

Rețele neuronale și deep learning

Cursul 5: Rețele recurente: RNN, LSTM și GRU pentru secvențe și serii de timp financiare

Prof. dr. Daniel Traian PELE

Academia de Studii Economice din București (ASE), Departamentul de Statistică și Econometrie
Academia Română, Institutul de Prognoză Economică (IPE)
IDA Institute Digital Assets ▪ AI4EFin ▪ MSCA Digital Finance

Obiectivele cursului

- Să scrieți o rețea neuronală recurentă (RNN) sub formă de ecuații, să o desfășurați în timp și să-i numărați parametrii.
- Să derivați propagarea înapoi în timp (BPTT) și să explicați de ce gradienții dispar sau explodează.
- Să citiți celulele long short-term memory (LSTM) și gated recurrent unit (GRU) poartă cu poartă.
- Să prognozați volatilitatea, randamentele și consumul de energie electrică cu un protocol fără scurgeri de informație, după modelele de referință: autoregresiv heterogen (HAR), autoregresiv generalizat cu heteroscedasticitate condiționată (GARCH), autoregresiv (AR) și autoregresiv cu variabile exogene (ARX).
- Să explicați de ce un perceptron multistrat (MLP) aplicat unei ferestre de valori trecute nu este suficient.
- Să decideți, cu un test de comparare a prognozelor, cînd merită costul unei rețele recurente.

Abrevieri

AI	inteligență artificială
AR	model autoregresiv
ARX	model autoregresiv cu variabile exogene
ASE	Academia de Studii Economice din București
BiLSTM	LSTM bidirecțional
BPTT	propagarea înapoi în timp
BVB	Bursa de Valori București
CEE	Europa Centrală și de Est
CNN	rețea neuronală convoluțională
CPU	procesor central
DM	testul Diebold–Mariano
ENTSO-E	Rețeaua europeană a operatorilor de transport și de sistem pentru energie electrică
EODHD	End of Day Historical Data (furnizor de date)
ES	pierderea medie așteptată peste VaR (expected shortfall)
GARCH	model autoregresiv generalizat cu heteroscedasticitate condiționată
GK	estimatorul Garman–Klass al variației, bazat pe amplitudine
GRU	unitate recurentă cu porți (gated recurrent unit)
GW	gigawatt
HAR	model autoregresiv heterogen
IPE	Institutul de Prognoză Economică, Academia Română
LSTM	memorie pe termen scurt extinsă (long short-term memory)
MAE	eroarea absolută medie
MAPE	eroarea procentuală absolută medie
MLP	perceptron multistrat
MSE	eroarea pătratică medie
MW	megawatt
MWh	megawatt-oră
RNN	rețea neuronală recurentă
TLV	Banca Transilvania (simbolul la Bursa de Valori București)
VaR	valoarea expusă la risc (value-at-risk)

Datele: trei secvențe reale

Seria	Sursa	Perioada	Sarcina
S&P 500, prețuri zilnice de deschidere, maxim, minim, închidere	Yahoo Finance	4 ianuarie 2005 – 30 octombrie 2026	volatilitate, randamente
Banca Transilvania (TLV), închidere ajustată	End of Day Historical Data (EODHD), Bursa de Valori București (BVB)	3 ianuarie 2008 – 30 octombrie 2026	randamente
Consumul orar din România	Rețeaua europeană a operatorilor de transport și de sistem pentru energie electrică (ENTSO-E), prin energy-charts.info	1 ianuarie 2016 – 31 octombrie 2026	consumul pentru ziua următoare

- Împărțiri în ordine cronologică: antrenare pînă la 31 decembrie 2016, validare pînă la 31 decembrie 2019, test de la 1 ianuarie 2020 (consum: antrenare pînă la 31 decembrie 2022, validare pînă la 31 decembrie 2023).
- Scalările, oprirea timpurie și penalizările sînt estimate doar pe anii de antrenare sau de validare; anii de test sînt folosiți o singură dată.
- Rețelele sînt antrenate pe procesorul central (CPU), determinist pentru o sămînță fixă (20261108 + indicele seminței).

(data/SOURCES.md)

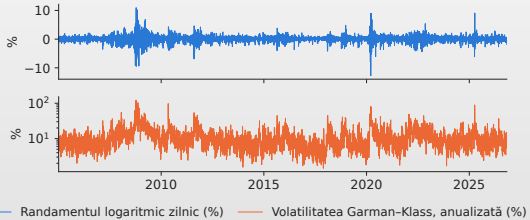
Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

Randamentele și gruparea volatilității

- **Pe scurt:** mărimea mișcării de azi spune mult despre mărimea celei de mîine; semnul ei nu spune aproape nimic.

$$r_t = 100 \ln \frac{P_t}{P_{t-1}}, \qquad \hat{\sigma}^2_{\mathrm{GK},t} = \tfrac{1}{2}\left(\ln \frac{H_t}{L_t}\right)^2 - (2\ln 2 - 1)\left(\ln \frac{C_t}{O_t}\right)^2$$



Sursa: Yahoo Finance, cotații zilnice. Date: 4 ianuarie 2005 - 30 octombrie 2026, zile de tranzacționare.

- Estimatorul Garman–Klass (GK), din prețurile de deschidere O_t, maxim H_t, minim L_t și închidere C_t
  - Volatilitatea anualizată are mediana 8,77 % și vîrful 122 %, la 10 octombrie 2008.

