



**Wydział Matematyki
i Nauk Informatycznych**

POLITECHNIKA WARSZAWSKA

4 czerwca 2024 roku

Activity Detection

Maciej Momot | Filip Langiewicz





Cel biznesowy

- pracujemy dla firmy produkującej urządzenia przeznaczone dla sportowców (np. zegarki)
- chcemy klasteryzować dane o aktywnościach użytkowników, aby wiedzieć, jakie aktywności preferują i na jakie warto zwracać uwagę w przyszłości



Podjęte działania



Analizowanie początkowych danych

Zapoznanie z czujnikami pomiarowymi

Redukcja wymiarów (x,y,z)

Usuwanie silnie skorelowanych zmiennych

Zastępowanie odstających wartości kwantylem 0,95

Wyliczanie całkowitego dystansu w aktywnościach

Wyznaczanie szczególnych cech dla zmiennych (mean, sd, max, min, median)

Znalezienie optymalnej liczby klastrów metodą łokcia



Modele:

K-means

DBSCAN

K-medoids

**Agglomerative
Hierarchical**

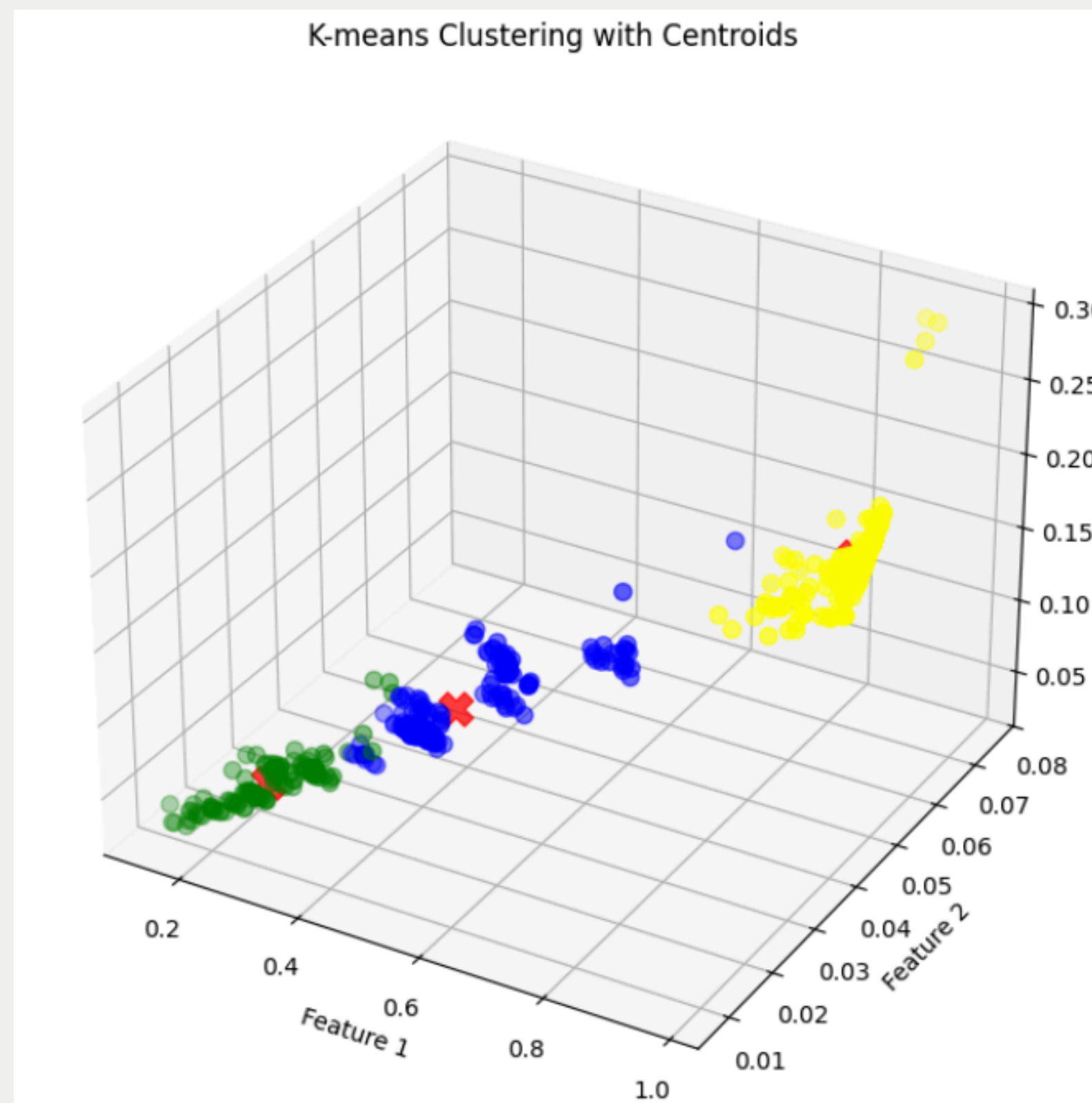


Wybrany model:

