

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

مقایسه عملکرد روش خوشه‌بندی k-means با شاخص‌های طیفی در بارزسازی کاربری و پوشش اراضی به کمک سنجش از دور با پایتون (مورد مطالعاتی: تصاویر چند طیفی ماهواره لندست ۸)

دکتر خلیل رضایی<sup>۱\*</sup>، مهدی فرمهینی فراهانی<sup>۲</sup>

۱- دانشیار رسوب شناسی دانشکده علوم زمین، دانشگاه خوارزمی تهران [khalil.rezaei@khu.ac.ir](mailto:khalil.rezaei@khu.ac.ir)

۲- دانشجوی کارشناسی زمین شناسی دانشگاه خوارزمی [mahdifarmahinifarahani@aol.com](mailto:mahdifarmahinifarahani@aol.com)

### چکیده

با وجود پیشرفت روزافزون فناوری و همه گیر شدن برنامه‌نویسی در شاخه‌های مختلف علوم پایه، الزام استفاده و یادگیری برنامه‌نویسی در شاخه‌های مختلف علوم زمین امری غیر قابل انکار است. یکی از موارد مورد توجه دنیای فناوری امروز، مبحث هوش مصنوعی و استفاده از یادگیری ماشین در انجام بدون نظارت کارهای روزمره است و با وجود توسعه بسته‌های مختلف در بستر زبان برنامه‌نویسی پایتون انجام کارهای سنجش از دور برای متخصصان این حوزه راحت‌تر از پیش شده است. در این مقاله تلاش بر آن شده است که با استفاده از الگوریتم خوشه‌بندی k-means و به کمک تصاویر مولتی اسپکترال لندست ۸ نمونه‌ای از توانایی زبان برنامه‌نویسی پایتون را در انجام فرایندهای خوشه‌بندی نمایش دهیم.

**کلمات کلیدی:** خوشه‌بندی، سنجش از دور، برنامه‌نویسی، پایتون، یادگیری ماشین

### ۱. مقدمه

امروزه اهمیت علوم زمین بدلیل افزایش نیاز صنایع و تکنولوژی‌های مختلف به مواد اولیه از یکسو و نیز مخاطرات گوناگون طبیعی برکسی پوشیده نیست. همچنین استفاده از هوش مصنوعی و تمام شاخه‌های آن از جمله یادگیری ماشین نیز با سرعت قابل توجهی در حال گسترش است. این بکارگیری شامل زمینه‌های مختلف از جمله پیش بینی رفتار آبی مواد طبیعی [10]، انجام کارها با دقت اما بدون نظارت [10]، وقوع یا عدم وقوع مخاطرات یا پیش بینی تقریبی زمان آنها (رفرنس)، کمک به جلوگیری از افزایش خسارات ناشی از مخاطرات طبیعی [29]، کمک به اکتشاف و تشخیص زون‌های حاوی مواد معدنی و یا نفت و گاز [18] می‌باشد.

پایتون یکی از نرم افزارهای قوی و جدید مورد استفاده از بسیاری از علوم از جمله زمین شناسی است که با دارا بودن زبان برنامه نویسی قدرتمند خود بطور روزافزونی در حال گسترش و تکمیل می‌باشد. این امر هم می‌تواند خود به انجام برخی تحقیقات منجر شده و هم برای مقایسه میزان دقت و صحت خروجی نرم افزارهای دیگر با هم و نیز خروجی خود زبان پایتون بکار رود. در همین راستا، در این مقاله تلاش بر آن شده است که با استفاده از الگوریتم خوشه‌بندی k-means و به

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

کمک تصاویر مولتی اسپکترال لندست ۸ نمونه‌ای از توانایی زبان برنامه‌نویسی پایتون را در انجام فرایندهای خوشه‌بندی نمایش داده شود.

