

---

## Table of Contents

Wine Dataset Klaszterezés - 5 csoport az összes paraméter alapján .....	1
1. Adatok betöltése .....	1
2. Adatelőkészítés, paraméterek meghatározása .....	1
3. Adatok normalizálása .....	2
4. Borok jellemzőinek ábrázolása oszlopdiagramokkal .....	2
5. K-means klaszterezés 5 klaszterre .....	4
6. Hierarchikus klaszterezés 5 klaszterre .....	4
7. Eredmények vizualizációja .....	5
8. Klaszterek jellemzői .....	5
9. Klaszterközpontok vizualizációja oszlopdiagramokkal .....	6
10. Eredmények exportálása .....	7

## Wine Dataset Klaszterezés - 5 csoport az összes paraméter alapján

```
clear; close all; clc;
```

```
Error using evalmxdm>instrumentAndRun (line 116)  
Error: Invalid text character. Check for unsupported symbol, invisible  
character, or pasting of non-ASCII characters.
```

```
Error in evalmxdm (line 21)  
[data,text,laste] =  
instrumentAndRun(file,cellBoundaries,imageDir,imagePrefix,options);
```

```
Error in publish
```

### 1. Adatok betöltése

```
data = readtable('winequality-red.csv');  
disp('Az adathalmaz első néhány sora:');  
disp(head(data));
```

### 2. Adatelőkészítés, paraméterek meghatározása

Minőség oszlop eltávolítása a dataset-ből

```
X = table2array(data(:,1:11));  
quality = data.quality;  
  
param_names = {'Fix savasság', 'Illékony savasság', 'Citromsav', ...  
               'Maradék cukor', 'Kloridok', 'Szabad SO2', ...  
               'Összes SO2', 'Sűrűség', 'pH érték', 'Szulfátok', 'Alkohol'};  
outputNames = {'Minőség'};
```

---

## 3. Adatok normalizálása

```
X_norm = zscore(X);
```

## 4. Borok jellemzőinek ábrázolása oszlopdiagramokkal

```
figure('Position', [100, 100, 1200, 800]);

% Átlagos értékek számítása minőségi szintek szerint
quality_levels = unique(quality);
num_levels = length(quality_levels);
colors = parula(num_levels); % Színskála minőségi szintekhez

% Alkohol tartalom minőség szerint
subplot(2,3,1);
mean_alcohol = zeros(num_levels, 1);
for i = 1:num_levels
    mean_alcohol(i) = mean(X(quality == quality_levels(i), 11));
end
bar(quality_levels, mean_alcohol, 'FaceColor', 'flat', 'CData', colors);
xlabel('Minőség');
ylabel('Átlagos alkohol tartalom (%)');
title('Alkohol tartalom minőség szerint');
grid on;

% Fix és illékony savasság
subplot(2,3,2);
mean_fixed_acid = zeros(num_levels, 1);
mean_volatile_acid = zeros(num_levels, 1);
for i = 1:num_levels
    mean_fixed_acid(i) = mean(X(quality == quality_levels(i), 1));
    mean_volatile_acid(i) = mean(X(quality == quality_levels(i), 2));
end
bar(quality_levels, [mean_fixed_acid, mean_volatile_acid], 'grouped');
xlabel('Minőség');
ylabel('Átlagos savasság (g/dm^3)');
title('Fix és illékony savasság');
legend({'Fix savasság', 'Illékony savasság'}, 'Location', 'best');
grid on;

% Citromsav és pH
subplot(2,3,3);
mean_citric = zeros(num_levels, 1);
mean_ph = zeros(num_levels, 1);
for i = 1:num_levels
    mean_citric(i) = mean(X(quality == quality_levels(i), 3));
    mean_ph(i) = mean(X(quality == quality_levels(i), 9));
end
yyaxis left;
bar(quality_levels, mean_citric);
```

