

UNIVERSITATEA TITU MAIORESCU

FACULTATEA DE INFORMATICĂ

**MASTER ȘTIINȚA DATELOR ȘI INTELIGENȚĂ ARTIFICIALĂ
FUNDAMENTELE ȘTIINȚEI DATELOR**

**CURSUL 1
INTRODUCERE ÎN DATA SCIENCE
Prezentarea cursului**

Motto

Schimbarea este prețul plătit progresului după
cum îndoiala este prețul plătit înțelepciunii.

**Autor:
Prof.univ.dr.ing.Titi PARASCHIV**

București
2024

CURSUL 1. INTRODUCERE ÎN DATA SCIENCE

DS este o știință interdisciplinară, apărută după anul 2000, care combină matematica, statistica și informatica pentru analiza și interpretarea datelor.

Obiectivul DS este identificarea modelelor, tiparelor și tendințelor din date pentru analiza descriptivă (diagnostic), predictivă și prescriptivă.

Aplicații ale DS sunt în business, sănătate, educație, societate, mediu și securitate prin transformarea datelor în cunoștințe utile pentru optimizarea proceselor și predicții.

Flux: **Date - informații - cunoștințe (înțelegere) - operaționalitate - înțelepciune.**

Evoluția DS: Tehnicile statistice ale secolului XIX, dezvoltarea tehnologiilor după 1947, apoi după 1990, popularizarea termenului în anii 2000, rezultatele anilor 2010 și întemeierea ca știință în 2016.

Contribuții ale unor cercetători din statistică: Fr. Galton, Ronald A. Fisher și Karl Pearson, apoi John Tukey (**fondator al analizei datelor exploratorii-EDA și a vizualizării datelor**), Geoffrey Hinton (**rețele neuronale și învățare profundă DL**), Yann LeCun (DL și RNC), Andrew Ng (ML și co-fondator al platformei Coursera), Peter Norvig (IA și DS).

Importanța DS: Facilitarea luării deciziilor. Impact în toate domeniile vieții sociale.

Domenii de aplicare ale DS: în retail, finanțe, sănătate, educație, tehnologie, robotică, autovehicule autonome.

Procesul de lucru în DS: înțelegerea problemei, colectarea datelor, modelarea și interpretarea rezultatelor, realizarea și validare modelului.

Instrumente și Tehnologii: Limbaje: Python, R, SQL, SPSS, fiecare cu biblioteci consacrate pe diverse domenii de cercetare. **Framework-uri:** TensorFlow, PyTorch.

Platforme de stocare: Hadoop, Spark.

CURSUL 2 FUNDAMENTELE MATEMATICE ȘI STATISTICE ÎN DS

- 1. Introducere:** concepte fundamentale: algebra liniară, calculul diferențial și integral, teoria grafurilor, tehnici de optimizare și învățare automată (ML).
- 2. Matematica în DS** este esențială pentru înțelegerea tehnicilor DS: algebra liniară, statistica, geometrie analitică și cercetări operaționale calculul diferențial și integral, și algoritmi de optimizare.
- 3. Algebra Liniară: Vectori:** Definiție, adunare, produs scalar și vectorial; **Matrici:** Definiție, operații și proprietăți; **Ecuații liniare:** Soluții prin metode de substituție, eliminare și utilizarea matricilor.
- 4. Statistica:** Colectarea, organizarea, analiza și interpretarea datelor. Include statistica descriptivă și inferențială.
Indicatori descriptivi: Măsurile ale tendinței centrale, variabilitate, simetrie și boltire.
Statistica inferențială: Estimarea parametrilor, testarea ipotezelor, niveluri și intervale de încredere.
- 5. Calculul Diferențial și Integral**
Diferențierea: Calcularea ratei de schimbare a funcțiilor, monotonie, punctele de extrem și gradientul.
Integrarea: Calculul ariei sub o curbă și aplicații în fizică, economie și inginerie.
- 6. Algoritmi de Optimizare:** gradient descendent, algoritmi genetici și metode de căutare;
- 7. Teoria Grafurilor: Concepte cheie:** Noduri, muchii, grafuri orientate și neorientate, arbori. Aplicații în optimizarea rețelelor de comunicații și științe sociale.
- 8. Statistica în DS:** Particularitățile includ volumul și varietatea datelor, abordarea interdisciplinară și integrarea tehnologiei. **Tehnici avansate:** Rețele neuronale, învățare supervizată și nesupervizată.
- 9. Modelare Statistică:** Construirea unui model care explică relațiile dintre variabile. Tehnicile sunt: regresia liniară, logistică și arbori de decizie.

