Syllabus

CEU Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

Data Science 2: Machine Learning Tools

Instructor	János Divényi (divenyij@ceu.edu)
Office hours	by appointment
Credits	2 US credits (4 ECTS credits)
Term	Winter 2023-2024
Course level	Master's
Prerequisites	Data Analysis 1-3, Data Science 1, Data Engineering 2
Course drop	Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

1. COURSE DESCRIPTION

This course aims to give a compass to the student on the quickly evolving field of Machine Learning. It builds on the basic concepts introduced in previous courses (like Data Analysis 3 and Data Science 1) and showcases state-of-the-art algorithms used in the industry. The goal is to understand which tool could help solving which type of problems rather than achieving proficiency in one specific tool. A large part of the course will be dedicated to getting hands-on experience using R supported by shorter intuition-based sessions covering the required theoretical background.

2. LEARNING OUTCOMES

Key outcomes. By the end of the course, students will be able to…

- Understand that ML methods are tools that can solve various business problems.
- Find the appropriate tool (supervised vs unsupervised vs bandit vs causal) for a given problem.
- Use the given tool at a basic level to address business problems.
- Evaluate the performance and limitations of such solutions.
- Deepen their knowledge in the given problem space on their own by building on the fundamentals learned in the course.

Other outcomes. The course will also help develop skills in the following areas:

Learning Area	Learning Outcome
Critical Thinking	Understand the steps of the quantitative research process, starting with business understanding, benchmark selection, incremental modeling, evaluation and final model assessment. Be aware of the limitations of the used solutions.
Quantitative Reasoning	Apply mathematical and statistical knowledge to plan and evaluate machine learning models.
Technology Skills	Implement solutions using the R language in an easily transferable way.
Interpersonal Communication Skills	Present data as evidence and explain methodology so that it is digestible for a non-technical audience as well.

3. READING LIST

Class materials will be available on Moodle.

Textbook (selected parts only):

- [ISLR] James, G., D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning.* Springer.
  A very accessible introduction to machine learning methods.
  URL: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
- [ESL] Hastie, T., R. Tibshirani and J. Friedman: *The Elements of Machine Learning.* Springer.
  This is the previous book's "big brother", gives a more advanced treatment of ML.
  URL: https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12_toc.pdf
- [NNDL] Nielsen, M.: *Neural Networks and Deep Learning.*
  An online book aiming to give solid understanding of the core principles of neural networks.
  URL: http://neuralnetworksanddeeplearning.com/
- [EMA] Przemysław Biecek, P. and T. Burzykowski: *Explanatory Model Analysis*. CRC Press.
  A handbook on model exploration, explanation and examination aimed for practitioners.
  URL: https://ema.drwhy.ai/
- [BA] Lattimore, T. and Cs. Szepesvári: *Bandit Algorithms*. Cambridge University Press.
  An in-depth textbook on bandits with a mathematical focus.
  URL: https://tor-lattimore.com/downloads/book/book.pdf

Video:

- Making Friends with Machine Learning by Cassie Kozyrkov:
  https://www.youtube.com/watch?v=1vkb7BCMQd0
- StatQuest by Josh Starmer. https://www.youtube.com/c/joshstarmer
- 3Blue1Brown by Grant Sanderson. https://www.youtube.com/c/3blue1brown
- deeplizard. https://www.youtube.com/c/deeplizard

4. TEACHING METHOD AND LEARNING ACTIVITIES

The course will involve a mix of presentations, discussions, and practical sessions. Learning objectives will be achieved through in-class discussions, reviewing the course materials, and solving the homeworks.

5. ASSESSMENT

Grading will be based on the total score out of 100, in line with CEU's standard grading guidelines

- 30%: Quizzes at the beginning of each session
- 40%: 2 assignments with practical tasks
- 30%: Take home exam

Assignment acceptance policy and achievable grades:

- 100% until the due date, 50% within 24 hours past the due date, 0% after that

6. TECHNICAL REQUIREMENTS

Students need to bring their own laptops to the sessions with python installed (preferably python 3.10).

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
1	Scoring problems: recap, feature engineering, model diagnostics. Model ensembles: bagging & boosting,	Videos, selected chapters from ISLR & ESL
2	Classification problems. Hyper-parameter tuning, autoML.	Videos, selected chapters from ISLR & ESL
3	Intro to deep learning with keras. Extensions: convolutional networks, transfer learning.	Videos, selected chapters from NNDL
4	Interpretability. Blackbox vs glassbox explainers. Interpretability vs causality. Uncertainty in predictions. Bandits: exploration-exploitation tradeoff. Bias in adaptively collected data.	Selected chapters from EMA and BA

8. SHORT BIO OF THE INSTRUCTOR

János Divényi earned his Ph.D. in economics in 2020 at the Central European University. He has been working as a data scientist since 2013 when he joined the interdisciplinary research group of CEU Microdata. Currently, he is leading a small team at Emarsys (an SAP company) focusing on the analytical capabilities of their marketing cloud platform. He is particularly interested in causality.