#### UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI" IAȘI FACULTATEA AUTOMATICĂ ȘI CALCULATOARE DOMENIUL: CALCULATOARE ȘI TEHNOLOGIA INFORMAȚIEI PROFIL DE STUDIU: TEHNOLOGIA INFORMAȚIEI

# PROIECT INTELIGENȚĂ ARTIFICIALĂ

Rezolvarea problemei de optimizare a unei mașini cu vectori supo	ort cu
ajutorul unui algoritm evolutiv	

#### Studenți:

Lungu Ștefania-Paraschiva Poleac Alexandra-Cătălina

An universitar 2022-2023

### **CUPRINS**

Capitolul 1. Descrierea problemei considerate Capitolul 2. Aspecte teoretice privind algoritmul	3 4
1. Cuantificarea unui set de date	5
2. Funcția de fitness	7
<ol><li>Determinarea soluției problemei – determinarea parametrilor alpha</li></ol>	7
4. Funcția de calcul pentru nucleul polinomial de grad II	8
5. Determinarea claselor	8
Capitolul 5. Concluzii	10
Capitolul 6. Bibliografie	10
Capitolul 7. Rolul fiecărui membru al echipei	11

# Capitolul 1. Descrierea problemei considerate

Problema de bază a proiectului este clasificarea familiilor dintr-o grădiniță care se împarte în 4 categorii: *not\_recommend*, *recommend*, *very\_recommend*, *priority*. Fiecare instanță conține un număr de atribute prin intermediul cărora se poate face clasificarea:

parents: ocupația părințilorhas nurs: grădinița copilului

- form: forma familiei

children: numărul de copii
housing: starea locuinței
finance: starea financiară
social: statutul social
health: starea de sănătate

Setul de date utilizat este cel din UCI Repository:

https://archive.ics.uci.edu/ml/datasets/Nursery?fbclid=lwAR0kX5DJ-OHfStlAkplc4kBjlNTo4ZnBfAKDLKjX-5DKwrF9CiWdR0oEb-s

Soluția își propun utilizarea unui algoritm evolutiv cu ajutorul căruia să se determine parametrii SVM-ului (alpha, bias și w) necesari în procesul de clasificare.

# Capitolul 2. Aspecte teoretice privind algoritmul

Problema duală a SVM presupune maximizarea a W(alfa) cu ajutorul algoritmului evolutiv. Trebuie să determinăm dreapta care maximizează marginea ce desparte cele două clase de obiecte, adica vectorii suport. Vectorii suport sunt o mulțime de puncte într-un spațiu n dimensional care sprijina "marginea".

Scopul acestui algoritm este de a determina valorile suport care maximizează marginea.

Vom folosi algoritmul evolutiv varianta standard prezentata la curs și la laborator.

Problema duală a SVM:

$$\max_{\alpha_i} \left( \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \right)$$

cu următoarele constrângeri:

$$\alpha_i > 0, i = 1..N$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0.$$

Am folosit nucleul polinomial de gradul 2:

$$K(x, z) = (x \cdot z + 1)^2$$

Genele din cromozomi sunt valorile alpha-urilor, iar funcția de fitness este următoarea:

$$W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

Apoi din acești alpha vom calcula parametrii w si b:

$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

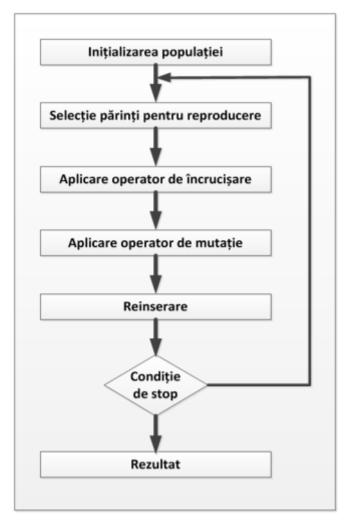
$$b = \frac{1}{|S|} \sum_{s \in S} \left( y_s - \sum_{t \in S} \alpha_t y_t(\mathbf{x}_t \cdot \mathbf{x}_s) \right)$$

unde S este multimea vectorilor suport iar |S| este numărul lor

De asemenea, am folosit ca metoda de optimizare algoritmul evolutiv asupra populației de date extrase din fișier, asupra cărora se aplică transformările specifice evoluției naturale:

- Selecția prin care determinăm părinții ce se vor reproduce pentru a forma următoarea generație. Se vor alege apoi indivizii cei mai adaptați (care se aproprie cel mai mult de soluția problemei).
- Încrucișarea prin care se generează un copil pornind de la doi părinți.
- Mutația ce se folosește pentru a asigura diversitatea populației. Asupra noilor copii generați se aplica transformări aleatorii, astfel se permite apariția unor noi trăsături care nu ar fi apărut în cadrul populației prin selecție și încrucișare.