CURSUL 3 PREPROCESAREA DATELOR

1. **Preprocesarea datelor** este un proces esențial în DS prin pașii: curățarea, transformarea și pregătirea datelor pentru a îmbunătăți performanța modelelor de ML și pentru a crește acuratețea analizei.

2. **Importanța preprocesării datelor** se realizează prin eliminarea valorilor eronate sau incomplete, pregătirea datelor pentru modelare și îmbunătățirea performanțelor algoritmilor de ML. Ajută la creșterea ponderii informațiilor în date.

Operațiile preprocesării datelor:

3. **Curățarea datelor** presupune eliminarea valorilor eronate, incomplete sau duplicate pentru a asigura un set de date curat și de calitate.

4. **Completarea valorilor lipsă** se face utilizând metodele: interpolarea, utilizarea regresiei sau algoritmi ML.

5. **Normalizarea datelor** transformă valorile într-un interval comun (de exemplu, între 0 și 1), eliminând astfel influențele valorilor extreme asupra modelelor.

6. **Transformarea datelor** - aplicarea unor funcții ex. logaritmice sau Box-Cox pentru a ajusta distribuția valorilor și a facilita analiza.

7. **Agregarea datelor** - combinarea mai multor seturi de date pentru a obține o viziune de ansamblu și pentru a simplifica analiza.

8. **Eliminarea zgomotului** - eliminarea informațiilor nerelevante sau eronate din date, obținând astfel un set de date mai curat și mai relevant pentru analiză.

9. **Codificarea valorilor categoriale** se realizează prin transformarea acestora în valori numerice, utilizând metode precum codificarea ordinală sau One-Hot.

CURSUL 4 EXPLORAREA DATELOR

- 1. Introducere.** Explorarea datelor este procesul de analiză preliminară a datelor pentru a înțelege și identifica modele, relații, tendințe, anomalii și outlier-uri.
- 2. Etapele Explorării Datelor:** încărcarea datelor, curățarea, analiza descriptivă, vizualizarea datelor și identificarea outlier-ilor.
- 3. Identificarea Outlier-ilor** este realizată cu urm. Metode: Boxplot, Z-score și IQR.
- 4. Transformarea Datelor** cuprinde tehnici precum normalizarea, agregarea și discretizarea.
- 5. Verificarea Ipotezelor** - formularea ipotezelor, metoda de testare, aplicarea testului statistic și interpretarea rezultatelor.
- 6. Analiza Distribuției** - face prin următoarele metode: histograme, grafice de densitate și diagrame Q-Q.
- 7. Analiza Corelațiilor:** Corelațiile sunt pozitive, negative, nule sau parțiale. Coef. de corelație: Pearson, Spearman și Kendall.
- 8. Vizualizarea Datelor** - grafice tip bară, scatter plots, diagrame de tip radier și hărți.
- 9. Agregare și Grupări** - Datele pot fi combinate sau clasificate în grupuri mai mari. Gruparea se face în funcție de valori și clusterizarea.
- 10. Statistici Descriptive** – Indicatorii statisticii descriptive sunt: centralitatea (media, mediana, modul) și dispersia (deviația standard, coeficientul de variație).
- 11. Aplicații Practice** - Explorarea datelor este aplicată în: Business Intelligence, marketing, sănătate și cercetare.
- 12. Rezumat**

Explorarea datelor este o bază solidă pentru analizele ulterioare și sprijină luarea deciziilor informate în procesele de analiză a datelor.