### ۲. روش و ابزار

#### ۲.۱.۱ یادگیری ماشین

یادگیری ماشین مبحثی است که بدون برنامه‌ریزی مستقیم (نظارت) به کامپیوترها قابلیت یادگیری می‌دهد. [1] در یادگیری بدون نظارت به علت کمبود اطلاعات موجود ماشین به دنبال اشتراکات است که با کمک آن‌ها بتواند داده‌ها را خوشه‌بندی کند، برآورد تراکم انجام دهد و در نهایت ابعاد را کوچک‌تر کند. [2, 3] در خوشه‌بندی، گروه بندی مجموعه‌ای از اشیاء انجام می‌شود. این کار به این صورتی که در یک خوشه قرار دارند، در مقایسه با اشیاء دیگر خوشه‌ها مشابه‌تر هستند. این یک روش معمول برای تجزیه و تحلیل داده‌های آماریه. خوشه‌ها شامل گروه‌هایی با فاصله‌های کم بین اعضای خوشه، مناطق متراکم فضای داده، فواصل و یا توزیع‌های آماری خاص هستند. در نتیجه خوشه بندی می‌تواند به عنوان یک مسئله بهینه سازی چند هدفه صورت بگیرد. انتخاب الگوریتم خوشه بندی مناسب و تنظیمات پارامتر (از جمله پارامترهایی مانند تابع فاصله مورد استفاده، آستانه تراکم یا تعداد خوشه مورد انتظار) بستگی به تنظیم مجموعه داده‌ها توسط فرد و استفاده خاص فرد از نتایج دارد. تجزیه و تحلیل خوشه‌ای یک روش اتوماتیک نیست، بلکه یک فرآیند تکراری از کشف دانش یا بهینه سازی چند هدفه تعاملی هست که در اون آزمایش و شکست وجود دارد. توی خوشه بندی اغلب لازم است که داده‌ها پیش پردازش بشوند و پارامترهای مدل اصلاح بشوند تا نتیجه حاصل، همان نتیجه دلخواه باشد. [4, 5, 6]

#### ۲.۱.۲ خوشه بندی

در روش خوشه بندی در مجموعه داده‌ای با  $n$  مشاهده و  $d$  بعد و با فرض تعداد خوشه‌های مورد نیاز  $k$  تایی هدف اصلی جدا کردن داده‌ها به  $k$  خوشه بصورت  $C = \{C_1, C_2, C_3, \dots, C_k\}$  است. هر خوشه نماینده‌ای مانند  $\mu_i$  دارد که معمولا مرکز خوشه نامیده می‌شود. [7, 8] نحوه محاسبه مرکز خوشه بصورت زیر است:

$$\mu_i = \frac{1}{n} \sum_{x_j \in C_i} x_j \quad (1)$$

همچنین نحوه محاسبه تعداد دقیق حالات ممکن برای خوشه‌بندی  $n$  داده در  $k$  خوشه بصورت زیر است [9]:

$$S(n, k) = \frac{1}{k!} \sum_{t=0}^k (-1)^t \binom{k}{t} (k-t)^n \quad (2)$$

تابع برای بررسی کیفیت خوشه بندی که بر اساس خطای مجموع مربعات ایجاد شده بصورت زیر است [12, 13, 14]:

$$SSE(C) = \sum_{i=1}^k \sum_{x_j \in C_i} |x_j - \mu_i|^2 \quad (3)$$

روش خوشه بندی  $k$ -means جزء روش‌های یادگیری بدون نظارت در نظر گرفته می‌شود که هدف آن خوشه‌بندی کردن داده‌ها بر اساس ویژگی‌های آن‌ها است. [10] این الگوریتم از  $k$  نقطه تصادفی در فضای حالت شروع می‌کند و بصورت تکراری داده‌ها را به نزدیک‌ترین مرکز خوشه نسبت می‌دهد و مراکز خوشه را با استفاده از آن بروزرسانی می‌کند. این عمل تا آن‌حایی ادامه پیدا می‌کند که فاصله میانگین‌های دو مرحله پیاپی کمتر از سطح حساسیت مورد نظر باشد و به گفته دیگر بهبودی برای خوشه‌بندی‌ها میسر نباشد. [11]

