

# Academic guidance of students based on cumulative hierarchical clustering algorithms (Case study: Junbazan High School, Sirvan, Ilam Province)

Mohammad reza karamolahi      rezakaramolahi1397@gmail.com  
Hossein Baloochian      hossein.baloochian@gmail.com

## Abstract

The purpose of this article is to present a model of academic guidance for first secondary students based on cumulative hierarchical clustering algorithms. Our goal in this article was to simultaneously consider the basic courses of each field along with the skills needed by students to succeed in studying in different fields. This research is of applied type. The statistical population of the research is 30 students of Zhanbazan high school in Sirvan city of Ilam province. Data is collected through the scores of students' report cards and the scores assigned by the school counselor in the field of skills required for each field of study. The main algorithm of the research is the single link algorithm and modeling, data analysis has been done using spss software. By clustering all the data and research variables, first 10 clusters and finally 3 main clusters containing the most similar research samples were formed. . Examining and comparing the obtained clusters with the actual results of students' academic guidance showed a high degree of similarity. Comparing the results of clustering by the algorithm and the actual results of students' choice of fields showed 100% agreement in the field of experimental sciences, 77% in the field of experimental sciences, and 85% in the field of humanities.

**Keywords:** academic guidance, clustering algorithm, cumulative hierarchical clustering

## هدایت تحصیلی دانش آموزان بر اساس الگوریتم های خوشه بندی سلسله مراتبی تجمعی (مطالعه موردی: دبیرستان جانبازان شهرستان سیروان ایلام)

محمد رضا کرم اللهی<sup>۱</sup>

حسین بلوچیان<sup>۲</sup>

صص ۲۳ - ۳۶

تاریخ دریافت: ۱۴۰۲/۸/۲۰ تاریخ پذیرش: ۱۴۰۲/۱۰/۲۰

### چکیده

هدف این مقاله ارائه یک الگوی هدایت تحصیلی دانش آموزان مقطع متوسطه اول بر اساس الگوریتم های خوشه بندی سلسله مراتبی تجمعی است. هدف ما در این مقاله در نظر گرفتن همزمان دروس پایه هر رشته در کنار مهارت های مورد نیاز دانش آموزان برای موفقیت در تحصیل در رشته های مختلف بود. این پژوهش از نوع کاربردی می باشد. جامعه آماری پژوهش ۳۰ نفر از دانش آموزان دبیرستان جانبازان شهرستان سیروان ایلام می باشد. جمع آوری داده ها از طریق نمرات کارنامه های دانش آموزان و نمرات اختصاص داده شده مشاور مدرسه در زمینه مهارت های مورد نیاز هر رشته تحصیلی است. الگوریتم اصلی پژوهش الگوریتم single link و مدلسازی، تجزیه و تحلیل داده ها با استفاده از نرم افزار spss انجام شده است. با خوشه بندی همه داده ها و متغیرهای پژوهش ابتدا ۱۰ خوشه و در نهایت ۳ خوشه اصلی شامل مشابه ترین نمونه های پژوهش شکل گرفت. بررسی و مقایسه خوشه های بدست آمده با نتایج واقعی هدایت تحصیلی دانش آموزان تقریب بالایی از مشابهت را نشان داد. مقایسه نتایج حاصل از خوشه بندی توسط الگوریتم و نتایج واقعی انتخاب رشته دانش آموزان در مورد رشته تحصیلی علوم تجربی ۱۰۰ درصد، در مورد رشته علوم تجربی ۷۷ درصد و در مورد رشته تحصیلی علوم انسانی ۸۵ درصد مطابقت را نشان داد.

**کلیدواژه‌ها:** هدایت تحصیلی، الگوریتم خوشه بندی، خوشه بندی سلسله مراتبی تجمعی

<sup>۱</sup> دانشجوی کارشناسی ارشد فناوری اطلاعات دانشگاه غیرانتفاعی سفیردانش ایلام

<sup>۲</sup> دکتری رایانه، استادیار دانشگاه غیرانتفاعی سفیردانش ایلام

## مقدمه

موضوع هدایت تحصیلی در زندگی افراد به ویژه دانش آموزان برای یافتن مسیر صحیح، کشف و ظهور استعدادها و فعلیت بخشیدن مهارت ها نقش بسیار مهمی دارد. از طرفی پیچیده تر شدن روابط و تحولات جامعه، توسعه مشاغل، رشته های تحصیلی و تخصصی شدن آنها، تحصیلات آموزش عالی و اهمیت انتخاب شغل متناسب با آموخته های تحصیلی همگی بر جایگاه مهم هدایت تحصیلی تاکید دارند. اجرای صحیح هدایت تحصیلی باعث تعالی نظام آموزشی و افزایش بازدهی سازمان ها و نهادها در ابعاد مختلف اقتصادی، اجتماعی، فرهنگی خواهد شد. میان هدایت تحصیلی کار آمد و هنجارهای اجتماعی، نوعی پیوستگی و رابطه معنادار وجود دارد، زیرا با اجرای صحیح هدایت تحصیلی دانش آموزان به سمت توانایی ها و مهارت های واقعی شان سوق داده می شوند و به هنجارهای مورد نظر جامعه و خانواده نزدیک می شوند.

تجربه اجرای هدایت تحصیلی در سال های قبل از سال تحصیلی ۱۳۹۴-۱۳۹۵ نشان می دهد که هدایت تحصیلی توانسته است به کارکردهای مطلوب خود دست یابد و ساختارهای بیرون از نظام آموزشی همچون اشتغال و بازار کار بیش از همه بر این فرآیند تاثیر گذاشته است. همین امر باعث شده تا در هر دهه تعداد زیادی از دانش آموزان به یک رشته تحصیلی روی آورده و پیامد آن نیز انباشت فارغ التحصیلان زیاد در آن رشته و خالی شدن سایر رشته ها و شاخه ها از فارغ التحصیلان علاقمند و مستعد می باشد.

## بیان مسئله

در این مقاله با استفاده از الگوریتم ها و تکنیک های داده کاوی و به طور خاص الگوریتم خوشه بندی سلسله مراتبی تجمعی و از طریق پیدا کردن شباهت ها در خوشه های به دست آمده بر اساس نمرات کسب شده دانش آموزان در زمینه دروس مهمی مانند ریاضی، زبان و علوم تجربی و نظرات و نمرات مشاور مدرسه در خصوص برخی از مهارت ها و توانایی های مورد نیاز دانش آموزان مانند مهارت تفکر، مهارت حل مسئله، مهارت تصمیم گیری، تفکر خلاقانه و تفکر انتقادی، همچنین در نظر گرفتن علایق خود دانش آموزان سعی در ارائه یک الگوی پیشنهادی در جهت بر طرف کردن ایراد های کنونی فرآیند هدایت تحصیلی، به منظور علمی و کارآمد شدن این فرآیند داشته ایم. نتایج حاصل از خوشه بندی ها منجر به گروه بندی دانش آموزان بر اساس نمرات کسب شده در دروس اصلی مورد نیاز هر رشته، مهارت های اصلی لازم برای رشته های گوناگون و شکل گیری خوشه های مورد نظر ما در پژوهش و در نتیجه ارائه یک الگوی پیشنهادی هدایت تحصیلی دانش آموزان بر اساس نتایج به دست آمده با کمترین خطا و حذف مداخلات انسانی و کارآمد تر شدن این فرآیند گردید.

