



12 August 2024

Klasteryzacja zdjęć galaktyk

Michał Piechota, Julia Kruk



**Wydział Matematyki
i Nauk Informacyjnych**
POLITECHNIKA WARSZAWSKA

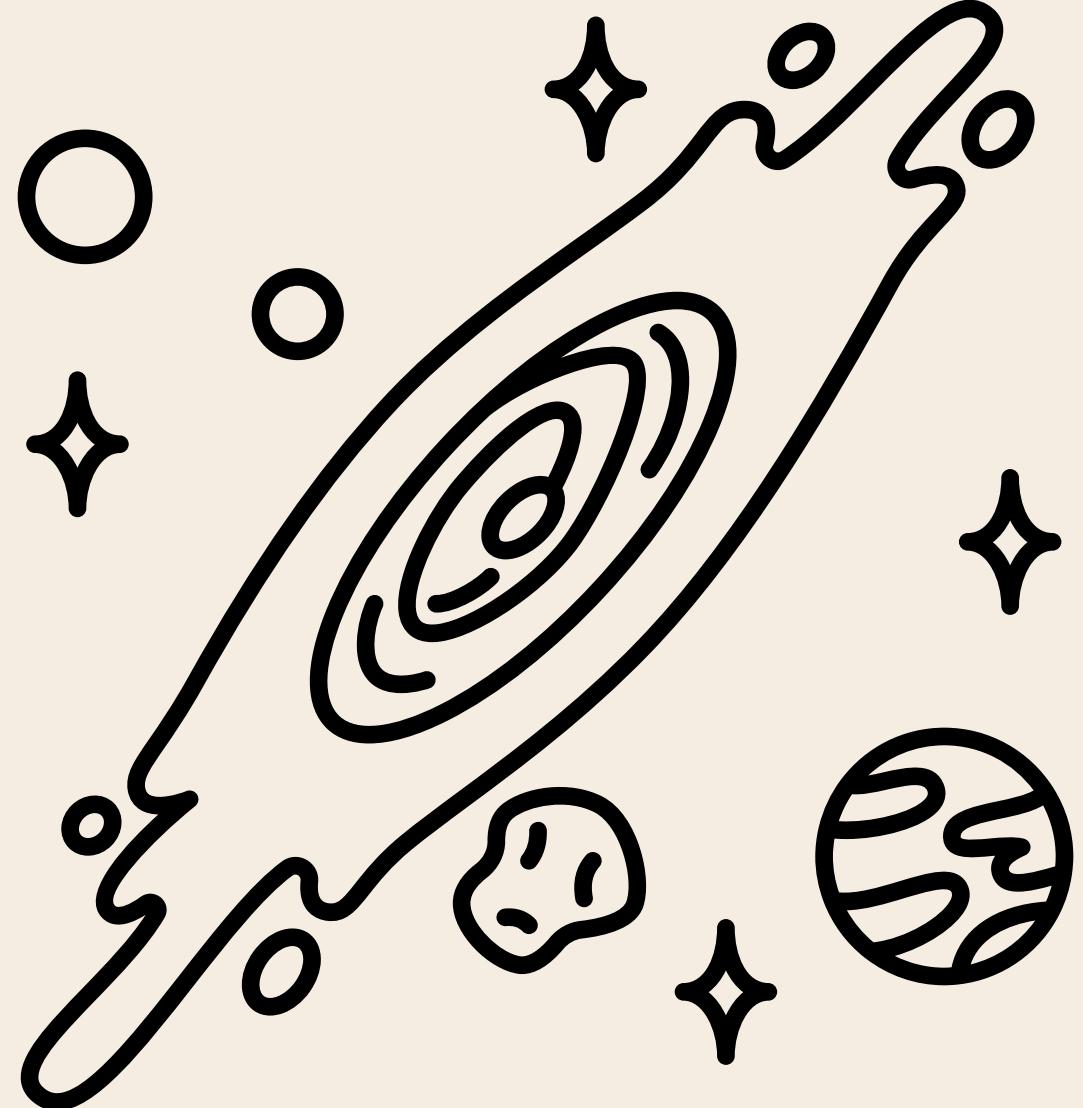
Agenda

1. Cel biznesowy
2. Eksploracyjna analiza danych
3. Inżynieria cech
4. Wybór modelu
5. Powstałe klastry



02





Cel biznesowy

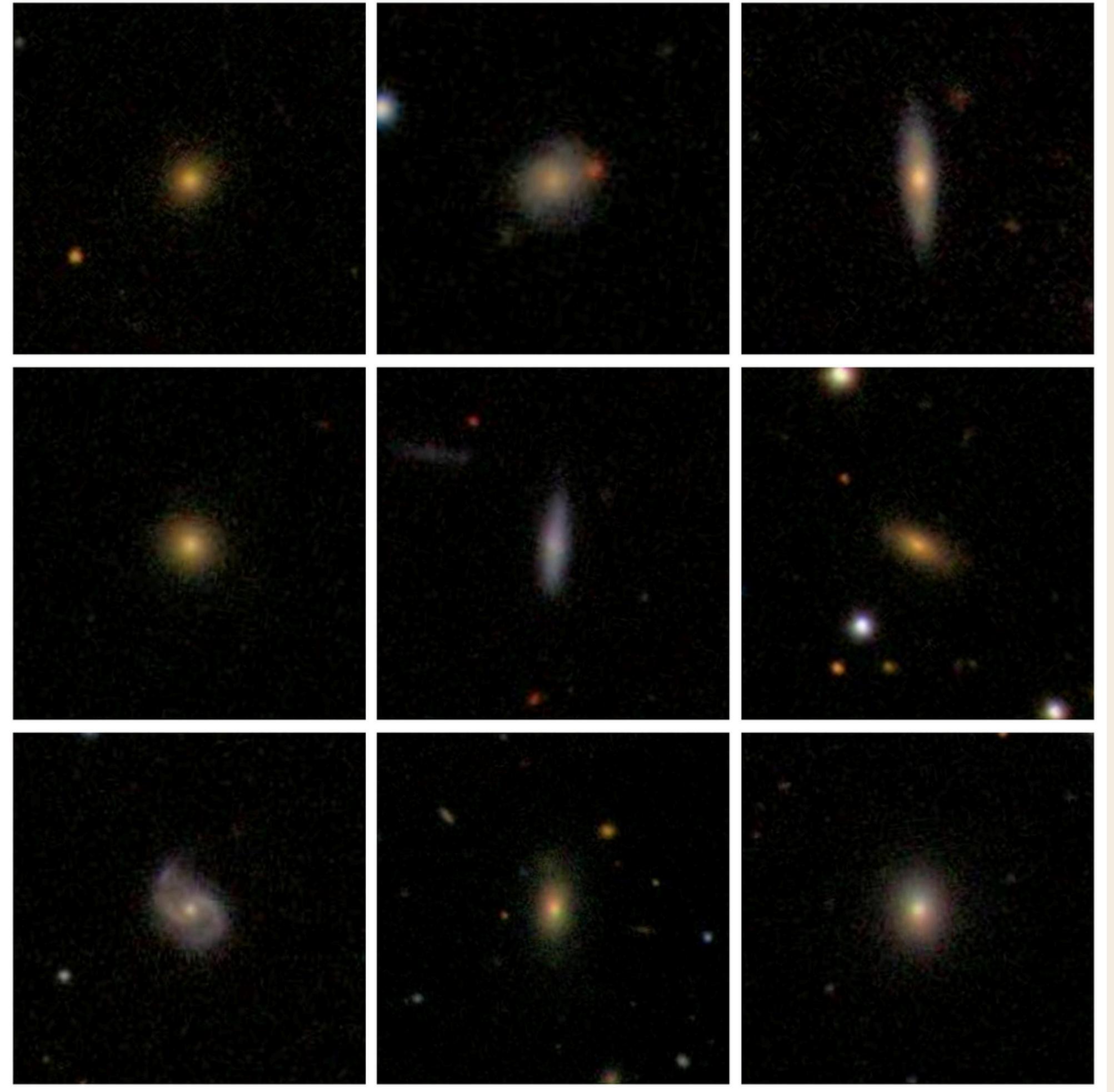
- odnalezienie cech wspólnych dla różnych galaktyk,
- usprawnienie procesu klasyfikacji

Dane

- Galaxy Zoo 2 | Kaggle
- 243 tys. zdjęć
- wszystkie wymiarów 424 x 424
- próbka 100 tys.