CURSUL 5 MODELAREA DATELOR

1. Introducere în Modelarea Datelor

Modelarea este procesul de creare a unui model sau a unei reprezentări a unui sistem real. **2.**

Etapele de bază ale procesului de modelare

1. Colectarea datelor; 2. Preprocesarea datelor; 3. Explorarea datelor; 4. Selecția modelului; 5. Antrenarea modelului; 6. Evaluarea modelului; 7. Afișarea rezultatelor; 8. Implementarea și întreținerea modelului

3. Tehnici și Metode de Modelare

Principalele tehnici sunt: - Regresie; - Clasificare; - Clustering; - Analiza factorilor latenți; - Algoritmi de învățare automata; - Analiza de componente principale (PCA)

4. Modele de Clasificare - Regresia logistică, Arborele de decizie, Random Forest, Naive Bayes, SVM, Rețele Neurale, KNN.

5. Modele de Regresie Tipuri: Regresie liniară simplă, multiplă, polinomială. Aceste modele se bazează pe relația liniară sau polinomială între VI și VD.

6. Modele de Clustering

Tehnici: K-Means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Model (GMM). Aceste tehnici sunt utilizate pentru gruparea datelor în funcție de similarități fără etichete prestabilite.

7. Validarea și Evaluarea Modelelor

Tehnici: K-fold cross-validation, Matricea de confuzie, Acuratețea, Precizia, F1-Score. Validarea modelelor este esențială pentru a asigura că modelul funcționează bine pe date noi.

8. Selecția Modelului și Evitarea Overfitting-ului

Tehnici de selecție: Forward-Backward, GridSearchCV. Se recomandă evitarea modelelor care sunt prea complexe (overfitting) sau prea simple (underfitting).

CURSUL 6 EVALUAREA ȘI VALIDAREA MODELELOR

1. Importanța Evaluării și Validării

Evaluarea și validarea modelelor sunt esențiale pentru a asigura acuratețea, precizia și fiabilitatea acestora.

2. Metode și Tehnici de Evaluare

1. Metrice Supervizate: Acuratețea, precizia, recall, F1-score, ROC curve.
2. Validare încrucișată (cross-validation): Împărțirea setului de date în multiple subseturi.
3. Evaluare cu set de testare independent: Modelul este testat pe un set separat de date.

3. Metrice de Evaluare Supervizate

Acuratețea (Accuracy): Proporția corectă a clasificărilor făcute de model.

Precizia (Precision): Proporția de exemple pozitive corect identificate.

Recall: Proporția de exemple pozitive identificate corect.

F1-Score: Media armonică între precizie și recall.

4. Validare Încrucișată

Împărțirea setului de date în k subseturi pentru a evalua performanța modelului și pentru a preveni supraantrenarea.

5. Evaluarea cu Set de Testare Independent

Separarea unui set de date pentru testare independentă, pentru a estima capacitatea modelului de a generaliza pe date noi.

6. Validarea cu Set de Validare

Folosirea unui set separat de date pentru ajustarea parametrilor modelului înainte de testarea finală.

7. Metrice de Evaluare Ne-Supervizate

Exemple: Silhouette score, Coerența, Rand Index, Dunn Index, Inertie. Aceste metrice sunt folosite în evaluarea clustering-ului.

8. Validarea prin Configurarea Hiperparametrilor

Ajustarea parametrilor modelului pentru a maximiza performanța. Tehnici utilizate: Grid search, Random search.

CURSUL 7 APLICAȚII PRACTICE ÎN DS

1. Domeniile de aplicare ale Data Science

- Marketing: Segmentarea pieței și optimizarea campaniilor publicitare.
- Finanțe: Analiza datelor și identificarea riscurilor financiare.
- Sănătate: Diagnosticare și identificarea pattern-urilor în date medicale.

2. Operațiile ce compun tehnicile de Data Science: identificarea pattern-urilor, automatizarea proceselor, luarea deciziilor bazate pe date și îmbunătățirea experienței clienților.

3. Particularitățile aplicării DS în Tehnologie: analiza comportamentului utilizatorilor, personalizare și predicții de calitate.

