

CZU: 37.016[004+004.42]:373.5

DOI: 10.36120/2587-3636.v29i3.32-41

## ASPECTE METODOLOGICE PRIVIND EXPLICITAREA PROGRAMĂRII PREDICȚIILOR ÎN CADRUL DISCIPLINEI INTELIGENȚA ARTIFICIALĂ

Andrei BRAICOV, dr., conf. univ.

<https://orcid.org/0000-0001-6416-2357>

Facultatea Fizică, Matematică și Tehnologii Informaționale  
Universitatea Pedagogică de Stat „Ion Creangă” din Chișinău

**Rezumat.** Studiarea regresiei liniare în liceu deschide oportunități pentru examinarea unui șir de probleme practice și poate constitui un obiect de studiu în cadrul disciplinei opționale *Inteligența artificială*, dar și în cadrul modulelor *Calcul numeric*, *Aplicații ale derivatei*. În articol sunt prezentate considerații metodologice pentru explicitarea programării predicțiilor cu ajutorul regresiei liniare prin identificarea argumentată a funcției de regresie, elaborarea algoritmului de implementare sau prin valorificarea instrumentarului unei foi de calcul.

**Cuvinte-cheie:** inteligența artificială, programare, predicții, liceu.

## METHODOLOGICAL ASPECTS REGARDING THE EXPLANATION OF THE PREDICTIONS PROGRAMMING WITHIN THE DISCIPLINE OF ARTIFICIAL INTELLIGENCE

**Abstract.** The study of linear regression in high school opens up opportunities for examining a range of practical problems and can form an object of study within the optional discipline Artificial Intelligence, but also within the modules Numerical Calculus, Applications of the Derivative. The article presents methodological considerations for explaining the programming of predictions with the help of linear regression through the reasoned identification of the regression function, the elaboration of the implementation algorithm or by capitalizing on the tools of a spreadsheet.

**Keywords:** artificial intelligence, programming, predictions, high school.

### Introducere

În ultimii ani cursul de *Inteligență artificială* este recomandat ca disciplină opțională pentru învățământul liceal în mai multe țări.

Ministerul Educației și Cercetării din Republica Moldova a aprobat în 2020 curricula pentru disciplina opțională „Inteligența artificială” [1], care se studiază în clasele XI – XII, iar Ministerul Educației din România a aprobat în 2022 programa pentru disciplina „Introducere în învățarea automată”, care se studiază opțional în clasele XI – XII [2].

Ambele curricula recomandă în calitate de instrument de lucru în domeniul Inteligenței artificiale limbajul de programare Python și urmăresc formarea la elevi a competențelor de elaborare a soluțiilor informatice pentru probleme sociale, culturale, personale care au la bază *învățarea automată*.

*Inteligența artificială* este un domeniu transdisciplinar al Informaticii, Matematicii și Ingineriei, pe cât de intrigant, interesant, atrăgător și captivant, pe atât de complex, care

este focusat pe elaborarea sistemelor inteligente – sisteme capabile să simuleze gândirea logică, critică, similară celei umane sau mai performantă.

*Învățarea automată* (în engleză *Machine Learning*) este un subdomeniu al Inteligenței artificiale, care a luat amploare în ultimii 20 de ani și care presupune construirea algoritmilor capabili să îmbunătățească aplicațiile-sisteme prin acumularea experienței. Astfel, un sistem inteligent bazat pe învățarea automată va persevera continuu spre maximalizarea performanței de funcționare, deci va deveni mai bun după fiecare utilizare.

La rândul ei, învățarea automată a generat subdomenii și mai complexe, cum ar fi *Rețelele neuronale* și *Învățarea profundă* (în engleză *Deep Learning*). Sigur, acestea din urmă constituie obiectul de studiu și de lucru al cercetătorilor și programatorilor profesioniști.