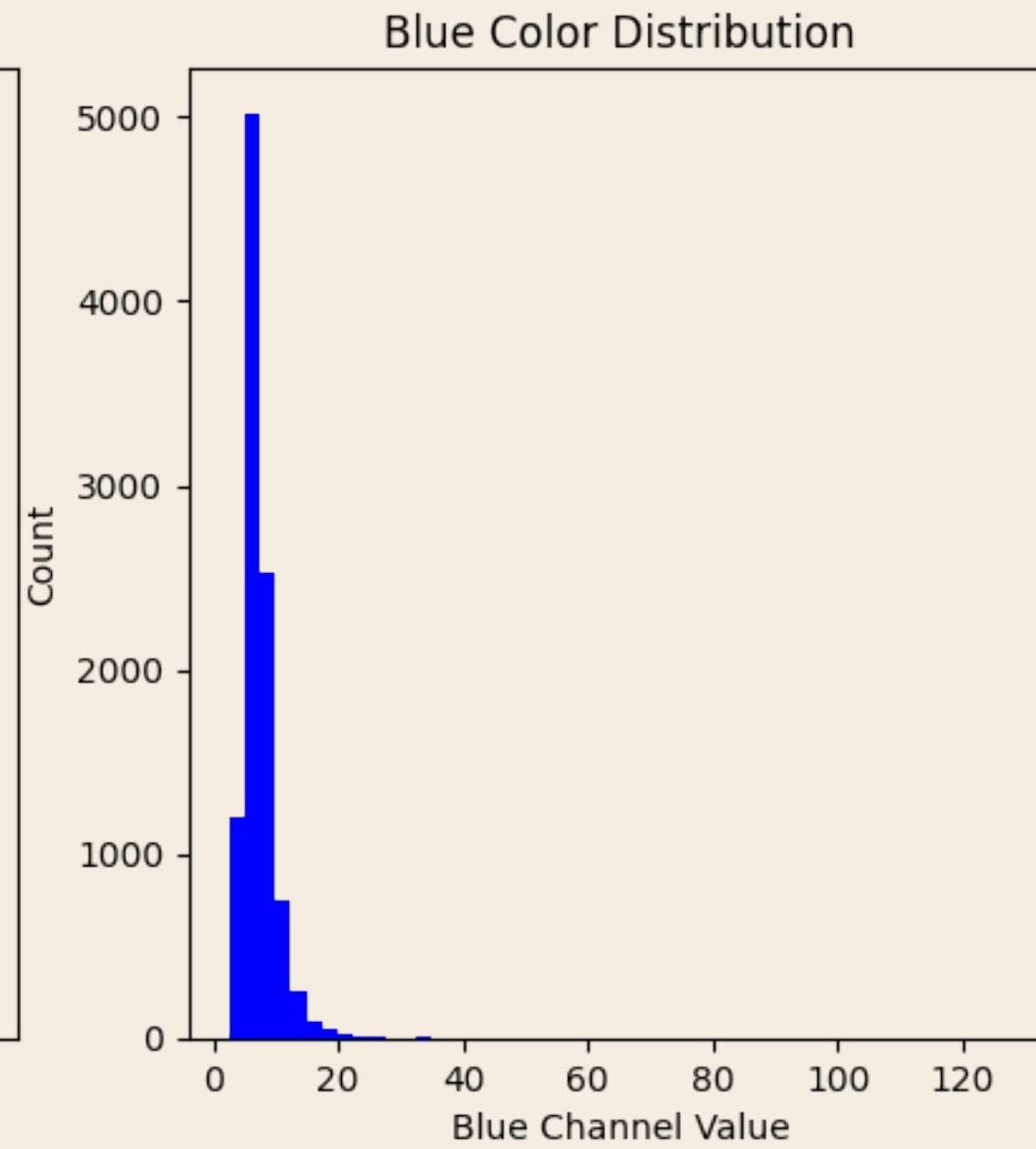
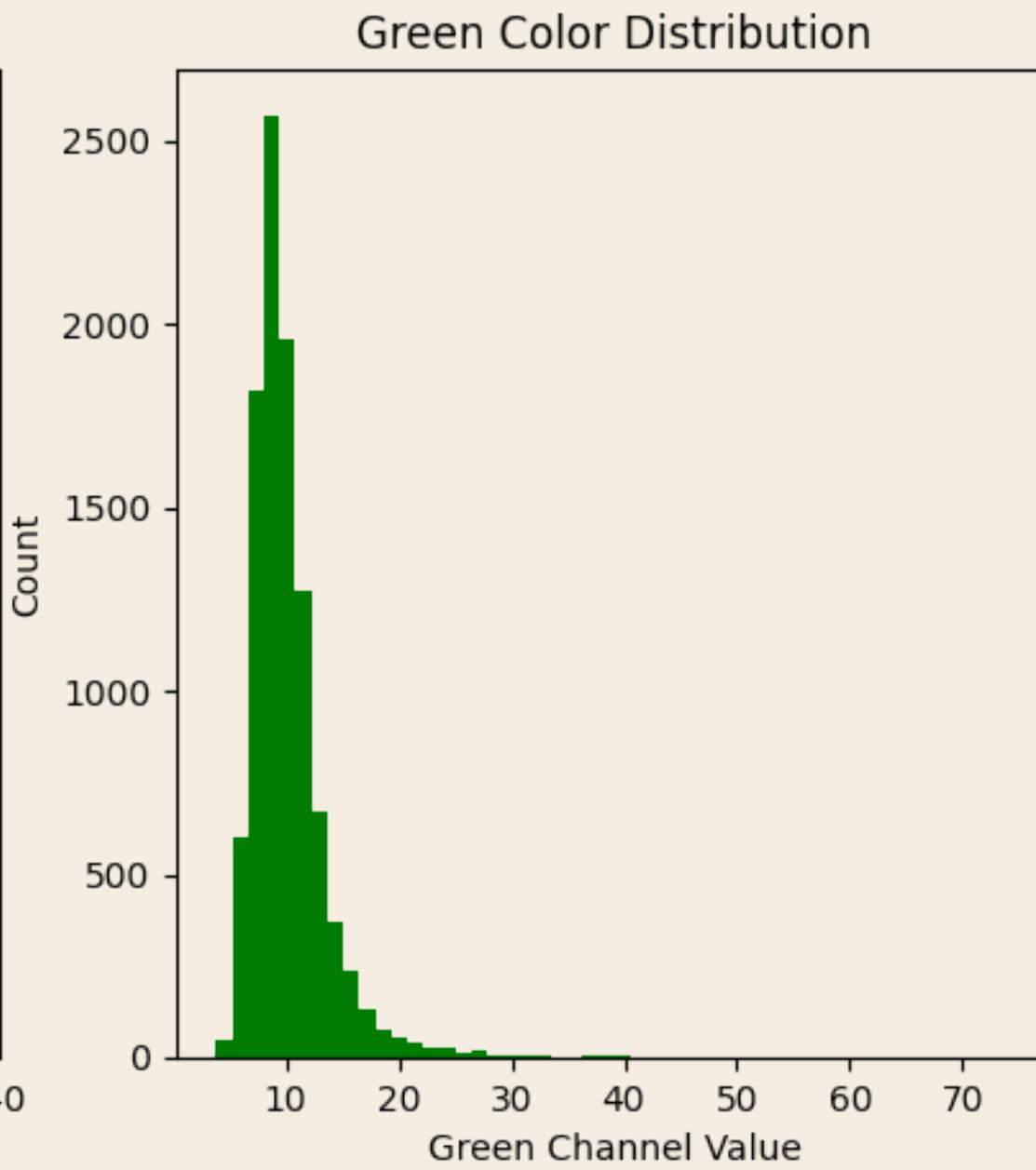