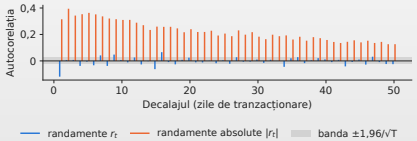
(Garman & Klass, 1980; calcul propriu)

NN_L5_sp500_vol

Autocorelația randamentelor și a randamentelor absolute

- **Pe scurt:** autocorelațiile lui $|r_t|$ rămîn pozitive luni întregi; cele ale lui r_t sînt în banda de zgomot.

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$



Sursa: Yahoo Finance, cotații zilnice. Date: 4 ianuarie 2005 - 30 octombrie 2026, zile de tranzacționare.

- $\hat{\rho}_r(1) = -0{,}119$, dar $\hat{\rho}_{|r|}(1) = 0{,}315$ și $\hat{\rho}_{|r|}(50) = 0{,}126$
  - Banda $\pm 0{,}026$
  - Decalaje semnificative din 50: 15 pentru r_t, 50 pentru $|r_t|$.

- **În practică:** un model cu memorie merită încercat pentru risc (volatilitate, valoarea expusă la risc (VaR)), mult mai puțin pentru direcția prețului.

(calcul propriu)

NN_L5_acf_abs

Memoria limitată a unui MLP pe fereastră fixă

- **Pe scurt:** un MLP aplicat ultimelor W observații are ponderi separate pentru fiecare decalaj și o limită strictă de memorie.

$$\hat{y}_{t+1} = g\Big(\mathbf{W}[\mathbf{x}_{t-W+1}; \ldots; \mathbf{x}_t] + \mathbf{b}\Big), \qquad \mathbf{W} \in \mathbb{R}^{m \times Wd}$$

- **unde**
  - $\mathbf{x}_t \in \mathbb{R}^d$ intrările zilei t
  - W lungimea ferestrei
  - m unitățile ascunse
  - Numărul de parametri crește liniar cu W

- **Ce ne dorim în schimb**
  - O stare $\mathbf{h}_t$ care rezumă trecutul
  - Aceeași regulă de actualizare la fiecare pas (ponderi partajate în timp, ca o rețea neuronală convoluțională (CNN) în spațiu)
  - Orice lungime a secvenței

- **În practică:** în producție, starea se actualizează la fiecare observație nouă, fără a recalcula întregul istoric.

(Bishop & Bishop, 2024, cap. 12)

Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

RNN simplă: o stare, actualizată după aceeași regulă

⊡ **Pe scurt:** noua stare combină starea veche cu intrarea nouă printr-o funcție neliniară.

$$\mathbf{h}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h), \qquad \hat{y}_t = \mathbf{W}_{hy}\mathbf{h}_t + b_y$$

⊡ **unde**

- $\mathbf{h}_t \in \mathbb{R}^{n_h}$ starea ascunsă, $\mathbf{h}_0 = \mathbf{0}$
- $\mathbf{W}_{xh} \in \mathbb{R}^{n_h \times d}$, $\mathbf{W}_{hh} \in \mathbb{R}^{n_h \times n_h}$ comune pentru toți t
- Parametri $n_h(d + n_h + 1) + n_h + 1$, independenți de lungimea T

⊡ Desfășurată în timp, este o rețea adîncă de adîncime T, cu ponderi legate.

⊡ Modelul nostru de volatilitate: $d = 3$ intrări (log-varianța, randamentul, randamentul absolut), $n_h = 32$, fereastră de $T = 22$ de zile de tranzacționare.

⊡ Many-to-one: doar $\hat{y}_T$ (log-varianța de mîine) intră în funcția de pierdere.

⊡ **În practică:** starea ascunsă este un analog învățat și neliniar al varianței condiționate dintr-un model GARCH.

(Rumelhart et al., 1986; Bishop & Bishop, 2024, sect. 12.2)

Propagarea înapoi în timp: o sumă peste drumuri

- **Pe scurt:** gradientul unei ponderi comune adună contribuția ei de la fiecare pas, fiecare trecînd printr-un produs de matrice jacobiene.

$$\frac{\partial \mathcal{L}_T}{\partial \mathbf{W}_{hh}} = \sum_{t=1}^{T} \frac{\partial \mathcal{L}_T}{\partial \mathbf{h}_T} \left(\prod_{s=t+1}^{T} \mathbf{J}_s \right) \frac{\partial^+ \mathbf{h}_t}{\partial \mathbf{W}_{hh}}$$

$$\mathbf{J}_s = \frac{\partial \mathbf{h}_s}{\partial \mathbf{h}_{s-1}} = \operatorname{diag}(1 - \mathbf{h}_s^2)\mathbf{W}_{hh}$$

- **unde**
  - ∂^+ derivata imediată, cu $\mathbf{h}_{t-1}$ fixat
  - $\mathbf{J}_s$ matricea jacobiană a unui pas
  - Derivarea completă este în anexă

- Cost: o trecere înapoi prin toți cei T pași, memorie $O(Tn_h)$.
- BPTT trunchiată: propagare înapoi doar pe $K < T$ pași
  - Mai ieftin, dar fără semnal de gradient dincolo de K.

- **În practică:** produsul celor $T - t$ matrice jacobiene decide ce poate învăța rețeaua din trecutul îndepărtat.

