

Raport de Analiză

January 12, 2024

1 Introducere

În cadrul acestei analize, am investigat și comparat diferiți algoritmi pentru a rezolva problema de clasificare a email-urilor spam în setul de date Ling-Spam [?]. Scopul nostru a fost să selectăm un algoritm adecvat, pe baza evaluării teoretice și experimentale, pentru a obține rezultate semnificative în rezolvarea acestei probleme.

2 Setul de Date

Am început analiza prin înțelegerea setului de date Ling-Spam, care conține email-uri clasificate ca fiind spam sau non-spam.

3 Alegerea Algoritmului

Am considerat algoritmul Naive Bayes ca un candidat potrivit, având în vedere următoarele aspecte teoretice:

- Eficiența sa în problemele de clasificare a textului.
- Simplitatea și eficiența computațională, făcându-l potrivit pentru seturile de date mari.

4 Comparativ cu Alți Algoritmi

Am evaluat performanța algoritmului Naive Bayes în comparație cu AdaBoost pentru clasificarea email-urilor spam. Cu toate că AdaBoost este cunoscut pentru adaptabilitatea sa, Naive Bayes a demonstrat o acuratețe remarcabilă în această problemă specifică. Implementarea simplă și ușor de înțeles a Naive Bayes a oferit rezultate competitive și a depășit performanța AdaBoost. Alegerea algoritmului depinde de natura specifică a setului de date și a caracteristicilor acestuia, iar în acest caz, Naive Bayes s-a dovedit a fi o opțiune robustă și eficientă. Este esențial să se evalueze performanța în contextul specific al problemei pentru a alege algoritmul potrivit.