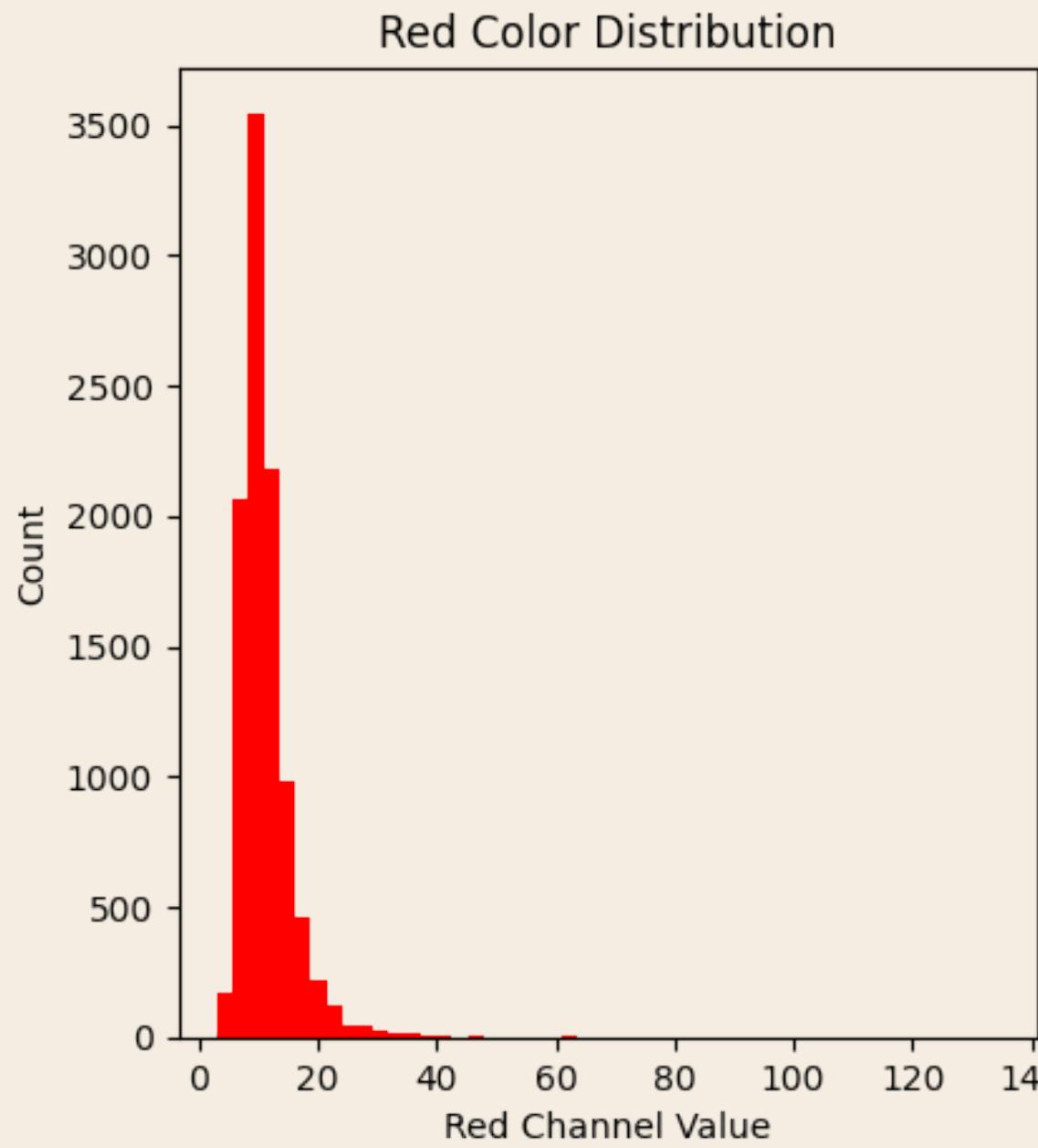
04



