

# 🏦 Modele de predicție a prețului Bitcoin

## ANALIZĂ DE DATE ȘI ALGORITMI DE ÎNVĂȚARE AUTOMATĂ PENTRU PREDICȚIA PREȚULUI BITCOIN

Autor: Cristiana Kovacs

### 🎯 Obiectivul proiectului

Scopul acestui proiect este dezvoltarea și evaluarea unor modele de învățare automată pentru **prognoza prețului de închidere al Bitcoin**, pe baza unui set extins de date:

- financiare (prețuri, indici bursieri),
- macroeconomice (inflație, rate ale dobânzilor),
- on-chain (indicatori blockchain),
- și parțial, date de sentiment.

Obiectivul este **obținerea unei erori de predicție cât mai reduse (MAE)** și identificarea variabilelor cu influență majoră asupra prețului Bitcoin.

### 💡 Aplicații practice posibile

- Suport pentru **investitori și traderi**, ca punct suplimentar de referință în luarea deciziilor.
- Analiză de **scenarii „ce s-ar întâmpla dacă?”** (impactul factorilor externi asupra prețului).
- Monitorizarea **factorilor cu cea mai mare influență** (variabile importante identificate de model).
- **Monitorizarea automată** a prețului BTC în timp real, cu alerte atunci când deviația depășește un prag (ex.  $\pm 2\%$ ).
- **Predicții zilnice** sau pe orizont scurt (1–3 zile) pentru sisteme de risk management sau pentru ajustarea pozițiilor de tranzacționare.
- **Suport pentru analiza** scenariilor („ce s-ar întâmpla dacă?”) prin simularea diferitelor evoluții de preț.
- Integrare într-un **dashboard de forecast** pentru vizualizarea simultană a predicției și a deviației procentuale în timp real.

99.2/99.2 MB 7.0 MB/s eta 0:00:00

Instalările au fost facute cu succes.

Bibliotecile au fost importate cu succes.

## 🧱 Etapa I: Construirea și curățarea bazei de date

### 💎 Surse de date utilizate

Tip date	Sursă	Variabile principale
Prețuri Bitcoin	Yahoo Finance	Open, High, Low, Close, Volume
Active financiare	Yahoo Finance	Gold, Oil, SP500, DXY, VIX
Randamente TIPS	FRED	TIPS 5Y, 10Y
Date macroeconomice	FRED	Fed Funds Rate, CPI, Unemployment
Date on-chain	Blockchain.com API	n-transactions, hash-rate, difficulty

Tip date	Sursă	Variabile principale
Sentiment (limitat)	CoinDesk, Reddit	scoruri pozitive/negative

## Preprocesare

1. Conversie coloanei `Date` la format `datetime`.
2. Conversie numerică a coloanelor relevante (float, 2 zecimale).
3. Completarea valorilor lipsă – metodă *forward fill*.
4. Detectarea outlierilor – metodă *IQR (Interquartile Range)*.
  - `Q1 - 1.5*IQR` și `Q3 + 1.5*IQR`
  - Marcarea rândurilor cu valori extreme, fără a le elimina imediat
5. Verificarea consistenței prețurilor Bitcoin:
  - `Low ≤ Open/Close ≤ High`.
6. **Unificarea** dataseturilor pe coloana `Date` (outer join).
7. Sortarea datelor cronologic.
8. Rezultat: dataset integrat cu toate variabilele relevante (financiare, macro, sentiment, on-chain).


## Curățarea dataset final

1. Eliminarea rândurilor incomplete la începutul seriei (care conțin NaN în majoritatea coloanelor).
2. Forward fill pentru completarea valorilor lipsă restante.
3. Asigurarea consistenței finale a datelor și a coloanelor critice


## Validareadata set

1. Vizualizarea primelor 20 de rânduri pentru verificarea consistenței.
2. Confirmarea respectării relațiilor logice între prețurile Bitcoin (**`Low ≤ Open/Close ≤ High`**).
3. Analiza procentului de outlieri pentru a evalua calitatea datasetului **\*\***(percentila 99.9%).

## Rezultatul final

 Baza de date finală conține **3967 rânduri** și **19 coloane**.

 Număr de linii care conțin cel puțin un **outlier: 925**

 **Procentul** rândurilor cu **outlieri: 23.32% din total**.

- valori care ies semnificativ din intervalul obișnuit (adică outlierii) nu înseamnă neapărat erori, ci reflectă evenimente reale și relevante pe piață.
- eliminarea lor ar fi echivalentă cu ștergerea unor evenimente importante din - setul de date, ceea ce ar distorsiona analiza statistică.
- practic, 1 din 4 sau 1 din 5 rânduri conține o valoare extremă, dar aceste valori pot fi chiar cele mai interesante pentru analize: ele indică crize, vârfuri sau scăderi abrupte, care sunt esențiale pentru înțelegerea comportamentului pieței.

**Nu elimin outlierii: pentru date financiare volatile precum Bitcoin, ei sunt parte din realitate!**

Primele 30 de rânduri din fișier:

	Date	Open	High	Low	Close	Volume	GOLD_Close	OIL_Close	SP500_Close	DXY_Close	VIX_Close	TIPS_5Y_Yield	TIPS_10Y_Yield	Fed Funds Rate	CPI	Unemployment Rate	n- transactions	hash-rate	difficulty
0	2014-09-17	465.86	468.17	452.42	457.33	21056800.0	1234.4	94.42	2001.57	84.70	12.65	0.19	0.56	0.09	237.48	5.9	80305	280257.53	2.982973e+10
1	2014-09-18	456.86	456.86	413.10	424.44	34483200.0	1225.7	93.07	2011.36	84.32	12.03	0.29	0.61	0.09	237.48	5.9	77364	274326.15	2.982973e+10
2	2014-09-19	424.10	427.83	384.53	394.80	37919700.0	1215.3	92.41	2010.40	84.80	12.11	0.27	0.56	0.09	237.48	5.9	69437	240220.74	2.982973e+10
3	2014-09-20	394.67	423.30	389.88	408.90	36863600.0	1215.3	92.41	2010.40	84.80	12.11	0.27	0.56	0.09	237.48	5.9	63482	256532.03	2.982973e+10
4	2014-09-21	408.08	412.43	393.18	398.82	26580100.0	1215.3	92.41	2010.40	84.80	12.11	0.27	0.56	0.09	237.48	5.9	59800	259497.71	2.982973e+10
5	2014-09-22	399.10	406.92	397.13	402.15	24127600.0	1216.8	91.52	1994.29	84.67	13.69	0.20	0.58	0.09	237.48	5.9	73289	232806.52	2.982973e+10
6	2014-09-23	402.09	441.56	396.20	435.79	45099500.0	1221.0	91.56	1982.77	84.66	14.93	0.15	0.52	0.09	237.48	5.9	77283	223909.46	2.982973e+10
7	2014-09-24	435.75	436.11	421.13	423.20	30627700.0	1218.6	92.80	1998.30	85.05	13.27	0.16	0.54	0.09	237.48	5.9	74974	241703.58	2.982973e+10
8	2014-09-25	423.16	423.52	409.47	411.57	26814400.0	1221.2	92.53	1965.99	85.20	15.64	0.13	0.50	0.09	237.48	5.9	78523	241703.58	2.982973e+10
9	2014-09-26	411.43	414.94	400.01	404.42	21460800.0	1214.1	93.54	1982.85	85.64	14.85	0.21	0.57	0.09	237.48	5.9	72555	260177.35	3.466143e+10
10	2014-09-27	403.56	406.62	397.37	399.52	15029300.0	1214.1	93.54	1982.85	85.64	14.85	0.21	0.57	0.09	237.48	5.9	62973	220547.69	3.466143e+10
11	2014-09-28	399.47	401.02	374.33	377.18	23613300.0	1214.1	93.54	1982.85	85.64	14.85	0.21	0.57	0.09	237.48	5.9	62605	239501.01	3.466143e+10
12	2014-09-29	376.93	385.21	372.24	375.47	32497700.0	1217.5	94.57	1977.80	85.59	15.98	0.16	0.55	0.09	237.48	5.9	77924	237777.98	3.466143e+10
13	2014-09-30	376.09	390.98	373.44	386.94	34707300.0	1210.5	91.16	1972.29	85.94	16.31	0.22	0.55	0.09	237.48	5.9	78006	261900.38	3.466143e+10
14	2014-10-01	387.43	391.38	380.78	383.61	26229400.0	1214.6	90.73	1946.16	85.97	16.71	0.09	0.46	0.09	237.43	5.7	78779	249839.18	3.466143e+10
15	2014-10-02	383.99	385.50	372.95	375.07	21777700.0	1214.2	91.01	1946.17	85.60	16.16	0.13	0.49	0.09	237.43	5.7	72059	255008.27	3.466143e+10
16	2014-10-03	375.18	377.70	357.86	359.51	30901200.0	1192.2	89.74	1967.90	86.69	14.55	0.17	0.52	0.09	237.43	5.7	75121	242947.06	3.466143e+10
17	2014-10-04	359.89	364.49	325.89	328.87	47236500.0	1192.2	89.74	1967.90	86.69	14.55	0.17	0.52	0.09	237.43	5.7	66632	246393.12	3.466143e+10
18	2014-10-05	328.92	341.80	289.30	320.51	83308096.0	1192.2	89.74	1967.90	86.69	14.55	0.17	0.52	0.09	237.43	5.7	67330	306699.13	3.466143e+10
19	2014-10-06	320.39	345.13	302.56	330.08	79011800.0	1206.7	90.34	1964.82	85.71	15.46	0.15	0.50	0.09	237.43	5.7	78049	275684.61	3.466143e+10
20	2014-10-07	330.58	339.25	320.48	336.19	49199900.0	1211.7	88.85	1935.10	85.67	17.20	0.12	0.44	0.09	237.43	5.7	79069	260177.35	3.466143e+10

	Date	Open	High	Low	Close	Volume	GOLD_Close	OIL_Close	SP500_Close	DXY_Close	VIX_Close	TIPS_5Y_Yield	TIPS_10Y_Yield	Fed Funds Rate	CPI	Unemployment Rate	n- transactions	hash-rate	difficulty
21	2014-10-08	336.12	354.36	327.19	352.94	54736300.0	1205.3	87.31	1968.89	85.30	15.11	0.03	0.39	0.09	237.43	5.7	78818	225716.78	3.466143e+10
22	2014-10-09	352.75	382.73	347.69	365.03	83641104.0	1224.6	85.77	1928.21	85.52	18.76	0.06	0.37	0.09	237.43	5.7	82789	247077.56	3.500248e+10
23	2014-10-10	364.69	375.07	352.96	361.56	43665700.0	1221.0	85.82	1906.13	85.91	21.24	0.03	0.35	0.09	237.43	5.7	81058	276657.27	3.500248e+10
24	2014-10-11	361.36	367.19	355.95	362.30	13345200.0	1221.0	85.82	1906.13	85.91	21.24	0.03	0.35	0.09	237.43	5.7	65020	248817.54	3.500248e+10
25	2014-10-12	362.61	379.43	356.14	378.55	17552800.0	1221.0	85.82	1906.13	85.91	21.24	0.03	0.35	0.09	237.43	5.7	63509	257517.46	3.500248e+10
26	2014-10-13	377.92	397.23	368.90	390.41	35221400.0	1229.3	85.74	1874.74	85.43	24.64	0.03	0.35	0.09	237.43	5.7	77942	266217.37	3.500248e+10
27	2014-10-14	391.69	411.70	391.32	400.87	38491500.0	1233.6	81.84	1877.70	85.82	22.79	0.04	0.29	0.09	237.43	5.7	81364	243597.59	3.500248e+10
28	2014-10-15	400.95	402.23	388.77	394.77	25267100.0	1244.1	81.78	1862.49	85.15	25.27	0.00	0.29	0.09	237.43	5.7	76257	255777.47	3.500248e+10
29	2014-10-16	394.52	398.81	373.07	382.56	26990000.0	1240.5	82.70	1862.76	85.00	25.20	0.02	0.28	0.09	237.43	5.7	75737	260997.42	3.500248e+10



