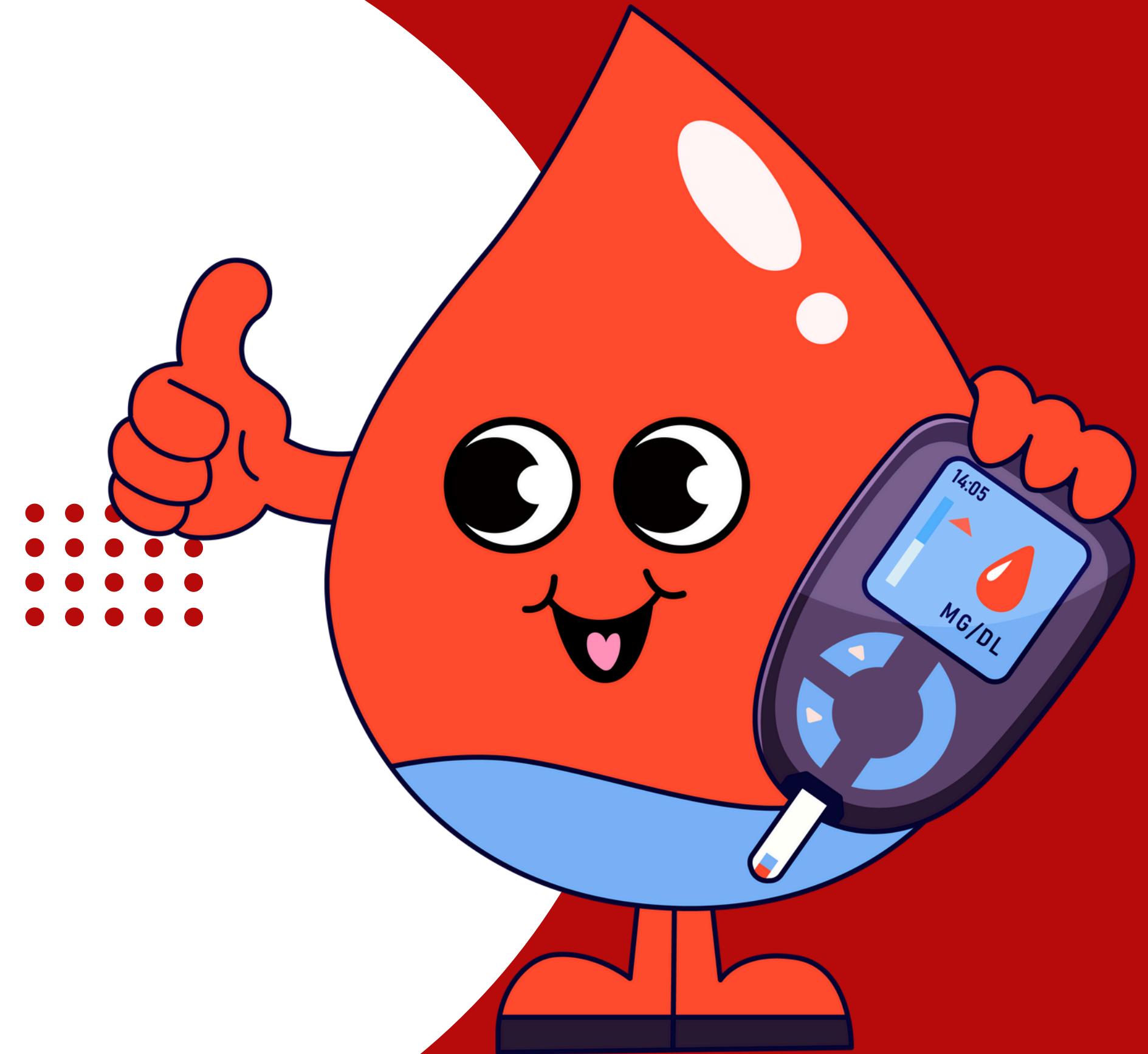




UFRPE

ANÁLISE DE AGRUPAMENTOS EM PACIENTES COM DIABETES

Março 2025
inteligência artificial



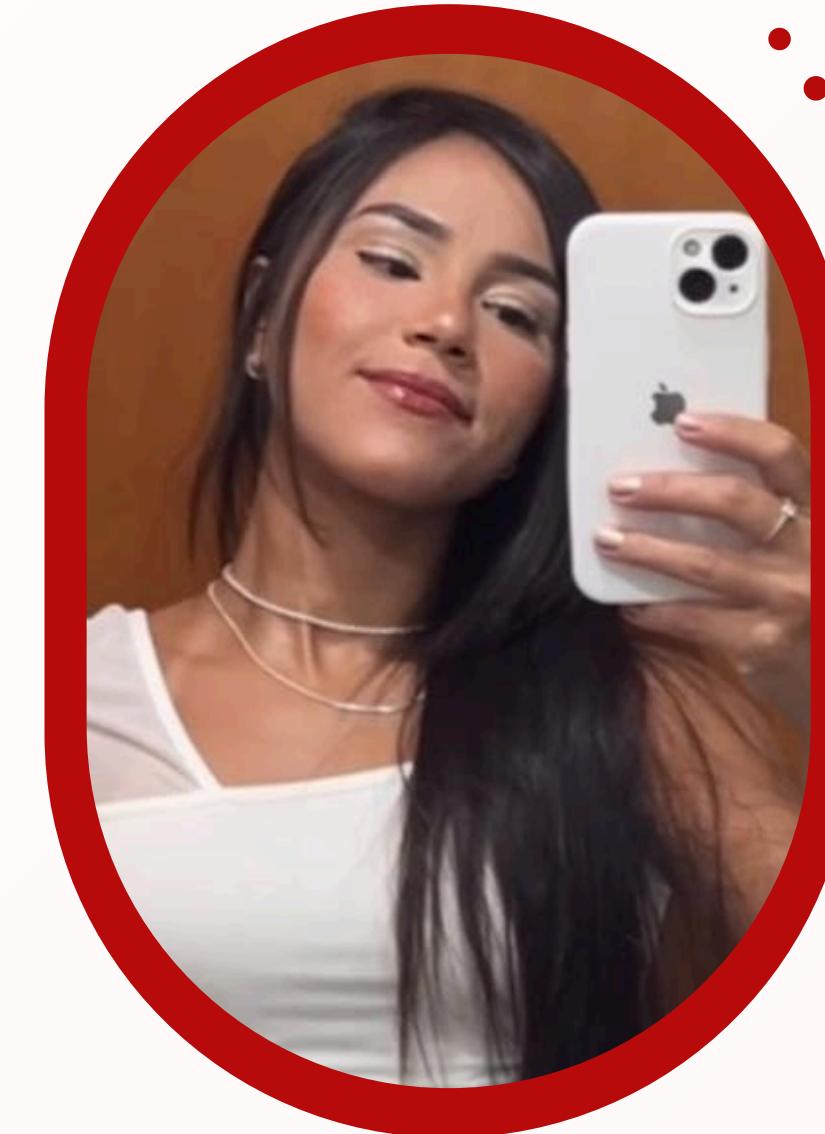
Integrantes



Allan Rêgo



Izabella Pereira



Marina Souza

Sobre o projeto

Objetivos do Projeto

- Aplicar algoritmos de Agrupamento para classificar pacientes com diabetes.
- Analisar diferentes perfis de pacientes com base em indicadores de saúde.
- Comparar resultados de algoritmos através de métricas específicas.

Base de Dados

- Fonte: diabetes_012_health_indicators_BRFSS2015.csv
- Atributos Selecionados: BMI, HbA1c, Age, DiabetesDuration
- Total de Registros: [243.833]
- Tratamento de Dados: Remoção de outliers (IQR), normalização (MinMaxScaler), redução de dimensionalidade (PCA)

Diabetes Health Indicators Dataset

# HighChol	# CholCheck	# BMI
0 = no high cholesterol = high cholesterol	0 = no cholesterol check in 5 years 1 = yes cholesterol check in 5 years	Body Mass Index
0	0	12
1	1	1
1.0	1.0	40.0
0.0	0.0	25.0
1.0	1.0	28.0
0.0	1.0	27.0
1.0	1.0	24.0
1.0	1.0	25.0
0.0	1.0	30.0

✓ Algoritmos Utilizados

- K-Means (k=5): Baseado na minimização da distância intra-cluster.
- Hierarchical Clustering: Agrupamento hierárquico aglomerativo.
- DBSCAN: Algoritmo baseado em densidade (eps=1.5, min_samples=5).

✓ Configuração dos Experimentos

- Execuções: 50 rodadas para cada algoritmo.
- Seeds: Aleatórias (0 a 10000) para garantir reproduzibilidade.
- Justificativa: Baseadas nas referências de clustering em diabetes (Zhang et al., 2022; Somolinos-Simón et al., 2024).



Métricas de Avaliação

- Índice de Silhueta: Mede coesão e separação dos clusters.
- Índice de Davies-Bouldin: Avalia similaridade intra e inter-clusters.
- Tempo de Execução: Média de tempo gasto por algoritmo.

