Data Science and Machine Learning in Python

Virtual international collaborative course offering between March to July 2022

Prof. Dr. Stephan Weyers





Background

This course is part of InduTwin funded by DAAD. The goal of this project is to establish and extend double degree agreements in the fields of Business Management, Engineering and Computer Science between FH Dortmund and partner universities in Latin America and China.

Students of the bachelor study programs International Business, Betriebswirtschaft and Betriebswirtschaftliche Logistik at FH Dortmund are invited to choose this course as an elective.

Students from the InduTwin partner universities UGTO Guanajuato, UTTEC Tecamac, UBA Buenos Aires, UDEM Medellin, Universidad de Valparaiso, and ESAN Lima are invited to participate in this course. Please apply at your International Office or via the InduTwin contact persons of your university. You can receive a certificate of attendance and a transcript of records. Whether or not this course can be acknowledged at your home university is out of the responsibility of FH Dortmund.

The main focus of this course is on collaborative case study work in international student teams. You will get to know people from other countries and continents and work together with them on realistic Data Science tasks.

Organization

Lectures

- Tuesdays between March 29th, 2022, and June 28th, 2022
- 14:30-19:20 German local time (CEST/UTC+2)
- German students can attend classes on-site, if pandemic situation allows
- International students attend lectures online (hybrid setup)
- Very interactive including break-out rooms in small groups
- Course language: English

Preparation and teamwork

- Self learning in preparation of lectures expected (DataCamp courses and textbooks)
- Weekly assignments
- Self organized collaboration in international student teams outside of classroom
- Students are recommended to start individual preparation mid of March 2022
- Deadline for last team assignment until mid of July 2022

Preparation and course material

- Google Colaboratory (free Google account required) <u>https://colab.research.google.com/</u>
- Optional: Jupyter Notebook <u>https://www.anaconda.com/products/individual</u>
- Textbooks (available in library of FH Dortmund, partner students to check own libraries)
 - McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media. <u>https://ebookcentral.proquest.com/lib/fh-dortmund/detail.action?docID=5061179</u>
 - Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python. O'Reilly Media. <u>http://ebookcentral.proquest.com/lib/fh-dortmund/detail.action?docID=4698164</u>
- Courses on DataCamp.com (free access available for participants between Mar-Aug 2022)
 - Introduction to Python <u>https://learn.datacamp.com/courses/intro-to-python-for-data-science</u>
 - Intermediate Python <u>https://learn.datacamp.com/courses/intermediate-python</u>
 - Supervised learning <u>https://learn.datacamp.com/courses/supervised-learning-with-scikit-learn</u>
 - Unsupervised learning https://learn.datacamp.com/courses/unsupervised-learning-in-python
- ILIAS learning platform at FH Dortmund (registration via TAN for partner students)

Syllabus

Part 1: Data Science Basics

	Date	Topics covered
1	Mar 29 th	How to use Google Colaboratory Python types and lists
2	Apr 5 th	Loops, if/else, functions Tuples, lists, dictionaries
3	Apr 12 th	Numpy basics and operations Image processing
4	Apr 26 th	Pandas Series, DataFrame Import/export files
5	May 3 rd	Principles of data visualization Data cleaning and preparation Join, combine and reshape data
6	May 10 th	Data visualization in Python How to write Data Science reports Data aggregation and grouping

Part 2: Machine Learning

	Date	Topics covered
7	May 24 th	(Un-)supervised learning in scikit-learn k-Nearest Neighbors Linear regression (ridge and lasso)
8	May 31 st	Linear classification models Ensembles of decision trees
9	Jun 7 th	Kernel support vector machines Neural networks
10	Jun 14 th	Preprocessing and scaling Dimensionality reduction Principal component analysis
11	Jun 21 st	k-means, hierarchical clustering, DBSCAN
12	Jun 28 th	Representing data, engineering features Model evaluation and improvement Text data analysis

Use Cases (selection) Customer Churn Prediction Market Basket Analysis Inventory analytics Social Irresponsibility Survey Motor insurance dataset Happiness dataset Fraud Detection World development indicators

Workload

- Total expected workload: 10 ECTS / 200-300 hours
- Part 1: Introduction to Data Science with Python (5 ECTS / 100-150 hours)
- Part 2: Machine Learning with Python (5 ECTS / 100-150 hours)
- https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en

Grading

- 10% Self-study courses on DataCamp.com
- 40% Individual assignments
- 50% Team case studies

This course is designed for Bachelor students in the field of Business and Management. However, students of other disciplines like Engineering and Computer Science are very welcome. Data Science is interdisciplinary and can be applied to many different types of problems.