(Werbos, 1990)

Gradienți care dispar și gradienți care explodează

- **Pe scurt:** un produs de multe matrice scade sau crește geometric cu distanța în timp.

$$\left\| \prod_{s=t+1}^{T} \mathbf{J}_s \right\| \leq \left(\gamma\, \sigma_{\max}(\mathbf{W}_{hh})\right)^{T-t}, \qquad \gamma = \max |\tanh'| \leq 1$$

- **unde**
  - $\sigma_{\max}$ cea mai mare valoare singulară
  - Dacă $\gamma\sigma_{\max} < 1$, marginea dispare geometric (condiție suficientă)
  - O valoare proprie maximă peste 1 este necesară pentru explozie

- **Explozia**: limitarea normei gradientului
  - $\mathbf{g} \leftarrow \mathbf{g} \cdot \min(1,\, c/\|\mathbf{g}\|)$; folosim $c = 1$ (ales a priori)
- **Dispariția**: schimbarea arhitecturii
  - Un drum aditiv prin timp: celulele LSTM și GRU, în secțiunea următoare

- **În practică:** în practică, o RNN simplă își amintește cîteva zeci de pași: suficient pentru volatilitatea zilnică, nu pentru tiparele săptămînale din datele orare.

(Bengio et al., 1994; Pascanu et al., 2013)

Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

Celula LSTM și porțile ei

- **Pe scurt:** o stare separată a celulei, $\mathbf{c}_t$, se actualizează aditiv; trei porți decid ce se uită, ce se scrie și ce se arată.

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_f), \quad \mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_i),$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_o), \quad \tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_c),$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$$

- **unde**
  - σ funcția logistică, cu valori în $(0, 1)$
  - $\odot$ produsul element cu element
  - $\mathbf{f}, \mathbf{i}, \mathbf{o}$ porțile de uitare, intrare și ieșire
  - Parametri $4n_h(n_h + d + 1)$

- Poarta de uitare aproape de 1: celula își păstrează memoria
  - Aproape de 0: o resetează.
- Poarta de intrare: cît din candidatul $\tilde{\mathbf{c}}_t$ se scrie.
- Poarta de ieșire: cît din memorie apare în $\mathbf{h}_t$.

- **În practică:** porțile sînt comutatoare învățate
  - Pe datele de piață ele reacționează la șocuri (două diapozitive mai încolo).

(Hochreiter & Schmidhuber, 1997)

De ce starea celulei duce gradienții departe în trecut

- **Pe scurt:** de-a lungul stării celulei, jacobianul este o diagonală de porți de uitare, nu o matrice de ponderi comprimată de $\tanh'$.

$$\left.\frac{\partial \mathbf{c}_t}{\partial \mathbf{c}_{t-1}}\right|_{\text{direct}} = \operatorname{diag}(\mathbf{f}_t), \quad \prod_{s=t+1}^{T} \operatorname{diag}(\mathbf{f}_s) \approx \mathbf{I} \text{ dacă } \mathbf{f}_s \approx \mathbf{1}$$

- **unde**
  - Celelalte drumuri trec prin $\mathbf{h}_{t-1}$ și prin porți
  - Drumul direct este cel care menține gradienții pe distanțe lungi

- Inițializarea pozitivă a termenului liber al porții de uitare (de exemplu 1) face $\mathbf{f}_t \approx \sigma(1) = 0{,}73$ la început: memorie implicită.
- În perioada de test a volatilității, media porții de uitare este 0,547, iar a porții de intrare 0,544.
- Actualizarea aditivă este aceeași idee ca a conexiunii reziduale din cursul 4.

- **În practică:** alegeți LSTM sau GRU atunci cînd trecutul util este lung (o săptămînă de ore) și o RNN simplă doar pentru memorii scurte.

(Hochreiter & Schmidhuber, 1997)

GRU: două porți și nicio stare separată

- **Pe scurt:** o poartă de actualizare interpolează între starea veche și un candidat; o poartă de resetare decide cît trecut intră în candidat.

$$\mathbf{z}_t = \sigma(\mathbf{W}_z[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_z), \qquad \mathbf{r}_t = \sigma(\mathbf{W}_r[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_r),$$
$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h[\mathbf{r}_t \odot \mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_h), \qquad \mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t$$

- **unde**
  - $\mathbf{z}_t$ poarta de actualizare
  - $\mathbf{r}_t$ poarta de resetare
  - Parametri $3n_h(n_h + d + 1)$, trei sferturi dintr-un LSTM cu același n_h

- Același drum aditiv: $\partial \mathbf{h}_t / \partial \mathbf{h}_{t-1} \supset \mathrm{diag}(1 - \mathbf{z}_t)$.
- LSTM sau GRU: nu există un cîștigător general
  - Comparați-le pe datele de validare, așa cum facem mai jos.

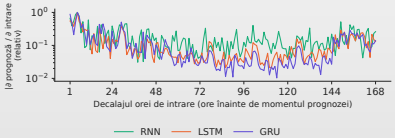
- **În practică:** numărul mai mic de parametri face GRU puțin mai rapidă la antrenare și mai puțin pretențioasă la date.

(Cho et al., 2014)

Sensibilitatea la intrări pe decalaje: rețelele se bazează pe ziua și pe săptămîna precedentă

- **Pe scurt:** sensibilitatea ieșirii rețelelor de consum antrenate la fiecare dintre cele 168 de ore de intrare, mediată pe 200 de zile de test.

$$S(k) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\partial \hat{y}_{i,12}}{\partial x_{i,\,T+1-k}} \right|$$



Sursa: energy-charts.info (Fraunhofer ISE), din datele burselor și ale ENTSO-E. Date: 200 de zile de test din 2024; modele cu sămînța 20261108; prognoza pentru ora 12 (timpul universal).

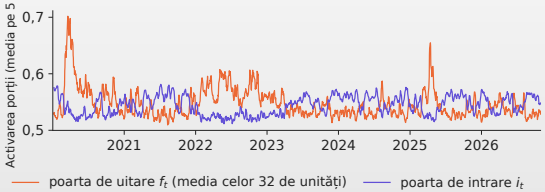
- Citirea a 7 profiluri zilnice în locul a 168 de ore scurtează drumul gradientului la 7 pași
  - La un decalaj de o săptămînă, sensibilitatea raportată la maximul ultimei zile este RNN 0,261, LSTM 0,131, GRU 0,140.

