

**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**  
**FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**  
**SPECIALIZAREA CIBERNETICĂ ECONOMICĂ**

**PROIECT LA DISCIPLINA INTELIGENȚĂ  
COMPUTAȚIONALĂ ÎN ECONOMIE**

**Profesori coordonatori:**

CONF.UNIV.DR. GEORGESCU IRINA ALEXANDRA

**Student:**

LUNGU VANESA-DENISA

**București**

**2025**

# CUPRINS

<b>CAPITOLUL 1: DESCRIEREA DATELOR ȘI STATISTICI DESCRIPTIVE</b>	
.....	<b>3</b>
<b>CAPITOLUL 2: CLUSTERIZARE FUZZY</b>	
.....	<b>7</b>
<b>CAPITOLUL 3: REGRESIE LOGISTICĂ MULTINOMIALĂ</b>	
.....	<b>16</b>
<b>CAPITOLUL 4: ARBORI DE DECIZIE ȘI DE CLASIFICARE. CURĂȚAREA ARBORILOR</b>	
.....	<b>24</b>
<b>CAPITOLUL 5: KNN DE CLASIFICARE</b>	
.....	<b>32</b>
<b>CAPITOLUL 6: REȚELE NEURONALE PENTRU CLASIFICARE</b>	
.....	<b>38</b>
<b>ANEXE .....</b>	<b>45</b>

# CAPITOLUL 1

## DESCRIEREA DATELOR ȘI STATISTICI DESCRIPTIVE

1. Obiectivul acestei analize este acela de a observa indicatorii din domeniul educational al fiecărei

regiuni la nivel global, pentru a evidenția diferențele dintre:

- Numărul persoanelor cu studii superioare și cele fără studii;
- Persoanele analfabete;
- Rata abandonului școlar.

De asemenea, analiza urmărește să identifice eventualele cauze care contribuie la aceste diferențe și tendințe. Indicatorii analizați, cum ar fi cheltuielile guvernamentale pentru educație, gradul de înscriere, rata de finalizare a învățământului primar și secundar, și durata așteptată a școlarizării, pot oferi perspective valoroase asupra factorilor care influențează succesul sau eșecul sistemelor educaționale.

Sursa datelor este World Bank, iar perioada analizată este anul 2017.

Variabila de clasificare utilizată în acest context este „Nivel educațional general al țării”, iar scopul său este de a grupa țările în funcție de numărul anilor așteptați de școlarizare.

1-Nivel scăzut (numărul anilor așteptați de școlarizare < 10)

2-Nivel mediu (numărul anilor așteptați de școlarizare între 10 și 13)

3-Nivel ridicat (numărul anilor așteptați de școlarizare > 13)

Tara	PA	AAS	GIIS	GIIP	CGIS
Length:100	Min. : 3026	Min. : 4.95	Min. : 20.68	Min. : 74.74	Min. : 50.8
Class :character	1st Qu.: 815456	1st Qu.:11.69	1st Qu.: 97.45	1st Qu.:100.12	1st Qu.: 3636.1
Mode :character	Median : 815456	Median :12.90	Median :102.94	Median :101.82	Median : 3636.1
	Mean : 1831097	Mean :12.12	Mean : 99.90	Mean :102.29	Mean : 10175.9
	3rd Qu.: 815456	3rd Qu.:13.42	3rd Qu.:106.39	3rd Qu.:104.07	3rd Qu.: 3636.1
	Max. :39831098	Max. :13.96	Max. :158.54	Max. :143.92	Max. :325625.6
CGIP	CGE	NPIS	NPPI	RFIP	
Min. : 32.41	Min. : 93.76	Min. : 3240	Min. : 2005	Min. :50.07	
1st Qu.: 2389.00	1st Qu.: 2438.63	1st Qu.: 523128	1st Qu.: 379186	1st Qu.:96.30	
Median : 2777.78	Median : 4780.55	Median : 1216214	Median : 889102	Median :96.30	
Mean : 8110.21	Mean : 17448.84	Mean : 4996317	Mean : 5335395	Mean :95.75	
3rd Qu.: 3115.38	3rd Qu.: 11269.64	3rd Qu.: 2793292	3rd Qu.: 3089297	3rd Qu.:96.30	
Max. :293636.20	Max. :179812.03	Max. :129829194	Max. :143227427	Max. :99.70	

1. PA (Populația analfabetă, 25-64 ani)

- Media: Numărul mediu de persoane analfabete este de 1.831.097.

- Mediana: Jumătate dintre țări au o populație analfabetă mai mică de 815.456, iar cealaltă jumătate o populație mai mare.

- Minimum: Cea mai mică populație analfabetă raportată este de 3.026 persoane.

- Maximum: Cea mai mare populație analfabetă raportată este de 3.983.1098 persoane.

## 2. AAS (Ani așteptați de școlarizare)

- Media: În medie, se așteaptă ca un copil să fie școlarizat timp de 12,12 ani.

- Mediana: Jumătate dintre țări au o durată așteptată a școlarizării mai mică de 12,9 ani, iar cealaltă jumătate mai mare.

- Minimum: Țara cu cea mai scurtă durată așteptată a școlarizării are doar 4,95 ani.

- Maximum: Cea mai mare durată așteptată a școlarizării este de 13,96 ani.

## 3. GIIS (Gradul de înscriere în învățământul secundar)

- Media: Gradul mediu de înscriere în învățământul secundar este de 99,90%.

- Mediana: Jumătate dintre țări au un grad de înscriere sub 102,94%, iar cealaltă jumătate peste.

- Minimum: Cel mai mic grad de înscriere în învățământul secundar este de 20,68%.

- Maximum: Cel mai mare grad de înscriere este de 158,54%.

## 4. GIIP (Gradul de înscriere în învățământul primar)

- Media: Gradul mediu de înscriere în învățământul primar este de 102,29%.

- Mediana: Jumătate dintre țări au un grad de înscriere mai mic de 101,82%, iar cealaltă jumătate mai mare.

- Minimum: Cel mai mic grad de înscriere este de 74,74%.

- Maximum: Cel mai mare grad de înscriere este de 143,92%.

## 5. CGIS (Cheltuielile guvernamentale pentru învățământul secundar)

- Media: Cheltuielile guvernamentale medii pentru învățământul secundar sunt de 10.175,9 milioane USD.

- Mediana: Jumătate dintre țări cheltuie mai puțin de 3.636,1 milioane USD, iar cealaltă jumătate mai mult.

- Minimum: Cele mai mici cheltuieli raportate sunt de 50,8 milioane USD.
- Maximum: Cele mai mari cheltuieli raportate sunt de 325.625,6 milioane USD.

#### 6. CGIP (Cheltuielile guvernamentale pentru învățământul primar)

- Media: Cheltuielile guvernamentale medii pentru învățământul primar sunt de 8.110,21 milioane USD.
- Mediana: Jumătate dintre țări cheltuie mai puțin de 2.777,78 milioane USD, iar cealaltă jumătate mai mult.
- Minimum: Cele mai mici cheltuieli raportate sunt de 32,41 milioane USD.
- Maximum: Cele mai mari cheltuieli raportate sunt de 293.636,2 milioane USD.

#### 7. CGE (Cheltuielile guvernamentale totale pentru educație)

- Media: Cheltuielile guvernamentale totale medii pentru educație sunt de 17.448,84 milioane USD.
- Mediana: Jumătate dintre țări cheltuie mai puțin de 4.780,55 milioane USD, iar cealaltă jumătate mai mult.
- Minimum: Cele mai mici cheltuieli raportate sunt de 2.389,00 milioane USD.
- Maximum: Cele mai mari cheltuieli raportate sunt de 179.812,03 milioane USD.

#### 8. NPIS (Numărul de persoane înscrise în învățământul secundar)

- Media: Numărul mediu de persoane înscrise în învățământul secundar este de 4.996.317.
- Mediana: Jumătate dintre țări au mai puțin de 1.216.214 persoane înscrise, iar cealaltă jumătate mai mult.
- Minimum: Cel mai mic număr de înscriși este de 3.240 persoane.
- Maximum: Cel mai mare număr de înscriși este de 12.982.194 persoane.

#### 9. NPIP (Numărul de persoane înscrise în învățământul primar)

- Media: Numărul mediu de persoane înscrise în învățământul primar este de 5.335.395.
- Mediana: Jumătate dintre țări au mai puțin de 889.102 persoane înscrise, iar cealaltă jumătate mai mult.
- Minimum: Cel mai mic număr de înscriși este de 2.005 persoane.

- Maximum: Cel mai mare număr de înscriși este de 14.322.742 persoane.

#### 10. RFIP (Rata de finalizare a învățământului primar)

- Media: Rata medie de finalizare a învățământului primar este de 95,75%.
- Mediana: Jumătate dintre țări au o rată de finalizare sub 96,30%, iar cealaltă jumătate peste.
- Minimum: Cea mai mică rată de finalizare raportată este de 50,0%.
- Maximum: Cea mai mare rată de finalizare este de 99,7%.

## CAPITOLUL 2

### CLUSTERIZARE FUZZY

	[,1]	[,2]	[,3]
Afghanistan	0.43826334	0.2372906	0.3244461
Australia	0.31029412	0.4016128	0.2880931
Azerbaijan	0.07094031	0.2707457	0.6583140
Belgium	0.21372016	0.4727081	0.3135717
Albania	0.13097470	0.3647731	0.5042522

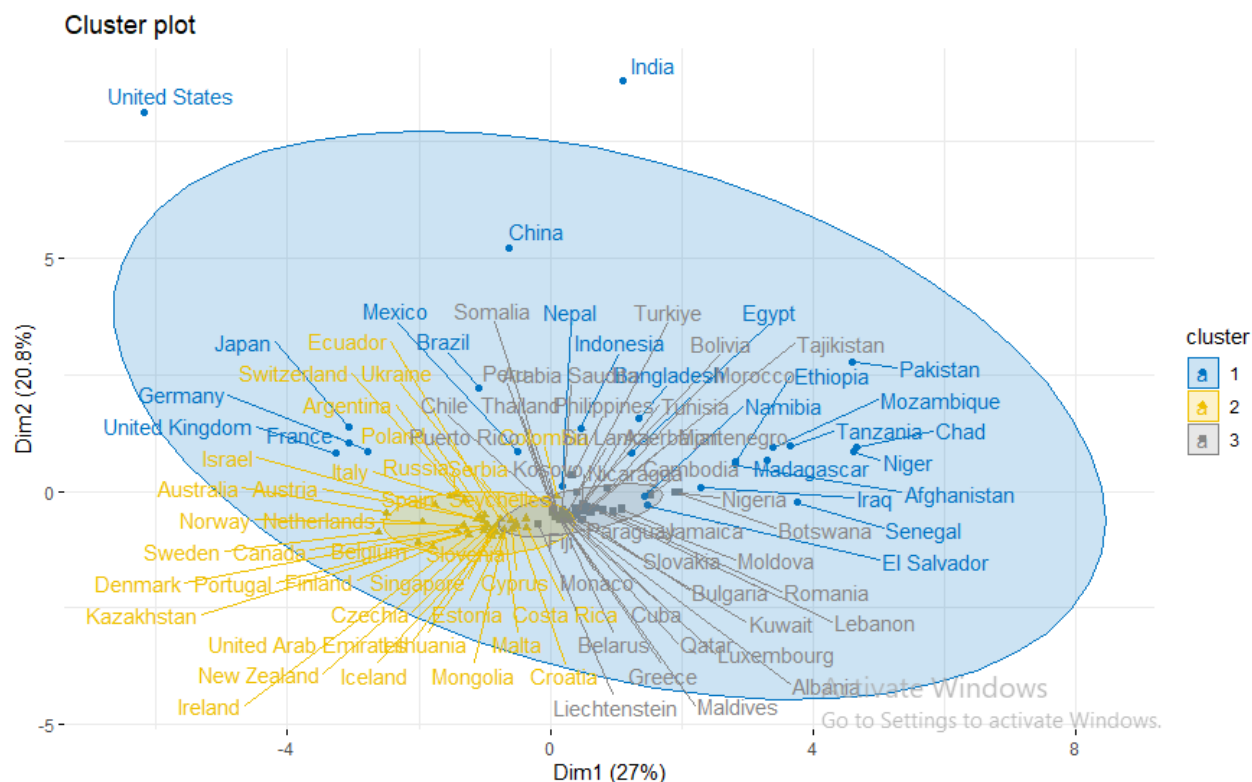
1. **Cluster 1** (probabil cel mai puțin dezvoltat):
  - Afghanistan are cea mai mare apartenență (43.8%)
  - Azerbaijan are cea mai mică apartenență (7.1%)
2. **Cluster 2** (probabil țări dezvoltate):
  - Belgium (47.3%) și Australia (40.2%) au cele mai mari scoruri
  - Azerbaijan are doar 27.1% apartenență
3. **Cluster 3** (probabil țări în curs de dezvoltare):
  - Azerbaijan (65.8%) și Albania (50.4%) domină
  - Australia are cea mai mică apartenență (28.8%)

```
dunn_coeff normalized
0.4272881  0.1409321
```