Resultados

K-Means:

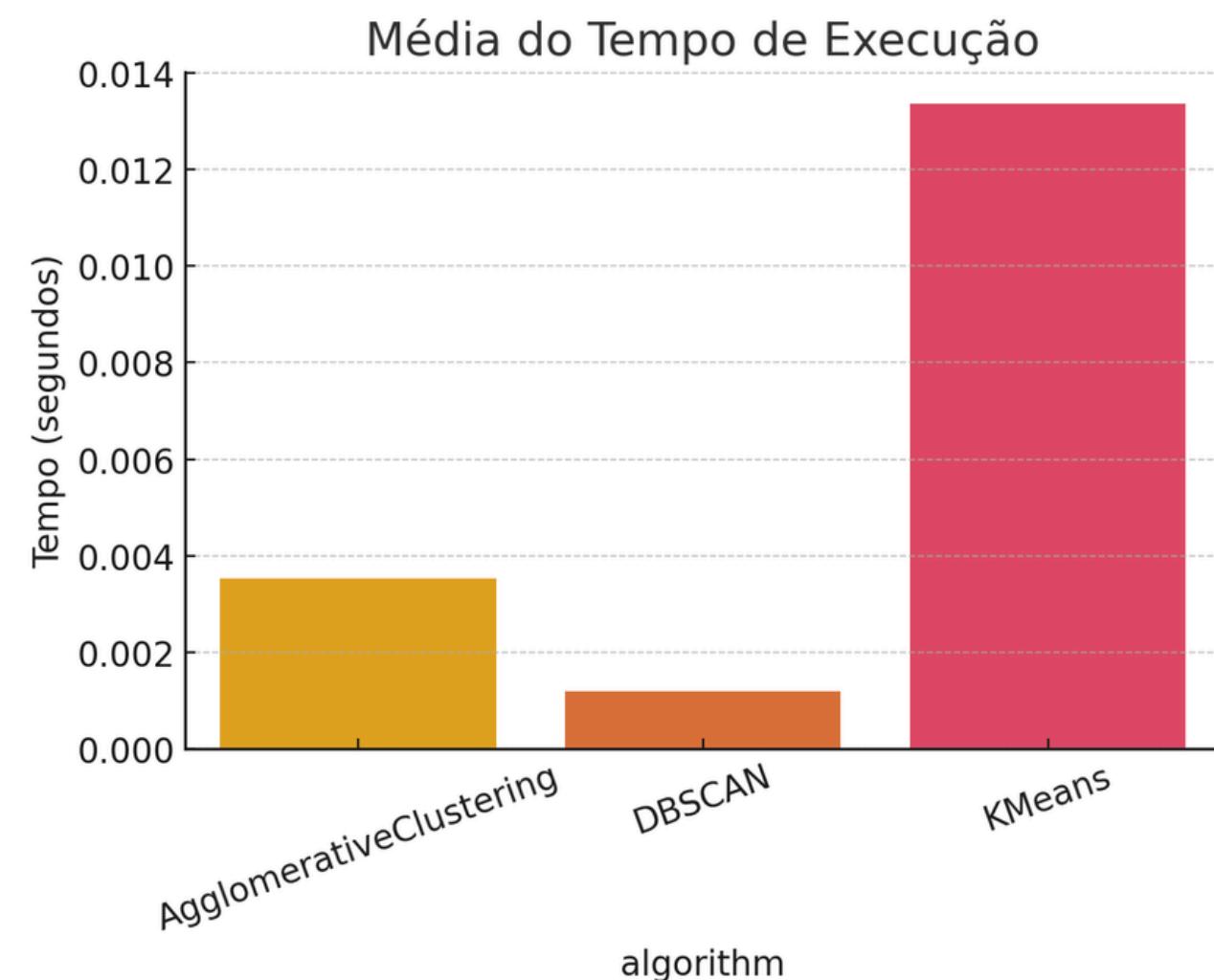
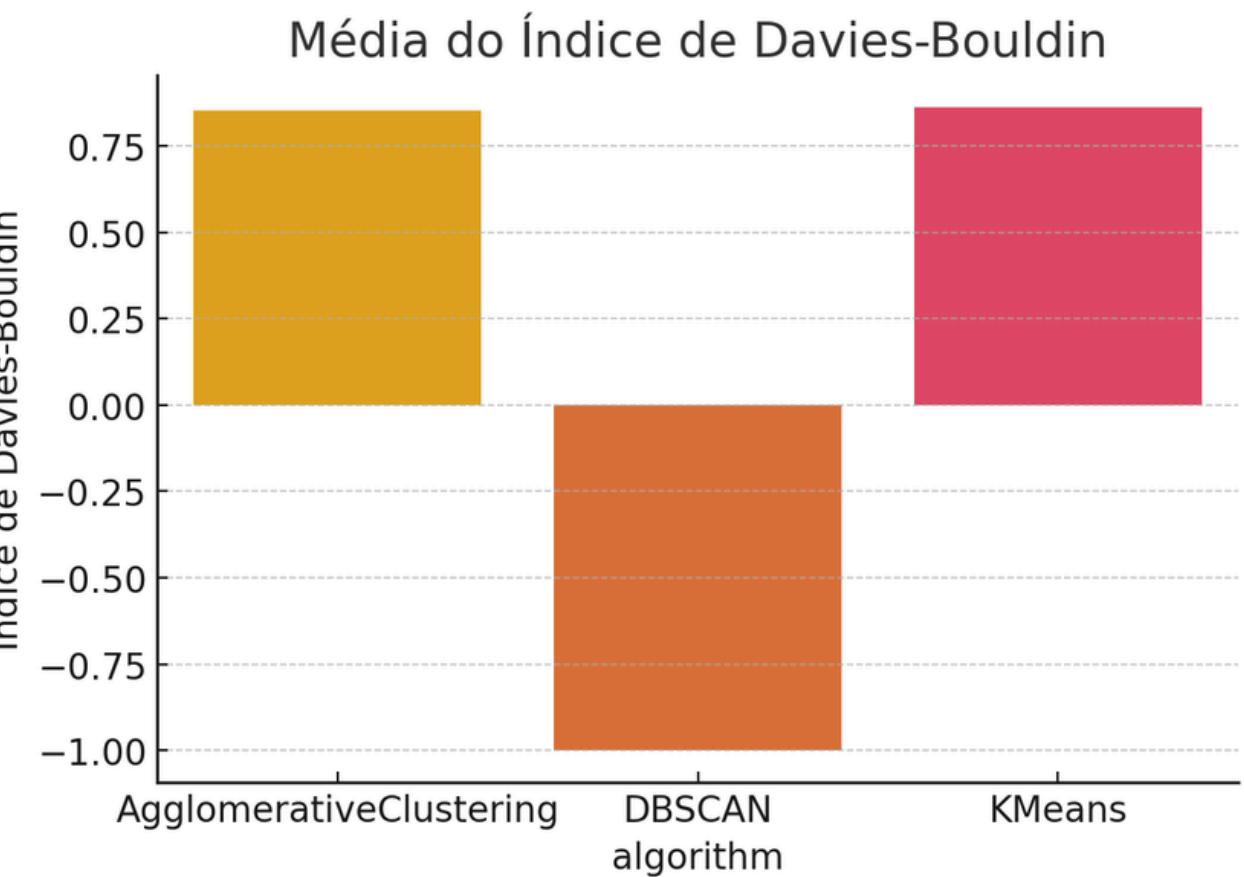