Pentru „începători” (acest termen este convențional, deoarece pentru a purcede la studierea învățării automate este practic obligator să ai o pregătire bună matematică și să fii inițiat într-un limbaj de programare de nivel înalt) se consideră accesibili într-o anumită măsură algoritmi care permit prognozarea, predicția evenimentului ca urmare a unor date colectate și prelucrate [3], cum ar fi:

- ✓ *Regresia liniară* – modelează relația dintre două mulțimi (două variabile), care grafic reprezintă o dreaptă în interiorul unui mulțimi de puncte în plan.
- ✓ *Regresia logistică* – algoritmul care modelează relația dintre o mulțime de variabile independente și o variabilă independentă. Este similară cu regresia liniară, doar că funcția care returnează predicția este neliniară și rezultatul este un număr real din intervalul [0, 1].
- ✓ *Arborii de clasificare și arborii de regresie* (CART) – propuși de Leo Breiman ș.a. în 1984 [4], care sunt, de fapt, arbori decizionali pentru împărțirea datelor în clase și care au drept scop producerea mulțimii de clasificatori (cei mai importanți parametri pentru formularea predicției).
- ✓ *Analiza normală discriminantă* (LDA) – funcție de clasificare a fenomenelor, bazată pe o mulțime de variabile și pe o probabilitate de apariție [5].
- ✓ *Algoritmul KNN* (k-Nearest Neighbors) sau *metoda celor mai apropiați vecini* – utilizată pentru clasificare și pentru regresie [6].
- ✓ *Clasificatorul Naive Bayes* etc.

Din punct de vedere a metodei de organizare învățarea automată se divizează în *învățarea automată nesupervizată* și *învățarea automată supervizată*.

Explicitarea în liceu a învățării automate supervizate ar trebui să includă punerea în evidență a caracteristicilor ei, cum ar fi seturile de date etichetate de antrenare, de test, corespondența dintre datele de intrare și etichete (datele de ieșire), relația de ordine, distanța dintre vecini (date), datele concludente etc.

Cea mai simplă metodă care exemplifică metodologia învățării automate supervizate este *regresia liniară*.

Vom descrie în continuare un exemplu ilustrativ de prezentare și de aplicare a regresiei liniare la rezolvarea unei probleme de predicție. Luând în calcul că la prezentarea acestei metode se aplică derivata funcției și aplicațiile ei în soluționarea problemelor de maxim și minim, problemele de predicție pot fi explicitate elevilor din clasele a XII-a și a XI-a, după studierea modulelor *Funcții derivabile*, *Aplicații ale derivatelor* (modulele 4 și 5 în manualul de Matematică, clasa a 11-a [7]).

### **Formularea problemei practice**

În tabelul care urmează este indicat consumul zilnic de gaz a unui apartament cu centrală termică autonomă și temperatura medie zilnică de afară pe parcursul a 9 zile. Stăpânul apartamentului este matematician și presupune că temperatura medie zilnică de afară este cel mai important factor care influențează consumul de gaz. Astfel, cunoscând prognoza meteo pe următoarele zile, el ar putea calcula consumul de gaz al apartamentului pentru aceste zile. Cu alte cuvinte, stăpânul-matematician vrea să determine relația dintre valoarea temperaturii de afară și volumul de caz consumat de centrala termică a apartamentului său. De exemplu, cât gaz va consuma centrala termică dacă în următoarele 3 zile temperatura medie va fi constantă și va fi egală cu  $-1^{\circ}\text{C}$ ?

