مطالعه تطبیقی الگوریتم‌های کلستربندی در زمینه‌ی دیتاکاوی تحصیلی

پوهنیار احمدضیا بهرامی[۱]، پوهنوال امیرکرور شهیدزی[۲]

[۱،۲]دیپارتمنت کمپیوترساینس، پوهنځی کمپیوترساینس، پوهنتون کابل، کابل، افغانستان
ایمیل: bahrami.qarabagh@gmail.com

چکیده

دیتاکاوی تحصیلی دیتاهای خام سیستم‌های تحصیلی را به اطلاعات مفیدی تبدیل می‌کند که به‌طور باالقوه می‌تواند تأثیر بیشتری بر عمل‌کرد تحقیقاتی و تحصیلی داشته باشد. بنابراین، سؤال این است که چگونه یک پوهنتون می‌تواند از قدرت این دیتاهای تحصیلی برای استفاده در استراتیژیک خود استفاده کند؟ برای یافتن پاسخ، این مقاله بر مطالعه‌ی تطبیقی الگوریتم‌های کلستربندی به‌عنوان کاربرد در دیتاکاوی تحصیلی تمرکز می‌کند. روش‌های مبتنی بر پارتیشن (K-Means)، مبتنی بر چگالی (DBSCAN) و سلسله‌مراتبی (BIRCH) را مقایسه کرده است تا تعیین کند که کدام تکنیک برای انجام تحلیل‌های کلستربندی در محیط تحصیلی مناسب‌تر است. مطالعات نشان می‌دهد که الگوریتم K-Means مبتنی برطبقه‌بندی بهتر از الگوریتم BIRCH سلسله‌مراتبی و الگوریتم DBSCAN مبتنی بر چگالی عمل می‌کند.

اصطلاحات کلیدی: دیتاکاوی تحصیلی؛ کلستربندی؛ الگوریتم؛ تکنیک؛ دیتاکاوی

Comparative Study of Clustering Algorithms in the Context of Education Data Mining

Jr. Teaching Asstt. Ahmad Zia Bahrami[1] and Associate Prof. Amir Kror Shahidzay[2]

[1,2]Department of Computer Science, Faculty of Computer Science, Kabul University, Kabul, Afghanistan
Email: bahrami.qarabagh@gmail.com

Abstract

Educational data mining turns raw data in education systems into useful information that could potentially have a greater impact on research and academic performance. So, the question is how a university can harness the power of this academic data for its strategic use. To find the answer, this paper focuses on a comparative study of clustering algorithms as an application in academic data mining. Compares partition-based (K-Means), density-based (DBSCAN), and hierarchical (BIRCH) methods to determine which technique is best suited for performing clustering analysis in an educational setting. Studies show that the classification-based K-Means algorithm performs better than the hierarchical BIRCH algorithm and the density-based DBSCAN algorithm.

Keywords: EDM; Clustering; Algorithm; Technique; Data Mining

مقدمه

در دهه‌ی گذشته، در نتیجه‌ی افزایش مؤسسات تحصیلات عالی، تعداد محصلان لیسانس و ماستری افـزایش یافته است. بسیاری از پوهنتون‌ها در روش تدریس یا برگـزاری امتحانات تغییراتی ایجاد کرده اند؛ اما هم‌چنان با مشکلات بیکاری و ترک تحصیل مواجه هستند. درک عوامل عمل‌کرد پایین یا افزایش نرخ ترک تحصیل، کار دشواری است که شامل عمل‌کرد گذشته و حال محصلان و سوابق رشته‌ی آن‌ها می‌شود. دیتاکاوی تحصیلیEDM یکـی از آن ابزارهای قـدرت‌مند برای تجزیه و تحلیـل و پیش‌بینی رفتار و عمل‌کرد آینده‌ی محصلان است (۱)

اگرچه مؤسسات تحصیلات عالی/ پوهنتون‌ها دیتابیس‌های معلوماتی تحصیلی خوبی دارند؛ اما در هیچ تصمیم‌گیری برای افزایش عمل‌کرد محصلان یا کاهش نرخ ترک تحصیل یا بیکاری از آن‌ها استفاده نمی‌شود. هدف EDM کشف روش‌هایی برای پیشرفت تحصیل و کارآیی محصلان با کاهش نرخ ترک تحصیل آن‌هاست. EDM از کلستربندی و طبقه‌بندی پیش‌پردازش شده برای استخراج دیتا استفاده می‌کند. کلستربندی روشی بدون نظارت برای بررسی معلومات است. این شامل گروپ‌بندی دیتاهای مشابه برای تشکیل یک کلستر یا یک گروپ است. بهترین الگوریتم‌های کلستربندی آن‌هایی در نظر گرفته می‌شوند که اقلام دیتا در همان کلستر بیشترین شباهت را داشته باشند و اقلام دیتا متعلق به کلسترهای مختلف حداقل یا اصلاً دارای شباهت نباشند (۲).

