



Analiza și Prognoza Seriilor de Timp

Capitolul 10: Recapitulare Completă

Analiză Completă cu Date Reale



Cuprins

- 1 Fluxul Complet de Analiză
- 2 Studiu de Caz 1: Analiza Volatilității Bitcoin
- 3 Studiu de Caz 2: Analiza Ciclului Petelor Solare
- 4 Studiu de Caz 3: Șomajul SUA cu Ruptură Structurală
- 5 Studiu de Caz 4: Analiza Multivariată VAR
- 6 Selecția Modelului: Ghid Practic
- 7 Rezumat și Concluzii Cheie

Metode Clasice

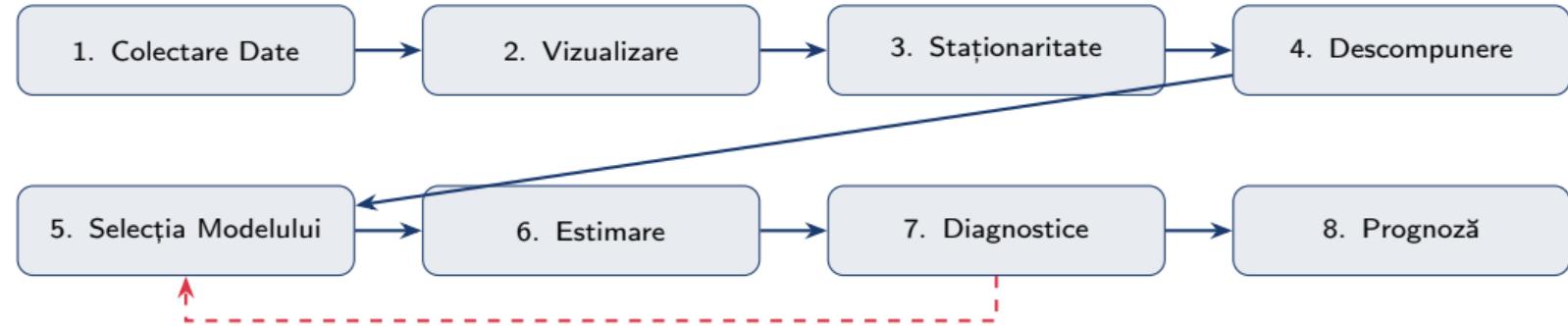
- Cap 1: Fundamentele Seriilor de Timp
- Cap 2: Modele ARMA
- Cap 3: Modele ARIMA
- Cap 4: Modele SARIMA
- Cap 5: Modele GARCH

Metode Avansate

- Cap 6: VAR & Cauzalitate Granger
- Cap 7: Cointegrare & VECM
- Cap 8: Extensiile Moderne
- Cap 9: Prophet & TBATS

Astăzi: Aplicăm TOATE pe Date Reale!

Fluxul Complet de Analiză



Principiu Cheie

Diagnosticele pot necesita revenirea la seleția modelului (proces iterativ)

Seturi de Date Reale pentru Acest Capitol

Bitcoin

- Zilnic 2019-2024
- Clustering volatilitate
- ARIMA + GARCH

Pete Solare

- Anual 1900-2023
- Ciclu de 11 ani
- Termeni Fourier

Economic VAR

- Trimestrial 2000-2023
- PIB, Inflație, etc.
- VAR Multivariat

Metodologie Cheie: Train / Validation / Test Split



De Ce Contează:

- **Training:** Estimare parametri model
- **Validare:** Comparare modele, tuning
- **Test:** Performanță finală nebiasată!

Regulă Critică

NICIODATĂ nu priviți datele de test în timpul selecției!
Pentru serii de timp: **NICIODATĂ** nu amestecați — păstrați ordinea temporală.

Date: Randamente zilnice Bitcoin (2019–2024)

- Sursă: Yahoo Finance (BTC-USD)
- 1.826 observații
- Randamente = $100 \times \ln(P_t/P_{t-1})$

Obiectiv: Prognoza volatilității (nu randamente!)

De ce volatilitatea?

- Randamentele sunt aproape imprevizibile
- Volatilitatea arată persistență puternică
- Critică pentru managementul riscului (VaR)

Împărțire Date:

Set	Zile	Perioadă
Training	1.278	2019–2022
Validare	274	2022–2023
Test	274	2023–2024

Statistică Cheie

Randament mediu: 0.12%/zi

Abatere std: 3.8%/zi

Kurtoză: 8.2 (cozi groase!)

