

# Analiza de Text

Capitolul 12 -  
Introducere și  
Concepte  
Fundamentale

# Introducere în Analiza de Text

- Procesul de extragere și înțelegere a informațiilor semantice și contextuale dintr-un text scris sau verbal.
- Scop: Obținerea unei înțelegeri profunde a textului și extragerea de informații utile.
- Aplicații: Identificarea entităților, analiza sentimentului, analiza tematică.



# Motivația și Importanța Analizei de Text

- Volumul uriaș de date: Textul este o sursă bogată de informații.
- Extracția informațiilor: Ajută la identificarea entităților, relațiilor și temelor.
- Îmbunătățirea deciziilor: Suport pentru decizii informate în diverse domenii.
- Monitorizarea opiniei publice: Evaluarea sentimentelor și opiniilor exprimate în texte.

# Domenii de Aplicare ale Analizei de Text

- Analiza social media: Monitorizarea sentimentelor și opiniilor în rețelele sociale.
- Servicii clienți: Gestionarea feedback-ului și îmbunătățirea experienței clienților.
- Monitorizarea presei: Analiza conținutului media pentru identificarea tendințelor.
- Monitorizarea și detectarea fraudei: Identificarea comportamentelor suspecte și a activităților ilegale.

# Procesul General al Analizei de Text

- Colectarea și curățarea datelor:  
Eliminarea elementelor nedorite din texte.
- Preprocesarea textului: Tokenizare, eliminarea stopwords, lematizare.
- Analiza semantică și gramaticală:  
Înțelegerea semnificației și structurii textului.
- Clasificarea și etichetarea: Atribuirea de etichete și categorii textului.
- Extracția de informații: Identificarea entităților și relațiilor relevante.

# Instrumente și Tehnologii Utilizate în Analiza de Text

- Biblioteci NLP: NLTK, SpaCy, TensorFlow.
- Algoritmi de analiză sentiment: Naive Bayes, SVM, Random Forests.
- Tehnici de extragere a informațiilor: Analiza de dependențe, expresii regulate.
- Platforme de analiză: IBM Watson, Google Cloud NLP, Microsoft Azure Text Analytics.

# Provocări și Limitări ale Analizei de Text

- Ambiguitatea și polisemia: Dificultatea de a înțelege sensurile multiple ale cuvintelor.
- Ironia și sarcasmul: Detectarea automată a tonului subtil.
- Volumul mare de date: Gestionarea și procesarea eficientă a datelor.
- Bias și subiectivitate: Influența seturilor de date și algoritmilor asupra rezultatelor.




# Preprocesarea Textului

- Curățarea textului: Eliminarea caracterelor speciale și a stopwords.
- Tokenizarea: Împărțirea textului în unități semantice.
- Lematizarea și stemming: Reducerea cuvintelor la forma lor de bază.
- Normalizarea: Conversia în litere mici, eliminarea spațiilor suplimentare.



A large orange circle on the left side of the slide, partially cut off by the edge.

# Analiza Sentimentului

- Definiția sentimentului: Atribuirea unei semnificații emoționale unui text.
  - Metode de evaluare: Dicționare de sentiment, machine learning, NLP.
  - Aplicabilitate: Monitorizarea opiniei publice, analiza feedback-ului clienților.
- 
- A blue dashed line in the bottom right corner, consisting of several curved segments.

# Analiza Tematică

- Definiție: Identificarea și înțelegerea temelor și subiectelor din texte.
- Metode: Analiza frecvenței cuvintelor, co-ocurența cuvintelor, topic modeling.
- Aplicații: Cercetare academică, analiza conținutului online, analiza discursului politic.

# Interpretarea și Vizualizarea Rezultatelor

- Distribuția cuvintelor-cheie: Identificarea celor mai semnificative cuvinte pentru fiecare temă.
- Harta de căldură: Vizualizarea importanței relative a cuvintelor în teme.
- Diagrame cu nori de cuvinte: Reprezentarea grafică a frecvenței cuvintelor în teme.



# Evaluarea Rezultatelor

- Coerența tematică: Măsura clarității și interpretabilității temelor.
- Acuratețea modelelor: Evaluarea performanței algoritmilor de învățare automată.
- Interpretarea subiectivă: Evaluarea temelor în funcție de expertiza umană.