(calcul propriu)

NN_L5_grad_lag

Porțile LSTM reacționează la stresul din piață

- **Pe scurt:** activările porților LSTM-ului de volatilitate antrenat, în anii de test 2020–2026 (medii pe 5 zile).



Sursa: Yahoo Finance, cotații zilnice. Date: perioada de test 2020-2026, LSTM cu sămînța 20261108, ultimul pas al fiecărei ferestre de 22 de zile.

- Corelația cu log-varianța de azi: poarta de uitare 0,743, poarta de intrare −0,616.

- **În practică:** porțile sînt un instrument de diagnostic, nu o explicație: folosiți-le pentru a vedea cînd își schimbă modelul regimul, apoi verificați pe date.

(calcul propriu)

NN_L5_lstm_gates

Arhitecturi pentru secvențe și scurgerea de informație din viitor

- **Many-to-one**
  - O fereastră la intrare, o valoare la ieșire: volatilitatea de mîine
- **Many-to-many (ieșire vectorială)**
  - O săptămînă de profiluri zilnice la intrare, cele 24 de ore de mîine la ieșire: modelul nostru de consum
- **Codificator–decodificator**
  - O rețea citește, alta scrie o secvență pas cu pas
- **Straturi suprapuse**
  - Stările stratului 1 sînt intrările stratului 2
- **Bidirecțional (BiLSTM)**
  - Citește fereastra înainte și înapoi
  - Potrivit pentru clasificarea unui text sau a unei secvențe încheiate
  - O scurgere de informație în prognoză dacă trecerea înapoi vede viitorul

- **În practică:** în prognoză, fiecare intrare trebuie să fie cunoscută în momentul prognozei: verificați acest lucru înaintea oricărei alegeri de arhitectură.

(Bishop & Bishop, 2024, cap. 12)

Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

Ferestre, scalare și împărțiri fără scurgeri de informație

- **Pe scurt:** fiecare număr pe care îl vede modelul trebuie să poată fi calculat în momentul prognozei t.

$$\mathcal{D} = \{(\mathbf{x}_{t-W+1:t},\, y_{t+1})\}, \qquad \tilde{x} = \frac{x - \bar{x}_{\text{train}}}{s_{\text{train}}}$$

- **unde**
  - W fereastra
  - $\bar{x}_{\text{train}}, s_{\text{train}}$ media și abaterea standard calculate numai pe anii de antrenare

- S&P 500: 3 020 de zile de antrenare, 754 de validare și 1 717 de test, în această ordine.
- Oprirea timpurie: ne oprim cînd pierderea de validare nu s-a îmbunătățit timp de 6 epoci
  - Păstrăm cele mai bune ponderi.
- Semințe: cîte 5 pentru fiecare rețea (volatilitate, randamente), 3 pentru consum
  - Raportăm dispersia și prognoza medie.

- **În practică:** o amestecare aleatoare a zilelor înainte de împărțire pune anul 2020 în setul de antrenare și face orice model să pară bun pe 2020.

(calcul propriu)

Exercițiu practic: §2

Modelele de referință: naiv, HAR și GARCH

- **Pe scurt:** o rețea trebuie să depășească modele pe care econometria le-a rafinat de decenii.

$$\text{HAR: } y_{t+1} = \beta_0 + \beta_d y_t + \beta_w \bar{y}_{t-4:t} + \beta_m \bar{y}_{t-21:t}$$

$$\text{GARCH(1,1): } \sigma^2_{t+1} = \omega + \alpha\,\varepsilon_t^2 + \beta\,\sigma_t^2, \quad \varepsilon_t = r_t - \mu$$

- Estimările modelului HAR: $\hat{\beta}_d = 0{,}192$, $\hat{\beta}_w = 0{,}439$, $\hat{\beta}_m = 0{,}293$.
- GARCH: $\hat{\alpha} = 0{,}117$, $\hat{\beta} = 0{,}873$, persistență $0{,}990$, $\hat{\nu} = 5{,}78$.

- **unde**
  - $y_t = \ln \hat{\sigma}^2_{\mathrm{GK},t}$
  - $\bar{y}$ medii pe o săptămînă și pe o lună
  - GARCH cu erori Student-t, raportat la y printr-o regresie liniară pe anii de antrenare
  - Naiv: $\hat{y}_{t+1} = y_t$

- **În practică:** HAR este un model liniar cu trei medii
  - Dacă un LSTM nu îl poate depăși, LSTM-ul nu merită întreținut.

(Corsi, 2009; Bollerslev, 1986)

Testul Diebold–Mariano pentru acuratețe predictivă egală

- **Pe scurt:** comparăm zi de zi pierderile a două prognoze și testăm dacă diferența lor medie este zero.

$$d_t = L(e_{1,t}) - L(e_{2,t}), \qquad \mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2_{\mathrm{NW}}/n}} \overset{H_0}{\sim} \mathcal{N}(0,1)$$

- **unde**
  - L eroarea pătratică (volatilitate) sau eroarea absolută (consum)
  - $\hat{\sigma}^2_{\mathrm{NW}}$ varianța de termen lung Newey–West a lui d_t
  - $\mathrm{DM} < 0$: modelul 1 este mai bun

- Testul Diebold–Mariano (DM) ține seama de autocorelația diferențelor de pierdere, pe care erorile de prognoză o au întotdeauna.
- Semințele măsoară zgomotul antrenării
  - DM măsoară zgomotul de eșantionare al perioadei de test.

