Lab session 7

Machine Learning for Behavioral Data (CS-421)

April 14, 2021

Today

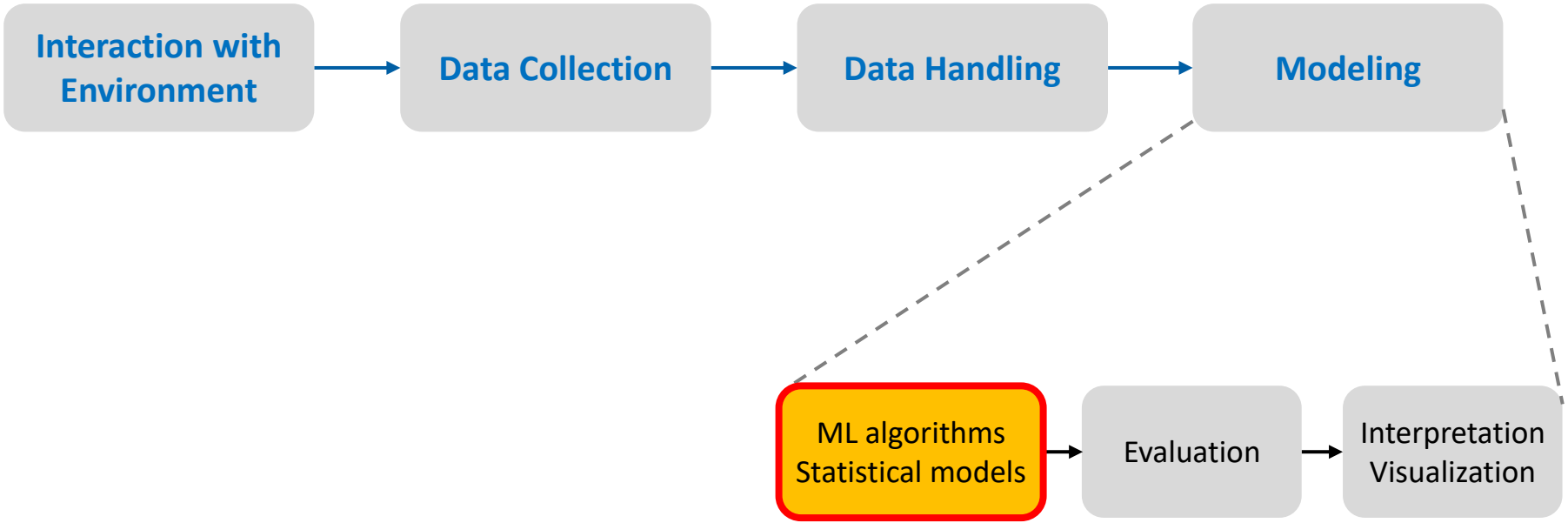
- **08:15 09:15** Tutorial on Recommender Systems Part I
- **09:15 09:25** SHORT BREAK
- **09:25 10:00** Homework and Project
  - Hands on and question time on HW4 solution
  - Question time on HW5
  - Question time on the class project
  - Upcoming events and tasks

Where we are

Week	Lecture	Lab Sessions	Projects
8	Recommender Systems	Tutorial + PO	M2: Research Questions and Exploratory Analysis
9	Neural Networks	Tutorial + PO	
10	Sequence Mining	Tutorial + PO	
11	Representation/ Feature Learning	Tutorial + PO	M3: Suggested Approach and Preliminary Results
12	Multimodal Analytics	Tutorial + PO	
13	Multimodal Analytics	Tutorial + PO	M4: Mature Approach and Results with Discussion
14	White Monday	PO	
15	Bias/Fairness		Project Presentations