## ضرورت و اهمیت موضوع

شاید بتوان گفت که پایان مقطع متوسطه اول و انجام فرآیند هدایت تحصیلی مهم ترین رویداد در زندگی دانش آموزان و حتی خانواده های آنها باشد که تا پایان عمر اثرات مثبت یا منفی این اتفاق در زندگی آنها اثرگذار است. الگو و فرآیند فعلی هدایت تحصیلی در نظام آموزشی ایران اشکال ها و ایراد های فراوانی دارد. از جمله اینکه نمره اصلی ترین ملاک برای گروه بندی دانش آموزان در گروه های تحصیلی است و مسائل بسیار مهمی مانند مهارت های مورد نیاز دانش آموزان برای تحصیل در رشته های مختلف، نظر مشاور و علایق دانش آموزان یا دیده نمی شود و یا به درستی ترتیب اثر داده نمی شود که این امر خود باعث بروز مشکلاتی می شود. به طور مثال دانش آموزانی که ضعیف ترین نمرات و مهارت های تحصیلی را دارند به سمت رشته های هنرستانی مانند الکترونیک و رایانه سوق داده می شوند در حالی که این رشته ها نیازمند مهارت کافی در زمینه ریاضی، آمار و زبان انگلیسی هستند. علاوه بر این نیاز به تغییر رشته های متعدد در طول سال تحصیلی، باعث لطمه زدن به وضعیت تحصیلی دانش آموزان و جریان کلی آموزش خواهد شد، این مورد نیز از دیگر اشکال های موجود در نظام فعلی هدایت تحصیلی است.

## اهداف کلی مقاله

- ۱- ارائه یک الگو و چارچوب پیشنهادی و نوین در زمینه هدایت تحصیلی
- ۲- ضرورت اصلاح و بازنگری در شیوه فعلی هدایت تحصیلی
- ۳- کاربست تکنیک ها و الگوریتم های داده کاوی در حل مسائل دنیای واقعی

۴- استفاده از تکنیک های علمی و دقیق در محاسبه خوشه ها و گروه ها و دقیق تر و علمی تر شده فرآیند هدایت تحصیلی

### چارچوب نظری

#### هدایت تحصیلی

بر اساس آیین نامه هدایت تحصیلی دانش آموزان مصوب شورای عالی آموزش و پرورش، هدایت تحصیلی عبارت است از فرآیند ارائه خدمات راهنمایی و مشاوره به دانش آموزان، تا ضمن آشنایی با استعدادها، علایق، توانایی ها و ویژگی های شخصیتی خویش، شاخه ها و رشته های تحصیلی، حرفه ها و مشاغل مورد نیاز جامعه را بشناسند و بر اساس آن، به صورت آگاهانه شاخه و رشته تحصیلی مناسب خود را انتخاب کنند.

#### داده کاوی

به استخراج اطلاعات نهان یا الگوها و روابط مشخص در حجم زیادی از داده ها در یک یا چند بانک اطلاعاتی بزرگ داده کاوی گفته می شود. بسیاری داده کاوی را مترادف واژه های رایج کشف دانش در پایگاه داده ها می دانند. داده کاوی، پایگاه ها و مجموعه حجیم داده ها را در پی کشف و استخراج، مورد تحلیل قرار می دهد.

#### خوشه بندی

فرآیند خوشه بندی از شاخه های داده کاوی و یادگیری بدون نظارت است. هدف از خوشه بندی کشف خودکار خوشه های موجود در نمونه های پژوهش است. در واقع خوشه همان نمونه های آموزشی نزدیک به هم و گروه های مشابه از اشیاء می باشد. معیارهای تعیین نزدیک بودن نمونه ها می تواند در هر پژوهش متفاوت باشد. در تجزیه و تحلیل خوشه یا خوشه بندی، گروه بندی مجموعه ای از اشیاء انجام می شود، این کار به این صورت است که اشیاء در یک گروه (به نام خوشه) در مقایسه با دیگر دسته ها (خوشه ها) مشابه تر هستند.

#### خوشه بندی سلسله مراتبی

یکی از روش های خوشه بندی نمونه ها خوشه بندی سلسله مراتبی است. در این روش یک سلسله مراتب مشخص از خوشه ها با هدف یادگیری تدریجی ایجاد می شود. دو رویکرد عمده در روش سلسله مراتبی وجود دارد. روش بالا به پایین یا تجزیه ای و روش پایین به بالا یا ترکیبی (تجمعی). در این مقاله از روش خوشه بندی سلسله مراتبی پایین به بالا یا تجمعی بهره گرفته شده است. روش های خوشه بندی سلسله مراتبی به دو دسته تقسیم می شوند:

- **تجمعی**: رویکرد این دسته «پایین به بالا» می باشد؛ با شروع از پایین، در هر مرحله دو خوشه با یکدیگر تجمیع شده و یک خوشه جدید تشکیل می دهند. خوشه های جدید در سطح های بالاتر قرار گرفته و این روند تکرار می شود.
- **تجزیه ای**: رویکرد این دسته «بالا به پایین» می باشد؛ با شروع از بالا، در هر مرحله یک خوشه به خوشه های کوچکتری تجزیه می شود که در سطح پایین تر قرار می گیرند.

#### خوشه بندی تک پیوندی

خوشه بندی تک پیوندی یکی از روش های خوشه بندی سلسله مراتبی است که با رویکرد از پایین به بالا یک سلسله مراتب از خوشه ها ایجاد می کند. خوشه بندی تجمعی که به نام تکنیک نزدیکترین همسایه نیز شناخته می شود برای اولین بار در سال ۱۹۵۱ توسط فلورک مطرح شد. مبنای کار این روش آن است که در هر مرحله از اجرا، دو خوشه ای که کمترین فاصله از یکدیگر را دارند با هم ادغام می کند و فاصله یک خوشه از خوشه ای دیگر برابر با کمترین فاصله از هر یک از اعضای آن خوشه با هر یک از اعضای خوشه دیگر تعریف می شود.

