

# TEHNOLOGII ȘI INSTRUMENTE UTILIZATE ÎN ȘTIINȚA DATELOR

## Capitolul 8

# TEHNOLOGII UTILIZATE ÎN DS

- ▶ 1. Limbaje de programare: Python, R, SQL
- ▶ 2. Biblioteci și framework-uri: NumPy, Pandas, TensorFlow
- ▶ 3. Baze de date: SQL, MongoDB
- ▶ 4. Instrumente de vizualizare: Tableau, PowerBI
- ▶ 5. Platforme cloud: AWS, Google Cloud, Azure

# LIMBAJE DE PROGRAMARE

- ▶ 1. Python: Simplu, flexibil, suport larg
- ▶ 2. R: Specializat în analize statistice
- ▶ 3. SQL: Manipularea bazelor de date relaționale
- ▶ 4. Scala: Performant pentru procesare Big Data
- ▶ 5. Julia: Creat pentru calcule numerice rapide

# BIBLIOTECI ȘI FRAMEWORK- URI

- ▶ 1. NumPy: Manipularea matricelor
- ▶ 2. Pandas: Analiza datelor tabelare
- ▶ 3. TensorFlow: Învățare automată
- ▶ 4. Keras: Interfață simplă pentru rețele neuronale
- ▶ 5. PyTorch: Flexibilitate pentru modele dinamice

# BAZE DE DATE ȘI DBMS

- ▶ 1. Baze de date relaționale: SQL, MySQL, PostgreSQL
- ▶ 2. Baze de date NoSQL: MongoDB, Cassandra
- ▶ 3. Baze de date in-memory: Redis
- ▶ 4. Sisteme de fișiere distribuite: Hadoop, Spark

# INSTRUMENTE DE ANALIZĂ ȘI VIZUALIZARE

- ▶ 1. Jupyter Notebooks: Explorare interactivă
- ▶ 2. Tableau: Vizualizare interactivă
- ▶ 3. Power BI: Raportare și vizualizare avansată
- ▶ 4. Matplotlib: Vizualizare grafică în Python
- ▶ 5. ggplot2: Vizualizare estetică în R

# PLATFORME CLOUD

- ▶ 1. AWS: Amazon Web Services
- ▶ 2. Microsoft Azure: Cloud flexibil
- ▶ 3. Google Cloud: BigQuery pentru analiza datelor
- ▶ 4. IBM Cloud: Soluții pentru inteligență artificială

# INSTRUMENTE DE ÎNVĂȚARE AUTOMATĂ

- ▶ 1. TensorFlow: Framework open-source
- ▶ 2. Keras: Ușor de utilizat pentru rețele neuronale
- ▶ 3. PyTorch: Suport dinamic pentru fluxuri de date
- ▶ 4. Scikit-learn: Algoritmi de clasificare și regresie
- ▶ 5. H2O.ai: Scalabil pentru seturi de date mari



# INSTRUMENTE DE PREPROCESARE A DATELOR

- ▶ 1. Pandas: Curățare și transformare date
- ▶ 2. NumPy: Operații matematice și matrici
- ▶ 3. Scikit-learn: Scalare, codificare, selecția caracteristicilor
- ▶ 4. Dask: Prelucrare paralelă a datelor mari
- ▶ 5. OpenRefine: Curățare date nestructurate

# INSTRUMENTE DE INTERPRETARE A MODELELOR

- ▶ 1. Scikit-learn: Vizualizarea importanței caracteristicilor
- ▶ 2. XGBoost: Analiza arborilor de decizie
- ▶ 3. LIME: Explicarea deciziilor modelului
- ▶ 4. SHAP: Atribuirea influenței caracteristicilor
- ▶ 5. ELI5: Explicarea predicțiilor modelelor