Advanced TM Project: Part I

As part of the Advanced TM course, you will do a group project in which you apply some of the techniques learned in the lecture. The project will have 4 parts:

- Part I: Finding the dataset and defining the question
- Part II: Developing the proof of concept
- Part III: Writing a short paper
- Part IV: Turning the paper into a poster

In the last lecture on the 14.5 we will have a poster session, during which you will present the outcome of your work to your fellow students.

Dates

We will meet 4 times in the tutoring sessions in BIN 1.D.29 at 1015, and there will be 4 deadlines for the individual project parts:

date	tutoring session	deadline
Friday, 21.2	kick-off part I	
Friday, 6.3		part I: project plan
Friday 13.3	kick off part II	
Friday 3.4	kick off part III	
Friday 10.4		part II: proof of concept
Friday 24.4	kick off part IV	
Friday 1.5		part III: paper
Friday 8.5		part IV: poster

Grading

The exam is worth 60 points, and the project is worth 40 points, which are distributed as follows on the individual parts:

part	points
I	10
II	15
III	10
IV	5

Part I

For every group, submit by email to colic@ifi.uzh.ch a document called `color_1.pdf` (according to your team color) and a BiBTeX document called `color_1.bib` by Friday, 6.3, at 1015 AM containing the following sections. For all sections, indicate the word count of the section at the end of the section in parentheses (57).

- Find a dataset that you would like to investigate using TM techniques. This dataset must be publicly available, and should be sufficiently large to lend itself to this project. When in doubt, consult with the tutor. For example, this could be a subset of PubMed[1] or you can check this page[2] if you don't have any good ideas. Make sure you can actually download the dataset. Describe your dataset in 90-110 words.
- Formulate a research question that you would like to answer using TM techniques. This section should include a concise question that could be the title of your paper in part III.
- Include a list of deliverables that will be the outcome of your research. While we will look at your code, the deliverables are *not* your code, but rather transformations and condensations of the data.
    - For example, you could run an ER tool on PubMed and perform a statistical correlation of diseases. In this case, a document listing the correlations and their frequency would be a good deliverable.
    - For example, given a dataset of doctor appointment dates, you could try to predict no-shows. In this case, a model and feature analysis would be a good deliverable.
- Choose 3 papers related to your research question and summarise them in a related work section that you can later use in the paper. Submit the references to the papers as

[1] https://www.ncbi.nlm.nih.gov/pubmed/
[2] https://course.fast.ai/datasets

a BiBTeX document, and include the related work section of 250-350 words in the PDF submission.

→ Explain BiBTeX
→ Related Work explanation