Structura algoritmului evolutiv este:



# Capitolul 3. Modalitatea de rezolvare

#### 1. Cuantificarea unui set de date

```
internal class DataManipulation
     public string[] lines = File.ReadAllLines(@"C:\Users\STEFANIA\Desktop\Facultate\IA\PROIECT 2023\nursery.data");
     public string[] Innes = File.ReadAllLines(@*C:\User:
public int[,] valuesClasified;
public int[,] valuesNonRecommend = new int[37, 8];
public int[,] valuesRecommend = new int[37, 8];
public int[,] valuesVeryRecommend = new int[22, 8];
public int[,] valuesPriority = new int[46, 8];
int NonRecomLines = 0;
      int RecomLines = 0;
     int VervRecomLines = 0:
      int PriorityLines = 0;
     int noInstances;
     public void fileRead()
            //string[] lines = File.ReadAllLines(@"C:\Users\STEFANIA\Desktop\Facultate\IA\PROIECT 2023\nursery.data");
            valuesClasified = new int[lines.Length. 9]:
           valuesclastried = new int[Innes.Lengtn, 9];
//int[,] valuesNonRecommend = new int[37, 8];
//int[,] valuesRecommend = new int[37, 8];
//int[,] valuesVeryRecommend = new int[22, 8];
//int[,] valuesPriority = new int[46, 8];
            noInstances = lines.Length;
            for (int i = 0; i < lines.Length; i++)
                  string[] values = lines[i].Split(',');
                  switch (values[0]) // parents = ocupatia parintilor
                         case "usual":
                              valuesClasified[i, 0] = 1;
                               break;
                        case "pretentious":
   valuesClasified[i, 0] = 2;
                        case "great_pret":
   valuesClasified[i, 0] = 3;
                         default:
                              break;
```

```
// Clasificare date in mai multi vectori in functie de ultimul parametru (very-recom, recom, non-recom, priority) for (int k = 0; k < lines.Length - 1; k++)
       switch (valuesClasified[k, 8])
           case -1:
               for (int j = 0; j < 7; j++)
                  valuesNonRecommend[NonRecomLines, j] = valuesClasified[k, j];
//Console.WriteLine("NR " + valuesNonRecommend[k, j] + " ");
               break;
           case 0:
               for (int j = 0; j < 7; j++)
                    \begin{tabular}{ll} values Recommend [Recomlines, j] = values Clasified [k, j]; \\ //Console.WriteLine ("R " + values Recommend [k, j] + " "); \\ \end{tabular} 
               Recomlines++:
               break;
           case 1:
               for (int j = 0; j < 7; j++)
                   valuesVeryRecommend[VeryRecomLines, j] = valuesClasified[k, j];
                   //Console.WriteLine("VR
                                             " + valuesVeryRecommend[k, j] + "
               VeryRecomLines++;
           break;
case 2:
               for (int j = 0; j < 7; j++)
                   valuesPriority[PriorityLines, j] = valuesClasified[k, j];
//Console.WriteLine("P " + valuesPriority[k, j] + " ");
               PriorityLines++;
               break;
           default:
      } //end switch
using (StreamWriter sw = new StreamWriter(@"C:\Users\STEFANIA\Desktop\Facultate\IA\PROIECT 2023\outDataNonRecomm.data"))
     sw.WriteLine("Populatia de date din categoria non-recommended.");
     sw.WriteLine("Numarul de populatii non-recomandate este " + NonRecomLines);
     for (int t = 0; t < valuesNonRecommend.GetLength(0); <math>t++)
           for (int s = 0; s < valuesNonRecommend.GetLength(1); s++)</pre>
                sw.Write(valuesNonRecommend[t, s] + " ");
          sw.WriteLine();
} //end using
using (StreamWriter sw = new StreamWriter(@"C:\Users\STEFANIA\Desktop\Facultate\IA\PROIECT 2023\outDataRecomm.data"))
     sw.WriteLine("Populatia de date din categoria recommended.");
     sw.WriteLine("Numarul de populatii recomandate este " + NonRecomLines);
for (int t = 0; t < valuesRecommend.GetLength(0); t++)</pre>
           for (int s = 0; s < valuesRecommend.GetLength(1); s++)</pre>
                sw.Write(valuesRecommend[t, s] + " ");
           sw.WriteLine();
     }
} //end using
```

