



**RESURSE EDUCAȚIONALE INTERACTIVE ȘI INOVATIVE  
DEDICATE DEZVOLTĂRII ABILITĂȚILOR PRACTICE ALE STUDENȚILOR  
ÎN DOMENIUL ANALIZEI DATELOR**

ATELIER DE INSTRUIRE APLICATIVĂ

- SINTEZĂ -

ASE

CENTRUL DE PREGĂTIRE PROFESIONALĂ COVASNA

PERIOADA 06 – 09 SEPTEMBRIE 2023



**ACADEMIA DE STUDII  
ECONOMICE DIN BUCUREȘTI**

CNFIS-FDI-2023-F-0159  
DEZVOLTAREA INOVATIVĂ A  
COMPETENȚELOR PRACTICE ȘI  
DIGITALE ALE STUDENȚILOR  
PENTRU INTEGRAREA ÎNTR-UN  
MEDIU DE AFACERI CARACTERIZAT  
DE INCERTITUDINE, BAZAT PE  
AGILITATE, CONECTIVITATE ȘI  
REZILIENȚĂ

DIRECTOR PROIECT: PROF. UNIV. DR.  
DOREL MIHAI PARASCHIV

PERIOADA DE DESFĂȘURARE  
APRILIE – DECEMBRIE 2023



# Resurse educaționale interactive și inovative dedicate dezvoltării abilităților practice ale studenților în domeniul analizei datelor

Competențele studenților ASE în domeniul Data Science

Experiența Datathon 2023 / Provocarea Datathon 2024

Utilizarea tehnicilor avansate de Web Scraping, curățare a datelor și Machine Learning pentru prognozele prețurilor pe piața imobiliară

Utilizarea analizei econometrice de tip panel pentru analiza mișcării migratorii

Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate în evaluarea dinamicii sistemului de învățământ superior și impactul pe piața muncii

Gamificarea - Drumul Inovației Educaționale

Explorarea complexității datelor: O taxonomie a modelelor (2 materiale)

Perspective inovatoare asupra interdependenței dintre inflație și nivelul de trai utilizând regresia logistică (metoda FORWARD CONDITIONAL)

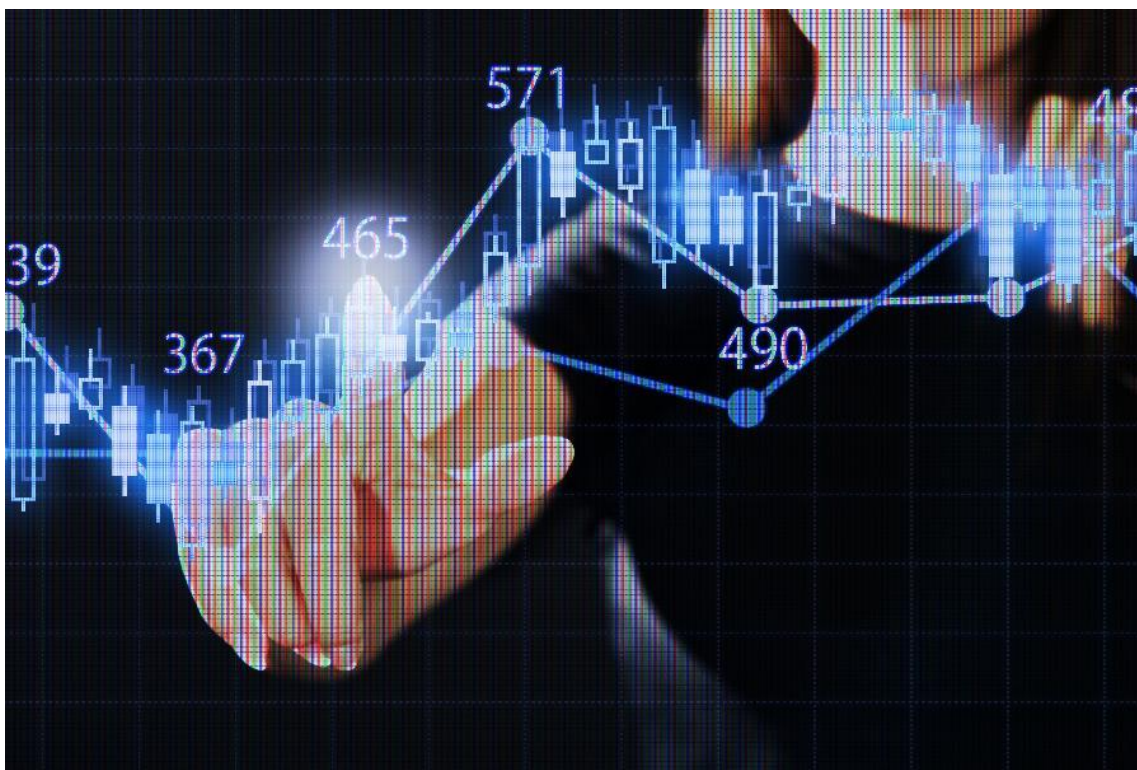
Analiza fluxurilor migratorii: abordarea modelului gravitațional în contextul utilizării Machine Learning

Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

Explorarea percepțiilor tinerilor despre sistemul educațional prin analiză avansată a datelor în contextul științelor sociale

## CNFIS-FDI-2023-F-0159

Dezvoltarea inovativă a competențelor practice și digitale ale studenților pentru integrarea într-un mediu de afaceri caracterizat de incertitudine, bazat pe agilitate, conectivitate și reziliență



### Obiectivul general:

Consolidarea pregătirii profesionale a studenților ASE în scopul creșterii capacității de integrare a acestora pe piața muncii, în contextul provocărilor de agilitate, conectivitate și reziliență.

- O1. Dezvoltarea și consolidarea abilităților practice și digitale ale studenților ASE prin intermediul unor instrumente educaționale inovative ce presupun integrarea cunoștințelor pluridisciplinare în simularea comportamentului antreprenorial, utilizând facilitățile infrastructurii educaționale modernizate în cadrul Centrelor de Pregătire Profesională ale ASE din Predeal, Covasna și București.

# Competențele studenților ASE în domeniul Data Science

## Experiența DATATHON 2023 / Provocarea DATATHON 2024

- Utilizarea de resurse educaționale interactive și inovative pentru dezvoltarea abilităților practice ale studenților în domeniul analizei datelor reprezintă o componentă esențială în dezvoltarea competențelor studenților ASE în Data Science.
- Resurse interactive pot include platforme de învățare online, simulări, aplicații de tip gaming și alte instrumente digitale care facilitează înțelegerea și aplicarea practică a conceptelor de analiză a datelor.
- Resurse interactive sunt concepute pentru a oferi o experiență de învățare captivantă și interactivă, punând accentul pe rezolvarea problemelor și pe aplicarea practică a cunoștințelor.

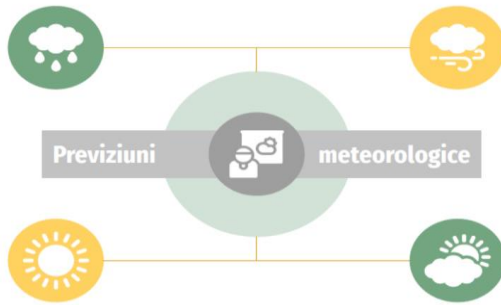
# Competențele studenților ASE în domeniul Data Science

## Experiența DATATHON 2023 / Provocarea DATATHON 2024

### TEMATICA WiDS 2023

#### Provocare globală

Evenimentele meteorologice extreme devin din ce în ce mai frecvente la nivel mondial.



#### Previziuni

Previzionarea intensității fenomenelor extreme este esențială.

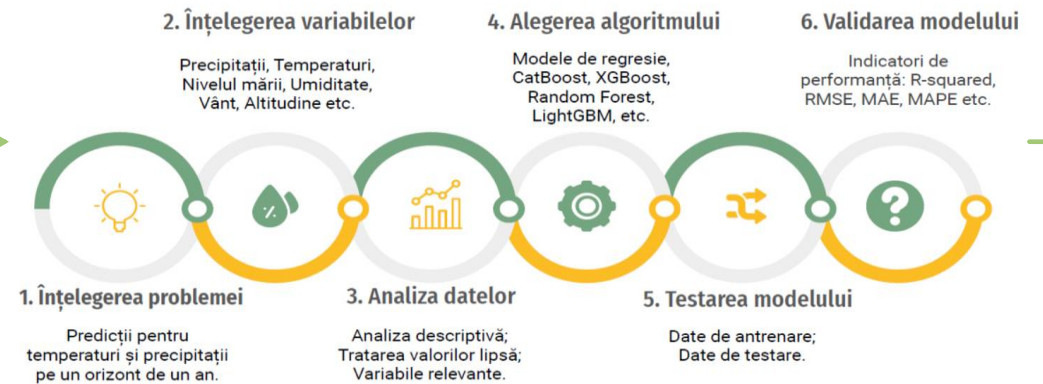
#### Fenomene extreme

Valuri de căldură, incendii, secetă, precipitații extreme și inundații.

#### Provocare WiDS 2023

Predicții pentru temperaturile și precipitațiile din SUA pe un orizont de un an.

### PAȘII GENERALI



#### SOLUȚIA

1 Analiza exploratorie a datelor

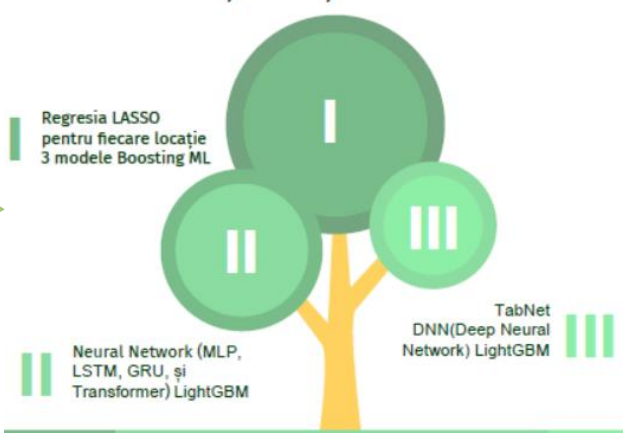
2 Imputarea valorilor lipsă pe baza formulor de calcul NME (North American Multi-Model Ensemble)

3 Împărțirea setului de date în subseturi de antrenare și testare.

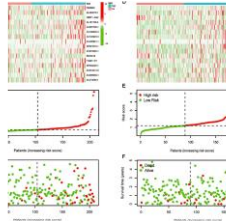
#### SOLUȚIA

- 4 Excluderea variabilelor redundante  
Pe baza matricei de corelație ( $r > 0.85$ )
- 5 Procesarea datelor  
One-Hot Encoding, MinMax Scaler, Crearea unor noi variabile
- 6 Utilizarea algoritmilor ML  
CatBoost, XGBoost, Random Forest, LightGBM, TabNet
- 7 Optimizarea hiper-parametrilor  
Grid Search, Random Search, Optuna, Bayesian Optimization
- 8 Identificarea celui mai bun model  
Obținerea unui RMSE minim pe setul de antrenare
- 9 Asamblarea unui model pe baza celor mai performante  
Medie ponderată

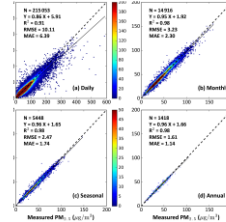
#### SOLUȚIILE CÂȘTIGĂTOARE



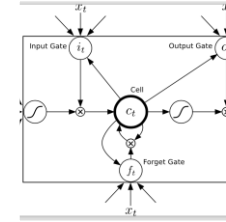




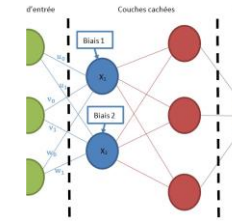
Regresia Lasso (Least Absolute Shrinkage and Selection Operator) este o tehnică de regresie utilizată în învățarea automată pentru a gestiona problema suprapunerii (overfitting) și pentru selecția automată a caracteristicilor. Ea face parte din familia de metode de regularizare, alături de alte tehnici precum regresia Ridge și Elastic Net.



LightGBM (Light Gradient Boosting) este un algoritm de învățare automată care se concentrează pe viteza de calcul și pe performanța modelului. Algoritmul poate fi utilizat atât pentru probleme de regresie, cât și pentru clasificare și a fost conceput pentru a funcționa cu seturi de date complexe.



LSTM (Long Short-Term Memory) este un tip de rețea neurală recurentă (RNN) care a fost dezvoltat pentru a gestiona problema dispariției gradientilor (vanishing gradients) în RNN-urile tradiționale. Acesta creează un modul suplimentar într-o rețea neuronală care învață când să-și amintească și când să uite informațiile pertinente.



În aceeași idee, MLP (Multi-Layer Perception) este o rețea neurală feedforward și constă în trei componente principale: stratul de intrare, stratul(e) ascuns și stratul de ieșire.

Competențele studenților ASE în domeniul Data Science  
 Experiența Datathon 2023 / Provocarea Datathon 2024

# Utilizarea tehnicilor avansate de Web Scraping, curățare a datelor și Machine Learning pentru prognozele prețurilor pe piața imobiliară

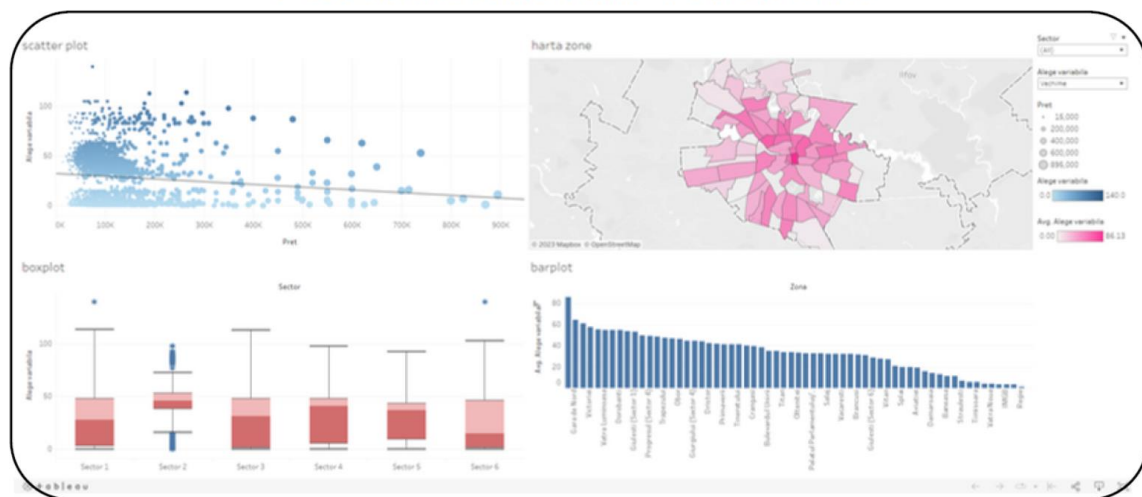
- Piețele imobiliare constituie un element fundamental în cadrul oricărei economii și înțelegerea dinamicii acestora reprezintă un aspect crucial pentru factorii de decizie, investitori și profesioniștii din domeniul imobiliar.
- Prezentarea resurselor educaționale inovative pentru a susține investigarea și înțelegerea profundă a aspectelor complexe ale pieței imobiliare referitoare la factorii determinanți pentru vânzarea unui imobil, beneficiile și caracteristicile zonei pentru vânzători, distribuția spațială a variabilelor pe piața imobiliară din zona vizată și modelele eficiente de învățare automată pentru prezicerea prețurilor locuințelor





# Utilizarea tehnicilor avansate de Web Scraping, curățare a datelor și Machine Learning pentru prognozele prețurilor pe piața imobiliară

## Dashboard în Tableau



Resursele educaționale inovative propuse oferă o abordare practică și interactivă pentru a contribui la analiza pieței imobiliare integrează diverse metode și instrumente tehnologice esențiale în procesul de colectare, curățare, vizualizare și analiză a datelor. Acestea integrează:

- utilizarea tehnicilor de web scraping pentru extragerea datelor relevante de pe diverse platforme online.
- utilizarea metodelor de curățare a datelor este pentru asigurarea calității și coerenței informațiilor, pregătind astfel datele pentru analiză.
- vizualizarea datelor în aplicații (Geoda, Tableau etc.) pentru a oferi o perspectivă intuitivă și accesibilă asupra tendințelor și distribuției spațiale a informațiilor din piața imobiliară.
- analiza de text pentru interpretarea și înțelegerea recenziilor sau descrierilor referitoare la proprietățile imobiliare. Această abordare permite extragerea de informații relevante din texte și identificarea tendințelor sentimentelor asociate cu diferite zone sau tipuri de proprietăți.
- utilizarea modelelor de Machine Learning oferă studenților (profesioniștilor imobiliari – în practică) instrumentele necesare pentru a prezice prețurile locuințelor și pentru a realiza analize mai avansate pe baza datelor disponibile.