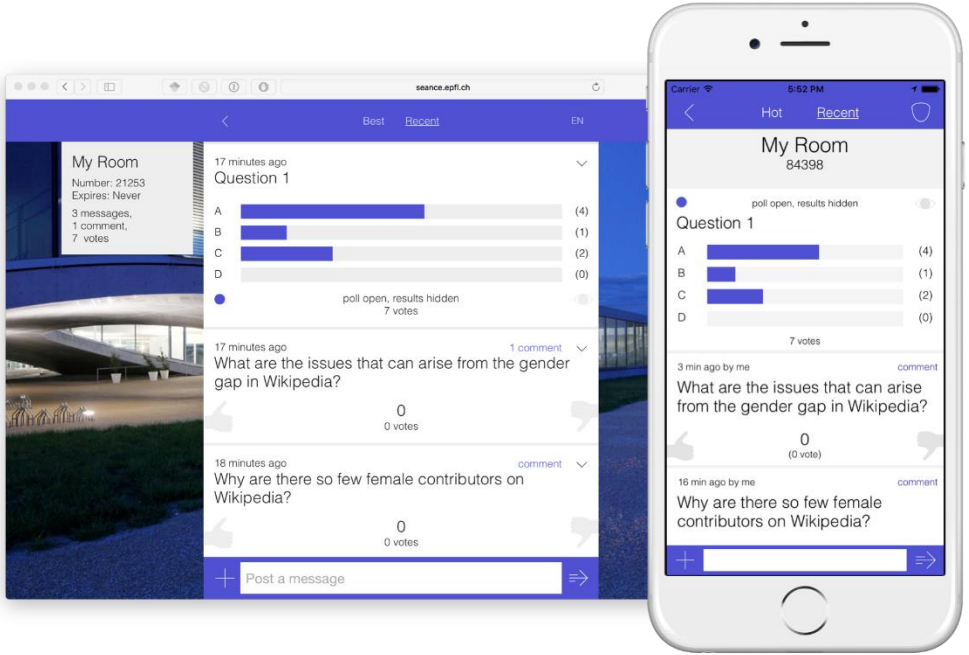
PO = project office hours

ML for Behavioral Data: Modeling



SpeakUp

- **Android / iOS:**
  http://speakup.info/
- **Web App:**
  https://web.speakup.info/
- **Room number:** 29315



Recommender systems theory

SpeakUp: How do you feel about recommender systems *theory*?

A: I have never heard of recommender systems.

B: I am **not confident at all** about recommender systems *theory*.

C: I am **slightly confident** about recommender systems *theory*.

D: I am **fairly confident** about recommender systems *theory*.

E: I am **very confident** about recommender systems *theory*.

Recommender systems in practice

SpeakUp: How do you feel about recommender systems *in practice*?

A: I have never heard of how to develop a recommender systems *in practice*.

B: I am **not confident at all** about developing recommender systems *in practice*.

C: I am **slightly confident** about developing recommender systems *in practice*.

D: I am **fairly confident** about developing recommender systems *in practice*.

E: I am **very confident** about developing recommender systems *in practice*.

User-based and Item-based CF

SpeakUp: How do you feel about user-based and item-based CF?

A: I have never heard of user-based and item-based CF.

B: I am **not confident at all** about user-based and item-based CF.

C: I am **slightly confident** about user-based and item-based CF.

D: I am **fairly confident** about user-based and item-based CF.

E: I am **very confident** about user-based and item-based CF.

Matrix factorization

SpeakUp: How do you feel about matrix factorization?

A: I have never heard of matrix factorization.

B: I am **not confident at all** about matrix factorization.

C: I am **slightly confident** about matrix factorization.

D: I am **fairly confident** about matrix factorization.

E: I am **very confident** about matrix factorization.

Glossary for recommendation

- **Item:** object to be recommended
- **Consumer or User**: person who receives the recommendations
- **Provider**: person or company who offers the item to be recommended
- **Feedback**: expression of preference
  - Explicit feedback (direct from the consumer)
  - Implicit feedback (inferred from consumer activity)
- **Prediction**: estimate of consumer-item preference
- **Recommendation**: ordered list of selected items for consumer
- **Attribute**: a descriptive characteristic of a consumer or item