- **În practică:** declarați un model mai bun doar cu valoarea p a testului DM și precizați perioada.

(Diebold & Mariano, 1995)

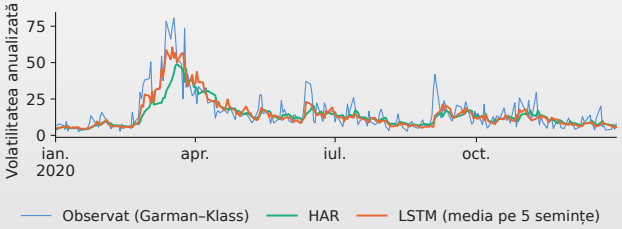
Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

În criza din 2020, LSTM s-a adaptat mai repede decît HAR

- **Pe scurt:** prognoze pentru ziua următoare ale volatilității anualizate, în primul an de test.

$$\hat{\sigma}_{t+1}^{\text{ann}} = 100\sqrt{252\,\exp(\hat{y}_{t+1})}$$



Sursa: Yahoo Finance, cotații zilnice. Date: anul de test 2020 (prognoze pentru ziua următoare), zile de tranzacționare.

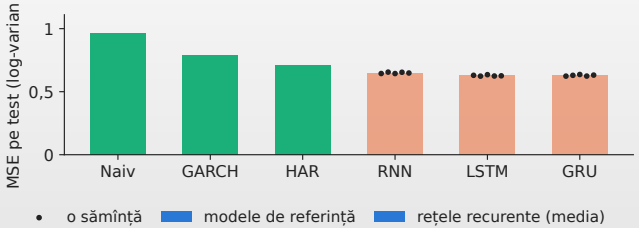
- Fereastra crizei (20 februarie – 30 aprilie 2020), eroarea pătratică medie (MSE) a lui $\hat{y}$: naiv 0,705, GARCH 1,032, HAR 1,014, LSTM 0,521.

(calcul propriu)

NN_L5_vol_forecast

Rețelele recurente depășesc HAR și GARCH pe volatilitate

- **Pe scurt:** MSE pe test a log-varianței, în perioada 2020–2026; punctele sînt semințe individuale.



Sursa: Yahoo Finance, cotații zilnice. Date: test 2 ianuarie 2020 - 31 octombrie 2026; 5 semințe pentru fiecare rețea.

- MSE: naiv 0,966, GARCH 0,786, HAR 0,706
  - RNN 0,649, LSTM 0,628, GRU 0,629 (medii pe semințe).

(calcul propriu)

NN_L5_vol_mse

Rezultatele pentru volatilitate, cu teste

Modelul	MSE	R^2	Semințe
Naiv	0,966	0,259	–
GARCH(1,1)-t	0,786	0,397	–
HAR	0,706	0,458	–
RNN (prognoza medie)	0,643	0,507	5
LSTM (prognoza medie)	0,625	0,521	5
GRU (prognoza medie)	0,624	0,521	5

- $R^2 = 1 - \mathrm{MSE}/\widehat{\mathrm{Var}}(y)$ pe cele 1 717 de zile de test.

- LSTM față de HAR: variația MSE $-11{,}6\,\%$
  - DM $= -5{,}52$, $p < 0{,}001$.
- LSTM față de GARCH: DM $= -6{,}80$, $p < 0{,}001$.
- LSTM față de modelul naiv: variația MSE $-35{,}3\,\%$.
- Cea mai bună rețea după prognoza medie: GRU
  - Dispersia LSTM pe semințe: între 0,623 și 0,635.

- **În practică:** cîștigul este semnificativ pe această perioadă de test
  - Înainte de a înlocui HAR într-un model VaR, verificați-l pe altă piață și pe subperioade și păstrați HAR ca soluție de rezervă.

(calcul propriu)

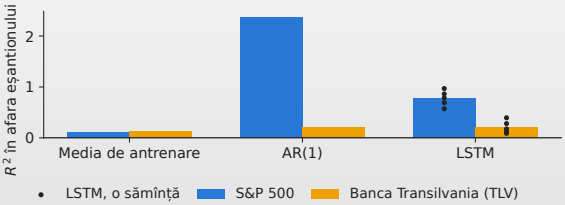
Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

Prognozele randamentelor explică foarte puțin din varianța în afara eșantionului

- **Pe scurt:** randamentele de a doua zi ale S&P 500 și ale Banca Transilvania (TLV), din ultimele 22 de randamente.

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{t \in \text{test}} (r_{t+1} - \hat{r}_{t+1})^2}{\sum_{t \in \text{test}} r^2_{t+1}}$$



Sursa: Yahoo Finance; EODHD. Date: test 2020 - octombrie 2026; S&P 500 (Yahoo Finance) și TLV (EODHD, Bursa de Valori București).

- S&P 500: AR(1) 2,366 %, LSTM 0,776 % (semințe între 0,572 și 0,969)
  - TLV: AR(1) 0,201 %, LSTM 0,213 %.

(calcul propriu)

NN_L5_returns_r2

Predictibilitatea redusă a randamentelor

- **Pe scurt:** pe o piață lichidă, partea previzibilă a randamentelor este infimă față de zgomotul lor.

$$r_{t+1} = \mu_t + \varepsilon_{t+1}, \qquad \frac{\mathrm{Var}(\mu_t)}{\mathrm{Var}(r_{t+1})} \approx 0$$

- **unde**
  - μ_t media condiționată pe care o poate învăța un model
  - ε_{t+1} știrile care apar după t

- Acuratețea semnului pentru LSTM pe testul S&P 500: 52,1 %, față de 54,0 % zile de creștere (prognoza "mereu în sus").
- Panta AR(1) estimată pe anii de antrenare: −0,102 (S&P 500), 0,015 (TLV).
- Modelele flexibile învață zgomotul în antrenare
  - Oprirea timpurie limitează paguba, dar nu poate crea un semnal.