### Analiza de Text

#### Despre Date:

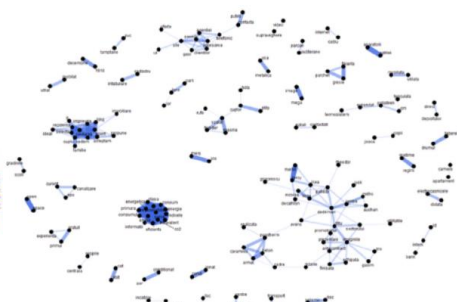
- 12 158 de date nestructurate
- reprezentate de descrierile textuale oferite de vânzatori în anunțurile publicate pe Publi24

#### Tehnologie:

tidytext, în R

#### Metode:

- analiza Word Cloud
- analiza perechilor celor mai comune de cuvinte
- analiza rețelelor de cuvinte



### Rețeaua de cuvinte

Combinările de cuvinte cele mai corelate sunt:

- indicele-consum-energetica-eficienta-regenerabilaechivalent-co2-surse-primaraclassa
- rezidenta-selecta-ideal-cuplu-familie-cautarea
- king-immobiliare-agent-propune-asteptam-li-vedem-impreuna
- jpsi-dedeman-jumbo-metro-suchan,theodor-pallady-ikea-merlin-leroy-decathlon-nicolae-teclu,park-lake-ior

Prin integrarea acestor metode și tehnologii într-un cadru educațional, se oferă oportunitatea de a înțelege și aplica conceptele teoretice într-un context practic și actual din piața imobiliară.

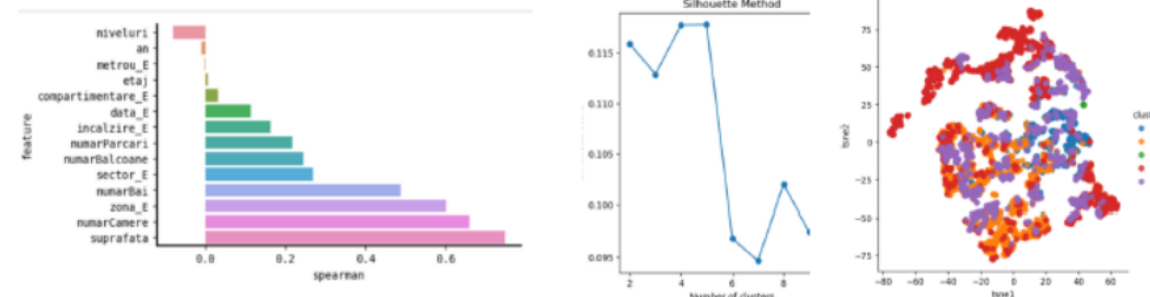
# Utilizarea tehnicilor avansate de Web Scraping, curățare a datelor și Machine Learning pentru prognozele prețurilor pe piața imobiliară

Prezentarea metodelor de analiza multivariată utilizate precum și a modelului de Machine Learning testat **Extreme Gradient Boosting (XGBoost)**.

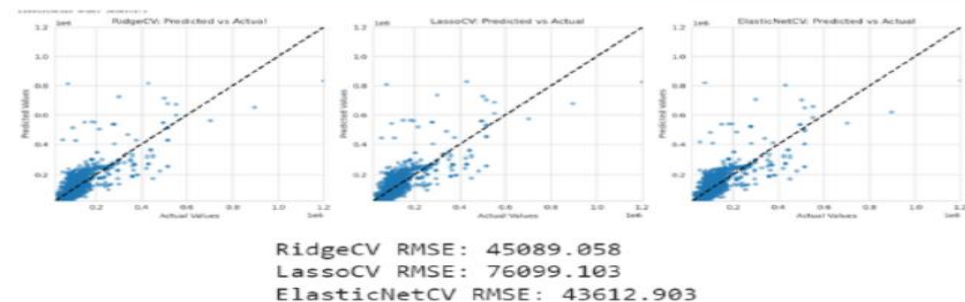
**XGBoost** - reprezintă o implementare a tehnologiei de boosting a arborilor de decizie și este recunoscut pentru performanțele sale remarcabile în predicția variabilelor țintă, în special în problemele de regresie și clasificare.

- este eficient în identificarea factorilor determinanți pentru prețurile imobiliare, ajutând la înțelegerea mai profundă a pieței imobiliare și a tendințelor acesteia. Prin analiza datelor imobiliare poate evidenția importanța diferitelor caracteristici ale proprietăților și poate determina modul în care acestea influențează prețurile.
- poate extrage caracteristici relevante din datele imobiliare (dimensiunea, localizarea, dotările și alte factori care influențează prețurile, contribuind astfel la o evaluare mai exactă a valorii proprietăților)

## Construirea modelelor de machine learning



## Modele de Machine learning



Exemplu: Apartament din 1978, 55m2, fara centrala proprie, Unirii. Pret previzionat - 80000  
Pret adevarat - 78307

# Utilizarea analizei econometrice de tip panel pentru analiza mișcării migratorii

## ANALIZA ECONOMETRICĂ DE TIP PANEL A NUMĂRULUI DE EMIGRANȚI DEFINITIVI

1996 -----2021 – INS – Tempo online

### Testul F pentru efecte fixe

DF	Den DF	F Value	Pr > m
41	1042	16.76	<.0001

Sursa: Prelucrare proprie în programul SAS

$H_0$ : Nu există efecte fixe individuale ( $\alpha_1 = \alpha_2 = \dots = \alpha_N = \alpha$ )

$H_1$ : Există  $i$  și  $j, i \neq j$  astfel încât  $\alpha_i \neq \alpha_j$



### Testul Breush-Pagan pentru efecte aleatoare

DF	m	Pr > m
1	1133.78	<.0001

Sursa: Prelucrare proprie în programul SAS

$H_0$ : Nu există efecte aleatoare ( $\sigma_u^2 = 0$ )

$H_1$ : Există efecte aleatoare ( $\sigma_u^2 > 0$ )



## ANALIZA ECONOMETRICĂ DE TIP PANEL A NUMĂRULUI DE EMIGRANȚI DEFINITIVI

2013 -----2021 – INS – Tempo online

$H_0$ : Nu există efecte fixe individuale ( $\alpha_1 = \alpha_2 = \dots = \alpha_N = \alpha$ )

$H_1$ : Există  $i$  și  $j, i \neq j$  astfel încât  $\alpha_i \neq \alpha_j$

### Testul F pentru efecte fixe

Num DF	Den DF	F	Pr > F
41	326	16.26	<.0001

Sursa: Prelucrare proprie în programul SAS

$H_0$ : Nu există efecte aleatoare ( $\sigma_u^2 = 0$ )

$H_1$ : Există efecte aleatoare ( $\sigma_u^2 > 0$ )



### Testul Breush-Pagan pentru efecte aleatoare

DF	m	Pr > m
1	339.29	<.0001

Sursa: Prelucrare proprie în programul SAS

Utilizarea analizei econometrice de tip panel pentru analiza mișcării migratorii reprezintă o abordare robustă și sofisticată, ce permite evaluarea și înțelegerea în profunzime a factorilor care influențează migrația într-o anumită regiune sau țară. Această metodă se distinge prin capacitatea sa de a examina comportamentul migrației pe multiple perioade de timp și în diverse unități geografice / sociale, oferind astfel o perspectivă completă asupra fenomenului de migrație.

Pe baza factorilor socio-economici precum nivelul salariilor, rata șomajului, condițiile economice, educația, infrastructura, dar și factori culturali sau politici, prin utilizarea analizei econometrice de tip panel, se pot identifica și evalua mai precis factorii care influențează deciziile de migrație.

Abordarea analizei pe date de tip panel, în analiza migrației permite:

- urmărirea evoluției în timp a migrației în diferite regiuni sau grupuri demografice. Astfel, se pot identifica particularitățile specifice și modul în care acestea influențează mișcarea migratorie.
- evaluarea impactului factoriilor externi (variabilelor explicative) asupra mișcării migratorii și a schimbărilor de tendințe în timp.
- gestionarea problemelor de autocorelație și heterogenitate, oferind estimări mai robuste și mai precise ale relațiilor dintre variabile.



# Utilizarea analizei econometrice de tip panel pentru analiza mișcării migratorii

## MODELUL CU EFECTE FIXE VS CEL CU EFECTE ALEATOARE

Statisticile modelului cu efecte fixe

SSE	154269.5168	DFE	326
MSE	473.2194	Root MSE	21.7536
R-Square	0.8252		

Sursa: Prelucrare proprie în programul SAS

Statisticile modelului cu efecte aleatoare

SSE	167114.0947	DFE	367
MSE	455.3518	Root MSE	21.3390
R-Square	0.5738		

Sursa: Prelucrare proprie în programul SAS

### Testul Hausman pentru efecte aleatoare

Coeficient	DF	m	Pr > m
10	10	18.95	0.0408

Sursa: Prelucrare proprie în programul SAS

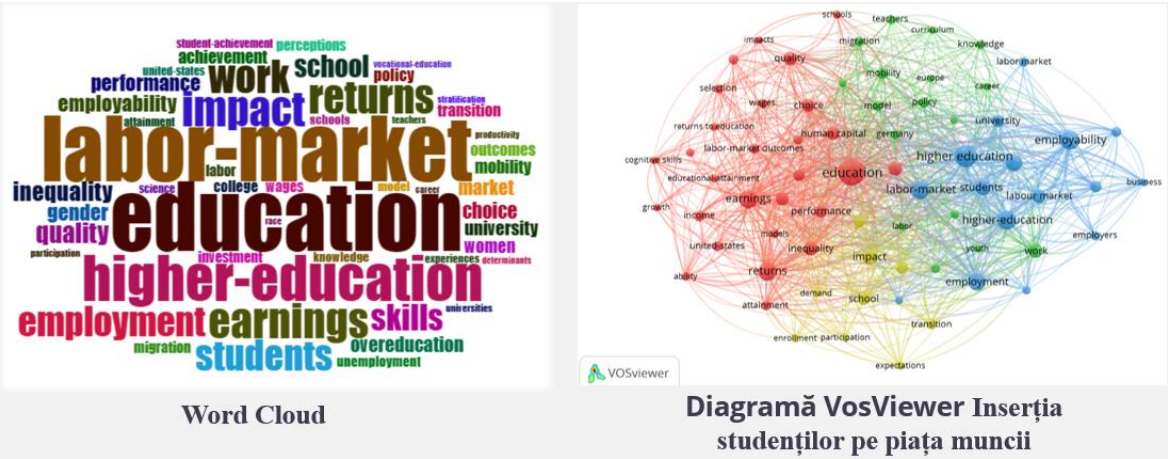
$H_0$ : Nu există corelație: FE consistent, RE consistent și eficient ( $E(x_i\alpha_i) = 0$ )

$H_1$ : Există corelație: FE consistent, RE inconsistent ( $E(x_i\alpha_i) \neq 0$ )

Această abordare inovatoare în analiza mișcării migratorii este esențială pentru elaborarea politicilor publice, pentru înțelegerea schimbărilor sociale și economice și pentru anticiparea tendințelor viitoare în ceea ce privește migrația.

Este un instrument necesar pentru luarea deciziilor informate în domeniul gestionării migrației și al dezvoltării socio-economice.

# Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate în evaluarea dinamicii sistemului de învățământ superior și impactul pe piața muncii

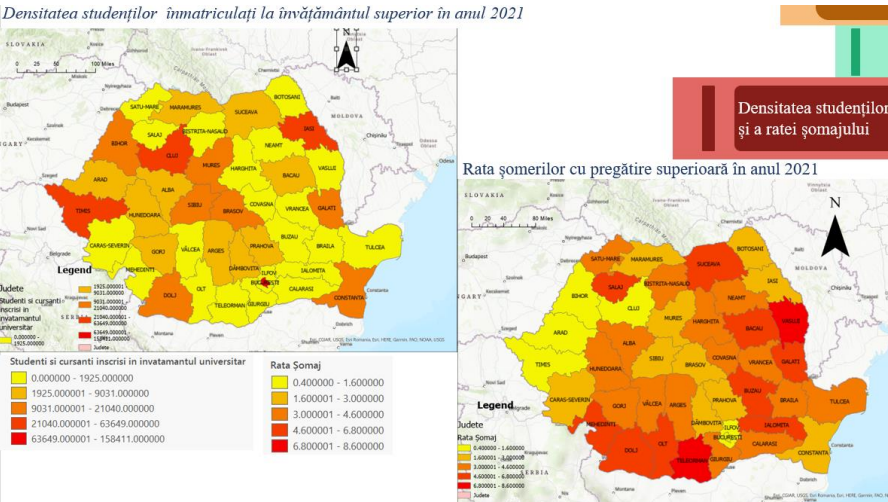


Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate reprezintă un demers complex și cuprinzător pentru a investiga dinamica sistemului de învățământ superior și impactul său asupra pieței muncii.

Această abordare permite o analiză detaliată și precisă a relației dintre evoluția numărului de studenți în perioada 2007-2021, ratele de șomaj în rândul absolvenților și calitatea pregătirii lor academice.

Vizualizarea avansată a datelor oferă o perspectivă clară și intuitivă asupra tendințelor și schimbărilor din sistemul de învățământ superior și din piața muncii de-a lungul timpului.

Aceste tehnici permit identificarea modelelor, a fluctuațiilor și a corelațiilor între diferitele variabile relevante, precum numărul de absolvenți, nivelul de ocupare a forței de muncă și calitatea pregătirii acestora.



# Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate în evaluarea dinamicii sistemului de învățământ superior și impactul pe piața muncii

## Ipoteza 1

Abilitățile practice sunt mai apreciate decât abilitățile teoretice

	<i>Cât de importante sunt următoarele aspecte care ar putea influența succesul unui absolvent ASE la integrarea pe piața forței de muncă? [abilitățile teoretice]</i>	<i>Cât de importante sunt următoarele aspecte care ar putea influența succesul unui absolvent ASE la integrarea pe piața forței de muncă? [abilitățile practice]</i>
Mean	1.816816817	3.501501502
Variance	0.746463331	0.564003763
Observations	333	333
df	332	332
F	1.323507714	
P(F<=f) one-tail	0.005425727	
F Critical one-tail	1.198181468	

## Ipoteza 2

Programul de studii dezvoltă semnificativ mai mult cunoștințe fundamentale decât deprinderi practice.

	<i>În ce măsură credeți că programul de studii pe care îl urmați vă ajută să vă dezvoltați următoarele? [Cunoștințe fundamentale]</i>	<i>În ce măsură credeți că programul de studii pe care îl urmați vă ajută să vă dezvoltați următoarele? Deprinderi practice]</i>
Mean	3.645645646	2.310240964
Variance	1.145139115	1.695000546
Observations	333	332
Hypothesized Mean Difference	0	
df	638	
t Stat	14.44691676	
P(T<=t) one-tail	1.95095E-41	
t Critical one-tail	1.647245473	
P(T<=t) two-tail	3.9019E-41	
t Critical two-tail	1.963689222	

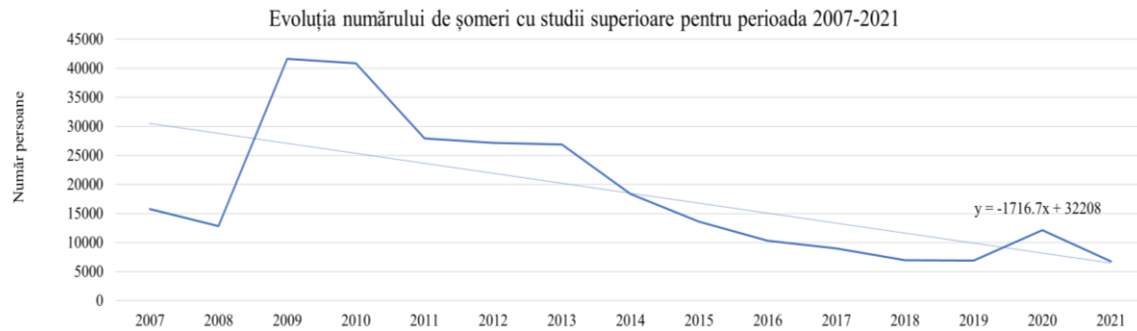
## Ipoteza 3

Persoanele din mediul Rural acordă o importanță mai sporită mediului de proveniență când vine vorba de influențele succesului unui absolvent al Academiei de Studii Economice din București la integrarea pe piața forței de muncă.

	<i>Mediul de proveniență</i>	<i>Cât de de sunt următoarele aspecte care ar putea influența succesul unui absolvent ASE la integrarea pe piața forței de muncă? [mediul de proveniență]</i>
Mean	1.394594595	2.183783784
Variance	0.240188014	0.770387779
Observations	185	185
Hypothesized Mean Difference	0	
df	289	
t Stat	-10.67781878	
P(T<=t) one-tail	5.90549E-23	
t Critical one-tail	1.650143229	
P(T<=t) two-tail	1.1811E-22	
t Critical two-tail	1.968206436	

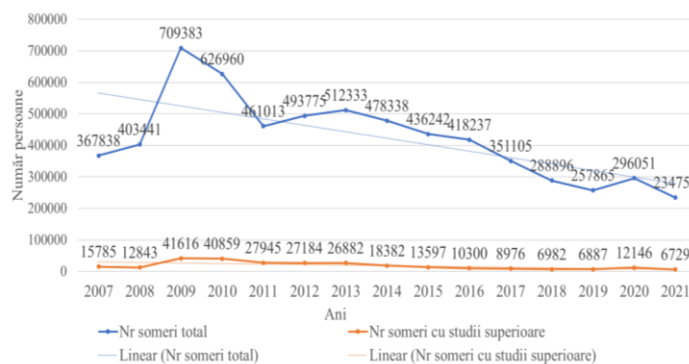


# Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate în evaluarea dinamicii sistemului de învățământ superior și impactul pe piața muncii

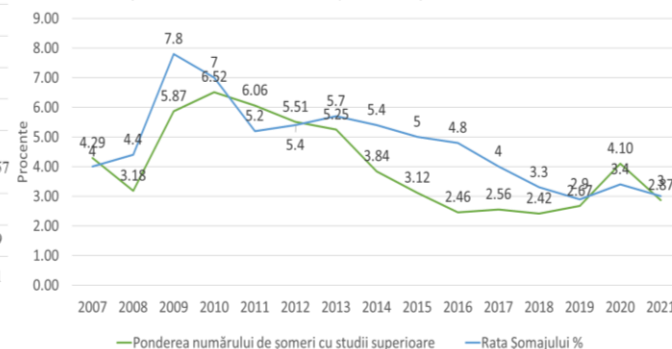


Utilizarea metodelor neparametrice avansate oferă posibilitatea de a analiza datele fără a fi nevoie să se presupună o anumită distribuție a acestora, ceea ce aduce o mai mare flexibilitate și precizie în evaluarea impactului sistemului de învățământ superior asupra pieței muncii. Aceste metode pot evidenția relații complexe și interdependențe între variabilele implicate, contribuind la înțelegerea mai profundă a dinamicii acestui sistem complex.

