

Simona-Vasilica OPREA
coordonator

Anca-Ioana ANDREESCU
Adela BÂRA

Gabriela DOBRIȚA
Andreea-Mihaela NICULAE
Alin-Gabriel VĂDUVA

**Dezvoltarea de aplicații
pentru prelucrarea și analiza datelor
utilizând algoritmi de învățare automată
în Python și SAS**

Colecția
Informatică

Editura ASE
București
2025



ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI

Copyright © 2025, Editura ASE

Toate drepturile asupra acestei ediții sunt rezervate editurii.

Editura ASE

Piața Romană nr. 6, sector 1, București, România

cod 010374

www.ase.ro

www.editura.ase.ro

editura@ase.ro

Descrierea CIP a Bibliotecii Naționale a României

**Dezvoltarea de aplicații pentru prelucrarea și analiza datelor
utilizând algoritmi de învățare automată în Python și SAS /**

Simona-Vasilica Oprea (coord.), Anca-Ioana Andreescu, Adela Bâra, –
București : Editura ASE, 2025

Conține bibliografie

ISBN 978-606-34-0593-8

I. Oprea, Simona-Vasilica

II. Andreescu, Anca Ioana

III. Bâra, Adela

004

Editura ASE

Redactor, tehnoredactor și copertă: Claudia-Marinela Dumitru

Autorii noștri își asumă întreaga responsabilitate pentru ideile exprimate,
pentru originalitatea materialului și pentru sursele bibliografice menționate.

Cuprins

Introducere	9
Partea I. Dezvoltarea de aplicații web și interactivitate în Python.....	13
1. Introducere în Streamlit	15
1.1 Introducere	15
1.1.1 Descrierea tehnologiilor similare	16
1.1.2 Când alegem Streamlit?.....	17
1.1.3 Alegerea tehnologiei în funcție de necesități	18
1.2 Funcții în Streamlit	18
1.2.1 Funcții pentru afișare în Streamlit	18
1.2.2 Funcții pentru afișarea datelor în Streamlit	20
1.2.3 Funcții pentru afișarea graficelor în Streamlit	20
1.2.4 Funcții pentru afișarea elementelor interactive (widget-uri) în Streamlit	22
1.2.5 Funcții pentru aranjarea interfeței și organizarea layout-ului în Streamlit	22
1.2.6 Funcții pentru interactivitate și stare	26
1.2.7 Funcții pentru fișiere media	29
1.2.8 Funcții pentru diverse opțiuni	30
1.3 Implementare (Deployment)	31
1.4 Aplicație cu mai multe pagini.....	36
1.6 Aplicație cu pagină de autentificare și altele	39
Întrebări recapitulative și exerciții	65
Referințe.....	66
Pagini utile.....	67
2. Preprocesarea datelor pentru machine learning în Streamlit	69
2.1 Importanța procesării datelor	69
2.2 Tratarea valorilor lipsă.....	72
2.2.1 Tratarea valorilor (numerice) lipsă	72
2.2.1.1 Analiză comparativă a metodelor de tratare a valorilor lipsă... 78	
2.2.1.2 Tratarea valorilor lipsă în Streamlit de la preprocesare simplă la cea avansată	80
2.2.2 Tratarea valorilor (non-numerice sau categorice) lipsă	89
2.3 Codificarea valorilor non-numerice	97

2.3.1 Ordinea aplicării diferitelor metode de preprocesare	97
2.3.2 Introducere în codificarea valorilor non-numerice.....	100
2.4 Tratarea valorilor extreme.....	106
2.5 Scalarea	118
Întrebări recapitulative și exerciții	125
Referințe.....	127
Pagini utile.....	128
3. eXplainable Artificial Intelligence (XAI) cu Streamlit.....	129
3.1 Introducere	129
3.2 SHAP și LIME în modelarea clasică.....	130
3.2.1 Analiza SHAP	131
3.2.2 Analiza LIME.....	139
3.3 XAI în procesarea limbajului natural	141
3.3.1 Aplicarea tehnicilor XAI în contextul analizei de sentiment.....	143
3.3.2 XAI în traducerea automată	152
3.3.3 XAI în generarea de text.....	166
Întrebări recapitulative și exerciții propuse	182
Referințe.....	184
Pagini utile.....	185
4. Flask	186
4.1 Introducere	186
4.2 Obiecte globale în Flask.....	187
4.3 Arhitectura aplicațiilor Flask: structura și organizarea	215
4.4 Gestionarea cererilor și rutarea în Flask.....	223
4.4.1 Cum funcționează WSGI.....	225
4.4.2 Ciclul de viață al unei cereri	228
4.4.3 Gestionarea fișierelor statice.....	233
4.5 Ecosistemul de extensii Flask: arhitectura modulară și integrarea avansată	235
4.5.1 Flask-SQLAlchemy: Arhitectura ORM integrată.....	235
4.5.2 Flask-Migrate: Versionarea schemei bazei de date	243
4.5.3 Flask-Login: Sistemul de autentificare pentru aplicații web	248
4.5.4 Flask-WTF: Gestionarea formularelor Web.....	260
4.5.5 Flask-Bootstrap: Integrarea cu framework-uri CSS	261
4.5.6 Flask-Admin: Interfețe pentru administrare.....	261
4.5.7 Flask-RESTX: Dezvoltarea API-urilor RESTful	262