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

$$\sum_{i=1}^k |\mu_j^t - \mu_{j-1}^{t-1}|^2 \leq \varepsilon$$

(4)

$\varepsilon > 0$  سطح حساسیت است و همواره باید بزرگ‌تر از ۰ باشد

$\varepsilon > 0$

(5)

البته لازم به ذکر است که روش خوشه بندی k-means دقت کمی در پردازش داده‌های غیر محدب دارد [11] الگوریتم‌های متعددی برای روش توسعه داده شده است که برخی از آن‌ها عبارتند از الگوریتم لوید، الگوریتم فورجی، الگوریتم هاتیگان - ونگ، الگوریتم مک کوئین. [15]

۲.۱.۳ پردازش تصویر

به مجموعه فرایندهایی که بر روی تصویر به منظور بهبود یا دستکاری تصویر صورت می‌گیرد پردازش تصویر می‌گویند. [19] پردازش تصویر فراندیست که به صورت پیکسل به پیکسل انجام می‌شود و با اعمال مجموعه‌ای از توابع (که با توجه به نوع فعالیت متفاوت است) روی هر پیکسل صورت می‌گیرد [19] در فرایندهای پردازش تصویر غالباً از الگوریتم‌های یادگیری استفاده می‌شود. [20]

۲.۲ شاخص‌های طیفی

در علم سنجش از دور استفاده از شاخص‌های طیفی امری مهم برای بارزسازی پدیده‌های سطح زمین (برای مثال پوشش گیاهی) است. شاخص‌ها مجموعه‌ای از فرمول‌های ریاضی هستند که با هدف بارزسازی یک پدیده‌ی سطح زمین، بر روی چند باند یک تصویر چند طیفی یا فراطیفی اعمال می‌شود. اساس کار شاخص‌های طیفی بر مبنای جذب طیفی است. در این مقاله ما به کمک شاخص‌های NDVI، SAVI، NDWI و UI محیط خود را مورد بررسی و مطالعه قرار می‌دهیم و پس از آن دقت نتیجه حاصله از روش خوشه بندی را مورد بحث قرار می‌دهیم

2.2.1 شاخص NDVI

شاخص NDVI (شاخص نرمال شده تفاوت پوشش گیاهی) برای محاسبه پوشش گیاهی مورد استفاده قرار می‌گیرد نحوه محاسبه شاخص NDVI به صورت زیر است:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

(6)

که در آن متغیر NIR مربوط به باند بازتابی نزدیک فرو سرخ است و متغیر Red مربوط به باند جذبی قرمز است. [27]

2.2.2 شاخص NDWI

شاخص NDWI (شاخص آب تفاضلی نرمال شده) برای محاسبه و بارزسازی آب تصویر مورد استفاده قرار می‌گیرد. نحوه محاسبه شاخص NDWI به صورت زیر است:

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

(7)

که در آن متغیر NIR مربوط به باند بازتابی نزدیک فرو سرخ است و متغیر Green مربوط به باند جذبی سبز است. [28]

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

### 2.2.3 شاخص SAVI

شاخص SAVI یکی از پرکاربردترین شاخص‌های گیاهی است که اثرات خاک را نیز در این شاخص در نظر می‌گیرد. اثر خاک در این شاخص با پارامتری به اسم L مشخص می‌شود. برای پوشش گیاهی ضعیف این مقدار صفر که در این صورت این شاخص با شاخص NDVI برابر است. نحوه محاسبه شاخص SAVI به صورت زیر است:

$$SAVI = \frac{(NIR - Red)}{(NIR + Red + L)} \times (1 + L) \quad (8)$$

که در آن متغیر NIR مربوط به باند بازتابی نزدیک فرو سرخ است، متغیر Red مربوط به باند جذبی قرمز است و پارامتر L به اثر خاک مربوط می‌شود و معمولا بر بروی 0.5 تنظیم می‌شود. [26]

