

Curs 9

Abordarea unei probleme prin ML

Pasii pentru abordarea unei probleme din sfera Data Science sunt:

1. Determinarea tipului de rezultat dorit: explorare de date sau determinarea unui pattern in date?
2. Exista un scop particular vizat? de exemplu, analiza de asocieri, clasificare, estimarea valorii viitoare a unei serii de date, regresie, grupare de date, detectare de anomalii (outlier detection)
 - A. Se defineste o masura de succes: acuratetea de clasificare sau scor F1, mean squared error etc.
 - B. Exista suficiente date pentru a masura succesul aplicatiei?
 - C. Care este impactul asupra business-ului? marirea acuratetei cu 1% duce la recuperarea investitiilor de timp pentru determinarea unui model mai performant?
3. Achizitionarea de date
4. Curatarea lor, tratarea situatiilor anormale - valori lipsa, valori in contradictie cu domain knowledge
5. elaborarea modelelor, evaluarea lor
6. Decizie: e nevoie de mai multe date sau de alte modele sau de alte valori ale hiperparametrilor pentru modele?

Humans in the loop

Uneori, prezenta oamenilor este necesara in procese de DS, in parti oarecare ale fluxului de procesare. Daca insa un algoritm instruibil poate sa obtina o automatizare a unui pas, cu un grad mare de succes, atunci se permite unui om sa isi dedice capacitatea unei alte sarcini.

Exemplu: diagnosticare; acolo unde rezultatele sunt cu incertitudine mare, se poate apela la cunostintele unui expert.

De la prototip la productie

Testarea sistemului in productie

Aderare la framework-uri

Pasii urmasi