

**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**  
Școala doctorală Informatică Economică



**TEZĂ DE DOCTORAT**

Prezentată și susținută public de către autor:  
**MIHAI-ȘERBAN V. AVRAMESCU**

Titlul tezei de doctorat:  
**ANALIZA ȘI PREDICȚIA DATELOR ECONOMICE  
UTILIZÂND TEHNICI DL (DEEP LEARNING)**

Conducător de doctorat:  
Prof. univ. dr. Cătălina-Lucia Cocianu

Comisia de susținere a tezei de doctorat:

Prof. univ. dr. <u>Claudiu Herțeliu</u> (președinte)	- Academia de Studii Economice din București
Prof. univ. dr. <u>Mihaela Muntean</u> (referent)	- Universitatea de Vest din Timișoara
Prof. univ. dr. <u>Gabriela Mircea</u> (referent)	- Universitatea de Vest din Timișoara
Prof. univ. dr. <u>Paul Pocatilu</u> (referent)	- Academia de Studii Economice din București
Prof. univ. dr. <u>Cătălina-Lucia Cocianu</u> (conducător de doctorat)	- Academia de Studii Economice din București

București, 2024

**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**

**Consiliul pentru Studii Universitare de Doctorat**

*Școala Doctorală*

**Informatică Economică**

**ANALIZA ȘI PREDICȚIA DATELOR ECONOMICE  
UTILIZÂND TEHNICI DL (DEEP LEARNING)**

Mihai-Șerban AVRAMESCU

Conducător de doctorat: Prof. univ. dr. Cătălina-Lucia Cocianu

**București, 2024**

# CUPRINS

<b>REZUMAT</b> .....	I
<b>MULȚUMIRI</b> .....	II
<b>PARTEA I: STADIUL ACTUAL AL CUNOAȘTERII ÎN DOMENIU</b> .....	1
<b>CAPITOLUL 1: INTRODUCERE</b> .....	1
1.1 INTRODUCERE.....	1
1.2 STRUCTURA LUCRĂRII.....	4
1.3 PRINCIPALELE CONTRIBUȚII .....	5
<b>CAPITOLUL 2: PREPROCESAREA DATELOR</b> .....	8
2.1 PREPROCESARE.....	8
2.2 SELECȚIA VARIABILELOR .....	9
2.3 EXTRACTORII DE CARACTERISTICI PCA ȘI ICA .....	16
2.4 FILTRARE.....	20
2.5 PARTICULARITĂȚI ALE ANALIZEI SERIILOR DE TIMP .....	24
<b>CAPITOLUL 3: METODE DE PREDICȚIE</b> .....	31
3.1 SERII DE TIMP .....	31
3.2 MODELELE MATEMATICE NAR ȘI NARX .....	35
3.3 METODE STATISTICE DE PREDICȚIE (AR, ARMA, ARIMA, GARCH) .....	38
3.4 PREDICȚIA DATELOR UTILIZÂND TEHNICI ML ȘI DL. CONSIDERAȚII GENERALE .....	43
3.5 EVALUAREA CALITĂȚII UNUI MODEL (MSE, MAPE, RMSE, POCID, F1- SCORE).....	47
3.6 OPTIMIZĂRI EVOLUTIVE .....	49
<b>PARTEA A II-A : CONTRIBUȚII PERSONALE: METODOLOGIE DE CERCETARE ȘI REZULTATE EXPERIMENTALE</b> .....	53
<b>CAPITOLUL 4: DEZVOLTAREA UNUI SISTEM DE PREDICȚIE HIBRID BAZAT PE     MODELUL NARX ȘI TEHNICA SVM PENTRU PREDICȚIA INDICILOR BURSIERI</b> ...	53
4.1 INTRODUCERE.....	53
4.2 TEHNICA DE INSTRUIRE SVM.....	53
4.3 ANALIZA SPECTRALĂ .....	56
4.4 MODELUL MATEMATIC DE PREDICȚIE ȘI METODOLOGIA PROPUȘĂ .....	59
4.5 REZULTATE EXPERIMENTALE .....	60