طبقه‌بندی یک روی‌کرد تحت نظارت یادگیری اسـت که مبتنی بر اسـتفاده از نتایج شـناخته شـده و پیش‌بینـــی نتایج آینده بر اساس محدودیت‌های داده شده است. EDM با استفاده از الگوریتم‌های طبقه‌بندی، زمینه های مختلف پیش‌بینی برای اثربخشـــی ســـازمانی را دارد. موفقیت الگوریتم های طبقه‌بندی به میزان متفاوت بودن دیتاهـای ارائه شده و بسیاری از عوامل دیگـر بستگی دارد.

پوهنتون‌هـــا باید بین خود جلسه بگـــذارند و دلایل رایج عمل‌کـــرد پایین محصلان و ترک تحصیل آن‌ها را شناسایی کنند تا بتوانند عمل‌کـرد و رفتار محصلان را پیش‌بینی کنند که در نهایت به محصلان و مؤسـسـات کمک خواهد کرد، تا بتوانند نقاط ضـعف خود را بشـناسـند و از قبل اقداماتی را انجام دهند. این کار به نفع همه شـرکت‌کنندگان مؤسسات تحصیلات عالی/ پوهنتون‌ها، مدیریت، استادان، محصلان و والدین مـــی شود. استادان می‌توانند در مورد مواد درسی و لکچرهای خود برنامه‌ریزی کنند و می‌توانند به چنین محصلان توصیه کنند که عمل‌کـــرد بهتری داشته باشند. محصلان می‌توانند روی مشـکلات خود کار کنند و برای بهبود آن از اسـتادان مشـوره بگیرند. والدین مـــی‌توانند از

عمل‌کـرد بهتر فرزندان خود در چنین راهنمایی‌هـا مطمئن باشند. مدیریت می‌تواند امکانات بهتری را هم برای استادان و هم برای محصلان فراهم کند.

پیشینه‌ی تحقیق

تجزیه و تحلیل کلستربندی یکی از تکنیک‌های مورد استفاده در دیتاکاوی است و شامل تقسیم مجموعه‌یی از اشیای دیتا به زیر مجموعه‌ها است. این زیرمجموعه‌ها یا کلسترهـا برای سازمان‌دهی اشیا به‌گونه‌یی استفاده می‌شوند که هر شیء درون یک کلستر مشابه یک‌دیگـر باشند، اما با اشیایی‌که متعلق به کلسترهای دیگـر هستند، متفاوت باشند (۳). تجزیه و تحلیل کلستربندی مـی‌تواند گـروپ‌های محصلان را با الگوهـای استفاده‌ی مشابه برای بهبود عمل‌کرد سایر انواع تحلیل در محیط‌هـای تحصیلی شناسایی کند (۴). بازخورد و تعامل گـروپ را با استفاده از تحلیل کلستربندی برای ارائه اطلاعات هدف‌مند بر اساس شباهت‌های استفاده‌ی استخـراج شده با کلستربندی و الگوکاوی متوالی به‌منظور بهبود سطح و کیفیت بازخورد بهبود می‌بخشد. حوزه‌ی تحقیقی دیتاکاوی تحصیلی (EDM) بر کاربرد تکنیک‌ها و روش‌های دیتاکاوی در محیط‌های تحصیلی متمـرکز است. EDM به توسعه، تحقیق و به‌کارگیری روش‌هـای یادگیری ماشین، دیتاکاوی و احصائیوی برای شناسایی الگوهایی در مجموعه‌های بزرگ دیتاهای تحصیلـی که در غیر این صورت تجزیه و تحلیل آن‌هـا غیرممکن است، می‌پردازد (۵).