Pas 1: Estimare variante GARCH pe Training

Pas 2: Comparare prognoze volatilitate pe Validare

Model	AIC	BIC	Val MAE
GARCH(1,1)	8542	8563	2.34
GARCH(2,1)	8540	8567	2.31
GJR-GARCH(1,1)	8531	8558	2.18
EGARCH(1,1)	8538	8565	2.25

⇒ Cel mai bun: GJR-GARCH (captează asimetria)

Modelul GJR-GARCH:

$$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Unde $I_{t-1} = 1$ dacă $\varepsilon_{t-1} < 0$

Efectul de Levier

$\gamma > 0$: Randamentele negative cresc volatilitatea
mai mult decât cele pozitive (frică vs lăcomie)

Pas 3: Re-estimare GJR-GARCH pe Training+Validare, evaluare pe Test

Rezultate Set Test:

Metrică	Valoare
Volatilitate MAE	2.21
Volatilitate RMSE	3.45

Interpretare:

- Persistență ridicată (≈ 0.98)
- Efect de levier semnificativ ($\gamma > 0$)
- Volatilitatea este previzibilă!

Parametri GJR-GARCH:

- $\alpha = 0.05$ (efect ARCH)
- $\gamma = 0.08$ (efect levier)
- $\beta = 0.89$ (persistență)
- $\alpha + \gamma/2 + \beta = 0.98$

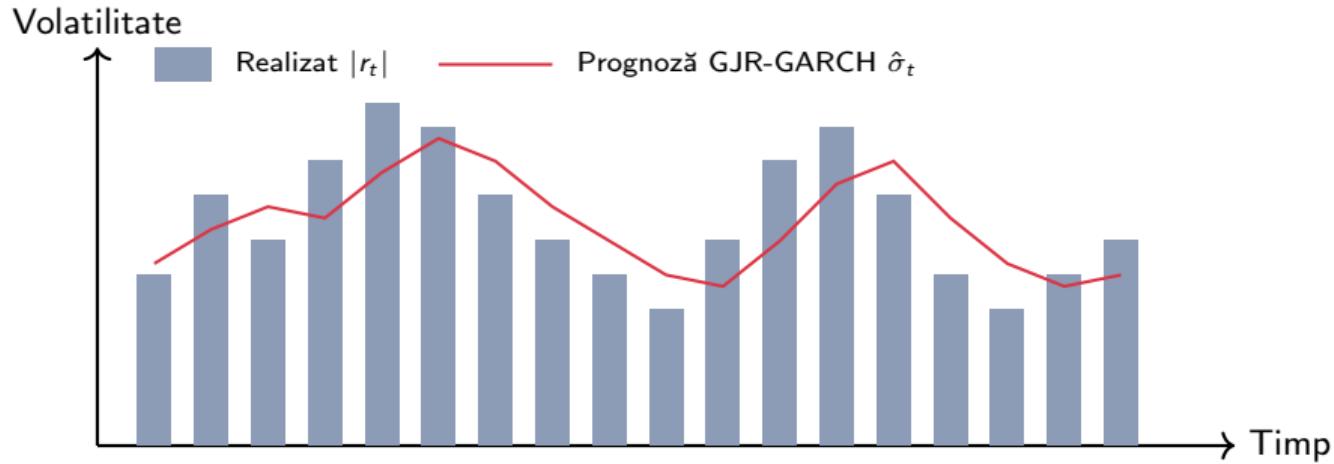
Aplicație Practică

VaR 1 zi (99%):

$$\text{VaR} = -2.33 \times \hat{\sigma}_{t+1}$$

Dacă $\hat{\sigma} = 4\%$, atunci $\text{VaR} = -9.3\%$

Perioada Test: Prognoză vs Volatilitate Realizată



- Modelul captează clusterele de volatilitate
- Prognoza se adaptează la condițiile de piață
- Test MAE = 2.21 (bun pentru prognoză zilnică)

Metodologie:

- ① Împărțire date: 70/15/15
- ② Comparare variante GARCH pe validare
- ③ Selectare cel mai bun model (GJR-GARCH)
- ④ Evaluare finală pe test

Constatare Cheie:

Randamentele sunt imprevizibile, dar **volatilitatea este previzibilă!**

Rezultate:

Cel mai bun model	GJR-GARCH(1,1)
Validare MAE	2.18
Test MAE	2.21
Insight cheie	Efect de levier

