

Course assignment

Deadline Friday 23rd August 9.00 am. Please read the instructions carefully. Incompliance may cause your work to be rejected.

The task is to make a plan how controlled experimentation could be used to improve a web service. You can work in pairs providing that the members of the pair represent different nationalities. If you wish, you may also work alone. The assignment has the following steps:

1. Read the articles
 - Ron Kohavi, Roger Longbotham, Dan Sommerfield, Randal M. Henne (2009) Controlled experiments on the web: survey and practical guide, *Data Mining and Knowledge Discovery*, Volume 18, Issue 1, pp 140–181. Available at <http://ai.stanford.edu/~ronnyk/2009controlledExperimentsOnTheWebSurvey.pdf>.
 - Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, Ya Xu (2012) Trustworthy Online Controlled Experiments: Five Puzzling Outcomes Explained, The 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Available at <http://www.exp-platform.com/Documents/puzzlingOutcomesInControlledExperiments.pdf>
2. Select a web service / company for your assignment. Any service where you can apply controlled experimentation explained in the articles will do but companies mentioned in the articles cannot be used. Familiarize yourself with the web page and the business logic of the company.
3. Make a plan how controlled experimentation could be applied in the case of the chosen web service.
4. Create a short (slides for 5 minutes talk) business presentation on your plan. This should be a sales pitch that concentrates on the benefits of proposal to the company. The intended audience is the executives of the company. Name your pdf file with the names of the authors.
5. Write a report (a white paper in a pdf format) where you summarize the articles, introduce the web service you have chosen and present

your plan how controlled experimentation could be applied in the case of the chosen web service. The intended audience is data scientists. The title page of the report should include:

- The title
- The names, the nationalities and the email addresses of the authors
- The name and the URL of the webpage / company considered
- A self-grading on the scale: 0 (fail), 1 (sufficient), 2 (satisfactory), 3 (good), 4 (very good), 5 (excellent)
- A summary of max 200 words

Name your pdf file with the names of the authors.

6. Send your report and business presentation in the pdf format to `juha.t.karvanen@jyu.fi` by Friday 23rd August 9.00 am. The title of the email should contain the code STAT1 followed by the family names of the authors. You will get a receipt after your email has been received.
7. No extensions for the deadline will be granted. Tight deadlines are an essential part of the every day life of a data scientist.
8. (optinal) Participate seminar on Friday 23rd August 10.15 – 11.45 am and prepare to give your business presentation.