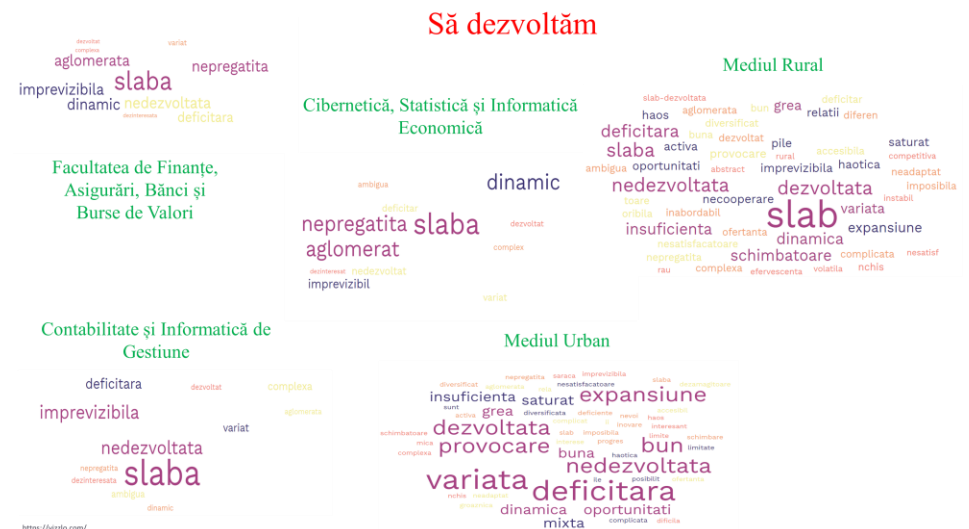
Evoluția comparativă a șomerilor cu studii superioare și total șomeri România din perioada 2007-2021



Rata și ponderea numărului de șomeri cu studii superioare în raport cu numărul total de șomeri în perioada 2007-2021



Sursă: [INSSE TEMPO Online \(insse.ro\)](https://insse.ro)



# Integrarea tehnicilor avansate de vizualizare a datelor și a metodelor neparametrice avansate în evaluarea dinamicii sistemului de învățământ superior și impactul pe piața muncii



Test Statistics<sup>a</sup>

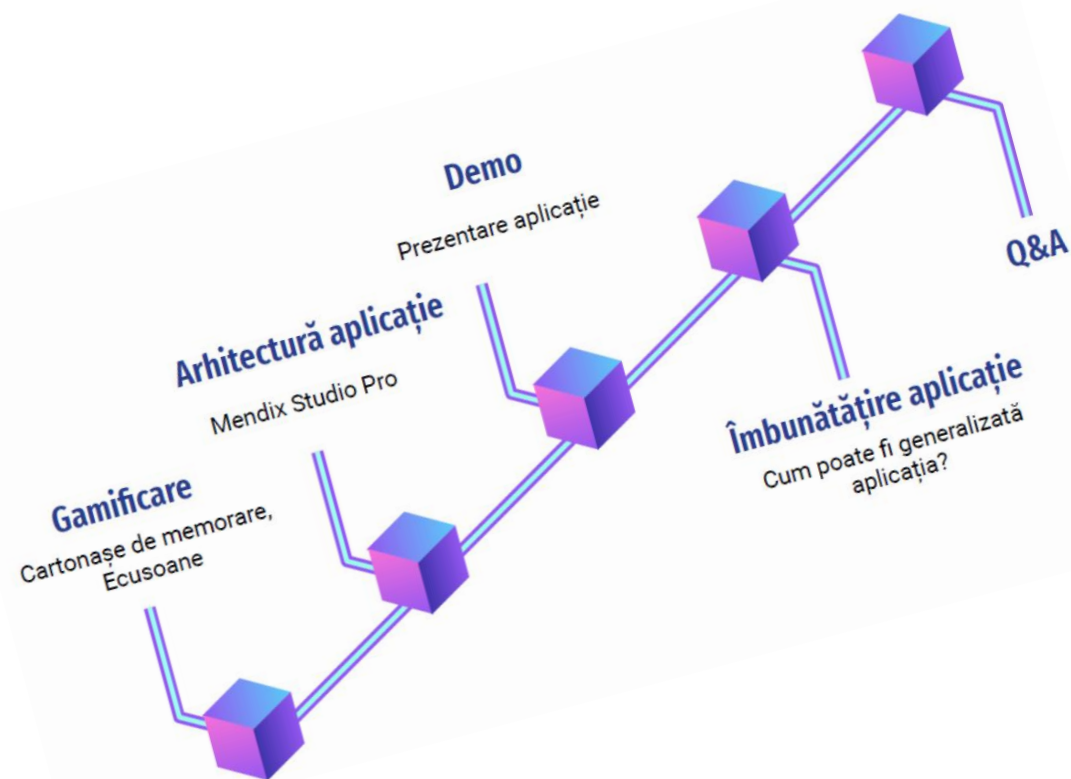
	Considerați că pe piața muncii sunt suficiente locuri de muncă în domeniul în care vă aflați?
Mann-Whitney U	1305.500
Wilcoxon W	2580.500
Z	-3.553
Asymp. Sig. (2-tailed)	.000



Există diferențe semnificative între studenții de la facultatea de Cibernetică, Statistică și Informatică Economică și Contabilitate când vine vorba de existența locurilor de muncă pe piața muncii.

	In cadrul căreia facultate ați studiat/ studiați?	N	Mean Rank	Sum of Ranks
Considerați că pe piața muncii sunt suficiente locuri de muncă în domeniul în care vă aflați?	CSIE	81	74.88	6065.50
	CIG	50	51.61	2580.50
	Total	131		

# Gamificarea - Drumul inovației educaționale



Gamificarea în cadrul procesului educațional reprezintă aplicarea principiilor și mecanismelor caracteristice jocurilor în contextul pedagogic, având ca scop stimularea și îmbunătățirea motivației intrinseci și extrinseci a elevilor, precum și optimizarea procesului de învățare.

Acest concept vizează transformarea paradigmei tradiționale a învățământului printr-o abordare mai dinamică, interactivă și captivantă folosind tehnologii actuale precum aplicații mobile, platforme online sau software specializat



# Gamificarea - Drumul inovației educaționale

Platformele de dezvoltare low-code (de ex. Mendix Studio Pro) oferă un spațiu de lucru intuitiv și accesibil pentru crearea rapidă a aplicațiilor software.

Conceptul se axează pe utilizarea unui mediu de dezvoltare centrat în jurul unui domeniu de model, ceea ce înseamnă că utilizatorii se concentrează mai mult pe definirea și proiectarea unui model conceptual al aplicației decât pe scrierea codului în sine.

Prin acest domeniu de model, utilizatorii pot:

- defini obiectele și structura datelor;
- configura logica aplicației;
- personaliza interfața utilizatorului și interacțiunea acestuia cu aplicația;
- gestiona și integra datele din diferite surse.

**Mendix Studio Pro**   
\*Versiune 9.7.0

Spațiu de dezvoltare low-code în  
jurul unui domeniu de model



Microflow



Nanoflow

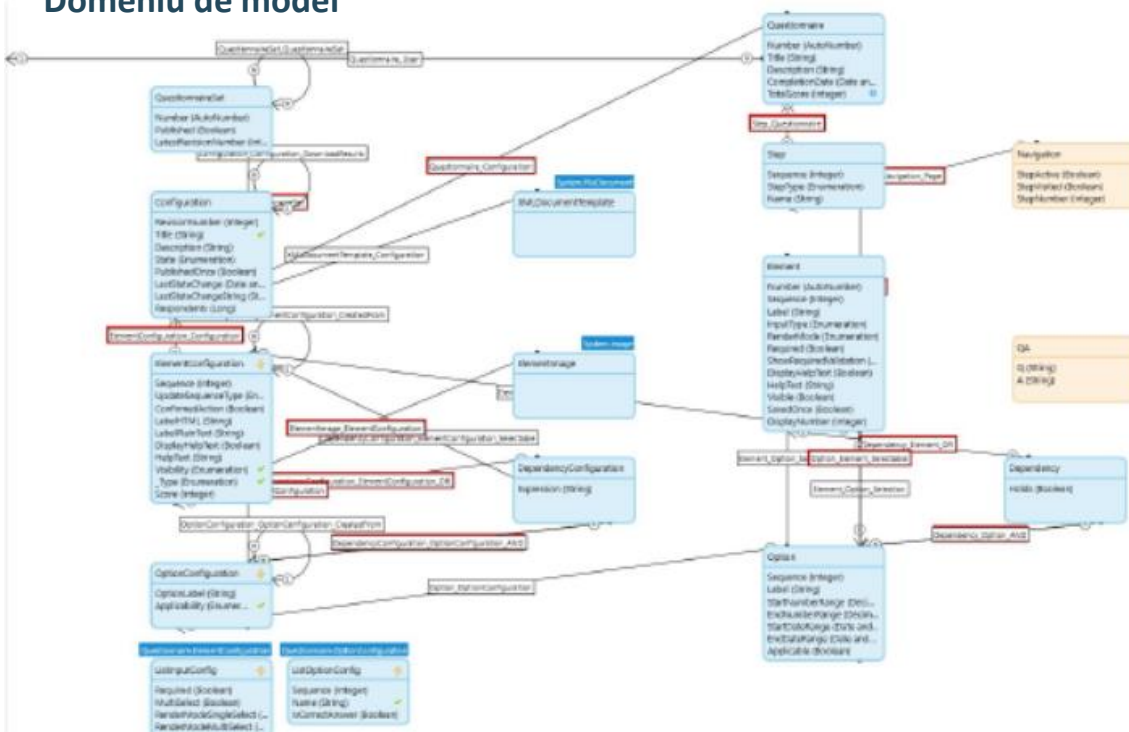


Workflow



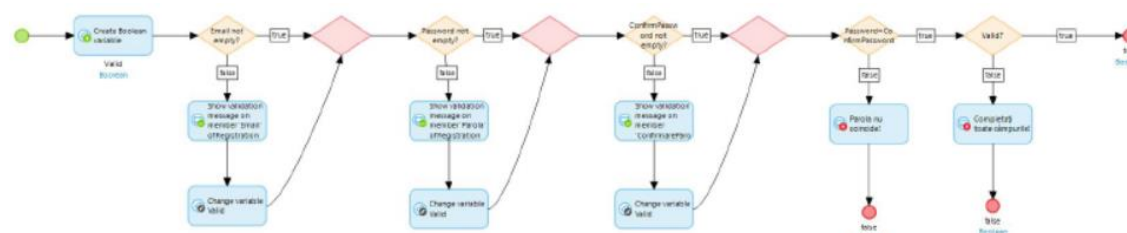
# Gamificarea - Drumul inovației educaționale

Domeniu de model



Abordarea centrată pe domeniul de modelare facilitează colaborarea între dezvoltatori și specialiști în afaceri, permițându-le acestora din urmă să participe activ la procesul de dezvoltare a aplicației fără a fi necesar să aibă cunoștințe profunde în programare.

Astfel, acest tip de platforme de dezvoltare low-code oferă un mediu prielnic pentru crearea rapidă și eficientă a aplicațiilor enterprise, accelerând ciclurile de dezvoltare și punând accentul pe rezolvarea problemelor specifice domeniilor de afaceri.



Microflow

# Gamificarea - Drumul Inovației Educaționale

## Limbaj de programare clasic

### Python, C++, Java

Utilizarea unui astfel de limbaj este facilitată de multitudinea de resurse de învățare (tutoriale, documentații), utile în dezvoltarea unei aplicații complexe.

## Dezvoltare

## Generalizare aplicație

### Design interfață

### Django, Tkinter, Pyside

Crearea unei interfețe dinamice și interactive este esențială pentru o experiență plăcută a utilizatorilor.

### MySQL

RDBMS utilizat adesea în stocarea unui volum mare de date pentru o aplicație web

### PostgreSQL

RDBMS avansat; poate stoca tipuri de date complexe (JSON, date spațiale)

### SQLite

Motor BD care are la bază un singur fișier și nu necesită procesarea pe un server.

### SQLAlchemy (py)

Librărie pentru interacțiunea cu BD, utilă în migrarea datelor.

## Stocare date

### Modul dinamic

pentru crearea în timp real a unor spații de lucru individuale.

### Elemente gamificate

Crearea unui leaderboard sau a unor jocuri interactive (individuale, de echipă).

### Monitorizare progres

Monitorizarea obiectivelor și stadiului de învățare.

## Funcționalități

### Personalizare

a experienței studenților în funcție de subiectele preferate și care necesită mai mult exercițiu.

### Dashboard

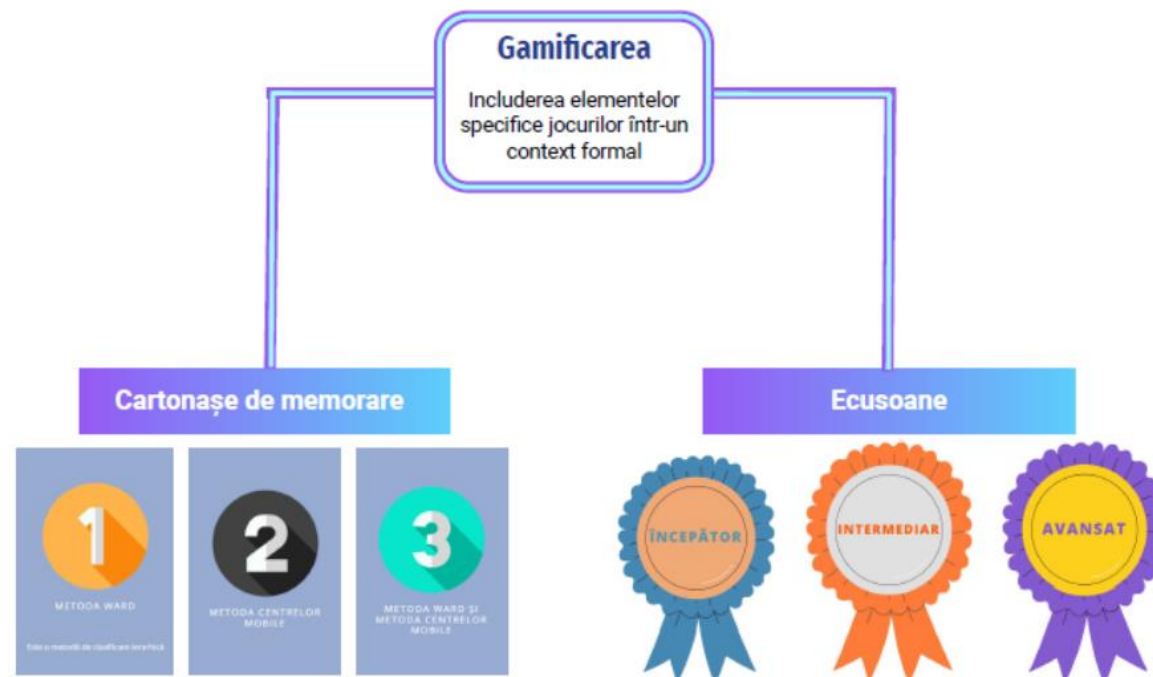
Analiza vizuală a principalilor indicatori de învățare (note, număr de accesări, teste completate)

### Feedback

Feedback-ul profesorilor pe baza unor elemente specifice poate facilita progresul educațional al studenților.

# Gamificarea - Drumul INOVAȚIEI educaționale

Este esențial ca integrarea gamificării să fie realizată într-un mod echilibrat, astfel încât să sprijine obiectivele educaționale și să ofere o experiență de învățare semnificativă și sustenabilă, evitând astfel transformarea excesivă într-un mediu pur distractiv, fără a aduce contribuții substanțiale la procesul de învățare.