## Etapa II: Analiză exploratorie a datelor (EDA)



### Obiective:

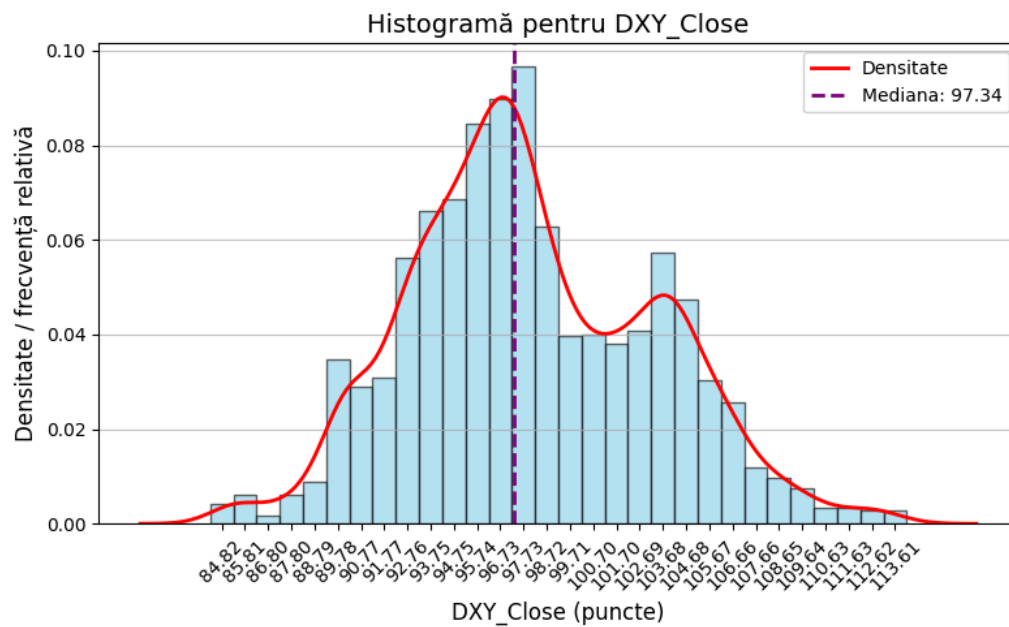
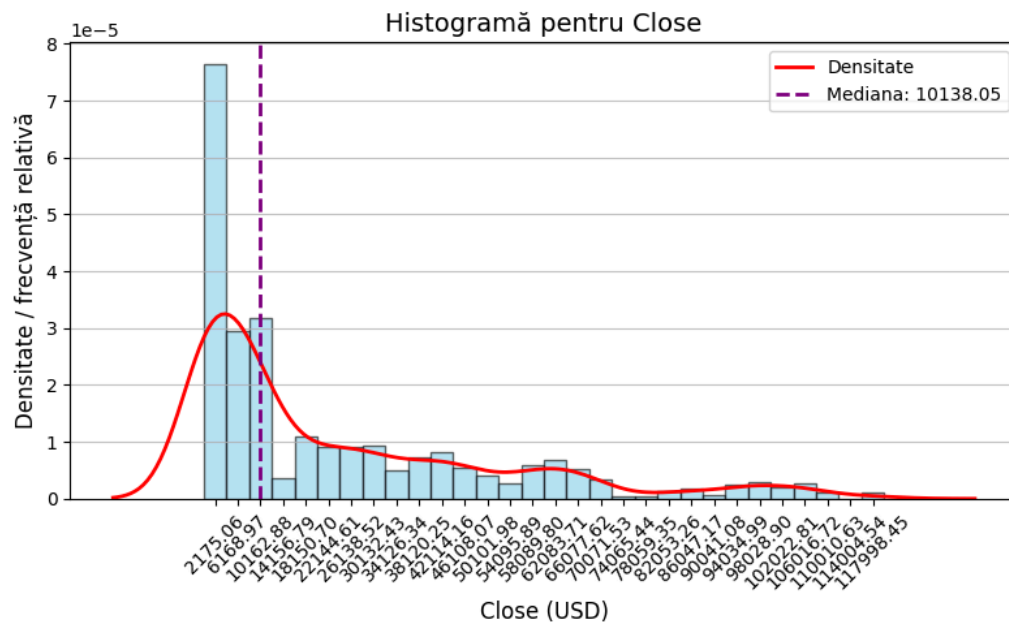
- Înțelegerea distribuției datelor și a relațiilor dintre variabile.
- Identificarea tendințelor, sezonality și pattern-urilor majore.



### 1. Histogramă pentru fiecare variabilă:

- Distribuția prețurilor Bitcoin este **puternic asimetrică spre dreapta (positively skewed)** – multe valori mici, puține valori extreme.
- Majoritatea valorilor sunt concentrate sub 20,000 USD, dar există valori extreme foarte mari (până la ~120,000 USD), **coadă lungă spre dreapta**.
- Majoritatea tranzacțiilor istorice au avut prețuri relativ mici, dar au existat câteva perioade cu creșteri foarte mari ale prețului.
- Această distribuție reflectă **volatilitatea ridicată a Bitcoin**.

- Puterea dolarului față de un coș de valute majore.
- **Mediana (~97.34 puncte)** → istoric, un nivel mediu, dar cu variații semnificative între 85 și 115.
- **Formă aproape normală, dar cu cozi:** sugerează episoade de întărire/ slăbire puternică a USD în crize sau perioade de politică monetară agresivă.
- dolar puternic (DXY↑) → capitalul global se refugiază în USD, presiune negativă pe Bitcoin și active riscante
- dolar slab (DXY↓) → stimulează lichiditatea și fluxurile spre Bitcoin și piețe emergente.