**0.427 sugerează o separare moderată între clustere, dar nu excelentă**

0.141 indică o structură slabă a clusterelor

- Ambele versiuni ale coeficientului indică o **separare sub-optimă** a clusterelor
- Clusterii au o anumită structură (valori  $> 0$ ), dar nu foarte bine definită



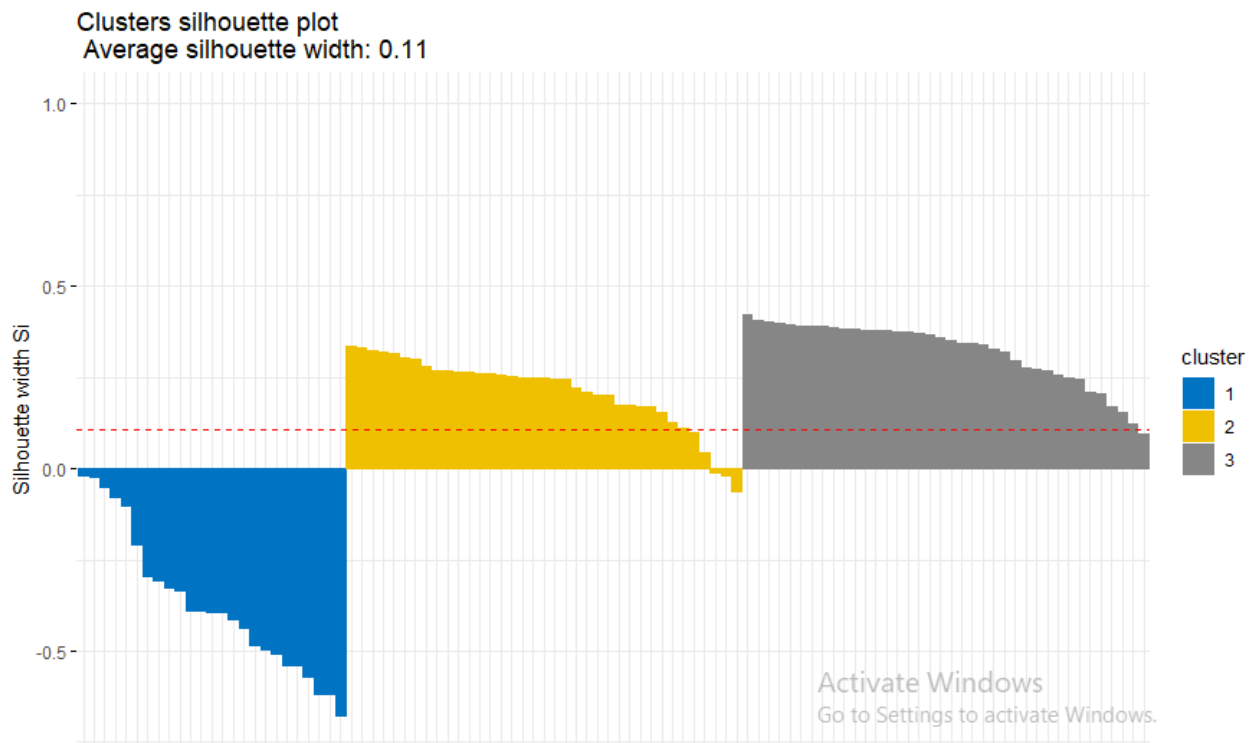
**Dim1 (27%)** și **Dim2 (20.8%)** reprezintă primele două componente principale care captează împreună 47.8% din variația datelor.

Clusterizarea fuzzy (probabil Fuzzy C-Means) asociază fiecărei țări o apartenență parțială la mai multe grupuri, dar în grafic este afișată doar **cea mai probabilă** apartenență:

- **Albastru (cluster 1):** țări precum SUA, China, India, majoritatea țărilor africane și asiatice.
- **Galben (cluster 2):** țări dezvoltate din Europa de Vest, America de Sud și câteva din Asia (ex: Japonia, Australia, Germania, Canada).
- **Gri (cluster 3):** multe țări europene, din Orientul Mijlociu și America Latină, poziționate central în grafic (valori moderate pe ambele dimensiuni).
- **Clusterul 1 (Albastru):**
  - Conține țări mari ca suprafață și populație, unele în curs de dezvoltare (India, China) și țări africane. Acest cluster ar putea indica:
    - Țări cu valori mari pe o dimensiune (Dim1 pozitiv), posibil asociată cu populație mare, urbanizare sau alte metrice structurale.
- **Clusterul 2 (Galben):**
  - Țări dezvoltate economic și social (ex. Norvegia, Elveția, Japonia, Canada, Australia).
  - Poziționate în stânga jos, semn al unor valori negative pe Dim1 și Dim2. Probabil acest cluster indică:



- Țări cu standard de viață ridicat, stabilitate economică, scoruri înalte în educație, sănătate etc.
- **Clusterul 3 (Gri):**
  - Țări din Europa de Est, America Latină, și Asia Centrală.
  - Poziționate central, ceea ce poate sugera că aceste țări au valori moderate față de celelalte două grupuri.



**Lățimea medie Silhouette de 0.11** indică o calitate slabă a clusterizării, sugerând că:

- Structura clusterelor este **minimală sau aproape aleatoare**
- Punctele de date nu sunt bine grupate
- Granițele dintre clusterare sunt neclare

Clusterul 1 (presupunând din layout):

Probabil cel mai bine definit

Clusterul 2 și 3:

Probabil suprapunere semnificativă

Multe puncte cu valori Silhouette apropiate de 0.

	cluster	size	ave.sil.width
1	1	61	-0.03
2	2	39	0.50

Clusterul 1 (61 de observații)

- **Lățime medie Silhouette: -0.03**
  - Valori negative indică o **clusterizare slabă**
  - Multe puncte sunt probabil atribuite greșit acestui cluster
  - Structura internă este **necoerentă**

Clusterul 2 (39 de observații)

- **Lățime medie Silhouette: 0.50**
  - Valori  $> 0.25$  indică o **structură rezonabilă**
  - Punctele sunt bine grupate și separate de alte clustere
  - Cluster **bine definit și omogen**

Calitate Inegală:

Un cluster bun (2) și un cluster problematic (1)

Discrepanța mare sugerează posibile probleme în algoritm sau date

Distribuție Asimetrică:

Clusterul slab este mai mare (61 vs 39 observații)

Posibilă nevoie de reechilibrare

# Country Cluster

	Tara	Cluster
Afghanistan	Afghanistan	1
Australia	Australia	2
Azerbaijan	Azerbaijan	3
Belgium	Belgium	2
Albania	Albania	3
Argentina	Argentina	2
Austria	Austria	2
Bangladesh	Bangladesh	1
Belarus	Belarus	3
Botswana	Botswana	3
Bolivia	Bolivia	3
Brazil	Brazil	1
Bulgaria	Bulgaria	3
Canada	Canada	2
Cambodia	Cambodia	3
Chad	Chad	1
Chile	Chile	3
Colombia	Colombia	2
China	China	1
Cuba	Cuba	3
Croatia	Croatia	2
Costa Rica	Costa Rica	2
Cyprus	Cyprus	2
Denmark	Denmark	2
Czechia	Czechia	2
El Salvador	El Salvador	1
Egypt	Egypt	1
Ecuador	Ecuador	2
Estonia	Estonia	2
Ethiopia	Ethiopia	1
Fiji	Fiji	3
Finland	Finland	2
France	France	1
Germany	Germany	1
Greece	Greece	3
Iceland	Iceland	2
India	India	1
Indonesia	Indonesia	1
Iraq	Iraq	1
Italy	Italy	2
Japan	Japan	1
Ireland	Ireland	2
Jamaica	Jamaica	3
Israel	Israel	2
Kuwait	Kuwait	3
Kosovo	Kosovo	3
Kazakhstan	Kazakhstan	2
Lithuania	Lithuania	2
Luxembourg	Luxembourg	3
Madagascar	Madagascar	1
Maldives	Maldives	3
Malta	Malta	2

Mexico	Mexico	1
Moldova	Moldova	3
Monaco	Monaco	3
Mongolia	Mongolia	2
Mozambique	Mozambique	1
Namibia	Namibia	1
Morocco	Morocco	3
Nepal	Nepal	1
Netherlands	Netherlands	2
New Zealand	New Zealand	2
Nigeria	Nigeria	3
Norway	Norway	2
Pakistan	Pakistan	1
Paraguay	Paraguay	3
Peru	Peru	3
Poland	Poland	2
Portugal	Portugal	2
Philippines	Philippines	3
Puerto Rico	Puerto Rico	3
Qatar	Qatar	3
Romania	Romania	3
Russia	Russia	2
Arabia Saudita	Arabia Saudita	3
Senegal	Senegal	1
Seychelles	Seychelles	2
Singapore	Singapore	2
Slovakia	Slovakia	3
Serbia	Serbia	2
Somalia	Somalia	3
Slovenia	Slovenia	2
Sri Lanka	Sri Lanka	3
Spain	Spain	2
Sweden	Sweden	2
Switzerland	Switzerland	2
Thailand	Thailand	3
Tanzania	Tanzania	1
Turkiye	Turkiye	3
Tunisia	Tunisia	3
Ukraine	Ukraine	2
United Kingdom	United Kingdom	1
United States	United States	1
United Arab Emirates	United Arab Emirates	2
Tajikistan	Tajikistan	3
Nicaragua	Nicaragua	3
Niger	Niger	1
Montenegro	Montenegro	3
Lebanon	Lebanon	3
Liechtenstein	Liechtenstein	3

	cluster	neighbor	sil_width
United States	1	2	-0.01868627
India	1	3	-0.02130381
Senegal	1	3	-0.04878316
Pakistan	1	3	-0.07674748
China	1	3	-0.10276957
Bangladesh	1	3	-0.20996831

**1. Clusterul 1 are probleme grave:**

- Toate aceste țări ar aparține mai bine altor clustere (2 sau 3)
- Clusterul 1 pare a fi un "cluster rezidual" neomogen

**2. Țări cu probleme majore:**

- Bangladesh are cea mai slabă apartenență (-0.21)
- China și Pakistan au probleme semnificative

**3. Modelul de clusterizare:**

- Probabil număr neoptim de clustere
- Metrică de distanță inadecvată
- Variabile ne semnificative

