

INFOMDAML: Course Syllabus 2026

Robert Bagheri

2026-07-23

1 Introduction

Data do not fall from heaven, but are collected and analysed within specific contexts. Data analysis and machine learning are central and complementary components of modern data-driven inquiry, and when employed in combination, they enable the extraction of valuable insights, the development of predictive models, and informed decision-making.

The course is designed to balance conceptual understanding with practical implementation (applied & real-world data), using both R and Python and widely adopted libraries for data analysis and machine learning. In this course, you will get acquainted with data analysis and machine learning, and implement a variety of techniques to go from data to analyses, visualizations and insights for science and business applications. This is an overview course designed to give you the tools and skills to use and evaluate data science/ML methods.

Prerequisites

We assume that students who will join the course will have knowledge of statistics up to regression and analysis of variance, as well as some experience in programming in languages such as R and Python.

Objectives

At the end of this course, students have attained the following objectives:

1. know, explain, and apply methodology to properly set-up data analysis experiments, such as train, validate, and test and the bias/variance trade-off.
2. know, explain, and apply supervised machine learning and deep learning algorithms, both for classification and regression purposes as well as their related quality measures, such as AUC and the confusion matrix.
3. know, explain, and apply non-supervised learning algorithms, such as clustering and other techniques that result in lower-dimensional data representations.
4. be able to choose between the different techniques learned in the course and be able to explain why the chosen technique fits both the data and the research question best

2 Course Policy

This course is a mandatory course in the ADS programme and worth 7.5 ECTS, representing a substantial part of the expected study workload during this period.

Weekly course flow

A regular week in this course consists of one lectures (Tuesday, first week on Monday) and three lab sessions (Thursday). The last week of the course (week 9) will take place over a full Thursday and will feature four lectures in the format of a student conference, organised jointly with the INFOMDW course.

The material of the course is introduced on a theoretical level in the lectures and then put into practice in the lab sessions. The practical work done in these labs is drawn from real life situations that allow the students to experience how to solve data science problems.

In addition, students will spend time during each week on take-home group assignments.

- The **lectures** are in-person. The required readings should be completed before each lecture. These are *not optional*. Attendance at the lectures is required.
- The **lab sessions** are in-person interactive sessions in which you apply the methods you learn about in the lectures. The answers to the lab exercises are discussed at the end of each session. Attendance at the lab sessions is required and is *not optional*.
- The skills acquired in the lectures and the labs provide the basis for doing the **take-home assignments**. Each assignment is made in groups of 4-5 students and handed in via Brightspace.

Synchronous course policy

- INFOMDAML is an offline-first course, with mostly in-person lectures and lab sessions.
- We find it important for interactive and collaborative learning that the course is offline-first.
- If you miss a session, e.g., due to sickness, you should catch up in the regular way:
 - Read the readings
 - Go through the lecture slides
 - Do the practicals
 - Ask your peers if you have questions
 - (after the above) ask the lab teacher for further explanation

Who to ask what

There are many teachers in this course. If you have questions, first **ensure the answer isn't in this syllabus** and then follow the table below:

Question type	How to ask
Course proceedings	Email course coordinators (Robert)
Content - general	Email / ask lab teachers
Practical content	Email / ask lab teachers
Assignment content	Email / ask lab teachers
Lecture content	Email Lecturer (Javier / Daniel / Robert)

Grading policy

Your final grade in the course consists of the following grading components:

- Assignments (30%): There are two group assignments. Each assignment is graded and worth 15% of the final grade.
- Final exam (70%): At the end of the course, there is a final exam. The content of this exam emphasizes all the teaching material of the course.

To pass the course:

- The weighted final grade of all grading components should be greater than or equal to 5.5.

Resit:

- If you obtain a final grade between 4.0 and 5.4, you are eligible for the resit.
- You can only retake the final exam.
- The grade attained in the resit will replace the grade of the final exam.

3 Course materials

Required Software

In this course, we will use a variety of software, but mainly Python and R. Try to install both on your computer by the start of the course; we will also have a set-up computer lab on the first lab session to help you with this process.