4. Particularitățile aplicării DS în Psihologie: colectarea și analiza datelor din studii de caz, evaluarea tratamentelor și realizarea de predicții.

5. Particularitățile aplicării DS în Sănătate: predicții medicale, personalizarea tratamentelor și optimizarea proceselor.

6. Proiecte practice - personalizarea ofertelor în marketing, prevenirea fraudelor în finanțe și recomandări de tratament în sănătate.

7. Tehnologiile utilizate în DS: limbaje de programare (Python, R), biblioteci și framework-uri (NumPy, Pandas, TensorFlow), precum și platforme de cloud computing (AWS, Azure, GCP).

8. Analiza de caz. Etapele unui proiect de DS: înțelegerea problemei, pre-procesarea datelor, modelarea și evaluarea modelului, fine-tuning și integrarea modelului în aplicații.

9. Comunitatea Data Science joacă un rol esențial în schimbul de cunoștințe și resurse, colaborare și dezvoltarea profesională.

10. Perspectivele Data Science IA, analiza Big Data în timp real și securitatea datelor.

CURSUL 8 TEHNOLOGII ȘI INSTRUMENTE ÎN DS

1. Tehnologii Utilizate în DS

- Limbaje de programare: Python, R, SQL
- Biblioteci și framework-uri: NumPy, Pandas, TensorFlow
- Baze de date: SQL, MongoDB
- Instrumente de vizualizare: Tableau, Power BI
- Platforme cloud: AWS, Google Cloud, Azure

2. Limbaje de Programare

- Python: Simplu, flexibil, suport larg
- R: Specializat în analize statistice
- SQL: Manipularea bazelor de date relaționale
- Scala: Performant pentru procesarea Big Data
- Julia: Creat pentru calcule numerice rapide

3. Biblioteci și Framework-uri

- NumPy: Manipularea matricelor
- Pandas: Analiza datelor tabelare
- TensorFlow: Învățare automată
- Keras: Interfață simplă pentru RNA
- PyTorch: Flexibilitate pentru modele dinamice

4. Baze de Date și DBMS

- Baze de date relaționale: SQL, MySQL, PostgreSQL
- Baze de date NoSQL: MongoDB, Cassandra
- Baze de date in-memory: Redis
- Sisteme de fișiere distribuite: Hadoop, Spark

5. Instrumente de Analiză și Vizualizare

- Jupyter Notebooks: Explorare interactivă
- Tableau: Vizualizare interactivă
- Power BI: Raportare și vizualizare avansată
- Matplotlib: Vizualizare grafică în Python
- ggplot2: Vizualizare estetică în R

6. Platforme Cloud

- AWS: Amazon Web Services
- Microsoft Azure: Cloud flexibil
- Google Cloud: BigQuery pentru analiza datelor
- IBM Cloud: Soluții pentru inteligență artificială

7. Instrumente de Învățare Automată

- TensorFlow: Framework open-source
- Keras: Ușor de utilizat pentru rețele neuronale
- PyTorch: Suport dinamic pentru fluxuri de date
- Scikit-learn: Algoritmi de clasificare și regresie
- H2O.AI: Scalabil pentru seturi de date mari

8. Instrumente de Preprocesare a Datelor

- Pandas: Curățare și transformare date
- NumPy: Operații matematice și manipularea matricelor
- Scikit-learn: Scalare, codificare, selecția caracteristicilor
- Dask: Prelucrare paralelă a datelor mari
- OpenRefine: Curățare date nestructurate

9. Instrumente de Interpretare a Modelelor

- Scikit-learn: Vizualizarea importanței caracteristicilor
- XGBoost: Analiza arborilor de decizie
- LIME: Explicarea deciziilor modelului
- SHAP: Atribuirea influenței caracteristicilor
- ELI5: Explicarea predicțiilor modelelor

CURSUL 9 MACHINE LEARNING

1. Definirea conceptelor fundamentale ale Machine Learning

ML cuprinde algoritmi care permit sistemelor să învețe din date. Tehnicile principale: regresia și clasificarea pentru prezicerea și categorisirea datelor.