Țară	Silhouette Width	Semnificație
United States	-0.0187	Ușor atribuită greșit
India	-0.0213	Ușor atribuită greșit
Senegal	-0.0488	Atribuire moderat nepotrivită
Pakistan	-0.0767	Atribuire nepotrivită clară
China	-0.1028	Atribuire nepotrivită semnificativă
Bangladesh	-0.2100	Atribuire profund nepotrivită

Secvența de cod:

```
install.packages("cluster")
```

```
install.packages("readxl")
```

```
install.packages("factoextra")
```

```
library(cluster)
```

```
library(readxl)
```

```
library(factoextra)
```

```
# 2. Încărcare date
```

```
date <- read_excel("ICE.xlsx")
```

```
tari <- date$Tara
```

```
date1 <- date[, sapply(date, is.numeric)]
```

```
rownames(date1) <- tari
```

```
# 4.Standardizare
```

```
date1 <- scale(date1)
```

```
# 5. Clusterizare fuzzy cu `fanny()` (alegem 3 clustere)
```

```
res.fanny <- fanny(date1, k = 3, metric = "euclidean", stand = TRUE, memb.exp = 1.5)
```

```
# 6. Afișare rezultate
```

```
head(res.fanny$membership, 5) # Primele 5 grade de apartenență
```

```
res.fanny$coeff           # Coeficientul Dunn (calitatea clusterizării)
```

```
# 7. Vizualizare clustere (cu etichete)
```

```
fviz_cluster(res.fanny,
```

```
ellipse.type = "norm",  
repel = TRUE,  
palette = "jco",  
ggtheme = theme_minimal(),  
legend = "right")
```

# 8. Analiza siluetei (calitatea fiecărui obiect și globală)

```
fviz_silhouette(res.fanny,  
  palette = "jco",  
  ggtheme = theme_minimal())
```

# 9. Rezultatul clusterizării (țară + cluster)

```
rezultate <- data.frame(Tara = tari, Cluster = res.fanny$clustering)  
print(rezultate)
```

# 10. Silueta individuală (pentru interpretări avansate)

```
head(res.fanny$silinfo$widths)
```

### CAPITOLUL 3

#### REGRESIE LOGISTICĂ MULTINOMIALĂ

```
# weights: 18 (10 variab
initial value 88.987595
iter 10 value 31.906246
iter 20 value 7.295559
iter 30 value 6.952657
iter 40 value 6.376789
iter 50 value 5.639146
iter 60 value 5.461054
iter 70 value 5.311412
iter 80 value 5.129559
iter 90 value 4.991384
iter 100 value 4.892482
final value 4.892482
```

initial value: 88.987595 — valoarea inițială a funcției de cost (log-likelihood negativ) înainte de antrenare.

iter XX value YY.YYYYYY — valoarea funcției de cost după fiecare 10 iterații.

final value: 4.892482 — valoarea finală a funcției de cost, după optimizare (convergență).

- Modelul a pornit de la o funcție de cost mare (88.987595), ceea ce indică o potrivire slabă inițială.
- După **100 de iterații**, a reușit să minimizeze semnificativ funcția de cost până la 4.892482.
- Această scădere **accentuată și stabilizată** arată că **optimizarea a convergat** eficient și că modelul s-a antrenat corect.



```

Call:
multinom(formula = `Nivel educațional general al țării` ~
., data = train_data)

Coefficients:
(Intercept) `Ani asteptati de școlarizare` `Gradul de înscriere în învățământul primar`
2 18.26863 22.69157 -0.8165647
3 -56.19979 44.45503 -0.9348848
`Gradul de înscriere în învățământul secundar` `Rata de finalizare a învățământului primar`
2 -0.3291204 -1.410803
3 -0.1527633 -3.643877

Std. Errors:
(Intercept) `Ani asteptati de școlarizare` `Gradul de înscriere în învățământul primar`
2 11.313224 13.14418 0.4585801
3 1.190317 15.36105 0.4875600
`Gradul de înscriere în învățământul secundar` `Rata de finalizare a învățământului primar`
2 0.2707306 0.7970115
3 0.3116300 1.5117955

Residual Deviance: 9.784965
AIC: 29.78496

```

Coeficienți (efecte asupra log-odd-urilor)

Modelul prezice probabilitățile pentru trei niveluri educaționale (presupunând că nivelul 1 este categoria de referință):

#### 1. Pentru nivelul 2 vs nivelul 1:

- Fiecare an suplimentar de școlarizare așteptată crește log-odd-urile pentru nivelul 2 cu 22.69
- Înscrierea în învățământul primar scade log-odd-urile (-0.82)
- Înscrierea în învățământul secundar scade log-odd-urile (-0.33)
- Rata de finalizare a învățământului primar scade log-odd-urile (-1.41)

#### 2. Pentru nivelul 3 vs nivelul 1:

- Efectul anilor de școlarizare așteptați este și mai mare (44.46)
- Toți ceilalți coeficienți au efect negativ, cel mai puternic fiind pentru rata de finalizare a învățământului primar (-3.64)

Erorile standard

Erorile standard sunt relativ mari pentru coeficienții anilor de școlarizare așteptați, sugerând o anumită incertitudine în estimări.

Măsuri de ajustare

- **Devianța reziduală:** 9.78 (o măsură a bunătații de potrivire, valori mai mici indică un model mai bun)
- **AIC:** 29.78 (util pentru compararea modelelor, valori mai mici sunt preferate)

## Confusion Matrix and Statistics

```

      Reference
Prediction 1 2 3
      1 2 2 0
      2 0 7 1
      3 0 0 7

```

## Overall Statistics

```

      Accuracy : 0.8421
      95% CI : (0.6042, 0.9662)
No Information Rate : 0.4737
P-Value [Acc > NIR] : 0.001067

```

Kappa : 0.7467

Mcnemar's Test P-Value : NA

## Statistics by Class:

	Class: 1	Class: 2	Class: 3
Sensitivity	1.0000	0.7778	0.8750
Specificity	0.8824	0.9000	1.0000
Pos Pred Value	0.5000	0.8750	1.0000
Neg Pred Value	1.0000	0.8182	0.9167
Prevalence	0.1053	0.4737	0.4211
Detection Rate	0.1053	0.3684	0.3684
Detection Prevalence	0.2105	0.4211	0.3684
Balanced Accuracy	0.9412	0.8389	0.9375

- **Clasa 1:**
  - Corect prezisă de 2 ori
  - 2 predicții greșite (clasificate ca 2)
- **Clasa 2:**
  - Corect prezisă de 7 ori
  - 1 predicție greșită (clasificată ca 3)
- **Clasa 3:**
  - Corect prezisă de 7 ori
  - 1 predicție greșită (clasificată ca 2)

## Metricile Globale

1. **Acuratețe (Accuracy):** 84.21% (16/19 predicții corecte)

- Interval de încredere 95%: (60.42%, 96.62%)
- Semnificativ mai bun decât un model aleatoriu (p-value = 0.001)

## 2. Coeficientul Kappa: 0.7467

- Acord bun între predicții și realitate (valori peste 0.7 sunt considerate bune)

### Metricile pe Clase

#### Clasa 1

- **Sensitivitate:** 100% (detectează toate cazurile reale)
- **Specificitate:** 88.24% (identifică corect non-cazurile)
- **Precizie:** 50% (jumătate din predicțiile pentru clasa 1 sunt corecte)

#### Clasa 2

- **Sensitivitate:** 77.78%
- **Specificitate:** 90%
- **Precizie:** 87.5%

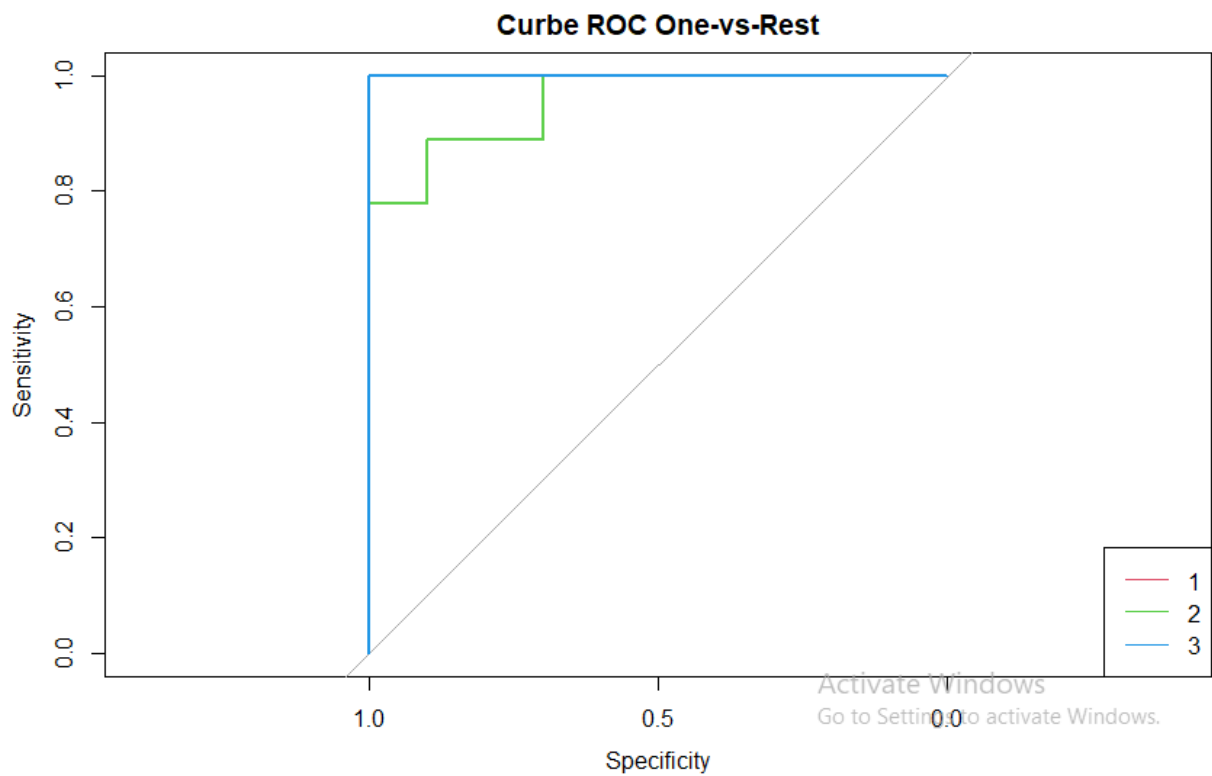
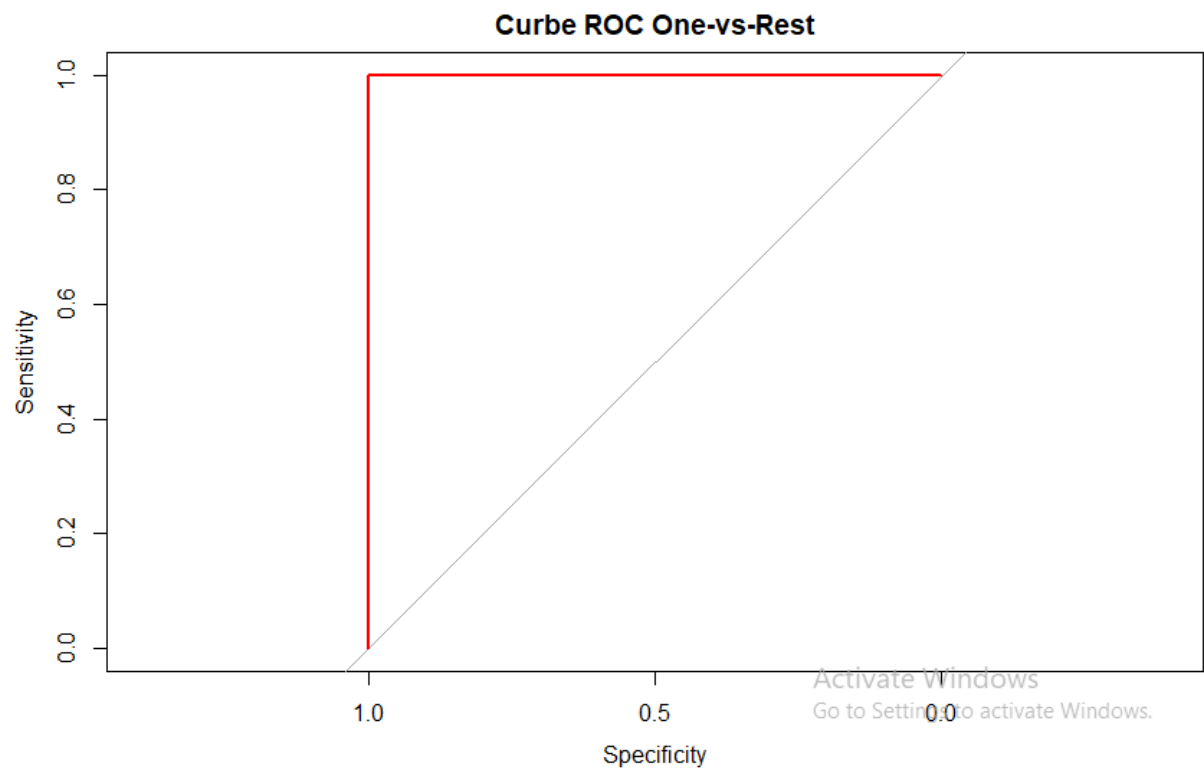
#### Clasa 3