4.5.8 Flask-CORS: Gestionarea Cross-Origin Resource Sharing	263
4.5.9 Flask-Mail: Sistemul de integrare pentru servicii de e-mail	263
4.5.10 Flask-Caching: optimizarea performanței	264
4.5.11 Flask-Session: Gestionarea avansată a sesiunilor	265
4.5.12 Flask-Babel: Internaționalizarea aplicațiilor	265
4.6 Construcția unei aplicații web cu Flask.....	266
4.5.1 Partea I. Structura de bază.....	266
4.5.2 Partea II. Configurarea conexiunii la baza de date.....	270
4.5.3 Partea III. Construirea modelelor pentru baza de date	272
4.5.4 Partea IV. Construirea sistemului de înregistrare și autentificare.....	287
4.5.5 Partea V. Gestionarea modelelor.....	300
4.5.6 Partea VI. Utilizarea serviciilor externe și a API-urilor.....	317
Întrebări recapitulative și exerciții.....	335
Referințe.....	336
Link-uri pentru resurse Flask:	337
Partea a II-a. Prelucrarea datelor și analiza scalabilă în SAS.....	339
5. Prelucrarea și analiza predictivă a datelor în SAS	341
5.1 SAS în prelucrarea analitică a datelor	341
5.2 Elemente specifice programării SAS	345
5.2.1 Organizarea și delimitarea secțiunilor de programe SAS	346
5.2.2 Tipuri de fișiere cu date	347
5.2.3 Reguli de sintaxă și convenții de denumire	348
5.2.4 Tipuri de date	350
5.2.5 Biblioteci	351
5.2.6 Formate SAS.....	352
5.3 Elemente generale de programare SAS cu aplicabilitate largă	355
5.4 Macro-programarea în SAS	362
5.4.1 Macro-variabile	362
5.4.2 Macro-proceduri.....	364
5.4.3 Macro-proceduri cu parametri	364
5.4.4 Controlul execuției cu instrucțiuni condiționale în macro-proceduri	365
5.4.5 Transfer de valori din date către macro-variabile: CALL SYMPUT.....	367
5.5 Machine learning în SAS: analiza retenției clienților.....	368
5.5.1 Introducere	368
5.5.2 Probleme de clasificare în machine learning	371
5.5.3 Clasificare binară în SAS ML.....	376

5.5.3.1 Importul datelor	376
5.5.3.2 Etapa de curățare a datelor – tratarea valorilor lipsă.....	377
5.5.3.3 Analiza exploratorie a datelor	380
5.5.3.4 Crearea de variabile noi (Feature creation).....	386
5.5.3.5 Transformarea variabilelor non-numerice în variabile numerice	387
5.5.3.6 Calculul matricei de corelație în SAS.....	389
5.5.3.7 Oversampling.....	390
5.5.3.8 Scalare și împărțirea train-test.....	393
5.5.3.9 Antrenarea modelelor de machine learning	395
5.5.3.10 Utilizarea modelelor de ML pe datele balansate	403
Întrebări recapitulative și exerciții	408
Referințe.....	410
Pagini utile.....	410
6. Analiza datelor utilizând SAS Viya	411
6.1 Introducere în SAS Viya.....	411
6.1.1 Principalele avantaje ale utilizării platformei.....	413
6.2 Arhitectura SAS Viya.....	415
6.2.1 Arhitectura SAS Viya 3.4	415
6.2.2 Arhitectura SAS Viya 4.0	417
6.2.3 SAS Viya 3.4 vs. SAS Viya 4.0	419
6.3 Funcționalități cheie în SAS Viya	420
6.3.1 JupyterLab	420
6.3.2 SAS Viya pentru învățare.....	422
6.4 Analiza preliminară a datelor în SAS Viya	426
6.4.1 Descoperirea activelor informaționale.....	426
6.4.2 Gestionarea surselor de date	430
6.5 Analiza vizuală și exploratorie în SAS Viya.....	431
6.6 Machine learning și Inteligența Artificială în SAS Viya.....	444
Întrebări recapitulative și exerciții	461
Referințe.....	463
Pagini utile.....	464

