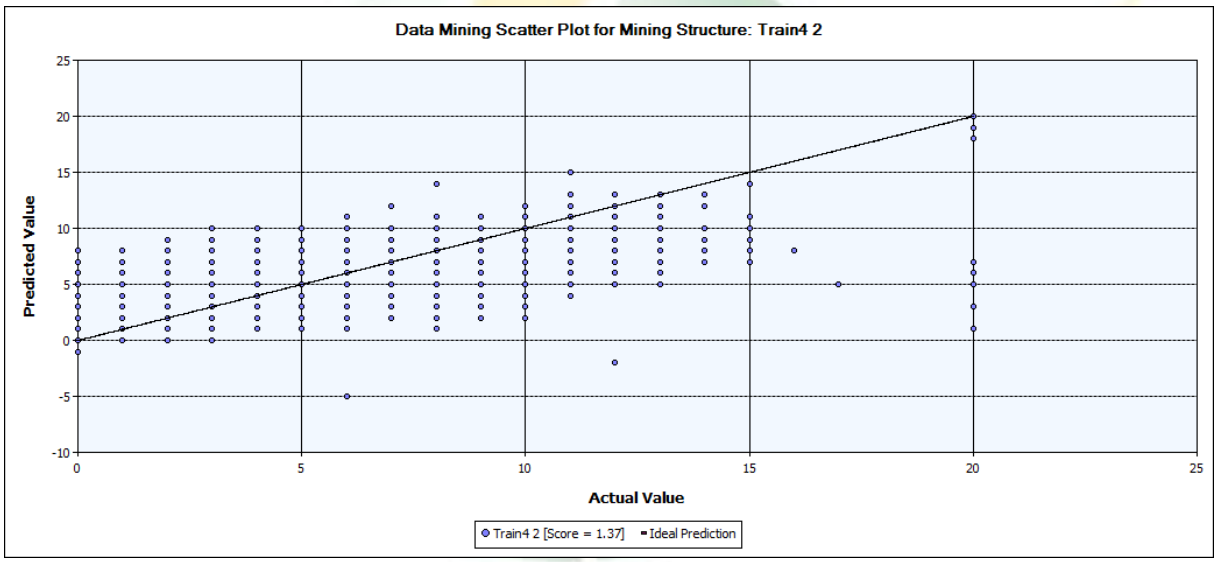
تصویر ۲ صفات ورودی الگوریتم و نوع داده ای آنها

Structure	Train4 2
	Microsoft_Decision_Trees
Chandominemtahan	Input
Comfort	Input
Crsnameavg	Ignore X
Do Emtahani 3	Input
Forsatfloat	Input
GPA	Input
Key Id Trm Crs	Key
Last Exam Avg	Ignore X
Last Exams Comfort	Input
Nomre Kam	PredictOnly
Ostadavg	Input
Sobzohr	Input
UNITS	Input
Yeruz Do Emtahan 3	Input