**Tabelul 1. Consumul de gaz și temperatura medie zilnică**

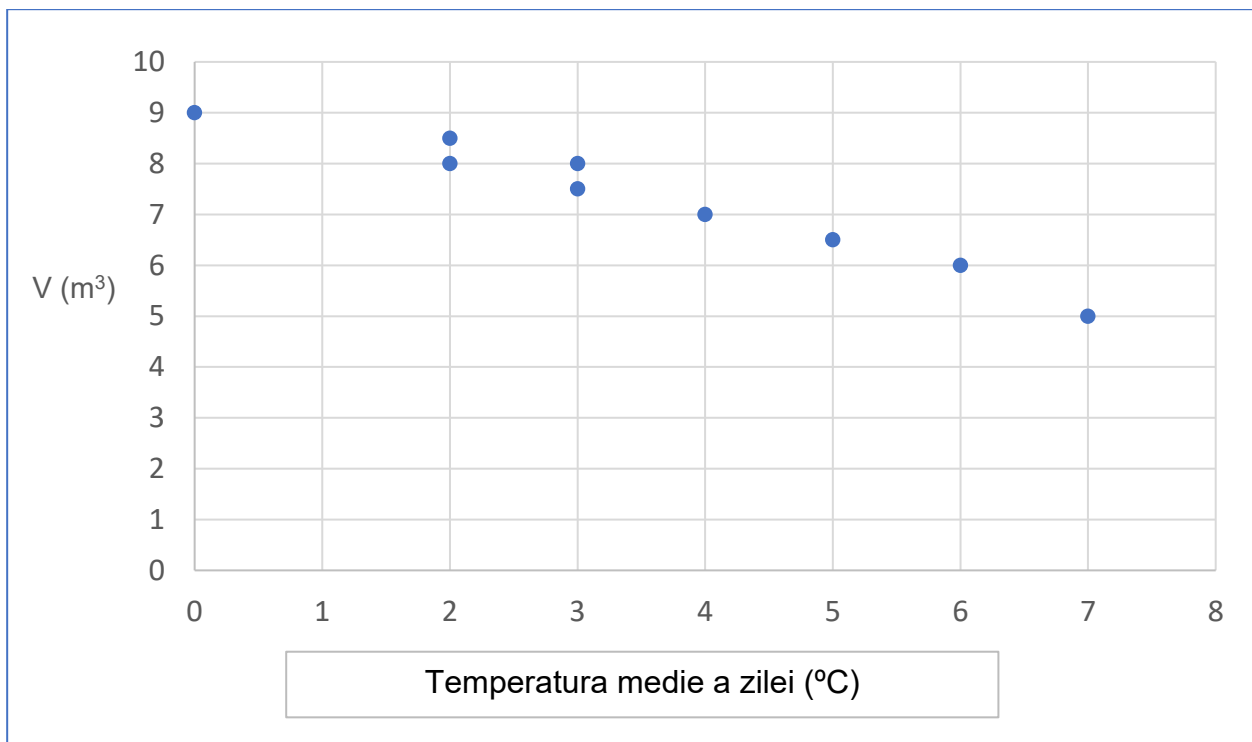
Nr. de ord. al zilei în lună		1	2	3	4	5	6	7	8	9
Temperatura medie zilnică de afară ( $^{\circ}\text{C}$ )	$x_i$	7	5	6	4	3	2	0	2	3
Consumul de gaz ( $\text{m}^3$ )	$y_i$	5	6,5	6	7	8	8,5	9	8	7,5

### **Crearea modelului matematic de rezolvare**

În termeni matematici valorile temperaturii se numesc *observații* sau *date de intrare* (input-uri), iar cantitățile de gaz – *obiectele predicției* sau *date de ieșire* (output-uri sau target-uri).

Sigur, datele din tabel pot fi reprezentate grafic cu ajutorul unui sistem de coordonate cartezian (fig. 1).

Examinând graficul din figura 1 observăm că punctele graficului sunt orientate de-a lungul unei drepte. Prin urmare, este firesc să identificăm această dreaptă, încât ea să ne sugereze/prezică consumul de gaz aproximativ pentru orice dată de intrare, adică pentru orice valoare a temperaturii de afară.



**Figura 1. Graficul dependenței consumului de gaz în raport cu temperatura medie a zilei**

Fie  $f(x) = ax + b$  ecuație acestei drepte. **(1)**

Așadar, trebuie să determinăm constantele  $a$  și  $b$  pe baza datelor din tabelul 1, astfel încât valorile-diferențe  $|f(x_i) - y_i|$ , altfel zis erorile, să fie minime (adică abaterile-distanțe dintre punctele graficului și linia drepte să fie minime). Luând în calcul că  $[f(x_i) - y_i] = \sqrt{(f(x_i) - y_i)^2}$  vom considera **eroare** în punctul  $x_i$  valoarea expresiei  $(f(x_i) - y_i)^2$ , iar **eroare medie** – media aritmetică a tuturor erorilor în punctele  $x_i$ , adică valoarea expresiei  $\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2$  sau, conform (1),

$$\frac{1}{n} \sum_{i=1}^n ((ax_i + b) - y_i)^2, \quad \mathbf{(2)}$$

unde  $n$  este numărul tuturor punctelor care formează graficul (în cazul exemplului nostru  $n = 9$ ).