→

05

Analiza kolorów





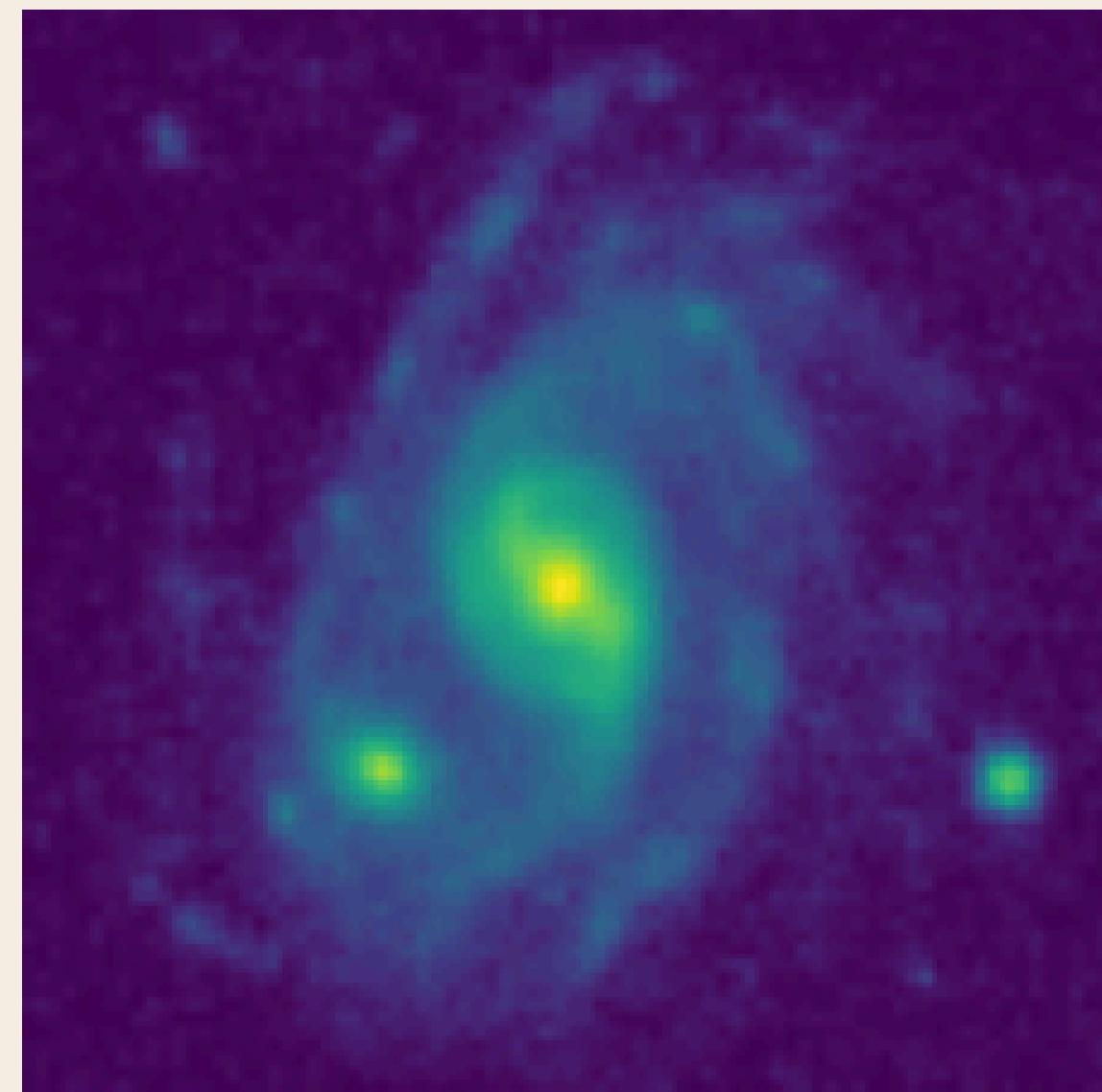
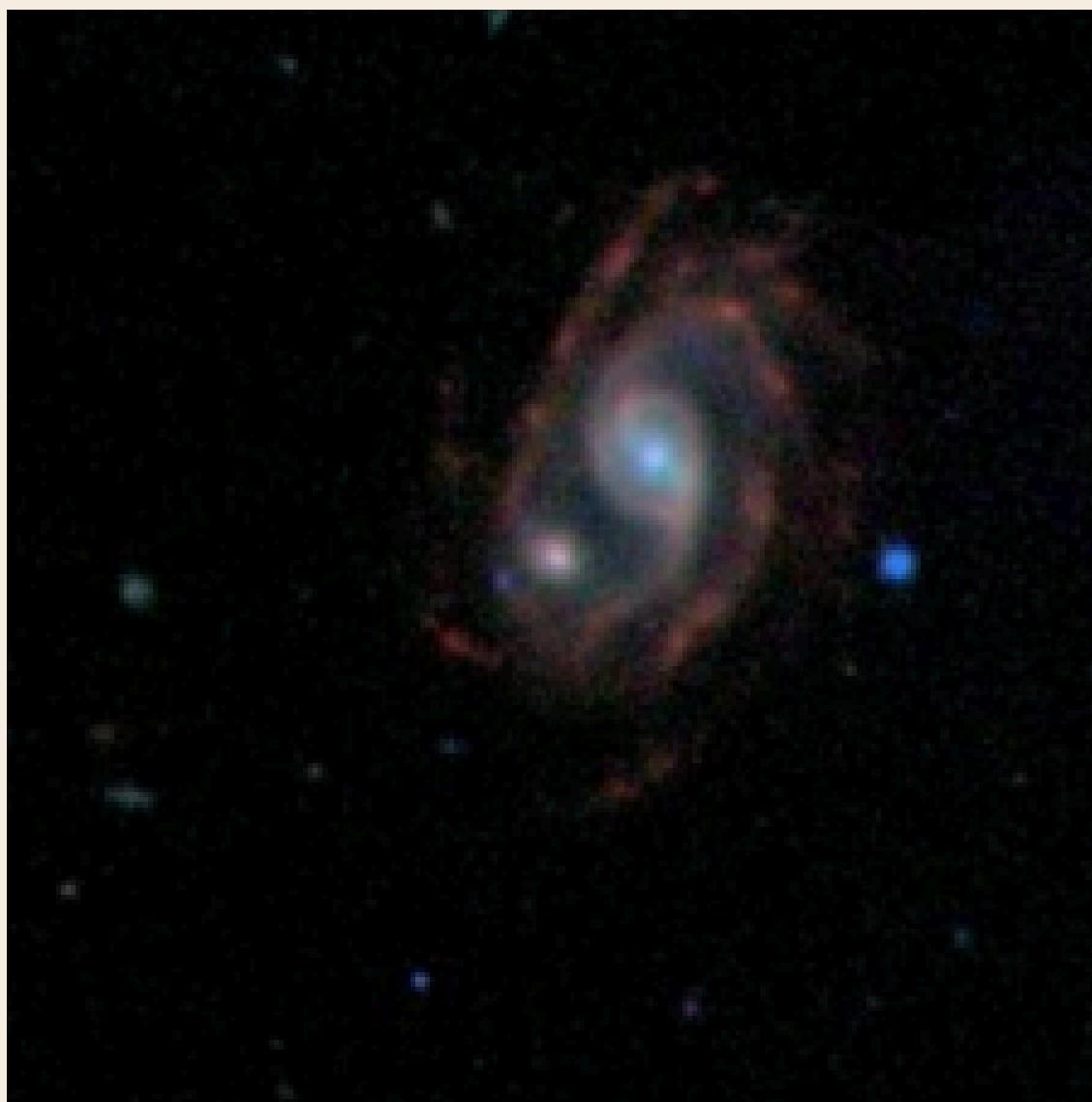
Inżynieria cech

- przycinanie obrazów
- kodowanie obrazów
- zamiana na wektory cech (feature vectors)
- skalowanie
- analiza głównych składowych (PCA)