این بخش، انواع مختلف کلستربندی را برجسته می‌کند و انتخاب الگوریتم‌های مورد استفاده برای نمایش هر یک از روش‌های شناسایی شده را توجیه می‌کند.

تکنیک کلستربندی

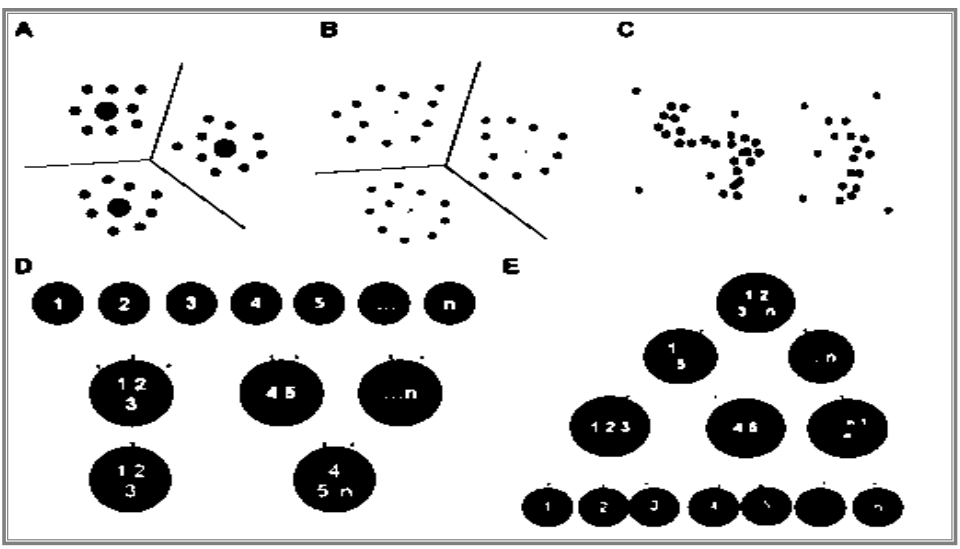
کلستربندی یکی از تکنیک‌های پرکاربرد در دیتاکاوی و یک تکنیک یادگیری بدون نظارت است. کلستربندی تقسیم دیتاها به گروپ‌های دارای اشیای مشابه است. هر گروپی که کلستر نامیده می‌شود، شامل اشیایی است که باهم متفاوت از اشیای گروپ‌های دیگر هستند. انواع مختلف الگوریتم‌های کلستربندی را می‌توان در ۳ گروپ، کلستربندی سلسله‌مراتبی، پارتیشن‌بندی و کلستربندی مبتنی بر تراکم تقسیم‌بندی نمود.

۱. **روش‌های پارتیشـــن‌بندی:** الگوریتم‌های پارتیشـــن‌بندی تلاش خواهند کرد n شـــیء را در k پارتیشن‌هـــای مشخص گروپ‌بندی کنند (k <= n) که در آن هر پارتیشن (زیر گـــروپ) یک کلستر را نشان مـــی‌دهد به‌طوری‌که اشیاء درون یک کلستر با یک‌دیگر "شبیه" و با اشیاء در کلستر

دیگر "متفاوت" هستند. هرکلستر را می‌توان با یک مرکز نمایش داد که می‌تواند اشیایی در مجموعه‌ی دیتا (k-Medoid) باشد که در شکل ۱ نشان داده شده است.

۲. **روش‌های سلسله‌مراتبی:** تکنیک کلستربندی سلسله‌مراتبی یکی از تکنیک‌های پرکاربرد در کلستربندی سلسه‌مراتبی فاصله‌ی جفت‌ها را از یک‌دیگر محاسبه و سپس کلسترهای مشابه را به روش پایین به بالا، بدون نیاز به تأمین تعداد کلسترها ادغام می‌کند. کلستربندی سلسله‌مراتبی یک درخت از کلسترها می‌سازد که به‌عنوان دندروگرام شناخته می‌شود. هرنود کلستر، شامل چند کلستر فرزند است.

