Machine Learning

Diplomatische Akademie Wien
Vienna School of International Studies
Second Trimester 2025/2026
Course Type: Advanced
Course Credits: 4 ECTS

Lecture times: Please consult the online schedule. The lecture times and any possible changes are posted in the online schedule.
Office hours: Please see "Office hours" on the DA intranet
Convenor: Dr. Marcel Neunhoeffer
marcel@marcel-neunhoeffer.com

Course Description

Syllabi contain binding guidelines for each course. NO alterations are permitted.

Machine learning algorithms have become essential tools for research and decision-making across social sciences. This graduate course introduces participants to foundational and advanced topics in machine learning, with a focus on applications relevant to social science research. We will cover mathematical foundations, algorithmic mechanics, and practical implementation of key machine learning techniques.

The goal of the course is to balance theoretical understanding with hands-on applications. Each session combines conceptual lectures with lab exercises where participants apply methods to real-world social science datasets in R. By the end of the course, students will be equipped to implement machine learning workflows in their own research, evaluate model performance, and critically assess machine learning applications in published work.

Learning Objectives

By the end of this course, students will be able to:

1. Understand core machine learning concepts, including bias-variance tradeoff, cross-validation, regularization, and model evaluation
2. Implement supervised learning methods for classification and regression problems
3. Apply tree-based methods, ensemble learners, and neural networks to social science data
4. Design and execute complete machine learning workflows from data preprocessing to model deployment
5. Critically evaluate machine learning models and understand their limitations
6. Communicate machine learning results effectively in academic research

Prerequisites

- Graduate-level statistics and quantitative methods
- Basic familiarity with at least one programming language (R or Python)

Programming: All lab sessions will use R. Students with no R experience should familiarize themselves with R and R Studio before the class. For example, with the following online

book: Ismay, Chester, Albert Y. Kim, and Arturo Valdivia. 2025. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse (Second Edition). (Available online at: https://moderndive.com/v2/)

Assessments

Participation [20%]
Students are expected to actively participate in the course, including bringing critical questions related to the assigned readings to class, participating in group work, and completing exercises.

In-class presentation [20%]
Students will give a 10-minute presentation (with at most three slides) on the topic of their final paper (to be determined in consultation with the professor) in class on March 06, 2026. Students are expected to present a substantive social science research question they want to answer using one or more of the studied machine learning techniques, and to give a brief overview of how they will approach it.

Final paper [60%]
Students will be expected to write a short, original paper (research note/letter format) that presents an application of machine learning to a social science research question. A wide variety of approaches and research questions is acceptable and encouraged. Papers should be written in a concise style similar to that of articles in the natural sciences and should include 1-3 small display items (figures, tables, or equations) that convey the main results.

The topic for the final paper must be approved by the professor directly in class or via email by March 05, 2026, at 5 pm.

Each paper must include the following elements: an abstract (up to 200 words), an introduction (including a brief overview of related literature), an explicit statement of the research question, details on data and methods, an analysis and results section, and a discussion/conclusion.

Length: 3,500 words, including tables, figures, and footnotes, but excluding bibliography and appendices

The final paper must be submitted electronically to Moodle by the end of the trimester. See section below.

Examination / Paper Deadline

Final written work (denoted as the main assessment above) is to be submitted by **April 2, 2026 (9:00 am) via Moodle**. Should you encounter any difficulty submitting written work on the due date via Moodle, please send final work directly to paper-submission@da-vienna.ac.at. Only final work submitted via Moodle or emailed to paper-submission@da-vienna.ac.at will be accepted.

Not being able to meet an assessment criteria must be documented by a **doctor's letter** to the Academic Administration. See section **8.2. Medical Leave** in the Student Handbook.

Re-take policy: Students can re-take the **main assessment** once in order to improve their performance dependent on the program. Re-takes have to be completed by the end of the next following trimester.

NB:

- Unexcused absence from examination/failure to submit a paper forfeits the second attempt.
- A paper without scientific/research content will not be accepted as a submission.
- For the award of a tuition waiver and/or participation in an exchange programme, only examination results for first attempts in core and advanced courses will be considered.
- The grade on the re-take examination/paper (second attempt) replaces the grade of the previously taken examination/submitted paper (first attempt) but contributes to the overall course assessment to the same extent. Having re-taken an examination/re-submitted a paper it is NOT possible to revert back to the initially awarded grade.
- A failing grade on a re-take examination/paper (second attempt) cannot be improved. The core/advanced/optional course may be repeated in the next following academic year (subject to availability). Both courses will be displayed on the transcript of records.
- Students who wish to re-take an examination/re-submit a paper for advanced/optional courses must send a request via email to paper-submission@da-vienna.ac.at immediately and no later than one week after the grade is entered into the student's records.