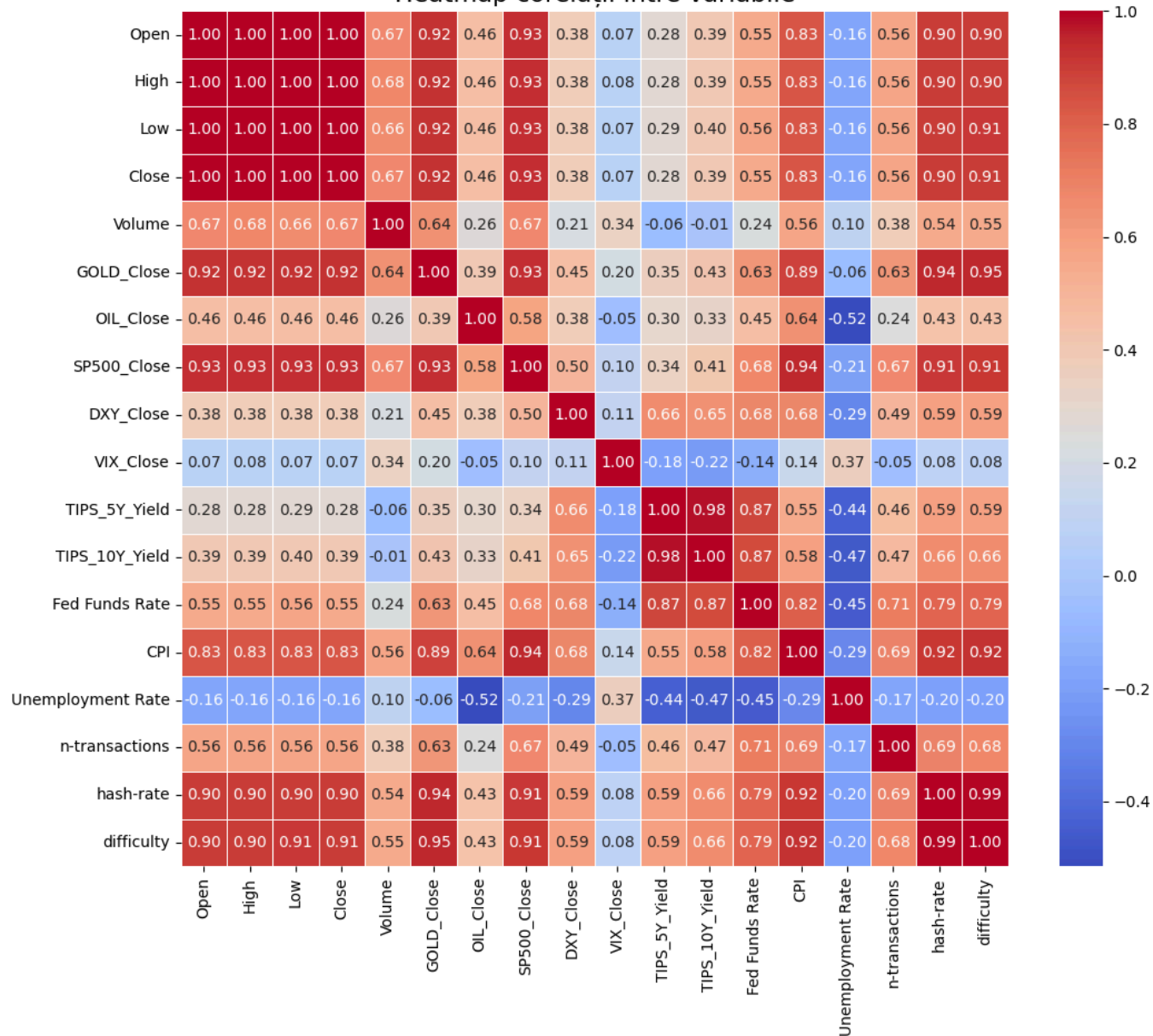


## 2. Heatmap pentru corelații:

**Matrice de corelații Pearson** pentru toate variabilele numerice selectate:

- Valorile merg de la **-1 (corelație negativă puternică)** la **+1 (corelație pozitivă puternică)**.
- **Zonele roșii** → corelație pozitivă mare (când o variabilă crește, crește și cealaltă).
- **Zonele albastre** → corelație negativă (una urcă, cealaltă coboară).
- **Zonele aproape albe** → corelație foarte slabă / aproape inexistentă.
- **Open, High, Low, Close** sunt corelate aproape perfect între ele (0.99–1.00). Asta e normal: toate sunt prețuri derivate din aceeași zi.
- **SP500\_Close (0.93)** → evoluția criptomonedei (Bitcoin) e foarte legată de bursa americană.

Heatmap corelații între variabile



⬆️ Top 5 corelații pozitive:

- **hash-rate** ↔ **difficulty** = **0.99** → aproape perfect corelate (logic, dificultatea urmează puterea de calcul).
- **TIPS\_5Y\_Yield** ↔ **TIPS\_10Y\_Yield** = **0.98** → randamentele obligațiunilor pe 5 și 10 ani merg mână în mână.



- **Open/High/Low/Close** ↔ **între ele = 0.99–1.00** → prețurile zilnice sunt practic identice ca tendință.
- **Close** ↔ **SP500\_Close = 0.93** → crypto urmărește foarte strâns bursa americană.
- **Close** ↔ **GOLD\_Close = 0.93** → crypto și aur cresc aproape simultan.

▼ **Top 5 corelații negative:**

- **Unemployment Rate** ↔ **GOLD\_Close = -0.52** → șomajul ridicat coincide cu scăderi la aur.
- **Unemployment Rate** ↔ **TIPS\_10Y\_Yield = -0.47** → șomaj mare = randamente reale scăzute.
- **Unemployment Rate** ↔ **TIPS\_5Y\_Yield = -0.45** → același efect, dar pe termen mai scurt.
- **Unemployment Rate** ↔ **Fed Funds Rate = -0.45** → șomaj ridicat forțează dobânzi mai mici.
- **DXY\_Close** ↔ **Close = -0.28** → un dolar mai puternic înseamnă prețuri crypto mai mici.

✦ **Concluzie generală pentru acest set de date:**

- Prețurile crypto sunt extrem de aliniate cu SP500, GOLD și inflația (CPI).
- Dificultatea și hash-rate-ul blockchain-ului sunt indicatori aproape perfect corelați cu prețurile.
- Dolarul puternic (DXY) și șomajul ridicat trag prețurile în jos.
- VIX (volatilitatea) nu pare să conteze aici.
- Aurul, surprinzător, merge mână în mână cu acțiunile și crypto în acest eșantion.

### 3. Evoluția în timp:

● **Bitcoin Close (linia albastră):**

- a pornit sub pragul de 1.000 USD în 2015, înregistrând o creștere exponențială în 2017, până la aprox. 20.000 USD, urmată de o corecție abruptă în 2018.
- 2020–2021 a atins din nou maxime istorice (peste 60.000 USD), susținut de creșterea interesului instituțional și politicile monetare expansive.
- 2024–2025, Bitcoin depășește pragul psihologic de 100.000 USD, consolidându-se ca activ de tip „store of value”, dar menținând o volatilitate ridicată.

● **Gold Close (linia galbenă):**

- Aurul a rămas un activ stabil, cu o creștere moderată de la circa 1.000 USD la peste 2.500 USD în 2025.
- În perioade de incertitudine (2020, 2022), aurul a înregistrat creșteri vizibile, confirmând rolul său tradițional de refugiu investițional („safe haven”).

■ **Oil Close (linia portocalie):**

- Petrolul a fost caracterizat de volatilitate accentuată, în special în 2020, când prețurile s-au prăbușit pe fondul crizei sanitare și scăderii cererii globale.
- După 2021, prețul a revenit treptat, stabilizându-se între 60–100 USD/baril, dar performanța pe termen lung rămâne sub cea a celorlalte active.

✅ **Indicele S&P500 (SP500\_Close – linie verde):**

- Indicele bursier american a înregistrat o creștere constantă de la aproximativ 2.000 puncte în 2015 la peste 6.000 în 2025.
- Corecțiile temporare (2020, 2022) au fost urmate de reveniri rapide, reflectând reziliența economiei SUA și expansiunea companiilor tehnologice.
- În ansamblu, analiza confirmă că, pe parcursul ultimului deceniu, Bitcoin s-a afirmat ca activ cu potențial investițional ridicat, dar cu un profil de risc semnificativ superior activelor tradiționale.
- Aurul și indicele S&P500 continuă să ofere stabilitate și predictibilitate pe termen lung, în timp ce petrolul rămâne sensibil la dinamica economică globală și la contextul energetic internațional.

Bitcoin vs Gold, Oil, S&P500 (axă secundară)



## 4. Sezonality sau trend:

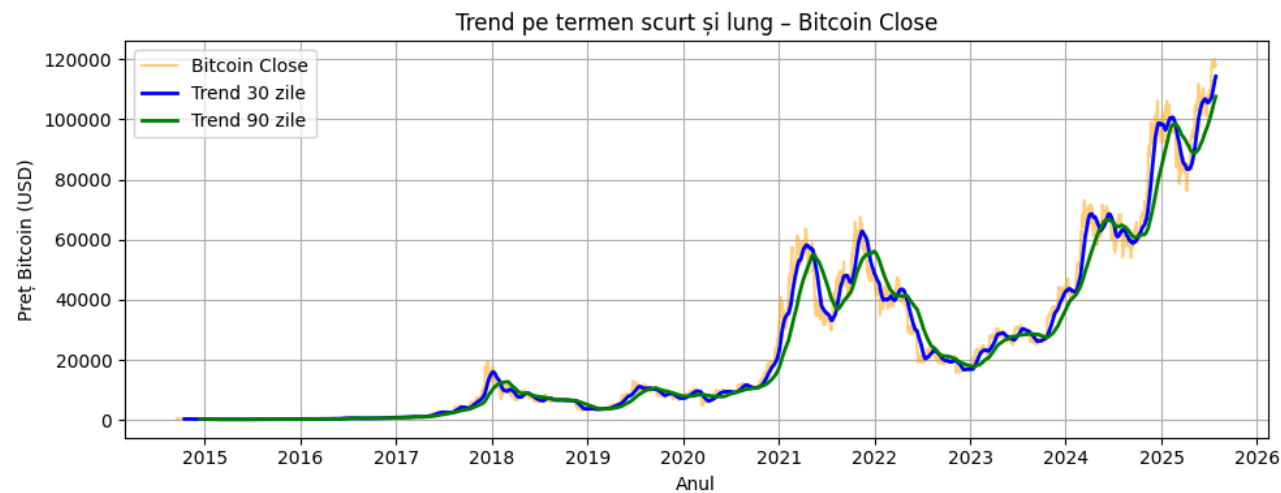
Număr zile în trend ascendent și descendent:

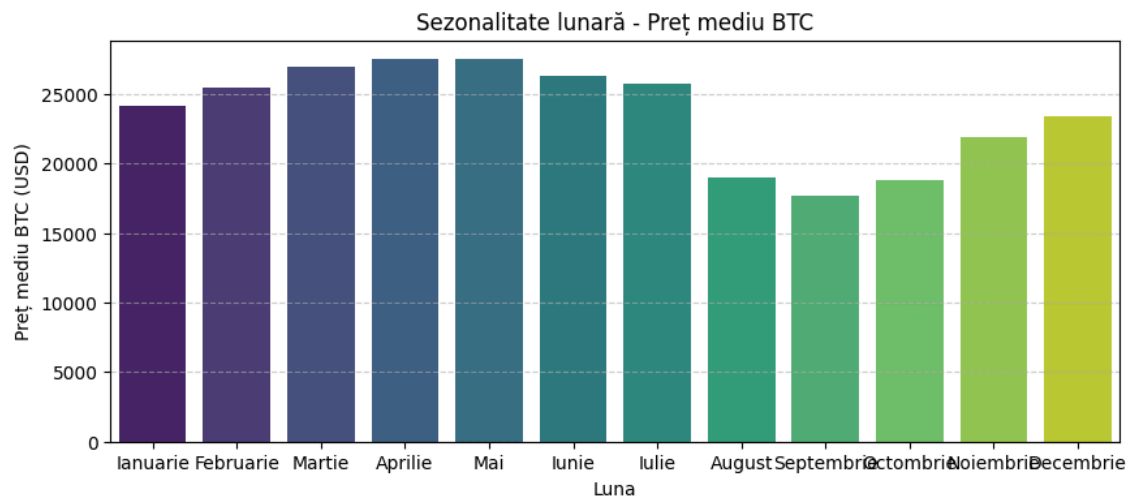
- *Uptrend*: 2559(zile)
- *Downtrend*: 1708(zile)

Sezonality lunară identificată:

- Luna cu preț mediu maxim: **Mai** - 27470.70 USD
- Luna cu preț mediu minim: **Septembrie** - 17635.85 USD

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).





## 5. Pattern-uri posibile în prețul Bitcoin:

### Inversări de trend în prețul Bitcoin:

- Trend ascendent (bullish): maxime și minime tot mai mari.
- Trend descendent (bearish): maxime și minime tot mai mici.
- Inversare de trend: momentul în care piața schimbă direcția (bullish → bearish sau invers).
- Pattern-uri tehnice semnificative: Double Top, Head & Shoulders, Zone de consolidare.
- Psihologia pieței: profit-taking după creșteri mari, panic-selling după scăderi, consolidări înainte de breakout.
- Cauze externe posibile: știri economice, reglementări guvernamentale, adopție instituțională, evenimente macroeconomice (ex: inflație, crize financiare), și sentimentul general al pieței crypto.

### Tabel cronologic cu evenimente externe majore și impact asupra BTC

Perioadă	Eveniment extern major	Impact asupra BTC / Trend
2017	Hype crypto și speculații masive	Creștere rapidă până la ~20.000 USD (bullish)
Dec 2017	Profit-taking și reglementări mai stricte în Coreea și China	Scădere rapidă sub 4.000 USD (bearish)
2020	Adoptarea instituțională (MicroStrategy, Tesla)	Creștere accelerată BTC (bullish)
2021	China interzice miningul și tranzacțiile crypto	Corecție severă ~30% (bearish)
2021–2022	Creșterea ratelor dobânzii și panică pe piețele de risc	Scădere până la ~20.000 USD (bearish)
2024	Inovații tehnologice și adopție largă	Consolidare 60.000–70.000 USD (neutral/bullish)
2025	Creștere infrastructură custodie și exchange-uri	Rupere rezistență și urcare până la ~120.000 USD (bullish)

### 2017–2018:

- BTC ajunge la ~20.000 USD → se formează un Double Top (roz) și un Head & Shoulders (verde/roșu).
- Confirmarea pattern-ului → scădere rapidă sub 4.000 USD → inversare bullish → bearish.
- Motive posibile:** hype-ul inițial și speculațiile masive, urmate de reglementări mai stricte și panic-selling.

### ✿ 2021–2022:

- BTC atinge ~64.000 USD și apoi ~69.000 USD → Head & Shoulders clar.
- După spargerea suportului → scădere până la ~20.000 USD → exemplu clasic de inversare de trend bearish.
- **Motive posibile:** adoptarea instituțională, modificări de reglementare (China interzice miningul), creșterea ratelor dobânzii și retragerea investitorilor din activele riscante.

### ✿ 2024–2025:

- Zone de consolidare între 60.000–70.000 USD (triunghiuri gri) → prețul rupe rezistența.
- Creștere până la ~120.000 USD → inversare bullish confirmată.
- **Motive posibile:** adoptarea largă a BTC, inovații tehnologice, creștere în infrastructura de custodie și exchange-uri, și sentiment pozitiv în rândul investitorilor.

### 🗉 Observații practice:

- Pattern-urile tehnice din grafic sunt foarte vizibile și pot fi folosite pentru planificarea intrărilor și ieșirilor.
- Zona de consolidare poate fi folosită ca punct de referință pentru identificarea breakout-ului.
- Confirmarea prin spargerea nivelurilor de suport/rezistență reduce riscul de semnale false.
- Analiza combinată a pattern-urilor + nivelurilor cheie + cauze externe oferă cea mai clară perspectivă asupra posibilei inversări de trend.

### ✅ Concluzie finală

- Graficul BTC arată că inversările de trend nu apar întâmplător; ele sunt semnalate clar de pattern-uri clasice și zone de consolidare.
- Investitorii pot combina teoria (Double Top, Head & Shoulders, suport/rezistență) cu observația concretă din grafic și analiza cauzelor externe pentru decizii mai informate.