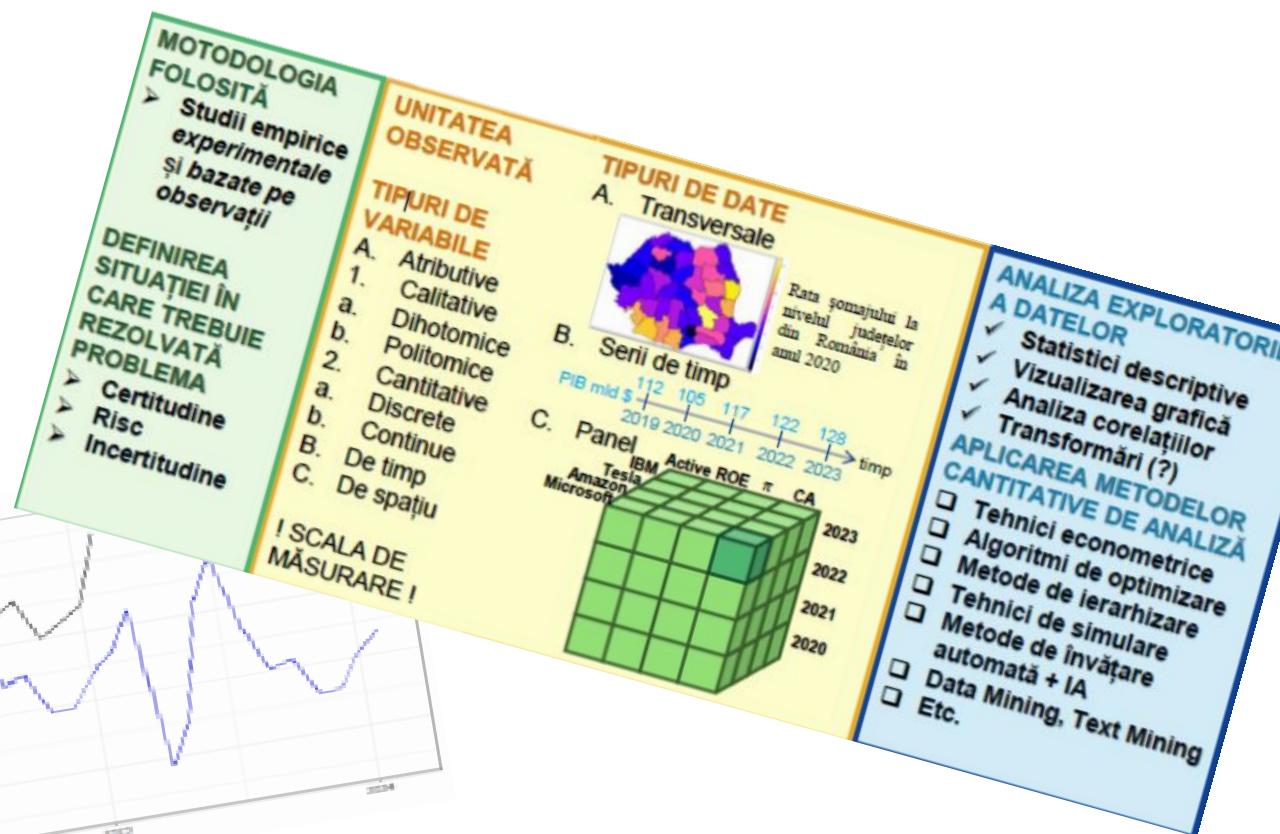
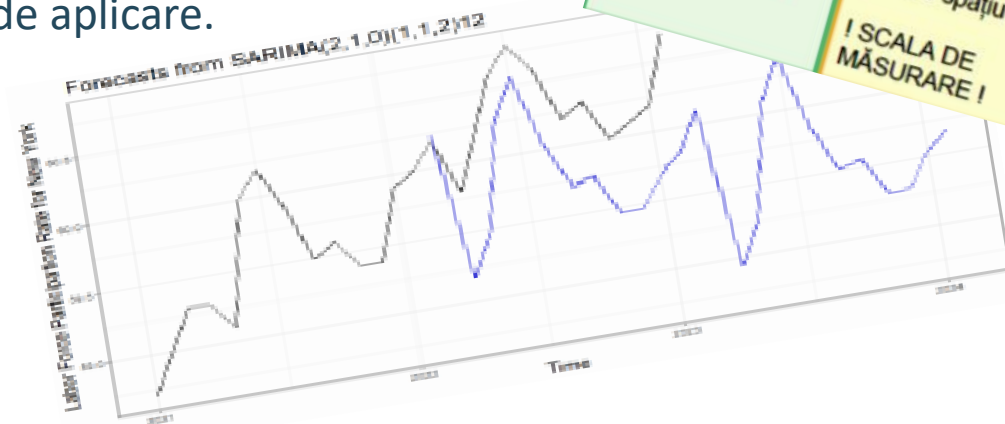




# Explorarea complexității datelor: O taxonomie a modelelor

Această taxonomie încearcă să ofere un cadru util pentru înțelegerea și explorarea modelelor și tehnicilor utilizate în explorarea și analiza datelor, acoperind atât datele structurate cât și cele nestructurate, precum și modelele interdisciplinare și combinate.

Este important să se țină cont că aceste categorii pot fi interconectate și pot exista suprapuneri între diferitele tipuri de modele în funcție de specificul și necesitățile fiecărui proiect sau domeniu de aplicare.



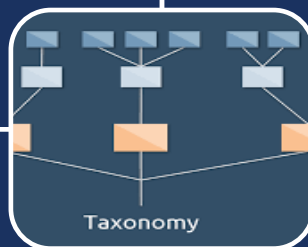
# Explorarea complexității datelor: O taxonomie a modelelor

## Modele de analiza a datelor structurate

- Modele statistice (regresie liniară și neliniară, modele de serii de timp, ANOVA, analiză de corespondență)
- Modele de învățare automată supervizată
  - Algoritmi de clasificare: arbori de decizie etc.
  - Algoritmi de regresie: Random Forest, XGBoost, rețele neurale etc.
- Modele de învățare automată nesupervizată:
  - Clustering (k-means etc.)
  - Reducerea dimensionalității (PCA etc.)

## Modele de analiza a datelor nestructurate

- Modele de prelucrare a limbajului natural (NLP):
- Modele de învățare automată supervizată
  - Algoritmi de clasificare: arbori de decizie etc.
- Modele pentru imagini și videoclipuri
  - Rețele neuronale convoluționale pentru recunoaștere de obiecte
- Modele pentru date audio:
  - Analiză a semnalelor audio și recunoaștere vocală
  - Modele de procesare a sunetului pentru diverse aplicații



## Modele de simulare și prognoză

- Modele matematice și de simulare:
  - Simulare Monte Carlo
  - Modele de rețele bayesiene
  - Modele de sisteme complexe
- Modele de prognoză și predicție:
  - Metode de forecasting (ARIMA etc.)
  - Algoritmi de Machine Learning pentru predicții a seriilor temporale

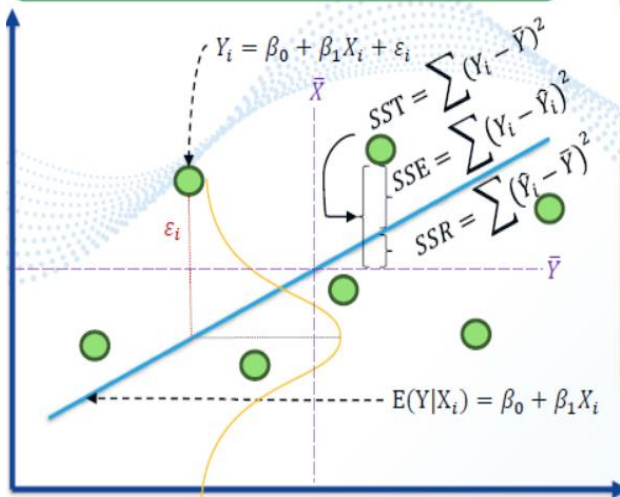
## Modele economice și financiare:

- Modele de evaluare a riscului
  - Modele de creștere economică
  - Modele financiare (portofolii etc.)
- Modele de învățare hibridă și ensemble:
  - Metode ensemble (bagging, boosting)
  - Integrarea mai multor tipuri de modele pentru a obține rezultate mai bune

# Explorarea complexității datelor: O taxonomie a modelelor

Extensii ale modelului clasic de regresie

## 1. Specificarea modelului



2. Estimarea parametrilor modelului: metoda celor mai mici pătrate (OLS)  $\rightarrow \min \sum \varepsilon_i^2$   
Th. Gauss Markov  $\rightarrow$  estimatorii sunt **BLUE**

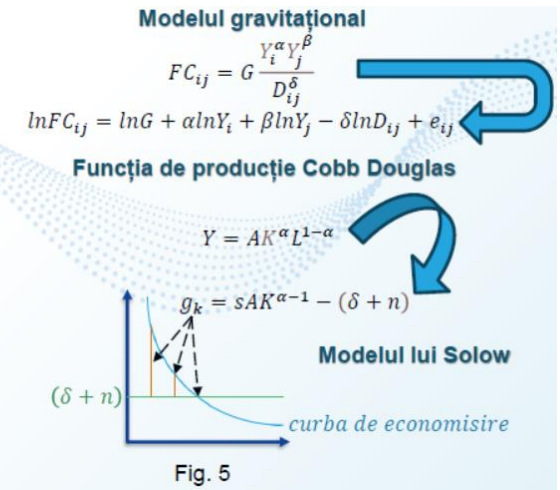
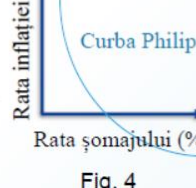
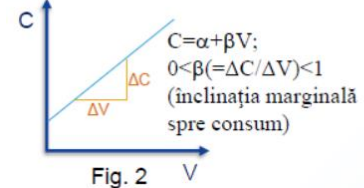
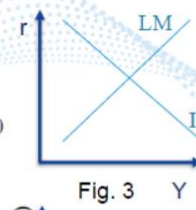
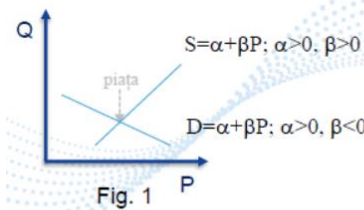
## 3. Ipotezele modelului de regresie

- 1)  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n$
- 2)  $E(\varepsilon_i) = 0, i = 1, 2, \dots, n$
- 3)  $Var(\varepsilon_i) = \sigma_\varepsilon^2 = \sigma^2, i = 1, 2, \dots, n$
- 4)  $cov(\varepsilon_i, \varepsilon_j) = 0$
- 5)  $cov(\varepsilon_i, x_i) = 0$
- 6)  $\varepsilon_i \sim N(0, \sigma^2)$
- 7)  $cor(x_i, x_j) = 0$

## Extensii ale modelului clasic de regresie

Extensiile modelului clasic de regresie oferă flexibilitate și adaptabilitate în gestionarea diferitelor tipuri de date și scenarii și sunt utilizate pentru a îmbunătăți performanța și acuratețea predicțiilor în analiza datelor.

### Exemple de modele aplicate în economie



### Modelul gravitațional

$$FC_{ij} = G \frac{Y_i^\alpha Y_j^\beta}{D_{ij}^\delta}$$

$$\ln FC_{ij} = \ln G + \alpha \ln Y_i + \beta \ln Y_j - \delta \ln D_{ij} + e_{ij}$$

### Funcția de producție Cobb Douglas

$$Y = AK^\alpha L^{1-\alpha}$$

### Modelul lui Solow

# Explorarea complexității datelor: O taxonomie a modelelor

Extensii ale modelului clasic de regresie

## Modele liniare generalizate

**Distribuția Bernoulli - Bernoulli(p), p ∈ (0,1)**

Variabila aleatoare X este distribuită Bernoulli de parametru p dacă:

$$X: \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} \quad M[X] = p \quad D^2[X] = p(1-p)$$

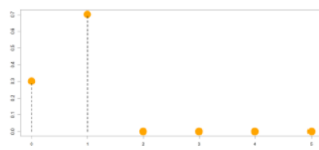
**Distribuția binomială - Binom(n,p), n ∈ N, p ∈ (0,1)**

Variabila aleatoare X este distribuită binomial de parametri n și p dacă:

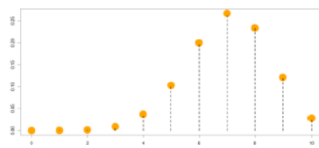
$$X: \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ C_n^0 p^0 q^n & C_n^1 p^1 q^{n-1} & \dots & C_n^k p^k q^{n-k} & \dots & C_n^n p^n q^0 \end{pmatrix},$$

unde q = 1 - p

$$M[X] = np \quad D^2[X] = npq$$



Bernoulli(0.7)



Binom(10,0.7)

**Distribuția Poisson - Poisson(λ), λ > 0**

Variabila aleatoare X este distribuită Poisson de parametru λ dacă:

$$X: \begin{pmatrix} 0 & 1 & \dots & n & \dots \\ e^{-\lambda} \frac{\lambda^0}{0!} & e^{-\lambda} \frac{\lambda^1}{1!} & \dots & e^{-\lambda} \frac{\lambda^n}{n!} & \dots \end{pmatrix}_{n \in \mathbb{N}}$$

$$M[X] = D^2[X] = \lambda$$

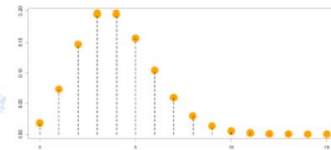
**Distribuția Gamma - γ(α, β); α, β > 0**

Variabila aleatoare X este distribuită Gamma de parametri α și β dacă:

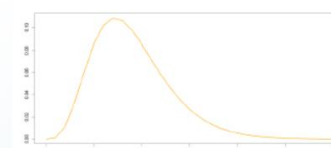
$$f_X(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & , \text{dacă } x \geq 0 \\ 0 & , \text{în rest} \end{cases}$$

unde  $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-t} dt$  pentru  $\alpha > 0$

$$M[X] = \alpha\beta \quad D^2[X] = \alpha\beta^2$$



Poisson(4)

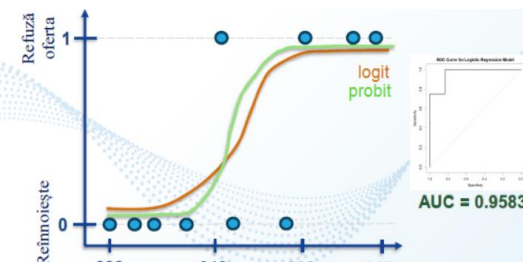


Gamma(5,0.56)

X	Y	Yhat - Logit	Yhat - Probit
800	0	0.0139	0.0053
808	0	0.0231	0.0124
811	0	0.0279	0.0167
823	0	0.0587	0.0486
844	0	0.1946	0.2002
850	1	0.2625	0.2719
876	0	0.6558	0.6574
891	1	0.8338	0.8388
915	1	0.9594	0.9729
920	1	0.9702	0.9830

LOGIT	Estimate	Std. Error	z value	Pr(> z )	AIC: 10
(Intercept)	-55.8811	32.1600	-1.7380	0.0823	
X	0.0645	0.0373	1.7310	0.0835	

PROBIT	Estimate	Std. Error	z value	Pr(> z )	AIC: 9.8
(Intercept)	-33.7140	17.1477	-1.9660	0.0493	
X	0.0390	0.0199	1.9580	0.0502	



AUC = 0.9583

$$\ln\left(\frac{p}{1-p}\right) = -55.8811 + 0.0645 \cdot X \Leftrightarrow p = \frac{e^{-55.8811 + 0.0645X}}{1 + e^{-55.8811 + 0.0645X}}$$

$$p = \phi(-33.714 + 0.039 \cdot X)$$

unde p = P(Y = 1|X)

## Modele de regularizare

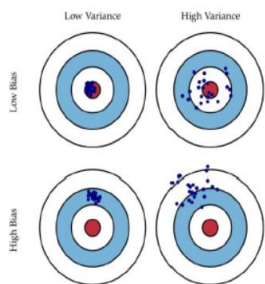
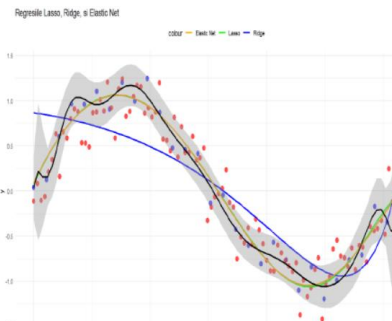
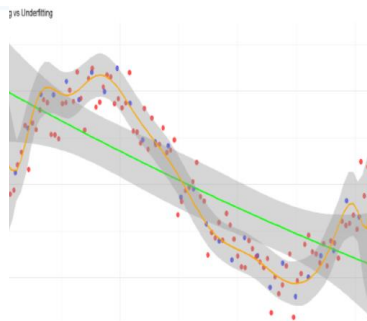
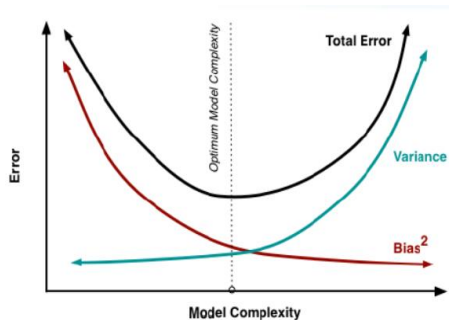


Fig. 1 Graphical illustration of bias and variance.