---

```

ylabel('Átlagos citromsav (g/dm^3)');
yyaxis right;
bar(quality_levels, mean_ph);
ylabel('Átlagos pH');
xlabel('Minőség');
title('Citromsav és pH érték');
legend({'Citromsav', 'pH'}, 'Location', 'best');
grid on;

% Maradék cukor és sűrűség
subplot(2,3,4);
mean_sugar = zeros(num_levels, 1);
mean_density = zeros(num_levels, 1);
for i = 1:num_levels
    mean_sugar(i) = mean(X(quality == quality_levels(i), 4));
    mean_density(i) = mean(X(quality == quality_levels(i), 8));
end
yyaxis left;
bar(quality_levels, mean_sugar);
ylabel('Átlagos maradék cukor (g/dm^3)');
yyaxis right;
bar(quality_levels, mean_density);
ylabel('Átlagos sűrűség (g/cm^3)');
xlabel('Minőség');
title('Maradék cukor és sűrűség');
legend({'Maradék cukor', 'Sűrűség'}, 'Location', 'best');
grid on;

% Szulfátok és kloridok
subplot(2,3,5);
mean_sulphates = zeros(num_levels, 1);
mean_chlorides = zeros(num_levels, 1);
for i = 1:num_levels
    mean_sulphates(i) = mean(X(quality == quality_levels(i), 10));
    mean_chlorides(i) = mean(X(quality == quality_levels(i), 5));
end
bar(quality_levels, [mean_sulphates, mean_chlorides], 'grouped');
xlabel('Minőség');
ylabel('Átlagos tartalom (g/dm^3)');
title('Szulfátok és kloridok');
legend({'Szulfátok', 'Kloridok'}, 'Location', 'best');
grid on;

% SO2 típusok
subplot(2,3,6);
mean_free_so2 = zeros(num_levels, 1);
mean_total_so2 = zeros(num_levels, 1);
for i = 1:num_levels
    mean_free_so2(i) = mean(X(quality == quality_levels(i), 6));
    mean_total_so2(i) = mean(X(quality == quality_levels(i), 7));
end
bar(quality_levels, [mean_free_so2, mean_total_so2], 'grouped');
xlabel('Minőség');
ylabel('Átlagos SO2 (mg/dm^3)');

```

---

---

```

title('Szabad és összes SO2');
legend({'Szabad SO2', 'Összes SO2'}, 'Location', 'best');
grid on;

sgtitle('Vörösborok jellemzőinek minőség szerinti eloszlása', 'FontSize',
16);

% Összes paraméter átlagos értékei minőség szerint
figure('Position', [100, 100, 1200, 600]);
all_means = zeros(num_levels, size(X,2));
for i = 1:num_levels
    all_means(i,:) = mean(X(quality == quality_levels(i), :));
end

% Normalizálás
norm_means = (all_means - min(all_means)) ./ (max(all_means) -
min(all_means));

% Heatmap
imagesc(norm_means);
colorbar;
set(gca, 'XTick', 1:length(param_names), 'XTickLabel', param_names);
set(gca, 'YTick', 1:num_levels, 'YTickLabel', quality_levels);
xtickangle(45);
ylabel('Minőség');
title('Normalizált paraméterértékek minőség szerint');

```

## 5. K-means klaszterezés 5 klaszterre

```

rng(123);
k = 5;
[km_idx, km_centroids, km_sumd] = kmeans(X_norm, k, 'Replicates', 20,
'Display', 'final');

```

## 6. Hierarchikus klaszterezés 5 klaszterre

Távolságmátrix számítása

```

distances = pdist(X_norm);

% Hierarchikus kapcsolódás
linkage_tree = linkage(distances, 'ward');

% Klaszterek létrehozása
hc_idx = cluster(linkage_tree, 'maxclust', k);

% Dendrogram megjelenítése
figure('Position', [100, 100, 800, 600]);
dendrogram(linkage_tree);
title('Hierarchikus klaszterezés dendrogramja', 'FontSize', 16);
xlabel('Mintaindex', 'FontSize', 12);
ylabel('Euklideszi távolság', 'FontSize', 12);
grid on;