Astfel, putem formula următoarea problemă:

Ce valori reale trebuie să aibă constantele  $a$  și  $b$ , astfel încât valoarea expresiei (2) să fie minimă?

### **Identificarea funcției de regresie liniară**

În limbaj matematic, spunem că funcția (1) este *funcția de regresie liniară* (a datelor de intrare). Considerăm expresia (2) drept funcție de variabilele  $a$  și  $b$ , notând-o prin  $E$ . Din cursul liceal de matematică elevii cunosc că pentru a afla minimul sau maximul unei funcții trebuie să examinăm derivata acestei funcții egalată cu 0.

- Derivăm funcția  $E$  după variabila  $a$  și o egalăm cu 0:

$$E(a)' = \left( \frac{1}{n} \sum_{i=1}^n ((ax_i + b) - y_i)^2 \right)' = \frac{1}{n} \sum_{i=1}^n 2((ax_i + b) - y_i)x_i = \\ = \frac{2}{n} (a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i) = 0. \quad (3)$$

- Derivăm funcția  $E$  după variabila  $b$  și o egalăm cu 0:

$$E(b)' = \left( \frac{1}{n} \sum_{i=1}^n ((ax_i + b) - y_i)^2 \right)' = \frac{1}{n} \sum_{i=1}^n 2((ax_i + b) - y_i) = \\ = \frac{2}{n} (\sum_{i=1}^n (ax_i - y_i)) + 2b = 0. \quad (4)$$

Dacă notăm  $X = \frac{1}{n} \sum_{i=1}^n x_i$ , iar  $Y = \frac{1}{n} \sum_{i=1}^n y_i$ , (5)

atunci egalitatea (4) poate fi scrisă astfel  $E(b)' = 2aX - 2Y + 2b = 0$ . (6)

Din (6) obținem că

$b = Y - aX. \quad (7)$
-------------------------

Egalitatea (3) poate fi scrisă astfel:

$$\frac{2}{n} a \sum_{i=1}^n x_i^2 = \left( \frac{2}{n} \sum_{i=1}^n x_i y_i - 2bX \right) \text{ sau } \frac{1}{n} a \sum_{i=1}^n x_i^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i y_i - bX \right) \quad (8)$$

Substituim în (8) valoare lui  $b$  din (7) și obținem:

$$\frac{1}{n} a \sum_{i=1}^n x_i^2 = \frac{1}{n} aX^2 = \frac{1}{n} \sum_{i=1}^n x_i y_i - XY + aX^2 \text{ sau} \\ a \left( \frac{1}{n} \sum_{i=1}^n x_i^2 - X^2 \right) = \frac{1}{n} \sum_{i=1}^n x_i y_i - XY. \quad (9)$$

Egalitatea (9) poate fi scrisă astfel:  $a \frac{1}{n} \sum_{i=1}^n (x_i^2 - X^2) = \frac{1}{n} \sum_{i=1}^n (x_i y_i - XY)$ . (10)

Obținem:

$a = \frac{\sum_{i=1}^n (x_i y_i - XY)}{\sum_{i=1}^n (x_i^2 - X^2)} \quad (11)$
---

Valoarea lui  $b$  se obține substituind valoarea lui  $a$  în egalitatea (7).

### **Algoritmul de soluționare și codul de program**

Pentru algoritmul de soluționare a problemei vom folosi formulele (5), (7), (11).