2. Modele de învățare automată

- Modele supervizate: regresie, rețele neuronale, arbori de decizie.
- Modele nesupervizate: clustering și reducerea dimensionalității.

Exemple: Regresie liniară, Random Forest, SVM, K-Means.

3. Pre-procesarea datelor include:

- Curățarea datelor: gestionarea valorilor lipsă și incorecte.
- Normalizarea datelor: scalarea datelor la un interval comun.
- Codificarea variabilelor categoricale: transformarea acestora în valori numerice.
- Divizarea setului în train-test pentru antrenament și evaluare.

4. Evaluarea modelelor de ML se realizează prin măsurători de performanță precum valoarea predictivă și incidența. Cross-validation este o metodă frecvent utilizată pentru împărțirea datelor în seturi de antrenament și testare.

5. Tehnicile de îmbunătățire a ML:

- Optimizarea hiper-parametrilor
- Selecția caracteristicilor relevante
- Ensemble learning și tehnici de regulare
- Transfer learning

6. Aplicații practice ale Machine Learning

- Clasificare: detectarea spam-ului și analiza sentimentelor
- Regresie: predicția prețului acțiunilor
- Clustering: segmentarea clienților
- Detectarea anomaliilor și sisteme de recomandare

1. DL este o ramură a învățării automate care utilizează rețele neuronale cu mai multe straturi pentru a analiza și înțelege date complexe. Rețelele neuronale extrag caracteristici și le reprezintă ierarhic din date.

2. Componentele Unui Sistem de Deep Learning

- Date de intrare: imagini, texte, semnale audio
- Arhitectura rețelei neuronale: straturi convoluționale, recurente
- Straturi ascunse pentru extracția caracteristicilor
- Strat de ieșire pentru predicțiile finale
- Funcții de activare: ReLU, Sigmoid, Tanh, Softmax
- Algoritmi de optimizare: Gradient Descent, Adam

3. Rețele Neuronale Artificiale sunt modele inspirate de creierul uman folosite în clasificare, regresie și procesarea limbajului. Tipuri de rețele includ Feedforward, RNN (recurente), CNN (convoluționale) și GAN (Generative Adversarial Networks).

4. Funcții de Activare în Deep Learning

5. Antrenarea Rețelelor Neuronale. Tehnicile de îmbunătățire sunt: regularizarea, dropout și normalizarea loturilor.

6. Transfer Learning - utilizarea modelelor pre-antrenate pe seturi mari de date pentru sarcini similare, ceea ce reduce timpul de antrenare și îmbunătățește performanța pe seturi mici de date.

7. Deep Learning a evoluat datorită dezvoltării algoritmului backpropagation, a GPU-urilor și disponibilității datelor masive. Tendințele actuale includ arhitecturi complexe și integrarea cu tehnologii AI avansate.

8. Aplicații Practice ale Deep Learning

- Computer Vision: Clasificarea și detectarea obiectelor
- Procesarea Limbajului Natural (NLP): Traducerea automată, analiza sentimentelor
- Speech Processing: Transcrierea și generarea vorbirii
- Predictive Analytics: Prezicerea vânzărilor și evenimentelor

9. Evaluarea Performanței Rețelelor Neuronale se realizează folosind metrice precum acuratețea, F1-score și AUC. Validarea include cross-validation și testarea pe date necunoscute, iar evaluarea vizuală poate fi folosită pentru a analiza rezultatele.

10. Perspectivele în Deep Learning: dezvoltarea de arhitecturi noi, integrarea cu IoT și Edge Computing.

1. Introducere

Big Data reprezintă seturi de date extrem de mari și complexe, caracterizate prin volum, viteză și varietate

2. Caracteristicile Big Data

- Volum: - Viteză: - Varietate: Diferite tipuri și surse de date, inclusiv text, imagini și video.

3. Tehnologii Big Data

- Hadoop: Framework open-source pentru stocare și procesare distribuită.
- Apache Spark: Motor pentru analiza datelor în memorie.
- NoSQL: Baze de date non-relaționale pentru gestionarea datelor nestructurate.