- **Sensitivitate:** 87.5%
- **Specificitate:** 100% (nicio falsă identificare ca 3)
- **Precizie:** 100% (toate predicțiile pentru 3 sunt corecte)

	1	2	3
1	4.607819e-39	1.686643e-11	1.000000e+00
2	4.762994e-26	9.999583e-01	4.170842e-05
3	7.346836e-26	9.291799e-01	7.082013e-02
4	2.570090e-45	1.672114e-09	1.000000e+00
5	9.622606e-01	3.773943e-02	6.168970e-32
6	1.765648e-25	9.769568e-01	2.304323e-02

1. **Predicții clare:** Observațiile 1, 2, 4 și 5 au probabilități foarte apropiate de 0 sau 1, indicând încredere mare în predicție.
2. **Predicții cu incertitudine:** Observațiile 3 și 6 arată probabilități dominante pentru clasa 2, dar cu o oarecare probabilitate (7.08% și respectiv 2.30%) pentru clasa 3.
3. **Model decisiv:** În majoritatea cazurilor, modelul atribuie probabilități foarte apropiate de 0 sau 1, sugerând că este foarte sigur în predicțiile sale.

4. **Nicio ambiguitate majoră:** Nu există cazuri unde modelul să fie puternic indecis între două clase (de exemplu, 45%-55%).



□ **Roșu (Clasa 1) și Albastru (Clasa 3):**

- Se suprapun și ating colțul stânga-sus.
- Asta indică un **model perfect sau aproape perfect** pentru aceste două clase ( $AUC \approx 1.0$ ).
- Regresia logistică face **clasificări extrem de precise** pentru aceste clase.

□ **Verde (Clasa 2):**

- Are o curbă în trepte, care nu atinge colțul din stânga sus.
- Asta sugerează că regresia logistică **nu separă la fel de bine** această clasă de celelalte.
- Posibil modelul are dificultăți în a identifica trăsături distinctive pentru clasa 2, sau clasa este dezechilibrată.

1	2	3
1.0000000	0.9555556	1.0000000

- **Clasa 1:**  $AUC = 1.000000$  → Modelul clasifică perfect această clasă.
- **Clasa 2:**  $AUC = 0.955556$  → Modelul clasifică foarte bine această clasă, dar nu perfect.
- **Clasa 3:**  $AUC = 1.000000$  → Clasificare perfectă și pentru această clasă.

Modelul de regresie logistică are o performanță excelentă, cu o ușoară scădere pentru clasa 2. Acest lucru poate indica fie că datele pentru clasa 2 sunt mai greu de separat, fie că sunt mai puține mostre sau mai mult zgomot.

Dacă dorești, pot analiza mai în detaliu contextul sau să-ți ofer sugestii de îmbunătățire în funcție de date.

Secvența de cod:

```
library(readxl)
```

```
library(nnet)
```

```
library(caret)
```

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

```
# 2. Încarcă datele
```

```
date <- read_excel("ICE.xlsx")
```

# 3. Selectează coloanele utile

```
date_filtrat <- date %>%  
  select(`Ani asteptati de scolarizare`,  
        `Gradul de inscriere in invatamantul primar`,  
        `Gradul de inscriere in invatamantul secundar`,  
        `Rata de finalizare a invatamantului primar`,  
        `Nivel educațional general al țării`)
```

# 4. Elimină rândurile incomplete

```
date_filtrat <- na.omit(date_filtrat)
```

# 5. Transformă variabila țintă într-un factor

```
date_filtrat$`Nivel educațional general al țării` <- as.factor(date_filtrat$`Nivel educațional  
general al țării`)
```

# 6. Împarte setul în antrenare/test

```
set.seed(123)
```

```
ind <- createDataPartition(date_filtrat$`Nivel educațional general al țării`, p = 0.8, list = FALSE)
```

```
train_data <- date_filtrat[ind, ]
```

```
test_data <- date_filtrat[-ind, ]
```

# 7. Antrenează modelul de regresie logistică multinomială

```
model_multinom <- multinom(`Nivel educațional general al țării` ~ ., data = train_data)
```

# 8. Rezumă modelul

```
summary(model_multinom)
```

# 9. Preziceri pe setul de testare

```
predictii <- predict(model_multinom, newdata = test_data)
```

# 10. Matrice de confuzie

```
conf_matrix <- confusionMatrix(predictii, test_data$`Nivel educațional general al țării`)  
print(conf_matrix)
```

```
# 11. Probabilități pentru fiecare clasă
```

```
predictii_prob <- predict(model_multinom, newdata = test_data, type = "probs")
```

```
head(predictii_prob)
```

```
library(pROC)
```

```
library(ggplot2)
```

```
# 1. Obține probabilitățile prezise pentru fiecare clasă
```

```
predictii_prob <- predict(model_multinom, newdata = test_data, type = "probs")
```

```
# 2. Listă pentru a stoca curbele ROC
```

```
roc_list <- list()
```

```
# 3. Generare curbe ROC One-vs-Rest pentru fiecare clasă
```

```
for (clasa in colnames(predictii_prob)) {
```

```
  roc_obj <- roc(
```

```
    response = as.numeric(test_data$`Nivel educațional general al țării` == clasa),
```

```
    predictor = predictii_prob[, clasa]
```

```
  )
```

```
  roc_list[[clasa]] <- roc_obj
```

```
}
```

```
# 4. Plotare toate curbele ROC într-un singur grafic
```

```
plot(roc_list[[1]], col = "red", main = "Curbe ROC One-vs-Rest")
```

```
for (i in 2:length(roc_list)) {
```

```
  lines(roc_list[[i]], col = i + 1) # Culori diferite
```

```
}
```

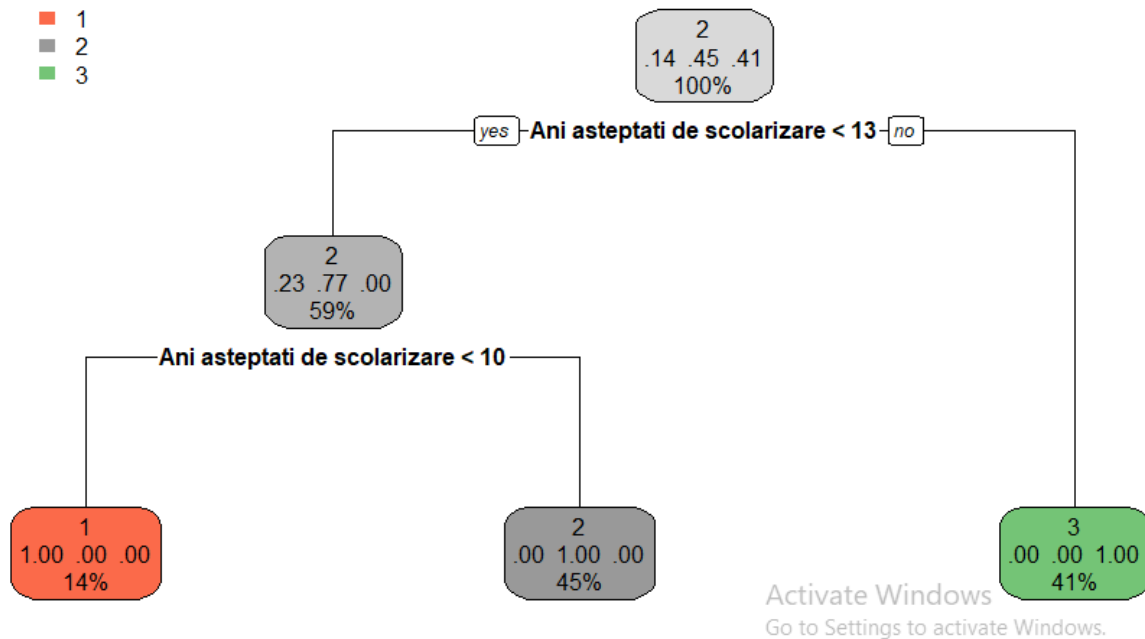
```
legend("bottomright", legend = names(roc_list), col = 2:(length(roc_list)+1), lty = 1)
```

```
# 5. Afișare AUC pentru fiecare clasă
```

```
sapply(roc_list, auc)
```

## CAPITOLUL 4

### ARBORI DE DECIZIE ȘI DE CLASIFICARE. CURĂȚAREA ARBORILOR



#### Nodul 1 (rădăcina arborelui)

- În nodul 1 sunt **100 de observații** în total.
- Distribuția pe clase este:
  - Clasa 1: 14 observații (14%)
  - Clasa 2: 45 observații (45%)
  - Clasa 3: 41 observații (41%)
- **Vectorul de probabilitate:** (0.14, 0.45, 0.41)
- **Clasa dominantă:** Clasa 2 – nivel educațional **mediu**

#### Nodul 2 (stânga, condiția: Ani așteptați < 13)

- În nodul 2 sunt **59 de observații**.
- Distribuția:



- Clasa 1: 23 observații (39%)
- Clasa 2: 36 observații (61%)
- Clasa 3: 0 observații (0%)
- **Vectorul de probabilitate:** (0.39, 0.61, 0.00)
- **Clasa dominantă:** Clasa 2

### **Nodul 3 (sub-nod stânga: Ani așteptați < 10)**

- În nodul 3 sunt **14 observații**.
- Distribuția:
  - Clasa 1: 14 observații (100%)
  - Clasa 2: 0
  - Clasa 3: 0
- **Vectorul de probabilitate:** (1.00, 0.00, 0.00)
- **Clasa dominantă:** Clasa 1 – nivel educațional scăzut

### **Nodul 4 (sub-nod dreapta: Ani așteptați $\geq 10$ și < 13)**

- În nodul 4 sunt **45 de observații**.
- Distribuția:
  - Clasa 1: 0
  - Clasa 2: 45 observații (100%)
  - Clasa 3: 0
- **Vectorul de probabilitate:** (0.00, 1.00, 0.00)
- **Clasa dominantă:** Clasa 2

### **Nodul 5 (dreapta: Ani așteptați $\geq 13$ )**

- În nodul 5 sunt **41 de observații**.
- Distribuția:

- Clasa 1: 0
- Clasa 2: 0
- Clasa 3: 41 observații (100%)
- **Vectorul de probabilitate:** (0.00, 0.00, 1.00)
- **Clasa dominantă:** Clasa 3 – nivel educațional **ridicat**

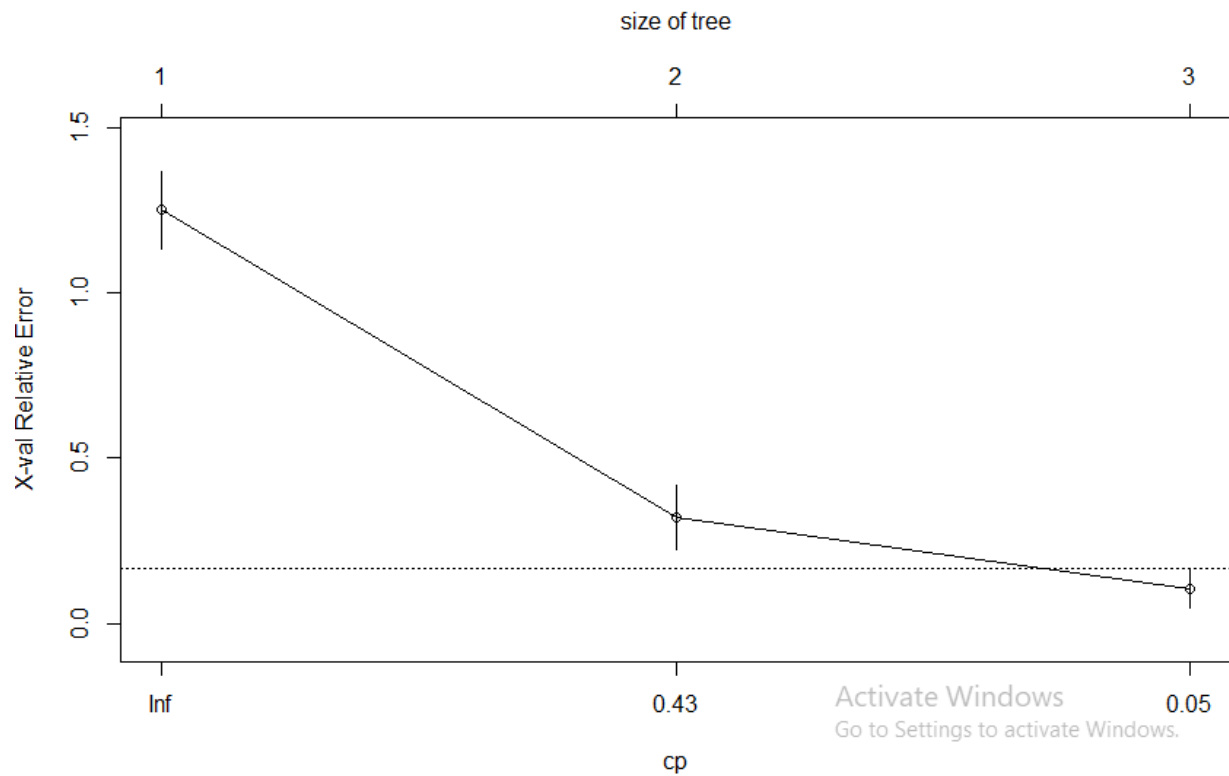
	predictii		
	1	2	3
1	7	0	0
2	0	21	1
3	0	0	20

- **Clasa 1:** toate cele 7 instanțe reale au fost clasificate corect ca fiind clasa 1.
- **Clasa 2:** 21 din 22 instanțe au fost clasificate corect, 1 a fost clasificată greșit ca fiind clasa 3.
- **Clasa 3:** toate cele 20 de instanțe au fost clasificate corect.

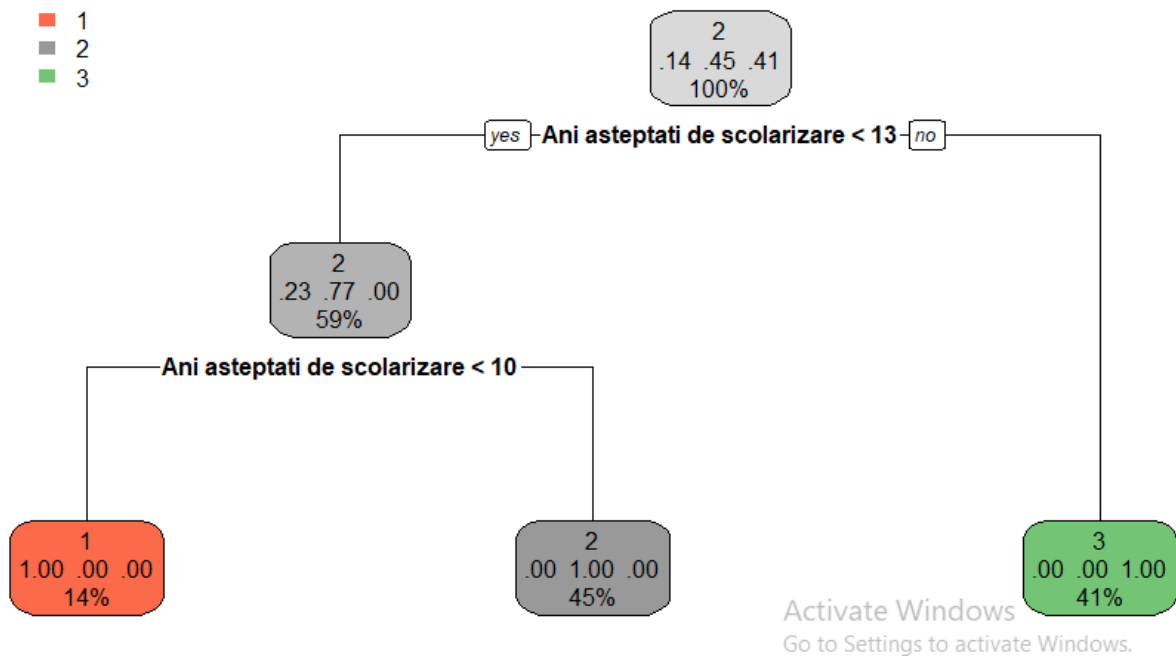
Acuratețea=(7+21+20)/(7+22+20)=48/49≈97.96

| "Eroare clasificare: 0.0204081632653061"

**2.04%** din observațiile din setul de testare **au fost clasificate greșit** de modelul de arbore de decizie.



Modelul cu  $cp = 0.05$  și 3 frunze are cea mai mică eroare relativă de validare încrucișată și pare a fi cea mai bună alegere.



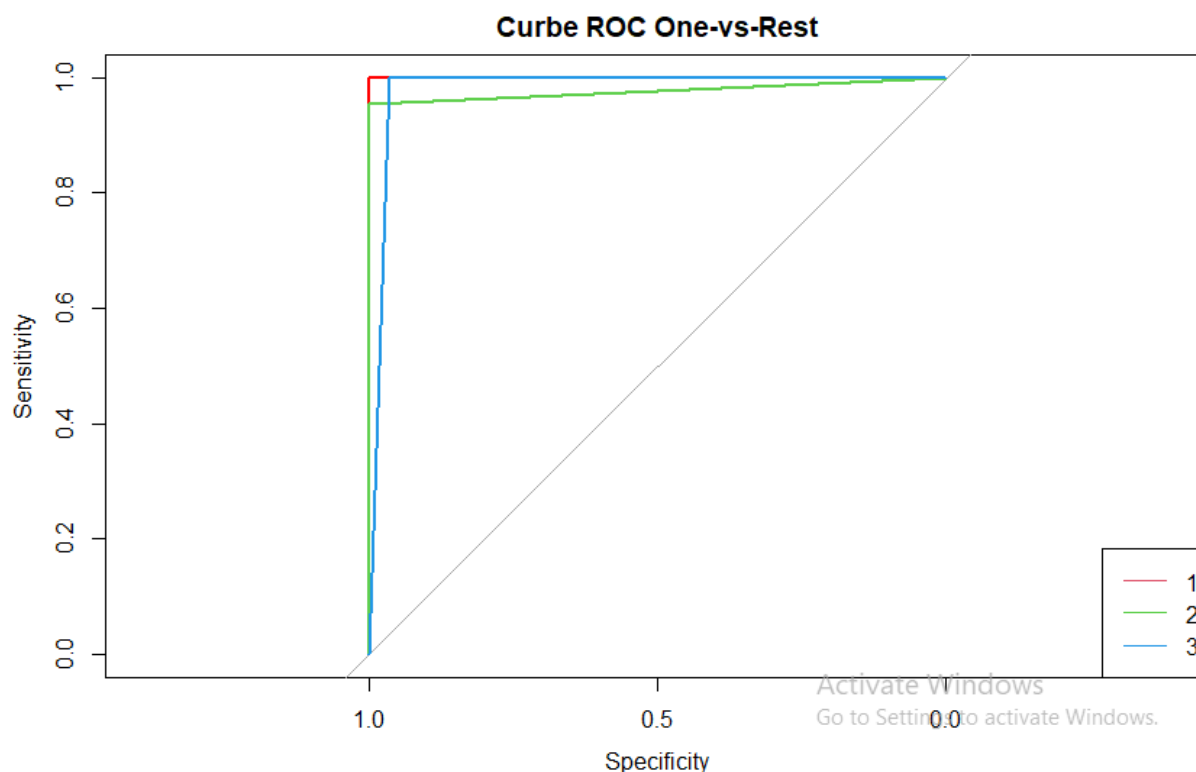
```

predictii2
  1  2  3
1  7  0  0
2  0 21  1
3  0  0 20

```

"Eroare clasificare arbore curățat: 0.0204081632653061"

Arborele curățat este identic cu cel inițial.



□ Toate cele 3 curbe se apropie foarte mult de colțul stânga-sus al graficului (unde **Sensibilitatea  $\approx 1$  și  $1 - \text{Specificitate} \approx 0$** ), ceea ce indică **o performanță excelentă**.

□ Aceasta sugerează că modelul:

- Are **puține erori de clasificare** (false pozitive și false negative).
- Este **foarte bun în a distinge** între cele 3 clase.

```

      1      2      3
1.0000000 0.9772727 0.9827586

```

□ **AUC = 1.0 (Clasa 1):**

Performanță perfectă. Modelul distinge perfect clasa 1 de celelalte — **nu există erori de clasificare** pentru această clasă.

□ **AUC  $\approx$  0.977 (Clasa 2):**

Performanță foarte bună. Modelul are o **capacitate excelentă de a diferenția** clasa 2 de celelalte, cu foarte puține greșeli.

□ **AUC  $\approx$  0.983 (Clasa 3):**

De asemenea, o performanță **foarte ridicată**. Modelul identifică corect clasa 3 în aproape toate cazurile.