Codul Pascal	Codul C++
<pre>program Consumption; Uses crt; type Vector= array[1..30] of real; var T, C: Vector; n, i, tmp: integer; S1, S2, X, Y, a, b: real;</pre>	<pre>#include &lt;iostream&gt; using namespace std; { int n, i, tmp; double S1, S2, X, Y, a, b; double T[30];</pre>

<pre> BEGIN ClrScr; write ('Write number of inputs: '); readln (n); S1:=0; S2:=0; for i:=1 to n do begin     write('T[' , i, ']=');     readln(T[i]); S1:=S1+T[i];     write('C[' , i, ']=');     readln(C[i]); S2:=S2+C[i]; end; X:=S1/n; Y:=S2/n; S1:=0; S2:=0; for i:=1 to n do begin     S1:=S1+T[i]*C[i]-X*Y;     S2:=S2+T[i]*T[i]-X*X; end; a:=S1/S2; b:=Y-a*X; writeln('We obtained the next linear function:'); writeln('y = ', a:2:2, 'x + ',b:2:2); write ('Input the new value of Temperature: '); readln(tmp); writeln('The consumption will be approximately equal to ', (a*tmp+b):2:1); readkey; END. </pre>	<pre> double C[30]; cout&lt;&lt;"Write number of inputs: "; cin &gt;&gt; n; S1=0; S2=0; for (int i = 1; i &lt;= n; i++) {     cout&lt;&lt;"T["&lt;&lt;i&lt;&lt;"]= ";     cin &gt;&gt; T[i];     S1=S1+T[i];     cout&lt;&lt;"C["&lt;&lt;i&lt;&lt;"]= ";     cin &gt;&gt; C[i];     S2=S2+C[i]; } X=S1/n; Y=S2/n; S1=0; S2=0; for (int i = 1; i &lt;= n; i++) {     S1=S1+T[i]*C[i]-X*Y;     S2=S2+T[i]*T[i]-X*X; } a=S1/S2; b=Y-a*X; cout&lt;&lt;"We obtained the next linear function:"&lt;&lt; endl; cout&lt;&lt;" y = "&lt;&lt; a &lt;&lt;" x "&lt;&lt; b &lt;&lt; endl; cout&lt;&lt;"Input the new value of Temperature: "; cin &gt;&gt; tmp; cout&lt;&lt;"The consumption will be approximately equal to: " &lt;&lt; (a*tmp+b)&lt;&lt; endl; return 0; } </pre>
<pre> Write number of inputs: 9 T[1]=7 C[1]=5 T[2]=5 C[2]=6.5 T[3]=6 C[3]=6 T[4]=4 C[4]=7 T[5]=3 C[5]=8 T[6]=2 C[6]=8.5 T[7]=0 C[7]=9 T[8]=2 C[8]=8 T[9]=3 C[9]=7.5 We obtained the next linear function: y = -0.57x + 9.31 Input the new value of Temperature: -1 The consumption will be approximately equal to 9.9 </pre>	<pre> Write number of inputs: 9 T[1]= 7 C[1]= 5 T[2]= 5 C[2]= 6.5 T[3]= 6 C[3]= 6 T[4]= 4 C[4]= 7 T[5]= 3 C[5]= 8 T[6]= 2 C[6]= 8.5 T[7]= 0 C[7]= 9 T[8]= 2 C[8]= 8 T[9]= 3 C[9]= 7.5 We obtained the next linear function: y = -0.572674 x 9.31395 Input the new value of Temperature: -1 The consumption will be approximately equal to: 9.88663 </pre>

### Codul Python

```
import numpy as np
n=9
mx = [7, 5, 6, 4, 3, 2, 0, 2, 3]
my = [5, 6.5, 6, 7, 8, 8.5, 9, 8, 7.5]
T = np.array(mx)
C = np.array(my)
S1=0
for i in range(0, n):
    S1=S1 + T[i]
X=S1/n
S2=0
for i in range(0, n):
    S2=S2 + C[i]
Y=S2/n
S1=0
S2=0
for i in range(0, n):
    S1=S1+T[i]*C[i]-X*Y
for i in range(0, n):
    S2=S2+T[i]*T[i]-X*X
a = S1/S2
b = Y-a*X
print("y = ", a, "X + ", b)
tmp = input("Input the new value of Temperature: ")
tmp = float(tmp)
r=a*tmp + b
print("The consumption will be approximately equal to ", r)
```

```
y = -0.572674418604651 X + 9.313953488372093
Input the new value of Temperature: -1
The consumption will be approximately equal to 9.886627906976743
```

### ***Determinarea erorii medii***

Analizând graficul funcției obținute (figura 2), observăm ca acesta este în apropierea punctelor-date de intrare.

