

# An Introduction to Machine Learning



HELMHOLTZAI

Anoushka Jain, Till Korten, Thorsten Auth, **Peter Steinbach**  
*Helmholtz-Zentrum Dresden-Rossendorf and elsewhere / March 11, 2024*

“Helping the students with less computing background succeed makes a much bigger difference for society long-term than does keeping entertained the most privileged students.” (Mark Guzdial, 2020)

“Helping the students with less computing background succeed makes a much bigger difference for society long-term than does keeping entertained the most privileged students.” (Mark Guzdial, 2020)

- **audience:**  
domain scientists

“Helping the students with less computing background succeed makes a much bigger difference for society long-term than does keeping entertained the most privileged students.” (Mark Guzdial, 2020)

- **audience:**  
domain scientists
- **goals:**
  1. learn existing tools
  2. adopt to domain data

“Helping the students with less computing background succeed makes a much bigger difference for society long-term than does keeping entertained the most privileged students.” (Mark Guzdial, 2020)

- **audience:**

- domain scientists

- **goals:**

- 1. learn existing tools
  - 2. adopt to domain data

- **content serving the majority of learners:**

- concepts
  - pipelines
  - metrics

## No Free Lunch (David Wolpert, 1996)

---

“In a noise-free scenario where the loss function is the misclassification rate, if one is interested in off-training-set error, then **there are no a priori distinctions between learning algorithms.**”

“In a noise-free scenario where the loss function is the misclassification rate, if one is interested in off-training-set error, then **there are no a priori distinctions between learning algorithms.**

How well you do is determined by how ‘aligned’ your learning algorithm  $P(f_{pred}|data)$  is with the actual posterior,  $P(f_{true}|data)$ . ”

## Modern (Teaching) Challenges


---




- 📖 content for 3 - 4 days  
(essential vs. important)
- ⚙️ slidedecks won't cut it  
(learner speed)
- 😐 zoom fatigue




## status quo

 instructor transports  
content  
(one speed for all)


## status quo

 instructor transports  
content


(one speed for all)

 learner consumes  
exercises  
(reinforces alone)


## status quo

 instructor transports  
content


(one speed for all)

 learner consumes  
exercises  
(reinforces alone)


### **status quo**

 **instructor transports content**

(one speed for all)


 **learner consumes exercises**  
(reinforces alone)

### **flipped**


 **learner consumes content**

(alone at individual speed)

### status quo


 instructor transports  
content

(one speed for all)


 learner consumes  
exercises

(reinforces alone)

### flipped

 learner consumes  
content

(alone at individual  
speed)

 learner performs  
exercises in group  
(reinforces with mentor)

# Machine learning in Python with scikit-learn MOOC



Scikit-learn course

🔍 Search this book...

Introduction

MACHINE LEARNING CONCEPTS

📖 Introducing machine-learning concepts

✅ Quiz Intro.01

THE PREDICTIVE MODELING PIPELINE

Module overview

Tabular data exploration

Fitting a scikit-learn model on numerical data

Handling categorical data

📄 Wrap-up quiz 1

Main take-away



## Introduction

### Course presentation

#### 1 Welcome!

The goal of this course is to teach machine learning with scikit-learn to beginners, even without a strong technical background.

Predictive modeling brings value to a vast variety of data, in business intelligence, health, industrial processes and scientific discoveries. It is a pillar of modern data science. In this field, scikit-learn is a central tool: it is easily accessible, yet powerful, and naturally dovetails in the wider ecosystem of data-science tools based on the Python programming language.

This course is an in-depth introduction to predictive modeling with scikit-learn. Step-by-step and didactic lessons introduce the fundamental methodological and software tools of machine learning, and is as such a stepping stone to more advanced challenges in artificial intelligence, text mining, or data science.

The course is more than a cookbook: it will teach you to be critical about each step of the design of a predictive modeling pipeline: from choices in data preprocessing, to choosing models, gaining insights on their failure modes and interpreting their predictions.

## Course Setup and Timetable

---

- Split audience into **teams of up to approx 10 learners**  
(*2 mentor(s) per team*)
- teams walk through modules of learning material **independently**
- learning modules:  
**jupyter notebooks and prerecorded videos**  
(on-premise jupyter service, mybinder, google colab)