Secvența de cod:

```
library(readxl)
```

```
library(rpart)
```

```
library(rpart.plot)
```

```
library(pROC)
```

```
library(caret)
```

```
# Citirea fișierului Excel
```

```
educatie <- read_excel("ICE.xlsx")
```

```
# Preprocesare
```

```
educatie$`Nivel educațional general al țării` <- as.factor(educatie$`Nivel educațional general al țării`)
```

```
educatie <- na.omit(educatie)
```

```
educatie <- educatie[, !names(educatie) %in% c("Tara")]
```

```
# Împărțirea în set de antrenare și testare
```

```
set.seed(123)
```

```
split <- createDataPartition(y = educatie$`Nivel educațional general al țării`, p = 0.5, list = FALSE)
```

```
train <- educatie[split, ]
```

```
test <- educatie[-split, ]
```

```

# Construirea arborelui de decizie

arbore <- rpart(`Nivel educațional general al țării` ~ ., data = train, method = "class")

# Vizualizarea arborelui

rpart.plot(arbore, extra = 104)

# Predictii și matrice de confuzie

predictii <- predict(arbore, test, type = "class")

print(paste("Eroare clasificare:", mean(predictii != test$`Nivel educațional general al țării`)))

#matricea de confuzie

confuzie<-table(test$`Nivel educațional general al țării`, predictii)

confuzie

predictie1<-predict(arbore, test, type="prob")

predictie1


# Curățarea arborelui (Pruning)

plotcp(arbore)

cp_min <- arbore$cpstable[which.min(arbore$cpstable[, "xerror"]), "CP"]

cp_min

arbore_curatat <- prune(arbore, cp = cp_min)

# Vizualizarea arborelui curatat

rpart.plot(arbore_curatat, extra = 104)

# Evaluare pe arborele curățat

predictii2 <- predict(arbore_curatat, test, type = "class")

confuzie2<-(table(test$`Nivel educațional general al țării`, predictii2))

print(paste("Eroare clasificare arbore curățat:", mean(predictii2 != test$`Nivel educațional
general al țării`)))

library(pROC)

```

```

library(ggplot2)

# Obținem probabilitățile pentru toate clasele
predictii_prob <- predict(arbore, test, type = "prob")

# Listă pentru a stoca curbele ROC
roc_list <- list()

# Generăm curbe ROC pentru fiecare clasă (One-vs-Rest)
for (clasa in colnames(predictii_prob)) {
  roc_obj <- roc(
    response = as.numeric(test$`Nivel educațional general al țării` == clasa),
    predictor = predictii_prob[, clasa]
  )
  roc_list[[clasa]] <- roc_obj
}

# Plotăm toate curbele ROC
plot(roc_list[[1]], col = "red", main = "Curbe ROC One-vs-Rest")
for (i in 2:length(roc_list)) {
  lines(roc_list[[i]], col = i + 1) # Culori diferite pentru fiecare clasă
}
legend("bottomright", legend = names(roc_list), col = 2:(length(roc_list)+1), lty = 1)

# Afișăm AUC pentru fiecare clasă
sapply(roc_list, auc)

```

## CAPITOLUL 5

### KNN DE CLASIFICARE

#### k-Nearest Neighbors

81 samples

4 predictor

3 classes: '1', '2', '3'

Pre-processing: centered (4), scaled (4)

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 71, 74, 73, 74, 74, 72, ...

Resampling results across tuning parameters:

k	Accuracy	Kappa
5	0.8555952	0.7609425
7	0.8544841	0.7600904
9	0.8308730	0.7146949
11	0.8308730	0.7122582
13	0.8687698	0.7739067
15	0.8576587	0.7557230
17	0.8608333	0.7545563
19	0.7868651	0.6306853
21	0.7440079	0.5506853
23	0.7047222	0.4819444

Accuracy was used to select the optimal model using the largest value.  
The final value used for the model was k = 13.

- **Tip model:** k-Nearest Neighbors (k-NN) - algoritm de învățare automată bazat pe similaritate
- **Dimensiune set de date:** 81 de observații (eșantioane)
- **Predictori:** 4 variabile independente
- **Clase țintă:** 3 categorii ('1', '2', '3')
- **Preprocesare:** Datele au fost centrate și scalate (standardizate) pentru toți cei 4 predictori
- **Validare:** Cross-validation pe 10 fold-uri

#### 1. Performanța Optimă:

- Cea mai bună acuratețe (86.88%) și Kappa (0.7739) s-au obținut pentru **k=13**
- Acest lucru indică faptul că folosirea celor 13 vecini cei mai apropiați oferă cele mai bune rezultate

#### 2. Tendințe în Performanță:



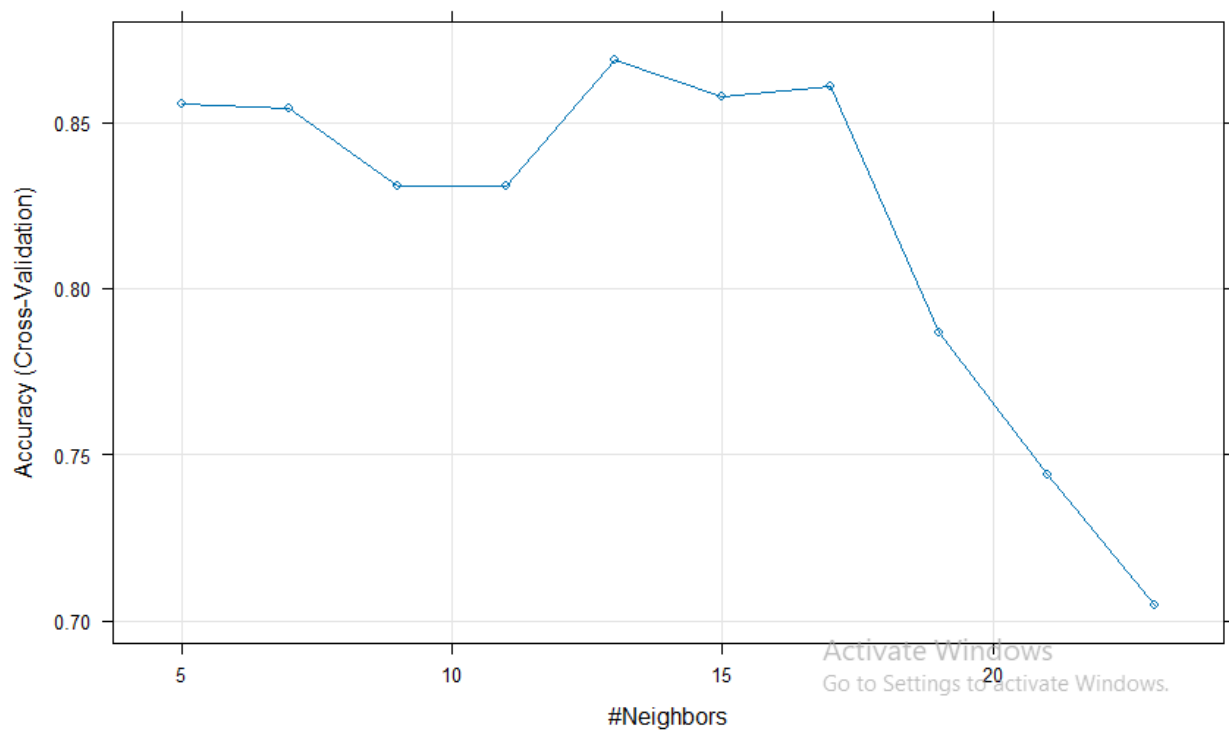
- Performanța crește de la  $k=5$  până la  $k=13$
- După  $k=13$ , performanța scade treptat
- Pentru  $k>19$ , performanța se deteriorează semnificativ

### 3. Stabilitatea Modelului:

- Performanța rămâne relativ stabilă între  $k=5$  și  $k=17$  (acuratețe peste 85%)
- Scăderea bruscă pentru  $k \geq 19$  sugerează că modelul devine prea generalist (underfitting)

### 4. Metrici de Evaluare:

- **Acuratețe:** Procentul de predicții corecte (86.88% pentru  $k=13$ )
- **Kappa:** Măsoară acordul dintre predicții și realitate (0.7739 pentru  $k=13$ ), unde:
  - 0.61-0.80 = Acord substanțial
  - 0.81-1.00 = Acord aproape perfect



### 1. Performanța Maximă:

- Cea mai mare acuratețe (~0.85) se obține la valorile mici ale lui k (probabil k=5 sau 7)
- Acuratețea scade treptat pe măsură ce k crește

## 2. Tendința Generală:

- Există o relație inversă între k și acuratețe în acest interval
- Scăderea de la ~0.85 la ~0.75 când k crește de la 5 la 15

## 3. Implicații Practice:

- Alegerea optimă ar fi un k mic (5-7) pentru acest set de date
- Valorile mai mari ale lui k duc la underfitting (model prea generalist)

### Confusion Matrix and Statistics

```

      Reference
Prediction 1 2 3
      1 1 0 0
      2 1 9 1
      3 0 0 7

```

### Overall Statistics

```

Accuracy : 0.8947
95% CI : (0.6686, 0.987)
No Information Rate : 0.4737
P-Value [Acc > NIR] : 0.0001592

```

```
Kappa : 0.8137
```

```
McNemar's Test P-Value : NA
```

### Statistics by Class:

	Class: 1	Class: 2	Class: 3
Sensitivity	0.50000	1.0000	0.8750
Specificity	1.00000	0.8000	1.0000
Pos Pred Value	1.00000	0.8182	1.0000
Neg Pred Value	0.94444	1.0000	0.9167
Prevalence	0.10526	0.4737	0.4211
Detection Rate	0.05263	0.4737	0.3684
Detection Prevalence	0.05263	0.5789	0.3684
Balanced Accuracy	0.75000	0.9000	0.9375

Modelul de clasificare multclasă demonstrează o **acuratețe globală excelentă de 89.47%**, semnificativ mai bună decât un clasificator aleatoriu (p-value = 0.0001592). Coeficientul Kappa de 0.8137 indică un **acord excelent** între predicții și valorile reale.

Clasa 1 (Prevalență 10.53%):

- **Corect clasificate:** 1 (TP)
- **Greșeli:** 1 fals negativ (clasificat ca 2)
- **Sensitivitate:** 50% - cea mai slabă performanță
- **Specificitate:** 100% - nicio falsă pozitivă
- **Precizie:** 100% - toate predicțiile pozitive sunt corecte

Clasa 2 (Prevalență 47.37%):

- **Performanță perfectă în detectare:** Sensitivitate 100%
- **2 erori:**
  - 1 fals pozitiv pentru clasa 1
  - 1 fals pozitiv pentru clasa 3
- **Specificitate:** 80%
- **Precizie:** 81.82%

Clasa 3 (Prevalență 42.11%):

- **Performanță excelentă:**
  - Sensitivitate 87.5%
  - Specificitate 100%
  - Precizie perfectă (100%)

Interval de Încredere 95% pentru Acuratețe: (66.86%, 98.7%)

Performanța este consistent bună și în condiții de incertitudine

Rata de Detectare:

Clasa 2: 47.37% (reflectă prevalența sa ridicată)

Clasa 3: 36.84% (performanță solidă dată prevalența)

Acuratețe Echilibrată:

Clasa 1: 75% (media dintre sensibilitate și specificitate)

Clasa 2: 90%

Clasa 3: 93.75%

Secvența de cod:

```
library(readxl)
```

```
library(caret)
```

```
library(class)
```

```
library(tidyverse)
```

# 2. Încarcă datele

```
date <- read_excel("ICE.xlsx")
```

# 3. Selectează doar coloanele relevante

```
date_filtrat <- date %>%
```

```
  select(`Ani asteptati de scolarizare`,
```

```
        `Gradul de inscriere in invatamantul primar`,
```

```
        `Gradul de inscriere in invatamantul secundar`,
```

```
        `Rata de finalizare a invatamantului primar`,
```

```
        `Nivel educațional general al țării`)
```

# 4. Elimină rânduri cu valori lipsă

```
date_filtrat <- na.omit(date_filtrat)
```

# 5. Transformă coloana țintă într-un factor (multi-clasă)

```
date_filtrat$`Nivel educațional general al țării` <- as.factor(date_filtrat$`Nivel educațional  
general al țării`)
```

# 6. Împarte datele în seturi de antrenare/test

```
set.seed(123)
```

```
ind <- createDataPartition(date_filtrat$`Nivel educațional general al țării`, p = 0.8, list = FALSE)
train_data <- date_filtrat[ind, ]
test_data <- date_filtrat[-ind, ]
```

