

# Seminar Data Analytics

## Summer Term 2026 - Institute of Labour Economics

Prof. Dr. Patrick Puhani | Dr. Franziska Braschke | M.Sc. Vangjel Bitu

### Seminar Description

In this seminar, we invite students to engage in practical exercises, gain hands-on experience with Python, and learn and practice machine learning (ML) skills within the field of economics and business analytics. Students will form teams and be assigned projects, which include a specific topic and dataset to work on throughout the semester. Additionally, python code, introduction on the methods and helpful instructions will be provided to start with (Programming experience is advantageous but not necessary).

Our projects will focus on leveraging the predictive advantages offered by machine learning for data-driven analysis. Specifically, we will utilize machine learning techniques to:

1. Estimate wage gaps (gender and union) using partially linear regression where ML predictions are used to find the optimal controls.
2. Identify racial bias and make accurate predictions of recidivism risk for individuals with prior convictions.
3. Explore heterogeneity in treatment effects of Conditional Cash Transfers (PKH) on child health outcomes in Indonesia.

Students in groups of up to 2 will choose a **Project Track** (see Table below) and a machine learning algorithm to apply to their data. The group assignment will take place via StudIP on a first-come, first-served basis. The concrete topic slots will be created on StudIP by March; we will notify you by an announcement on StudIP.

### Contents

<b>1 Objectives &amp; Learning Outcomes</b>	<b>2</b>
<b>2 Project Tracks (Topics &amp; Data)</b>	<b>2</b>
<b>3 Course Structure</b>	<b>2</b>
<b>4 Books and Literature</b>	<b>3</b>
<b>5 Machine Learning Algorithms</b>	<b>3</b>
<b>6 Examination</b>	<b>3</b>
<b>7 Seminar Paper Guidelines</b>	<b>4</b>
<b>8 Presentation Guidelines</b>	<b>5</b>
<b>9 Submission of Your Work</b>	<b>5</b>
<b>10 Important Dates and Deadlines</b>	<b>6</b>

# 1 Objectives & Learning Outcomes

The primary aim is to turn students into knowledgeable users of ML methods with awareness for potential pitfalls when transferring methods to real-world applications (e.g., data quality, over-fitting, causal inference, and ethical considerations).

By the end of the course, students should be able to:

- Work effectively with data and program in Python.
- Perform data manipulation and visualization.
- Understand and apply ML methods like Trees, Random Forests, etc.
- Use ML for causal analysis (choosing best controls to overcome omitted variable bias).
- Use ML methods for optimal and fair prediction.
- Use ML for the analysis of heterogeneous treatment effects (Meta Learners).

# 2 Project Tracks (Topics & Data)

To clarify the assignment of topics, methods, and data, the seminar is divided into three specific Project Tracks. Each group will be assigned one track.

Track Name	Analysis Goal	Analytical Approach	Dataset
1. Wage Gap Analytics	Estimate Gender OR Union Wage Gaps.	Partial Linear Regression (Double ML) for optimal controls.	CPS Earnings
2. AI Bias & Predictive Fairness	Detect machine bias & calibrate predictions for recidivism.	Classification Algorithms & Fairness Calibration.	COMPAS Recidivism
3. Heterogeneous Policy Effects	Analyze impact of Cash Transfers (PKH) on birth weight.	Meta Learners (S-Learner / T-Learner).	Indonesia Family Life Survey (Triyana, 2016)

# 3 Course Structure

The seminar will consist of the five following meetings.

(For specific dates, see the "Important Dates" section on the last page!!!).

1. **Seminar Kick-off:** Presentation of the general seminar structure, the Project Tracks (Table 1), tasks, and student expectations.
2. **Methods Introduction:** Hands-on introduction to the topics and machine learning methods (Double ML, Fairness, Meta Learners).
3. **Coding Introduction:** Hands-on introduction to coding in Python and implementing the specific ML algorithms.
4. **Feedback Meeting:** Opportunity for questions and guidance two weeks before submission.
5. **Presentations:** Final presentation of results.

## 4 Books and Literature

Relevant literature per Project Track:

1. **Track 1 (Wage Gap Analytics):** Double/debiased machine learning for treatment and structural parameters.
2. **Track 2 (AI Bias & Predictive Fairness):** COMPAS Recidivism Risk Score Data and Analysis
3. **Track 3 (Heterogeneous Policy Effects):** Do Health Care Providers Respond to Demand-Side Incentives? Evidence from Indonesia

Main Sources (General):

1. [DMBA] Shmueli, Bruce, Gedeck, Patel. Data mining for business analytics: concepts, techniques and applications in Python (Data Analytics Course Book, covers both ML methods and coding in Python.)
2. [PDSH] VanderPlas: Python Data Science Handbook (Good for programming in Python.)