Sursă <http://scott.fortmann-roe.com/docs/BiasVariance.html>



**Lasso (L1)**

$$\min \sum \epsilon_i^2 + \lambda \sum |A_i|$$

**Ridge (L2)**

$$\min \sum \epsilon_i^2 + \lambda \sum \beta_i^2$$

**Elastic Net**

$$\min \sum \epsilon_i^2 + \lambda_1 \sum |A_i| + \lambda_2 \sum \beta_i^2$$



panta = 0



panta ≠ 0

Grupează și diminuează parametrii asociați cu variabilele corelate, lăsându-i sau eliminându-i din ecuație.



# Explorarea complexității datelor: O taxonomie a modelelor

Extensii ale modelului clasic de regresie

## Modele cu ecuații simultane

În domeniul științelor sociale și economice, modelele cu ecuații simultane au evoluat în timp, iar inovațiile recente au implicat integrarea unor tehnici și metode mai avansate pentru a identifica relațiile complexe între variabile.

- Modelele bazate pe rețele neuronale - oferă o abordare flexibilă și puternică pentru a surprinde relațiile complexe și interdependențele dintre variabile.
- Modelele bayesiene structurale - permit estimări și previziuni robuste, luând în considerare incertitudinea și variabilitatea din date.
- Metodele de inferență causală - ajută la înțelegerea și identificarea efectelor cauzale între variabile și la evaluarea impactului intervențiilor sau politicii asupra sistemului.
- Modelarea cu date de frecvență ridicată - în contextul noilor tehnologii și creșterii disponibilității datelor în timp real, modelele cu ecuații simultane au început să integreze date de frecvență ridicată, permițând o analiză dinamică și actualizată a relațiilor dintre variabile.
- Abordări de Machine Learning - poate aduce o mai mare flexibilitate și precizie în identificarea și înțelegerea interdependențelor dintre variabile.

Aceste inovații și adaptări recente în modelele cu ecuații simultane au scopul de a face analiza relațiilor complexe între variabile mai robustă, mai precisă și mai adaptată la evoluția tehnologică și la necesitățile din diferitele domenii de aplicare.

**Ce sunt variabilele explicativ endogene?**  
**Cauzele endogenității**  
**Soluția? → Introducerea variabilelor instrumentale**

**EXEMPLE**

- efectul șomajului ( $X_e$ ) asupra PIB ( $Y$ ) → 0
- variabila corelată cu rata șomajului este "rata șomajului" ( $Z$ ): PIB ( $X_e$ ) și Consumul ( $Y$ ) sunt în relație de cauzalitate inversă → Investițiile ( $Z$ ) ca variabilă instrumentală inversă → Introducerea cotei de impozitare
- nu influențează deocarece sunt corelate cu PIB, dar estimarea efectului direct consumului ( $X_e$ ) asupra sănătății ( $Y$ ) → Introducerea cotei de impozitare ( $Z$ ) care influențează decizia de a fuma.

**Metoda celor mai mici pătrate în 2 etape (2SLS)**  
$$X_e = Z + X \rightarrow \hat{X}_e$$
$$Y = X_e + X \rightarrow \hat{Y}$$

**Când se folosește 2SLS?**  
 $RT_i + 1 = g$  ecuație exact identificată  
 $RT_i + 1 > g$  ecuație supra-identificată

**Unde**  
 $g_{1,1} = 1, m$  Variabile endogene  
 $m_{1,1} = 1, g$  Variabile exogene

$$\begin{cases} a_{11}y_{1t} + \dots + a_{1m}y_{mt} + b_{11}x_{1t} + \dots + b_{1m}x_{mt} = \varepsilon_{1t} \\ a_{21}y_{1t} + \dots + a_{2m}y_{mt} + b_{21}x_{1t} + \dots + b_{2m}x_{mt} = \varepsilon_{2t} \\ \vdots \\ a_{n1}y_{1t} + \dots + a_{nm}y_{mt} + b_{n1}x_{1t} + \dots + b_{nm}x_{mt} = \varepsilon_{nt} \end{cases}$$

# Explorarea complexității datelor: O taxonomie a modelelor

Extensii ale modelului clasic de regresie

## Modele spațiale

**Tipologia modelelor spațiale**

$\theta = 0$  Kolesjan-Prucha, SARAR, SAC, Cliff-Ord  
 $y = \rho W y + X \beta + u, u = \lambda W u + \epsilon$   
 $y = X \beta + u, u = \lambda W u + \epsilon$   
 $\lambda = 0$  Spatial Lag, Lag Y, SAR  
 $y = \rho W y + X \beta + \epsilon$   
**OLS**

**"Manski Model"**  
 $y = \rho W y + X \beta + W X \theta + u, u = \lambda W u + \epsilon$   
 $\lambda = 0$ , Spatial Durbin Model (SDM)  
 $y = \rho W y + X \beta + W X \theta + \epsilon$

$\rho = 0$ , Spatial Durbin Error Model  
 $y = X \beta + W X \theta + u, u = \lambda W u + \epsilon$   
 $\theta = 0$ , Spatial Error Model (SEM)  
 $y = X \beta + u, u = \lambda W u + \epsilon$

$\rho = 0$ , Spatially Lagged X (SLX)  
 $y = X \beta + W X \theta + \epsilon$   
 $\lambda = 0$ , Spatially Lagged Y, SAR  
 $y = \rho W y + X \beta + \epsilon$   
**OLS**

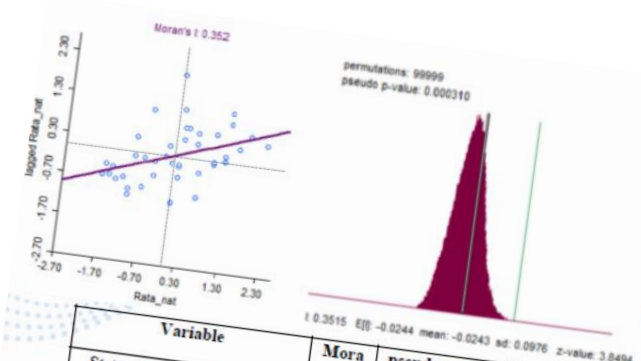
**Cine sunt vecinii?**

**Cum se măsoară autocorelația spațială?**

Indicele Moran — măsoară gradul de similaritate  
 ia valori de la [-1; 1]

$$IM = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot (x_i - \bar{x})(x_j - \bar{x})}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

**Testul Moran**

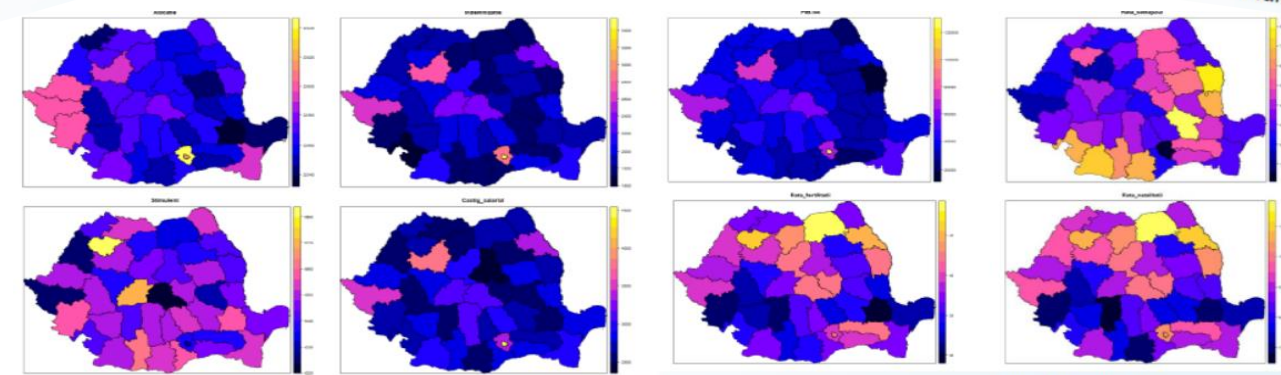


Variable	Moran's I	pseudo p-value
State allowance for children	0.153	0.0394
Benefit for child raising	0.242	0.0083
Incentive	-0.127	0.1390
Net earnings	0.112	0.0717
GDP per capita	0.254	0.0032
Unemployment rate	0.292	0.0017

Indicii Moran pentru variabilele modelului

Variable	Model 1		Model 2	
	Coefficient (Std. Error)	Probability	Coefficient (Std. Error)	Probability
Intercept	-151.0050 (189.7000)	0.4314	-25.4787 (9.0432)	0.0076
Ln (Benefit for child raising)	7.1419 (3.6273)	0.0569	7.6835 (2.0893)	0.0007
Ln (GDP per capita)	-2.4322 (1.0113)	0.0221	-2.2828 (0.9209)	0.0187
Ln (Net earnings)	-1.3558 (3.8312)	0.7255		
Ln (Incentive)	-12.8005 (10.1519)	0.2157		
Ln (State allowance for children)	29.1028 (24.1836)	0.2369		
Unemployment rate	-0.0347 (0.1292)	0.7896		
Adjusted R2				
AIC	0.2404			
F	128.871		0.2453	
Moran's I (error)	3.1629	0.0139	125.146	
Lagrange Multiplier (lag)	2.6909	0.0073	7.6613	0.0016
Robust LM (lag)	7.4169	0.0064	3.0714	0.002
Lagrange Multiplier (error)	3.2861	0.0698	8.3307	0.004
Robust LM (error)	4.3463	0.0370	5.9592	0.113
Lagrange Multiplier (SARMA)	0.2155	0.8425	0.1339	0.714
	7.6324	0.0220	8.4648	0.015

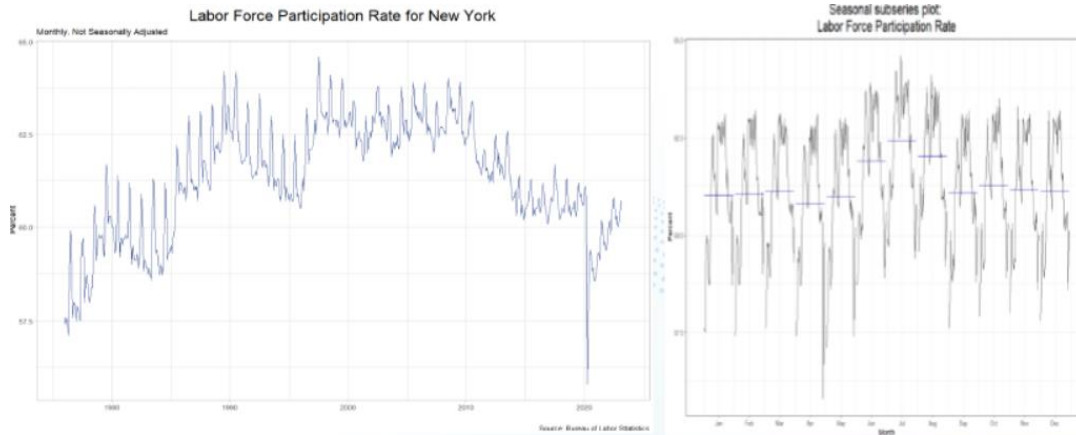
Modelele lin-log de regresie



# Explorarea complexității datelor: O taxonomie a modelelor

Extensii ale modelului clasic de regresie

## Modelarea seriilor de timp



### Holt Winters Aditiv

$$L_t = \alpha * (y_t - S_{t-m}) + (1 - \alpha) * (L_{t-1} + b_{t-1})$$

$$b_t = \gamma * (L_t - L_{t-1}) + (1 - \gamma) * b_{t-1}$$

$$S_t = \delta * (y_t - L_t) + (1 - \delta) * S_{t-1}$$

### Holt Winters Multiplicativ

$$L_t = \alpha * \frac{y_t}{S_{t-m}} + (1 - \alpha) * (L_{t-1} + b_{t-1})$$

$$b_t = \gamma * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1}$$

$$S_t = \delta * \frac{y_t}{L_t} + (1 - \delta) * S_{t-1}$$

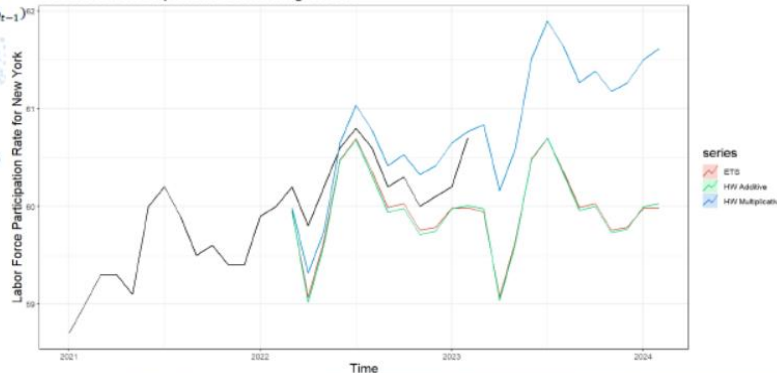
### ETS

$$Y_t = L_{t-1} * S_{t-M} * (1 + \epsilon_t)$$

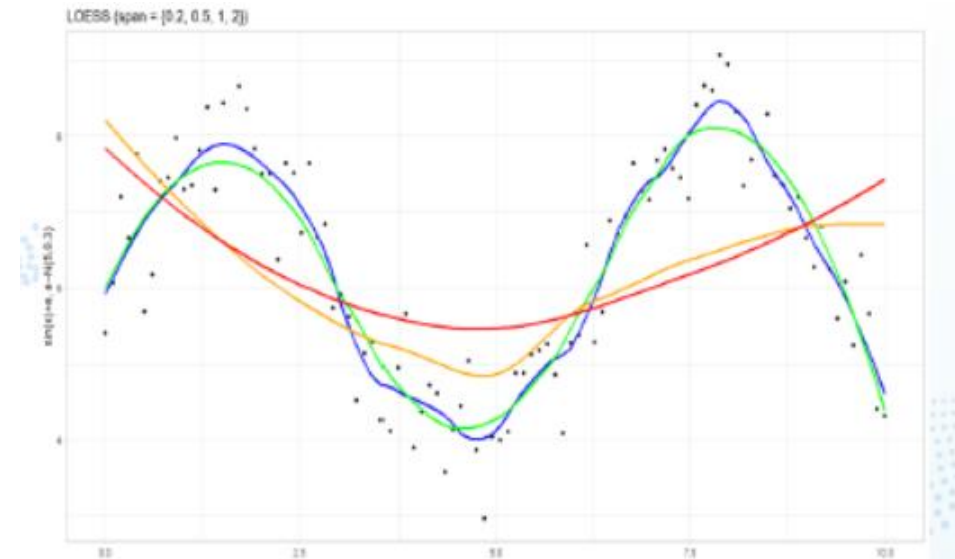
$$L_t = L_{t-1} * (1 + \alpha \epsilon_t)$$

$$S_t = S_{t-M} * (1 + \gamma \epsilon_t)$$

Forecasts from exponential smoothing models



## Modele neparametrice



Modelele neparametrice sunt metode statistice care nu impun o formă specifică a relației dintre variabile, ci permit datelor să definească forma funcției. Aceste modele sunt flexibile și pot captura relații complexe și nonlineare fără să necesite o ecuație specifică. Un exemplu este metoda LOESS, care ajustează mici funcții de regresie în jurul fiecărui punct de date.



# Explorarea complexității datelor: O taxonomie a modelelor

**Ce este simularea proceselor?**

**Tipuri de simulare**

1. Simularea probabilistică
2. Simularea matematică: pură și Monte Carlo
3. Simularea analogică

**Metoda transformatei inverse** folosește funcția de distribuție cumulativă pentru a transforma numerele aleatoare uniforme în numere aleatoare cu o anumită distribuție. Este o tehnică pentru generarea de date aleatoare conform distribuțiilor dorite.

**Necesitatea simulării**

- ☐ Nu este posibilă găsirea unei soluții analitice;
- ☐ Experimentarea reală este prea scumpă sau greu de realizat;
- ☐ Simularea este unica metodă disponibilă deoarece modelul real de studiat este dificil.

$x_i = F^{-1}(u_i) \cdot i = \overline{1, n}$

**Monte Carlo Simulation**

Interpretat în mod simplu, este un model utilizat pentru a prezice probabilitatea unei game de rezultate atunci când există variabile aleatoare în prezent.

## Tehnici de simulare

Modelele matematice și de simulare sunt instrumente esențiale folosite pentru a reprezenta și analiza comportamentul sistemelor complexe. Aceste modele utilizează concepte matematice, ecuații și algoritmi pentru a crea reprezentări virtuale ale sistemelor reale. Aceste modele permit anticiparea evoluției unui sistem, testarea ipotezelor sau optimizarea proceselor într-un mediu controlat.

Simulările matematice pot varia de la simple modele statistice la simulări complexe bazate pe ecuații diferențiale sau algoritmi de simulare, contribuind la dezvoltarea și progresul în multe domenii științifice și tehnologice.

Ele sunt fundamentale în diverse domenii, oferind o perspectivă utilă pentru înțelegerea fenomenelor complexe și pentru luarea deciziilor informate.