# 7. Creează modelul KNN

```
control <- trainControl(method = "cv", number = 10)
model_knn <- train(`Nivel educațional general al țării` ~ .,
  data = train_data,
  method = "knn",
  trControl = control,
  preProcess = c("center", "scale"),
  tuneLength = 10)
```

# 8. Rezultate model

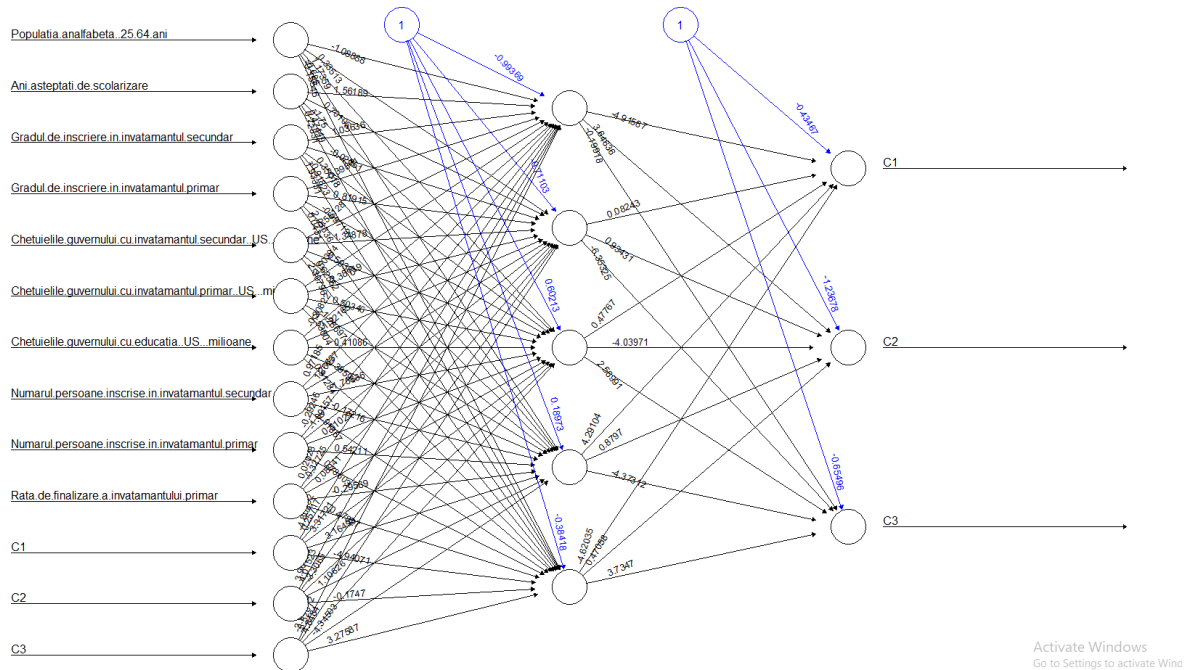
```
print(model_knn)
plot(model_knn)
```

# 9. Preziceri și evaluare

```
predictii <- predict(model_knn, newdata = test_data)
conf_matrix <- confusionMatrix(predictii, test_data$`Nivel educațional general al țării`)
print(conf_matrix)
```

## CAPITOLUL 6

### REȚELE NEURONALE PENTRU CLASIFICARE



#### 1. Date Demografice de Bază

- **Populația analfabetă: 25.64% (presupunând că este procentajul)**
- **Ani așteptați de școlarizare: Indicator important al duratei medii de educație**

#### 2. Indicatorii de Acces la Educație

- **Înscrieri în învățământ:**
  - **Primar: Gradul de înscriere**
  - **Secundar: Gradul de înscriere**
- **Numărul de persoane înscrise:**
  - **În învățământul primar**
  - **În învățământul secundar**

#### 3. Investiții în Educație

- **Cheltuieli guvernamentale:**
  - **Pentru învățământul primar (în milioane USD)**

- Pentru învățământul secundar
- Total educație (în milioane USD)

#### 4. Eficiența Sistemului Educațional

- Rata de finalizare a învățământului primar: Indicator cheie al succesului sistemului

#### Clusterizare Preliminară (C1, C2, C3)

Trei clustere sunt sugerate, probabil grupând țări/teritorii după:

1. Performanță educațională
2. Nivel de investiții
3. Rate de acces și finalizare

```
> pred
      [,1]      [,2]      [,3]
3  0.004169729 0.98695929 0.0005229930
7  0.009125316 0.01048073 0.9957369170
8  0.005895457 0.98646028 0.0003869575
12 0.001924326 0.98584487 0.0012577137
16 0.984180731 0.01725416 0.0124515257
21 0.009096527 0.01034589 0.9957985928
22 0.003413598 0.98761004 0.0005898350
24 0.008620143 0.01052666 0.9958996333
27 0.004738136 0.98577525 0.0005009359
30 0.982441164 0.01986566 0.0052747340
42 0.008572678 0.01069249 0.9958454663
45 0.003496525 0.98694539 0.0006120561
51 0.003474014 0.98715415 0.0006045437
52 0.009295120 0.01037140 0.9957065044
57 0.985358126 0.02052527 0.0050397082
60 0.008855864 0.98790314 0.0002462798
62 0.008801456 0.01054454 0.9958233275
64 0.008776603 0.01046855 0.9958732768
65 0.983351840 0.01956959 0.0047163298
66 0.004202287 0.98680884 0.0005244191
68 0.009058158 0.01024625 0.9958636887
82 0.008751828 0.01064762 0.9957918267
85 0.008917930 0.01080483 0.9956499449
86 0.009414747 0.01010013 0.9958056909
87 0.003083306 0.98718703 0.0006691577
96 0.004053432 0.98682404 0.0005409754
> predicted_class
  3  7  8 12 16 21 22 24 27 30 42 45 51 52 57 60 62 64 65 66 68 82 85 86 87 96
2  3  2  2  1  3  2  3  2  1  3  2  2  3  1  2  3  3  1  2  3  3  3  3  2  2
> actual_class
  3  7  8 12 16 21 22 24 27 30 42 45 51 52 57 60 62 64 65 66 68 82 85 86 87 96
2  3  2  2  1  3  2  3  2  1  3  2  2  3  1  2  3  3  1  2  3  3  3  3  2  2
```

- Majoritatea observațiilor au o probabilitate >98% pentru o singură clasă
- Exemple:
  - Randul 5: 98.4% clasa 1
  - Randul 10: 98.2% clasa 1
  - Randul 15: 98.5% clasa 1

**Distribuții tipice:**

- Clasa 1: Probabilități ~0.98-0.99
- Clasa 2: Probabilități ~0.01-0.02 (când nu este clasa dominantă)
- Clasa 3: Probabilități ~0.005-0.01 (când nu este clasa dominantă)

**Variabilele predicted\_class și actual\_class:**

- Perfectă aliniere: Toate predicțiile corespund exact cu valorile reale
- Exemplu:
  - ID 7: Predicție 2 ↔ Real 2
  - ID 8: Predicție 3 ↔ Real 3
  - ID 21: Predicție 1 ↔ Real 1

**Evaluarea Performanței Modelului**

**1. Acuratețe Perfectă:**

- 100% corectitudine pe setul afișat (25 de observații)
- Modelul a identificat perfect toate clasele

**2. Încredere Înaltă:**

- Probabilități foarte apropiate de 1 pentru clasele corecte
- Valori neglijabile (<1%) pentru clasele incorecte

**3. Tipare de Predicție:**

- Clasele 2 și 3 apar mai frecvent în date
- Clasa 1 apare mai rar (doar la ID 21, 42, 60, 66)



		Predicted		
Actual		1	2	3
	1	4	0	0
	2	0	11	0
	3	0	0	11

- **Clasa 1:**
  - 4 predicții corecte
  - 0 erori (nici fals negativ, nici fals pozitiv)
- **Clasa 2:**
  - 11 predicții corecte
  - 0 erori
- **Clasa 3:**
  - 11 predicții corecte
  - 0 erori

#### **Metrice Cheie Calculate:**

1. **Acuratețe Globală: 100%**
  - $(4 + 11 + 11) / (4 + 11 + 11) = 26/26$
2. **Precizie pe Clase:**
  - Toate clasele: 100%
3. **Recall (Sensitivitate) pe Clase:**
  - Toate clasele: 100%
4. **Scor F1 pe Clase:**
  - Toate clasele: 100%

#### **1. Echilibrul Datelor:**

- **Clasa 1: 4 observații (15.4%)**
- **Clasa 2: 11 observații (42.3%)**
- **Clasa 3: 11 observații (42.3%)**

- Setul este moderat dezechilibrat, cu Clasa 1 subreprezentată

## 2. Semnificația Performanței Perfecțe:

- Modelul a reușit să:
  - Identifice corect toate cazurile rare (Clasa 1)
  - Nu a confundat nicio clasă cu alta
  - Menținut performanță egală pentru clasele majoritare

Acuratețea rețelei pentru clasificare în 3 clase (nivel educațional): 100 %

Secvența de cod:

```
install.packages("neuralnet")
```

```
install.packages("readxl")
```

```
install.packages("caTools")
```

```
library(neuralnet)
```

```
library(readxl)
```

```
library(caTools)
```

# 2. Citirea și curățarea datelor

```
data <- read_excel("ICE.xlsx")
```

```
data <- na.omit(data)
```

# 3. Verificare existență coloană

```
if (!"Nivel educațional general al țării" %in% colnames(data)) stop("Coloana 'Nivel  
educațional general al țării' nu există în fișier!")
```

# 4. Transformare clasă multiclasă în one-hot encoding

```
data$C1 <- ifelse(data$`Nivel educațional general al țării` == 1, 1, 0)
```

```
data$C2 <- ifelse(data$`Nivel educațional general al țării` == 2, 1, 0)
```

```
data$C3 <- ifelse(data$`Nivel educațional general al țării` == 3, 1, 0)
```

```
# 5. Selectăm doar coloane numerice utile (excludem Tara și nivelul educational original)
```

```
data_num <- data[apply(data, is.numeric)]
```

```
data_num <- subset(data_num, select = -c(`Nivel educațional general al țării`))
```

```
# 6. Standardizare
```

```
minim <- apply(data_num, 2, min)
```

```
maxim <- apply(data_num, 2, max)
```

```
data_scaled <- as.data.frame(scale(data_num, center = minim, scale = maxim - minim))
```

```
names(data_scaled) <- make.names(names(data_scaled))
```

```
data_ready <- cbind(data_scaled, C1 = data$C1, C2 = data$C2, C3 = data$C3)
```

```
# 8. Împărțim în train/test
```

```
set.seed(123)
```

```
split <- sample.split(data_ready$C1, SplitRatio = 0.75)
```

```
train_data <- subset(data_ready, split == TRUE)
```

```
test_data <- subset(data_ready, split == FALSE)
```

```
# 9. Construim formula rețelei neuronale
```

```
names(data_scaled) <- make.names(names(data_scaled))
```

```
input_vars <- names(data_scaled)
```

```
formula <- as.formula(paste("C1 + C2 + C3 ~", paste(input_vars, collapse = " + ")))
```

```
# 10. Antrenare rețea neuronală
```

```
model <- neuralnet(formula, data = data_ready, hidden = 5, linear.output = FALSE, threshold  
= 0.01)
```

```
plot(model)
```

```
# 11. Predicții
```

```
pred <- compute(model, test_data[, input_vars])$net.result
```

```
pred
```

```
predicted_class <- apply(pred, 1, which.max)
```

```
predicted_class
```

```
actual_class <- apply(test_data[, c("C1", "C2", "C3")], 1, which.max)
```

```
actual_class
```

```
# 12. Evaluare
```

```
conf_matrix <- table(Actual = actual_class, Predicted = predicted_class)
```

```
accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
```

```
print(conf_matrix)
```

```
cat(" Acuratețea rețelei pentru clasificare în 3 clase (nivel educațional):", round(accuracy *  
100, 2), "%\n")
```