4. Arhitectura Big Data este distribuită pentru a asigura scalabilitate și redundanță. Sistemul HDFS (Hadoop Distributed File System) și procesarea MapReduce sunt componente cheie.

5. Pre-Procesarea Datelor

Pre-procesarea include curățarea, transformarea și agregarea datelor pentru a îmbunătăți calitatea și acuratețea acestora.

6. Analiza Datelor:

- Învățare automată: Algoritmi pentru predicție și clasificare.
- Analiza statistică: Teste de ipoteze și regresii.
- Analiza textuală: Procesarea limbajului natural și clasificarea textului.

7. Stocarea Datelor în depozite distribuite sau în memorie principală pentru viteză și acces rapid, sau în depozite bazate pe coloane pentru eficiență.

8. Securitatea și Protecția Datelor criptarea datelor, autentificarea și autorizarea accesului, precum și monitorizarea activităților suspecte.

9. Aplicații Practice ale Big Data

- Marketing: Personalizarea ofertelor și campaniilor publicitare.
- Sănătate: Prevenirea epidemiilor și optimizarea tratamentelor.
- Finanțe: Identificarea fraudelor și optimizarea investițiilor.

10. Perspective Privind Big Data

Tendențele viitoare includ analiza de date în cloud, procesarea în timp real și creșterea utilizării IoT pentru a gestiona volumele mari de date generate de dispozitive conectate.

CURSUL 12 ANALIZA DE TEXT

- 1. Introducere în Analiza de Text** este procesul de extragere și înțelegere a informațiilor semantice și contextuale dintr-un text scris sau verbal.
- 2. Motivația și Importanța Analizei de Text**
- 3. Domenii de Aplicare ale Analizei de Text**
 - Social Media: Monitorizarea sentimentelor și opiniilor.
 - Servicii clienți: Gestionarea feedback-ului.
 - Monitorizarea presei: Identificarea tendințelor din conținutul media.
 - Detectarea fraudei: Identificarea comportamentelor suspecte.
- 4. Procesul General al Analizei de Text:** - Colectarea și curățarea datelor
 - Preprocesarea textului: tokenizare, eliminarea stopwords, lematizare
 - Analiza semantică și gramaticală; - Clasificarea și etichetarea textului; - Extracția informațiilor relevante
- 5. Instrumente și Tehnologii Utilizate în Analiza de Text**
 - Biblioteci NLP: NLTK, SpaCy, TensorFlow
 - Algoritmi de analiză sentiment: Naive Bayes, SVM, Random Forests
 - Platforme de analiză: IBM Watson, Google Cloud NLP, Microsoft Azure Text Analytics
- 6. Provocări și Limitări ale Analizei de Text:** ambiguitatea, polisemia, detectarea ironiei și gestionarea volumului mare de date. De asemenea, bias-ul și subiectivitatea pot influența rezultatele.
- 7. Preprocesarea Textului:** curățarea textului, tokenizarea, lematizarea, stemming-ul și normalizarea datelor textuale pentru a le face ușor de analizat.
- 8. Analiza Sentimentului** presupune atribuirea unei semnificații emoționale unui text, prin metode precum dicționare de sentiment și algoritmi de machine learning.
- 9. Analiza Tematică** identifică și înțelege temele și subiectele dintr-un text, utilizând metode precum analiza frecvenței cuvintelor și topic modeling.
- 10. Interpretarea și Vizualizarea Rezultatelor**

Rezultatele analizei textului pot fi vizualizate prin distribuția cuvintelor-cheie, hărți de căldură și diagrame cu nori de cuvinte.
- 11. Evaluarea Rezultatelor** se face prin măsurători precum coerența tematică și acuratețea modelelor, în combinație cu interpretarea umană a temelor.

