Inteligenta Artificiala

Laborator 11

**Clasificarea datelor de tip text**

Tema: Analiza sentimentelor

Considerati un set de date care contine recenzii despre filme (spre exemplu <https://ai.stanford.edu/~amaas/data/sentiment/> sau <https://www.kaggle.com/oumaimahourrane/imdb-reviews>).

[0.6] a. Invatarea reprezentarii vectoriale (constructia modelului word2vec)

- utilizati libraria gensim (python) pentru a construi modelul word2vec

<https://radimrehurek.com/gensim/models/word2vec.html#gensim.models.word2vec.Word2Vec>

- pentru o lista de 10 cuvinte, afisati cuvintele cele mai similare cu acestea

Optional: vizualizarea vectorilor de cuvinte (utilizati t-SNE pentru reducerea dimensionalitatii)

[0.4] b. Clasificarea datelor utilizand un algoritm de invatare automata

- utilizati libraria scikit learn pentru clasificare <https://scikit-learn.org/stable/>

Observatie: se pot folosi si alte librarii pentru constructia reprezentarilor vectoriale si pentru clasificare