30E03000 – Data Science for Business I (6cr)

Syllabus

Version 4.3 (10.01.2020)

Instructors' contact information	Course information
Names: Pekka Malo **Teaching Assistant(s):** Philipp Back, Lina Siltala-Li **E-mail:** firstname.surname@aalto.fi **Office:** T203 **Instructors' webpages:** https://people.aalto.fi/pekka_malo https://people.aalto.fi/philipp.back	**Status of the course:** Advanced Studies in Master's degree program in Information and Service Management (DR2013); Business application course in the Aalto level minor on Analytics and Data Science **Academic Year, Period:** 2019-2020, Period III **Location:** Ekonominaukio 1, U006 (R038) **Language of Instruction:** English **Course Website:** https://mycourses.aalto.fi/course/view.php?id=23730

1. Overview

"Data Science for Business – What You Need to Know About Data Mining and Data Analytic Thinking"

The objective of the course is to provide an introduction to practical data science from the perspective of business analysts. The course takes a problem-based approach for teaching data-analytic thinking and fundamental concepts/tools that are needed for data-driven decision making in business. As such, the course is not a replacement for algorithm-centered courses that give deeper insights to the data mining techniques.

The course will discuss the following topics:

- **Fundamentals of Predictive Analytics** – Topics covered include basics of predictive modeling, classification, shopping-basket analysis, evaluation of models, expected value framework, problem of over-fitting and its avoidance. Application examples range from modeling marketing responses to prediction of credit risks, and mining of transaction datasets.
- **Data Science Tools for Business Analysts** – A technical objective of the course is to introduce python programming for business analysts. The module does not require prior knowledge in python but benefits from prior experience in scripting/programming. In addition to fundamentals of data analysis with python, the module offers an opportunity to experiment with more advanced techniques such as solving of high-dimensional regression problems via regularization techniques. Given the diverse background of course participants, parts of the module are designed to be suitable for self-study to allow progress at different paces and at varying levels of challenge. Other tools (including commercial platforms) may be introduced during the visiting lectures from different companies.

2. Target group and prerequisites

The course is intended for participants with diverse backgrounds:

- Business analysts and developers who will be implementing and evaluating data science solutions;
- Aspiring future data scientists;
- Business people who will be working with data scientists, managing data science-oriented projects, or investing in data-driven ventures.

The course has a strong focus on empirical assignments, which require prior knowledge in statistics and basic skills in programming/scripting (or at least willingness to learn). However, more theoretical or mathematical aspects of data analytics are beyond the scope of the course.

3. Learning outcomes

After completing the course, the students will be able to

- identify the role of data as a business asset;
- understand the principles of predictive modeling;
- recognize how different data science methods can support business decision-making;
- learn basic data analytic techniques for solving business problems;
- understand the promises and limitations of big data;
- gain some experience in using data analytic tools that are widely used in companies.

4. Assessment, assignments, and grading

The course assessment is comprised of the following parts:

- Week 1 assignment (A1): DataCamp courses (2) 5%
- Week 2 assignment (A2): MyCourses quiz on theory and Credit Risk coding task 10%
- Week 3 assignment (A3): DataCamp course (1, tbd) 5%
- Week 4 assignment (A4): MyCourses quiz on theory 10%
- Week (1-) 6: Course project (**team case**; cannot be done individually; no exceptions) 40%
- Exam in computer lab 30%

All assignments must be completed to pass the course, and late assignments will not be accepted.

Note that the starting level of the student teams will be taken into account in the grading of the team case, and thus special attention is paid to the teams' development in knowledge sharing and learning.

Additional details and instructions about this will be provided during the course.

5. Readings

Course book:

- Provost, F. and Fawcett, T. (2013) "Data Science for Business: What you need to know about data mining and data-analytic thinking." O'Reilly Media, 1st Edition.

Complementary reading: In addition to learning python, it maybe helpful to look also other popular tools such as R. Therefore, the following list will include references covering different types of tools that can be used in a complementary fashion.

- Grus, J. (2015): "Data Science from Scratch". O'Reilly
- Müller, A. and Guido, S. (2017): "Introduction to Machine Learning with Python: A Guide for Data Scientists". O'Reilly.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) "An Introduction to Statistical Learning: with Applications in R." Springer Texts in Statistics.
- Miller, T. (2013) "Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R." FT Press, 1st Edition.
- Rajaraman, A. and Ullman, J. (2011) "Mining of Massive Datasets." Cambridge University Press.

