ΘΕΩΡΙΑ

**Στόχος**: feature extraction με tf-idf και μετά evaluation με multi layer perception classifier MLP

Με βάση την βιβλιογραφία: 4476, [**https://arxiv.org/pdf/1802.05322.pdf**](https://arxiv.org/pdf/1802.05322.pdf)

**1ο Βήμα:** pre-processing the data

Source: <https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089>

* Removing the genres that is not needed (Family, history, music, tv movie, animation, documentary)
* Lower case
* Stop words
* Punctuation
* Apostrophe
* Stop words
* Single characters
* Convert numbers
* Punctuation and stop words
* Stemming

**2ο Βήμα:** tf-idf

* Split into training (70%) and test set (30%) the description and the corresponding genres (source: <https://arxiv.org/pdf/1802.05322.pdf> )
* Transform and fit into the training set

Transform into the test set

(source: <https://arxiv.org/pdf/1802.05322.pdf> )

code:

1.[Creating a Movie Reviews Classifier Using TF-IDF in Python (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2021/09/creating-a-movie-reviews-classifier-using-tf-idf-in-python/?) here he use the transform and the TfidfVectorizer like the source

2. <https://medium.com/geekculture/creating-content-based-movie-recommender-with-python-7f7d1b739c63> (not the same recommendation systems for the genres)

3. <https://github.com/AlexanderNixon/Machine-learning-reads/blob/master/Movie-content-based-recommender-using-tf-idf.ipynb> (not the same recommendation systems for the genres)

4. <http://www.sefidian.com/2022/02/28/understanding-tf-idf-with-python-example/>

(general code, with transform, two methods??)

Λογικά θέλουμε κάτι τέτοιο (αντί για movie, τα genres)

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

Από τα λινκ για κώδικα παίρνουμε κάτι τέτοιο

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

Table με τους συνδιασμους από τα combo με τα genres είναι αρκετα σημαντικό να γίνει