# Course Setup and Timetable

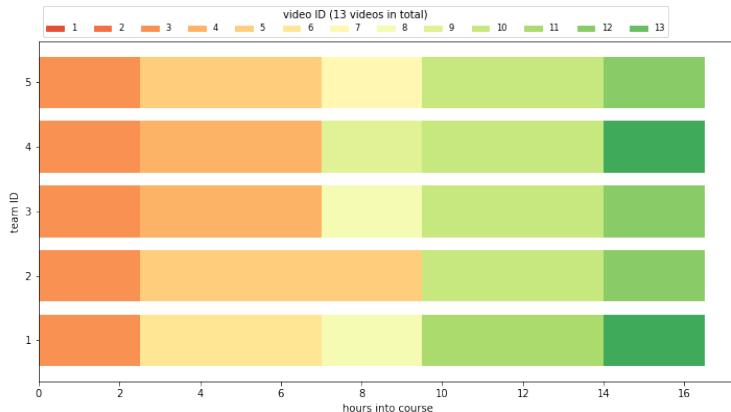
---

- Split audience into **teams of up to approx 10 learners**  
(*2 mentor(s) per team*)
- teams walk through modules of learning material **independently**
- learning modules:  
**jupyter notebooks and prerecorded videos**  
(on-premise jupyter service, mybinder, google colab)

- 09:00 all-hands good morning  
(feedback from last day)
- 09:15 split into teams (rinse and repeat)
  - learners **study notebooks** autonomously
  - mentor polls team to rejoin  
(or for more time)
  - **general discussion** on the video
  - learners conduct **exercises**
- 16:15 general questions and feedback
- 16:30 good bye



## Evidences for different learning speeds



**Figure:** Use of videos of five teams across the course duration (hours into course). Teams expose different learning speeds.

# Learner Assessment

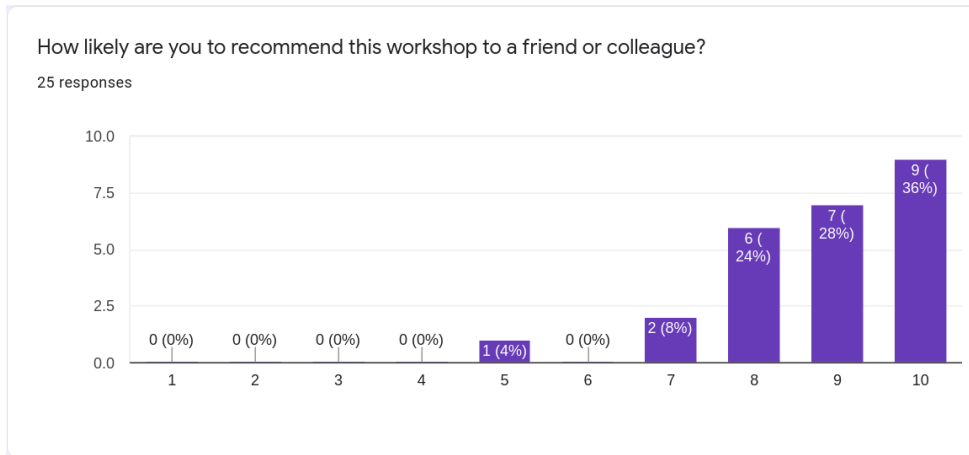


Figure: Net Promoter Score: results from Oct 2021 GSI workshop survey, 25/36 learners replied

## Code of Conduct

Please use positive language and be supportive to your peers. *In case of issues, contact the facilitators.*

## Code of Conduct

Please use positive language and be supportive to your peers. *In case of issues, contact the facilitators.*

## Central (learner) Pad

Use the central pad to get orientation. Add anything that may be of general interest.

# Housekeeping

---

## Code of Conduct

Please use positive language and be supportive to your peers. *In case of issues, contact the facilitators.*

## Central (learner) Pad

Use the central pad to get orientation. Add anything that may be of general interest.

## Zoom

Use our Zoom room for communication.

# Housekeeping

---

## Code of Conduct

Please use positive language and be supportive to your peers. *In case of issues, contact the facilitators.*

## Central (learner) Pad

Use the central pad to get orientation. Add anything that may be of general interest.

## Zoom

Use our Zoom room for communication.

## Notebooks

If you can, work through the notebooks on your local laptop. A backup solution is described on the central learner pad.

## To all mentors ...

---



**Thank you!**  
(especially given the chaotic preparation)

## Summary

---

- virtual “flipped” classroom approach



## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)

## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)
- scalable compact courses  
(open educational resources, reusable video+notebooks)

## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)
- scalable compact courses  
(open educational resources, reusable video+notebooks)
- convergence of MOOC and classroom

## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)
- scalable compact courses  
(open educational resources, reusable video+notebooks)
- convergence of MOOC and classroom

## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)
- scalable compact courses  
(open educational resources, reusable video+notebooks)
- convergence of MOOC and classroom

**Questions, Comments, Feedback or  
Concerns are highly welcome!**

## Summary

---

- virtual “flipped” classroom approach
- small learner groups  
(in-class notebook-discuss-exercise cycles)
- scalable compact courses  
(open educational resources, reusable video+notebooks)
- convergence of MOOC and classroom

**Questions, Comments, Feedback or  
Concerns are highly welcome!**

**Have fun learning together!**