CAPITOLUL 5: NOI ABORDĂRI NEURONALE ALE MODELULUI DE PROGNOZĂ BAZAT PE NARX. ANALIZA ȘI PREDICȚIA CURSURILOR DE SCHIMB.....	65
5.1 REȚELE NEURONALE NARX .....	65
5.2 UTILIZAREA REȚELELOR NEURONALE LSTM PENTRU IMPLEMENTAREA MODELULUI NARX .....	75
5.3 REȚELE CONVOLUTIVE .....	82
CAPITOLUL 6: METODE EVOLUTIVE PENTRU SETAREA PARAMETRILOR LSTM PENTRU PREDICȚIA SERIILOR TIMP. ANALIZA EVOLUȚIEI PRINCIPALELOR CRIPTOMONEDE .....	98
6.1 VERSIUNI ÎMBUNĂȚĂȚITE ALE ALGORITMULUI FIREFLY .....	98
6.2 METODA DE PREDICȚIE LSTM CU ALGORITMI EA.....	100
6.3 METODA ES-LSTM.....	102
6.4 EXPERIMENTE .....	104
CAPITOLUL 7: CONCLUZII SI DIRECȚII VIITOARE .....	121
7.1 REZUMAT ȘI CONTRIBUȚII.....	121
7.2 APLICAȚII ÎN CARE POT FI UTILIZATE CONCEPTELE PREZENTATE.....	124
7.3 DIRECȚII VIITOARE .....	126
<b>BIBLIOGRAFIE .....</b>	<b>128</b>
<b>LISTA DE ABREVIERI .....</b>	<b>143</b>
<b>LISTA DE TABELE .....</b>	<b>144</b>
<b>LISTA DE FIGURI .....</b>	<b>146</b>

# REZUMAT

Scopul principal al lucrării de doctorat este dezvoltarea unor metodologii bazate pe tehnici de învățare profundă (DL - Deep Learning) pentru analiza și predicția datelor economice. Datele care vor fi analizate sunt de forma seriilor de timp.

Primul tip de date este reprezentat de indicii bursieri. Aceștia sunt instrumente esențiale în evaluarea performanței unei piețe bursiere sau a unui segment specific al acesteia. Indicii bursieri sunt importanți deoarece oferă investitorilor un punct de referință pentru performanța generală a pieței și permit evaluarea comparativă a performanței fondurilor de investiții și a altor portofolii.

A doua categorie de date studiate o reprezintă cursurile de schimb valutar. Acestea sunt rata la care o monedă poate fi schimbată pentru o altă monedă, fiind utilizați peste tot în lume, atât la nivelul economiei generale, cât și la nivel de individ. Este, astfel, un indicator raportat la nivel global. În mod natural, ratele valutare influențează prețurile bunurilor și serviciilor, afectând astfel comerțul internațional, fluxurile de capital și investițiile.

O categorie de date financiare relativ nouă este reprezentată de criptomonede. Acestea sunt monede digitale (virtuale) care utilizează criptografia pentru securitate. Criptomonedele sunt extrem de volatile, cu prețuri care pot fluctua dramatic într-un interval scurt de timp, ceea ce poate fi atribuit speculațiilor, reglementărilor variabile și percepțiilor pieței. Această volatilitate ridicată poate oferi oportunități de câștiguri rapide, dar și riscuri semnificative pentru investitori.

## STRUCTURA LUCRĂRII

Lucrarea este structurată în două părți principale. Acestea sunt reprezentate de trecerea în revistă literaturii teoretice în domeniu și, respectiv, metodele practice dezvoltate. Organizarea este în capitole, după cum urmează.

### Partea I. STADIUL ACTUAL AL CUNOAȘTERII ÎN DOMENIU

Prima parte a tezei cuprinde, pe lângă introducere, două capitole, în care sunt prezentate studii asupra principalelor direcții de cercetare și metodologiilor clasice, pe baza literaturii de specialitate relevante pentru domeniul abordat.