Introducere

Contextul actual este definit de o abundență de date, ceea ce a condus la utilizarea algoritmilor complecși de învățare automată în luarea deciziilor în domenii variate, de la sectorul financiar-bancar și medicină, până la marketing și administrație publică. Totuși, în mediul de lucru curent, există o discontinuitate între etapa de creare a unui model performant într-un mediu izolat și transformarea acestuia într-un produs utilizabil, tangibil și generator de valoare.

Cartea își propune să unească aceste etape, servind nu doar ca un ghid tehnic, cât și ca manual de arhitectură a soluțiilor analitice. Având ca obiectiv operaționalizarea analizei tradiționale, manualul oferă instrumentele necesare pentru ca experții în date să poată gestiona întregul ciclu de viață al unui proiect, de la prelucrarea statistică riguroasă, specifică analizei de date, până la implementarea unor soluții software interactive, dinamice și scalabile, specifice ingineriei datelor.

Conținutul cărții se adresează în special studenților din anii terminali de licență, oferindu-le instrumentele necesare pentru a trece de la teorie la practică (deployment). Premisa centrală a cărții este aceea că cititorul stăpânește deja fundamentele matematice și statistice ale algoritmilor de analiză a datelor, fie prin studierea acestora în cadrul facultății, prin lectura manualelor dedicate acestora sau prin învățare individuală din diverse surse. Prin urmare, discuția nu se concentrează asupra teoriei cuprinzătoare din spatele regresiei sau clasificării, ci explorează operaționalizarea acestor concepte prin intermediul unor interfețe moderne și fluxuri de lucru automatizate.

Structura cărții se bazează pe doi piloni fundamentali ai analizei de date moderne: ecosistemul Python și ecosistemul SAS. Abordarea celor două direcții are ca scop reflectarea realității hibride întâlnită în marile corporații și instituții. Pentru a excela în știința datelor este importantă capacitatea de a selecta instrumentul optim pentru fiecare etapă a proiectului de analiză. Astfel, familiarizarea cu ambele paradigme oferă libertatea de a exploata atât flexibilitatea bibliotecilor Python pentru dezvoltarea de aplicații, cât și rigoarea, securitatea și capacitatea de procesare masivă/nișată specifice mediului SAS.

Cartea este organizată în două părți distincte, dar interconectate, axate pe prelucrarea avansată a datelor și dezvoltarea aplicațiilor, Partea I: Dezvoltarea de aplicații web și interactivitate în Python, și Partea a II-a: Prelucrarea datelor și analiza scalabilă în SAS.

Prima parte a lucrării, capitolele 1-4, explorează dezvoltarea web orientată către date, utilizând limbajul Python. Accentul cade pe două framework-uri esențiale pentru prototipizare și producție: Streamlit și Flask.

Capitolele 1 și 2 introduc Streamlit, un instrument principal pentru prototipizare rapidă. Streamlit a revoluționat modul de prezentare a modelelor de învățare automată permițând transformarea scripturilor de date în aplicații web partajabile. Aceste capitole analizează structura unei aplicații, inclusiv implementarea paginilor de autentificare și importanța preprocesării datelor în interfață. Procese precum tratarea valorilor lipsă, codificarea variabilelor sau scalarea datelor devin procese interactive, oferind utilizatorului final control, transparență, și mai ales reproductibilitatea rezultatelor.

Capitolul 3 se concentrează pe eXplainable Artificial Intelligence (XAI). În contextul în care algoritmi de tip "Black Box" au un rol important în procesul decizional, interpretabilitatea rezultatelor devine o cerință obligatorie. Capitolul detaliază integrarea bibliotecilor precum SHAP și LIME în aplicațiile Streamlit, arătând cum se pot explica deciziile modelelor, atât pentru date structurate, cât și în procesarea limbajului natural, de la analiza de sentiment la generarea de text.