### 2.2.4 شاخص UI

شاخص UI یکی از پرکاربردترین شاخص‌های شهری است. نحوه محاسبه شاخص SAVI به صورت زیر است:

$$UI = \frac{(SWIR2 - NIR)}{(SWIR2 + NIR)} \quad (9)$$

که در آن متغیر SWIR2 مربوط به باند بازتابی فروسرخ با طول موج کوتاه است و متغیر NIR مربوط به باند جذبی نزدیک فرو سرخ است. [25]

### ۲.۳ لندست ۸

ماهواره لندست ۸ یکی ماهواره‌های ناسا می‌باشد که در تاریخ ۱۱ فوریه ۲۰۱۳ پرتاب شد و به فعالیت‌های رصد زمین مشغول شد. لندست ۸ دارای دو سری سنجنده است: تصویربرداری زمین عملیاتی (OLI) و حسگر حرارتی مادون قرمز (TIRS) که در جدول زیر به عملکرد و مشخصات آن‌ها اشاره شده است. [16, 17]

Table 2. Landsat 8 bands properties and applications

Landsat 8 Sensors	Band	Band Name	Wavelength (μm)	Resolution (m)	Applications	
Operational Land Imager (OLI)	1	Coastal / Aerosol	0.433-0.453	30	Coastal and Aerosol studies	
	2	Blue	0.450-0.515		Bathymetric mapping, distinguishing soil from vegetation and deciduous from coniferous vegetation	
	3	Green	0.525-0.600		Emphasizes peak vegetation, which is useful for assessing plant vigour	
	4	Red	0.630-0.680		Discriminates vegetation slopes	
	5	Near Infrared (NIR)	0.845-0.885		Emphasizes biomass content and shorelines	
	6	Short-wave Infrared (SWIR) 1	1.560-1.660		Discriminates moisture content of soil and vegetation; penetrates thin clouds	
	7	Short-wave Infrared (SWIR) 2	2.100-2.300		Improved moisture content of soil and vegetation and thin cloud penetration	
	8	Panchromatic	0.500-0.680		15	Sharper image definition
	9	Cirrus	1.360-1.390		30	Improved detection of cirrus cloud contamination
Thermal Infrared Sensor (TIRS)	10	Long-wave Infrared (LWIR) 1	10.30 - 11.30	100*	Thermal mapping and estimated soil moisture	
	11	Long-wave Infrared (LWIR) 2	11.50-12.50		Improved thermal mapping and estimated soil moisture	
	BQA	Quality Assessment			Quality assessments for every pixel in the scene	

\* TIRS bands are acquired at 100 meter resolution, but are resampled to 30 meter in delivered data product.

جدول ۱: مشخصات و عملکرد سنسورهای ماهواره لندست ۸ [16]

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

۲.۴ پایتون

پایتون که یک زبان برنامه‌نویسی ماژولار و سطح بالا است. [18] ماژولار بودن آن به ما کمک می‌کند تا بتوانیم از کتابخانه‌های موجود و توسعه داده شده در آن استفاده کنیم. در این مطالعه تلاش بر آن شده است تا با بهره‌گیری از بسته‌های Matplotlib, NumPy, scikit-learn و scikit-image فرایندهای محاسبات علمی، پردازش تصویر و یادگیری ماشین برای سنجش از دور صورت بگیرد. [21, 22, 23, 24]