Utilizare Practică

- Calcul Value-at-Risk
- Dimensionare poziții
- Evaluare opțiuni

Pete Solare: Date și Obiectiv

Date: Numere anuale pete solare (1900–2023)

- Sursă: statsmodels (dataset Wolfer)
- 124 observații anuale
- Celebrul ciclu Schwabe de 11 ani

Obiectiv: Prognoza activității solare

Provocare:

- SARIMA standard cu $m = 11$ este complex
- Soluție: Termeni Fourier ca regresori

Împărțire Date:

Set	Ani	Perioadă
Training	87	1900–1986
Validare	18	1987–2004
Test	19	2005–2023

Statistică Cheie

Medie: 76.4

Abatere std: 56.8

Perioadă ciclu: ≈ 11 ani

Idee: Capturăm ciclul de 11 ani cu regresori sinus/cosinus

$$y_t = \mu + \sum_{k=1}^K \left[a_k \sin\left(\frac{2\pi k t}{11}\right) + b_k \cos\left(\frac{2\pi k t}{11}\right) \right] + \text{erori ARMA}$$

Câte armonici (K)?

- K=1: Formă de ciclu de bază
- K=2: Vârfuri/văi mai clare
- K=3,4: Mai mult detaliu
- Prea multe \Rightarrow suprareprezentare

Selectăm K folosind setul de validare!

De Ce Fourier?

- Doar $2K$ parametri (nu 11)
- Nu e nevoie de diferențiere sezonieră
- Funcționează pentru orice perioadă
- Se combină natural cu ARIMA

Pete Solare: Selectia Modelului pe Validare

Model: ARIMA(2,0,1) + K armonici Fourier

K	AIC	BIC	Val RMSE	Val MAE
1	892	905	28.4	22.1
2	878	896	24.2	18.9
3	875	898	25.1	19.5
4	873	901	26.8	21.2

⇒ Cel mai bun: K=2 armonici Fourier

K=3,4 supraesantionează: AIC mai mic dar eroare validare
mai mare!

Model Selectat:

ARIMA(2,0,1) + 2 perechi Fourier

Parametri:

- $\phi_1 = 1.35, \phi_2 = -0.68$
- $\theta_1 = -0.42$
- 4 coeficienți Fourier

Total: 8 parametri

Re-estimare pe Training+Validare, prognoză Test (2005–2023)

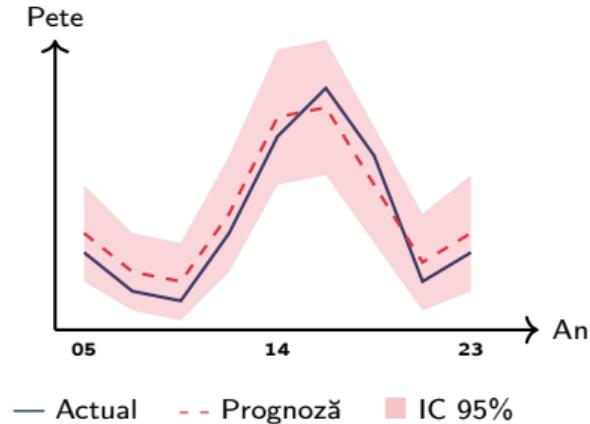
Rezultate Set Test:

Metrică	Valoare
Test RMSE	26.8
Test MAE	21.4
Test MAPE	42.3%

Interpretare:

- Modelul captează timing-ul ciclului
- Amplitudinea mai greu de prezis
- Intervale de încredere largi

Prognoză vs Actual (2005–2023):



Metodologie:

- ① Împărțire: 70/15/15 (ani)
- ② Folosire termeni Fourier pentru ciclu 11 ani
- ③ Selectare K pe validare (K=2 optim)
- ④ Evaluare finală pe test

Constatare Cheie:

Termenii Fourier captează eficient tiparele sezoniere lungi fără a pierde date prin diferențiere.