Re-take papers are to be submitted by June 8, 2026 (9:00 am) via Moodle. Should you encounter any difficulty submitting written work on the due date via Moodle, please send final work directly to paper-submission@da-vienna.ac.at. Late submissions will not be accepted. Only final work submitted via Moodle or emailed to paper-submission@da-vienna.ac.at will be accepted.

For further information, please check your student account and the Student Handbook (Chapter 9, "Regulations for Examinations, Seminar and Course Papers, Interviews").

Course Topics and Meetings

Session 1: Introduction & Machine Learning Basics – Learning, Generalization & Hyperparameters

Session 2: Predictions & Coefficients – Regression & Regularization

Session 3: Weak Learners, Ensembles & Boosting – Decision Trees, Random Forests & Gradient Boosting

Session 4: Neural Networks & Deep Learning

Session 5: Modern Neural Network Architectures – The Transformer

Session 6: Student Presentations & Wrap Up

Readings

Fundamental readings

Fundamental readings are for orientation and for students lacking the preferred background knowledge for this course.

Ismay, Chester, Albert Y. Kim, and Arturo Valdivia. 2025. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse (Second Edition). (Available online at: https://moderndive.com/v2/)

Textbooks

Throughout the course, we will work with the following two textbooks:

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT Press. (Available online at: https://www.deeplearningbook.org)

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. An Introduction to Statistical Learning with Applications in R. Springer New York, NY. (Available online at: https://www.statlearning.com)

Session 1: Introduction & Machine Learning Basics – Learning, Generalization & Hyperparameters

Required readings:

Deep Learning (DL): Chapter 5, until and including Section 5.4

An Introduction to Statistical Learning (ISL): Chapter 2.1, 2.2, and Chapter 5

Arnold, Christian, Luka Biedebach, Andreas Küpfer, and Marcel Neunhoeffer. 2024. The role of hyperparameters in machine learning models and how to tune them. Political Science Research and Methods, 12(4): 841-848.

Cranmer, Skyler J., and Bruce A. Desmarais. 2017. "What can we learn from predictive modeling?" Political Analysis 25(2): 145-166.

Optional readings:

Neunhoeffer, Marcel, and Sebastian Sternberg. 2019. "How cross-validation can go wrong and what to do about it." Political Analysis 27(1): 101-106.

Wrap-up:

Session 2: Predictions & Coefficients – Regression & Regularization

Required readings:

DL: Re-read Chapters 5.1.4 and 5.2, read Chapter 5.7.1

ISL: Chapters 3.1, 3.2, 6.1, and 6.2

Kim, In Song. 2017. "Political Cleavages within Industry: Firm-Level Lobbying for Trade Liberalization." American Political Science Review 111(1): 1–20.

Session 3: Weak Learners, Ensembles & Boosting – Decision Trees, Random Forests & Gradient Boosting:

Required readings:

DL: Chapter 5.7.3

ISL: Chapter 8

Kaufman, Aaron Russell, Peter Kraft, and Maya Sen. 2019. "Improving supreme court forecasting using boosted decision trees." Political Analysis 27(3): 381-387.

Optional readings:

Castanho Silva, Bruno, Danielle Pullan, and Jens Wäckerle. 2025. "Blending in or standing out? Gendered political communication in 24 democracies." American Journal of Political Science 69(2): 653-668.

Session 4: Neural Networks & Deep Learning:

Required readings:

DL: Chapters 5.9, 5.11

ISL: Chapter 10

Torres, Michelle, and Francisco Cantú. 2022. "Learning to see: Convolutional neural networks for the analysis of social science data." Political Analysis 30(1): 113-131.

Optional readings:

DL: Chapter 6

Session 5: Modern Neural Network Architectures – The Transformer:

Required readings:

Turner, Richard E. 2024. "An introduction to transformers." arXiv preprint arXiv:2304.10557.

Alvarez, R. Michael, and Jacob Morrier. 2025. "Measuring the Quality of Answers in Political Q&As with Large Language Models." Political Analysis: 1–18.

Optional readings:

Vaswani, Ashish, et al. 2017. "Attention Is All You Need". Advances in Neural Information Processing Systems, vol. 30.

Session 6: Student Presentations & Wrap Up:

Optional readings:

Blackwell, Matthew, and Michael P. Olson. 2022. "Reducing model misspecification and bias in the estimation of interactions." Political Analysis 30(4): 495-514.

Athey, Susan, and Guido W. Imbens. 2019. "Machine learning methods that economists should know about." Annual Review of Economics 11(1): 685-725.