CURSUL 13 ANALIZA DE IMAGINI

- 1. Introducere în Analiza de Imagini.** Operațiile sunt: preprocesarea, segmentarea și clasificarea imaginilor digitale. Aceasta joacă un rol esențial în domenii precum medicină, securitate, vehicule autonome și entertainment.
- 2. Preprocesarea Imaginilor:** redimensionarea imaginilor, eliminarea zgomotului și segmentarea regiunilor de interes pentru analiza ulterioară.
- 3. Extracția caracteristicilor vizuale și semantice** se realizează prin algoritmi statistici și transformări matematice. Caracteristici precum culoarea, textura și formele sunt esențiale pentru recunoașterea obiectelor și analiza contextului imaginii.
- 4. Segmentarea Imaginilor** presupune identificarea și separarea regiunilor semnificative, folosind metode bazate pe praguri, umbre sau contururi. Aceasta ajută la detectarea și clasificarea formelor și obiectelor din imagine.
- 5. Detectarea și Recunoașterea de Obiecte** se realizează prin metode precum Sliding Windows și Deep Learning, iar recunoașterea se bazează pe algoritmi clasici (Haar-like, SIFT, HOG) și arhitecturi de rețele neuronale convoluționale (CNN).
- 6. Clasificarea Imaginilor.** Algoritmii utilizați pentru clasificarea imaginilor includ SVM, CNN și Random Forests.
- 7. Recuperarea și Indexarea Imaginilor** se bazează pe conținutul vizual, iar indexarea eficientă se realizează prin tehnici de hashing și utilizarea de embeddings vizuale.
- 8. Aplicații și Studii de Caz**
Aplicațiile analizei de imagini includ domenii precum medicina, securitatea, vehiculele autonome și realitatea augmentată. Studiile de caz includ detectarea obiectelor, diagnosticarea medicală și recunoașterea conținutului vizual.
- 9. Generarea de Imagini** - rețele generative adversariale (GAN) și autoencodere, utilizate pentru generarea de imagini realiste sau bazate pe text.

CURSUL 14 ANALIZA SERIILOR TEMPORALE

1. Introducere în Analiza Seriilor Temporale. O serie temporală este o secvență de date ordonate în funcție de timp, formată din observații preluate la intervale de timp egale sau inegale. Caracteristici principale: ordinea temporală, dependența temporală, sezonalitatea, tendințele și fluctuațiile severe.

2. Importanța și Aplicațiile Analizei Seriilor Temporale

Analiza seriilor temporale este importantă pentru prognoză, optimizarea resurselor și detectarea anomaliilor.

3. Domenii de Aplicare ale Analizei Seriilor Temporale

- Economie și finanțe: Modelarea și previziunea piețelor financiare.
- Meteorologie: Predicții meteorologice și climatice.
- Medicină: Monitorizarea semnalelor fiziologice.
- Logistică: Planificarea producției și a lanțului de aprovizionare.

4. Preprocesarea Seriilor Temporale curățarea datelor (eliminarea zgomotului, tratarea valorilor lipsă), eșantionarea și interpolarea pentru completarea datelor lipsă.

5. Vizualizarea Seriilor Temporale prin grafice de linie și diagrame temporale, separând componentele de trend, sezonalitate și reziduale.

6. Analiza Exploratorie a Seriilor Temporale prin măsurători de tendință centrală, dispersie, autocorelație și schimbări de regim pentru a evidenția fluctuațiile.

7. Modele de Analiză a Seriilor Temporale

8. Prognozarea Seriilor Temporale

Prognozarea utilizând modele ARIMA implică estimarea parametrilor și evaluarea performanței prognozelor prin metrici precum MAE, MSE, RMSE și MAPE.

9. Analiza Sezonieră a Seriilor Temporale

Modele sezoniere precum SARIMA separă componentele sezoniere și tendințele, ajutând la realizarea prognozelor sezoniere.

10. Analiza Evenimentelor în Seriile Temporale

Aceasta presupune identificarea evenimentelor precum schimbările bruste sau anomaliile, folosind metode precum deviația standard și intervalele de încredere.

11. Metode Avansate de Analiză a Seriilor Temporale

Modele de regresie și analiza componentelor multiple (coerență și cointegrare) sunt utilizate pentru analiza avansată a seriilor temporale multidimensionale.