Rezultate:

Cel mai bun model	ARIMA(2,0,1)+Fourier(2)
Validare RMSE	24.2
Test RMSE	26.8
Insight cheie	K=2 optim

Lecție

Pentru perioade sezoniere lungi:

- Fourier > SARIMA
- Selectați K prin validare încruciată

Şomaj: Date și Provocare

Date: Rata Șomajului SUA (2010–2023)

- Sursă: FRED (UNRATE)
- 168 observații lunare
- Ruptură structurală majoră: COVID-19

Provocarea:

- Aprilie 2020: 3.5% → 14.7% (+10.3 pp)
- Cea mai mare creștere lunară din istorie
- ARIMA tradițional nu face față

Soluție: Prophet cu detectie changepoints

Împărțire Date:

Set	Luni	Perioadă
Training	118	2010–2019
Validare	25	2020–2021
Test	25	2022–2023

Notă

Validarea include şocul COVID — testează adaptabilitatea modelului

Prophet: Detectia Changepoints

Parametru Cheie: `changepoint_prior_scale`

Ce controlează:

- Flexibilitatea schimbărilor de trend
- Mic (0.01): Trend neted, rigid
- Mare (0.5): Multe changepoints permise

Pentru date COVID:

- Nevoie de flexibilitate pentru salt brusc
- Dar nu prea mult (supraesantionare)
- Selectăm prin validare!

Modelul Prophet:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Unde:

- $g(t)$: Trend liniar pe bucăți
- $s(t)$: Sezonalitate (Fourier)
- $h(t)$: Efecte sărbători
- Changepoints: Unde panta $g(t)$ se schimbă

Testare valori diferite changepoint_prior_scale:

Scale	Changepoints	Val RMSE	Val MAE
0.01	2	2.85	2.31
0.05	4	1.92	1.54
0.10	5	1.24	0.98
0.30	8	1.31	1.05
0.50	12	1.48	1.19

⇒ Cel mai bun: scale = 0.10

Changepoints detectate:

- Martie 2020 (început COVID)
- Aprilie 2020 (vârf)
- Iunie 2020 (început recuperare)

De ce 0.10 câştigă:

- Destul de flexibil pentru COVID
- Nu supraesantionează zgomotul
- Echilibrează bias/varianță

Insight Cheie

scale = 0.50 a detectat 12 changepoints — prea multe!

Modelul urmărea zgomotul pe termen scurt.

Re-estimare Prophet (scale=0.10) pe Training+Validare, prognoză Test (2022–2023)

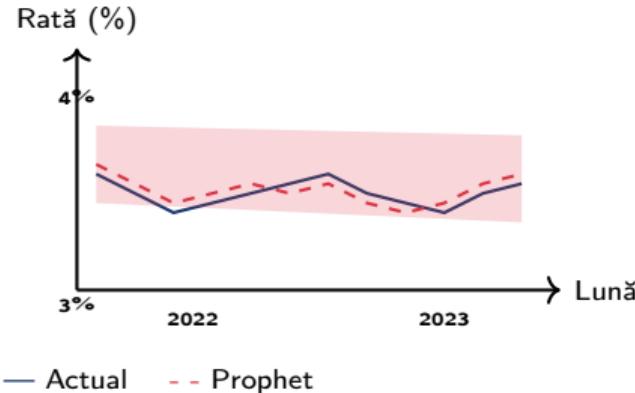
Rezultate Set Test:

Metrică	Valoare
Test RMSE	0.42
Test MAE	0.35
Test MAPE	9.8%

Comparatie cu ARIMA:

Model	Test RMSE
ARIMA(2,1,2)	0.89
Prophet	0.42

Vizualizare Prognoză (2022–2023):



Metodologie:

- ① Împărțire: 70/15/15 (luni)
- ② Tuning changepoint_prior_scale
- ③ Selectare scale optim pe validare
- ④ Evaluare finală pe test

Constatare Cheie:

Prophet gestionează rupturile structurale mai bine decât ARIMA prin detecția automată a changepoints.

Rezultate:

Cel mai bun model	Prophet (scale=0.10)
Validare RMSE	1.24
Test RMSE	0.42
vs ARIMA	53% mai bun

Lecție

Pentru date cu rupturi structurale:

- Prophet > ARIMA tradițional
- Ajustați flexibilitatea changepoints

VAR: Date și Obiectiv

Date: Indicatori Economici SUA (2001–2023)

- Sursă: FRED
- 92 observații trimestriale
- 4 variabile (PIB, Șomaj, Inflație, Rata Fed)

Obiectiv: Prognoză conjunctă + analiza relațiilor

De ce VAR?