K-means
(Normalization)



Klastry - train:

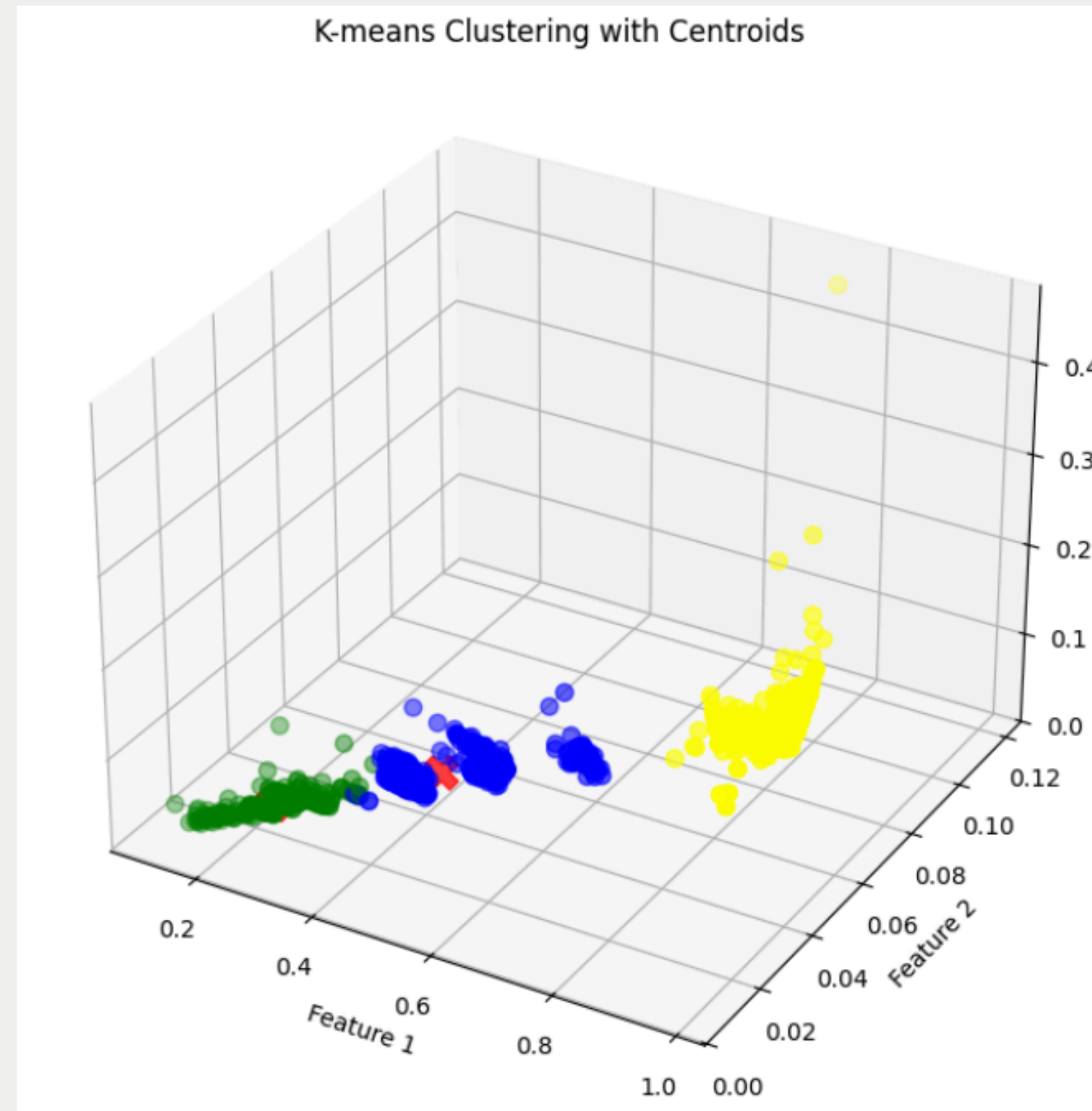


Wyniki - train

Model inertia:	25.360528130719
Silhouette coefficient:	0.638789708874
Davies Bouldin Score:	0.5601056774615
Calinski-Harabasz Score:	10024.824574626



Klastry - valid:



Wyniki - valid

Model inertia:	15.917658975201
Silhouette coefficient:	0.618092146733
Davies Bouldin Score:	0.5573002284270
Calinski-Harabasz Score:	5659.3424254992