Pentru a calcula eroarea medie vom utiliza formula (2) și vom adăuga în program următorul cod C++ (unde variabila  $E_m$  este de tip *double*):

```
S1=0;
for (int i = 1; i <= n; i++) {
    S1 = S1+((a*T[i]+b)-C[i])*((a*T[i]+b)-C[i]);
}
Em = S1/n;
cout<<"The mean error is equal to: "<< Em << endl;
```

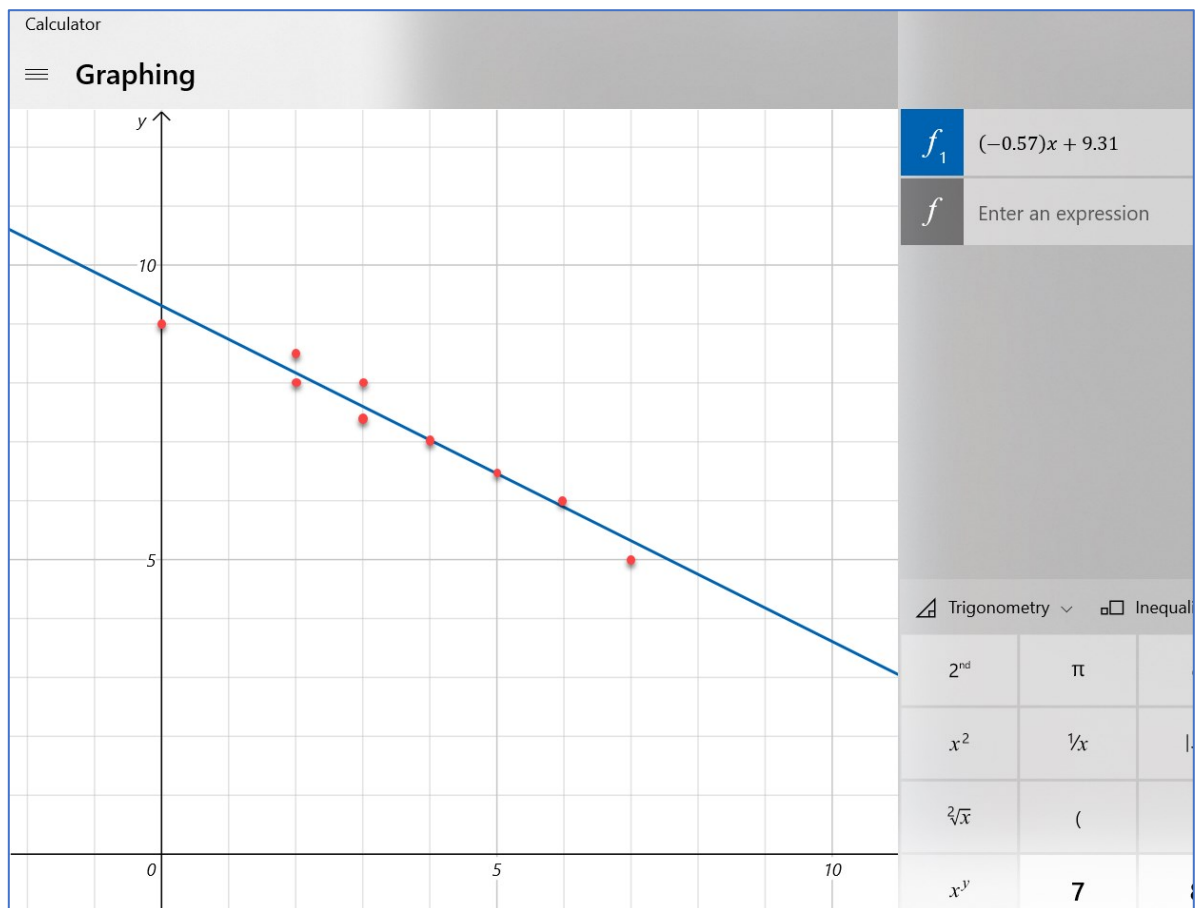


Figura 2. Graficul funcției  $y = -0,57x + 9,31$

Executând codul de program vom obține eroarea medie egală cu 0,0578165 (figura 3).

```
We obtained the next linear function:
y = -0.572674 x 9.31395
Input the new value of Temperature: -1
The consumption will be approximately equal to: 9.88663
The mean error is equal to: 0.0578165
```