۳. **کلستربندی مبتنی بر تراکم:** این روش از تراکم‌های محلی برای تعیین نحوه‌ی گروپ‌بندی اشیا برای ایجاد کلسترها استفاده می‌کند. این تراکم‌های محلی که به‌عنوان همسایگی‌ها نامیده می‌شوند، امکان شناسایی کلسترهایی از اشکال دل‌خواه را فراهم می‌کنند، همان‌طور که در شکل۱ نشان داده شده است، در مقایسه با شکل‌های بیضی و محدب روش‌های تقسیم‌بندی که صرفاً به فاصله بین اشیاء در داخل متکی هستند.



شکل۱: خانواده الگوریتم‌های کلستربندی (۱)

الگوریتم‌های کلستربندی

کلستربندی K-Means

در این الگوریتم دیتا را به K کلستر مجزا تقسیم کنیم. این الگوریتم به دو فاز مجزا تقسیم می‌شود. فاز اول برای هر کلستر یک نقطه را به عنوال نقطه‌ی ثقل کلستر یا نقطه‌ی مرکزی کلستر به دست می‌آوریم. در فاز بعدی به‌دست می‌آوریم که هر نقطه از مجموعه، به کدام مرکز کلستر نزدیک‌تر است و آن نقطه

را به کلستر مربوطه نسبت می‌دهیم، در حالت کلی برای به‌دست آوردن فاصله بین نقاط دیتا و مراکز کلسترها از فاصله‌ی اقلیدسی استفاده می‌شود. زمانی‌که تمام نقاط در کلسترها قرار گرفتند، مرحله‌ی اول به اتمام رسیده و کلستربندی اولیه انجام شده است. سپس دوباره برای کلسترها مراکز جدید به‌دست می‌آوریم و فاصله‌ی هر نقطه را نسبت به این نقاط مرکزی اندازه می‌گیریم تا کلسترها آپدیت شوند و این کار تا زمانی ادامه پیدا می‌کند که کلسترها همگرا شوند.

اما مشکل عمده‌یی این الگوریتم این است که با توجه به مقدار مراکز اولیه کلسترهای متفاوتی تولید می‌شود و در نتیجه کیفیت کلسترهای نهایی شدیداً به انتخاب مراکز اولیه کلسترها وابسته است. این الگوریتم از لحاظ محاسباتی گران و متناسب با تعداد نقاط، تعداد کلسترها و تعداد تکرارها نیاز به زمان دارد. مرکز برای الگوریتم K-Means به‌عنوان مقدار اوسط در فضای دیتا محاسبه شده از مقادیر ویژگی اشیای درون کلستر نشان داده می‌شود.

درحالی‌که الگوریتم‌های تقسیم‌بندی دیگر مانند Partitioning Around Medoids وجود دارد که از یک شـــی واقعی در فضای شـــی به‌عنوان مرکز استفاده می‌کند و Clustering Large Applications based upon Randomized Search (Ng & Han 2002), از ویژگی‌های فضایی اشیاء استفاده می‌کند که تجزیه و تحلیل نه‌تنها اشیای نقطه‌یـی بلکه اشیاب مبتنی بر چندضلعی پیچیده‌تر را تسهیل می‌کند.

کلستربندی BIRCH

الگوریتم کلستربندی BIRCH یکی از الگوریتم‌های مطرح در کلستربندی سلسله مراتبی است. این روش با کمک تکنیک کلستربندی سلسله‌مراتبی و تکنیک دیگر (معمولاً k-means) برای کاربر روی دیتاهای عددی با حجم بالا طراحی شـده اسـت. این الگوریتم از ویژگی کلسـتر یا به اختصـار CF جهت انجـام کلستربندی استفاده مـی‌کند و همین امر سبب شده تا الگوریتم از سـرعت و قابلیت مقیاس‌پذیری بالایی برخوردار باشد. ویژگی کلستر، یک روش خلاصه‌سازی اطلاعات است که نقاط دیتا برای هر زیر کلستر را نگهداری می‌نماید. الگوریتم BIRCH از درخت متوازن ا ستفاده می شود. درخت متوازن CFها را جهت اجرای الگوریتم ذخیره مـی‌کند. هر گروپ از این درخت که دارای گره فرزند اسـت، در خود مجموع CFهای فرزندانش نگهداری می نماید. بدین‌ترتیب اطلاعات خلاصـه شـده‌یی از فرزندان خود را خواهد داشـت. درخت دارای دو پارامتر به نام‌های فاکتور کلستربندی B واحد آستانه T ا ست. حد اکثر تعداد فرزندان گره‌های غیرپایانی با فاکتور کلستربندی B مشخص می‌شوند و حد آستانه T به حداکثر فاصله‌ی که میان دو نمونه از کلسترهـای برگ قرار دارد، اطلاق می‌شود.