#### پیشینه موضوع

نجفی، افضلی و مرادی در پژوهش کاربرد داده کاوی در شناسایی عوامل موثر در افت تحصیلی دانش آموزان، با استفاده از الگوریتم های درخت تصمیم و خوشه بندی و قوانین انجمنی موضوع را به صورت یک مسئله بهینه سازی درآورده و به دنبال به دست آوردن چارچوبی با اطمینان بالا، عمومیت و قابلیت

درک بالابودند. الگوریتم پیشنهادی برای استخراج قوانین از مجموعه داده های افت تحصیلی از اطلاعات فردی ۲۵۶ دانش آموز مقطع هنرستان در شهر زنجان جمع آوری و مورد آزمایش قرار گرفت. نتایج نشان داد که الگوریتم  $z48$  درخت تصمیم با دقت ۰,۹۵، برای مجموعه داده افت تحصیلی و خوشه بندی داده ها با استفاده از الگوریتم K-means با ضریب اطمینان ۰,۹۵ و به دست آوردن قوانین مد نظر با ضریب اطمینان بالا با استفاده از قوانین انجمنی که از الگوریتم Apriori برای مجموعه داده ها اجرا و تولید شد. (نجفی، افصلی و مرادی، ۱۴۰۰)

کمرئی و منتظر در مقاله گروه بندی همسان یادگیرندگان در محیط یادگیری الکترونیکی به کمک روش خوشه بندی شورایی از ۵ روش خوشه بندی شورایی برای ترکیب نتایج خوشه بندی استفاده کردند. نتایج ارزیابی تجربی روش خوشه بندی پیشنهادی بر اساس ۳ شاخص (دیویس-بولدین)(خلوص-تجمع) و (واریانس) نشان می دهد که این روش کاهش هزینه محاسباتی و دقت و سرعت بیشتری نسبت به سایر روش های متداول در شناسایی گروه ها داشته است. (کمرئی و منتظر، ۱۳۹۴)

رضایی و منتظر در مقاله گروه بندی یادگیرندگان در سامانه های تطبیق یادگیری به کمک روش خوشه بندی پیوندی فازی، روشی مبتنی بر نظریه فازی برای انتخاب خوشه های مناسب در روش خوشه بندی پیوندی معرفی کردند. برای ارزیابی دقت روش پیشنهادی روش های k-means و fcm خوشه بندی فازی تکاملی، و شبکه عصبی مصنوعی انتخاب و نتایج خوشه بندی آنها با نتایج حاصل از روش پیشنهادی مقایسه شده است. از نتایج حاصل ملاحظه می شود که خوشه بندی در همه ابعاد سبک یادگیری روش خوشه بندی پیوندی فازی نتایج بهتری را نسبت به سایر روش ها داده است. (رضایی و منتظر، ۱۳۹۸)

هدایتی و جوانمرد در پژوهش دسته بندی دانش آموزان سال اول دبیرستان بر اساس نمرات سالانه با استفاده از داده کاوی آموزشی، با اعمال تکنیک های داده کاوی بر روی نمرات دانش آموزان سال اول دبیرستان مهدیه شهر یزد، دسته بندی مناسبی برای این دانش آموزان ارائه داده اند. داده های جمع آوری شده معدل کل، نمرات سالانه، نمرات مستمر و پایانی کسب شده توسط دانش آموزان سال اول دبیرستان دخترانه مهدیه شهر یزد می باشد. جهت کشف الگوهای پنهان در داده ها، تکنیک ها و الگوریتم های داده کاوی مورد استفاده قرار می گیرند. (هدایتی و جوانمرد، ۱۳۹۲)

دی پیر و رابو در پژوهش استفاده از داده کاوی آموزشی جهت گروه بندی یادگیرندگان در محیط یادگیری الکترونیکی به منظور شخصی سازی برنامه آموزش، مدلی به منظور گروه بندی یادگیرندگان بر اساس سبک یادگیری آنها ارائه کردند به این ترتیب که ابتدا با استفاده از پرسشنامه فلدر-سولومون ابعاد مختلف سبک یادگیری یادگیرندگان در قالب مدل فلدر - سولومون سنجیده شده، سپس با استفاده از الگوریتم K میانگین به گروه بندی آنها پرداخته شد. در نهایت مدل ارائه شده به منظور برنامه ریزی بهتر آموزشی، در داده های واقعی مورد استفاده و ارزیابی قرار گرفته است. اطلاعات گردآوری شده شامل ۶۵ یادگیرنده در بازه سنی ۱۹ تا ۳۴ سال با میانگین سنی ۲۴ سال می باشد. نتایج حاصل از پرسشنامه فلدر - سولومون برای هر یادگیرنده به صورت بردارهایی دودویی با ۴۴ عنصر است. به منظور حاصل شدن مشخصه های مورد نظر در این تحقیق به پردازش این بردارها پرداخته شده است. در نهایت نتایج حاصل نشان از کارایی مدل پیشنهادی داشت چون یادگیرندگان با استفاده از مدل پیشنهادی به موفقیت تحصیلی بالاتری دست یافتند. همچنین دوره ی برگزار شده جذابیت بالاتری برای آنها داشت. از طرفی آنها تمایل بیشتری برای حضور در سامانه یادگیری از خود نشان دادند. (دی پیر و رابو، ۱۳۹۷)

رستمی، آیت، صاغری و یعقوبی در پژوهش پیش بینی میزان پیشرفت تحصیلی دانشجویان با روش خوشه بندی فازی در محیط های آموزشی به دنبال ارائه الگویی جهت پیش بینی عملکرد و افزایش کارایی و موفقیت یادگیری دانشجویان در یک محیط آموزشی با استفاده از داده کاوی بودند. در این پژوهش با تکیه بر روش های کتابخانه ای و پرسشنامه ای و مشاوره با افراد خبره تعدادی از ویژگی های تاثیرگذار در یادگیری دانشجویان شناسایی شد و با استفاده از روش انتخاب ویژگی، موثرترین آنها انتخاب شدند و برای روشن تر شدن روابط بین ویژگی های انتخاب شده، خوشه بندی فازی بر روی آنها انجام گرفت. بر مبنای الگوهای به دست آمده می توان هر دانشجو را در راستای ویژگی های تاثیرگذار بر روی آنها (دانشجویان) از ابتدای ترم راهنمایی و با توجه به نمراتی که در طول ترم کسب می کند، او را از محدوده نمره نهایی خود آگاه کرد و بر طبق توانایی هایش برنامه ریزی مناسب تحصیلی نمود. این الگوها می توانند برای کارآمد تر ساختن فرآیند یادگیری در سیستم موثر باشند. نتایج آزمایش ها حاکی از دقت مطلوب روش پیشنهادی به میزان ۰,۹۳۹، نسبت به روش های قبلی (کشف قوانین همبستگی، کلاس بندی و تشخیص ناهمگونی) داشت. (رستمی، آیت، صاغری و یعقوبی ۱۳۹۴)

