Notes DM 2

What? (Slides college)

* Information filtering system to predict user rating or preference of/for something
* Filtering information by determining relevance
* User based, item based, knowledge based(?) or user-item interaction based
* Collaborative filtering
  + Wisdom of the crowd
  + Based on preferences in the past 🡪 will be similar in the future
  + Output; rating of 1 item or top-N list of items
  + Find users like Alice who have rated the item and use average
    - By means of K-nearest neighbours:
      * Measure similarity between users with pearson correlation
      * Trial and error to find K
      * How to aggregate k rating values while avoiding a runtime that is too long?
  + Find items that are similar to the item and use Alice’s ratings for those
    - Use singular value decomposition (U and Vt enzo)
  + Problems:
    - How to rate new items and how to help new users? (“cold start problem”)
* Content-based recommendations
  + Use more information about the items
  + Learn user preferences
  + locate/recommend items that are "similar" to the user preferences
    - DICE coefficient
    - Use text (Piek’s lecture)

General:

* Look at distributions of variables
* Use different methods and their outcome as new columns and combine recommendations of methods.
* Use part of training set as validation set
* Use only part of training set to actually train?
* User based vs Item based
* Take running time into account

Susan Li @ towardsdatascience.com:

* Nothing seems to correlate linearly 🡪 which models are not suitable?
* Combinations of search destinations, hotel country, hotel market will definitely help finding the hotel cluster
* Only booking is of interest?
* Naïve Bayes, K-nearest neighbours classifier, random forest classifier, Multi-class logistic regression, SVM classifier (time consuming but good performance)

Ashkan Ebadi thesis:

* users clustering, matrix factorization module, and hybrid recommender system

Owen Zhang from previous competition:

* Ensemble of Gradient boosting machines
* Kijk in de slides
  + <https://www.dropbox.com/sh/5kedakjizgrog0y/_LE_DFCA7J/ICDM_2013?preview=3_owen.pdf>

Example/walk through:

<https://www.dataquest.io/blog/kaggle-tutorial/>

Files on collaborative filtering uit college zijn pdf’s in de map