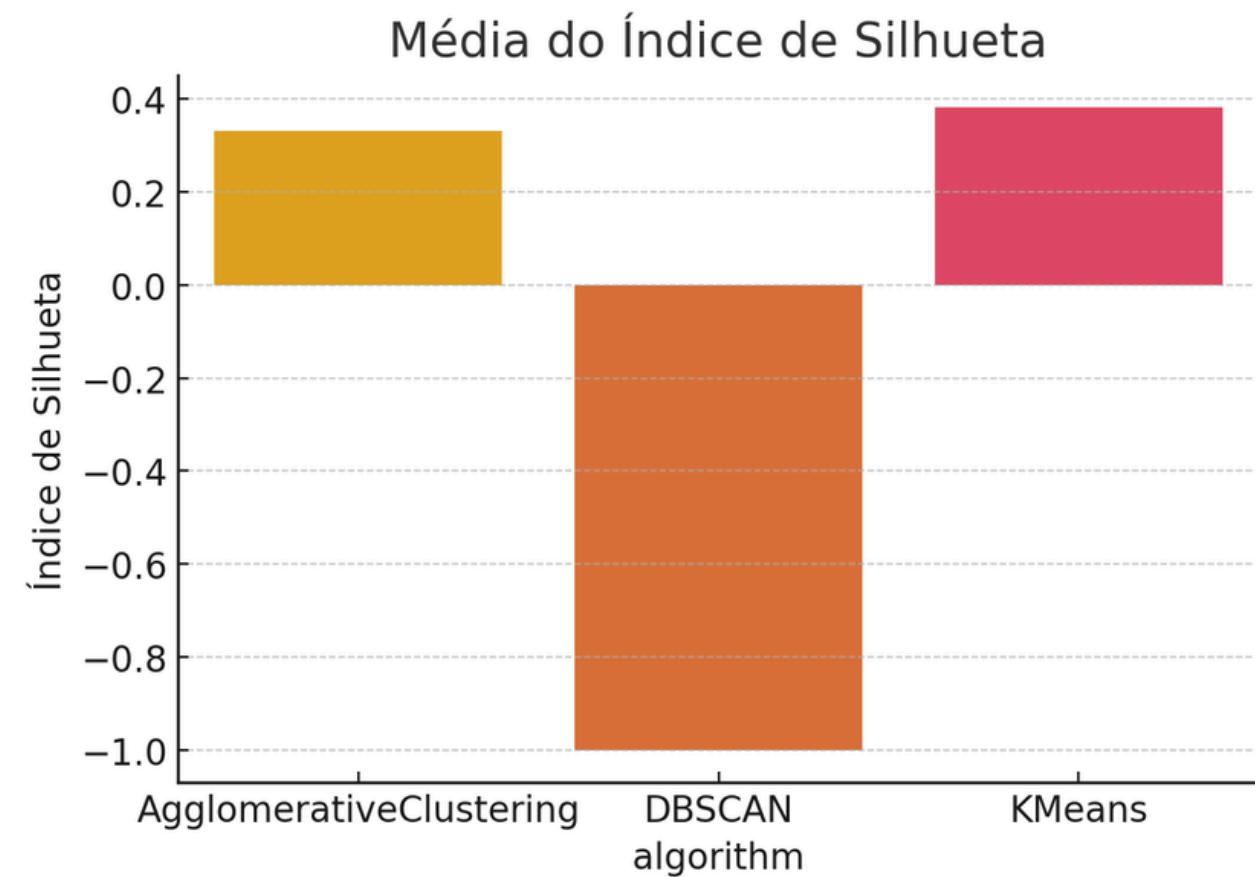
- Índice de Silhueta: 0.3815
- Índice de Davies-Bouldin: 0.8627
- Tempo de Execução: 0.0134 segundos