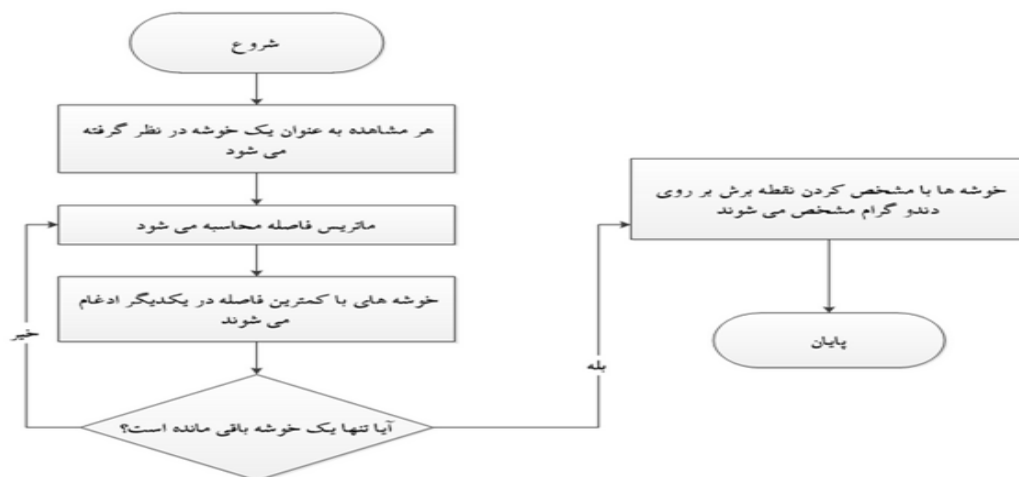
### شرح روش تحقیق

این پژوهش از نوع پژوهش کاربردی است. تحقیقات کاربردی به تحقیقاتی اطلاق می‌شود که به طور مستقیم بر روی یک موضوع کاربردی و عملی تمرکز می‌کنند و به دنبال یافتن راه‌حلی برای مشکلات حاضر هستند. هدف این مقاله نیز کاربرد الگوریتم‌های خوشه‌بندی سلسله‌مراتبی تجمعی در زمینه حل مسائل دنیای واقعی است.

جامعه آماری پژوهش عبارت است از ۳۰ نفر از دانش‌آموزان پایه سوم مقطع متوسطه اول دبیرستان جانبازان شهرستان سیروان ایلام که در آستانه فرآیند هدایت تحصیلی هستند. دبیرستان مذکور بخشی از یک مجتمع آموزشی تربیتی واقع شده در مسیر ایلام به لومار است که شامل سه مقطع متوسطه اول (راهنمایی)، متوسطه دوم (دبیرستان) و هنرستان (کاردانش) می‌باشد. این مجتمع در بخش کارزان شهرستان سیروان، در یک بافت روستایی و با تعداد حدوداً ۱۵۰ نفر دانش‌آموز قرار دارد.

ابزار جمع‌آوری اطلاعات عبارت است از نمرات دانش‌آموزان مقطع متوسطه اول مندرج در کارنامه نهایی خرداد ماه و نمرات طبقه‌بندی شده مشاوران مدرسه در زمینه مهارت‌های پایه و اساسی برای موفقیت در رشته‌های تحصیلی مختلف. برای خوشه‌بندی نمرات از الگوریتم‌های خوشه‌بندی سلسله‌مراتبی تجمعی با هدف یافتن نزدیک‌ترین نمرات و مهارت‌ها و گروه‌بندی و خوشه‌بندی دانش‌آموزان انجام شده است. برای محاسبه فاصله بین نمرات نیز از روش متهتن یا بلوک شهری به منظور منفی نبودن نمرات حاصل استفاده شده است. با تجزیه و تحلیل خوشه‌های حاصل شده، دانش‌آموزان در گروه‌های مختلف دسته‌بندی و الگوی پیشنهادی ارائه خواهد شد.

طبقه‌بندی داده‌ها و تعریف متغیرها با استفاده از الگوریتم اصلی این مقاله یعنی الگوریتم درخند حذخ‌رخز یا الگوریتم تک‌پیوندی خواهد بود. پارامترهای اصلی پژوهش عبارتند از: نمرات دروس پیش‌نیاز هر رشته تحصیلی، نمرات دبیران و مشاوران مدرسه در زمینه مهارت‌های اصلی مورد نیاز هر رشته. این الگوریتم ابتدا تعداد پ خوشه را به تعداد نمونه‌های پژوهش در نظر می‌گیرد و برای همه این پ خوشه ماتریس فاصله را محاسبه می‌کند. روش محاسبه فاصله هم برای منفی نبودن مقادیر روش متهتن یا بلوک شهری خواهد بود. در گام بعدی الگوریتم خوشه‌های با کمترین فاصله را ترکیب کرده تا در نهایت به تعداد معینی از خوشه‌ها برسیم که در واقع همان الگو و چارچوب پیشنهادی ما در این مقاله است. فلوجارت مراحل اجرایی این الگوریتم در تصویر زیر نشان داده شده است.



شکل ۱- فلوجارت مراحل اجرایی الگوریتم درخند حذخ‌رخز

### طبقه‌بندی داده‌ها

بعد از جمع‌آوری داده‌های مورد استفاده در پژوهش برای دسته‌بندی داده‌ها و انجام مراحل خوشه‌بندی توسط الگوریتم و همچنین بدست آوردن خوشه‌های نهایی و نمودارهای مربوط به آن از نرم‌افزار تجزیه و تحلیل آماری SPSS STATISTICS استفاده شده است. ابتدا متغیرهای اصلی پژوهش یعنی دروس و مهارت‌های اصلی مربوط به گروه‌های تحصیلی انتخاب شدند. این متغیرها همگی به صورت کمی یعنی عددی تعریف و اندازه‌گیری شدند.