- **În practică:** testul retrospectiv al unui furnizor de "predicții de randamente bazate pe inteligență artificială (AI)" trebuie să treacă acest test pe date nevăzute, după costuri.

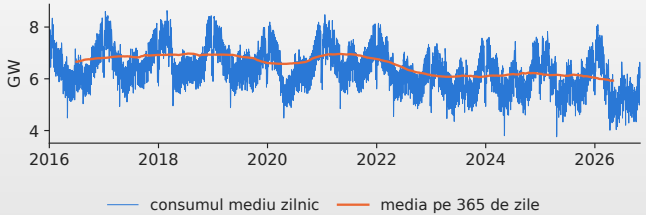
(calcul propriu)

Cuprins

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

Consumul are cicluri zilnice, săptămînale și anuale

- **Pe scurt:** o secvență cu memorie lungă și regulată: un candidat natural pentru rețelele recurente.



Sursa: energy-charts.info (Fraunhofer ISE), din datele burselor și ale ENTSO-E. Date: 1 ianuarie 2016 - 31 octombrie 2026, orar, MW medii pe zi.

- Consum orar de la 1 ianuarie 2016 la 31 octombrie 2026, în medie 6 510 megawatt (MW)
  - Cele 129 de ore lipsă rămîn lipsă (zilele respective sînt omise).

(calcul propriu)

NN_L5_load_series

Sarcina pentru ziua următoare și modelele ei de referință

- **Pe scurt:** la sfîrșitul zilei d, prognozăm cele 24 de valori orare ale consumului din ziua $d+1$, pe baza ultimei săptămîni de ore.

$$\hat{\mathbf{y}}_{d+1} = \mathbf{y}_{d-6} + \mathbf{W}_o[\mathbf{h}_7; \mathbf{s}_{d+1}] + \mathbf{b}_o \in \mathbb{R}^{24}$$

$$\text{ARX: } \hat{\mathbf{y}}_{d+1} = \mathbf{y}_{d-6} + \mathbf{B}\,\mathbf{x}_{d,1:168} + \grave{}\mathbf{s}_{d+1}$$

- 2 186 de zile de antrenare, 365 de validare, 1035 de test
  - 3 semințe pentru fiecare rețea.

- **unde**
  - $\mathbf{y}_{d-6}$ prognoza naivă sezonieră (modelele învață corecția ei)
  - $\mathbf{h}_7$ ultima stare după cele 7 zile anterioare, cîte un pas pe zi (24 de valori orare + ziua săptămînii)
  - $\mathbf{s}_{d+1}$ ziua săptămînii și sezonul zilei $d+1$
  - ARX prin regresie ridge (penalizarea 1, aleasă pe anul de validare)

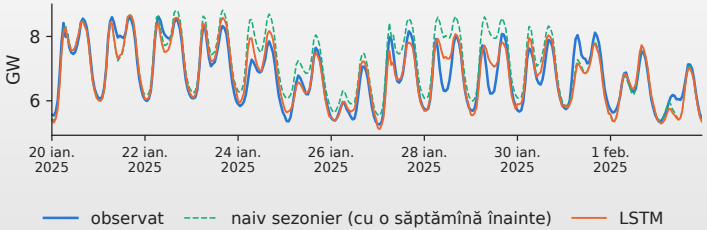
- **În practică:** un furnizor transmite cele 24 de volume orare înainte de prînz
  - Fiecare MW de eroare se cumpără sau se vinde la prețul de dezechilibru.

(calcul propriu)

Exercițiu practic: §5

Două săptămîni de test din iarnă: prognozele față de consumul observat

- **Pe scurt:** consumul orar în două săptămîni de test din iarnă, observat față de două prognoze.



Sursa: energy-charts.info (Fraunhofer ISE), din datele burselor și ale ENTSO-E. Date: 14 zile de test, 20 ianuarie 2025 - 2 februarie 2025.

- Prognoza naivă sezonieră copiază săptămîna trecută, inclusiv nivelul și sărbătorile ei
  - Rețelele învață doar o corecție a acesteia.

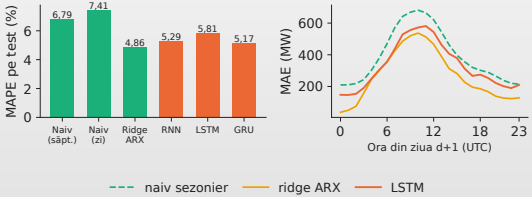
(calcul propriu)

NN_L5_load_fortnight

Pe consum, modelul liniar ARX depășește rețelele recurente

- **Pe scurt:** stînga: eroarea procentuală absolută medie (MAPE) pe toate orele de test; dreapta: eroarea absolută medie (MAE) pe ore ale zilei.

$$\mathrm{MAPE} = \frac{100}{24n} \sum_{d} \sum_{h=1}^{24} \frac{|y_{d,h} - \hat{y}_{d,h}|}{y_{d,h}}$$



Sursa: energy-charts.info (Fraunhofer ISE), din datele burselor și ale ENTSO-E. Date: 1 ianuarie 2024 - 31 octombrie 2026 (zile de test).