Figura 3. Rezultatele execuției împreună cu eroarea medie

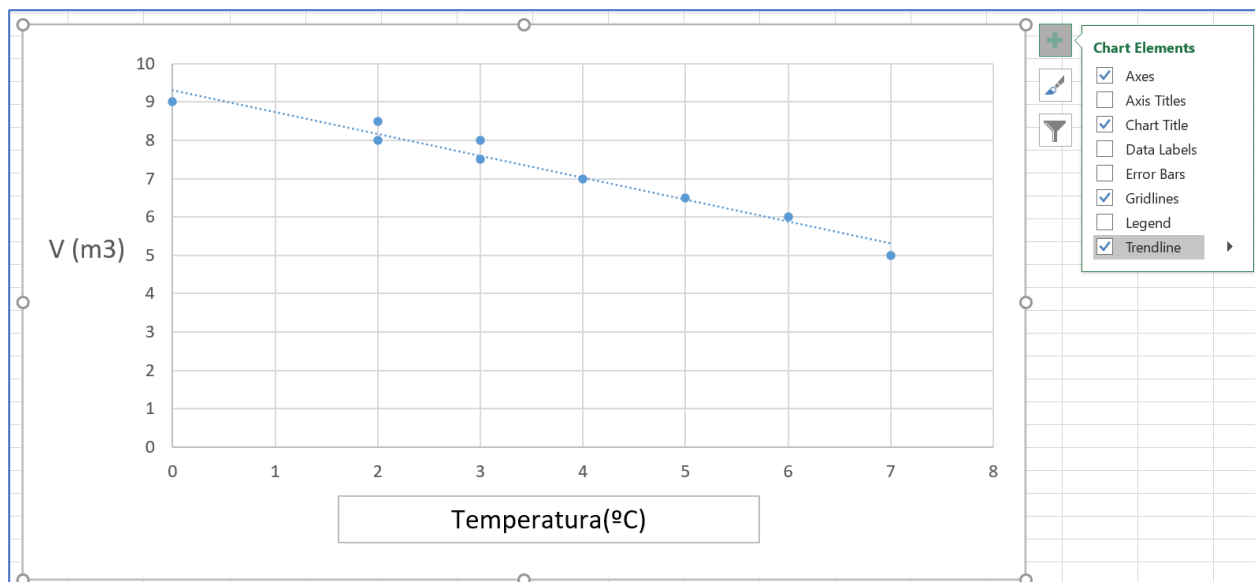
### Utilizarea foilor de calcul pentru rezolvarea probleme

Pentru soluționarea rapidă a problemei se poate folosi o foaie de calcul. Astfel, linia de regresie liniară poate fi generată cu ajutorul opțiunii **Trendline** (figura 4).

Pentru a obține predicția pentru altă valoare a temperaturii medii zilnice folosim funcția **FORECAST.LINEAR(X, Seria\_Y, Seria\_X)**, unde **X** este valoarea nouă de intrare (în cazul nostru  $-1^{\circ}\text{C}$ ), **Seria\_Y** – setul de valori de ieșire (în cazul nostru cele 9 valori ale consumului de gaz, mai exact E5:M5, conform figurii 5), **Seria\_X** – setul de valori de intrare (în cazul nostru cele 9 valori ale temperaturii, mai exact E4:M4, conform figurii 5). Observăm, că aplicând această funcție obținem același rezultat ca și al algoritmul nostru.