- Variabilele se influențează reciproc
- Bucle de feedback ($\text{PIB} \leftrightarrow \text{\v{S}omaj}$)
- Transmisia politicii ($\text{Fed} \rightarrow \text{Economie}$)

Împărțire Date:

Set	Trimestre	Perioadă
Training	64	2001–2016
Validare	14	2017–2020
Test	14	2021–2023

Variabile:

- Creștere PIB (YoY %)
- Șomaj (%)
- Inflație (CPI YoY %)
- Rata Fed Funds (%)

VAR: Selectia Ordinului Lag pe Validare

Model VAR(p): $y_t = c + A_1y_{t-1} + \dots + A_py_{t-p} + \epsilon_t$

Selectare ordine lag p :

Lag	AIC	BIC	Val RMSE
1	12.4	13.1	1.85
2	11.8	13.0	1.42
3	11.6	13.4	1.51
4	11.5	13.8	1.68

⇒ Cel mai bun: VAR(2)

Lag-uri mai mari: AIC mai mic dar eroare validare mai mare
(supraesantionare)

Parametri VAR(2):

- 4 variabile \times 2 lag-uri = 8 coef per ecuație
- + 4 intercepte
- Total: 36 parametri

BIC vs AIC

BIC selectează $p=2$ (mai simplu)

AIC selectează $p=3$ (complex)

Validarea confirmă BIC!

VAR: Rezultate Cauzalitate Granger

Testare relații predictive (pe date Training):

Cauză → Efect	F-stat	p-value	Sig.
PIB → Șomaj	4.82	0.012	**
Șomaj → PIB	1.45	0.243	
Inflație → Rata Fed	6.21	0.004	**
Rata Fed → Inflație	2.88	0.065	*
PIB → Inflație	3.12	0.052	*
Fed → Șomaj	2.15	0.127	

** p<0.05, * p<0.10

Interpretare:

- PIB conduce șomajul (Legea Okun confirmată)
- Inflația conduce politica Fed (Regula Taylor)
- Bidirectional: Fed ↔ Inflație

Atenție

Cauzalitate Granger = predictivă, nu cauzalitate reală!

VAR: Evaluare Finală pe Test

Re-estimare VAR(2) pe Training+Validare, prognoză Test (2021–2023)

Rezultate Test (per variabilă):

Variabilă	RMSE	MAE	vs AR(2)
Creștere PIB	1.82	1.45	-18%
Șomaj	0.58	0.44	-25%
Inflație	1.24	0.98	-12%
Rata Fed	0.89	0.72	-31%
Medie	1.13	0.90	-22%

VAR(2) bate AR(2) univariat cu 22% în medie!

De ce VAR câștigă:

- Folosește informația inter-variabile
- PIB ajută la predicția șomajului
- Inflația ajută la predicția Fed

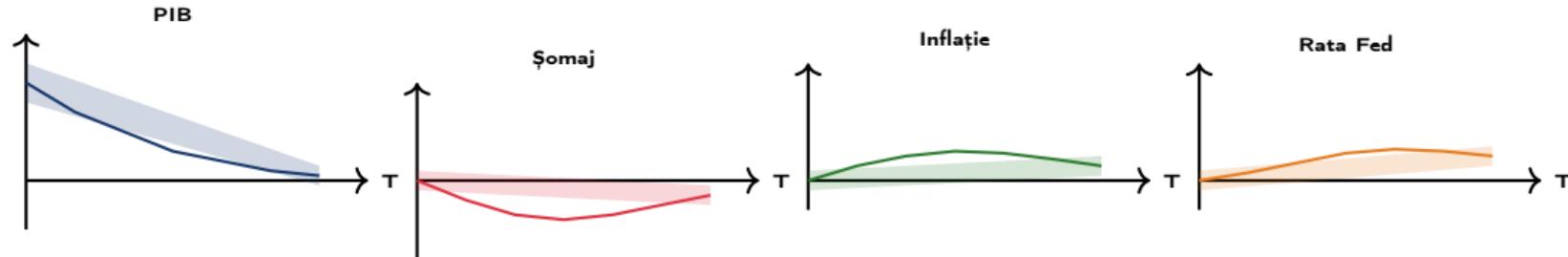
Cea Mai Mare Îmbunătățire

Rata Fed: -31% RMSE

De ce? Inflația este un indicator puternic pentru politica Fed.

VAR: Analiză Funcții Răspuns la Impuls

Cum afectează un soc de 1% PIB celelalte variabile?



Interpretare (din IRF):

- Soc PIB \Rightarrow Șomajul scade 4-6 trimestre (Legea Okun)
- Soc PIB \Rightarrow Inflația crește, vârf la T3-T4 (cerere-push)
- Soc PIB \Rightarrow Rata Fed crește cu întârziere (răspuns politică)

Metodologie:

- ① Împărțire: 70/15/15 (trimestre)
- ② Selectare ordine lag pe validare ($p=2$)
- ③ Testare cauzalitate Granger
- ④ Analiză IRF
- ⑤ Prognoză finală pe test

Constatare Cheie:

VAR captează interdependențele economice pe care modelele univariate le pierd.