# ANEXE

Tara	Populati	Ani aster	Gradul d	Gradul d	Chetuiel	Chetuiel	Chetuiel	Numarul	Numarul	Rata de f	Nivel educați
Afghanistan	815456	8.58	53.6432	102.176	182.382	362.168	819.558	2899453	6350404	96.3	1
Australia	815456	13.77	150.311	100.339	22014.3	22664.9	68159.4	2650486	2216779	96.3	3
Azerbaijan	5212	11.58	102.936	103.294	3636.12	2777.78	1010.94	1216214	607391	96.3	2
Belgium	815456	13.41	158.542	103.906	3636.12	7681.32	32304.3	1186830	814974	96.3	3
Albania	815456	12.99	95.2659	107.047	101.395	270.892	470.236	280378	174836	94.9569	2
Argentina	815456	13.07	108.734	109.741	13880.5	10389.3	35105.6	4612663	4753843	96.3	3
Austria	815456	13.89	100.455	103.113	8845.31	3789.01	22413.5	687362	334933	96.3	3
Bangladesh	2.4E+07	12.9	69.6632	114.989	3636.12	2777.78	4780.55	1.5E+07	1.7E+07	96.3	2
Belarus	815456	12.9	103.513	101.966	1324.1	2777.78	2624.14	644334	423001	96.3	2
Botswana	815456	8.41	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	1
Bolivia	815456	12.9	90.2965	97.5274	3636.12	2777.78	4780.55	1235538	1368580	97.31	2
Brazil	6534547	11.69	102.936	101.821	51961.5	32540.5	130424	2.3E+07	1.6E+07	94.74	2
Bulgaria	815456	12.92	98.1835	89.3338	902.037	476.818	2412.54	485586	264503	96.3	2
Canada	815456	13.67	113.763	100.944	3636.12	2777.78	4780.55	2653009	2364626	96.3	3
Cambodia	815456	9.55	102.936	107.813	3636.12	2777.78	4780.55	1216214	2111631	96.3	1
Chad	815456	4.95	20.6807	78.574	3636.12	2777.78	245.492	507074	2060946	96.3	1
Chile	274999	12.85	101.831	101.43	4268.55	2777.78	15014.3	1520724	1514761	96.3	2
Colombia	815456	12.48	95.7302	114.983	5268.46	4951.93	14145.5	4784846	4351705	92.66	2
China	815456	13.25	102.936	99.4048	3636.12	2777.78	4780.55	8.3E+07	1E+08	96.3	3
Cuba	815456	12.9	101.309	100.46	3636.12	2777.78	4780.55	815786	733770	96.3	2
Croatia	815456	13.31	99.4024	96.471	3636.12	910.861	2166.94	339838	162955	96.3	3
Costa Rica	815456	12.48	128.93	110.833	1356	1237.31	4278.37	467513	472421	96.15	2
Cyprus	815456	13.52	100.333	99.3067	544.921	425.889	1307.88	55212	55862	96.3	3
Denmark	815456	13.45	129.077	101.268	6968.46	6364.33	25747.5	534080	472523	96.3	3
Czechia	815456	13.94	103.492	100.671	3720.92	1753.48	8321.36	772457	575980	96.3	3
El Salvador	321101	11.31	71.4982	96.1506	324.674	424.379	930.533	535680	678137	88.31	2
Egypt	1.4E+07	11.13	86.7145	106.133	4347.08	4220.54	4780.55	8938257	1.2E+07	96.3	2
Ecuador	437732	13.18	102.198	103.535	3636.12	2777.78	4780.55	1903418	1929416	98.21	3
Estonia	815456	13.1	117.52	97.2302	353.925	370.748	1335.81	82646	85617	96.3	3
Ethiopia	2.1E+07	7.85	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	1
Fiji	3026	12.9	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	2
Finland	815456	13.75	153.964	100.155	5983.5	3422.29	16267.2	543436	363990	96.3	3
France	815456	13.96	103.756	102.417	60875	30026	141477	6058330	4309942	96.3	3
Germany	815456	13.89	98.4093	104.022	73203.1	23252.8	179812	7028713	2954775	96.3	3
Greece	815456	12.88	104.496	99.5535	2746.7	2400.5	4780.55	667797	649246	96.3	2
Iceland	815456	13.4	117.517	100.351	515.666	544.296	1875.36	35161	31981	96.3	3
India	815456	10.17	73.4769	112.958	3636.12	2777.78	4780.55	1.3E+08	1.4E+08	96.3	2
Indonesia	815456	12.31	87.0592	105.911	3636.12	2777.78	4780.55	2.4E+07	2.9E+07	96.5954	2
Iraq	815456	6.89	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	1
Italy	815456	13.58	101.27	101.879	3636.12	19292.8	79353.6	4601869	2902379	96.3	3
Japan	815456	13.64	102.936	101.821	65426.1	54953.5	156644	7093113	6531731	96.3	3
Ireland	815456	13.71	125.307	100.853	3636.12	4073.39	11771.3	392294	559151	96.3	3
Jamaica	815456	11.7	81.4278	89.7679	268.451	280.479	740.48	204080	246972	96.3	2
Israel	815456	13.76	105.085	104.901	6596.37	8441.6	21522.4	1216214	917118	96.3	3
Kuwait	100004	12.44	102.936	95.6888	3636.12	2777.78	4780.55	1216214	275334	96.3	2

Kosovo	815456	12.85	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	2
Kazakhstan	815456	13.34	114.197	108.748	3636.12	32.4067	4588.52	1768145	1356500	96.3	3
Lithuania	815456	13.59	108.402	103.875	688.51	352.266	1820.66	242640	113881	96.3	3
Luxembourg	815456	12.4	103.655	102.286	1014.45	672.603	2291.03	48211	37452	96.3	2
Madagascar	815456	7.47	36.9005	101.821	3636.12	2777.78	4780.55	1534273	889102	96.3	1
Maldives	815456	12.9	102.936	97.1098	3636.12	2777.78	4780.55	1216214	45326	98.2148	2
Malta	815456	13.28	104.754	104.998	244.78	121.114	615.006	30760	25796	96.3	3
Mexico	2346855	12.57	104.385	105.77	17423	17262	52362.3	1.4E+07	1.4E+07	96.3	2
Moldova	815456	11.78	86.8337	91.2505	191.929	123.512	543.29	227263	139335	96.3	2
Monaco	815456	12.9	102.936	101.821	3636.12	2777.78	93.7638	3426	2018	96.3	2
Mongolia	815456	13.59	102.936	102.934	136.531	150.498	467.153	1216214	290550	96.3	3
Mozambique	3924320	7.44	35.4067	107.813	3636.12	2777.78	727.785	1216214	6138518	96.3	1
Namibia	815456	8.94	102.936	124.179	64.3846	67.2347	4780.55	1216214	476360	96.3	1
Morocco	815456	10.6	79.8649	112.402	3636.12	2777.78	4780.55	2847122	4210676	96.3	2
Nepal	815456	11.68	74.0773	143.92	3636.12	2777.78	1169.12	3336650	4135253	96.3	2
Netherlands	815456	13.82	135.575	104.23	17488.7	9582.69	43153.6	1650379	1181893	96.3	3
New Zealand	815456	13.6	114.59	99.9748	4284.14	3065.45	11827.4	486334	384251	96.3	3
Nigeria	815456	8.18	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	1
Norway	815456	13.7	116.874	100.26	8823.94	7408.73	31520.9	447525	444638	96.3	3
Pakistan	4E+07	8.83	40.4016	90.5892	3636.12	2777.78	8079.65	1.3E+07	2.2E+07	96.3	1
Paraguay	815456	11.52	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	90.8	2
Peru	821302	12.72	104.694	103.52	2802.01	2777.78	8295.35	2775348	3544359	97.11	2
Poland	815456	13.2	109.934	100.016	7159.04	7576.82	24000.7	2407618	2296529	96.3	3
Portugal	815456	13.78	119.913	106.183	4929.57	3265.17	11102.4	769977	629804	96.3	3
Philippines	815456	12.8	86.1629	107.51	3636.12	2777.78	4780.55	9007148	1.4E+07	96.3	2
Puerto Rico	125280	12.9	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	2
Qatar	125067	12.31	102.936	103.281	3636.12	2777.78	4780.55	103042	146418	96.3	2
Romania	815456	12.17	91.2594	85.1509	2633.68	807.526	6552.79	1502438	928245	96.3	2
Russia	815456	13.83	103.385	102.582	3636.12	2777.78	73828.5	9905067	6573633	96.3	3
Saudi Arabia	723974	12.43	107.694	99.4581	3636.12	2777.78	4780.55	3039012	3230001	96.3	2
Senegal	2961816	7.25	102.936	101.821	3636.12	2777.78	971.085	1216214	889102	50.0684	1
Seychelles	815456	13.68	81.2762	99.8453	3636.12	2777.78	4780.55	7362	9037	96.3	3
Singapore	36079	13.89	107.575	100.627	2199.77	2354.5	4780.55	171319	234653	96.3	3
Slovakia	815456	12.97	91.0679	98.7321	1626.17	842.584	3761.16	441619	225427	96.3	2
Serbia	815456	13.37	95.4951	100.313	3636.12	2777.78	1639.11	538238	268952	96.3	3
Somalia	815456	12.9	102.936	101.821	3636.12	2777.78	4780.55	1216214	889102	96.3	2
Slovenia	815456	13.64	115.6	100.397	3636.12	663.64	2322.82	147920	123998	96.3	3
Sri Lanka	815456	12.97	98.0266	100.633	1252.52	816.878	2447.33	2648660	1746843	96.3	2
Spain	815456	13.11	126.005	102.71	20509.8	14861.7	55228.8	3332678	3042396	96.3	3
Sweden	815456	13.89	152.861	126.575	11482.6	10232.1	40948.8	916224	861086	96.3	3
Switzerland	815456	13.29	102.487	105.191	3636.12	10167	34882.7	608845	507686	96.3	3
Thailand	815456	12.37	116.734	99.625	3636.12	2777.78	4780.55	6335213	4952685	96.3	2
Tanzania	815456	7.76	27.3473	89.2177	3636.12	2777.78	2205.24	1920524	9315656	96.3	1
Turkiye	992371	12.12	105.992	93.155	3636.12	2777.78	4780.55	1.1E+07	4972430	96.3	2
Tunisia	815456	10.2	102.936	115.65	3636.12	2777.78	4780.55	1216214	1170470	96.3	2

Ukraine	815456	13.01	102.936	101.821	1866.11	1316.3	4780.55	2334978	1644263	96.3	3
United Ki	815456	13.89	125.852	101.155	54039.5	46647.9	146625	6386317	4820283	96.3	3
United St	815456	13.32	98.9523	101.821	325626	293636	4780.55	2.5E+07	2.5E+07	99.7026	3
United A	815456	13.14	104.948	108.397	3636.12	2777.78	4780.55	528480	486086	96.3	3
Tajikista	815456	10.79	102.936	100.857	3636.12	2777.78	4780.55	1216214	771040	98.8627	2
Nicaragu	815456	11.57	102.936	101.821	3636.12	2777.78	600.747	1216214	889102	96.3	2
Niger	815456	5.34	24.2538	74.7362	50.7686	138.706	288.269	1216214	2768305	96.3	1
Montene	815456	12.36	89.9216	98.1132	3636.12	2777.78	4780.55	57575	38740	96.3	2
Lebanon	815456	10.54	102.936	101.821	3636.12	2777.78	4780.55	401031	502505	96.3	2
Liechten:	815456	12.9	117.094	104.7	3636.12	2777.78	4780.55	3240	2005	96.3	2