Przycinanie obrazów



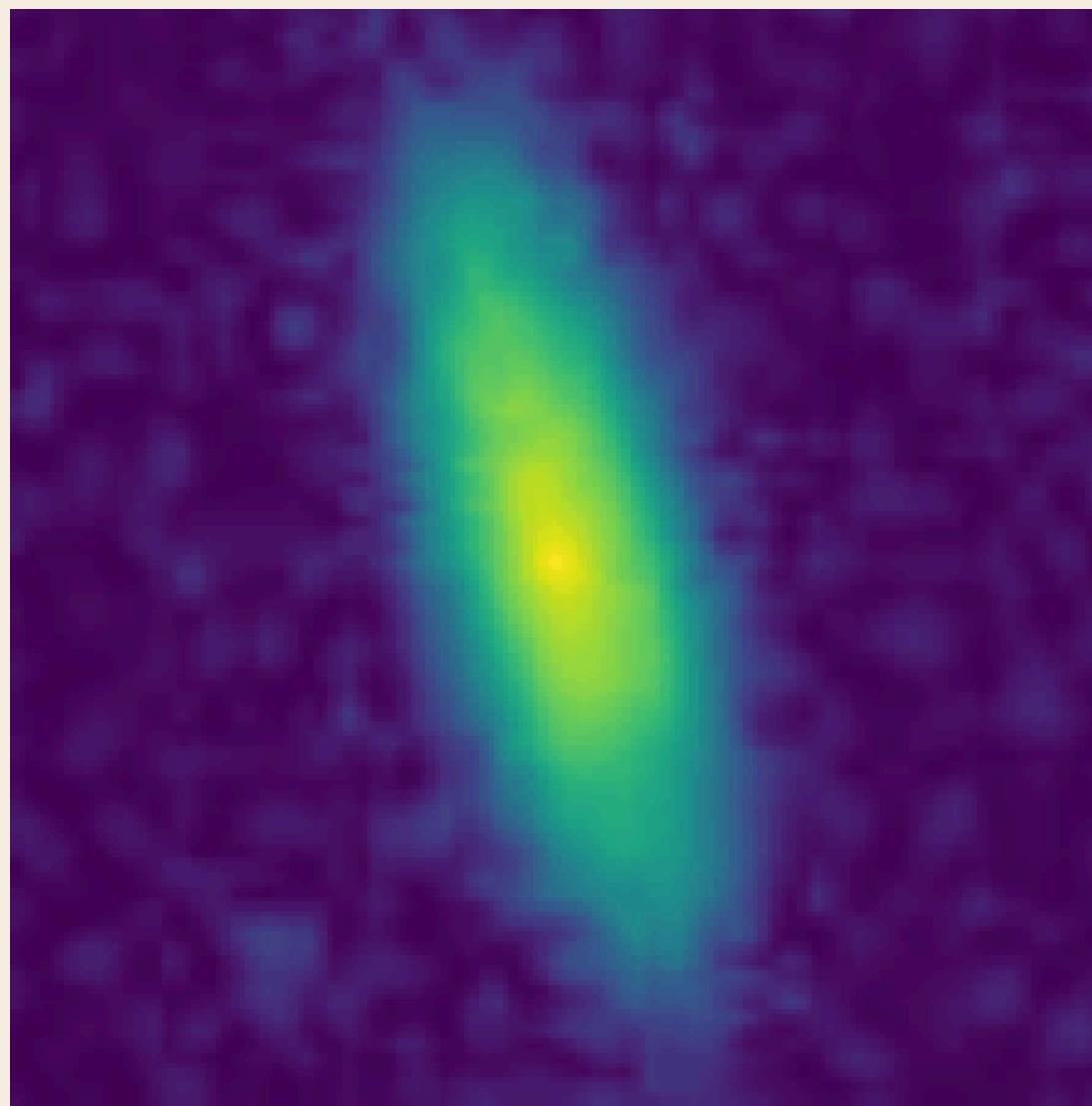
Przetwarzanie obrazów



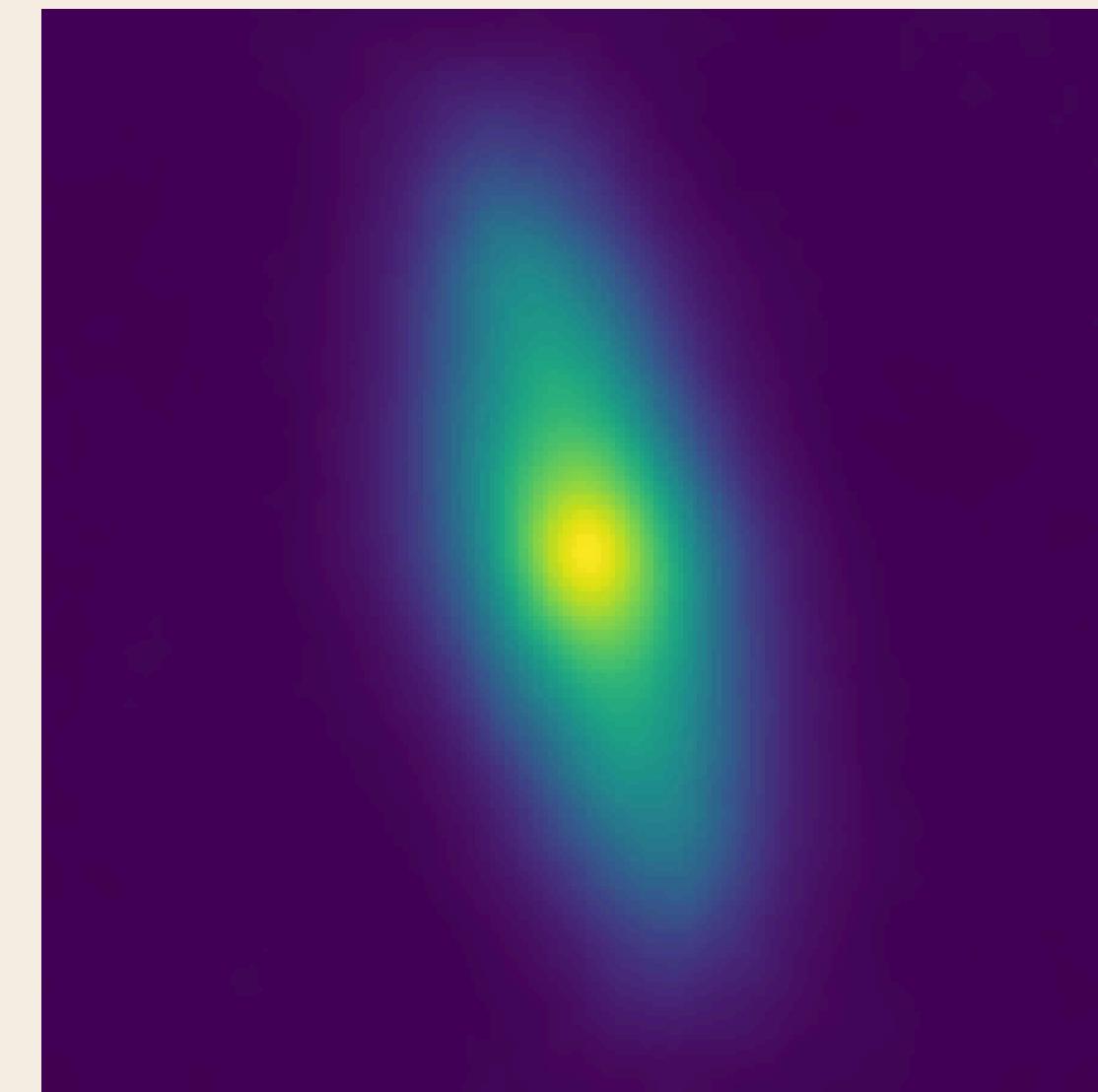
Usuwanie szumów



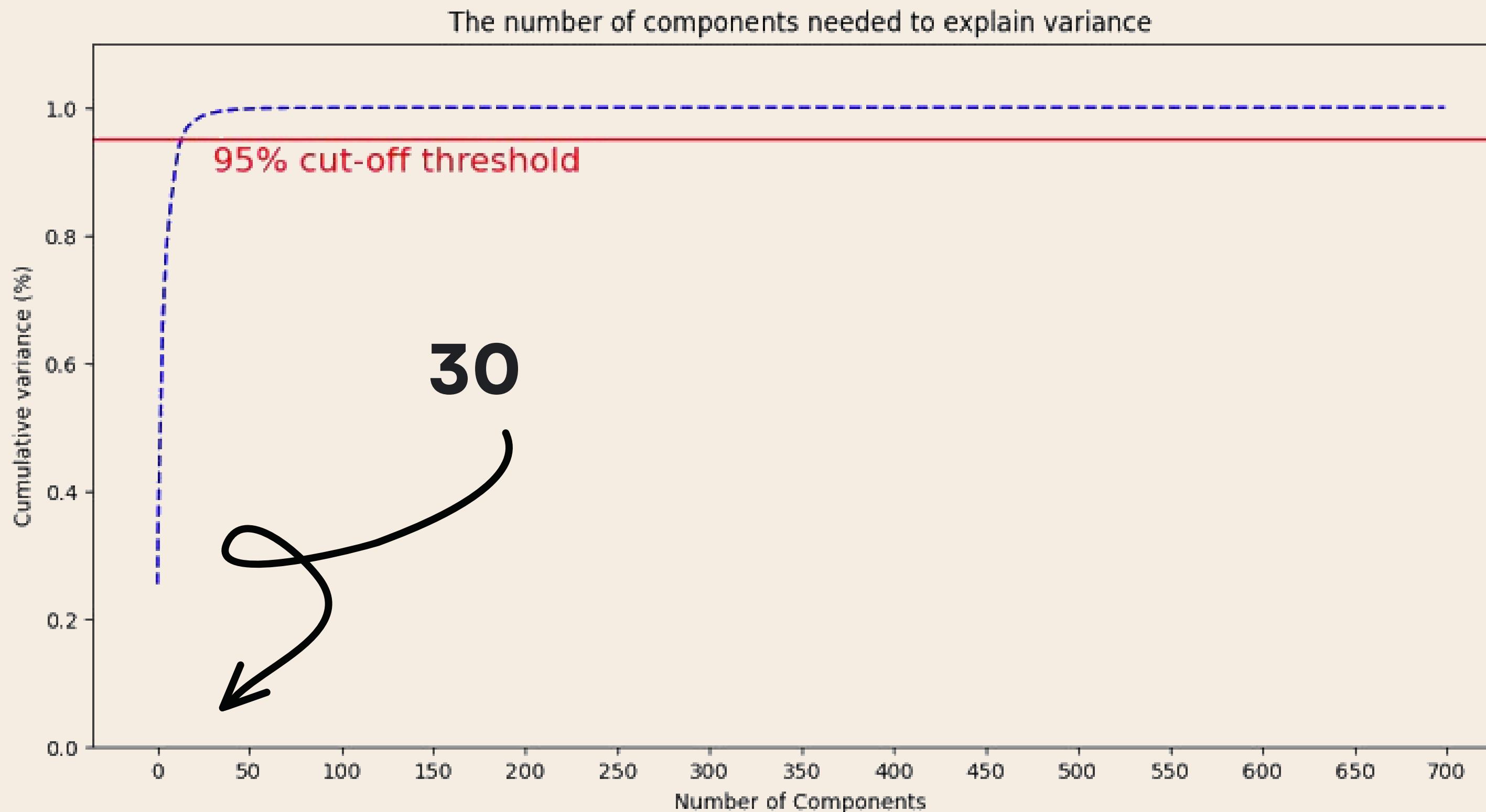
Before



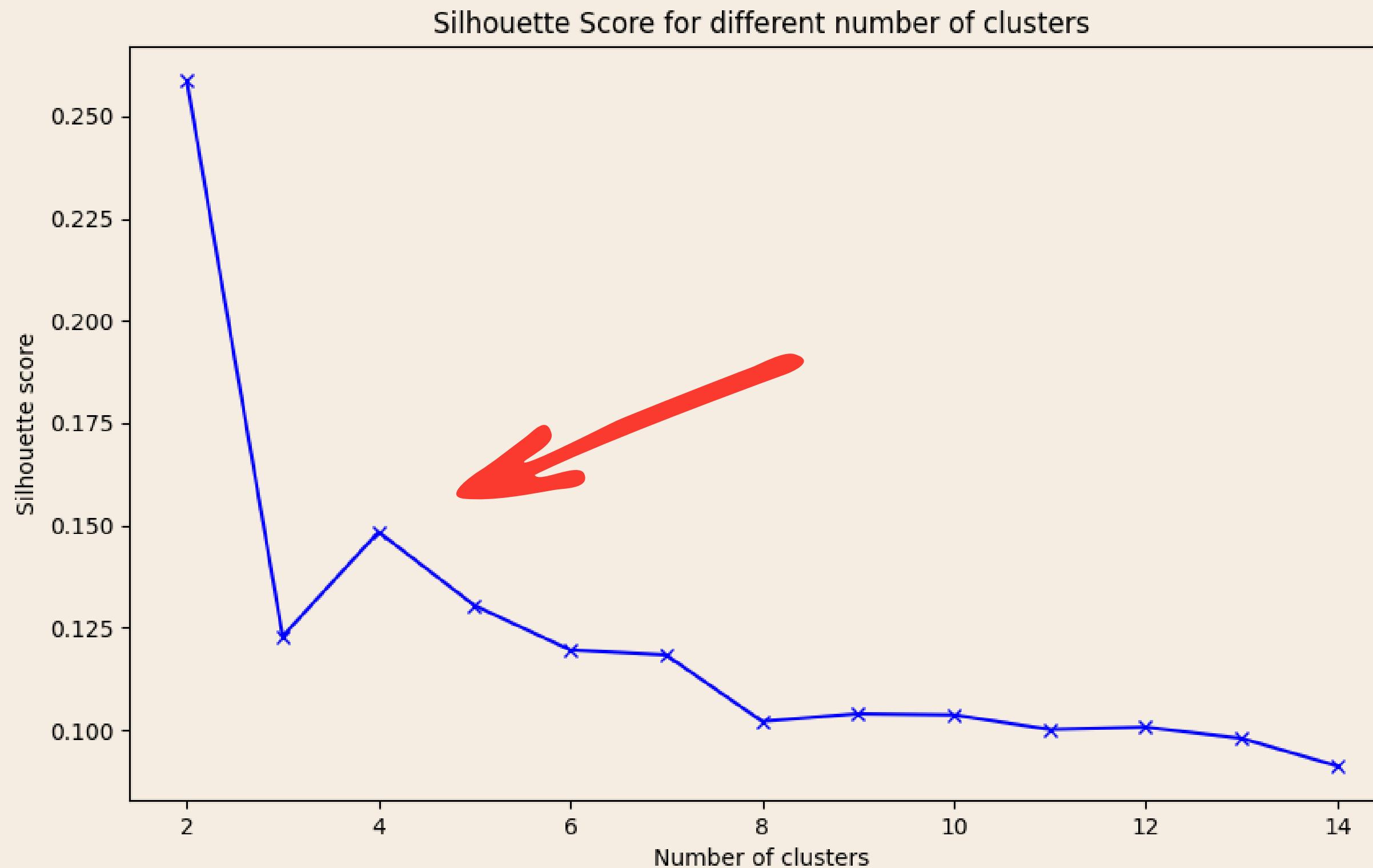
After Auto Encoder



Redukcja wymiarów



KMeans



wybór liczby klastrów