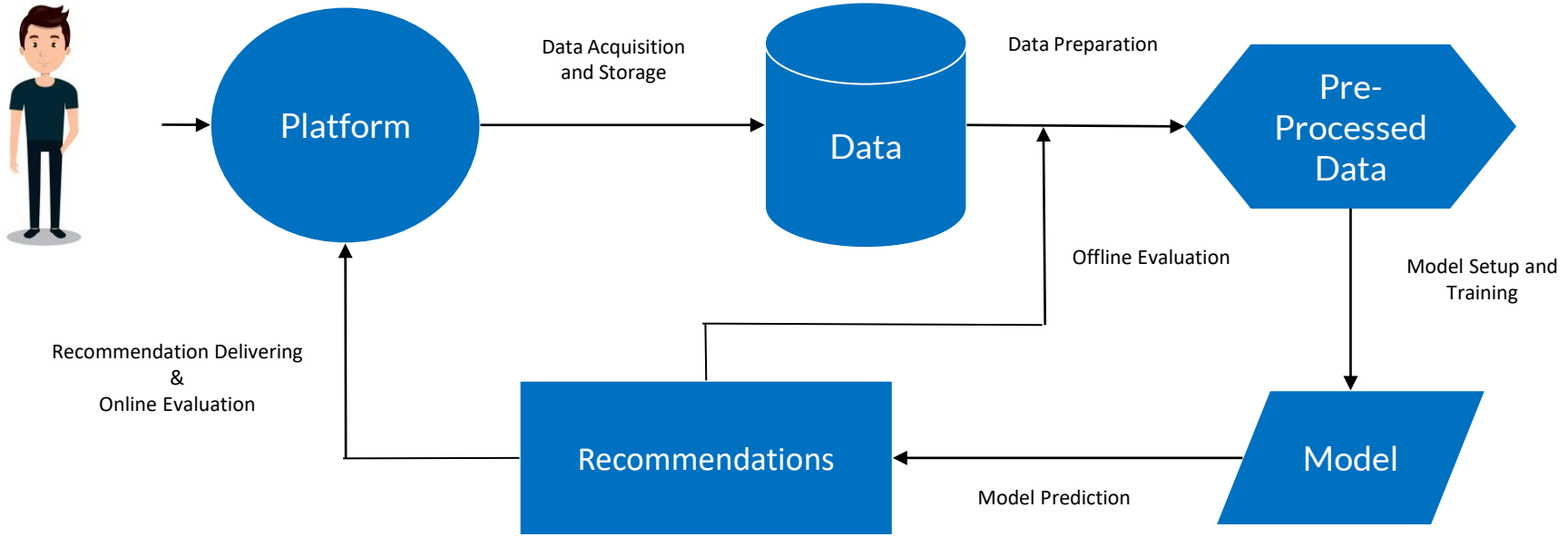
Recommendation task formalization

- Given:
  - a set of **consumers** $C = \{c_1, c_2, ..., c_M\}$
  - a set of **items** $I = \{i_1, i_2, ..., i_N\}$
- Let $R \subseteq R^{M \times N}$ be the **consumer-item feedback matrix**:
  - $R(c,i) \geq 0$ if consumer c expressed interest in item i
  - $R(c,i) = 0$ otherwise
- The objective is to predict **unobserved consumer-item feedback** $R(c,i) = f(c,i \mid \vartheta)$ in R:
  - ϑ denotes model parameters
  - f denotes the function that maps model parameters to the predicted relevance
- Given a consumer c, items not rated by c are ranked by decreasing relevance:

$$i^* = \arg\max f(c,j \mid \vartheta)$$

$$j \in I \setminus I_c$$

Recommendation pipeline



Taxonomy of preference modes

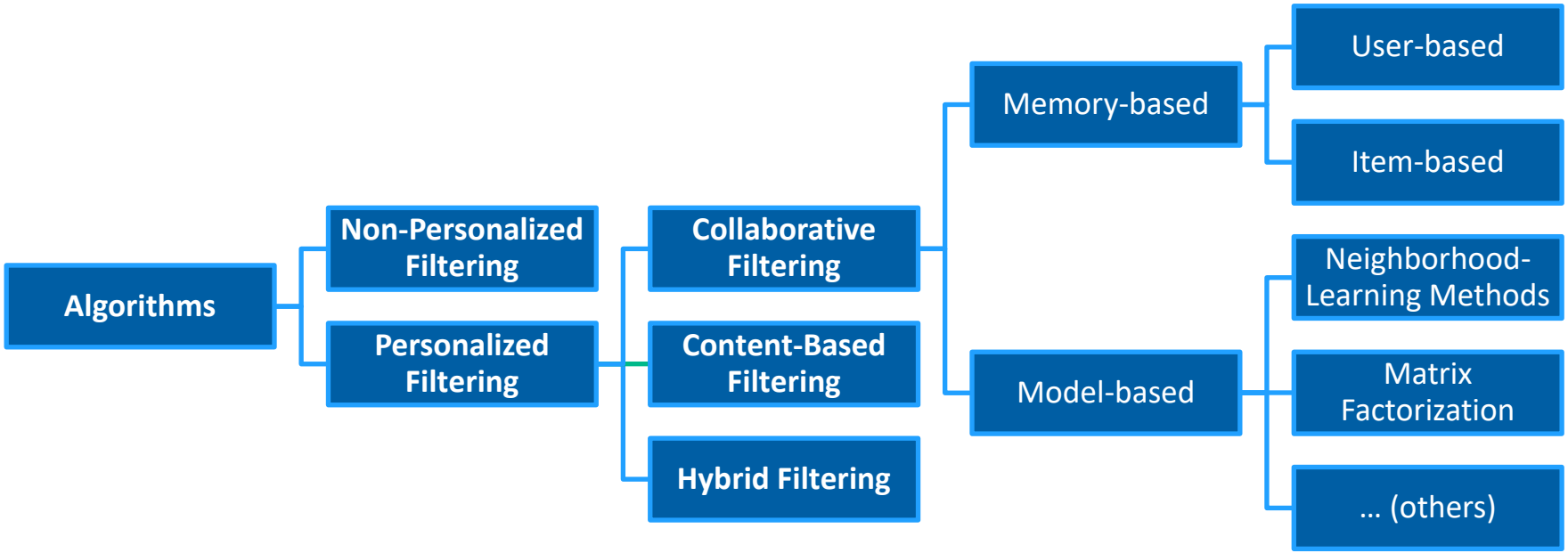
Explicit

- Rating
- Review
- Vote
- Like
- …

Implicit

- Click
- Purchase
- Follow
- Wish
- …

Taxonomy of recommendation algorithms



Tutorial 7 Hands on

Feel free to use Noto or your own environment:

- Go to https://noto.epfl.ch/.
- Login with your GASPAR.
- If you have **NOT** already cloned the repository:
  - Go to Git → Clone → https://github.com/d-vet-ml4ed/mlbd.
- Go to Git → Pull.
- Go through Tutorials/Tutorial07/Recommender_Systems_PartI.ipynb.

Questions?

Extensive debriefing on homework 4

Homework 04: KNN for Churn Prediction

Introduction

This week, we will go more deeply into the churn prediction task introduced in the previous homework. As you already know, churn prediction is one of the most popular use cases of machine learning in business, consisting of detecting customers who are likely to cancel a subscription to a service. Churn can be triggered by better price offers, more interesting packages, bad service experiences or personal situation changes experienced by customers. To timely prevent customers' churn, companies might adopt a machine learning classifier able to predict churn on an individual customer basis, and then use the predictions of this classifier to know when countermeasures (e.g., discounts or special offers) against churn are needed, to prevent the churn event.

In the previous homework, you were expected to use a range of demographic and basic contract-related features and create five additional behavioral as well as an appropriate evaluation method and performance metric(s) to assess the goodness of your classifier in predicting churn for customers.

In this homework, we ask you again to work on the machine-learning pipeline presented at the end of lecture 4, by instantiating, exploring, and fine-tuning a series of k-Nearest-Neighbors (kNNs) classifiers. Specifically, we will ask you to:

- Experiment with distance matrices and KNN classifiers fed with only demographic features or only behavioral features.
- Experiment with distance matrices and KNN classifiers fed with both demographic and behavioral features.
- Report, visually, the performance of three fine-tuned KNN classifiers (demographic, tuned), (behavioral, tuned), and (combined, tuned), and discuss.

The focus of this homework is on modelling and not on the quality of features. Therefore, note that there should be **no** need to perform any feature engineering in this homework. You are expected to re-use the features (and possibly a part of the source code) that you have created in homework 3. If you have not submitted homework 3, please do get in touch with the TAs, and they will provide five example behavioral feature to use in this homework.

Submission

The homework is due **Mar 30, 2021 23:59 CET**. The notebook must be pushed to your GitHub classroom repository.

If you have any questions, feel free to use the Q&A forum in Moodle.

Some Instructions

The complete solution of this homework has been uploaded on the public GitHub repository at:

https://github.com/d-vet-ml4ed/mlbd/blob/main/Homework/Homework04/Homework04-KNNChurnPrediction-Example-Solution.ipynb

Questions?

Class project milestones

- **M01** on Preferences on Tracks and Group Members
  - due March 23, 2021 23:59 CET - **MANDATORY**
- **M02** on Research Questions and Exploratory Analysis
  - due April 13, 2021 23:59 CET - optional
- **M03** on Implemented Approach and Preliminary Results
  - due May 04, 2021 23:59 CET - optional
- **M04** on Mature Approach and Results with Discussion
  - due May 18, 2021 23:59 CET – optional
- **Project Presentation** for Course Evaluation
  - to be given on May 31, 2021 - **MANDATORY**
- **Final Project Deliverable** for Course Evaluation
  - due June 11, 2021 23:59 CET - **MANDATORY**

Your current idea for the project

SpeakUp: Which type of machine-learning approach are you planning to apply?

A: Traditional supervised learning

B: Traditional unsupervised learning

C: Knowledge tracing

D: Recommendation

E: Others

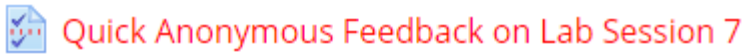
Questions?

Important upcoming dates

- Apr 16, 2021 15:00 16:00
  - Office Hour
- By Apr 16, 2021 23:59 CET
  - Submission Deadline for Homework #5
- Apr 19, 2021 15:15 17:00
  - Lecture #8
- Apr 21, 2021 08:15 10:00
  - Lab Session #8

Your feedback is essential

- It is a new course, please give feedback on how to improve it.
- Short anonymous feedback forms on Moodle.



https://moodle.epfl.ch/mod/questionnaire/view.php?id=1145589