Methodological aspects regarding the explanation  
of the predictions programming within the discipline of artificial intelligence



**Figura 4. Construirea liniei de regresie liniară într-o foaie de calcul**

	1	2	3	4	5	6	7	8	9	10
<b>T (°C)</b>	7.0	5.0	6.0	4.0	3.0	2.0	0.0	2.0	3.0	-1.0
<b>V (m3)</b>	5.000	6.500	6.000	7.000	8.000	8.500	9.000	8.000	7.500	9.887

**Figura 5. Prezicerea consumului într-o foaie de calcul cu ajutorul funcției  
FORECAST.LINEAR**

Sigur, datele cunoscute ale unei probleme de predicție grafic nu sunt întotdeauna amplasate de-a lungul unei linii drepte. Din aceste considerente în afară de regresia liniară, se folosește regresia pătratică, interpolarea prin polinoame (Lagrange, Newton) etc.

În plus, datele cunoscute grafic pot reprezenta figuri multidimensionale, astfel încât deseori generarea prezicerii constituie o problemă dificilă.

### Concluzii

Studierea regresiei liniare în liceu deschide oportunități pentru examinarea unui șir de probleme practice și poate constitui un obiect de studiu în cadrul disciplinei opționale *Inteligența artificială*, dar și în cadrul modulelor *Calcul numeric*, *Aplicații ale derivatei*. Putem trage concluzia că aplicând metodologia de rezolvare a problemelor de regresie liniară se va stimula dezvoltarea competențelor transdisciplinare, drept consecință și a celor transversale.

Această transdisciplinaritate cuprinde preponderent domeniile științelor exacte (deci domeniile STEM): Matematică, Informatică, Tehnologii Informaționale, Fizică, Chimie, Biologie și Geografie.

*Articol realizat în cadrul proiectului de cercetări științifice „Metodologia implementării TIC în procesul de studiere a științelor reale în sistemul de educație din Republica Moldova din perspectiva inter/transdisciplinarității (concept STEAM)”, inclus în „Program de stat” (2020-2023), Prioritatea IV: Provocări societale, cifrul 20.80009.0807.20, cu suportul financiar oferit de Agenția Națională pentru Dezvoltare și Cercetare*

## **Bibliografie**

1. Planul-cadru pentru învățământul primar, gimnazial și liceal, anul de studii 2022 – 2023. [ordin\\_mec\\_nr\\_123\\_din\\_28.02.2022\\_plan-cadru\\_2022-2023.pdf \(gov.md\)](#)
2. Ordinul 4049 al Ministrului Educației din România privind programa pentru disciplina Învățare automată [https://rocnee.eu/images/rocnee/fisiere/curriculum/Ordin\\_nr\\_4049\\_14\\_06\\_2022\\_programa\\_invatare%20automata%20.pdf?fbclid=IwAR3E9orKnB5azSaqHIZ\\_K8\\_sQe5ykwB-JcZEmFsvUdF6E4exgBq4phBfp5g](https://rocnee.eu/images/rocnee/fisiere/curriculum/Ordin_nr_4049_14_06_2022_programa_invatare%20automata%20.pdf?fbclid=IwAR3E9orKnB5azSaqHIZ_K8_sQe5ykwB-JcZEmFsvUdF6E4exgBq4phBfp5g)
3. Alfredo Junior. 7 Most Popular Machine Learning Algorithms for Beginners. <https://morioh.com/p/f199016e23bb?fbclid=IwAR07GKDWVfw73UDL7pfh8fVB5aa4H8abJbTt1CciM6wmDs5OPLI6l3lJLRg>
4. BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. *Arborii de clasificare și regresie*. Monterey, CA: Wadsworth & Brooks / Cole Advanced Books & Software, 1984. ISBN 978-0-412-04841-8.
5. <https://ro.economy-pedia.com/11040389-discriminant-analysis>
6. [https://webpace.ulbsibiu.ro/daniel.morariu/html/StudentDoc/DM/Lab4\\_KNN.pdf](https://webpace.ulbsibiu.ro/daniel.morariu/html/StudentDoc/DM/Lab4_KNN.pdf)
7. ACHIRI, I.; CIOBANU, V.; EFROS, P.; GARIT, V.; NEAGU, V.; PRODAN, N.; TARAGAN, D.; TOPALĂ, A. *Matematică. Manual pentru clasa a 11-a*. Chișinău: Prut Internațional. 304 p. ISBN 978-9975-54-514-3.