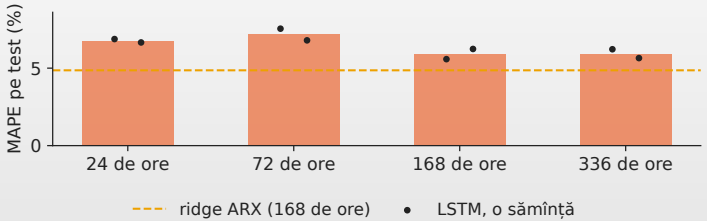
- MAPE (%): naiv sezonier 6,79, naiv 7,41, ridge ARX 4,86, RNN 5,29, LSTM 5,81, GRU 5,17.

(calcul propriu)

NN_L5_load_errors

Ferestrele mai lungi ajută LSTM, pînă la o săptămînă

⊡ **Pe scurt:** eroarea LSTM pe test în funcție de fereastra de intrare, cu același protocol în rest.



Sursa: energy-charts.info (Fraunhofer ISE), din datele burselor și ale ENTSO-E. Date: zile de test 2024-2026; LSTM cu 2 semințe pentru fiecare fereastră.

⊡ MAPE (%): 24 de ore 6,77, 72 de ore 7,17, 168 de ore 5,91, 336 de ore 5,93.

(calcul propriu)

NN_L5_load_window

Valoarea economică a unei prognoze mai precise a consumului

- **Semnificația**
  - Cea mai bună rețea după prognoza medie: GRU, MAPE cu $+6{,}4\,\%$ față de ridge ARX: cu 2 186 de zile de antrenare, modelul liniar cîștigă
  - DM pe MAE zilnic, LSTM față de ridge ARX: 9,80, $p < 0{,}001$
- **În megawați**
  - Eroarea absolută medie pe oră: naiv sezonier 396 MW, ridge ARX 269 MW, LSTM 331 MW
  - LSTM în locul prognozei naive sezoniere: cu aproximativ 1 560 megawați-oră (MWh) mai puțină eroare absolută pe zi, la scară națională
- **Costul**
  - Antrenarea tuturor modelelor de consum și studiul ferestrei: 67 de secunde pe procesorul central al unui laptop

- **În practică:** înmulțiți MWh economisiți cu diferența de preț de dezechilibru și comparați cu costul modelului: aici ridge ARX economisește cel mai mult cu cel mai mic cost
  - Variabilele meteo ar aduce un plus (cursul 13).

(calcul propriu)

Test 1: parametrii

- Cîți parametri are un strat LSTM cu $d = 3$ intrări și $n_h = 32$ de unități?

Răspuns

- $4n_h(n_h + d + 1) = 4 \cdot 32 \cdot 36 = 4608$ (PyTorch păstrează doi vectori de termeni liberi, deci raportează $4608 + 128 = 4736$).

Test 2: gradienții

- De ce are o RNN simplă dificultăți cu tiparul săptămînal al consumului orar?

Răspuns

- Pe 168 de pași orari, gradientul este un produs de 168 de matrice jacobiene și dispare
  - Citirea a 7 profiluri zilnice scurtează drumul la 7 pași, iar așa sînt construite modelele noastre de consum.

Test 3: afirmația unui furnizor

- Un furnizor prezintă un LSTM cu acuratețe de 60 % a semnului randamentelor S&P 500, antrenat pe zile amestecate
  - Ce întrebați?

Răspuns

- Un test în ordine cronologică, pe ani nevăzuți, față de prognoza "mereu în sus" (54,0 %) și după costuri
  - LSTM-ul nostru, evaluat corect, atinge 52,1 %.

Test 4: HAR sau LSTM

- LSTM schimbă MSE a volatilității cu $-11{,}6\,\%$ față de HAR, cu DM $= -5{,}52$
  - Înlocuiește HAR în sistemul VaR?

Răspuns

- Cîștigul este semnificativ aici, deci LSTM este un candidat serios
  - Schimbați modelul doar după ce cîștigul se confirmă pe altă piață și pe subperioade, cu monitorizare și cu HAR ca soluție de rezervă.

Test 5: bidirecțional

- Puteți folosi un LSTM bidirecțional pentru a prognoza consumul de mîine dintr-o fereastră care se termină azi?

Răspuns

- Da, dacă fereastra se termină în momentul prognozei: trecerea înapoi vede atunci doar trecutul
  - Apare o scurgere doar dacă fereastra conține ore de după acel moment.

Recapitulare

- Secvențe în finanțe și energie
- Rețeaua recurentă și propagarea înapoi în timp
- Celule cu porți: LSTM și GRU
- Un protocol pentru serii de timp financiare
- Studiul de caz A: volatilitatea S&P 500
- Studiul de caz B: prognoza randamentelor cu rețele neuronale
- Studiul de caz C: consumul de energie electrică din România

- O RNN partajează ponderile în timp
  - BPTT înmulțește matrice jacobiene, deci gradienții dispar sau explodează
  - LSTM și GRU adaugă un drum aditiv al memoriei.
- Pe date reale: rețelele depășesc HAR pe volatilitate, prognozele randamentelor nu explică aproape nimic, iar un ARX liniar cîștigă pe consumul orar
  - Urmează textul (cursul 6).