Capitolul 2: PREPROCESARE. În acest capitol sunt prezentate elemente teoretice despre metode de selecție a variabilelor, extragerea de caracteristici, tehnici de filtrare a datelor și indicatori tehnici specifici analizei pe serii de timp. Sunt amintiți principalii extractori de caracteristici: analiza în componente principale și analiza în componente independente. De asemenea, este propusă o variantă a mediei mobile ponderate, indicator utilizat în preprocesarea

seriilor de timp pentru studiul practic prezentat în capitolele următoare. Această variantă de indicator este bazată pe distribuția de probabilitate rang exponențial.

Capitolul 3: METODE DE PREDICȚIE. Capitolul al treilea are rolul de a prezenta modelele matematice NAR și NARX, metodele statistice de predicție, precum și tehnicile recente de ML și DL. De asemenea sunt prezentate principalele metrice pentru evaluarea calității modelelor propuse. În ultima parte a capitolului sunt introduse concepte despre optimizările evolutive care se pot atașa modelelor clasice de predicție, precum algoritmi genetici, algoritmi de tip inteligența roiului și strategiile evolutive.

#### Partea a II-a. CONTRIBUȚII PERSONALE, METODOLOGIE DE CERCETARE ȘI REZULTATE EXPERIMENTALE

A doua parte a lucrării cuprinde trei capitole și o secțiune de concluzii. Aceasta este dedicată contribuțiilor personale și conține modelele și algoritmi propuși, evidențiind inovațiile aduse în domeniul studiat.

Capitolul 4: DEZVOLTAREA UNUI SISTEM DE PREDICȚIE HIBRID BAZAT PE MODELUL NARX ȘI TEHNICA SVM PENTRU PREDICȚIA INDICILOR BURSIERI. În acest capitol este propusă o metodologie pentru selectarea indicatorilor tehnici și predicția indicilor bursieri care utilizează un model hibrid în mai multe etape.

Capitolul 5: NOI ABORDĂRI NEURONALE ALE MODELULUI DE PROGNOZĂ BAZAT PE NARX. ANALIZA ȘI PREDICȚIA CURSURILOR DE SCHIMB. În cadrul capitolului sunt propuse trei metodologii care utilizează metoda NARX, o metodologie care utilizează LSTM în implementarea modelului de predicție NARX și o metodă care utilizează rețele neuronale convolutive (CNN) în implementarea modelului de predicție NARX.

Capitolul 6: METODE EVOLUTIVE PENTRU SETAREA PARAMETRILOR LSTM PENTRU PREDICȚIA SERIILOR TIMP. ANALIZA EVOLUȚIEI PRINCIPALELOR CRIPTOMONEDE. Este propusă o tehnică evolutivă inovatoare de învățare profundă pentru analiza și prognoza seriilor temporale financiare.

Capitolul 7: Capitolul final rezumă întreg fluxul lucrării, al modelelor utilizate și al rezultatelor obținute, descriind câteva posibile direcții pentru dezvoltarea ulterioară a domeniului.

## PRINCIPALELE CONTRIBUȚII

Principalele contribuții originale propuse în această lucrare sunt următoarele:

- O metodologie pentru selectarea indicatorilor tehnici și predicția indicilor bursieri care utilizează un model hibrid în patru etape. Prima etapă constă în calcularea și selectarea indicatorilor tehnici pe baza datelor istorice ale prețurilor de pe piața bursieră. În etapa următoare este calculată variabila întârziere. Cea de-a treia etapă presupune utilizarea instrumentelor de analiză spectrală pentru a elimina zgomotul conținut în date. Au fost utilizate pentru eliminarea zgomotului reprezentarea wavelet și filtrul trece-jos gaussian, derivat pe baza transformării Fourier. Etapa de predicție propriu-zisă utilizează tehnica SVM neliniară.
  - Metodologii de predicție dezvoltate pe baza rețelelor neuronale NARX. Scopul acestor abordări este predicția cursului de schimb valutar USD-RON pe baza cursului CHF-RON.
    - Prima metodologie utilizează metoda NARX standard la care adăugăm un pas suplimentar de testare pentru validarea rețelei.
    - Cea de-a doua abordare propusă este o variantă îmbunătățită a primei, prin includerea unei etape suplimentare de optimizare. În cadrul acesteia este menținută configurația rețelei NARX pentru cazurile în care ea returnează rezultate satisfăcătoare în etapa de instruire dar rezultate slabe în etapa de testare și este re-antrenată pentru îmbunătățirea capacităților de generalizare.
    - A treia metodologie vine ca o completare pentru primele două și aduce nou o funcționalitate care permite modificări ale parametrilor pentru configurațiile deja conservate ale rețelei NARX.
- Următoarele metode dezvoltate au avut la bază trei obiective principale: revizuirea rețelelor de tip LSTM, prin aplicarea acestora pe modelul de predicție NAR și NARX pentru studiul rețelelor de tip convolutiv și analiza performanțelor abordărilor LSTM și CNN. Scopul principal a fost predicția cursului de schimb valutar USD-RON. Abordările propuse includ:
  - O metodă care utilizează LSTM în implementarea modelului de predicție NARX, alternativă viabilă și mai performantă comparativ cu rețelele neuronale standard NARX.

- O metodă care utilizează rețele neuronale convolutive (CNN) în implementarea modelului de predicție NARX, de asemenea o alternativă mai exactă de predicție comparativ cu rețelele neuronale NARX.

Concluzia studiului este că se obțin rezultate mai bune și mai stabile decât LSTM pe partea de metrice cantitative. Pentru măsura calitativă POCID, care reflectă tendința evoluției, rezultatele sunt similare dar mai stabile pentru LSTM. Scopul principal a fost predicția cursului de schimb valutar USD-RON.

- O variantă îmbunătățită a algoritmului FA (Firefly Algorithm), prin definirea unor noi reguli de actualizare și tratare a situațiilor de părăsire a spațiului soluțiilor.

- Un algoritm pe baza căruia este realizată antrenarea de rețele neuronale profunde (Deep Learning) de tip LSTM care sunt configurate utilizând metode specifice strategiilor evolutive. Abordarea propusă îmbină o procedură de căutare locală ES cu doi membri cu un optimizator ADAM pentru antrenarea rețelelor neuronale LSTM în vederea implementării modelului NARX. Obiectivul principal este de a îmbunătăți predicțiile în funcție de scorul F1, fără a crește semnificativ valoarea metricii de eroare MAPE. Scopul principal ale acestor metode este predicția cursurilor de schimb ale principalelor criptomonede (BTC și ETH) și USD și a cursului de schimb USD-EURO. Studiile realizate au avut ca punct de plecare volatilitatea semnificativă a criptomonedelor comparativ cu principalele valute și necesitatea obținerii unor variante de regresori cu putere de generalizare superioară celor clasici. Principalul indice de acuratețe urmărit a fost cel al tendințelor de evoluție (predicția corectă a unei creșteri/descreșteri de preț).

## **APLICAȚII POTENȚIALE**

Există două direcții în care temele tratate în cadrul acestei teze au aplicabilitate pentru sfera financiară reală. Prima se referă la analiza datelor economice, ale cărei principale componente sunt amintite în continuare.

Înțelegerea comportamentului piețelor: analiza datelor istorice permite identificarea și cuantificarea riscurilor asociate investițiilor. Volatilitatea trecutului, de exemplu, este o măsură esențială derivată din mișcările anterioare ale prețurilor și este folosită pentru a evalua riscul unui activ.

Identificarea și limitarea riscurilor: comercianții folosesc mișcările anterioare ale prețurilor pentru a crea modele și algoritmi care exploatează ineficiențele pieței. Testarea retroactivă a



acestor strategii pe baza datelor istorice ajută la evaluarea viabilității și rentabilității acestora înainte de aplicarea lor în tranzacțiile reale.