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

=== Rezumat pattern-uri tehnice detectate ===

Double Tops: 25 (confirmate: 0)

Head & Shoulders: 79

Triunghiuri de consolidare: 144

Triunghi #1: 2015-05-09 - 2015-05-29

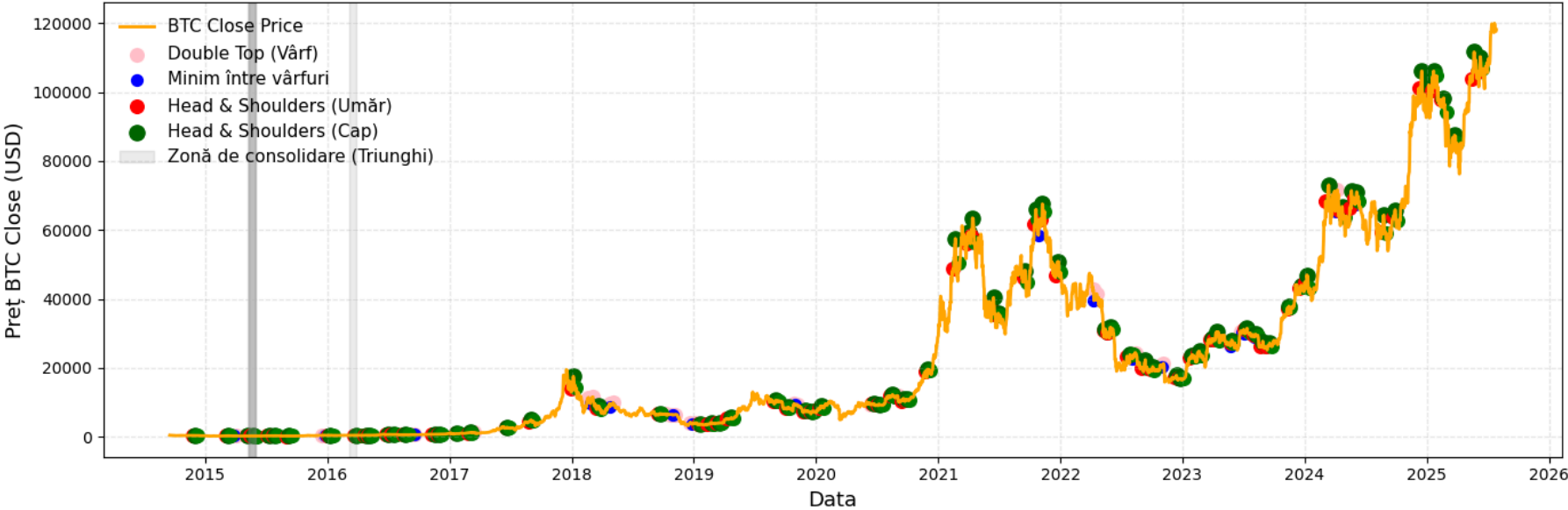
Triunghi #2: 2015-05-10 - 2015-05-30

Triunghi #3: 2015-05-11 - 2015-05-31

Triunghi #4: 2015-05-12 - 2015-06-01

Triunghi #5: 2016-03-06 - 2016-03-26

Pattern-uri tehnice detectate în prețul Bitcoin



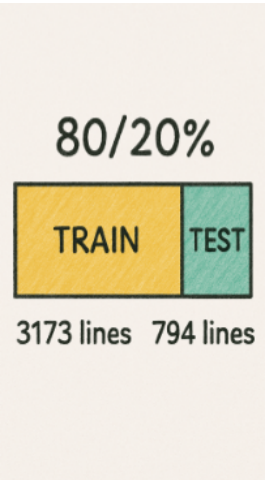
### ⚙️ Etapa III: Feature Engineering

#### 🔧 Variabile construite

Tip caracteristică	Descriere	Scop
Lag features	Valori anterioare (1–3 zile)	Captarea dependenței temporale, identifică pattern-uri și trenduri bazate pe istoricul prețului.
Moving averages	MA7, MA14, MA30	Evidențiază trenduri pe termen scurt și lung, reducând zgomotul din datele zilnice.
Volatilitate	Std dev pe ferestre mobile	Măsoară riscul și instabilitatea prețului, util pentru strategie de gestionare a riscului.
Variabile derivate	BTC/SP500, BTC/Gold	Analizează corelațiile și performanța relativă a Bitcoin față de piețele tradiționale.
Indicatori tehnici	RSI, MACD, Bollinger Bands	Surprind supracumpărarea/supravânzarea, momentum-ul și volatilitatea.

#### 🎯 Împărțirea datelor

- 80% antrenare
- 20% testare
- Datele au fost scalate numeric și aliniate temporal.