### ۳. نتایج و بحث

در این پژوهش از تصاویر چند طیفی تصاویر باندهای ۲، ۳، ۴، ۵، ۶ و ۷ (آبی، سبز، قرمز، نزدیک فرو سرخ و فرو سرخ با طول موج کوتاه ۱ و ۲) ماهواره لندست ۸ استفاده شده است. در ابتدا تصاویر را با استفاده از اسکرپیت زیر وارد محیط Jupyter Lab می‌کنیم.

```
import matplotlib.pyplot as plt
import numpy as np
from matplotlib import cm
from skimage import io
from skimage import exposure
import skimage as skm
from skimage import exposure, img_as_float
```

```
image1 = io.imread('E:/Landsat8/Blue.tif')
image2 = io.imread('E:/Landsat8/Green.tif')
image3 = io.imread('E:/Landsat8/Red.tif')
image4 = io.imread('E:/Landsat8/NIR.tif')
image5 = io.imread('E:/Landsat8/SWIR1.tif')
image6 = io.imread('E:/Landsat8/SWIR2.tif')
```

```
B1 = np.array(image1).astype(float)
B2 = np.array(image2).astype(float)
B3 = np.array(image3).astype(float)
B4 = np.array(image4).astype(float)
B5 = np.array(image5).astype(float)
B6 = np.array(image6).astype(float)
```

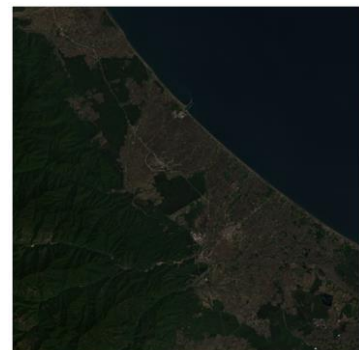
سپس با استفاده از اسکرپیت زیر تصاویر را نرمالایز می‌کنیم.

```
NB1 = (B1-np.min(B1))/(np.max(B1)-np.min(B1))
NB2 = (B2-np.min(B2))/(np.max(B2)-np.min(B2))
NB3 = (B3-np.min(B3))/(np.max(B3)-np.min(B3))
NB4 = (B4-np.min(B4))/(np.max(B4)-np.min(B4))
NB5 = (B5-np.min(B5))/(np.max(B5)-np.min(B5))
NB6 = (B6-np.min(B6))/(np.max(B6)-np.min(B6))
```

با استفاده از اسکرپیت زیر تصویر اولیه را نمایش می‌دهیم.

```
rgb = np.stack([NB3,NB2,NB1], axis=2)

plt.figure(figsize=(15,10))
plt.imshow(rgb)
plt.title('RGB Image')
plt.axis('off')
plt.show()
```



شکل ۱: تصویر RGB حاصل شده از ترکیب باندهای قرمز، سبز و آبی

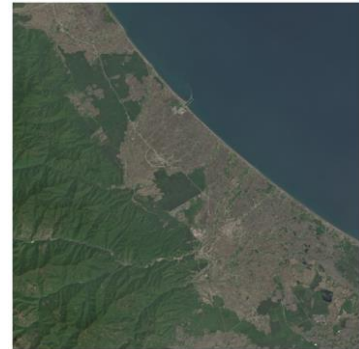
# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

با استفاده از اسکرپت زیر توسط روش Gamma تصویر را واضح‌تر می‌کنیم.

```
GEnNB1 = exposure.adjust_gamma(NB1 , gamma=0.5, gain=1)
GEnNB2 = exposure.adjust_gamma(NB2 , gamma=0.5, gain=1)
GEnNB3 = exposure.adjust_gamma(NB3 , gamma=0.5, gain=1)
GNstack = np.stack([GEnNB3,GEnNB2,GEnNB1], axis=2)

plt.figure(figsize=(15,10))
plt.imshow(GNstack)
plt.axis('off')
plt.show()
```