#### 2. Funcția de fitness

#### 3. Determinarea soluției problemei

```
public Chromosome Solve(IOptimizationProblem p, int populationSize, int maxGenerations, double crossoverRate, double mutationRate, int[,] inputValues)
    // throw new Exception("Aceasta metoda trebuie completata");
    Chromosome[] population = new Chromosome[populationSize];
    for (int i = 0; i < population.Length; i++)
         population[i] = p.MakeChromosome();
p.ComputeFitness(population[i], inputValues);
    for (int gen = 0; gen < maxGenerations; gen++)
         Chromosome[] newPopulation = new Chromosome[populationSize];
newPopulation[0] = Selection.GetBest(population); // elitism
         for (int i = 1; i < populationSize; i++)
              // selectare 2 parinti: Selection.Tournament
              Chromosome mama = Selection.Tournament(population);
Chromosome tata = Selection.Tournament(population);
              while (mama == tata)
                   mama = Selection.Tournament(population);
              // generarea unui copil prin aplicare crossover: Crossover.Arithmetic
Chromosome copil = Crossover.Arithmetic(mama, tata, crossoverRate);
              // aplicare mutatie asupra copilului: Mutation.Reset
              Mutation.Reset(copil, crossoverRate);
              // calculare fitness pentru copil: ComputeFitness din problema p
              p.ComputeFitness(copil, inputValues);
               // introducere copil in newPopulation
              newPopulation[i] = copil;
         for (int i = 0; i < populationSize; i++)
    population[i] = newPopulation[i];</pre>
    return Selection.GetBest(population);
```

#### 4. Funcția de calcul pentru nucleul polinomial de grad II

#### 5. Determinarea parametrilor SVM-ului - alfa, w (weights) și bias

```
EvolutionaryAlgorithm ea = new EvolutionaryAlgorithm();
Chromosome solution = ea.Solve(new Equation(), 120, 100, 0.95, 0.2, valuesClasified);
int nrInstante| = 120;
double[] alfa = new double[nrInstante]; //nr de instante = 120
Console.WriteLine("Coeficientii alfa sunt ");
for(int i = 0; i< nrInstante; i++)</pre>
    alfa[i] = solution.Genes[i];
Console.WriteLine("Pe linia " + i + " avem coeficientul " + alfa[i] + " ");
Console.WriteLine("-----");
double w = 0;
for (int i = 0; i < nrInstante; ++i)</pre>
   w += alfa[i] * valuesClasified[i, 8];
Console.WriteLine("Coeficientul w este " + w);
Console.WriteLine("
int[] xi = new int[8]; // 8 atribute (fara ultimul)
int[] xj = new int[8]; // vectori in care vom avea valorile din inputValues pentru a le transmite ca paramtru in functia calculata de Kernel
double bias = 0;
double s2 = 0;
Kernel kernel = new Kernel();
for (int i = 0; i < nrInstante; i++)
     for (int j = 0; j < nrInstante; j++)
         for (int t = 0; t < 8; t++)
             //c.Genes = alpha, inputValues = y conform formulei matematice din Cursul 12 pag 50
s2 += valuesClasified[j, 8] * alfa[j] * kernel.PolinomialKernel(xi, xj);
    bias += valuesClasified[i, 8] - s2;
bias /= nrInstante;
Console.WriteLine("Bias este " + bias);
```