An	kWh/ap.	Nr. ap. obs.	f	fc	Interval
2011	2217	643	0.0899	0.0899	[0;0.09]
2012	2468	726	0.1015	0.1915	[0.09;0.191]
2013	2136	652	0.0912	0.2827	[0.191;0.283]
2014	2865	635	0.0888	0.3715	[0.283;0.371]
2015	3184	969	0.1355	0.5070	[0.371;0.507]
2016	3230	740	0.1035	0.6105	[0.507;0.61]
2017	3279	824	0.1152	0.7257	[0.61;0.726]
2018	3125	678	0.0948	0.8206	[0.726;0.821]
2019	3180	721	0.1008	0.9214	[0.821;0.921]
2020	3563	562	0.0786	1.0000	[0.921;1]

**Generarea de nr. aleatoare**

u_i [0,1]			Cantități simulate		
0.5864	0.4269	0.7605	3230	3184	3125
0.0027	0.5137	0.5273	2217	3230	3230
0.8735	0.7872	0.3369	3180	3125	2865
0.6114	0.6557	0.2543	3279	3279	2136
0.2639	0.5473	0.1272	2136	3230	2468
0.4929	0.3265	0.4314	3184	2865	3184
0.0447	0.6793	0.7069	2217	3279	3279
0.0129	0.1729	0.1389	2217	2468	2468
0.9537	0.0166	0.5928	3563	2217	3230
0.5866	0.8726	0.2046	3230	3180	2136

Medie: 2887.7  
AS: 459.8  
CV: 16%

IC:  $(\bar{x} - t_{1-\frac{\alpha}{2}, N-1} \frac{s}{\sqrt{N}} < \mu < \bar{x} + t_{1-\frac{\alpha}{2}, N-1} \frac{s}{\sqrt{N}})$

IC: (2716 < μ < 3059.39)

7150 apartamente observate în cei 10 ani



# Explorarea complexității datelor: O taxonomie a modelelor

- Orice problemă decizională poate fi formulată astfel:
- se alege o variantă dintre **variantele** avute la dispoziție:  $\{V1, V2, \dots, Vi, \dots, Vm\}, \forall i = \overline{1, m}$
  - ținând cont de **criteriile decizionale** avute în vedere:  $\{C1, C2, \dots, Cj, \dots, Cn\}, \forall j = \overline{1, n}$
  - și în funcție de **stările naturii** care pot avea loc:  $\{N1, N2, \dots, Nk, \dots, Nr\}, \forall k = \overline{1, r}$

A	$N_1$	$\dots N_k \dots$	$N_r$
	$c_1, c_2, \dots, c_n$	$c_1, c_2, \dots, c_n$	$c_1, c_2, \dots, c_n$
$v_1$			
$\vdots$			
$v_i$	$\dots$	$a_{ijk}$	$\dots$
$\vdots$			
$v_n$			
$\pi_j$	$\pi_1, \pi_2, \dots, \pi_n$	$\pi_1, \pi_2, \dots, \pi_n$	$\pi_1, \pi_2, \dots, \pi_n$

CERTITUDINE $p_k \in (0,1)$	INCERTITUDINE $p_k ?$	RISC $p_k \in (0,1)$
<b>Exemple de metode:</b> <ul style="list-style-type: none"> <li>➢ ELECTRE</li> <li>➢ PERMUTĂRIILOR SUCCESIVE</li> <li>➢ PONDERĂRII SIMPLE ADITIVE</li> <li>➢ PRODUSULUI PONDERAT</li> <li>➢ MOMENTELOR</li> <li>➢ TOPSIS</li> <li>➢ ONICESCU</li> <li>➢ MAXIMIZĂRII UTILITĂȚII GLOBALE</li> <li>➢ ELECTRE-BOLDUR</li> </ul>	<b>Exemple de metode:</b> <ul style="list-style-type: none"> <li>➢ Pesimistă</li> <li>➢ Optimistă</li> <li>➢ Hurwicz</li> <li>➢ Laplace</li> <li>➢ Savage</li> </ul>	<b>Exemple de metode:</b> <ul style="list-style-type: none"> <li>➢ Principiul maximizării valorii așteptate</li> <li>➢ Principiul maximizării valorii monetare așteptate</li> <li>➢ Principiul maximizării utilității așteptate</li> <li>➢ Analiza Bayesiană</li> <li>➢ Arbori de Decizie</li> </ul>

## Metode de ierarhizare

Metodele de ierarhizare - tehnici utilizate pentru a organiza și a clasifica datele, informațiile sau entitățile într-o structură ierarhică bazată pe anumite criterii sau relații.

Aceste tehnici sunt folosite în diverse domenii, de la știința computerelor până la științele sociale, pentru a evidenția relațiile și dependențele dintre elemente într-un mod organizat și comprehensibil.

Metode de ierarhizare sunt utilizate pentru a simplifica și a înțelege mai bine structura și relațiile din datele sau informațiile complexe, contribuind la organizarea eficientă a datelor și la extragerea de înțelesuri utile din acestea.

Se folosește **normalizarea vectorială (euclidiană)**:  $r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^n a_{ij}^2}}$

**PAS 2:** Se determină **matricea consecințelor ponderate**:  $R = (r_{ij})_{i=\overline{1,m}, j=\overline{1,n}}$  unde:  $t_{ij} = \pi_j \cdot r_{ij}$ , reprezintă consecința alegerii variantei  $i$ , după criteriul  $j$ , având ponderea (importanța)  $\pi_j$ .

**PAS 3:** Se determină **SOLUȚIA IDEALĂ POZITIVĂ (SIP)** și **SOLUȚIA IDEALĂ NEGATIVĂ (SIN)**:  
 $T^+ = (t_1^+, t_2^+, \dots, t_n^+)$  (SIP)  $t_j^+ = \begin{cases} \max_{i=\overline{1,m}} t_{ij} \rightarrow \text{criteriu MAX} \\ \min_{i=\overline{1,m}} t_{ij} \rightarrow \text{criteriu MIN} \end{cases}$   $t_j^- = \begin{cases} \min_{i=\overline{1,m}} t_{ij} \rightarrow \text{criteriu MAX} \\ \max_{i=\overline{1,m}} t_{ij} \rightarrow \text{criteriu MIN} \end{cases}$

**PAS 4:** Se calculează **distanța euclidiană față de SIP și SIN**:  
 $S_i^+ = \sqrt{\sum_{j=1}^n (t_{ij} - t_j^+)^2}, (\forall i = \overline{1,m}, j = \overline{1,n})$   $S_i^- = \sqrt{\sum_{j=1}^n (t_{ij} - t_j^-)^2}, (\forall i = \overline{1,m}, j = \overline{1,n})$

**PAS 6: CLASAMENTUL OPTIM** al variantelor decizionale va fi dat de ordonarea descrescătoare a valorilor corespunzătoare apropierei relative față de soluția ideală,  $R_i^+$ .

## Metoda TOPSIS

# Explorarea complexității datelor: O taxonomie a modelelor

## Algoritmi de optimizare

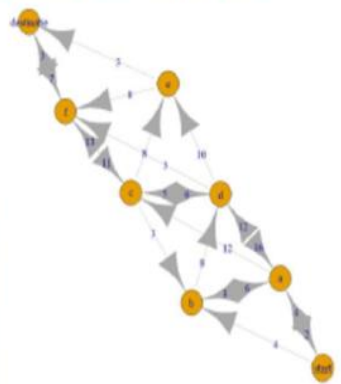
Algoritmii de optimizare sunt seturi de proceduri sau reguli matematice utilizate pentru a găsi soluții optime pentru diverse probleme de optimizare. Aceste probleme pot fi variate, cum ar fi minimizarea sau maximizarea unei funcții obiectiv în condițiile unor constrângeri specifice.

**Algoritmul lui J. B. Kruskal (1956)** – algoritm greedy care selectează o muchie la fiecare pas, identifică arborele parțial de valoare minimă.

**Algoritmul lui Dijkstra** – determină drumurile de valoare minimă de la un nod arbitrar  $s$  al unui graf la toate celelalte noduri ale acestuia.

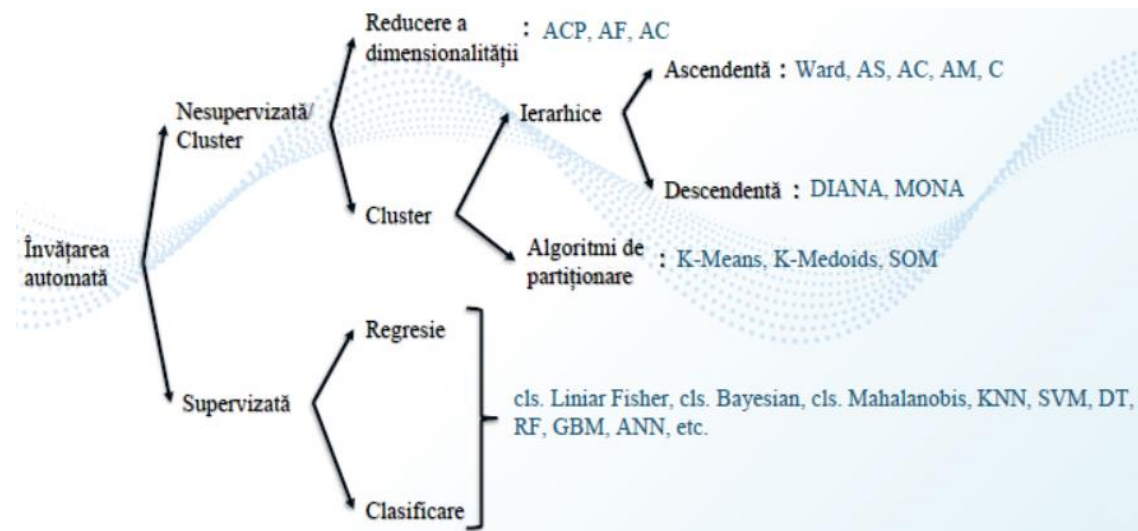
**Algoritmul Bellman Kalaba** – determină drumurile de valoare minimă sau maximă de la fiecare nod al unui graf la un nod arbitrar  $t$ .

EXEMPLU R – alg. dijkstra



"start" → "a" → "b" → "c" → "f" → "destinatie"

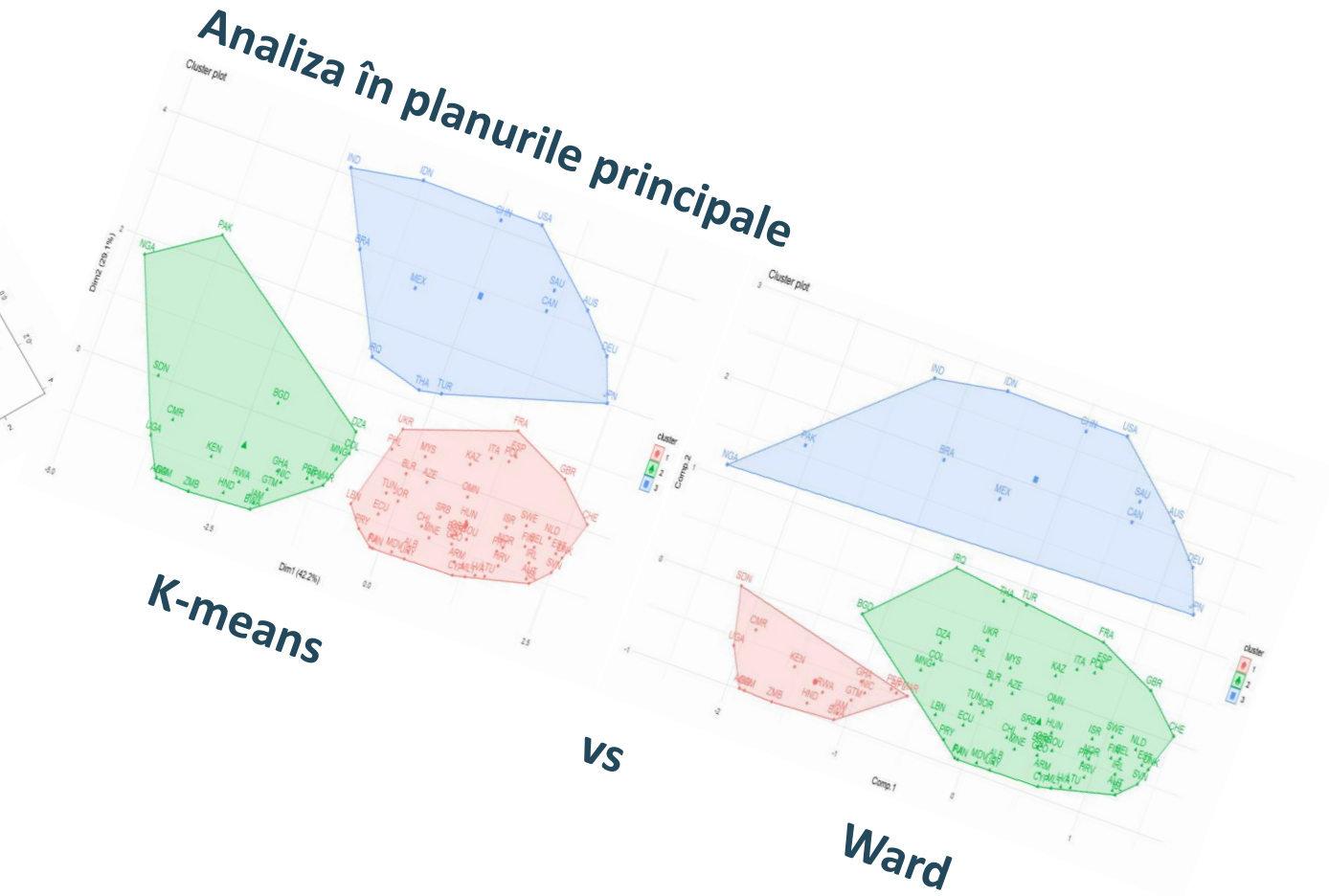
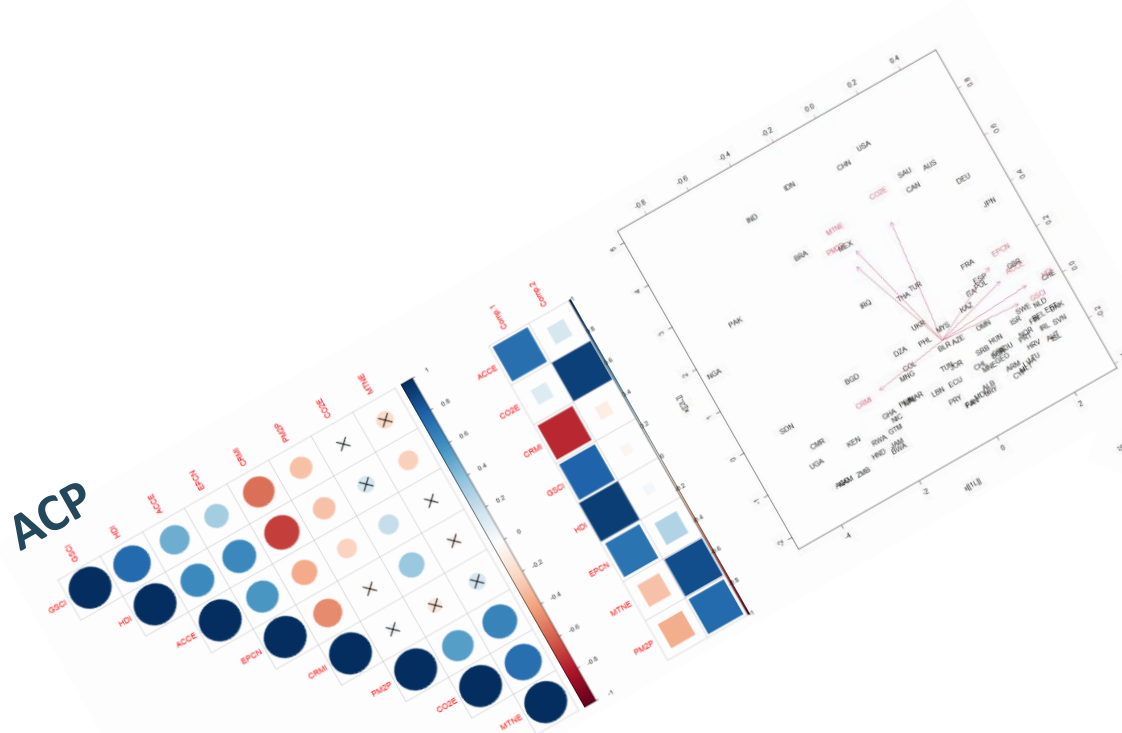
"destinatie" → "e" → "d" → "b" → "start"



Învățarea automată este o ramură a inteligenței artificiale care se ocupă de dezvoltarea modelelor și algoritmilor care permit calculatoarelor să învețe și să facă predicții sau să ia decizii fără a fi explicit programate pentru fiecare situație. În cadrul învățării automate, există două abordări principale: nesupervizată și supervizată.

**Învățarea automată: nesupervizată și supervizată**

# Explorarea complexității datelor: O taxonomie a modelelor





# Explorarea complexității datelor: O taxonomie a modelelor

## Date nestructurate

Modelele pentru date nestructurate folosesc tehnici avansate de procesare a datelor și de învățare automată pentru a extrage informații valoroase din aceste tipuri de date diverse și fără o structură clar definită, facilitând înțelegerea și interpretarea acestora în contextul diferitelor domenii de aplicare.