k=4

12

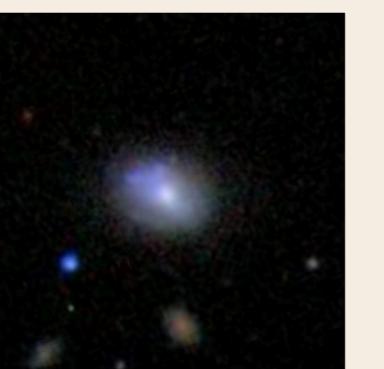
Cluster 0



Cluster 0



Cluster 0



Cluster 0



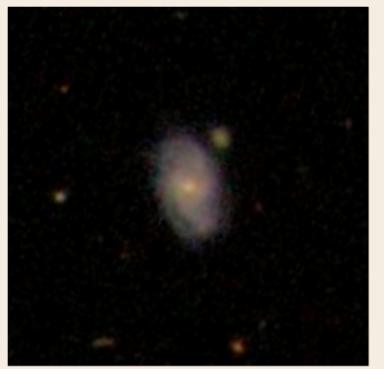
Cluster 0



Cluster 1



Cluster 1



Cluster 1



Cluster 1



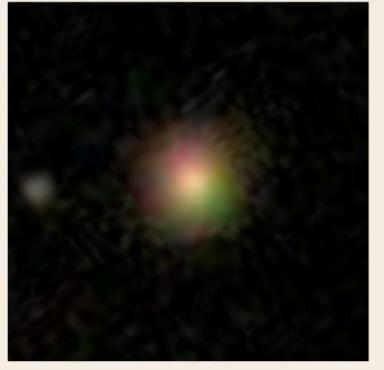
Cluster 1



Cluster 2



Cluster 2



Cluster 2