برای دروس مورد نظر به صورت یک نمره شامل عدد صحیح و اعشار از بیست نمره و برای مهارت ها هم یک عدد بین مقادیر ۱ و ۵ که هر کدام نشان دهنده یک عملکرد خاص در زمینه داشتن، ضعف و یا فقدان در آن مهارت است. متغیرهای مقاله عبارتند از:

جدول ۱ - متغیرهای اصلی پژوهش

مهارت ها	دروس
تفکر	ریاضی
حل مسئله	علوم تجربی
کار گروهی	زبان انگلیسی
فنی و مهندسی	زبان عربی
کارآزمایشگاهی	ادبیات فارسی
تجزیه و تحلیل	مطالعات اجتماعی
	کار و فناوری

در مورد نمرات، نمره سالانه خرداد ماه (نهایی) هر درس انتخاب شد، دلیل این کار استاندارد و کشوری بودن آزمون های خرداد ماه سال سوم مقطع متوسطه اول بود، و در مورد مهارت ها نیز بر اساس نمرات مشاور مدرسه در مورد هر مهارت و هر فرد که یک نمره بین اعداد ۱ تا ۵ بود انتخاب شد. نمره ۱ نشان دهنده عملکرد بسیار ضعیف در هر مهارت، نمره ۲ نشان دهنده عملکرد ضعیف، نمره ۳ نشان دهنده عملکرد متوسط، نمره ۴ نشان دهنده عملکرد خوب و نمره ۵ نشان دهنده عملکرد بسیار خوب یا عالی هر دانش آموز در مورد آن مهارت است.

### تجزیه و تحلیل داده ها

به منظور این کار از آزمون های آماری موجود در برنامه SPSS استفاده کردیم. در میان آزمون های این برنامه از دسته آزمون classify به منظور خوشه بندی داده ها استفاده شد، سپس تنظیمات مربوط به مراحل اجرای الگوریتم، تعداد خوشه های مورد انتظار در خروجی، تعیین روش محاسبه فاصله بین نمونه ها و رسم نمودار دندوگرام و ماتریس های مربوط به مراحل اجرا و محاسبه فاصله هم انجام شد تا داده ها آماده خوشه بندی شوند.

### خوشه بندی داده

در مرحله اول به منظور قرار گرفتن نمونه های مشابه در نمرات و مهارت ها و جلوگیری از شکل گیری یک خوشه خیلی بزرگ، ابتدا تعداد ۱۰ خوشه به عنوان خروجی اولیه پژوهش در نظر گرفته شد. برای محاسبه فاصله بین نمونه ها روش منهتن یا بلوک شهری استفاده شد، دلیل این امر در درجه اول منفی نبودن فاصله بین نمونه ها و در درجه بعدی دقیق تر بودن محاسبه فاصله به این روش نسبت به روش های دیگر محاسبه فاصله بود. این ۱۰ خوشه به صورت ۳ خوشه شامل مجموعه ای از نمونه های شبیه به هم و ۷ خوشه مجزا بودند. با مشاهده ۷ خوشه مشخص شد که همه آنها از نظر نمرات و مهارت ها شباهت بالایی به قرار گرفتن در یکی از ۳ خوشه قبلی داشتند. بنابراین در مرحله بعدی در همان خوشه ادغام شدند. به این ترتیب خروجی کار الگوریتم ۳ خوشه بود.

### مراحل اجرای الگوریتم

اگر در همان ابتدا خروجی برنامه را روی ۳ خوشه تنظیم کنیم با مشکلاتی روبرو خواهیم شد. از جمله اینکه در یکی از این ۳ خوشه ۲۸ نمونه و در دو خوشه دیگر هر کدام تنها یک نمونه قرار خواهد گرفت. علاوه بر این امکان دارد کسانی در یک خوشه باشند که هیچ شباهت خاصی در نمرات و مهارت ها نداشته باشند. به عنوان مثال کسانی که نمرات دروس مختلف آنها روی مقادیر ۱۲ و ۱۳ باشد با کسانی در یک خوشه باشند که در همان دروس نمرات ۱۷ و ۱۸ گرفته باشند تنها به خاطر اینکه الگوریتم ممکن است در تشخیص موارد شباهت دچار این خطا شود.

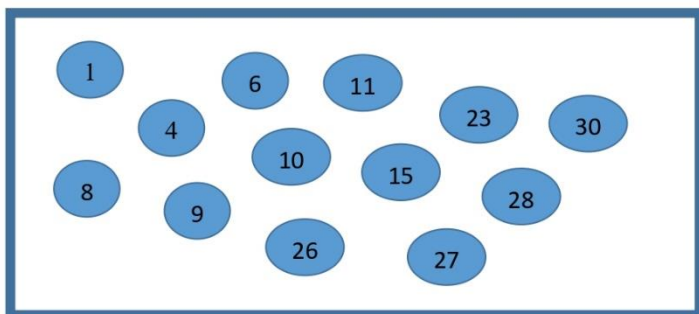
در نتیجه خروجی اولیه الگوریتم را روی ۱۰ خوشه قرار داده ایم. با مشاهده این ۱۰ خوشه خروجی مشخص شد که ۳ خوشه بزرگتر و ۷ خوشه مجزا و تک نمونه ای ایجاد می شود. با بررسی ۳ خوشه بزرگتر از نظر نمرات و عملکرد دانش آموزان مشخص شد که هر کدام نمایانگر یکی از ۳ خوشه مد نظر ما

بودند، پس این ۳ خوشه به عنوان همان ۳ رشته تحصیلی مبنا در نظر گرفته شدند. بعد از بررسی ۷ خوشه مجزا هم معلوم گردید که همه ۷ خوشه شباهت فراوانی به قرارگیری در یکی از خوشه ها یعنی علوم انسانی داشتند، پس همه ۷ خوشه در خوشه سوم یعنی علوم انسانی ادغام شدند. در اینجا منظور از مشابهت، مشابهت در نمرات دروس کلیدی رشته علوم انسانی یعنی دروس ادبیات فارسی، مطالعات اجتماعی و زبان عربی و مشابهت در عملکرد دانش آموز در مهارت های اصلی مورد نیاز این رشته مانند توانایی تجزیه و تحلیل متون مختلف بوده است. این شباهت ها معیار قرارگیری همه ۷ خوشه مجزا در خوشه مربوط به رشته علوم انسانی بود. در جدول بعد مراحل اجرای الگوریتم و ادغام نمونه ها با هم، سپس ادغام خوشه ها با همدیگر نشان داده شده است. در اولین مرحله نمونه های ۳ و ۷ که کمترین فاصله را در بین نمونه ها داشتند با هم ترکیب شده و یک خوشه را تشکیل دادند. در مرحله بعدی نمونه های ۱۸ و ۲۱ ترکیب و یک خوشه را ایجاد کردند.