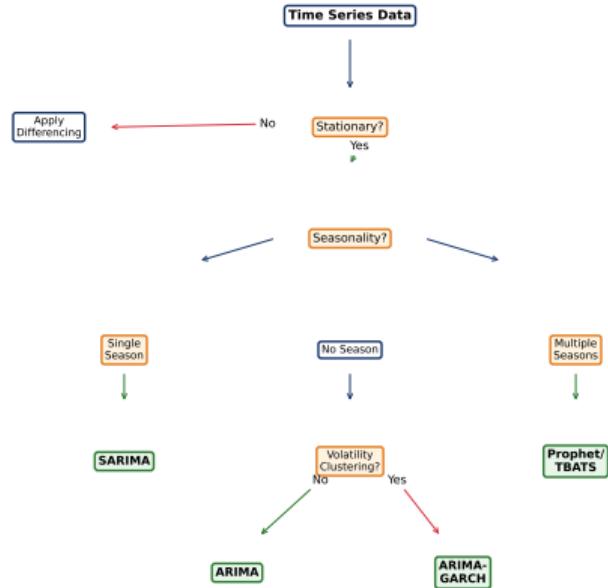
Rezultate:

Cel mai bun model	VAR(2)
Validare RMSE	1.42
Test RMSE	1.13
vs Univariat	22% mai bun

Aplicații

- Prognoză macroeconomică
- Analiză impact politici
- Risc portofoliu (multi-activ)

Cadrul de Decizie



Rezumat Selectie Model

Tip Date	Caracteristici	Model Recomandat	Alternative
Randamente finan- ciare	Fără trend, clustering vol.	ARIMA-GARCH	EGARCH, GJR
Sezonalitate simplă	Trend + o perioadă sezon.	SARIMA	ETS, Prophet
Cicluri lungi	Pete solare, cicluri business	AR + Fourier, TBATS	Metode spectrale
Rupturi structurale	COVID, schimbări regim	Prophet	ARIMA intervenție
Serii multiple	Interdependențe	VAR, VECM	Modele factoriale

Metrici pentru Prognoze Punctuale:

RMSE (Root Mean Square Error):

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

MAE (Mean Absolute Error):

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAPE (Mean Absolute % Error):

$$\frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Când să Folosim Fiecare:

- **RMSE**: Penalizează erorile mari
- **MAE**: Robust la outlieri
- **MAPE**: Independent de scală

Validare Încrucișată

Folosiți întotdeauna CV pentru serii de timp:

- Fereastră rulantă
- Fereastră expandabilă
- Nu amestecați niciodată!

Înțelegerea Datelor

- Vizualizare mai întâi!
- Testați staționaritatea (ADF, KPSS)
- Identificați tipare sezoniere
- Verificați rupturi structurale

Modele Clasice

- ARIMA: Date nesezoniere
- SARIMA: Sezonalitate simplă
- GARCH: Modelarea volatilității

Abordări Moderne

- Prophet: Interpretabil, gestionează rupturi
- TBATS: Sezonalități multiple/lungi
- VAR/VECM: Serii de timp multiple

Cele Mai Bune Practici

- Verificați întotdeauna diagnosticile
- Folosiți validare încrucișată
- Comparați modele multiple
- Cunoașterea domeniului contează!

Recomandări Finale

- ① **Începeți Simplu:** Începeți cu vizualizare și statistici de bază
- ② **Testați Ipotezele:** Staționaritate, normalitate, independență
- ③ **Iterați:** Model → Diagnostics → Îmbunătățire
- ④ **Comparați:** Nu vă bazați niciodată pe un singur model
- ⑤ **Validați:** Testarea out-of-sample este esențială
- ⑥ **Comunicați:** Vizualizări și interpretări clare

Amintiți-vă

“Toate modelele sunt greșite, dar unele sunt utile.” — George Box

Scopul nu este predicția perfectă, ci perspective utile și prognoze rezonabile.

Întrebări?

Pași Următori:

- Exersați cu notebook-ul Jupyter
- Aplicați aceste metode pe propriile date
- Comparați diferite modele pe același set de date

Materiale Curs: github.com/danpele/Time-Series-Analysis