Discipline: Information Systems

1 Language
English

2 Title
Data Science as a Research Method

3 Lecturer
Professor Dr. Oliver Müller, Universität Paderborn
https://wiwi.uni-paderborn.de/dep3/mueller/team/
oliver.mueller@uni-paderborn.de

4 Date and Location
19.-22.02.2024

Universität Paderborn
LS für Wirtschaftsinformatik, insb. Data Analysis
Warburger Str. 100
33098 Paderborn

5 Course Description
5.1 Abstract and Learning Objectives
The course is targeted at PhD students and young researchers who want to apply data science methods in their research. It covers various data preparation, statistical modeling, and visualization techniques for extracting knowledge from the vast and complex data sets that have emerged in business over the past years. The learning objective of the course is to enable participants to apply these techniques in design-oriented and/or quantitative empirical research projects.

5.2 Content
Topics include supervised machine learning for regression and classification (linear and non-linear models), unsupervised machine learning (e.g., clustering, dimensionality reduction), ensemble models, AutoML, interpretability of machine learning, and natural language processing (e.g., dictionary-based methods, supervised text classification, topic modeling).
5.3 Schedule

Day I
- 10:00 – 10:30: Welcome
- 10:30 – 12:00: Introduction to data science as a research method
- 12:00 – 13:00: Lunch
- 13:00 – 14:00: Introduction to data science with Python
- 14:00 – 15:00: Supervised learning with linear models (i.e., linear and logistic regression)
- 15:00 – 16:00: Exercises with Python
- 16:00 – 17:00: Tree-based models (incl. bagging, random forest, and boosting)
- 17:00 – 18:00: Exercises with Python

Day II
- 09:00 – 10:30: Deep learning (i.e., single- and multilayer neural networks)
- 10:30 – 12:00: Exercises with Python
- 12:00 – 13:00: Lunch
- 13:00 – 13:45: Ensemble models and AutoML
- 13:45 – 14:30: Interpretable machine learning
- 14:30 – 15:15: Exercises with Python
- 15:15 – 16:15: Unsupervised learning with clustering and dimensionality reduction
- 16:15 – 17:00: Exercises with Python

Day III
- 09:00 – 10:00: Natural language processing with dictionaries and the bag-of-words model
- 10:00 – 11:00: Exercises with Python
- 11:00 – 13:00: Natural language processing with neural networks (incl., word embeddings, LSTM)
- 13:00 – 14:00: Lunch
- 14:00 – 15:00: Exercises with Python
- 15:00 – 16:00: Natural language processing with probabilistic topic models
- 16:00 – 17:00: Exercises with Python
Day IV

- 09:00 – 12:00: Group work on data science mini-project (Hackathon)
- 12:00 – 13:00: Lunch
- 13:00 – 14:00: Group work on data science mini-project (Hackathon)
- 14:00 – 15:30: Presentation of data science mini-project
- 15:30 – 16:00: Wrap-up and feedback

5.4 Course format
The course consists of a mix of conceptual lectures, hands-on exercises with Python, and discussions of research papers. Participants get the chance to directly apply the methods taught in the lectures to real-life datasets. In addition, we will discuss a number of papers applying data science methods. Hence, we strongly recommend that participants familiarize themselves with the papers listed in the reading list. The course language is English.

6 Preparation and Literature

6.1 Prerequisites
Master-level education in Business, Economics, Computer Science, Engineering, or a related field.

6.2 Essential Reading Material


6.3 To prepare
All participants are required to read the essential reading material prior to the course. In addition, participants will receive a list of Python programming assignments that have to be completed before the course.
7 Administration

7.1 Max. number of participants
The number of participants is limited to 20.

7.2 Assignments
The course will include a number of in-class assignments. These assignments will not be graded.

7.3 Exam
After the course, students are required to finalize and write up the results of their Data Science Mini-Project (Day IV) in a working paper (10 pages). The working paper will be graded.

7.4 Credits
The course corresponds to a scope of 6 LP/ECTS.

8 Working Hours

<table>
<thead>
<tr>
<th>Working Hours</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandatory readings</td>
<td>40 h</td>
</tr>
<tr>
<td>Programming assignments (to be completed before the course)</td>
<td>40 h</td>
</tr>
<tr>
<td>Active participation in class</td>
<td>30 h</td>
</tr>
<tr>
<td>Final exam (working paper to be written after the course)</td>
<td>70 h</td>
</tr>
<tr>
<td>TOTAL</td>
<td>180 h</td>
</tr>
</tbody>
</table>

ECTS: 6