

30E03500 – Data Science for Business II (6cr)

Syllabus

Version 2022.A.3 (21.10.2022)

Instructors' contact information	Course information
<p>Teacher: Pekka Malo</p> <p>Teaching Assistants: Iaroslav Kriuchkov, Lina Siltala-Li</p> <p>E-mail: firstname.surname@aalto.fi</p> <p>Office: T203, BIZ</p> <p>Instructor's webpages: https://people.aalto.fi/pekka_malo</p>	<p>Status of the course: Advanced Studies in Master's degree program in Information and Service Management. Business application course in the Aalto level module on Analytics and Data Science.</p> <p>Academic Year, Period: 2022-2023, Period II</p> <p>Location: Almost fully online (see details in MyCourses)</p> <p>Language of Instruction: English</p> <p>Course Website: https://mycourses.aalto.fi/course/view.php?id=36983</p>

1. Overview

"From Data to Decisions – What You Need to Know About Data-driven Decision Making"

The course presents further topics in predictive modeling (e.g. quantile regression, survival analysis, text analytics, and deep learning) and provides a brief introduction to prescriptive analytics in the form of linear programming models. In terms of content, **the course consists of five modules**. Each of the modules lasts for 1 week. The topics covered in the modules vary on a yearly basis. The material will involve programming assignments with practical applications. Python is used for the examples as well as the assignments. For Autumn 2022 implementation Google Colab will be used as the coding environment.

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.

As a prerequisite for enrolling in this course, we assume that you have completed 30E03000 Data Science for Business I (or a similar course). The course has a strong focus on empirical assignments, which requires prior knowledge in the following topics:

- Fundamentals in statistics (e.g., 30A02000 or preferably 30C00600/ISM-C1100)
- Knowledge in business mathematics; in particular, linear algebra and optimization (e.g., 30A03000 or preferably 30C00355)
- Basic skills in programming/scripting (teaching and assignment templates are in Python)

3. Learning outcomes

After completing the course, the students will

- understand the importance of prescriptive analytics in business decision-making,
- be able to apply basic optimization techniques to solve prescriptive analytics problems,
- be able to apply modern regression techniques,
- be able to implement (program) their solutions with suitable software,
- be able to apply basic text analytics tools, and understand fundamental concepts in natural language processing,
- know the basic concepts of deep learning, and how neural networks can be applied in predictive modeling,

4. Assessment, assignments, and grading

The course assessment is comprised of the assignments of 5 modules.

Grading will be based on the best 4 completed. Each assignment has equal weight in the final grade.

New assignments will be available every Monday morning. Deadlines for the assignments are given in the table below. Late submission is allowed: if you submit **0-12 hours** after the deadline, **20% of points** will be deducted; if you submit **12-24 hours** after the deadline, **40% of points** will be deducted; if you submit **24-36 hours** after the deadline, **60% of points** will be deducted; if you submit **36-48 hours** after the deadline, **80% of points** will be deducted.

Assignment 3, Assignment 4 and Assignment 5 (the report part) can be done in groups of max. 3 students. In your submission, you will have to mention the names of all group members and contribution of each student. Higher quality of work will be expected from group work, thus grading of group submissions will be more strict.

NOTE: DataCamp certification(s) maybe required as one of the assessments in some modules, further instruction will be given in MyCourses. Free DataCamp access will be provided to all enrolled students for the period of 6 months from the beginning of the course.

- **Assignments deadlines**

Module 1: Introduction to prescriptive modelling	
Assignment 1 (max. 25 points)	DL 02.11.2022 23:59
Module 2: Convex quantile regression	
Assignment 2 (max. 25 points)	DL 09.11.2022 23:59
Module 3: Natural Language Processing	
Assignment 3 (max. 25 points)	DL 16.11.2022 23:59
Module 4: Predictive tools for health care analytics	
Assignment 4 (max. 25 points)	DL 23.11.2022 23:59
Module 5: Introduction to Deep Learning with Neural Networks	
DataCamp certification, (max. 10 points)	DL 30.11.2022 23:59
Report on Deep Learning (max. 15 points)	DL 30.11.2022 23:59

- **Grading rule**

5. Readings

All materials needed for completing the course are provided during the lectures, tutorials and Q&A sessions. The following list of complementary reading is recommended but not required.

5	[85%, 100%] of the total points
4	[75%, 85%)
3	[65%, 75%)
2	[55%, 65%)
1	[35%, 55%)

Complementary readings:

- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) “An Introduction to Statistical Learning: with Applications in R.” Springer Texts in Statistics.
- Anderson, D., Sweeney, D., and Williams, T. (2000) “An Introduction to Management Science: Quantitative Approaches to Decision Making.”