```

---

```
% Vágási vonal hozzáadása az 5 klaszterhez
hold on;
cutoff_line = median([linkage_tree(end-k+2,3) linkage_tree(end-k+1,3)]);
line([0,size(X_norm,1)*1.1], [cutoff_line, cutoff_line], ...
     'Color','r', 'LineStyle','--', 'LineWidth', 1.5);
legend('Klaszterhatár', 'Location', 'northeast');
hold off;
```

## 7. Eredmények vizualizációja

PCA alkalmazása dimenziócsökkentéshez

```
[coeff, score, ~, ~, explained] = pca(X_norm);

figure('Position', [100, 100, 1200, 500]);

% K-means eredmények PCA térben
subplot(1,2,1);
gscatter(score(:,1), score(:,2), km_idx, [], 'o', 10);
title('K-means klaszterek (első két főkomponens)');
xlabel(['PC1 (' num2str(round(explained(1),2)) '%)']);
ylabel(['PC2 (' num2str(round(explained(2),2)) '%)']);
legend('Location', 'best');

% Hierarchikus eredmények PCA térben
subplot(1,2,2);
gscatter(score(:,1), score(:,2), hc_idx, [], 'o', 10);
title('Hierarchikus klaszterek (első két főkomponens)');
xlabel(['PC1 (' num2str(round(explained(1),2)) '%)']);
ylabel(['PC2 (' num2str(round(explained(2),2)) '%)']);
legend('Location', 'best');
```

## 8. Klaszterek jellemzői

```
figure('Position', [100, 100, 1400, 800]);

% Fontos paraméterek kiválasztása
selected_params = [1 2 3 10 11]; % Fix sav, Illékony sav, Citromsav,
Szulfátok, Alkohol
selected_names = param_names(selected_params);

for i = 1:length(selected_params)
    subplot(2,3,i);
    boxplot(X(:,selected_params(i)), km_idx);
    title(selected_names{i});
    xlabel('Klaszter');
    ylabel('Érték');
    grid on;
end

% Minőség ábra
subplot(2,3,6);
```

---

```
boxplot(quality, km_idx);
title('Minőség');
xlabel('Klaszter');
ylabel('Érték (1-10)');
grid on;

sgtitle('Klaszterenkénti jellemzők eloszlása', 'FontSize', 16);
```

## 9. Klaszterközpontok vizualizációja oszlopdiagramokkal

```
figure('Position', [100, 100, 1200, 800]);

% Denormalizáljuk a központokat az eredeti skálára
centroids_original = km_centroids .* std(X) + mean(X);

% Oszlopdiagramok készítése minden klaszterhez
for cluster = 1:k
    subplot(2, ceil(k/2), cluster);
    bar(centroids_original(cluster,:));
    title(['Klaszter ' num2str(cluster) ' jellemzői']);
    set(gca, 'XTick', 1:length(param_names), 'XTickLabel', param_names);
    xtickangle(45);
    ylabel('Átlagos érték');
    grid on;
end
sgtitle('Klaszterközpontok jellemzői', 'FontSize', 16);

% Összes klaszter egy ábrán

figure('Position', [100, 100, 1200, 600]);
colors = lines(k); % Különböző színek minden klaszterhez

% Oszlopdiagram csoportosított módon
for param = 1:length(param_names)
    subplot(3, 4, param);
    bar(centroids_original(:,param), 'FaceColor', 'flat', 'CData', colors);
    title(param_names{param});
    set(gca, 'XTickLabel', 1:k);
    xlabel('Klaszter');
    ylabel('Érték');
    grid on;

    if param == 12
        break;
    end
end
sgtitle('Klaszterenkénti paraméterértékek összehasonlítása', 'FontSize', 16);
```

---

## 10. Eredmények exportálása

```
data.KMeans_Cluster = km_idx;  
data.Hierarchical_Cluster = hc_idx;  
writetable(data, 'wine_clusters.xlsx');
```

*Published with MATLAB® R2024a*