

Syllabus

Course name

Introduction to Data Science

Semester

Spring 2022 semester

Name of the main instructor

Vahan Sargsyan, Ph.D.

Date

- Classes: April 25 – June 3, 2022
- Final exam week: June 6 – 10, 2022
- Make-up exam week: June 13 – 17, 2022

Time

- Lectures: Monday 7.00-8.30 a.m. CEST, Wednesday 7.00-7.45 a.m. CEST.
- Exercise session: Wednesday 7.45-8.30 a.m. CEST

Office hours:

To be determined.

Course description

This course provides an introduction to the data science as a profession and focuses on the theoretical methodologies of the most widely applied machine learning models. The main topics covered include: data preparation (data mining, cleaning and exploring strategies), statistical modeling with the application of appropriate machine learning methodologies (data segmentation, predictive analytics, recommendation systems), and mathematical evaluation.

Course Objective

The objective of this course is to present and elucidate the basic concepts of Data Science. It can be best seen as a course that provides the foundations of machine learning and therefore heavily relies on econometrics and data management, and opens the door to other econometric courses, both applied and theoretical.

Prerequisites

Statistics and Econometrics at introductory levels, experience with at least one programming language for statistical data analytics (Python, R, Stata or others).

Textbook and Materials

Burkov, A. (2019). "The Hundred-page Machine Learning Book".

Online documentations and materials for machine learning algorithms and analytics in the software of choice.

Grading

You will have **unannounced** quizzes during seldom classes which will cover all the topics up to that day. There is no retake of a quiz (even if you have excused absence). You will have one Final Exam, one Make-up Exam, and one Team Project in the end of the course. If you spend time on practicing the provided materials at home, you can excel in quizzes, project and exam. If you come on time, are attentive at classes and do the required reading, you shall have no difficulties throughout the course.

Grades:

Quizzes	20%
Participation/Attendance	15%
Team Project	30%
Final Exam	35%

ACADEMIC DISHONESTY: (THIS IS NOT A JOKE – READ CAREFULLY)!!!

Any form of academic dishonesty **will not** be tolerated at all. Cheating on exam and quizzes will automatically get you **ZERO** the first time from a particular exam or quiz and the second (last) time you will get **ZERO** from the class. Academic dishonesty includes receiving and/or providing unauthorized help from/to other students on exams or quizzes. Note also that **LATE QUIZ** or **EXAM WILL NOT BE ACCEPTED**.

Course Outline

I. Part 1 Introduction

1. Data Science as a Profession
2. Machine Learning in Data Science
3. Statistics and Econometrics in Machine Learning

II. Part 2 Notations and Definitions

4. Types of learning
5. Main models used in Machine Learning
6. Shallow vs. Deep Learning

III. Part 3 Working with Data

7. Data Sources
 8. Data Cleaning
 9. Specifications of Software/programming languages
-

IV. Part 4 Fundamental Algorithms

- 10. Linear Regression
- 11. Logistic Regression
- 12. Random Forest
- 13. Ensemble Models

V. Part 5 Analytics and Evaluation

- 14. Confusion Matrix
 - 15. Precision/Recall
 - 16. Accuracy
-