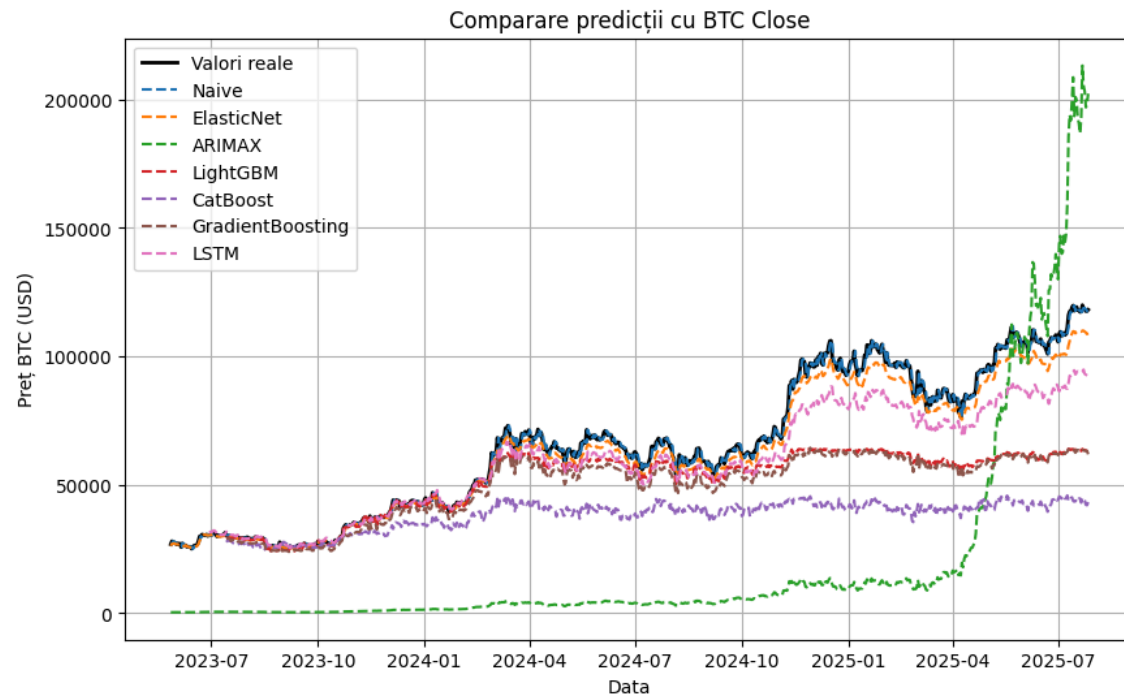


🗨 Etapa IV: Selectarea și antrenarea modelelor

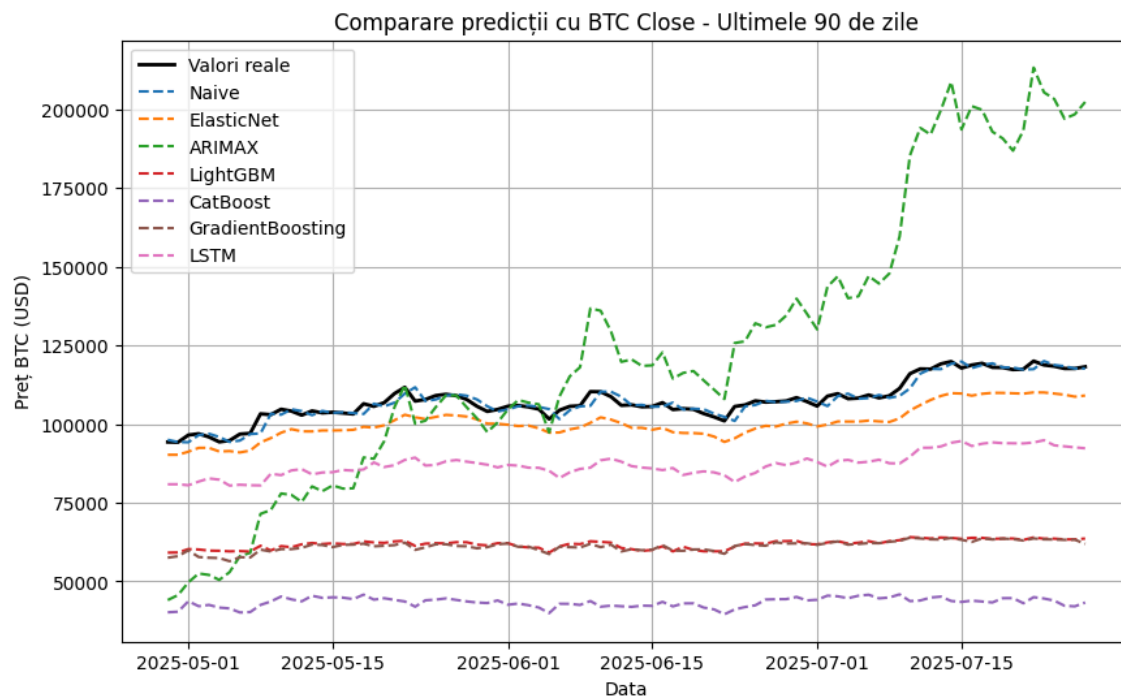
Nr.	Model / Componentă	Descriere & Insight	Parametri / Detalii Cheie	Avantaje	Limitări / Atenționări
1	Naive Baseline	Predicție simplă: <code>valoarea de ieri = valoarea de azi</code> . Benchmark de bază, surprinde stabilitatea dar nu volatilitatea.	-	Foarte rapid, util ca referință.	Nu captează trenduri sau șocuri.
2	Elastic Net Regression	Model liniar cu penalizare L1 + L2. Identifică indicatorii care influențează direct prețul BTC.	<code>alpha</code> ∈ [0.0001–1], <code>l1_ratio</code> ∈ [0–1], scaler: <code>StandardScaler</code>	Interpretabil, bun pentru feature selection.	Limitat la relații liniare; sensibil la scaling.
3	ARIMAX pe randamente logaritmice	Modelează randamentele ( <code>log-returns</code> ) și folosește variabile exogene. Captură autocorelații și efecte exogene.	AIC/BIC pentru (p,d,q); teste ADF; exogene lagged.	Modelează dependențe temporale clare.	Sensibil la staționaritate; tuning dificil.
4	LightGBM Regressor	Boosting pe arbori, captează relații non-liniare complexe și interacțiuni între indicatori.	<code>num_leaves</code> (31–512), <code>learning_rate</code> (0.01–0.1), <code>max_depth</code> , <code>n_estimators</code>	Rapid, precis, ușor de tunat.	Poate supraînvăța; necesită early stopping.
5	CatBoost Regressor	Boosting optimizat pentru date secvențiale și categorice; predicții robuste pe termen mediu.	<code>iterations</code> (500–2000), <code>learning_rate</code> (0.01–0.1), <code>depth</code> (4–10)	Robust la overfitting; interpretabile SHAP.	Timp de antrenare mai mare.
6	Gradient Boosting Regressor	Boosting iterativ care corectează erorile recente; optimizează MAE.	<code>learning_rate</code> , <code>n_estimators</code> , <code>max_depth</code> , <code>subsample</code>	Echilibru între precizie și stabilitate.	Necesită tuning fin; lent pe seturi mari.
7	LSTM (Long Short-Term Memory)	Rețea neuronală recurentă pentru secvențe temporale. Surprinde dependențe pe termen lung.	1–3 straturi, 32–256 units, <code>dropout</code> , <code>batch_size</code> (32–128), fereastră 30–90 zile.	Captură pattern-uri complexe, laguri multiple.	Necesită multe date și resurse; risc mare de overfitting.

💡 Observații:

- Pe întreaga perioadă de testare, modelele de tip **boosting (LightGBM, CatBoost, GradientBoosting)** și **LSTM** oferă predicții mai conservatoare, **subestimând creșterile puternice din 2024 - 2025**.
- ElasticNet** și **Naive** urmează cel mai bine tendința generală, menținând o eroare relativ mică.
- ARIMAX** deviază semnificativ spre finalul perioadei, **amplificând erori** în momentele de volatilitate accentuată.
- În **ultimele 90 de zile**, **Naive** și **ElasticNet** rămân cele mai apropiate de valorile reale ale BTC, captând **corect trendul recent**.
- Modelele de **boosting** și **LSTM** oferă predicții stabile, dar rămân sub nivelul real, pierzând din sensibilitatea la fluctuațiile scurte.
- ARIMAX** prezintă variații mari și **pierderi de precizie** în această perioadă.







## ⚖️ Etapa V: Evaluare, comparare și selecția modelului final

### 🔍 Rezultate comparate

- Cel mai bun model simplu pe acest set de date: Naive baseline. Eficient pe termen scurt și surprinde bine trendul mediu.
- Modelele complexe (LSTM, boosting, ARIMAX) nu performează mai bine decât Naive fără optimizare suplimentară sau feature engineering avansat.

### 🤔 Idei de îmbunătățire:

- Combinarea modelelor (stacking) Naive, ElasticNet și LSTM, pentru a obține stabilitate și reducerea erorilor extreme.
- Prin combinarea lor, stacking-ul reușește să reducă drastic erorile comparativ cu modelele individuale:
  - a) MAE: 877 vs 1,143 (Naive) și 3,147 (ElasticNet) și 7,838 (LSTM) → reducere semnificativă.
  - b) RMSE: 1,226 vs 1,725 (Naive) și 4,080 (ElasticNet) și 10,575 (LSTM).
  - c) Max Error: 5,008 vs 8,227 (Naive) → chiar extremele sunt mai bine estimate.
  - d) Std Dev Error: 857 vs 1,292 (Naive) și 2,596 (ElasticNet) → stabilitate mai mare.

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

	Model	MAE	RMSE	Max Error	Min Error	Std Dev Error
0	Naive	1143.801148	1725.853652	8227.290000	0.000000	1292.396906
1	ElasticNet	3147.251588	4080.128162	11668.627542	10.740664	2596.584924
2	ARIMAX	51672.035118	56019.200147	93229.234344	59.408317	21636.810577
3	LightGBM	15357.057006	22533.540530	56026.466889	0.755077	16490.034837
4	CatBoost	29150.290052	36366.204330	76067.315566	49.111022	21743.077226
5	GradientBoosting	17339.135939	23519.670069	56338.601277	656.881002	15891.168774
6	LSTM	7838.591162	10575.925005	26385.680000	1.272000	7099.766075

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

	Model	MAE	RMSE	Max Error	Min Error	Std Dev Error
0	Stacking (Naive+ElasticNet+LSTM)	877.421938	1226.636123	5008.298605	1.678317	857.185465

## Analiză comparativă Predicții vs. Realitate în următoarele 7 zile:

### Predicția Naive:

- Model simplu, extrapolează valorile viitoare pe baza trendului recent.
- Eroarea (Error\_Naive) variază **între ~89 USD și 2,990 USD**, surprinde trendul general, dar nu reacționează suficient la schimbări bruște de preț.
- Primele 3 zile (28-30 iulie) au eroare procentuală mică (<1%), dar crește ușor ulterior (1.45% - 2.64%), modelul deviază în perioadele de volatilitate mai mare.

