

CAPITOLUL 3: PREPROCESAREA DATELOR

O privire
generală
asupra
procesului de
preprocesare a
datelor în
analiza datelor

3.1. IMPORTANȚA PREPROCESĂRII DATELOR

- Preprocesarea datelor este importantă pentru curățarea, transformarea și pregătirea datelor pentru modelare și analiză. Aceasta îmbunătățește performanța modelelor de învățare automată și precizia analizei.

3.1.1. CURĂȚAREA DATELOR

- ▶ Curățarea datelor presupune eliminarea valorilor eronate, incomplete sau duplicate din seturile de date pentru a asigura calitatea acestora.

3.1.2. COMPLETAREA VALORILOR LIPSĂ

- Completarea valorilor lipsă poate fi realizată prin tehnici precum interpolarea, utilizarea regresiei sau algoritmi de învățare automată.

3.1.3. NORMALIZAREA DATELOR

- Normalizarea datelor transformă valorile într-un interval comun, cum ar fi $[0,1]$, pentru a elimina influențele nedorite ale valorilor extreme asupra modelelor.

3.1.4.

TRANSFORMAREA DATELOR

- Transformarea datelor implică aplicarea de funcții, cum ar fi transformările logaritmice sau Box-Cox, pentru a ajusta distribuția valorilor și a facilita analiza.

3.1.5. AGREGAREA DATELOR

- ▶ Agregarea datelor presupune combinarea mai multor seturi de date pentru a oferi o viziune de ansamblu asupra informațiilor și pentru a simplifica analiza.

3.1.6. ELIMINAREA ZGOMOTULUI

- Eliminarea zgomotului din date constă în îndepărtarea informațiilor nerelevante sau eronate pentru a obține date mai curate și mai relevante pentru analiză.

3.3. CODIFICAREA CATEGORIILOR

- Codificarea categoriilor implică transformarea valorilor categorice în valori numerice, utilizând metode precum codificarea ordinală sau One-Hot.