پس از درج و یا حذف یک CF در درخت و یا درصورت افزایش مقادیر B و T از مقدار تعیین شده توسط کاربر، گره‌های درخت متوازن تقسیم و یا ادغام می‌شوند. واضح است که مقادیر B و T در اندازه‌ی درخت تولیدشده نهایی، نقش به‌سزایی ایفا می‌کند.

کلستربندی DBSCAN

الگوریتم Density-based spatial clustering of applications with noise به اختصار DBSCAN به معنای کلستربندی بر اساس تراکم فضایی از کاربرد با نویز، نوعی از کلستربندی است که در سال ۱۹۹۶ توسط مارتین استر و همکاران پای به عرصه وجود نهاد. همان طوری‌که از نام این الگوریتم پیدا است، این الگوریتم برمبنای الگوریتم بر تراکم نقاط دیتا استوار است؛ به این معنا که نقاط دیتا در هر کجای فضا دارای بیشترین تراکم باشد، آن‌ها را به‌عنوان یک کلستر در نظر می‌گیرد.

در الگوریتم DBSCAN نیز مجموعه‌یی از نقاط در برخی فضاها، نقاط با همسایگان مجاور خود به‌عنوان یک گروپ دسته‌بندی می‌شوند و تنها در مناطقی که دارای کم‌ترین تراکم اند (که از همسایگی بسیار دور هستند)، به‌عنوان دیتاهای خارج از محدوده آن‌ها را حساب می‌کند. برخلاف بسیاری از روش‌های کلستربندی، DBSCAN نسبت به‌شکل دیتا حساس نیست و می‌تواند دیتا شکل‌های بسیار پیچیده را نیز گروپ‌بندی کند.

استفاده از کلستربندی در دیتاکاوی تحصیلی

این بخش یک نمای کلی از تجزیه و تحلیل کلستربندی در زمینه‌ی تحصیلی ارائه می‌دهد. ما چندین نشریه را از سال ۲۰۱۱ تا ۲۰۱۹ بررسی کردیم و تحقیقات خویش را بر اساس هدف و الگوریتم‌های مورد استفاده سازماندهی نمودیم. هدف این است که این تحقیق بینش‌های ارزشمندی را در باره‌ی این‌که چطور می‌توانیم از تجزیه و تحلیل کلستربندی در محیط تحصیلی استفاده نماییم، ارائه می‌دهد.

جدول۱: بررسی تحلیل کلستربندی در تعلیم و تربیه از سال ۲۰۱۱ تا ۲۰۲۰

نوی‌سنده	هدف	الگوریتم
(6)	طبقه‌بندی	K-Means
(7)	ارزیابی	K-Means
(4)	ارزیابی	EM
(8)	ارزیابی	Hierarchical EM, X-Means, K-Means,
(9)	ارزیابی	Ward, K-Means
(5)	ارزیابی	K-Means
(10)	طبقه‌بندی	,K-Means, X-Means, EM, Hierarchical
(11)	ارزیابی	K-Means
(3)	طبقه‌بندی	SOM, Fuzzy Clustering K-Means,
(12)	طبقه‌بندی	Unknown
(13)	طبقه‌بندی	K-Means
(14)	اکتشاف	K-Means
(15)	اکتشاف	K-Means
(16)	اکتشاف	clustering K-Means, Fuzzy

کلستربندی در تعلیم و تربیه به‌طور گسترده توسط تعدادی از نویسنـدگان بررسـی شده و در جـدول۱ خلاصه شده است. در جدول۱ اهداف مختلفی را که تجزیه و تحلیل کلستربندی برای آن‌ها به‌کار گرفته شده است و در صورت وجود، الگوریتمی‌را که برای تجزیه و تحلیل واقعی استفاده می‌شود، برجسته شده است. در این جدول، هدف تجزیه و تحلیل را به سه دسته طبقه‌بندی، ارزیابی و اکتشاف گروپ‌بندی شده است. گروپ طبقه‌بندی با استفاده از الگوریتم‌های کلستربندی برای فورمول‌بندی مدل‌هایی اشاره دارد که می‌توانند متعاقباً برای کارهای طبقه‌بندی مانند پیش‌بینی مورد استفاده قرار گیرند. گروپ ارزشیابی به استفاده از تکنیک‌های کلستربندی برای ایجاد گروپ‌ها به‌منظور دستیابی به درک بهتر از رفتارهای کاربران و محتوا در محیط تحصیلی اشاره دارد. اکتشاف از تکنیک‌های کلستربندی برای ارائه بلوک ساختمانی برای تجزیه و تحلیل بیشتر دیتاها استفاده می‌کند. ممکـن است برای تجـزیه و تحلیل کلستربندی بعدی یا ورودی به سایر تکنیک‌های تجزیه و تحلیل باشد.