Installing Python & Jupyter For the python parts of the course, we will use [Google Colab](#), which is an interactive online notebook environment; this means no installation is necessary! However, you do need a google account, so make sure you have one (or make one specifically for the course).

Installing R & RStudio First, install the latest version of R for your system (see <https://cran.r-project.org/>). Then, install the latest (desktop open source) version of the RStudio integrated development environment ([link](#)). We will make extensive use of the *tidyverse* suite of packages, which can be installed from within R using the command `install.packages("tidyverse")`.

Required readings

Freely available sections from the following books:

Book	Title (Authors)	URL
R4DS	R for Data Science (Grolemund & Wickham)	r4ds.hadley.nz
ISLR	Introduction to Statistical Learning (James et al.)	statlearning.com
DLBK	Deep Learning (Goodfellow, Bengio, Courville)	deeplearningbook.org
FIMD	Flexible Imputation of Missing Data (van Buuren)	stefvanbuuren.name/fimd
SLP3	Speech and language processing (Jurafsky & Martin)	stanford.edu/~jurafsky/slp3
TTMR	Text mining with R: A tidy approach (Silge & Robinson)	tidytextmining.com/

And (parts of) the following references:

- Oberski, D.L. (2016). Mixture models: Latent profile and latent class analysis. Modern statistical methods for HCI, 275-287. [URL](#)
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 54(6), 1-35. [URL](#)
- Hennig, C. (2015). Clustering strategy and method selection. arXiv preprint arXiv:1503.02059. [URL](#)
- Bouveyron, C., Celeux, G., Murphy, T. B., & Raftery, A. E. (2019). Model-based clustering and classification for data science: with applications in R (Vol. 50). Cambridge University Press.
- Some other freely available articles & chapters

4 Class Schedule

You can find the up-to-date class schedule with locations on mytimetable.uu.nl.

Week	Date	Topic	Type	Reading
1	2026-09-07	Introduction to machine learning	Lecture	Syllabus; R4DS 1, 2, 4, 6; ISLR 5.1, 8.1, 8.2.1, 8.2.2, 8.2.3
1	2026-09-10	Set up & know your tools	Lab	
1	2026-09-10	Supervised learning: regression analysis	Lab	

Week	Date	Topic	Type	Reading
2	2026-09-15	Supervised learning: classification & evaluation	Lecture	ISLR 4.1, 4.2, 4.3, 4.4.1, 4.4.2
2	2026-09-17	Supervised learning: classification & evaluation	Lab	
3	2026-09-22	Data analysis & visualization	Lecture	R4DS 1, 3, 9, 10 (optionally 2, 4, 6)
3	2026-09-24	Data visualization using ggplot	Lab	
4	2026-09-29	Missing data mechanisms & solutions	Lecture	FIMD 1.1, 1.2, 1.3, 1.4
4	2026-10-01	Missing data mechanisms & solutions	Lab	
5	2026-10-05	16:00 PM Assignment 1	Deadline	
5	2026-10-06	Unsupervised learning	Lecture	ISLR 12.1, 12.4; Oberski (2016), optional Hennig (2016), Bouveyron et al. (2019)
5	2026-10-08	Unsupervised learning: clustering	Lab	
6	2026-10-13	Deep learning 1	Lecture	ISLR 10, DLBK 11, (optionally 6)
6	2026-10-15	Deep learning 1	Lab	
7	2026-10-20	Deep learning 2	Lecture	
7	2026-10-22	Deep learning 2	Lab	
8	2026-10-26	16:00 PM Assignment 2	Deadline	
8	2026-10-27	Natural language processing	Lecture	SLP3 2.1, 2.4, 6.2, 6.3, 6.5, 6.8; TTMR 3
8	2026-10-29	Natural language processing	Lab	
9	2026-11-03	Q&A (13:00-14:00)	Lecture	
9	2026-11-05	Student conference day (09:00 - 18:00)	Lecture	
9	2026-11-05	Student conference day (09:00 - 18:00)	Social	
10	2026-11-10	Final exam (13:30 - 16:30)	Exam	

Week	Date	Topic	Type	Reading
12	2026-11-24	Optional exam Inspection	Exam Inspection	
18	2027-01-05	Resit exam	Exam	