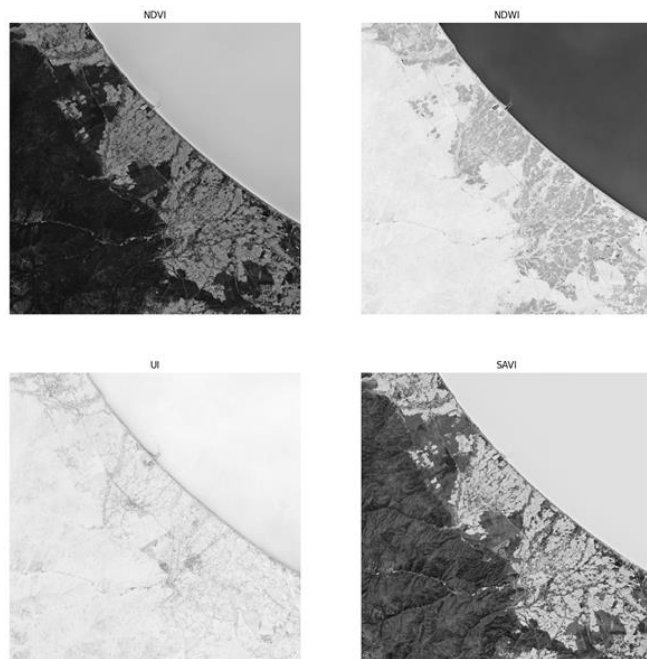
شکل ۲: تصویر RGB با اعمال اصلاحات متد Gamma

شاخص‌های طیفی معرفی شده را توسط اسکرپت‌های زیر معرفی می‌کنیم.

```
NDVI = (NB4-NB3) / (NB4+NB3)
NDWI = (NB2 - NB4) / (NB2 + NB4)
SAVI = ((NB4-NB3) / (NB4+NB3+0.5)) * (1+0.5)
UI = (NB6 - NB4) / (NB6 + NB4)
```

و در نهایت با استفاده از اسکرپت زیر تصویر حاصل از شاخص‌های مختلف را نمایش می‌دهیم.

```
fig , ax = plt.subplots(nrows=2, ncols=2, figsize=(15,15))
ax[0,0].imshow(NDVI, cmap=cm.Greys)
ax[0,0].set_title('NDVI')
ax[0,0].axis('off')
#####
ax[0,1].imshow(NDWI, cmap=cm.Greys)
ax[0,1].set_title('NDWI')
ax[0,1].axis('off')
#####
ax[1,0].imshow(UI, cmap=cm.Greys)
ax[1,0].set_title('UI')
ax[1,0].axis('off')
#####
ax[1,1].imshow(SAVI, cmap=cm.Greys)
ax[1,1].set_title('SAVI')
ax[1,1].axis('off')
plt.show()
```



شکل ۳: بالا سمت چپ و راست به ترتیب NDVI و NDWI ، پایین سمت چپ و راست به ترتیب UI و SAVI

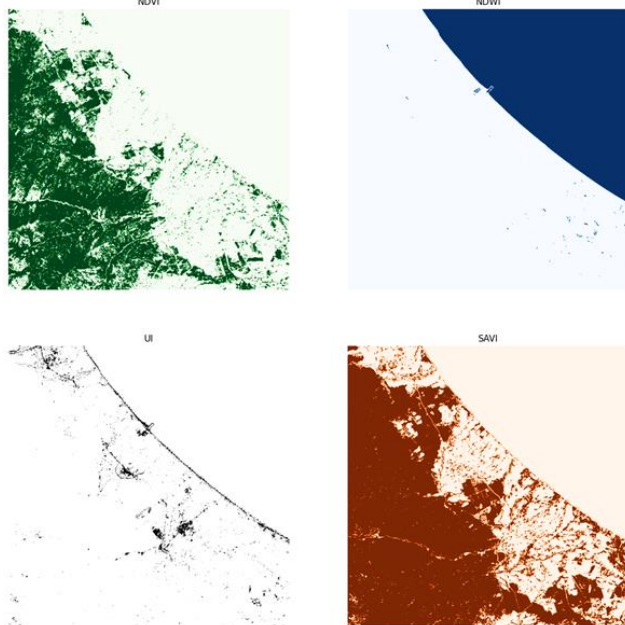
# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

تصویر چهارگانه فوق نتیجه حاصل شده از اعمال شاخص‌های طیفی NDVI، NDWI، UI و SAVI است.