**Analiza reglementărilor și a conformității:** Pentru organismele de reglementare și instituțiile financiare analiza datelor istorice este utilizată pentru asigurarea conformității cu reglementările și standardele financiare. Aceste date oferă o înregistrare detaliată a tranzacțiilor și activităților de pe piață, care pot fi auditate și revizuite pentru a detecta eventualele nereguli sau activități frauduloase.

**Îmbunătățirea educației financiare:** datele istorice sunt un instrument aplicat și în scopuri educaționale. Studenții și profesioniștii în finanțe și economie utilizează datele istorice de pe piață pentru a învăța despre mecanismele pieței, evaluarea activelor, gestionarea riscurilor și impactul evenimentelor economice.

**Evaluarea performanței portofoliilor:** investitorii și administratorii de fonduri folosesc date istorice pentru a compara performanța investițiilor sau portofoliilor lor cu indicii de piață și alte repere. Această comparație ajută la evaluarea eficienței strategiilor de investiții și la efectuarea ajustărilor necesare. Analiza performanțelor anterioare oferă, de asemenea, o bază pentru stabilirea obiectivelor și așteptărilor legate de performanțele viitoare

Cea de-a doua direcție principală se referă la predicția datelor. Principalele scenarii reale în care aceste informații pot fi utilizate sunt descrise mai jos:

**Gestionarea portofoliilor de acțiuni:** investitorii folosesc previziunile pentru a lua decizii informate privind alocarea activelor, diversificarea și re-echilibrarea portofoliilor. Previziunile precise pot ajuta la maximizarea randamentelor și la minimizarea riscurilor prin identificarea acțiunilor, monedelor sau criptomonedelor care ar trebui cumpărate sau vândute.

**Politica monetară:** băncile centrale utilizează cursurile de schimb și previziunile economice pentru a decide asupra politicilor monetare (de exemplu stabilirea ratelor dobânzilor). Previzionarea tendințelor inflaționiste și a creșterii economice ajută la decizii politice pentru stabilizarea economiei.

**Planificarea strategică:** corporațiile folosesc prognozele pieței pentru a lua decizii strategice pe termen lung, cum ar fi intrarea pe noi piețe, investiția în noi proiecte sau planificarea emisiunii de obligațiuni. Previziunile precise ale condițiilor economice au un impact pozitiv asupra profiturilor viitoare.

**Gestionarea lanțului de aprovizionare:** întreprinderile se folosesc de previziuni pentru a anticipa schimbările în ratele de schimb, care pot influența costul bunurilor importate și al materiilor prime. Aceasta facilitează optimizarea strategiilor de achiziții și gestionarea relațiilor cu furnizorii.

## **DIRECȚII VIITOARE**

O primă idee constă în integrarea și mai mult a modelelor hibride care combină elemente ale diferitelor metodologii și care ar putea duce către îmbunătățiri ale performanței de predicție. De exemplu, îmbinarea modelelor tradiționale de învățare automată cu arhitecturi de învățare profundă sau algoritmi evolutivi ar putea spori capacitatea de a surprinde atât modelele liniare, cât și cele neliniare din datele financiare.

O altă direcție potențială de luat în calcul pentru cercetările ulterioare este utilizarea modelelor propuse într-un sistem în care fluxurile de date să fie primite în timp real, practic să existe o învățare în timp real sau cu o întârziere limitată. Piețele financiare reale sunt dinamice și în continuă evoluție, iar modelele care se pot adapta la noile date în timp real vor putea aduce un plus de valoare.

În plus, extinderea capacităților de aplicare a metodelor de selecție și procesare a caracteristicilor poate îmbunătăți și mai mult performanța modelului. Utilizarea tehnicilor avansate de procesare a limbajului natural (NLP) și un proces mai amplu de selecție a variabilelor dintr-un set mult mai mare de potențiale surse de date, fără a întârzia foarte mult procesul de predicție, ar putea reprezenta o optimizare a stadiului curent în domeniu.