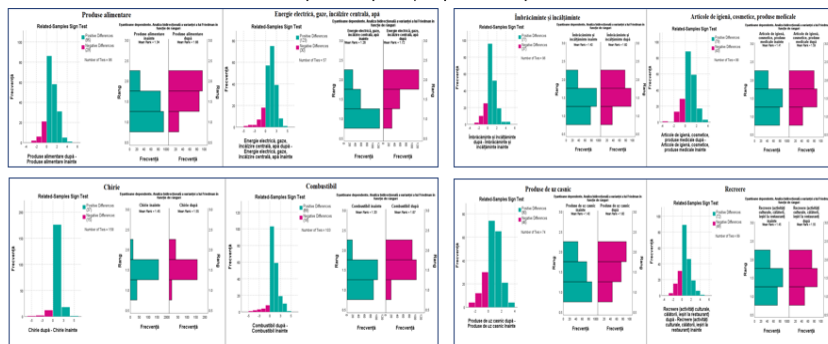


# Perspective inovatoare asupra interdependenței dintre inflație și nivelul de trai utilizând regresia logistică (metoda forward conditional)

Efectul inflației asupra nivelului de trai poate fi semnificativ, deoarece inflația poate eroda puterea de cumpărare a banilor și poate afecta accesibilitatea și disponibilitatea bunurilor și serviciilor pentru consumatori. Astfel, menținerea inflației la un nivel moderat este esențială pentru stimularea creșterii economice sănătoase și pentru asigurarea unui nivel de trai stabil și sustenabil pentru populație. Politicile monetare și fiscale pentru a controla inflația și pentru a menține stabilitatea economică.

Abordarea analitică propusă, oferă oportunitatea de a investiga în profunzime modul în care variabilele socio-economice modelează percepția și comportamentul indivizilor în contextul variațiilor inflației și implicațiilor acestora asupra nivelului de trai. Regresia logistică, în particular metoda forward conditional, permite investigarea relației complexe dintre inflație și consum, oferind o înțelegere mai profundă a modului în care preferințele și obiceiurile de consum sunt afectate de schimbările de prețuri. De asemenea, prin identificarea factorilor care influențează percepția indivizilor asupra inflației, această abordare analitică inovatoare aduce în discuție modul în care mediul socio-economic, venitul gospodăriei și alți factori socio-demografici contribuie la modul în care indivizii percep și reacționează la variabilitatea prețurilor.

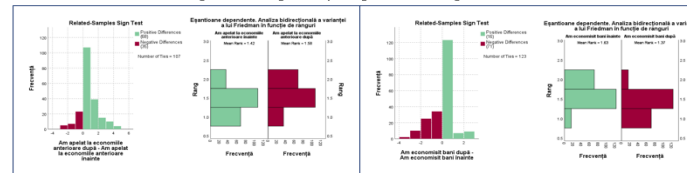
Modificarea părerilor respondenților spre acord sau spre dezacord



Sursă: Prelucrări proprii în SAS pe baza datelor din chestionar

Eșantioane dependente	Testul semnului	Testul Wilcoxon	Testul Friedman
Am contractat un credit înainte - după	0.134	0.419	0.102
Am apelat la economiile anterioare înainte - după	0.002	0.001	0.001
Am economisit bani înainte - după	0.000	0.000	0.000

Modificarea părerilor respondenților spre acord sau spre dezacord



Modificarea părerilor respondenților spre acord sau spre dezacord

Sursă: Prelucrări în SAS pe baza datelor din chestionar

Identificarea factorilor de influență asupra percepției oamenilor privind nivelul inflației



Prezentul răspunsurilor la întrebarea „Care este nivelul mediu anual al inflației percepute de dvs., în prezent?”

- Din cei 210 respondenți, 142 de persoane au afirmat că nu știu care este nivelul mediu anual al inflației și doar 68 de persoane, au răspuns care este nivelul anual al inflației percepute.
- Doar 27 de persoane au afirmat că nu se interesează sau nu se informează despre inflație
- 50.6% din cei care au răspuns că nu știu care este nivelul actual al inflației, au afirmat la întrebarea ulterioară că se informează despre inflație din diverse surse.

Sursă: Site-ul BNR și prelucrări în Excel pe baza datelor culese din chestionar

Sursă de informare ale respondenților despre nivelul inflației

# Perspective inovatoare asupra interdependenței dintre inflație și nivelul de trai utilizând regresia logistică (metoda FORWARD CONDITIONAL)

## Regresia logistică metoda FORWARD CONDITIONAL

### 1. Convergența modelului

5 iterații pentru obținerea soluției optime

### 2. Validarea modelului de regresie

Ipotezele testului Chi-square sunt :

- H0: modelul nu este valid
- H1: modelul este valid

### 3. Calitatea modelului

Medie

### 4. Procentul previziunilor corecte

Senzitivitatea modelului: 36 de respondenți identificați corect care se simt pregătiți în cazul unor cheltuieli neașteptate, adică 56,3% dintre aceștia.

Specificitatea modelului -> 89,7%

### Omnibus Tests of Model Coefficients

Step	Chi-square	df	Sig.
Step 5	11.327	4	.023
Block	85.096	11	.000
Model	85.096	11	.000

### Validarea modelului de regresie

Sursă: prelucrări proprii în SPSS pe baza datelor din chestionar

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	219.992*	.166	.235
2	203.500*	.229	.324
3	194.194*	.263	.371
4	184.466*	.296	.419
5	173.139*	.333	.471

### Convergența și calitatea modelului

### Classification Table<sup>a</sup>

Observed	Q25binara	Predicted		Percentage Correct
		0	1	
Step 5	0	131	15	89.7
	1	28	36	56.3
Overall Percentage				79.5

### Procentul previziunilor corecte

Sursă: prelucrări proprii în SPSS pe baza datelor din chestionar

## Semnificația parametrilor

Step 5	Variabila	B	Sig.	Exp(B)
	Care este venitul dvs. lunar?	0.857	0.02	2.356
	Am făcut anumite împrumuturi pentru a face față cheltuielilor	-2.07	0.003	0.126
	Am economisit bani.	0.187	0	1.206
	În prezent, resursele dvs sunt suficiente pentru acoperirea nevoilor lunare?	-1.499	0.001	0.223
	Care este nivelul dvs. de educație cel mai recent finalizat	-1.205	0.02	0.300

### Tabelul 7.

Sursă: prelucrări proprii în SPSS pe baza datelor din chestionar

Influențează pozitiv probabilitatea de a fi pregătit în cazul unor cheltuieli neprevăzute

Coefficientul  $\hat{\beta}_1 = 0.857 > 0$  & Coefficientul  $\hat{\beta}_3 = 0.187 > 0$

• Creșterea venitului lunar cu o unitate este asociată cu o creștere de 2.35 de ori a raportului de șanse (odds ratio) de a fi pregătit în cazul unor situații neprevăzute.

• Trecerea de la un nivel la altul în ceea ce privește dezechilibrul/acordul pentru afirmația "Am economisit bani" crește de 1.206 ori, adică cu 20.6%, raportul de șanse (odds ratio) de a fi pregătit în cazul unor cheltuieli neprevăzute.

Influențează negativ probabilitatea de a fi pregătit în cazul unor cheltuieli neprevăzute

Coefficientul  $\hat{\beta}_2 = -2.070 < 0$

Trecerea de la un nivel la altul în ceea ce privește acordul/dezacordul pentru afirmația "Am făcut anumite împrumuturi pentru a face față cheltuielilor" scade cu 87,4% raportul de șanse de a fi pregătit în cazul unor cheltuieli neprevăzute.

Coefficientul  $\hat{\beta}_4 = -1.499 < 0$

Trecerea de la un nivel la altul în ceea ce privește aprecierea resurselor pentru acoperirea nevoilor diminuează cu aproximativ 80% raportul de șanse de a fi pregătit în cazul unor cheltuieli neprevăzute.

Coefficientul  $\hat{\beta}_5 = -1.205 < 0$

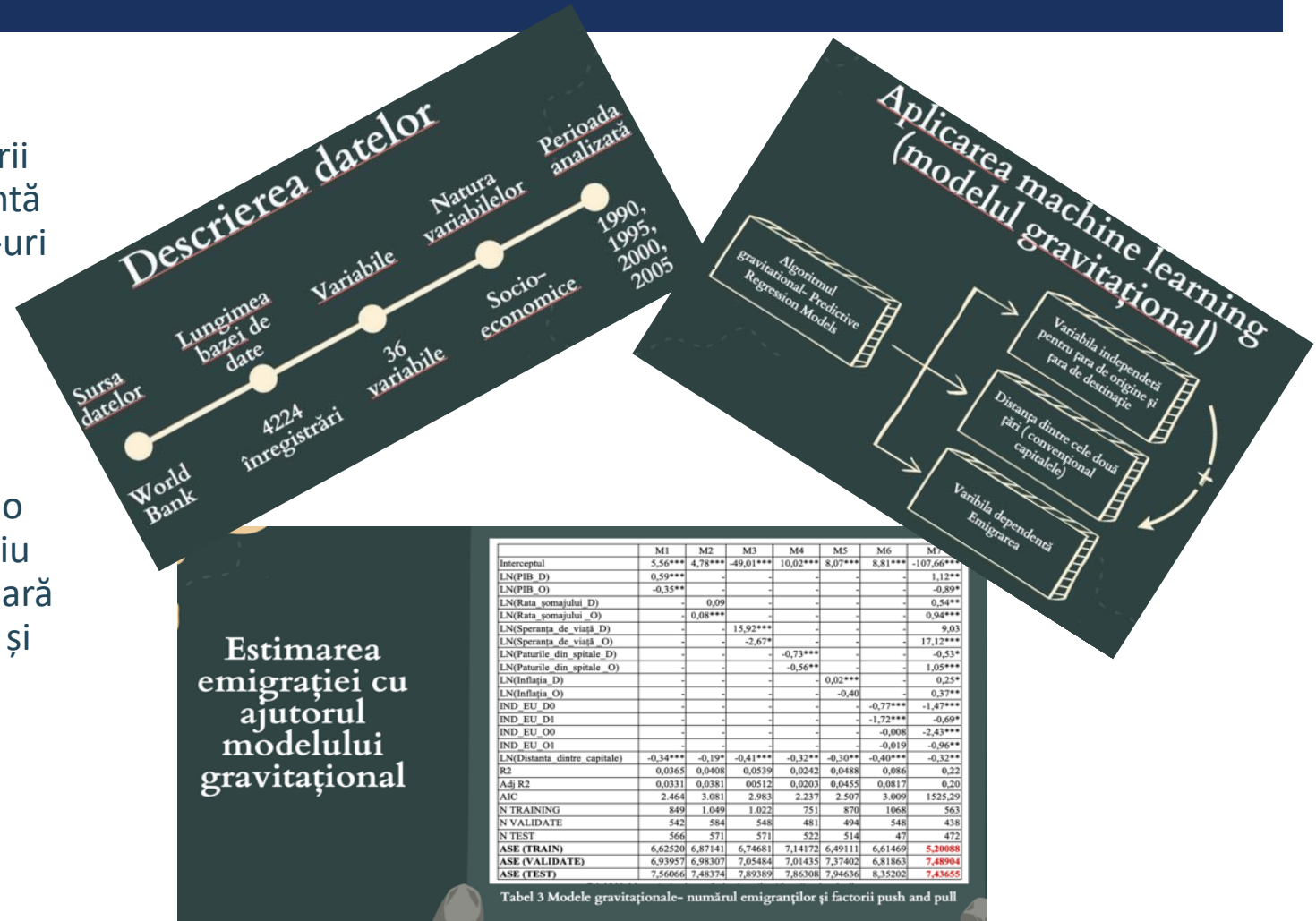
Trecerea de la un nivel la altul în ceea ce privește nivelul de educație scade cu 70% raportul de șanse ca o persoană să fie pregătită în cazul unor situații neprevăzute.

Această metodă permite și investigarea modului în care nivelul inflației influențează nivelul de trai al populației. Prin evaluarea modului în care creșterea prețurilor influențează opțiunile și accesibilitatea produselor și serviciilor, se poate obține o imagine mai clară asupra modului în care inflația afectează puterea de cumpărare și distribuția veniturilor într-o societate.

Metoda utilizată în analiză poate ajuta la identificarea segmentelor de populație care sunt mai susceptibile la impactul inflației asupra nivelului de trai și a percepției economice. Aceste informații sunt esențiale pentru elaborarea politicilor publice și a strategiilor economice care să abordeze eficient efectele inflației asupra nivelului de trai și să promoveze o mai mare stabilitate economică și socială în comunități.

# Analiza fluxurilor migrației: abordarea modelului gravitațional în contextul utilizării Machine Learning

Utilizarea Machine Learning în contextul modelării fluxurilor migrației poate fi o abordare interesantă și eficientă pentru a înțelege și a prezice pattern-uri de migrație între diferite locații. Integrarea modelului gravitațional în analiza fluxurilor migrației poate oferi o perspectivă interesantă și poate îmbunătăți precizia prognozelor, dar este important să se ia în considerare și alți factori relevanți pentru a avea o înțelegere mai completă a fenomenului migrației. Totodată, este necesară o abordare interdisciplinară care să integreze expertiza în domeniul migrației și învățarea automată pentru a obține rezultate relevante și utile.



# Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

Modelarea de tip Ensemble reprezintă o abordare în care se combină rezultatele multiplelor modele pentru a obține performanțe mai bune decât cele obținute de la un singur model individual.

Cele două metode de bază pentru modelele de tip Ensemble sunt Bagging și Boosting. Modelele de tip Bagging îmbunătățesc performanța combinând multiple modele independente antrenate pe subseturi aleatoare din setul de date, folosind tehnici precum Random Forests sau Bagged Decision Trees. În schimb, modelele de tip boosting construiesc modele secvențial, fiecare încercând să corecteze erorile modelelor anterioare, cum ar fi în cazul AdaBoost sau Gradient Boosting Machines (GBM), pentru a obține o performanță superioară.

## Modele de tip "Ensemble"

### Concepte de bază

- Aprecierea performanțelor modelelor. Compromisul bias/variance
- Weak learners, strong learners
- Modelarea de tip ensemble. Concept, elemente de bază, avantaje

### Modalitățile de modelare de tip ensemble

- Modelarea de tip bagging
- Descriere generală
- Tehnica bootstrapping. Elemente

### Modele din clasa random forests

- Descriere generală algoritmul
- Probleme specifice legate de utilizarea practică

## Modele de tip "Ensemble"

### Modelarea de tip boosting

- Modele de tip boosting. Elemente de bază
- Istoricul modelelor de tip boosting
- Modelul adaboost M1

### Modelarea cu metoda Gradient boosting

- Particularități și forma generală
- Noțiuni elementare. Funcția de eroare, funcția de cost, optimizarea gradient descent
- Formalizarea algoritmului GBM. Elemente de optimizare
- Modelul XGBoost
- Optimizarea hiperparametrilor

### Aprecieri generale despre modelele de tip ensemble



# Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

**Weak learners, strong learners și modelarea de tip ensemble**

**Weak learners:** modele ale căror performanțe sunt slabe. Rezultatele sunt ceva mai bune decât o predicție aleatoare.

Pe acest tip de modele se bazează modelele de tip ensemble, fiind folosite pentru obținerea unui model superior ca performanțe.

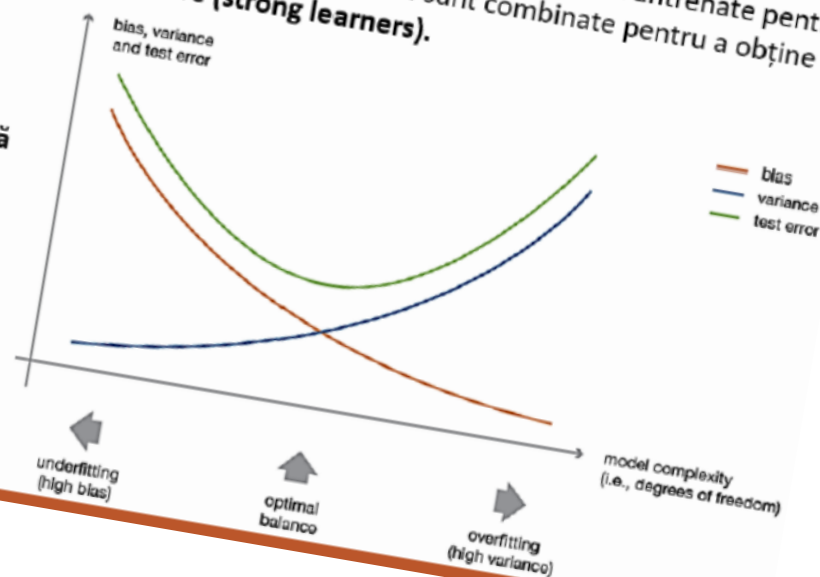
**Strong learners:** modele cu performanțe superioare, ale căror putere de predicție este mult superioară față de cea a unui model aleatoriu.

Printre aceste modele se numără și cele de tip ensemble.

## Modelarea de tip ensemble

**Modelarea de tip ensemble** (paradigmă de machine learning): tehnica prin care mai multe modele de bază cu performanță scăzută (**weak learners**), sunt antrenate pentru a rezolva **aceeași problemă**, iar rezultatele acestora sunt combinate pentru a obține modele cu performanțe superioare (**strong learners**).

Scopul este estimarea unor modele cu performanță sporită prin obținerea unui compromis optim între **bias** și varianță, ținând cont de complexitatea modelelor.