برای درک هرچه بهتر تصویر حاصل شده از شاخص‌ها برای هر کدام از تصاویر یک بازه مشخص می‌کنیم بصورتی که تمام مقادیر بالای آن مقدار باقی بماند و باقی مقادیر حذف شود. برای این کار اسکریپت زیر را وارد می‌کنیم.

```
lai = 0.۷۰
veg = NDVI > lai
non_veg = NDVI <= lai
#####
water = 0.2
watero = NDWI > water
non_watero = NDWI <= water
#####
soilveg = 0.44
soil = SAVI > soilveg
non_soil = SAVI <= soilveg
#####
urbaniz = -0.28
urban = UI > urbaniz
non_urban = UI <= urbaniz
fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15,15))
ax[0,0].imshow(veg, cmap=cm.Greens)
ax[0,0].set_title('NDVI')
ax[0,0].axis('off')
#####
ax[0,1].imshow(watero, cmap=cm.Blues)
ax[0,1].set_title('NDWI')
ax[0,1].axis('off')
#####
ax[1,0].imshow(urban, cmap=cm.Greys)
ax[1,0].set_title('UI')
ax[1,0].axis('off')
#####
ax[1,1].imshow(soil, cmap=cm.Oranges)
ax[1,1].set_title('SAVI')
ax[1,1].axis('off')
plt.show()
```



شکل ۴: بالا سمت چپ و راست به ترتیب NDVI و NDWI، پایین سمت چپ و راست به ترتیب UI و SAVI

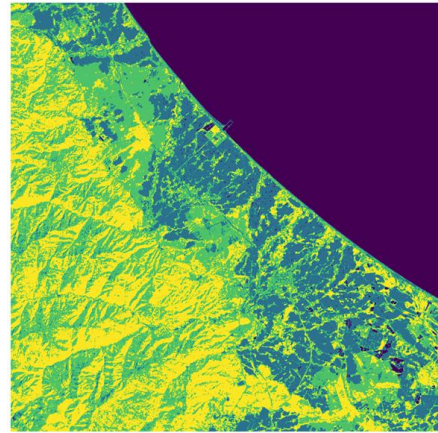
در ادامه برای استفاده از روش خوشه‌بندی k-means اسکریپت زیر را وارد می‌کنیم.

```
from sklearn.cluster import KMeans
from skimage import io
import matplotlib.pyplot as plt
image = io.imread('E:/Landsat8/NIR.tif')
row = image.shape[0]
col = image.shape[1]
X = image.reshape((row*col, 1))
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(X)
cluster_centers = kmeans.cluster_centers_
cluster_labels = kmeans.labels_
```

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

```
clusterd_image = cluster_centers[cluster_labels].reshape(row, col)
plt.figure(figsize=(15,15))
plt.imshow(clusterd_image)
plt.axis('off')
plt.show()
```



شکل 5: نتیجه حاصل شده از خوشه‌بندی K-means

تصویر فوق نتیجه در نظر گرفتن 4 دسته با استفاده از روش خوشه‌بندی k-means برای منطقه است.

#### ۴. نتیجه‌گیری

همان گونه که از تصاویر حاصل شده از اعمال الگوریتم خوشه‌بندی k-means پیداست، این الگوریتم توانسته است عملکرد خوبی در بارزسازی پوشش اراضی با توجه به نتایج حاصل شده از شاخص‌های طیفی داشته باشد. اگر چه با دقت بیشتر در نتیجه حاصل شده می‌توانیم شاهد خطاهایی در دسته‌بندی باشیم، اما با در نظر گرفتن نویزهای موجود در تصاویر می‌توانیم از آن‌ها صرف نظر کنیم. در کل می‌توان گفت استفاده از این روش می‌تواند برای شناسایی و دسته‌بندی برای شناخت منطقه مد نظرمان کاربردی باشد اما نیازمند اعمال مراحل برای حذف نویزهای موجود در تصویر می‌باشد.