#### 6. Determinarea sub-claselor / împărțirea vectorului pe categorii

# Capitolul 4. Rezultatele obținute prin rularea programului

Rezultatele obținute în urma rulării script-ului de clasificare se pot observa în imaginea de mai jos:

```
ocentajul pentru gradinitele nerecomandate: 30%.
ocentajul pentru gradinitele recomandate: 11%.
ocentajul pentru gradinitele foarte recomandate: 18%.
ocentajul pentru gradinitele prioritare: 38%.
Procentajul pentru gradinitele foarte recomandate Procentajul pentru gradinitele prioritare: 38%.

Deficientii alfa sunt
e linia 0 avem coeficientul 114,83578099304563
el linia 1 avem coeficientul 113,0997835734393
el linia 2 avem coeficientul 103,0997835734393
el linia 2 avem coeficientul 112,80405052995455
el linia 3 avem coeficientul 1112,80405052995455
el linia 4 avem coeficientul 113,437421747834367
el linia 6 avem coeficientul 113,437421747834367
el linia 6 avem coeficientul 189,14700315546815
el linia 10 avem coeficientul 189,14700315546815
el linia 8 avem coeficientul 7,942085357707893
el linia 9 avem coeficientul 199,20286211808084
el linia 10 avem coeficientul 162,8407666002006
el linia 11 avem coeficientul 162,8407666002006
el linia 13 avem coeficientul 199,3831570255384
el linia 13 avem coeficientul 199,331316735198456
el linia 15 avem coeficientul 199,3831570255384
el linia 16 avem coeficientul 199,383136735198456
el linia 17 avem coeficientul 16,733422186069636
el linia 17 avem coeficientul 18,734322186069636
el linia 19 avem coeficientul 11,79327888041555
el linia 20 avem coeficientul 11,79327888041555
el linia 20 avem coeficientul 11,2937888041555
el linia 20 avem coeficientul 11,2937888041555
el linia 20 avem coeficientul 17,2937888041555
el linia 20 avem coeficientul 17,2937888041555
el linia 20 avem coeficientul 18,80666657742849
el linia 21 avem coeficientul 18,80666657742849
el linia 23 avem coeficientul 17,293757689319
el linia 24 avem coeficientul 17,293757689319
el linia 25 avem coeficientul 17,293757689319
el linia 27 avem coeficientul 18,81140996582834
el linia 29 avem coeficientul 18,9363879904951
el linia 20 avem coeficientul 18,9363879904558284
el linia 31 avem coeficientul 18,9363879904558284
el linia 32 avem coeficientul 18,937497355514
el linia 33 avem coeficientul 18,937497355318
```

```
Microsoft Visual Studio Debug Console

linia 78 avem coeficientul 27,5491926142124

linia 79 avem coeficientul 77,80961928310452

linia 80 avem coeficientul 83,0785787103363

linia 81 avem coeficientul 89,1549841730251

linia 82 avem coeficientul 189,1549841730251

linia 83 avem coeficientul 116,68410971092707

linia 83 avem coeficientul 116,68410971092707

linia 85 avem coeficientul 18,32746092196049

linia 86 avem coeficientul 42,22493415487612

linia 88 avem coeficientul 42,22493415487612

linia 88 avem coeficientul 48,013897767681854

linia 89 avem coeficientul 18,3897767681854

linia 90 avem coeficientul 17,37504226141056

linia 91 avem coeficientul 178,37504226141056

linia 93 avem coeficientul 27,118801574953605

linia 94 avem coeficientul 36,199854023964

linia 95 avem coeficientul 36,1998854023964

linia 96 avem coeficientul 176,95204158242327

linia 97 avem coeficientul 178,95204158242327

linia 98 avem coeficientul 178,95204158242327

linia 98 avem coeficientul 37,95204158242327

linia 98 avem coeficientul 38,988060274745

linia 100 avem coeficientul 38,848141271599775

linia 101 avem coeficientul 38,848141271599775

linia 101 avem coeficientul 38,848141271599775
         linia 98 avem coeficientul 48,9209624966808
linia 90 avem coeficientul 49,0209624966808
linia 100 avem coeficientul 33,48141271599775
linia 101 avem coeficientul 33,28344864429085
linia 102 avem coeficientul 28,490360162445807
linia 103 avem coeficientul 28,490360162445807
linia 104 avem coeficientul 19,31167089876562
linia 105 avem coeficientul 19,31167089876562
linia 107 avem coeficientul 117,67682760799255
linia 108 avem coeficientul 117,67682760799255
linia 108 avem coeficientul 114,587883006048191
linia 110 avem coeficientul 42,88942601642307
linia 111 avem coeficientul 42,88942601642307
linia 111 avem coeficientul 33,88012934996011
linia 113 avem coeficientul 34,98806167596051
linia 115 avem coeficientul 34,9776498836956
linia 116 avem coeficientul 38,67845239867808
linia 117 avem coeficientul 19,4980616395687328
linia 118 avem coeficientul 19,4036693304472
linia 119 avem coeficientul 19,4036693304472
linia 119 avem coeficientul 29,4036693304472
peficientul w este 5017,801730930123
```