Cluster 2



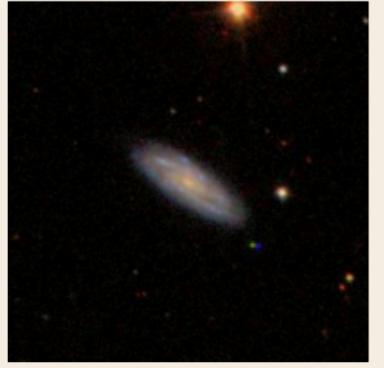
Cluster 2



Cluster 3



Cluster 3



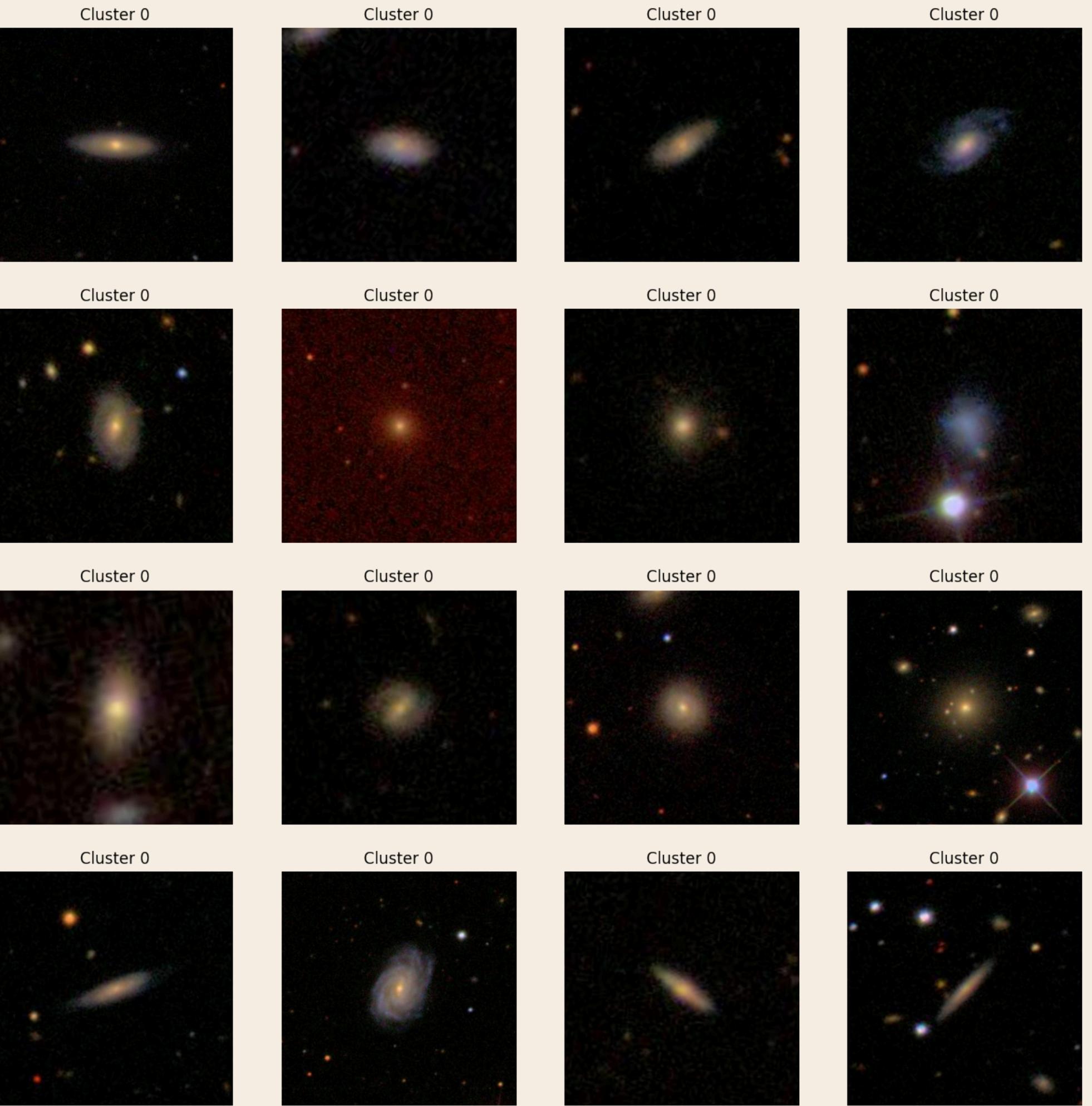
Cluster 3



Cluster 3

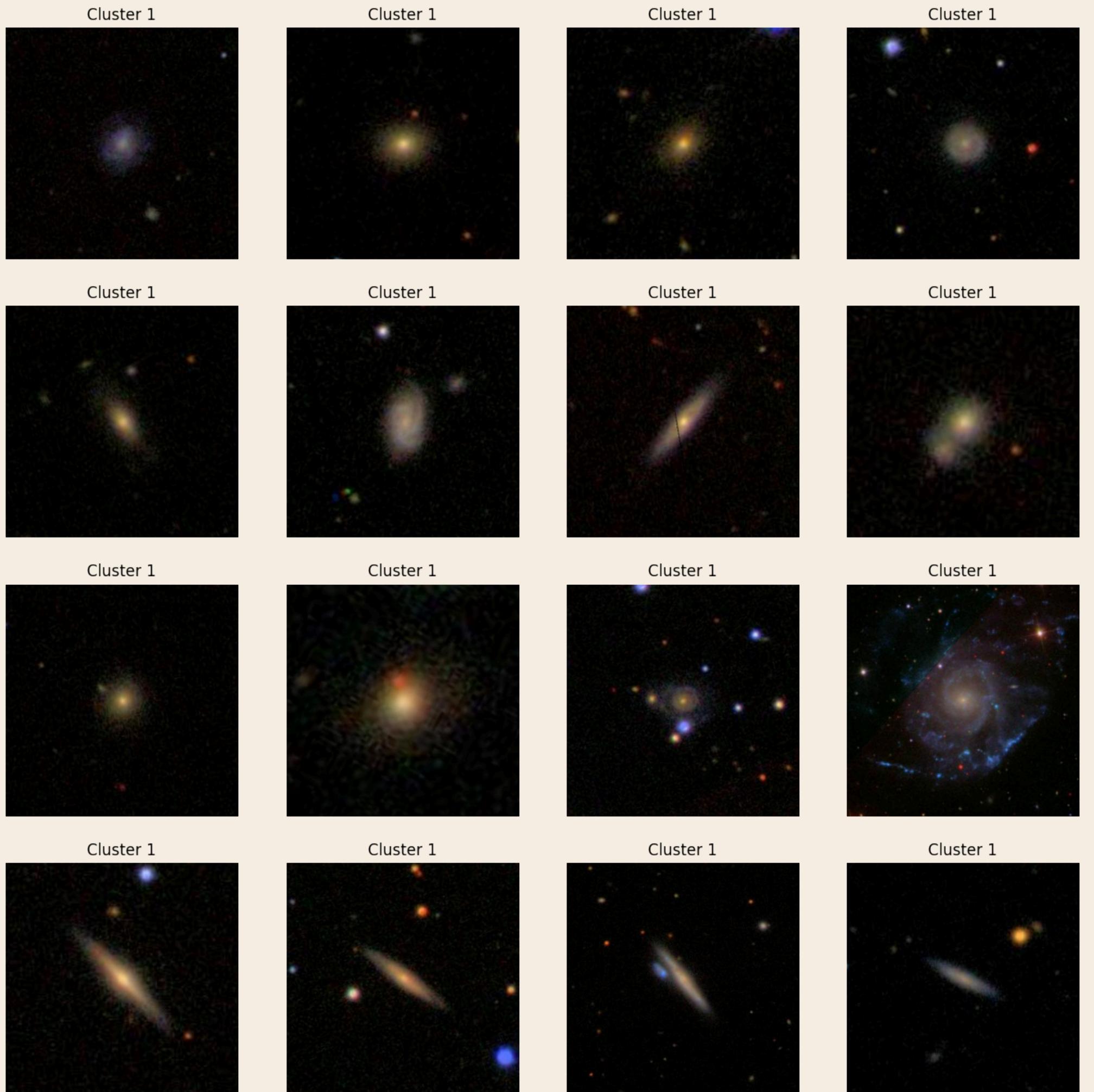


Klastry

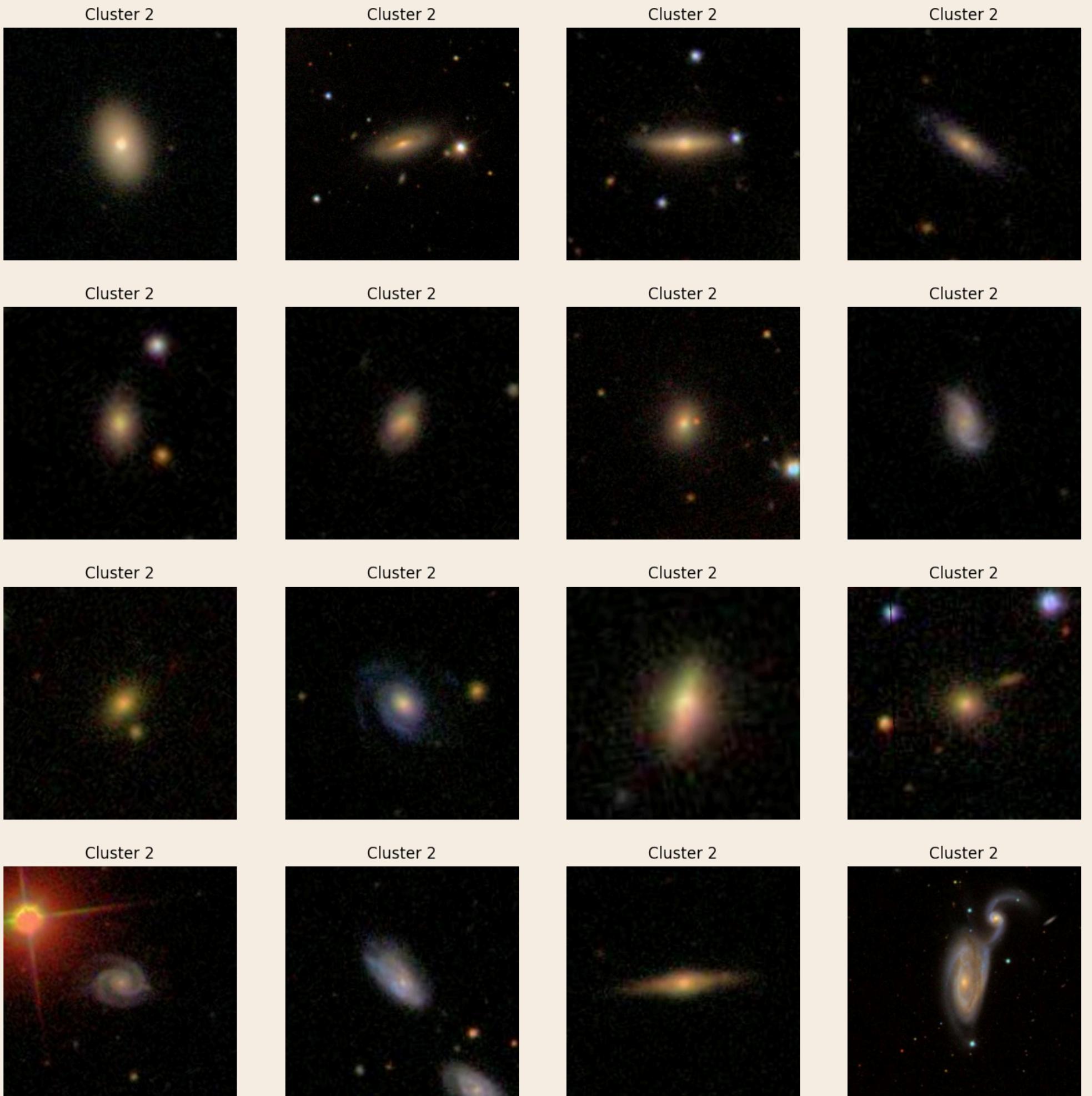


Klaster 0

- eliptyczne,
- mniej wyraźne
spiralne