Further Sources: look at your topic handout.

## 5 Machine Learning Algorithms

Regardless of the Project Track, groups will choose an algorithm to implement:

- Decision Trees
- Random Forests
- Boosted Trees

## 6 Examination

1. Submission of Code script: Code that reproduces graphs, tables, and the results of your analysis in Python.
2. Term Paper: 8 - 10 pages. For more details see section Seminar Paper Guidelines
3. Presentation: 15 Minutes. For more details see section Presentation Guidelines
4. **Grade Determination:**
  - Seminar Paper: 50%
  - Oral Presentation: 40%
  - Python Code: 5%
  - Active Participation in the Seminar: 5 %

## 7 Seminar Paper Guidelines

### Essential Components of the Seminar Paper

- Cover page
- 8 to 10 pages of text and essential tables and figures
- References
- Appendix with (additional) figures or tables from your data work

### Formal Aspects

- Text font and size: Times New Roman, Arial or Calibri 12 point, or LaTeX standard 11 point
- Line spacing: 1.5-line
- Alignment: Justify
- Margins: all sides 2.54 cm (1 inch)
- Page numbers
- Citation: Harvard citation style, i.e. in the text as ‘surname (year)’ respectively ‘surname (year, p. xy)’ for direct quotes
- References: Include every reference cited in the text and no others
- Language: German or English

### Detailed Structure and Guidelines for Seminar Paper

#### Introduction

- State the analysis goal, relevance, and objective clearly.
- Briefly describe your choice of variables and chosen methodology.

#### Descriptive Statistics and Data Preparation

- Provide key summary statistics and visualizations.
- Clearly document data cleaning and variable selection reasons.

#### Methodology

- Explain your group’s ML algorithm and how it work, its strengths and weaknesses(if any).
- Detail the chosen method (Double ML, calibration vs balance, Meta Learners) as per your Project Track.

#### Results

- Explain your steps on applying the chosen method and your ML algorithm.
- Clearly present and compare baseline regression and machine learning results.
- Include graphical representations of key outcomes.

- Describe your hyper-parameter tuning strategy and show graphically.

### Discussion

- Reflect on the robustness and significance of the results. What do they mean?
- Identify possible methodological and data limitations, if any.

### Conclusion

- Summarize key findings and their implications.

## 8 Presentation Guidelines

- Duration: 15 minutes + 10 minutes discussion.
- Suggested structure:
  - Intro Slide
  - Show your contributions on data preparation and data cleaning and present interesting descriptive statistics or data visualization.
  - Describe your implementation and any challenges of your Track's method (Partial linear models, Meta Learners, etc.).
  - Describe your implementation and any challenges of the ML algorithm of your group (Trees, random forest, boosting).
  - Show your hyper-parameter tuning strategy (which target values tuned, process, results).
  - Present your results, tables and graphs.
  - Discussion slide (How do your results compare to the baseline? How do you interpret your results in this context?).

## 9 Submission of Your Work

Submit your seminar paper and presentation as PDF files via email to: **seminare@aoek.uni-hannover.de**. Submit also your python code script as .py or as .ipynb files. If possible, submit as a zip file containing dataset, code, result tables, and PDF of your seminar paper.

**Subject line format:** *"Submission (or Abgabe) Seminar DATA ANALYTICS"*

### File naming conventions:

- *DA GroupNumber TrackName YourLastName S* (seminar paper)
- *DA GroupNumber TrackName YourLastName P* (presentation)
- *DA GroupNumber TrackName YourLastName C* (python code script)

### Example:

- DA 1 WageGap Mustermann S/P/C

## 10 Important Dates and Deadlines

- **22.04.2026:** Registration deadline (Send forms to [seminare@aoek.uni-hannover.de](mailto:seminare@aoek.uni-hannover.de)).
- **14.04.2026:** Kick-off Meeting (Seminar introduction and Topic presentation).
- Methods Introduction Meeting (project track specific).
  - **24.04.2026:** Methods Introduction Meeting for Track 2.
  - **28.04.2026:** Methods Introduction Meeting for Track 1.
- Coding Introduction Meeting (project track specific).
  - **29.04.2026:** Coding Introduction for Track 2.
  - **05.05.2026:** Coding Introduction for Track 1.
- **17.05.2026 (23:59):** Draft Code submission before feedback meeting (optional)
- **19.05.2026:** Feedback meeting before submission for questions and guidance.
- **14.06.2026 (23:59):** Term Paper and Presentation submissions.
- **16-17.06.2026:** Seminar Presentations.

*Exact times and rooms to be announced on StudIP.*