#### Fișierele generate:

outDataNonRecomm.data

#### - outDataPriority.data

```
Populatia de date din categoria priority.
Numarul de populatii recomandate este 46
    1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 0 \cdot 0 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 0 \cdot 0
     1.1.1.1.1.1.2.0
    \begin{array}{c} 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 2 \cdot 0 \cdot 0 \\ 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 0 \cdot 0 \\ 2 \cdot 1 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 2 \cdot 0 \end{array}
     1 · 1 · 1 · 1 · 2 · 1 · 2 · 0
    3 · 2 · 4 · 1 · 1 · 1 · 3 · 0 · 3 · 2 · 4 · 1 · 1 · 1 · 3 · 0
     1 · 1 · 1 · 1 · 2 · 1 · 3 · 0
    1 \cdot 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 3 \cdot 0
1 \cdot 1 \cdot 1 \cdot 1 \cdot 2 \cdot 2 \cdot 0 \cdot 0
     1 · 1 · 1 · 2 · 1 · 1 · 0 · 0
    1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 0
1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 3 \cdot 0
     1.1.1.2.1.1.3.0
    \begin{array}{c} 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 0 \cdot 0 \\ 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 2 \cdot 0 \cdot 0 \\ 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 0 \end{array}
    1 · 1 · 2 · 2 · 1 · 3 · 0 · 2 · 1 · 2 · 2 · 1 · 3 · 0 · 1 · 1 · 1 · 2 · 2 · 2 · 1 · 3 · 0 · 1 · 1 · 1 · 2 · 2 · 2 · 0 · 0 · 0
    2 · 1 · 4 · 4 · 3 · 2 · 3 · 0 · 1 · 1 · 1 · 2 · 2 · 2 · 2 · 0 · 2 · 3 · 1 · 1 · 1 · 1 · 1 · 0 · 0
    1 · 1 · 1 · 2 · 3 · 2 · 3 · 0 · 1 · 1 · 1 · 3 · 1 · 1 · 0 · 0 · 3 · 1 · 1 · 1 · 1 · 1 · 1 · 0 · 0 · 0
    \begin{array}{c} 1 \cdot 1 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 2 \cdot 0 \\ 1 \cdot 1 \cdot 1 \cdot 4 \cdot 1 \cdot 1 \cdot 0 \cdot 0 \\ 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 0 \end{array}
    1.1.2.2.1.1.0.0

\begin{array}{c}
1 \cdot 1 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 1 \cdot 3 \cdot 0 \\
1 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 0 \\
2 \cdot 2 \cdot 1 \cdot 4 \cdot 2 \cdot 1 \cdot 2 \cdot 0
\end{array}
```

# Capitolul 5. Concluzii

În concluzie, mașinile cu vectori suport (SVM) sunt foarte des utilizate pentru probleme de clasificare, regresie, dar și pentru alte tipuri de probleme. Metoda se bazează pe o fundamentare matematică riguroasă, dar demonstrează o capacitate foarte bună de generalizare.

# Capitolul 6. Bibliografie

- 1. Prof. Florin Leon, PhD, *Inteligență artificială: mașini cu vector suport*, Editura Tehnopress
- 2. Prof. Florin Leon, PhD, *Sinteză Curs Inteligență artificială*, <a href="http://florinleon.byethost24.com/Curs\_IA/SintezaIA12.pdf?i=1">http://florinleon.byethost24.com/Curs\_IA/SintezaIA12.pdf?i=1</a>
- 3. Prof. Florin Leon, PhD, *Curs Inteligență artificială*, <a href="http://florinleon.byethost24.com/Curs\_IA/IA12\_RN2.pdf">http://florinleon.byethost24.com/Curs\_IA/IA12\_RN2.pdf</a>
- 4. *UCI Machine Learning Repository*, <a href="https://archive.ics.uci.edu/ml/datasets/Nursery?fbclid=IwAR0kX5DJ-OHfStIAkpIc4kBjlNTo4ZBfAKDLKjX-5DKwrF9CiWdR0oEb-s">https://archive.ics.uci.edu/ml/datasets/Nursery?fbclid=IwAR0kX5DJ-OHfStIAkpIc4kBjlNTo4ZBfAKDLKjX-5DKwrF9CiWdR0oEb-s</a>
- 5. *Microsoft.ML.Data*, *SvmLightLoader Class*, <a href="https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.data.svmlightloader?view="ml-dotnet">https://learn.microsoft.com/en-us/dotnet</a>
- 6. Prof. Florin Leon, PhD, *Laborator Inteligență artificială*, <a href="http://florinleon.byethost24.com/lab\_ia.html">http://florinleon.byethost24.com/lab\_ia.html</a>

### Capitolul 7. Rolul fiecărui membru al echipei

- Lungu Ştefania-Paraschiva:
  - o class Solve
  - class Equation
  - o class Kernel
  - class DataManipulation.fileRead()
  - class DataManipulation.writeInFile()
  - Documentatie
  - Coeficient bias
- Poleac Alexandra-Cătălina:
  - o class Selection
  - o class Crossover
  - o class Mutation
  - class DataManipulation.procent()
  - o Documentatie
  - o Coeficient alpha și w