Deși Streamlit excelează în prototipizare rapidă și demonstrarea conceptelor analitice, aplicațiile de producție necesită adesea arhitecturi web mai flexibile și mai robuste. Capitolul 4 răspunde acestei nevoi prin introducerea framework-ului Flask, care permite construirea aplicațiilor web scalabile destinate mediilor profesionale. Capitolul demonstrează cum modelele de machine learning pot fi transformate în servicii web accesibile și sigure, capabile să proceseze cereri în timp real și să se integreze cu alte sisteme software. Prin exemple concrete, de la aplicații simple de management al datelor până la sisteme complete care utilizează API-uri externe pentru generarea automată de conținut educațional, cititorul va dobândi competențele necesare pentru a transforma munca analitică în produse software funcționale. Trecere de la Streamlit la Flask reflectă evoluția naturală a unui proiect, de la validarea conceptului către implementarea în producție, unde fiabilitatea, securitatea și performanța devin la fel de importante ca și acuratețea modelelor.

Dacă prima parte a cărții explorează capacitățile ecosistemului open-source, a doua parte, capitolele 5 și 6, se concentrează pe mediul SAS, un standard în industriile reglementate, precum sectorul bancar și asigurări. Această parte nu este o simplă alternativă la Python, ci o completare pentru un profil profesional, urmând un parcurs didactic logic, de la înțelegerea mecanismelor interne de procesare (SAS), către arhitecturile moderne de tip cloud (SAS Viya). Prezența SAS alături de tehnologiile open-source este justificată de fiabilitatea, suportul tehnic și capacitatea acestuia de a gestiona volume masive de date cu o stabilitate superioară.

Capitolul 5 stabilește fundația analitică, prezentând programarea SAS tradițională dintr-o perspectivă orientată spre analiza predictivă. Înainte de adoptarea instrumentelor vizuale de automatizare, este necesară consolidarea competențelor de definire și prelucrare a datelor prin intermediul programelor specifice și al macro-programării. Este permis, astfel, controlul asupra execuției, asigură trasabilitatea pașilor operaționali și sprijină automatizarea proceselor complexe. De asemenea, un studiu de caz privind analiza retenției clienților ilustrează aplicabilitatea practică a învățării automate în mediul SAS.

Evoluția tehnologică este cuprinsă în Capitolul 6, dedicat platformei SAS Viya. Aici, rigurozitatea codului întâlnește puterea procesării distribuite și a analizei vizuale. Trecerea de la SAS clasic la Viya este justificată de necesitatea manipulării volumelor mari de date cu o viteză superioară. Capitolul demonstrează cum funcționalitățile de inteligență artificială din Viya pot fi exploatate la maxim doar având baza de înțelegere a datelor formată în capitolul anterior.

Ultimele două capitole oferă o abordare practică, ce reflectă realitatea din marile companii: folosirea stabilității programării SAS pentru gestionarea datelor importante și a puterii SAS Viya pentru analiză vizuală și modelare avansată. Astfel, cititorul va fi capabil să înțeleagă și să îmbunătățească întreg procesul analitic, combinând cunoștințele din sfera dezvoltării software cu cele specifice analizei de date.

Deși tehnologiile prezentate vizează două tehnologii distincte, ideea de bază a acestei cărți este capacitatea de a construi sisteme decizionale integrate. Indiferent dacă se utilizează ecosistemul Python pentru dezvoltarea de aplicații web modulare și implementarea metodelor XAI sau se utilizează programarea SAS pentru realizarea unor analize predictive robuste, extinzând ulterior capacitățile prin SAS Viya pentru vizualizare scalabilă și procesare distribuită, obiectivul rămâne același: extragerea informației utile din date prin intermediul unor interfețe accesibile și a unor fluxuri de lucru transparente.

Cartea este concepută ca un instrument de studiu practic. Fiecare capitol se încheie cu o bibliografie selectată de autori și cu exerciții recapitulative menite să testeze cunoștințele și să stimuleze gândirea analitică a studenților. Accentul este pus pe aplicabilitatea limbajelor Python și SAS, prin întrebări recapitulative și exerciții practice. Dincolo de aspectele teoretice, lucrarea vizează formarea capacității de a dezvolta aplicații pentru transformarea datelor în resurse decizionale utile. Eficiența procesului este susținută de utilizarea tehnologiilor de inteligență artificială, abordarea fiind orientată spre obținerea de rezultate concrete și relevante în contextul profesional actual.