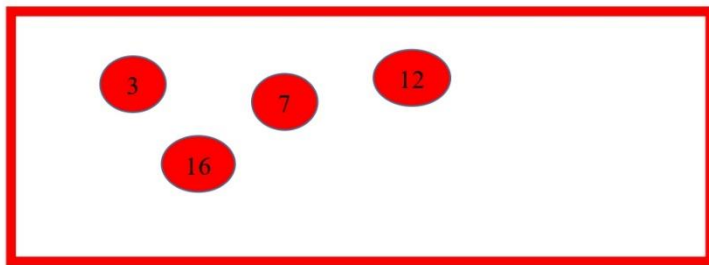
جدول ۲- مراحل اجرای الگوریتم

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster ۱	Cluster ۲		Cluster ۱	Cluster ۲	
۱	۳	۷	۵,۵۰۰	۰	۰	۴
۲	۱۸	۲۱	۸,۲۵۰	۰	۰	۹
۳	۹	۱۵	۹,۰۰۰	۰	۰	۵
۴	۳	۱۲	۹,۰۰۰	۱	۰	۱۴
۵	۹	۲۷	۹,۵۰۰	۳	۰	۸
۶	۶	۸	۹,۵۰۰	۰	۰	۷
۷	۶	۲۳	۹,۷۵۰	۶	۰	۱۰
۸	۹	۲۸	۱۱,۰۰۰	۵	۰	۱۲
۹	۱۸	۲۲	۱۱,۷۵۰	۲	۰	۱۳
۱۰	۶	۱۰	۱۱,۷۵۰	۷	۰	۱۱
۱۱	۴	۶	۱۱,۷۵۰	۰	۱۰	۱۵
۱۲	۹	۲۶	۱۲,۰۰۰	۸	۰	۱۵
۱۳	۵	۱۸	۱۲,۰۰۰	۰	۹	۱۷
۱۴	۳	۱۶	۱۲,۰۰۰	۴	۰	۲۲
۱۵	۴	۹	۱۲,۰۰۰	۱۱	۱۲	۱۶
۱۶	۱	۴	۱۲,۵۰۰	۰	۱۵	۱۹
۱۷	۵	۲۴	۱۲,۷۵۰	۱۳	۰	۱۸
۱۸	۵	۱۷	۱۲,۷۵۰	۱۷	۰	۲۱
۱۹	۱	۱۱	۱۲,۷۵۰	۱۶	۰	۲۰
۲۰	۱	۳۰	۱۳,۵۰۰	۱۹	۰	۲۱
۲۱	۱	۵	۱۳,۵۰۰	۲۰	۱۸	۲۳
۲۲	۳	۱۹	۱۴,۰۰۰	۱۴	۰	۲۵
۲۳	۱	۲	۱۴,۰۰۰	۲۱	۰	۲۴
۲۴	۱	۲۹	۱۴,۵۰۰	۲۳	۰	۲۵
۲۵	۱	۳	۱۴,۵۰۰	۲۴	۲۲	۲۶
۲۶	۱	۲۵	۱۵,۷۵۰	۲۵	۰	۲۷
۲۷	۱	۱۳	۱۶,۰۰۰	۲۶	۰	۲۸
۲۸	۱	۲۰	۱۶,۷۵۰	۲۷	۰	۲۹
۲۹	۱	۱۴	۱۶,۷۵۰	۲۸	۰	۰

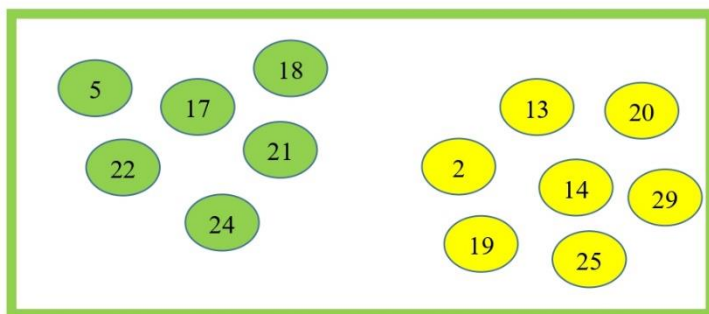
جدول ۳ - تعداد خوشه ها و اعضای هر کدام بر انجام فرآیند خوشه بندی (خوشه های اولیه) - بصری سازی خوشه های نهایی



خوشه پیشنهادی علوم تجربی (خوشه شماره 1)



خوشه پیشنهادی ریاضی فیزیک (خوشه شماره 3)



خوشه پیشنهادی ادبیات و علوم انسانی (خوشه شماره 4+خوشه های مجزا)

\*\* نمونه های زرد رنگ در ابتدا خوشه های مجزایی بودند اما مشابهت فراوانی به قرارگیری در رشته علوم انسانی داشتند.

Cluster Membership	
Case	۱۰ □□□□□□□□
۱	۱
۲	۲
۳	۳
۴	۱
۵	۴
۶	۱
۷	۳
۸	۱
۹	۱
۱۰	۱
۱۱	۱
۱۲	۳
۱۳	۵
۱۴	۶
۱۵	۱
۱۶	۳
۱۷	۴
۱۸	۴
۱۹	۷
۲۰	۸
۲۱	۴
۲۲	۴
۲۳	۱
۲۴	۴
۲۵	۹
۲۶	۱
۲۷	۱
۲۸	۱
۲۹	۱۰
۳۰	۱

### ارزیابی خوشه بندی

بعد از انجام فرآیند خوشه بندی و مشخص شدن خوشه های نهایی برای محاسبه و بررسی صحت و دقت خوشه های ایجاد شده باید به ارزیابی نتایج خوشه بندی پرداخت. روش های متعددی برای ارزیابی خوشه بندی وجود دارد. در این مقاله از روش ارزیابی محاسبه ضریب نیمرخ استفاده کرده ایم. در این روش برای محاسبه ضریب نیمرخ برای شی  $i$  که  $S(i)$  نشان داده می شود، متوسط فاصله  $i$  تا تمام اعضای خوشه خودش را با  $a(i)$  نشان می دهیم. و همینطور میانگین فاصله شی  $i$  با تمام خوشه های دیگر به جز خوشه خودش را محاسبه نموده و کمترین مقدار را با  $b(i)$  نشان می دهیم. خوشه ای که کمترین فاصله را داشت خوشه همسایه نامیده می شود. فاصله بین اعضا هم به روش فاصله اقلیدوسی محاسبه می شود. اگر خوشه دارای یک عضو باشد آنگاه ضریب نیمرخ برابر است با :

$$S(i) = [b(i) - a(i)] / \max \{a(i), b(i)\} \quad -1 \leq s(i) \leq 1$$

هرچه  $S(i)$  به یک نزدیک تر باشد یعنی  $i$  در خوشه درست قرار گرفته است. متوسط پهنای نیمرخ  $S$  عبارت است از متوسط  $S(i)$  تمام اشیاء که این شاخص می تواند برای تعیین تعداد خوشه ها استفاده شود. بیشترین متوسط پهنای نیمرخ را ضریب نیمرخ می نامند. مقادیر بالای ۰,۵۰ و مقادیر بالای ۰,۷۰ ساختار عالی خوشه بندی را نشان می دهند. جدول زیر نحوه تفسیر نتایج حاصل از محاسبه فاصله ها را نشان می دهد.