Agglomerative Clustering:

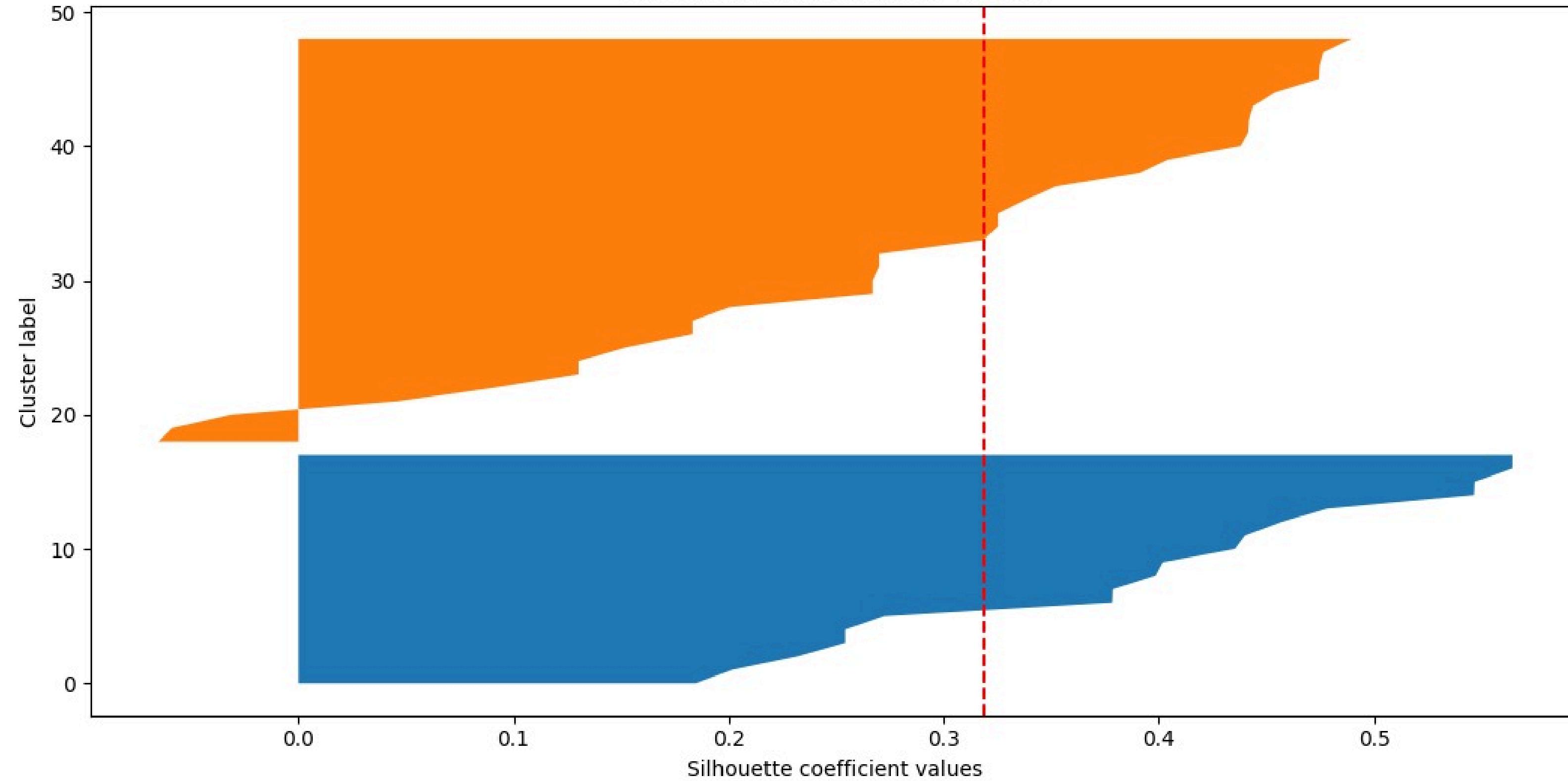
- Índice de Silhueta: 0.3308
- Índice de Davies-Bouldin: 0.8528
- Tempo de Execução: 0.0035 segundos