با استفاده از این جدول، می‌توانیم تعدادی روند جالب را یادداشت کنیم. اولاً، الگوریتم K-Means یک انتخاب محبوب برای انجام تجزیه و تحلیل کلستربندی در بین نویسندگان باقی مانده است. همان‌طور که انتظار می‌رود، اکثر مطالعات از تجزیه و تحلیل کلستربندی برای ارزیابی استفاده می‌کنند، با این‌حال بیست‌درصد (۲۰%) از کلستربندی برای ایجاد مدل‌هایی برای کاوش بیشتر استفاده می‌کنند و سی و شش درصد (۳٦%) از آن به‌عنوان مبنایی برای طبقه‌بندی بعدی استفاده می‌کنند. درحالی‌که

برخی از نویسندگان به تنهایی از تجزیه و تحلیل کلستربندی استفاده می‌کنند، تعدادی از نویسندگان مانند Campagni و همکاران، (۲۰۱۵)، از تکنیک‌های کلستربندی در ارتباط با الگوهای متوالی یا سایر الگوهای غیر مرتبط برای استخراج اطلاعات استفاده می‌کنند.

یکی از مشاهدات جالب بررسی مقالات این است که بسیاری از نویسندگان پیشنهاد می‌کنند که برای تعمیم نتایج به‌دست‌آمده از آزمایش‌های‌شان، نیاز به آزمایش نتایج در سطح مقطع وسیع‌تری از دروس وجود دارد. آن‌ها هم‌چنین نشان دادند که تعداد صفات رابطه‌ی قوی با دقت الگوریتم‌ها دارد. برخی از عمل‌کرد الگوریتم هنگام حذف ویژگی‌ها کاهش می‌یابد.

نتیجه‌گیری

تجزیه و تحلیل کلستربندی به ویژه در طول ارزیابی و تجزیه و تحلیل دیتاهای اکتشافی مهم است. جایی‌که محققان تلاش می‌کنند روندهای اساسی را که بدون دانش قبلی در مورد دیتاهای تولید شده در مؤسسات تحصیلات عالی/ پوهنتون‌ها وجود دارد، کشف کنند. با این‌حال، انتخاب تکنیک و الگوریتم کلستربندی با دانش ساختار دیتا، انواع تحلیل‌هایی که باید ترسیم شوند و اندازه‌ی مجموعه دیتای ارزیابی شده تعیین می‌شود. بررسی ما از تجزیه و تحلیل کلستربندی در تحصیلات در جدول۱ نشان می‌دهد که اکثر مقالات تحقیقی از کلستربندی برای ارزیابی و اکتشاف استفاده کرده‌اند و در بسیاری از موارد فقط از الگوریتم K-Means استفاده شده است. هدف این مطالعه گسترش دانش در مورد عمل‌کرد الگوریتم‌های کلستربندی با در نظر گرفتن یک تحلیل مبتنی بر موردی از عمل‌کرد تطبیقی در بین دسته‌های تکنیک کلستربندی است.

برای به‌دست آوردن بینشی در مورد این‌که چگونه تجزیه و تحلیل کلستربندی می‌تواند در دوره‌های مختلف انجام شود، عمل‌کرد سه الگوریتم نماینده در اندازه‌های مختلف مجموعه‌دیتاها مورد ارزیابی قرار گرفت.