CURSUL 15 ETICA ȘI SECURITATEA DATELOR

1. Introducere în Etica și Securitatea Datelor

Etica și securitatea datelor se referă la protecția și gestionarea responsabilă a datelor personale. Aceste concepte sunt esențiale pentru respectarea confidențialității și integrității datelor și pentru construirea încrederii.

2. Principiile Etice în Prelucrarea și Utilizarea Datelor

Principiile etice includ respectarea confidențialității, consimțământul informat și transparența în procesarea și utilizarea datelor. Aceste principii sprijină dreptul la acces, rectificare și ștergere a datelor.

3. Protecția Datelor și Securitatea Informațiilor

Protecția datelor implică colectarea și stocarea securizată a datelor, utilizarea criptării și controlul accesului la date. Autentificarea utilizatorilor și protecția comunicațiilor sunt măsuri esențiale pentru asigurarea securității.

4. Gestionarea Riscurilor de Securitate

Gestionarea riscurilor implică evaluarea vulnerabilităților și implementarea măsurilor de securitate precum criptarea, autentificarea și monitorizarea activităților. Reziliența și recuperarea în caz de incidente sunt, de asemenea, elemente critice.

5. Reglementări și Standardele în Etica și Securitatea Datelor

Legislații precum GDPR și CCPA și standardele ISO 27001 și NIST Cybersecurity Framework sunt esențiale pentru conformitatea și securitatea datelor. Organizațiile trebuie să aibă politici clare și să asigure formarea continuă pentru respectarea reglementărilor.

6. Etica în Analiza și Interpretarea Datelor

Analiza corectă a datelor implică transparența metodelor și evitarea bias-ului și a distorsiunilor. Utilizarea responsabilă a algoritmilor și monitorizarea echității sunt priorități etice în interpretarea datelor.

7. Securitatea Datelor în Colectarea, Stocarea și Transferul Acestora

8. Etica și Securitatea Datelor în Domenii Specifice

1. Introducere în Inteligența Artificială în Știința Datelor

Inteligența Artificială (IA) în Știința Datelor combină tehnici de învățare automată, rețele neuronale, procesarea limbajului natural (NLP) și analiza imaginilor. Aplicațiile includ finanțele, sănătatea și retailul, unde explicabilitatea IA este esențială.

2. Ramurile Inteligenței Artificiale: - Învățare Automată (ML) - Învățare Profundă - Procesarea Limbajului Natural (NLP) - Viziune Computerizată - Robotică - Sisteme Expert - Planificare Automatizată - Sisteme de Recomandare
- IA Distribuită - Etica și Reglementarea IA

3. Ramurile Științei Datelor - Analiza Descriptivă - Analiza Predictivă - Analiza Prescriptivă - Explorarea Datelor (EDA) - Vizualizarea Datelor - Managementul Datelor - Învățare Automată și IA

4. Fundamentele Inteligenței Artificiale

Inteligența Artificială se bazează pe IA Limitată, Generală și Superioară, utilizând rețele neuronale și tehnici ML. Provocările includ confidențialitatea datelor, bias-ul algoritmic și interpretabilitatea.

5. Evoluția Inteligenței Artificiale

- Origini (1950-1960) - Perioada de declin (1960-1980) - Revigorarea (1980-2000) - Era modernă
- Viitorul IA: IA Generală și Superioară

6. Paradigme și Abordări în IA

Principalele abordări includ Învățarea Supravegheată, Nesupravegheată, prin Întărire și Învățarea Bazată pe Reguli.

7. Procesarea Limbajului Natural (NLP)

Tehnici NLP: tokenizarea, stemming, reprezentările Bag-of-Words, TF-IDF și modele avansate: Transformers (BERT, GPT).

8. Rețele Neuronale și Învățarea Automată RNA, cum ar fi MLP, CNN și RNN, modelează relațiile complexe dintre date.

9. Explicabilitatea Modelelor IA Tehnici precum LIME și SHAP sunt folosite pentru explicabilitate locală, iar mapările de saliență vizualizează activările rețelelor neuronale.