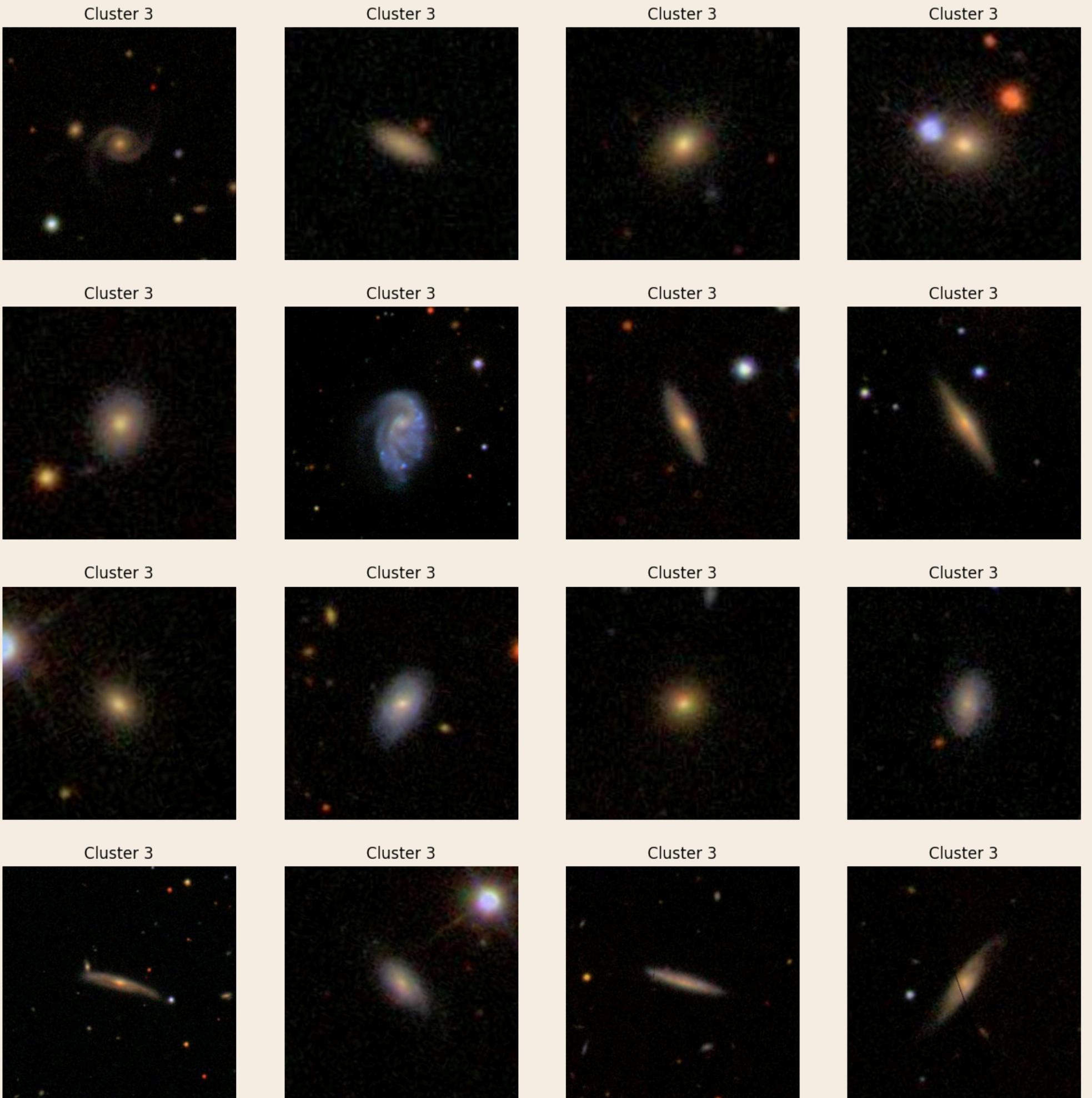


Klaster 1



Klaster 2

- podobny do pierwszego



Klaster 3

- wąskie galaktyki

Indeks Daviesa - Bouldina

1.801



Klasteryzacja zdjęć galaktyk

12 August 2024

Michał Piechota, Julia Kruk



**Wydział Matematyki
i Nauk Informacyjnych**

POLITECHNIKA WARSZAWSKA