DBSCAN:

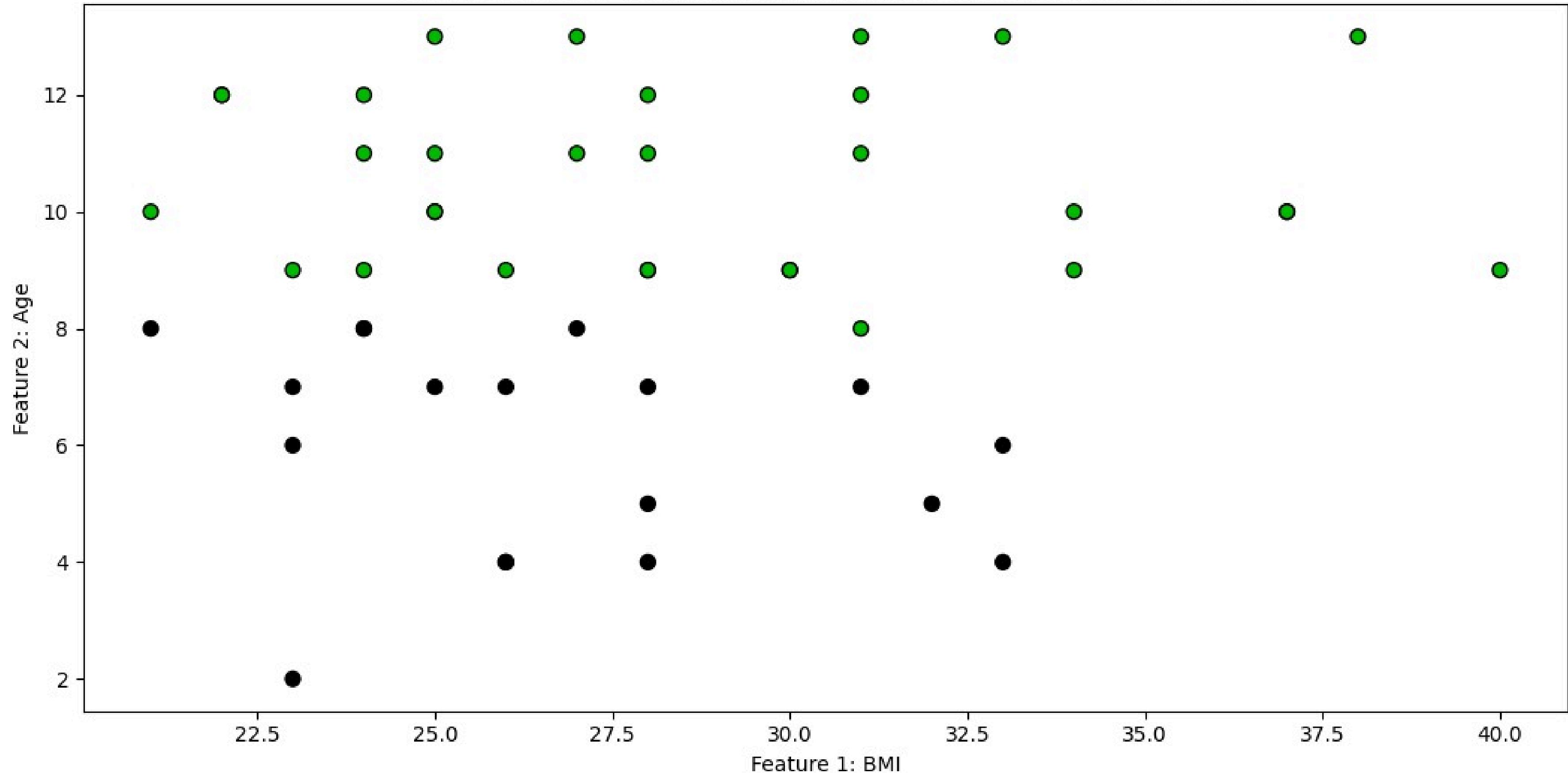
- Índice de Silhueta: -1.0000 (indicando que a clusterização pode não ter sido bem-sucedida)
- Índice de Davies-Bouldin: -1.0000
- Tempo de Execução: 0.0012 segundos



