

## داده‌کاوی نمونه‌های مانیتور شده هدایت الکتریکی و اسیدیته خاک برای پهنه‌بندی دشت قم با دو روش خوشه‌بندی فازی

سید مهدی هاشمی<sup>۱</sup>، مهدی نورزاده حداد<sup>۲</sup>، محمد جواد منعم<sup>۳</sup>، مصطفی کریمیان اقبال<sup>۴</sup>

<sup>۱</sup>دانشجوی دکتری سازه‌های آبی دانشگاه تربیت مدرس، <sup>۲</sup>دانشجوی کارشناسی ارشد خاک‌شناسی دانشگاه تربیت مدرس،

<sup>۳</sup>دانشیار سازه‌های آبی دانشگاه تربیت مدرس، <sup>۴</sup>دانشیار خاک‌شناسی دانشگاه تربیت مدرس

### مقدمه

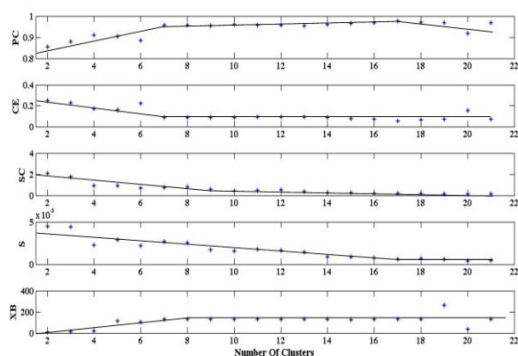
خوشه‌بندی فازی یکی از روش‌های متداول داده‌کاوی است. فازی بودن این الگوریتم خوشه‌بندی سبب می‌شود که در مواقعی که نتوان یک مرز مشخص و معلومی بین خوشه‌ها اعمال نمود، یک داده در یک زمان متعلق به چندین زیر مجموعه باشد. در این تحقیق از روش خوشه‌بندی برای دسته‌بندی نمونه‌های خاک برداشت شده استفاده شده است. هدف از خوشه‌بندی، دسته‌بندی اشیاء بر اساس میزان مشابهت بین آنها و سازماندهی گروهی خوشه‌ها می‌باشد. اصلی‌ترین ظرفیت خوشه‌بندی، شناسایی ساختار اصلی نهفته در داده‌ها می‌باشد [۱]. در این مطالعه با استفاده از دو روش فازی **C-Means** و **GK**، داده‌های هدایت الکتریکی و اسیدیته خاک که دو ویژگی بسیار مهم شیمیایی خاک برای کشاورزی می‌باشند را خوشه‌بندی و در نهایت پهنه‌بندی آنها به صورت نقشه‌های فازی برای ۱۱۴۰۰۰ هکتار از اراضی کشاورزی دشت قم به دست آمده است.

### مواد و روش‌ها

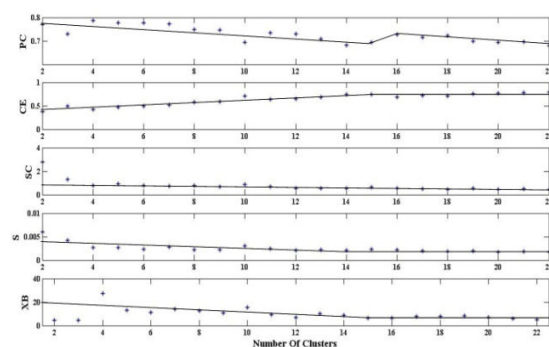
منطقه مورد مطالعه شامل ۱۱۴۰۰۰ هکتار از اراضی کشاورزی دشت قم می‌باشد. میانگین بارندگی سالیانه منطقه مورد مطالعه ۱۴۵ میلیمتر و متوسط درجه حرارت روزانه ۲۰/۴ درجه سانتیگراد گزارش شده است [۲]. در این تحقیق با استفاده از دو روش فازی **C-means** و **GK** خوشه‌بندی برای ۴۶۵ داده‌های هدایت الکتریکی و اسیدیته خاک به دست آمده است. همچنین به منظور صحت‌سنجی و مقایسه این دو روش از معیارهای معیار ضریب تقسیم‌بندی (**PC**)، معیار گشتاور رده‌بندی (**CE**)، شاخص تقسیم‌بندی **SC**، شاخص جداشدگی (**S**) و شاخص زای و بنی (**XB**) استفاده شده است [۳].

### نتایج و بحث

هر کدام از مدل‌های خوشه‌بندی ۲۲ مرتبه به کمک نرم‌افزار **MATLAB** اجرا گردید. شکل‌های ۱ و ۲ مقادیر ۵ شاخص صحت‌سنجی خوشه‌بندی فازی در مقابل تعداد خوشه‌ها را ترسیم کرده است.



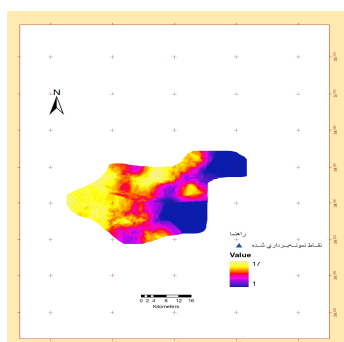
شکل ۲. مقادیر شاخص‌های صحت‌سنجی در مقابل تعداد خوشه‌ها در روش