جدول ۴- نحوه تفسیر مقادیر ضریب نیمرخ

تفسیر مقادیر ضریب نیمرخ	
تفسیر	Sc
ساختار عالی پیدا شده است.	۱,۰۰-۰,۷۱
ساختار قابل قبولی پیدا شده است.	۰,۷۱-۰,۵۱
ساختار ضعیف و شاید مصنوعی پیدا شده است، توصیه می شود سایر روش های ارزیابی انجام شود.	۰,۵۰-۰,۲۶
ساختار قابل توجهی به دست نیامده است.	< ۰,۲۵

### نتیجه ارزیابی خوشه ها

برای ارزیابی دقت و صحت نتایج حاصل از اجرای الگوریتم و خوشه های بدست آمده به صورت تصادفی و با استفاده از فرمول ضریب نیمرخ، ارزیابی فاصله بین خوشه های ریاضی فیزیک و علوم تجربی انجام گرفت و نتایج به شرح زیر بدست آمد:

فاصله ۳ تا اعضای خوشه خودش و خوشه دیگر	فاصله ۴ تا اعضای خوشه خودش و خوشه دیگر	فاصله ۱۲ تا اعضای خوشه خودش و خوشه دیگر	فاصله ۱۶ تا اعضای خوشه خودش و خوشه دیگر
۱(۳)= ۴/۲۳	۱(۷)= ۴/۰۹	۱(۱۲)= ۴/۷۷	۱(۱۶)= ۵/۷۶
۱(۳)= ۱۴/۲۶	۱(۷)= ۱۳/۱۲۴	۱(۱۲)= ۱۲/۵۷	۱(۱۶)= ۱۰/۵۳
ت(۳)= ۰/۷۰	ت(۷)= ۰/۶۸	ت(۱۲)= ۰/۶۲	ت(۱۶)= ۰/۴۵
قابل قبول	قابل قبول	قابل قبول	نیاز به ارزیابی مجدد

### مروری بر مساله پژوهش

هدف ما در این مقاله ارائه یک چارچوب پیشنهادی برای هدایت تحصیلی دانش آموزان بر اساس الگوریتم خوشه بندی سلسله مراتبی تجمعی بود. با توجه به اهمیت بالای فرآیند هدایت تحصیلی و نقش و تاثیر آن بر آینده شغلی و زندگی دانش آموزان و ایرادات فعلی روش فعلی انتخاب رشته تحصیلی در مدارس از جمله در نظر نگرفتن مهارت های لازم برای هر رشته در کنار نمرات دروس مختلف و دخالت عامل انسانی بر نتایج، تصمیم گرفتیم از الگوریتم کمک بگیریم. علاوه بر آن به دنبال کاربرد الگوریتم در حل یکی از مسائل دنیای واقعی و به عبارت دیگر کاربردی کردن پژوهش بودیم. برای این کار نمرات دروس پایه و کلیدی رشته های تحصیلی و نمرات عملکردی دانش آموزان در زمینه برخی از مهارت های لازم را جمع آوری کرده و بعد از تعریف متغیرهای پژوهش، به کمک برنامه ت-ت-پت به خوشه بندی داده ها پرداختیم.

### نوآوری پژوهش

مهارت آموزی به دانش آموزان در نظام های آموزشی نوین بیش از نمره و معدل اهمیت دارد. دانش آموزان با مهارت یابی می توانند به درک عمیق تری از آموخته های خود دست یابند و برای ساختن آینده ای بهتر آماده شوند. مهارت آموزی مقوله ای است که همواره مورد توجه نظام های آموزشی پیشرفته قرار دارد؛ چرا که تنها آموزش سطحی منابع درسی به دانش آموزان نمی تواند آنان را برای زندگی آینده مهیا سازد. از این رو، برای پرورش نسلی کارآمد که بتواند آینده روشنی را برای کشور رقم زند، باید به فکر راه های افزایش مهارت افراد در نظام آموزشی بود. هدف ما در این مقاله در نظر گرفتن همزمان دروس پایه هر رشته در کنار مهارت های مورد نیاز دانش آموزان برای موفقیت در تحصیل در رشته های مختلف بود. حتی به جای اینکه در فرآیند خوشه بندی رشته ها را تفکیک کرده و برای هر کدام فقط دروس و مهارت های خاص همان رشته را در نظر بگیریم، خوشه بندی را بر روی همه متغیرها انجام دادیم. دلیل این کار در کنار هم ارزیابی کردن همه نمرات دروس و مهارت های دانش آموزان به صورت یک مجموعه موثر بر هم بود تا در واقع نقاط ضعف و قوت دانش آموز با هم و در کنار هم دیده شود و مورد بررسی قرار گیرد.

### نتایج خوشه بندی

بعد از اتمام فرآیند خوشه بندی به منظور ارزیابی کار به مقایسه نتایج حاصل از الگوریتم با نتایج واقعی مربوط به فرآیند هدایت تحصیلی دانش آموزان پرداختیم. مقایسه این نتایج رابطه معناداری را میان دو گونه نتایج نشان داد. جدول زیر نتایج خوشه بندی توسط الگوریتم و همچنین نتایج واقعی مربوط به انتخاب رشته دانش آموزان دبیرستان را نشان می دهد.

علوم انسانی	علوم تجربی	ریاضی فیزیک	خوشه های پژوهش
۲-۵-۱۳-۱۴-۱۷-۱۸-۱۹- ۲۰-۲۱-۲۲-۲۴-۲۵-۲۹	۱-۴-۶-۸-۹-۱۰-۱۱-۱۵- ۲۷-۲۸-۳۰-۲۳-۲۶	۳-۷-۱۲-۱۶	نمونه های بدست آمده از الگوریتم
۲-۹-۱۳-۱۴-۱۷-۱۸-۱۹- -۲۵-۲۶-۲۱-۲۲-۲۳-۲۴ ۲۹	۶-۸-۱۰-۱۱-۲۳-۱-۴ ۲۶-۲۷-۲۸	۱۲-۱۵-۱۶-۳-۷	نمونه های واقعی بر اساس انتخاب رشته دانش آموزان

با مقایسه جدول فوق نتایج بدست آمده از پژوهش به صورت زیر عنوان می گردد:

**خوشه ریاضی فیزیک:** در مورد این خوشه هر ۴ نمونه ای که توسط الگوریتم در آن قرار گرفتند در رشته ریاضی فیزیک مشغول به تحصیل هستند. در مورد این خوشه ۱۰۰ درصد قرار گرفتن نمونه ها در خوشه صحیح بود.