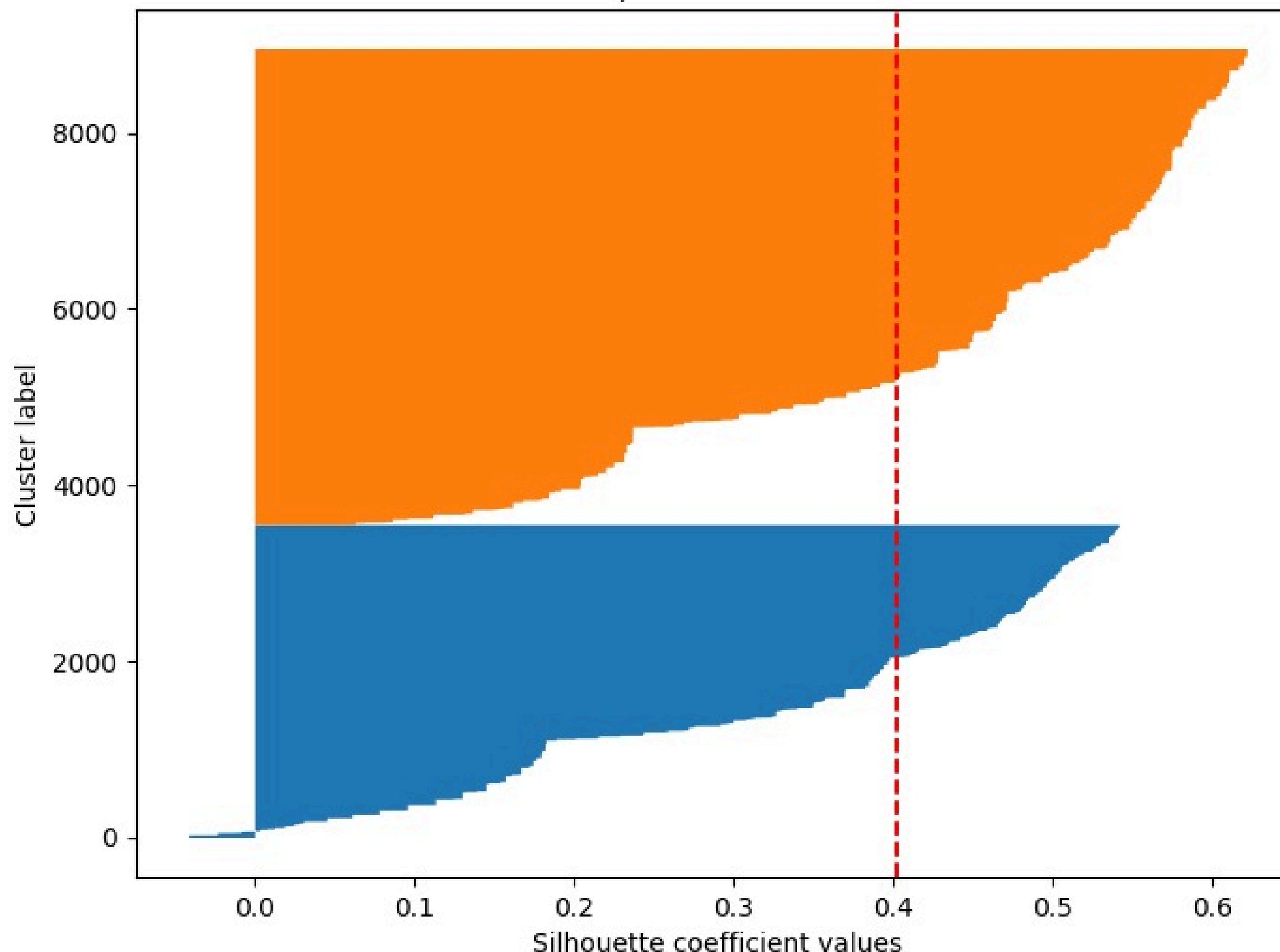
Silhouette plot for various clusters



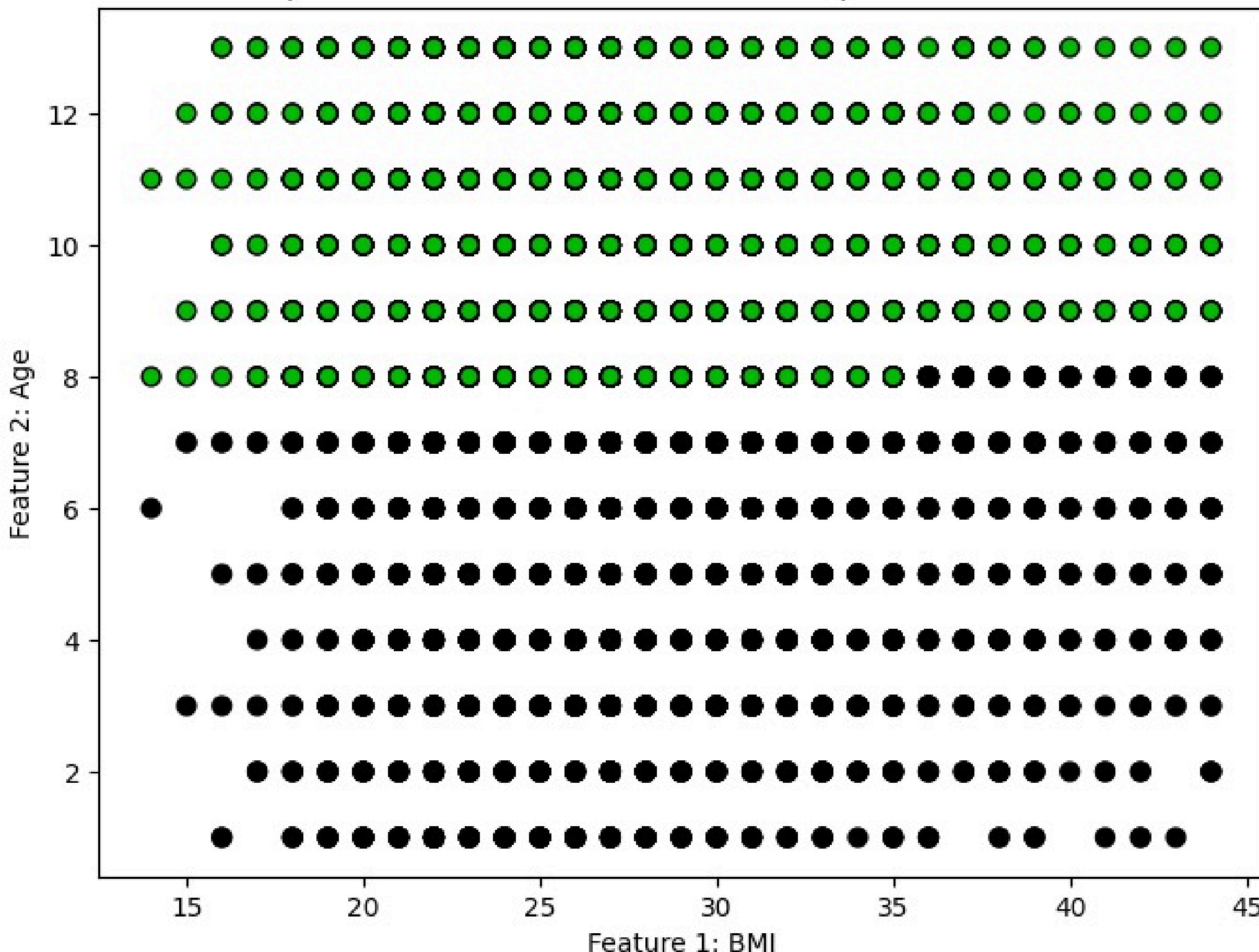
Feature space for the 1st feature x Feature space for the 2nd feature



Silhouette plot for various clusters



Feature space for the 1st feature x Feature space for the 2nd feature



Referencias Bibliograficas

- ✓ Zhang et al. (2022). Diabetes duration and types of diabetes treatment in data-driven clusters of patients with diabetes. *Frontiers in Endocrinology*.
- ✓ Somolinos-Simón et al. (2024). Cluster analysis of adult individuals with type 1 diabetes: Treatment pathways and complications over a five-year follow-up period. *Diabetes Research and Clinical Practice*.