6. Schedule

All lectures and tutorials will be pre-recorded and uploaded on Monday each week. Recordings will be available till the end of the course. Live Q&A session will be held in Zoom on Tue - keep an eye on the schedule and check the MyCourses page of the respective module. [There will be one on-campus event: the introduction meeting, Tue, 25.10.2022, 16.30-18.30, Room T2 - C105, Computer Science building, Konemiehentie 2](#)

NOTE: In the table below: M - Module, A - Assignment

Week	Dates	Topic
1	24.10. – 30.10.2022 Tue 16:30-18:30 Room T2-C105 Konemiehentie 2	<p>(M1) Introduction to prescriptive modeling</p> <p>Mon 24.10 — Recordings and material upload:</p> <ul style="list-style-type: none"> • (M1 Lecture) Optimization for decision support (Pekka) • (M1 Lecture) Role of data in decision analytics (Pekka) • (M1 Tutorial) Tutorial session: Optimisation in Python and Introduction to (A1) (Iaroslav) • (A1) Materials available - DL 02.11 23:59, Late DL 04.11 23:59 <p>Tue 25.10 — Introduction meeting and Q&A (M1)</p> <ul style="list-style-type: none"> • Course introduction meeting (Pekka & Iaroslav & Lina) • (M1) Q&A session for the lecture and the tutorial (Pekka & Iaroslav)

Week	Dates	Topic
2	31.10. – 06.11.2022 Tue 16:30-18:30 Zoom	<p>(M2) Convex Quantile Regression</p> <p>Mon 31.10 — Recordings and material uploading:</p> <ul style="list-style-type: none"> • (M2 Lecture) Convex Quantile Regression (Iaroslav) • (M2 Tutorial) Convex Quantile Regression (Iaroslav) • (A2) Materials available - DL 09.11 23:59, Late DL 11.11 23:59 <p>Tue 01.11 — Live Q&A on (M1, M2)</p> <ul style="list-style-type: none"> • (M1, M2) Live Q&A session for the lecture and the tutorial (Pekka & Iaroslav)
3	07.11 – 13.11.2022 Tue 16:30-18:30 Zoom	<p>(M3) Natural Language Processing</p> <p>Mon 07.11 — recordings and material uploading:</p> <ul style="list-style-type: none"> • (M3 Lecture) Fundamentals of NLP (Pekka) • (M3 Lecture) Introduction to sentiment analysis in finance (Pekka) • (M3 Tutorial) Tutorial on Text analytics and intro to (A4) (Iaroslav) • (A3) Materials available - DL 16.11 23:59, Late DL 18.11 23:59 <p>Tue 08.11 — Live Q&A on (M2, M3)</p> <ul style="list-style-type: none"> • (M2, M3) Live Q&A session for the lecture and the tutorial (Pekka, Iaroslav)
4	14.11. – 20.11.2022 Tue 16:30-18:30 Zoom	<p>(M4) Predictive tools for health care analytics</p> <p>Mon 14.11 — recordings and material uploading:</p> <ul style="list-style-type: none"> • (M4 Lecture) Modeling time to event risks (Pekka) • (M4 Lecture) Introduction to survival analysis (Pekka) • (M4 Tutorial) Tutorial on survival analysis and introduction to (A4) (Lina) • (A4) Materials available - DL 23.11 23:59, Late DL 25.11 23:59 <p>Tue 15.11 — Live Q&A on (M4)</p> <ul style="list-style-type: none"> • (M4) Live Q&A session for the lecture and the tutorial (Pekka & Lina)

Week	Dates	Topic
5	21.11. – 27.11.2022 Tue 16:30-18:30 Zoom	<p>(M5) Introduction to Deep Learning with Neural Networks</p> <p>Mon 21.11 — recordings and material uploading:</p> <ul style="list-style-type: none"> • (M5 Lecture) Learning to use neural networks in predictive modeling (Pekka) • (M5 Tutorial) Further topics on deep learning (example: deep survival analysis) (Lina) • (A5) Introduction to quiz on deep learning (Lina) - DL 30.11 23:59, Late DL 02.12 23:59 • (A5) Datacamp self-study - DL 30.11 23:59, Late DL 02.12 23:59 <p>Tue 22.11 — Live Q&A on (M4, M5)</p> <ul style="list-style-type: none"> • (M5) Live Q&A session for the lecture and the tutorial (Pekka & Lina)
6	27.11 – 30.11.2022 Tue 16:30-18:30 Zoom	<p>Tue 28.11 — Live Q&A on (M5) and final meeting</p> <ul style="list-style-type: none"> • (M5) Live Q&A session for the lecture and the tutorial (Pekka & Lina) • Final meeting - Q&A (Pekka & Iaroslav & Lina)

7. Course workload

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

Lectures and tutorials	18h
Assignment sessions	36h
Class preparation	12h
Assignments (independent work)	94h
Total	160h (6 op)

8. Ethical rules

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

9. Other issues

- Registration to the course via SISU
- Evaluation rubrics will be available in MyCourses