شکل ۱. مقادیر شاخص‌های صحت‌سنجی در مقابل تعداد خوشه‌ها در روش **FCM**

**GK**

بر اساس شکل ۱ شاخص **PC** تعداد ۱۶ خوشه را بهترین تعداد خوشه‌ها معرفی می‌کند. شاخص **CE** از تعداد ۱۵ خوشه به بعد تقریباً در مقدار ماکزیمم خود ثابت شده و تغییر چشمگیری در مقادیر آن رخ نداده است. بنابراین ۱۵ خوشه تعداد بهینه خوشه‌ها از دید شاخص **CE** می‌باشد. شاخص **SC** روند ثابتی داشته بنابراین تعداد بهینه خاصی را برای خوشه‌ها نشان نمی‌دهد. شاخص **S** از تعداد ۱۴ و ۱۵ خوشه به حداقل خود رسیده و ثابت شده و همچنین شاخص **XB** در تعداد ۱۵ خوشه به حداقل خود رسیده و از این تعداد به بعد مقادیرش ثابت گردیده است. بنابراین با توجه به نتایج بدست آمده از شاخص‌های صحت‌سنجی فازی، تعداد بهینه خوشه‌ها برای روش **FCM** برابر با ۱۵ خوشه است ولی بر اساس شکل ۲ شاخص **PC** در تعداد ۱۷ خوشه به حداکثر مقدار خود رسیده و ۱۷ خوشه را به عنوان تعداد بهینه خوشه‌ها معرفی می‌کند. شاخص **CE** از تعداد ۷ خوشه به بعد تقریباً در مقدار حداقل خود ثابت شده و از آن آنجا که این شاخص در مقدار حداکثر خود تعداد بهینه خوشه‌ها را نشان می‌دهد، بنابراین این شاخص تعداد خوشه مشخصی را معرفی نمی‌کند. شاخص **SC** روند ثابت نزولی را طی می‌کند و تغییر شیب محسوس در مقادیر این شاخص در تعداد ۹ خوشه حاصل گردیده است بنابراین تعداد بهینه خوشه‌ها برابر ۹ خوشه است. شاخص **S** از تعداد ۱۷ خوشه به حداقل خود رسیده و ثابت شده است. اما شاخص **XB** از تعداد ۸ خوشه به بعد در مقدار حداکثر ثابت گردیده و چون این شاخص به مقدار حداقل خود متمایل نگردیده است پس این شاخص تعداد بهینه خوشه‌ها را نشان نمی‌دهد. با توجه به نتایج بدست آمده از شاخص‌های صحت‌سنجی فازی، تعداد بهینه خوشه‌ها برای روش فازی **GK** برابر با ۱۷ خوشه است. برای تعیین مناسب‌ترین روش خوشه‌بندی فازی از معیار صحت‌سنجی **SC** استفاده می‌شود. بدین منظور مقادیر این شاخص در تعداد ۱۵ و ۱۷ خوشه برای دو روش خوشه‌بندی **FCM** و **GK** مقایسه گردیدند. همان طور که از شکل‌های ۱ و ۲ نیز قابل مشاهده است مقادیر شاخص مذکور برای روش **GK** به طور محسوسی کمتر از مقادیر همین شاخص در روش **FCM** می‌باشد. بنابراین روش خوشه‌بندی فازی **GK** روش مناسب‌تری جهت خوشه‌بندی داده‌های مانیتور شده دشت قم می‌باشد. مناسب‌تر بودن روش **GK** در خوشه‌بندی را می‌توان به استفاده از معیار فاصله تطابق‌پذیر جهت تعیین خوشه‌ها با اشکال مختلف در فضای داده‌ها، در این روش نسبت داد. جهت ترسیم نقشه پهنه‌بندی دشت قم بر اساس توزیع دو خاصیت هدایت الکتریکی و اسیدیته خاک، از نرم‌افزار **Arc GIS** استفاده گردیده است. بر این اساس همان طور که در شکل شماره ۳ مشخص شده است، هر کدام از خوشه‌ها با یک



رنگ مشخص شده و میزان درجه عضویت هر کدام از نقاط نمونه‌گیری به هر کدام از خوشه‌ها با درجه پررنگ بودن و کم‌رنگ بودن آن نقطه معلوم شده است. در بالای سمت راست شکل ۳ نقاط بیشتر به خوشه شماره ۱، که با رنگ آبی تیره معلوم شده، اختصاص دارند. با نزدیک شدن به مرکز دشت، درجه عضویت نقاط نمونه‌گیری به خوشه ۱ کاسته می‌شود که این پدیده با کم‌رنگ شدن رنگ آبی در این نقاط قابل مشاهده می‌باشد نقشه تهیه شده به این روش این قابلیت را دارد که تعلق هر نقطه نمونه‌برداری شده را به هر خوشه تعیین و بررسی کند در حالی که در پهنه‌بندی‌هایی که به طور معمول برای خصوصیات خاک تهیه می‌شود، تعداد و تعلق هر نقطه به یک

خوشه مشخص نمی‌شود. همچنین پهنه‌بندی صورت گرفته برای هدایت الکتریکی و اسیدیته خاک، این قابلیت را دارد که به‌طور همزمان این دو خصوصیت را پهنه‌بندی و خوشه‌بندی آن را به طور قابل فهم نمایش دهد. شکل ۳. پهنه‌بندی میزان درجه عضویت هر کدام از نقاط نمونه‌گیری به هر خوشه.

## منابع

[1] Han, J. and M. Kamber, 2006. Data Mining Concepts and Techniques. San Francisco, U.S.A, Morgan Kaufman Publishers.

[۲] گزارش نیمه تفصیلی قم-مسئله، مؤسسه تحقیقات خاک و آب، مردادماه ۱۳۶۲.

[3] Kim, D.W. and Lee, K.H. and Lee,D.(2004). On cluster validity index for estimation of the optimal number of fuzzy clusters. Journal of Pattern Recognition Society, 37,pp 2009-2025.