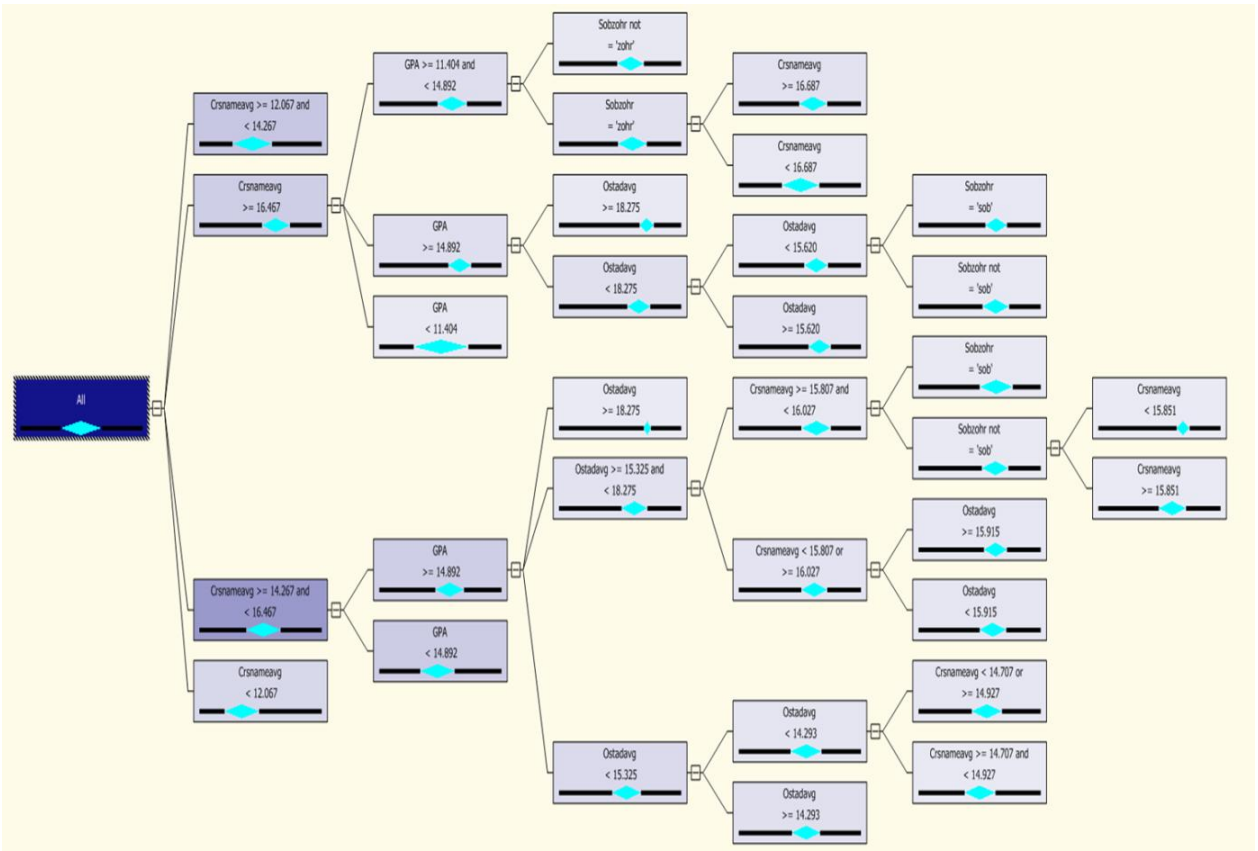
تصویر ۳



نمودار۳



نمودار۴



تصویر۴ درخت تصمیم

۸. مراجع

[۱] فرهادی، محسن؛ عطفائی، مهسا ؛ **" به‌کارگیری ابزارهای داده‌کاوی جهت پیش بینی موفقیت یا عدم موفقیت دانشجویان رشته کامپیوتر در فارغ التحصیلی "**، کنفرانس داده کاوی ایران، دوره ششم، دانشگاه صنعتی امیرکبیر، ۱۳۹۱.

[۲] توحیدی مقدم، مریم؛ فرهادی،محسن ؛ **" پیش‌بینی میزان موفقیت و عدم موفقیت دانشجویان رشته فناوری اطلاعات در فارغ‌التحصیلی "**، کنفرانس داده کاوی ایران، دوره ششم، دانشگاه صنعتی امیرکبیر، ۱۳۹۱.

[۳] نوری، بهاره؛ مقصودی، بهروز؛ شیخ احمدی، سیدامیر؛ **"پیداکردن دروس مرتبط از طریق الگوریتم سبد خرید و تاثیر آن در پیش بینی نمره از طریق الگوریتم درخت تصمیم"** چهارمین کنفرانس داده‌کاوی، دانشگاه صنعتی شریف،۱۳۸۹.

[۴] شهرابی، جمال؛ شکورنیاز، ونوس؛ **داده کاوی در SQL Server** (اولین مرجع داده کاوی کاربردی به زبان فارسی)، انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر، تهران، اول، ۱۳۸۸.

[۵] http://www.kdnuggets.com/