Walidatorzy



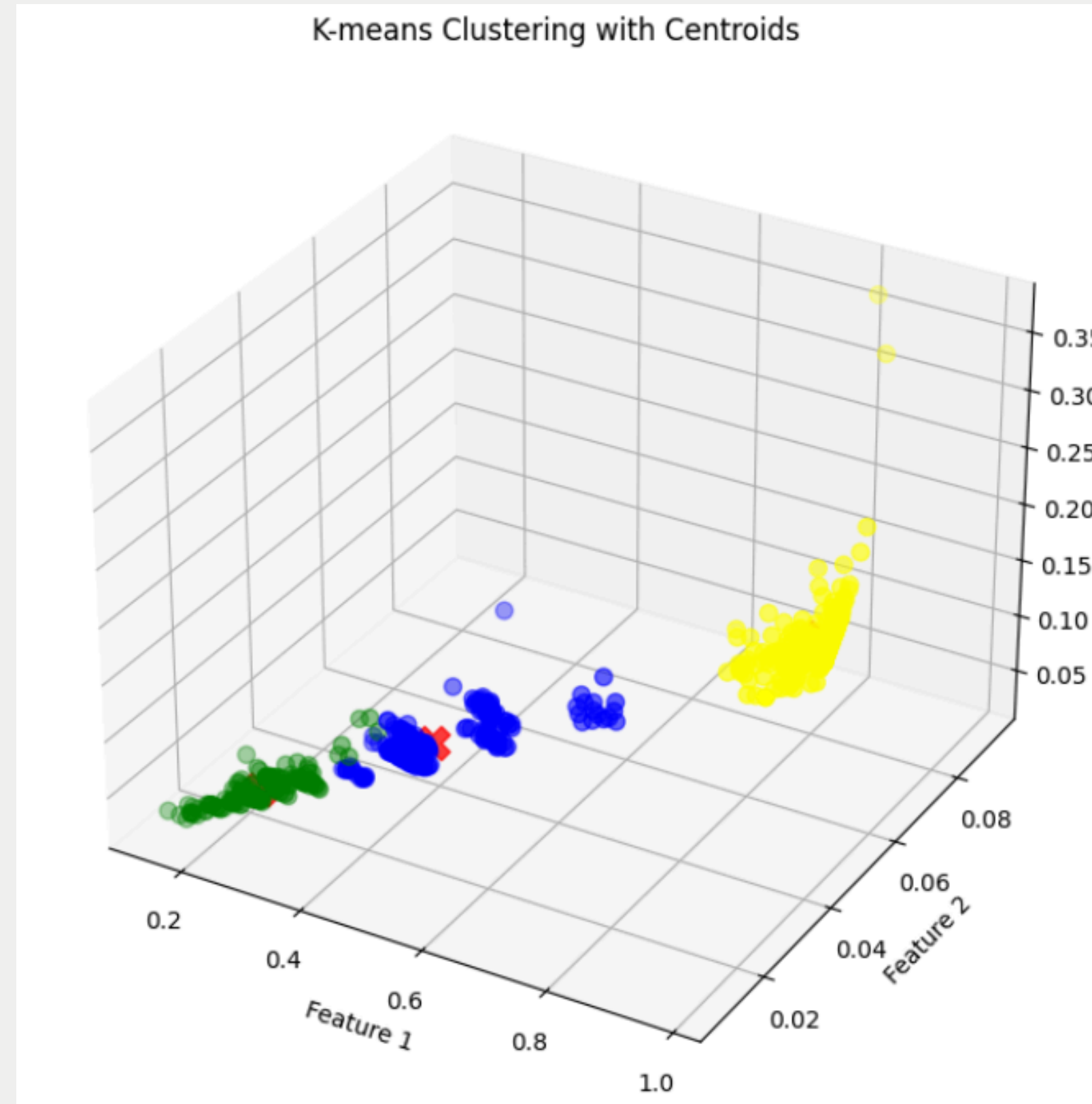
Bogumiła Okrojek



Aleksandra Kwiatkowska



Klastry - test



Wyniki - test

Model inertia:	13.927511329252
Silhouette coefficient:	0.657719583620
Davies Bouldin Score:	0.5181559533625
Calinski-Harabasz Score:	6930.4555402001



Interpretacja klastrów



chodzenie



siedzenie



jazda na rowerze



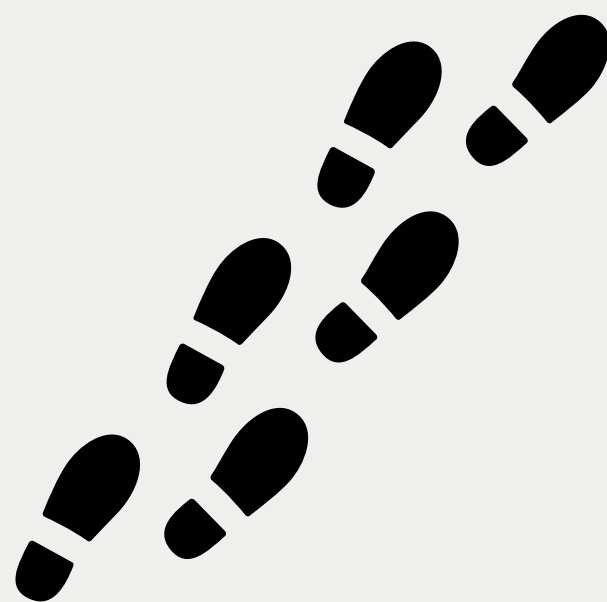
Zmienne decydujące



średnia prędkość



średnie przyspieszenie



liczba kroków na minutę



Dziękujemy!