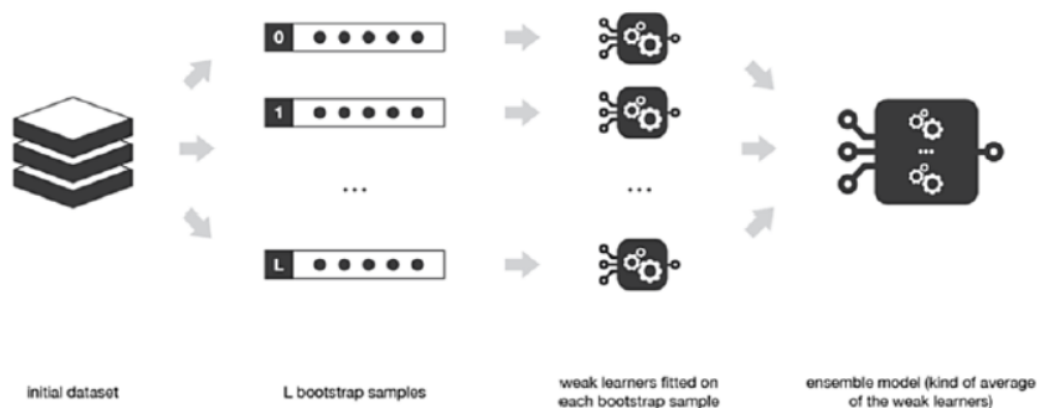
Rocca (2019)

# Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

## Modele de tip bagging

Modelarea de tip ensemble s-a structurat în trei mari tipuri/clase de modele

**Modele de tip bagging (bootstrap aggregating):** folosesc un singur tip de model de bază (cel mai adesea arbori de decizie). Algoritmii sunt implementați folosind estimări independente, realizate pe eșantioane diferite, iar rezultatele acestora sunt combinate într-o manieră deterministă pentru a obține estimarea finală.



Rocca (2019)

## Modalități de agregare a estimărilor

Rezultatele obținute prin estimarea modelelor de bază se agregă apoi pentru obținerea rezultatelor modelului de tip bagging. Considerând că s-au estimat L modele de bază, fiecare pe subeșantioane  $\bar{x}$  extrase din setul de date de antrenare, iar rezultatele acestora sunt de forma  $d_i(\bar{x})$ ,  $i \in \{1, L\}$

Rezultatele se agregă în felul următor:  
 - pe baza mediei în cazul arborilor de regresie (variabilă dependentă de tip continuu)

$$\hat{y} = \frac{1}{L} \sum_{i=1}^L d_i(\bar{x})$$

- pentru modelele cu variabilă dependentă de tip discret (cazul arborilor de clasificare) prin vot majoritar.

$$\hat{y} = \operatorname{argmax}_{d_i(\bar{x})} (d_i(\bar{x}))$$

## Random Forests- Optimizare

Estimarea se poate optimiza (hyperparameter tuning) prin ajustarea următorilor parametri (denumiți în practică hiperparametri):

- $m_{try}$  numărul variabilelor predictive care trebuie folosite în estimarea fiecărui model de bază. Pentru variabile țintă de tip categorial de regulă  $m_{try} = P/3$ , iar pentru cele de tip continuu,  $m_{try} = \sqrt{P}$
- numărul de modele de bază folosite (minimum 500, de preferat peste 1000).

Mai există și alte posibilități de optimizare:

- Numărul minim de observații dintr-un nod terminal
- Numărul maxim de noduri terminale

Deși pot fi utile pentru evitarea supraantrenării, cele două opțiuni pot impune restricții nejustificate asupra modelului

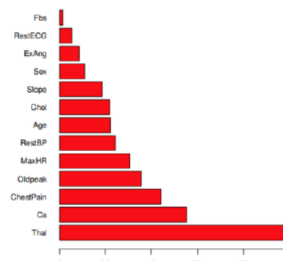
## Indicatorul de importanță a variabilei

Indicatorul variabile importance este esențial în evaluarea puterii explicative a variabilelor predictor la nivelul întregului model.

Pentru modelul bagged trees (similar cu random forests), indicatorul se calculează astfel:

- Pentru arborii de regresie; însumarea reducerii sumei pătratelor reziduurilor ca urmare a folosirii variabilei respective, calculată ca valoare medie pentru toate modelele componente

- Pentru arborii de clasificare, prin însumarea scăderii indicelui Gini obținută în urma folosirii variabilei respective, calculată ca medie pentru toate modelele componente

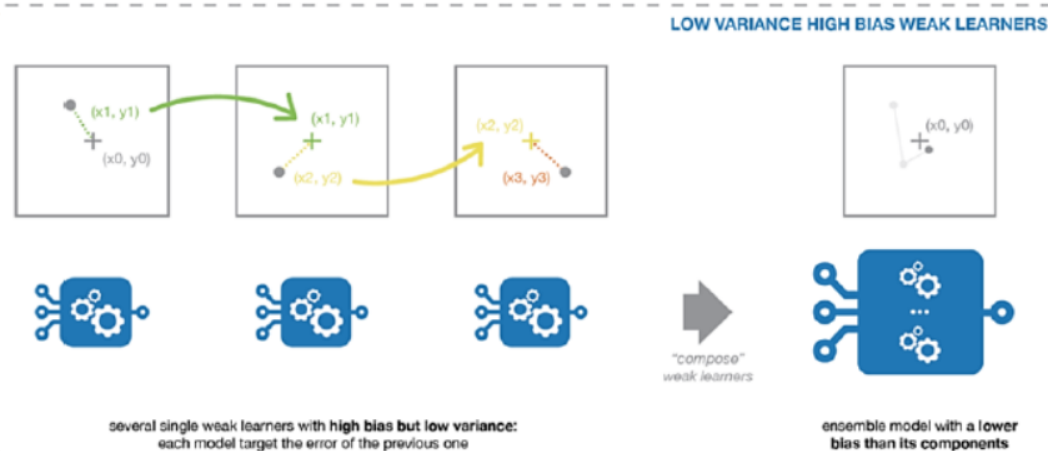


Exemplu: graficul variabile importance pentru variabilă categorială, exprimând scăderea indicelui Gini ca proporție din valoarea maximă.

# Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

## Modele de tip boosting

**Modele de tip boosting:** folosesc **un singur tip de model de bază**. Algoritmii sunt implementați într-o **manieră secvențială, adaptivă**, astfel încât estimările generate de către un model de bază depind de estimările modelelor de bază anterioare. Rezultatul final al estimării se obține prin combinarea rezultatelor **folosind o metodă deterministă**.



Rocca (2019)

## XGBoost în practică I

Pachetul xgboost implementează algoritmul care, pe lângă proprietățile descrise mai sus, oferă și alte funcționalități:

- posibilitatea paralelizării estimăției pe mai multe resurse informatice, putând rula de 10 ori mai rapid comparativ cu GBM
- Formate de date specifice, pentru optimizarea estimării pe matrici rare (sparse matrices) unde predomină valoarea 0.

Găsirea valorilor optime pentru hiperparametri este esențială pentru o estimare eficientă și relativ rapidă:

eta – learning rate, valori în intervalul  $[0,1]$   
gamma – valoare minimă a îmbunătățirii valorii funcției de cost necesară efectuării unei splitări, valori în intervalul  $[0, \infty)$   
max\_depth – adâncimea maximă a arborilor de decizie estimați  
subsample – fracția de observații de eșantionat din setul de date de antrenare pentru a fi folosite în dezvoltarea modelelor de bază

<https://cran.r-project.org/web/packages/xgboost/vignettes/xgboostPresentation.html>

## XGBoost în practică II

lambda și alpha – coeficienți de regularizare a rezultatelor estimării  $w$ , folosite pentru modele de bază de tip liniar. Valori mai mari => estimări mai conservatoare

**Alte opțiuni importante pentru funcția de estimare `xgb.train`**  
objective – modelul de bază folosit (valoare implicită: regresia liniară)  
eval\_metric – indicator de evaluare a rezultatelor estimării în funcție de modelul de bază folosit. De exemplu `eval_metric="rmse"` pentru modele de regresie, `eval_metric="auc"` pentru modele de clasificare  
maximize – legat de valorile specificate. Valoarea TRUE înseamnă că o valoare mai mare este considerată a fi mai bună. Ex. pentru `eval_metric="rmse"`, `maximize=FALSE`

early\_stopping\_rounds – numărul de iterații după care estimarea se oprește dacă rezultatele obținute pe un set de date de validare nu se îmbunătățesc

nrounds – numărul maxim de iterații E

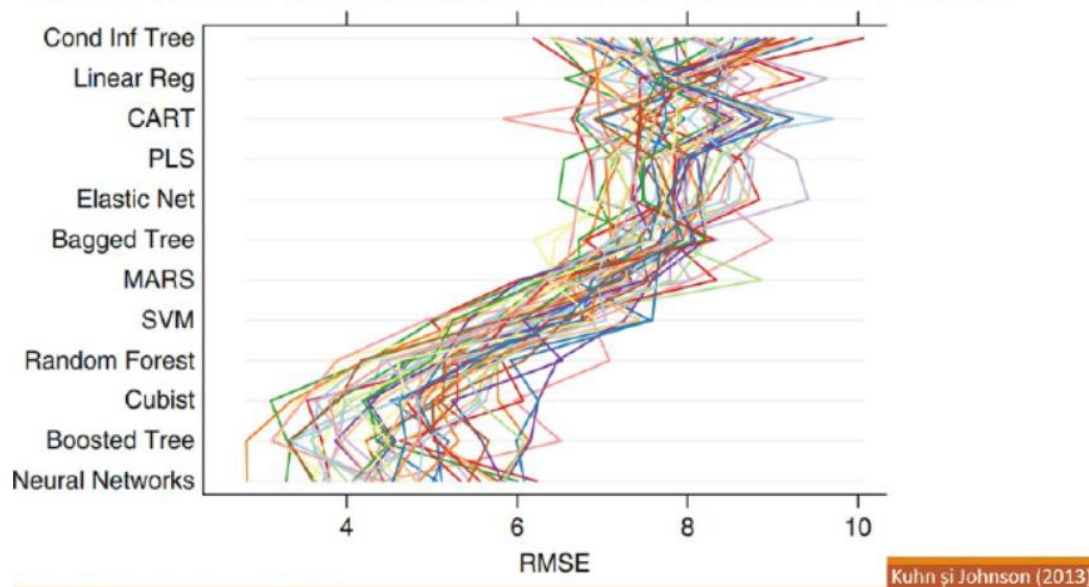
nthread – numărul procesărilor paralele permise. Pe un calculator personal valoarea maximă este dată de numărul de procesoare. Valoarea implicită este setată pentru numărul maxim de procesări permise

<https://xgboost.readthedocs.io/en/latest/parameter.html>

# Modelarea de tip Ensemble: Modele de tip Bagging, Modele de tip Boosting

## Modelarea de tip ensemble. Aprecieri generale I

Modelele de tip ensemble s-au impus în practică datorită performanțelor superioare, datorate gestionării optime a compromisului dintre bias și varianță.



## Modelarea de tip ensemble. Aprecieri generale II

Considerente pentru alegerea celui mai potrivit model

- **Ajustarea hiperparametrilor.** Singurul parametru important pentru modelele random forests este numărul de variabile selectate aleator pentru fiecare splitare, în timp ce modelele de tip gradient boosting folosesc mai mulți hiperparametri, dintre care **learning rate**, **numărul de iterații** și **adâncimea modelelor de bază**.
- **Supraantrenarea**, poate apărea mai rar la modelele random forests deoarece modelele de bază sunt estimate de o manieră independentă
- **Complexitatea modelului.** Modelele random forests ating performanța maximă în urma estimării unui număr mic de arbori de complexitate mare ( $D > 20$ ). Cele de gradient boosting o fac prin antrenarea unui număr mare de arbori de dimensiune redusă ( $D \leq 10$ ).
- **Performanța.** În multe cazuri, modelele de boosting au cea mai bună performanță, cu condiția setării hiperparametrilor de o manieră eficientă

Modelele random forests dau rezultate robuste într-un timp relativ scurt. Modelele de tip boosting au cea mai bună performanță, obținută cu prețul timpului necesar ajustării hiperparametrilor la valori optime.

Kuhn și Johnson (2013) Rokach (2019)



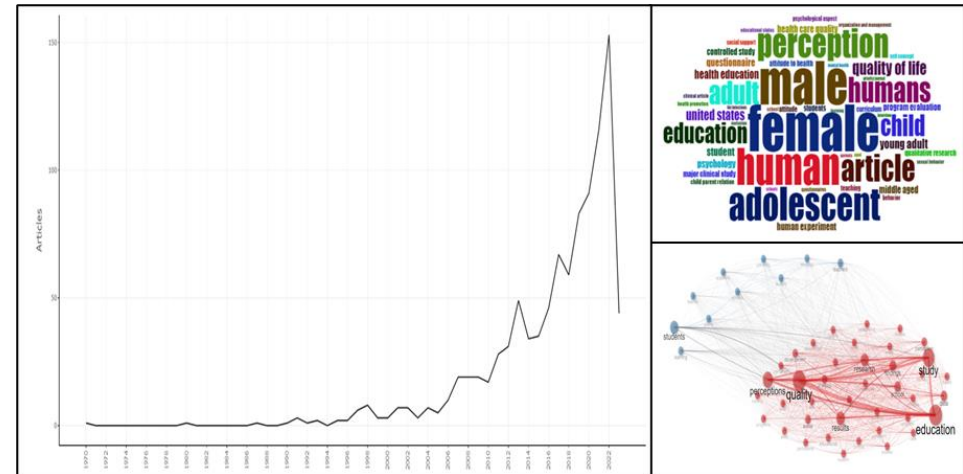
# Explorarea percepțiilor tinerilor despre sistemul educațional prin analiză avansată a datelor în contextul științelor sociale

Utilizând metode avansate de analiză a datelor, se pot identifica și analiza relațiile dintre variabile socio-economice și percepțiile tinerilor despre sistemul educațional. Astfel, se pot evidenția factorii care influențează atitudinile și perspectivele acestora în legătură cu educația.

Prin combinarea analizei bibliometrice cu regresia logistică în studiul percepțiilor tinerilor asupra sistemului educațional din țările balcanice, se poate obține o perspectivă cuprinzătoare și detaliată asupra acestui subiect complex.

- Analiza bibliometrică va furniza o privire retrospectivă asupra literaturii existente, identificând tendințe, domenii de cercetare și contribuții anterioare. Această evaluare a contextului va oferi un fundament solid pentru înțelegerea evoluției subiectului și pentru identificarea direcțiilor de cercetare viitoare în acest domeniu specific din regiunea balcanică.
- În paralel, utilizarea regresiei logice pentru analiza datelor specifice percepțiilor tinerilor va permite identificarea și evaluarea factorilor care influențează percepțiile lor în legătură cu sistemul educațional. Această metodă statistică va ajuta la înțelegerea relațiilor dintre variabilele independente, cum ar fi accesul la resurse educaționale, statutul socio-economic și experiența anterioară în sistemul educațional, și variabila dependentă - percepțiile tinerilor.

Interpretarea și compararea rezultatelor din ambele metode va oferi o imagine comprehensivă și detaliată a factorilor și influențelor care modelează perspectivele tinerilor în ceea ce privește sistemul educațional din țările balcanice



	B	S.E.	Wald	df	Sig.	Exp(B)
Varsta respondentului (1) Female	0.130	0.052	6.255	1	0.021	1.139
Câți oameni trăiesc în total în gospodăria ta?	-0.021	0.008	6.425	1	0.021	0.979
Crezi că în țara ta, pregătirea, învățământul școlar și universitar sunt bine adaptate sau nu la lumea actuală a muncii?(1) Nu este bine adaptată	0.028	0.017	2.624	1	0.105	1.029
Care este cel mai înalt nivel de studii finalizat până acum?	-1.568	0.057	761.233	1	0.000	0.208
Scoală primară	-0.063	0.161	34.242	6	0.000	
Liceu profesional sau tehnic	-0.202	0.175	0.151	1	0.697	0.939
Liceu: tip universitar-pregătitor	-0.046	0.170	1.338	1	0.247	0.817
Educație la nivel universitar: diplomă de licență sau similar	0.061	0.190	0.075	1	0.784	0.955
Educație la nivel universitar: superioară diplomă de licență (diplomă de masterat/MSC)	0.453	0.187	0.107	1	0.744	1.063
Doctorat sau post-doctorat	1.261	0.205	4.895	1	0.027	1.574
Sunteți de acord că există cazuri în care notele și examenele sunt „cumpărate” în institute/universități din țara voastră?	-0.131	0.747	2.852	1	0.091	3.528
Country			35.566	1	0.000	0.878
Bih	1.202	0.107	417.171	9	0.000	
Bulgaria	1.173	0.118	125.168	1	0.000	3.328
Croatia	0.207	0.104	99.598	1	0.000	1.232
Kosovo	-0.390	0.120	3.944	1	0.047	0.684
Macedonia	-0.658	0.120	10.073	1	0.002	2.132
Montenegro	0.757	0.133	24.578	1	0.000	0.518
Romania	0.339	0.138	30.175	1	0.000	2.132
Serbia	0.783	0.117	8.431	1	0.004	1.403
Slovenia	0.805	0.113	47.568	1	0.000	2.188
Slovenia	0.810	0.113	50.794	1	0.000	2.237
Slovenia	0.810	0.113	50.794	1	0.000	2.237
Constant	0.239	11.453	1	0.001	2.247	