Participants are expected to have basic knowledge in Mathematics, Statistics and Microsoft Excel.

Some pre-knowledge in programming is recommended. Although the course starts from scratch, quite a lot of topics will be covered in a relatively small amount of time. If you don't have any pre-experience in programming, you can still take part and can be very successful, but you should expect a high workload. It is definitely helpful to already start with the DataCamp courses before the first lecture. All materials and detailled information will be provided well in advance.

Student Feedback 2021

Fachhochschule Dortmund

University of Applied Sciences and Arts

not at all

8. Were your expectations fulfilled?	completely	not at all
The content of the event/course matches my field of study/research.	I totally agree	l totally disagree
The event/course offered was excellent.	I totally agree	l totally disagree
The event/course had a recognizable goal/ outcome.	I totally agree	l totally disagree
Attending the event/course has increased my knowledge of the subject matter considerably.	I totally agree	l totally disagree
The content level of the event/course was high.	I totally agree	l totally disagree
The amount of work for the event/course was substantial.	I totally agree	l totally disagree
The materials and tools provided were helpful for the learning achievement.	I totally agree	l totally disagree
The atmosphere during the stream encouraged me to ask questions.	I totally agree	l totally disagree
The event/course had a clearly recognizable connection to practice.	I totally agree	l totally disagree
It was very helpful that the teaching and learning materials for this event/course were made available on the internet.	I totally agree	l totally disagree
Digital formats should be offered more often in similar events/ courses.	I totally agree	l totally disagree
The digital format was used well.	I totally agree	l totally disagree
organized the lecture/course well	I totally agree	l totally disagree
explained the subject matter clearly	I totally agree	I totally disagree
clearly defined the learning objectives offered.	I totally agree	I totally disagree
presented summaries that supported the learning process.	I totally agree	l totally disagree
been interested in the learning progress of the participants	I totally agree	l totally disagree

given clear and detailed answers to questions	I totally agree		I totally
			disagree
spoken clearly and concisely.	I totally agree		l totally disagree
used varied teaching methods.	I totally agree		I totally disagree
encouraged us to ask questions	I totally agree		I totally disagree
deepened my interest in the subject.	I totally agree		l totally disagree
practiced gender equality	I totally agree	+ + -	l totally disagree
The ability to work in cross-disciplinary teams, where team members come from different study programs.	to a very high degree		not at all
The ability to motivate and coordinate a team work with students from different study programs.	to a very high degree		not at all
The ability to learn new approaches and complete group assignments with students from other study programs.	to a very high degree		not at all
The ability to work in international project teams.	to a very high degree		not at all
The ability to adapt to different strengths and weaknesses of other team members in order to achieve goals.	to a very high degree		not at all
The ability to consider different points of view and approaches for problem-solving in teams.	to a very high degree		not at all
The ability to bring different ideas together and to develop common study/ research topics.	to a very high degree		not at all
The ability to communicate and work effectively with people from different cultural backgrounds.	to a very high degree		not at all
The ability to bridge intercultural differences.	to a very high degree	$\left \begin{array}{c} \bullet \\ \bullet $	not at all
The ability to communicate in English with non- native speakers on research/ study topics with different native and non-native English speakers.	to a very high degree		not at all
The ability to communicate in written English with non-native speakers on research/study related	to a very high degree		not at all

to a very high degree

topics.

The ability to present and discuss research/study topics in English with non-native speakers.

Registration deadline and selection process

Students from Dortmund may choose this course as Wahlpflichtmodul. If you are a student of FH Dortmund, please consider the due dates of the internal selection process. About 45 German students may participate in this class.

InduTwin partners may send the contact details of the interested preselected students as soon as possible but not later than March 15th, 2022.

In total at most 100 students may participate in this course. That means, each of the 6 Latin American partner universities can send about 10 students. If more students are interested, feel free to send a longer list.

Lecturer

Stephan Weyers has been a professor for Mathematics, Statistics and Supply Chain Management at Fachhochschule Dortmund since March 2019. From 2014-2019 he was professor for Mathematics and Didactics at Technische Hochschule Mittelhessen in Gießen. In his professional career he worked as a Senior Analytic Specialist at McKinsey & Company between 2007-2014.

Stephan Weyers is Project Director of InduTwin.

Contact

stephan.weyers@fh-dortmund.de