**خوشه علوم تجربی:** در این خوشه ۱۰ مورد از ۱۳ نمونه ای که توسط الگوریتم در این خوشه قرار گرفتند در حال تحصیل در رشته علوم تجربی هستند. نمونه شماره ۱۵ در حال تحصیل در رشته ریاضی فیزیک و نمونه های شماره ۹ و ۳۰ در حال تحصیل در رشته علوم انسانی هستند. در مورد این خوشه ۷۷ درصد قرار گرفتن نمونه ها در خوشه صحیح بود.

**خوشه علوم انسانی:** در این خوشه ۱۱ مورد از ۱۳ نمونه ای که الگوریتم در این خوشه قرار داده، در حال تحصیل در همین رشته در دبیرستان هستند. نمونه های شماره ۵ و ۲۰ در حال تحصیل در مقطع هنرستان می باشند. در مورد این خوشه ۸۵ درصد قرار گرفتن نمونه ها در خوشه صحیح بود.

### میزان تحقق اهداف پژوهش

اهداف کلی این مقاله عبارت بودند از:

- ۱- ارائه یک الگو و چارچوب پیشنهادی و نوین در زمینه هدایت تحصیلی
  - ۲- ضرورت اصلاح و بازنگری در شیوه فعلی هدایت تحصیلی
  - ۳- کاربست تکنیک ها و الگوریتم های داده کاوی در حل مسائل دنیای واقعی
  - ۴- استفاده از تکنیک های علمی و دقیق در محاسبه خوشه ها و گروه ها و دقیق تر و علمی تر شده فرآیند هدایت تحصیلی
- از بررسی نتایج بدست آمده از خوشه بندی داده ها و تجزیه و تحلیل نتایج بدست آمده از خوشه بندی، می توان نتیجه گرفت که همه اهداف اصلی این پژوهش محقق شدند. در واقع ما با در کنار هم ارزیابی کردن نمرات و عملکرد دانش آموزان در مهارت ها هم به اصلاح و بازنگری در روند فعلی انتخاب رشته و هدایت تحصیلی دانش آموزان کمک کرده و هم یک الگو و چارچوب جدید را در این زمینه پیشنهاد دادیم. یعنی تحقق اهداف اول و دوم پژوهش علاوه بر این، با علمی و کاربردی کردن فرآیند هدایت تحصیلی، به حداقل رساندن نقش و دخالت عامل انسانی و تاثیرگذاری آن بر این فرآیند و همچنین کاربست عملی یکی از تکنیک های علم داده کاوی در زمینه خوشه بندی و تجزیه و تحلیل آن هم به علمی شدن این فرآیند کمک کرده و هم از یکی از الگوریتم های علم داده کاوی در حل یکی از مسائل بسیار مهم و تعیین کننده در آینده علمی، اقتصادی و اجتماعی دانش آموزان به مثابه آینده سازان کشور کمک کردیم. یعنی تحقق اهداف سوم و چهارم پژوهش

### پیشنهادات پژوهش

- ۱- جامعه آماری این پژوهش عبارت بود از تعداد ۳۰ نفر دانش آموزان پایه سوم مقطع متوسطه اول که در آستانه فرآیند هدایت تحصیلی بودند، در واقع حجم نمونه آماری کوچک و محدود بود. به منظور بررسی بهتر و دقیق تر نتایج این مقاله، پیشنهاد می شود که در پژوهش های آینده جامعه آماری گسترده تری و شامل دانش آموزان دختر و پسر در نظر گرفته شود.
- ۲- پیشنهاد می شود که سازمان آموزش و پرورش شهرستان به صورت آزمایشی (پایلوت) به اجرای این پژوهش در یکی از مدارس تحت حوزه خود پرداخته تا بتوان با مشاهده روند اجرای فرآیند هدایت تحصیلی بر اساس این الگوریتم دقیق تر نسبت به استخراج نقاط ضعف و قوت و تکمیل طرح پرداخت.
- ۳- پیشنهاد می شود در پژوهش های آینده علاوه بر متغیرهای این پژوهش، متغیرهایی شامل: علایق خود دانش آموزان، شرایط فرهنگی و اجتماعی هر منطقه و زیست بوم، نیازسنجی رشته های تحصیلی مختلف به منظور اشباع نشدن برخی از رشته ها و خالی از ظرفیت شدن دیگر رشته ها، همچنین در نظر گرفتن متغیر تاثیر خانواده دانش آموز بر فرآیند هدایت تحصیلی پرداخت. چون تمامی این متغیرها مربوط به حوزه علوم تربیتی و مدیریت آموزشی می باشند، پیشنهاد می شود در پژوهش های مرتبط آینده که بین رشته ای هستند مورد نظر قرار گیرند.

## References

Abbasi, L.; Bayat, M. (۲۰۱۵), Examining the issue of academic guidance in Iran's education with an emphasis on evaluating the performance of the new academic guidance program, Social Studies Office, Islamic Council Research Center, March ۲۰۱۵ [in Persian]

De Pierre, M.; Rabo, A.; (۲۰۱۷), using educational data mining to group learners in the e-learning environment in order to personalize the educational program, Journal of Management and Planning in Educational Systems, Volume ۱۱, Number ۱

Hedayati, B.; Javanmard, M. (۲۰۱۳), Classification of first-year high school students based on annual grades using educational data mining, National Conference on Computers and Sustainable Development, ۸th period, November ۲۰۱۳. [in Persian]

Kamrei, M.; Mantezar, G.; (۲۰۱۸), similar grouping of learners in the e-learning environment using the council clustering method, Iran Information and Communication Technology Scientific Research Quarterly, year ۱۱, numbers ۳۹ and ۴۰

Najafi, M.; Afzali, M.; Moradi, M.; (۲۰۲۲), the use of educational data mining to identify factors affecting students' academic failure, the Quarterly Journal of Intelligent Multimedia Processing and Communication Systems, second year, first issue. [in Persian]

Regulations on academic guidance of students, approved on ۲۰۱۴/۲۴/۱۲, Supreme Council of Education of the Islamic Republic of Iran

Rezaei, M.; Mantezar, G.; (۲۰۱۴), grouping of learners in adaptive learning systems using fuzzy clustering method, Iranian Electrical and Computer Engineering Journal, year ۱۳, number ۲. [in Persian]

Rostami, M.; Ayat, S.; Sagheri, F.; Yaqoubi, F.; (۲۰۱۴), predicting the level of academic progress of students in educational environments, Shahid Rajaee University of Tehran's educational technology scientific research journal, volume ۹, number ۴. [in Persian]

The document on the fundamental transformation of the country's education system, approved on ۲۰۱۳/۰۶/۱۰, by the Supreme Council of the Cultural Revolution of the Islamic Republic of Iran Xu, c.; zheng, l.; effectiveness analysis of the application of clustering in student grouping (2013), international conference on education technology and information system