### Stacking (Naive + ElasticNet + LSTM):

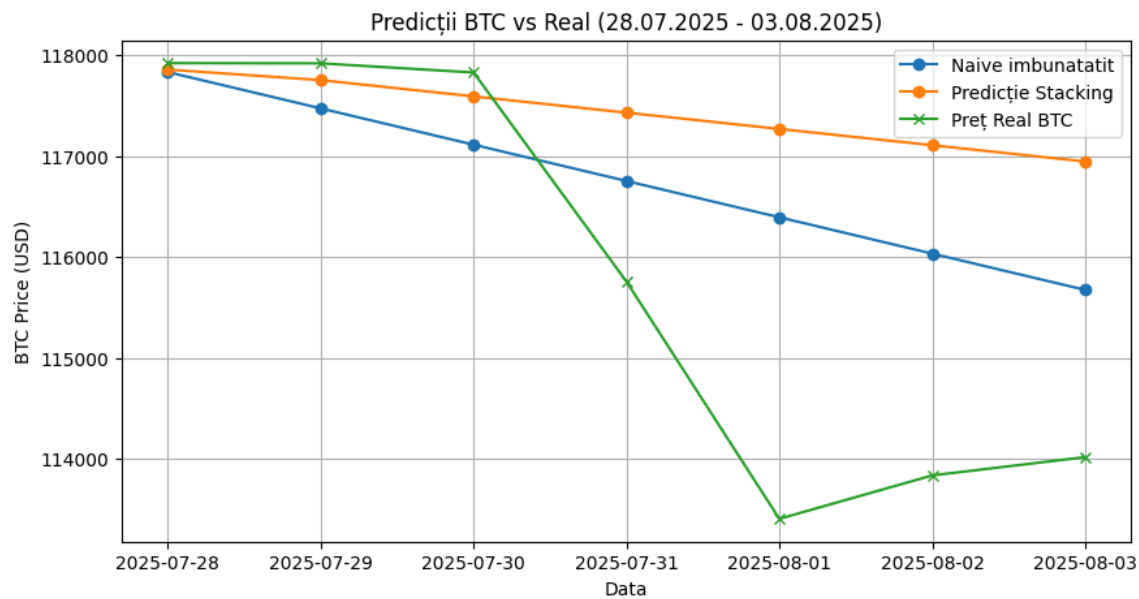
- Eroarea absolută (Error\_Stacking) este mai mică pentru toate zilele, **între 64 USD și 3,866 USD**.
- Eroarea procentuală este foarte mică, între 0.05% și 3.41%, arătând predicții **foarte apropiate de valorile reale**.
- Chiar și în perioadele cu volatilitate ridicată (1-3 august), Stacking-ul limitează deviațiile

### În final:

- Stacking-ul este un model solid, echilibrat și scalabil, capabil să ofere predicții precise într-un context volatil precum piața Bitcoin.
- Cu ajustări suplimentare de tip adaptiv și integrarea unor surse informaționale externe, poate deveni un instrument practic și de încredere în analiza și prognoza financiară.

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

MAE următoarea săptămână: 1744.82  
RMSE următoarea săptămână: 2303.81



	Date	Pred_Naive	Pred_Stacking	BTC_Real	Error_Stacking	Error_Naive	Error_%_Stacking	Error_%_Naive
0	2025-07-28	117835.37	117860.42	117924.47	64.05	89.10	0.05	0.08
1	2025-07-29	117475.37	117755.71	117922.15	166.44	446.78	0.14	0.38
2	2025-07-30	117115.36	117594.29	117831.19	236.90	715.83	0.20	0.61
3	2025-07-31	116755.36	117432.87	115758.20	1674.67	997.16	1.45	0.86
4	2025-08-01	116395.35	117271.45	113405.34	3866.11	2990.01	3.41	2.64
5	2025-08-02	116035.34	117110.03	113836.00	3274.03	2199.34	2.88	1.93
6	2025-08-03	115675.34	116948.61	114017.03	2931.58	1658.31	2.57	1.45

## 🌟 Concluzii și direcții viitoare

### ✅ Rezumatul rezultatelor

- Modelul **Stacking** combină avantajele mai multor algoritmi individuali, reușind să producă o predicție agregată stabilă și cu **eroare medie absolută de aproximativ 1.500 USD**, respectiv o **eroare procentuală medie de 1,5%** raportat la prețul real al BTC (~115.000–118.000 USD).
- Această performanță arată că modelul este **suficient de precis pentru predicții operaționale pe termen scurt**, cum ar fi monitorizarea tendinței zilnice sau anticiparea ușoarelor corecții de piață.
- Stacking-ul a demonstrat o **bună rezistență la volatilitate**, menținând deviații moderate chiar și în intervalele de scădere sau creștere bruscă (ex. 31 iulie–2 august 2025).
- Prin combinarea mai multor surse de informație (ex. modele de tip regresie și boosting), el reușește să echilibreze supraestimările și subestimările, reducând varianța totală a erorii.
- Totuși, modelul prezintă o **ușoară tendință de netezire a predicțiilor**, pierzând precizie în fața mișcărilor extreme ale pieței. Aceasta indică faptul că metamodelul (modelul final care combină predicțiile) poate fi optimizat în direcția unei ponderări dinamice a intrărilor.

## Direcții de îmbunătățire

- **Reantrenare adaptivă (rolling window)** – recalibrarea modelului la fiecare 30 de zile pentru a capta noile regimuri de piață.
  - **Includerea variabilelor exogene:** volatilitate istorică, date de sentiment, activitate pe rețele sociale, știri.
  - **Optimizarea metamodelului de stacking** – testarea unor meta-learneri diferiți (ex. Ridge, XGBoost, CatBoost) și ajustarea ponderilor pe baza performanței recente.
  - **Predicție pe randamente logaritmice** în loc de prețuri brute pentru reducerea dependenței de scara numerică și pentru o stabilitate mai bună.
  - **Validare pe ferestre mobile (walk-forward validation)** pentru estimarea realistă a performanței în condiții de piață variabile.
  - **Implementarea unei măsuri de încredere** (confidence interval / prediction interval) pentru a cuantifica incertitudinea fiecărei predicții.
- 

## Concluzie finală

Proiectul demonstrează că, prin integrarea mai multor surse de date și a modelelor combinate, se pot obține **predicții robuste și precise** pentru o piață volatilă precum cea a criptomonedelor.

---

## Mulțumiri

Vă mulțumesc pentru atenție!

Feedback-ul și întrebările sunt binevenite 😊

 *Cristiana Kovacs*