Cercetări ale grupului nostru: secvențe și risc

- **Pot gestiona riscul modelele fundaționale?** (Jheng, Tak, Găman, Mazurencu-Marinescu-Pele & Pele, 2026, Economic Computation and Economic Cybernetics Studies and Research)
  - Modele preantrenate pentru serii de timp, folosite fără reantrenare pentru prognoza VaR și a pierderii medii peste VaR (ES) pe piețele din Europa Centrală și de Est (CEE)
  - O calibrare conformală face fiabile prognozele de coadă; succesorii transformer ai LSTM sînt prezentați în cursurile 7 și 13
- **Avertizarea timpurie pentru criptoactivele "zombie"** (Będowska-Sójka, Wójcik & Pele, 2026, North American Journal of Economics and Finance)
  - Un sistem de avertizare timpurie bazat pe învățare automată pentru criptoactivele care își pierd activitatea de piață
- **BitMood** (Conda, Găman, Bâg, Mazurencu-Marinescu-Pele, Pele & Härdle, 2026, Digital Finance)
  - Sentimentul de pe Facebook și volumul Bitcoin: regresia liniară și un LSTM bidirecțional au performanțe comparabile pe test, deci complexitatea modelului nu garantează prognoze mai bune
- Rezultatele de pe acest diapozitiv sînt cele raportate în lucrări; celelalte cifre din curs provin din calculele noastre.

(Academia de Studii Economice din București; Academia Română, Institutul de Prognoză Economică)

Lecturi și resurse suplimentare

- **Bishop & Bishop (2024)**, cap. 12: modele de secvențe, de la RNN la transformere.
- **Pascanu, Mikolov & Bengio (2013)**: cea mai clară analiză a gradienților care dispar sau explodează.
- **Corsi (2009)**: modelul HAR, reperul cu care trebuie să se măsoare orice rețea de volatilitate.
- **Acest curs**: cod, date, notebook-uri Jupyter — github.com/danpele/NEURAL_BIZ.
- **Exercițiu practic**: notebook-ul reproduce graficele acestui curs — notebooks/.
- **Seminarul 5**: un GRU pentru volatilitatea altui indice, față de HAR, cu testul DM.

Bibliografie (1/2)

- Bengio, Y., Simard, P. & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. doi:10.1109/72.279181
- Bishop, C. M. & Bishop, H. (2024). *Deep Learning: Foundations and Concepts*, ch. 12. Springer. doi:10.1007/978-3-031-45468-4
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. doi:10.1016/0304-4076(86)90063-1
- Będowska-Sójka, B., Wójcik, P. & Pele, D. T. (2026). Early warning systems for cryptocurrency markets: Predicting 'zombie' assets using machine learning. *North American Journal of Economics and Finance*, 81, 102543. doi:10.1016/j.najef.2025.102543
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. *EMNLP*, 1724–1734. doi:10.3115/v1/D14-1179
- Conda, A. I., Găman, Ș., Bâg, R. C., Mazurencu-Marinescu-Pele, M., Pele, D. T. & Härdle, W. K. (2026). BitMood: AI analysis of Bitcoin trends via Facebook emotions. *Digital Finance*, 8(3), 42. doi:10.1007/s42521-026-00205-3
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196. doi:10.1093/jjfinec/nbp001
- Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. doi:10.1080/07350015.1995.10524599

Bibliografie (2/2)

- Garman, M. B. & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1), 67–78. doi:10.1086/296072
- Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Jheng, S.-L., Tak, R., Găman, Ș., Mazurencu-Marinescu-Pele, M. & Pele, D. T. (2026). Can foundation models manage risk? Zero-shot VaR and ES forecasting with conformal calibration in CEE markets. *Economic Computation and Economic Cybernetics Studies and Research*, 60(2), 267–290. doi:10.24818/18423264/60.2.26.14
- Pascanu, R., Mikolov, T. & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *ICML*, PMLR 28(3), 1310–1318. arXiv:1211.5063
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. doi:10.1038/323533a0
- Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550–1560. doi:10.1109/5.58337

Anexă: BPTT pentru RNN simplă, pas cu pas

- Preactivarea $\mathbf{a}_t = \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h$, starea $\mathbf{h}_t = \tanh(\mathbf{a}_t)$, pierderea $\mathcal{L} = \sum_t \ell_t(\hat{y}_t, y_t)$.
- Recurența înapoi: $\boldsymbol{\delta}_t \equiv \partial\mathcal{L}/\partial\mathbf{h}_t = \mathbf{W}_{hy}^\top \partial\ell_t/\partial\hat{y}_t + \mathbf{W}_{hh}^\top \left[(1 - \mathbf{h}_{t+1}^2) \odot \boldsymbol{\delta}_{t+1}\right]$, cu $\boldsymbol{\delta}_{T+1} = \mathbf{0}$.
- Gradienții ponderilor: $\partial\mathcal{L}/\partial\mathbf{W}_{hh} = \sum_t \left[(1 - \mathbf{h}_t^2) \odot \boldsymbol{\delta}_t\right] \mathbf{h}_{t-1}^\top$, $\partial\mathcal{L}/\partial\mathbf{W}_{xh} = \sum_t \left[(1 - \mathbf{h}_t^2) \odot \boldsymbol{\delta}_t\right] \mathbf{x}_t^\top$.
- Desfășurarea recurenței dă produsul de matrice jacobiene din textul principal
  - La modelele many-to-one, $\partial\ell_t/\partial\hat{y}_t = 0$ pentru $t < T$.

Anexă: modelele în PyTorch

```
class RecNet(nn.Module):
    def __init__(self, cell, n_in, hidden=32, n_out=1, n_static=0):
        super().__init__()
        self.rnn = {'rnn': nn.RNN, 'lstm': nn.LSTM, 'gru': nn.GRU}[cell](n_in, hidden, batch_first=True)
        self.out = nn.Linear(hidden + n_static, n_out)
    def forward(self, x, s=None):                # x: batch x time x features
        h, _ = self.rnn(x)
        z = h[:, -1] if s is None else torch.cat([h[:, -1], s], 1)
        return self.out(z)
# training step: loss.backward(); nn.utils.clip_grad_norm_(model.parameters(), 1.0); opt.step()
```

- Codul complet: lectures/L5/analysis_L5.py și notebook-ul acestui curs.