6. Preliminary schedule

Week	Dates	Topic
1	**7.1. – 8.1.2020** **Tue 15:15-18:00** **Wed 15:15-18:00** **Session legend:** **Compulsory** **Not compulsory** **Extra if needed**	**Introduction to Predictive Analytics** Lecture, Tuesday 7.1.2020: • Fundamental concepts and definitions • Data mining as a process (CRISP-DM) • Data mining tasks and popular algorithms • Supervised vs. unsupervised learning Theory to practice: Tutorial 1, Wednesday 8.1.2020: • Decision trees (theory by Prof. Malo) • Insurance fraud detection with decision trees in Python • Classification as a predictive modeling task

2	**13.1. – 15.1.2020**
	Mon 15:15-18:00
	Tue 15:15-18:00
	Wed 15:15-18:00
		Data-Driven Decision-Making
		Lecture, Monday 13.1.2020:
		• Measuring value from predictive analytics
		  ◦ Expected value framework for model building
		  ◦ Visualizing model performance
		Q&A session, Tuesday 14.1.2020
		Theory to practice: Tutorial 2, Wednesday 15.1.2020
		• Support Vector Machines (theory by Prof. Malo)
		• Python for predicting customer responses to marketing campaigns
		• Model evaluation
		• Balanced vs. unbalanced datasets in classification

Week	Dates	Topic

3	**20.1. – 22.1.2020**
	Mon 15:15-18:00
	Tue 15:15-18:00
	Wed 15:15-18:00
		Pattern Mining and Shopping Basket Analysis
		Lecture, Monday 20.1.2020:
		• Reaktor (guest lecture)
		• What is pattern recognition?
		• Association rule mining as a predictive modeling task:
		  ◦ Shopping basket analysis
		  ◦ Apriori algorithm
		Q&A session, Tuesday 21.1.2020
		Theory to practice: Tutorial 3, Wednesday 22.1.2020:
		• Python for shopping basket analysis a.k.a. "who buys what?"

4	**27.1. – 29.1.2020**
	Mon 15:15-18:00
	Tue 15:15-18:00
	Wed 15:15-18:00
		Feature selection and model building
		Lecture, Monday 27.1.2020:
		• Linear regression modeling
		• Forward / backward feature selection
		• Regularization techniques for feature selection
		  ◦ Lasso regression
		  ◦ Ridge regression
		• Cross-validation and model evalution
		Q&A session, Tuesday 28.1.2020
		Theory to practice: Tutorial 4, Wednesday 29.1.2020
		• Feature selection for linear regression modeling

5	**3.2. – 4.2.2020**
	Mon 15:15-18:00
	Tue 15:15-18:00
		Visiting lecture and time for project work
		Lecture, Monday 3.2.2020:
		• IBM (guest lecture & demos by Jukka Ruponen)
		  ◦ Data science in hybrid/multiclouds (business value & how to)
		  ◦ Information architecture & data governance
		  ◦ Multi-modal AI platforms for business
		  ◦ AI ethics (model bias, transparency & explainability)
		Q&A session on project work, Tuesday 4.2.2020
		(No session on Wednesday 5.2.2020!)

6	**10.2. – 12.2.2020** **Mon 15:15-18:00** **Tue 15:15-18:00** **Wed 15:15-18:00**	**Team Case Presentations** Team Case Presentations - session 1, Monday 10.2.2020 Team Case Presentations - session 2, Tuesday 11.2.2020 Q&A session on course and exam, Wednesday 12.2.2020

7. Course workload

The following is a tentative breakdown of the workload of the course. All hours are academic hours.

Contact sessions	
• Lectures and tutorials (1-2 × 3h / week)	18h
• Exercise demos and workshops (2 × 3h / week)	36h
Class preparation	12h
Assignments	48h
Course project	46h
Total	**160h (6op)**

8. Ethical rules

Aalto University Code of Academic Integrity and Handling Thereof:
https://into.aalto.fi/pages/viewpage.action?pageId=3772443

9. Student selection

The maximum number of seats on this course is 60 (not 50 as indicated in Oodi). Note that Oodi allows more than 60 people to register, so a successful registration does not equal a spot on the course (hence no "first come first serve"). Instead, the selection is made based on the course pre-survey. All registered students will receive a MyCourses notification when the link to the pre-survey is available. The survey has to be answered by **December 31st 2019**. The priority for the student selection is as follows:

1. Aalto ISM MSc students whose specialization area is Business Analytics
2. Aalto Analytics and Data Science minor students & Students in Master's Programme in ICT Innovation (EIT digital)
3. Other Aalto MSc students (incl. ISM students without specialisation in Business Analytics)

10. Attendance

Updated (10.1.2020): No mandatory attedance.

11. Other issues

- Students will be divided into working teams by the teachers in charge; preference stated in the pre-course survey will be taken into account **if possible**.
- Evaluation rubrics will be available in MyCourses