#### ۵. مراجع

1. Samuel, Arthur L. "Machine learning." *The Technology Review* 62, no. 1 (1959): 42-45
2. Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018, October). Unsupervised learning based on artificial neural network: A review. In *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)* (pp. 322-327). IEEE.
3. Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, 4, 51-62.
4. Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664-681.
5. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
6. Madhulatha, T. S. (2012). An overview on clustering methods. *arXiv preprint arXiv:1205.1117*.



# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

7. Fleiss, J. L., & Zubin, J. (1969). On the methods and theory of clustering. *Multivariate Behavioral Research*, 4(2), 235-250.
8. Hartigan, J. A. (1985). Statistical theory in clustering. *Journal of classification*, 2, 63-76.
9. Ali I, Rehman AU, Khan DM, Khan Z, Shafiq M, Choi J-G. Model Selection Using K-Means Clustering Algorithm for the Symmetrical Segmentation of Remote Sensing Datasets. *Symmetry*. 2022; 14(6):1149.
10. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), 100-108.
11. Tou, J. T., & Gonzalez, R. C. (1974). Pattern recognition principles.
12. Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019, November). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. In *Journal of Physics: Conference Series* (Vol. 1361, No. 1, p. 012015). IOP Publishing.
13. Thinsungnoena, T., Kaoungkub, N., Durongdumronchaib, P., Kerdprasopb, K., & Kerdprasopb, N. (2015). The clustering validity with silhouette and sum of squared errors. *learning*, 3(7).
14. Kwedlo, W. (2011). A clustering method combining differential evolution with the K-means algorithm. *Pattern Recognition Letters*, 32(12), 1613-1621.
15. Gan, G., Ma, C., & Wu, J. (2020). *Data clustering: theory, algorithms, and applications*. Society for Industrial and Applied Mathematics.
16. Acharya, T. D., & Yang, I. (2015). Exploring landsat 8. *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, 4(4), 4-10.
17. Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., ... & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment*, 145, 154-172.
18. Petrelli, M. (2021). *Introduction to Python in Earth Science Data Analysis: From Descriptive Statistics to Machine Learning*. Springer Nature.
19. Huang, T. S., Schreiber, W. F., & Tretiak, O. J. (1971). Image processing. *Proceedings of the IEEE*, 59(11), 1586-1609.
20. Pandey, R., Naik, S., & Marfatia, R. (2013). Image processing and machine learning for automated fruit grading system: A technical review. *International Journal of Computer Applications*, 81(16), 29-39.
21. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(03), 90-95.
22. Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. Nature 585, 357–362 (2020).
23. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

# دومین کنفرانس بین‌المللی دانشجویان جغرافیا و علوم زمین ایران

## 2nd International Conference for Iranian Geography and Earth Sciences' Students

24. Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... & Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, 2, e453.
25. Javed, A., Cheng, Q., Peng, H., Altan, O., Li, Y., Ara, I., ... & Saleem, N. (2021). Review of Spectral Indices for Urban Remote Sensing. *Photogrammetric Engineering & Remote Sensing*, 87(7), 513-524.
26. Mróz, M.; Sobieraj, A. Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series and their suitability for land cover and agricultural crop identification. *Tech. Sci.* 2004, 7, 39–66.
27. Karnieli, A.; Qin, Z.; Wu, B.; Panov, N.; Yan, F. Spatio-Temporal Dynamics of Land-Use and Land-Cover in the Mu Us Sandy Land, China, Using the Change Vector Analysis Technique. *Remote Sens.* 2014, 6, 9316–9339.
28. Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14), 3025-3033.
29. Gidden, M. J., & Huppmann, D. (2019). pyam: a Python Package for the Analysis and Visualization of Models of the Interaction of Climate, Human, and Environmental Systems. *Journal of Open Source Software*, 4(33), 1095.