نتایج تجربی تأیید کرد که الگوریتم K-Means مبتنی بر طبقه‌بندی بهتر از الگوریتم BIRCH سلسله‌مراتبی و الگوریتم DBSCAN مبتنی بر تراکم عمل کرد. کلسترهایی با ضرایب بالاتر نسبت به دو الگوریتم دیگر برای مجموعه دیتاهای ۲ بُعدی و ۷ بُعدی تولید می‌کند. بنابراین، روش‌های مبتنی بر طبقه‌بندی را می‌توان مناسب‌ترین تکنیک برای انجام تحلیل کلستربندی دیتاهای محیط تحصیلی دانست. عمل‌کرد خوب روش مبتنی بر طبقه‌بندی و عمل‌کرد نسبتاً پایین روش مبتنی بر تراکم نشان می‌دهد که دیتاهای تولید شده توسط گزارش‌های مؤسسات تحصیلات عالی، تمایل به تشکیل

گروپ‌هایی در اطراف مرکز (نقطه یا شی) دارند و تجزیه و تحلیل با استفاده از این فرض در برخی از موارد، نتایج مفیدتر برای تحلیل اکتشافی گروپ‌بندی‌ها را تولید می‌کند.

عمل‌کرد جالب الگوریتم BIRCH نیاز به بررسی این امکان، اشاره دارد که ممکن است ساختارهای سلسله‌مراتبی در دیتاهای گزارش تحصیلی وجود داشته باشد. درک این سلسله‌مراتب ممکن بینش جالبی برای درک روابط موجود بین ابعاد مختلف و تعامل کاربر در تحصیل ارائه دهد.

منابع

(1) Martins MPG, Miguéis VL, Fonseca DSB. Data Mining Educacional: Uma Revisão da Literatura Educational Data Mining: A Literature Review (Internet]. Available from: www.educationaldatamining.org.

(2) Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. In: Proceedings - IEEE International Conference on Data Mining, ICDM. 2010. pp. 911–6.

(3) Mamdouh Fouad M, Yu PS, Fattah Mashat A, Fouad MM, Arabia Tarek Gharib SF. Efficient Clustering Technique for University Admission Data [Internet]. Vol. 45, International Journal of Computer Applications. 2012. Available from: https://www.researchgate.net/publication/235333407

(4) Bogarín A, Romero C, Cerezo R, Sánchez-Santillán M. Clustering for improving Educational process mining. In: ACM International Conference Proceeding Series. Association for Computing Machinery; 2014. pp. 11–5.

(5) Dutt A, Ismail MA, Herawan T. A Systematic Review on Educational Data Mining. Vol. 5, IEEE Access. Institute of Electrical and Electronics Engineers Inc.; 2017. pp. 15991–6005.

(6) Harwati, Alfiani AP, Wulandari FA. Mapping Student's Performance Based on Data Mining Approach (A Case Study). Agriculture and Agricultural Science Procedia. 2015; 3:173–7.

(7) Campagni R, Merlini D, Sprugnoli R, Verri MC. Data mining models for student careers. Expert Systems with Applications. 2015 Aug 1; 42(13):5508–21.

(8) Liu Y, Li Z, Xiong H, Gao X, Wu J, Wu S. Understanding and enhancement of internal clustering validation measures. IEEE Transactions on Cybernetics. 2013 Jun; 43(3): 982–94.

(9) Ghorpade-Aher J, Metre VA. PSO based Multidimensional Data Clustering: A Survey. Vol. 87, International Journal of Computer Applications. 2014.

(10) López MI, Luna JM, Romero C, Ventura S. Classification via clustering for predicting final marks based on student participation in forums.

(11) El-Halees A, Tair MMA, El-Halees AM. Mining Educational Data to Improve Students' Performance: A Case Study Cite this paper Mining Educational Data to Improve Students' Performance: A Case Study. 2012; 2(2). Available from: http://www.esjournals.org.

(12) DeFreitas K, Bernard M. COMPARATIVE PERFORMANCE ANALYSIS OF CLUSTERING TECHNIQUES IN EDUCATIONAL DATA MINING. IADIS International Journal on Computer Science and Information Systems. 10(2): 65–78.

(13) Durairaj M, Vijitha C. Educational Data mining for Prediction of Student Performance Using Clustering Algorithms [Internet]. Available from: www.ijcsit.com.

(14) Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. Vol. 41, Expert Systems with Applications. 2014. pp. 1432–62.

(15) Romero C, Ventura S. Educational data science in massive open online courses. Vol. 7, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Wiley-Blackwell; 2017.

(16) Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining