

# Machine Learning (IN2064)

## Lecture 1: Introduction

---

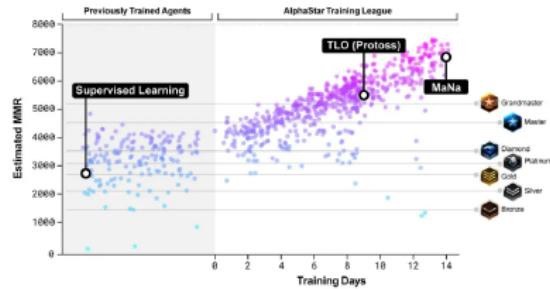
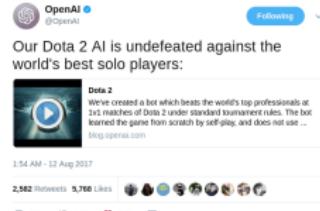
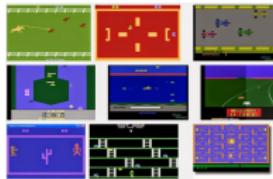
Prof. Dr. Stephan Günnemann

Data Analytics and Machine Learning  
Technical University of Munich

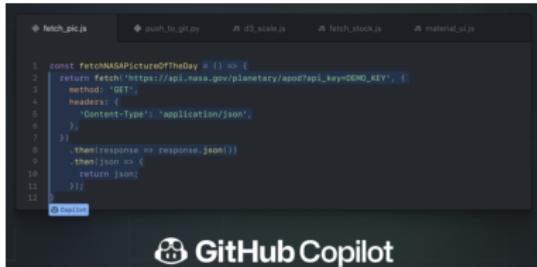
[cs.cit.tum.de/daml/](http://cs.cit.tum.de/daml/)

Winter term 2022/2023

## Game playing



# Natural language processing



A screenshot of the GitHub Copilot interface. At the top, there's a navigation bar with tabs like 'fetch\_nasa.js', 'push\_to\_github.py', 'd3\_scale.js', 'fetch\_stock.js', and 'material.js'. Below the tabs is a code editor window containing the following JavaScript code:

```
1 const fetchNASAPictureOfDay = () => {
2   return fetch('https://api.nasa.gov/planetary/apod?api_key=DEMO_KEY', {
3     method: 'GET',
4     headers: {
5       'Content-Type': 'application/json',
6     },
7   })
8   .then(response => response.json())
9   .then(json => {
10     return json;
11   })
12 }
```

At the bottom left of the code editor is a GitHub Copilot logo.



---

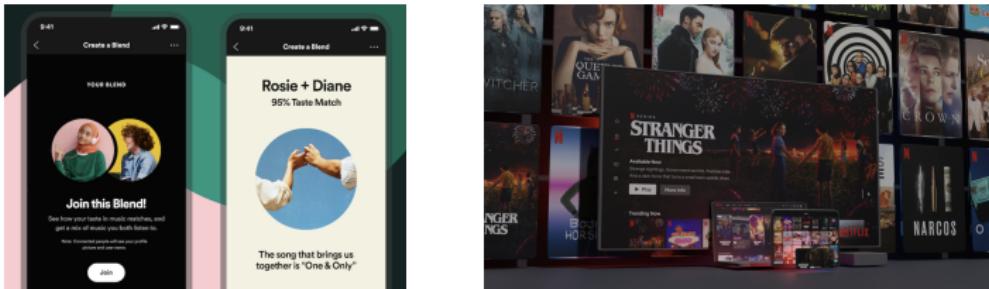
COPYWRITING    SUMMARIZATION    PARSING UNSTRUCTURED TEXT    CLASSIFICATION    TRANSLATION

---

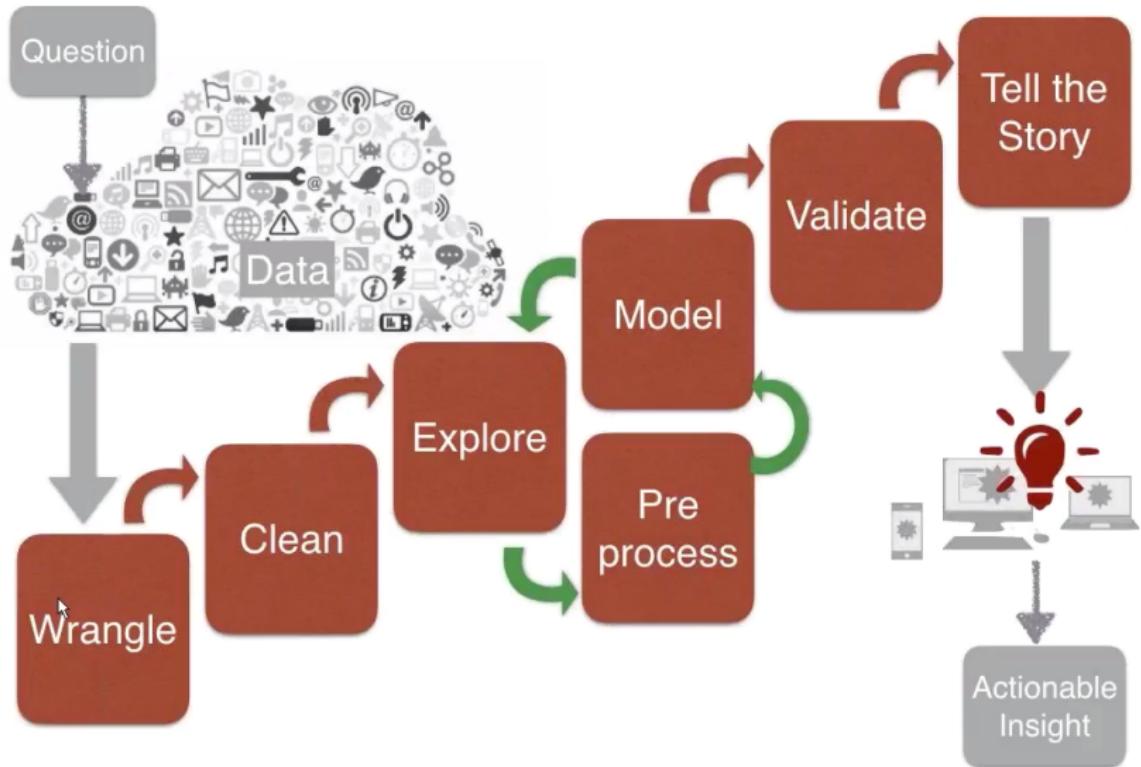
Create promo copy for the FamilyTime mobile application. It allows unlimited uploading, special filters and makes it easy to create albums of photos and videos. It runs on iOS and Android:

**FamilyTime is the perfect app to share and create memories with your family. FamilyTime is a free, unlimited photo and video editor that allows you to create albums of your memories. It has a special filter for the perfect summer look and is so simple to use.**

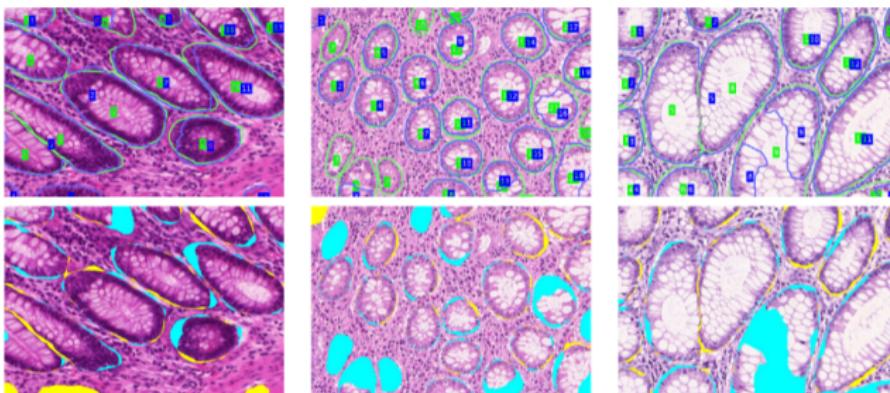
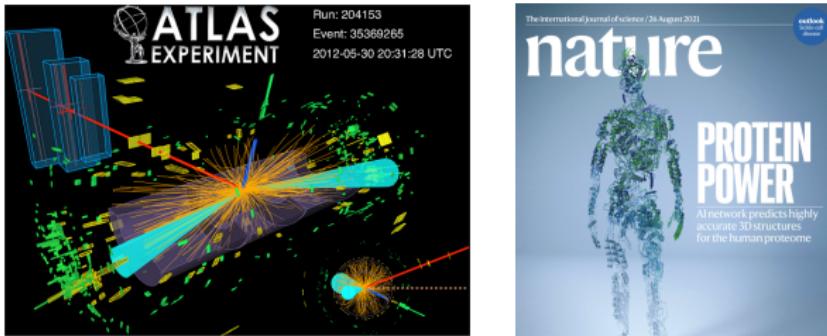
# Web, ads and recommendations



# Data science



# Physics, biology and medicine



(d) benign

(e) benign

(f) malignant

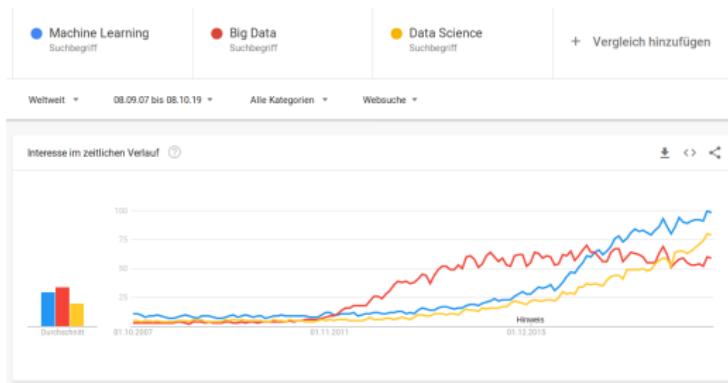
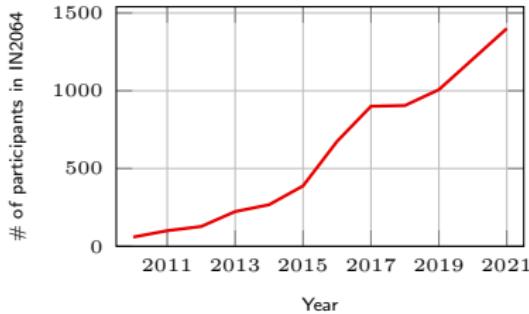
# What unites all these technologies?

- Computer vision
- Natural language processing
- Recommender systems
- Computational advertising
- Robotics
- Artificial intelligence
- Data science
- Bioinformatics
- Many other fields



All are using Machine learning

# Hot topic

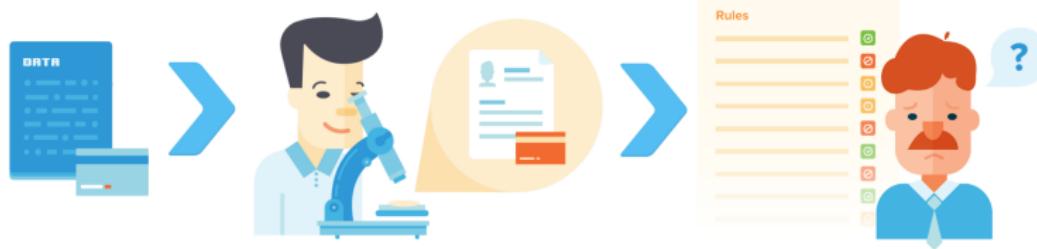


# What is Machine Learning?

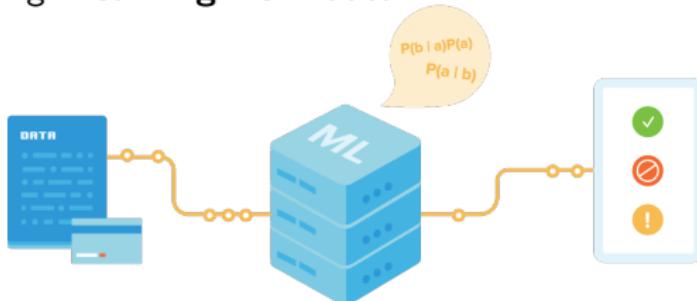
将交易分为合法交易和欺诈性交易。

Simple example - classify transactions into legitimate and fraudulent.

Rule-based approaches - rules handcrafted by human experts

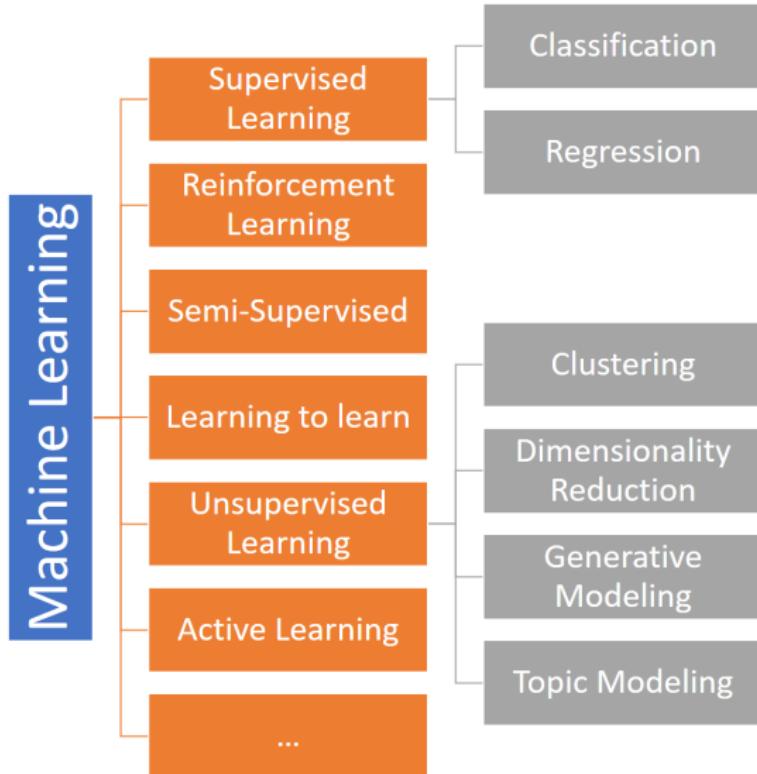


Machine learning - learning from data



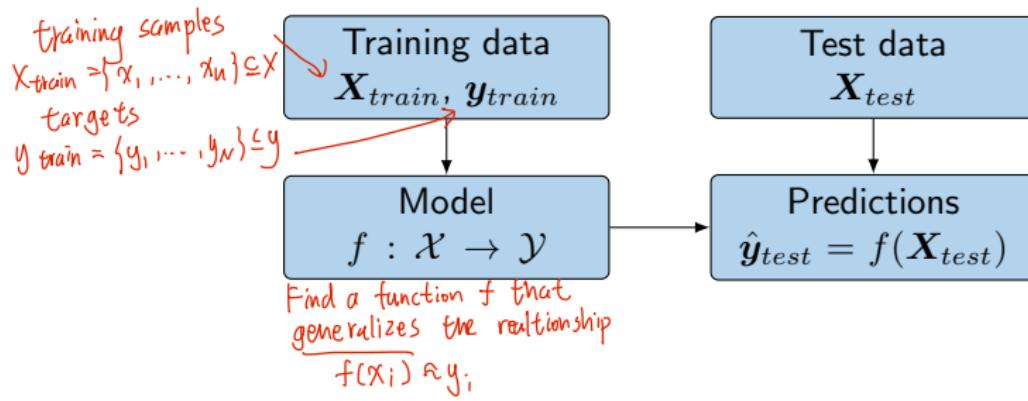
Figures adapted from <https://siftscience.com/sift-edu/prevent-fraud>

# Types of ML problems



# Supervised learning

- Given **training samples**  $\mathbf{X}_{train} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \mathcal{X}$
- with corresponding **targets**  $\mathbf{y}_{train} = \{y_1, \dots, y_N\} \subseteq \mathcal{Y}$
- Find a function  $f$  that generalizes this relationship, i.e.  $f(\mathbf{x}_i) \approx y_i$ .
- Using  $f$ , make predictions  $\hat{\mathbf{y}}_{test}$  for the **test data**  $\mathbf{X}_{test}$ .



# Supervised learning: Classification

If the targets  $y_i$  represent categories, the problem is called classification.

## Examples

- Handwritten digit recognition
- Transaction classification  
(**fraud**, **valid**)
- Object classification  
(cat, dog, hotdog, ...)
- Cancer detection



# Supervised learning: Regression

If the targets  $y_i$  represent **continuous numbers**, the problem is called regression.

## Examples

- Stock market prediction
- Demand forecasting
- User involvement measurement
- Revenue analysis

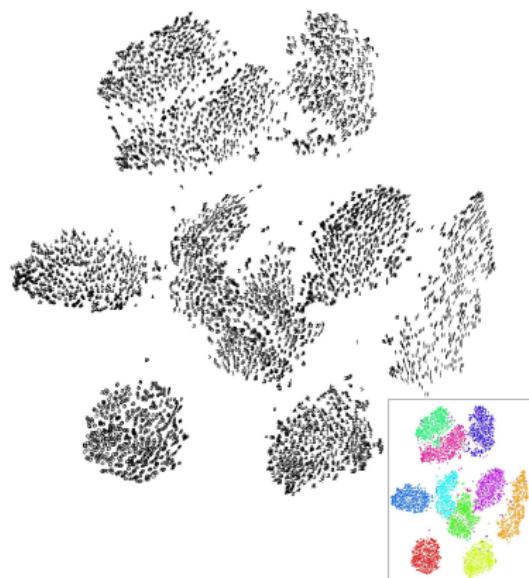


# Unsupervised learning

Unsupervised learning is concerned with finding structure in **unlabeled** data.

## Typical tasks

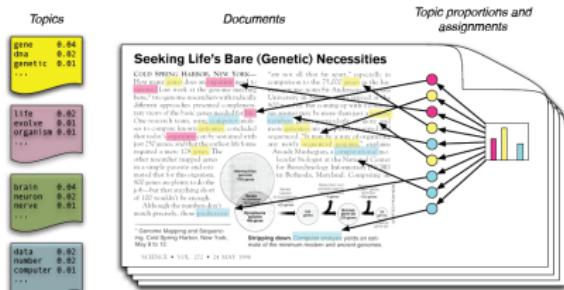
- Clustering
  - Group similar objects together
- Dimensionality reduction 降低维度
  - Project down high-dimensional data



# Unsupervised learning

## Typical tasks - continued

- Generative modeling
  - (Controllably) generate new "realistic" data
- Topic models
  - Discover hidden semantic structures in text.



TEXT PROMPT      an armchair in the shape of an avocado....

AI-GENERATED IMAGES



Edit prompt or view more images↓

# Other categories

- Reinforcement learning
  - Learning by interacting with a **dynamic environment**. Goal is to maximize **rewards** obtained by performing “desirable” actions.
- Semi-supervised learning
  - Learning to **combine lots of unlabeled data with a few labeled examples** for further prediction tasks.
- Active learning
  - Learn while obtaining labels by **querying an oracle**.
- Learning to learn (meta-learning)
  - Learning to **construct better models** for ML. Operates one level above the standard ML techniques.
- Learning to rank
  - What are the relevant items for a given query?  
(e.g. Netflix, web search, ad placement)
- And many more...

# General information

## Staff

- Lecturer: Prof. Dr. Stephan Günnemann
- Teaching assistants:  
Marin Bilos, Nicholas Gao, Simon Geisler, Lukas Gosch, Jan Schuchardt, Johanna Sommer
- Group website: [cs.cit.tum.de/daml/](http://cs.cit.tum.de/daml/)

## Details

- 8 ECTS, Language - English
- Open for both Bachelor and Master students
- Doesn't count for Wirtschaftsinformatik (Information Systems) students
- Doesn't stack with IN2332 in your curriculum

# Piazza

- Register at <http://piazza.com/tum.de/fall2022/in2064>  
Access code: ml2022
- All announcements will be made on Piazza.
- All course material will be uploaded to Moodle.
- You will miss important information if you don't register.

Use Piazza to ask questions - your emails will likely not be answered.

## Schedule & Logistics

- Lecture slides, lecture video and exercise sheet for topic of week  $N$  are uploaded before Monday of week  $N$ .
- Monday & Tuesday: lecture for week  $N$
- Wednesday morning: Q & A for week  $N$
- Wednesday afternoon: exercise for week  $N - 1$
- Exercise solutions are published on Moodle.

## Exam

- Written final exam, probably in February
- Most likely in-person exam
- 120 minutes
- One handwritten two-sided A4 sheet with notes

# Planned weekly schedule

Week	Date	Topic
1	Oct 17	Introduction, basic concepts
2	Oct 24	k-nearest neighbors, decision trees
3	Oct 31	Probabilistic inference
4	Nov 7	Linear regression
5	Nov 14	Linear classification
6	Nov 21	Optimization
7	Nov 28	No Lecture
8	Dec 5	Deep learning 1
9	Dec 12	Deep learning 2
10	Dec 19	Support Vector Machines
11	Jan 9	Dimensionality Reduction 1
12	Jan 16	Dimensionality Reduction 2
13	Jan 23	Clustering
14	Jan 30	Buffer

# Contents

- This is an introductory, **theoretical** Machine Learning course
  - There will be a fair amount of theory and mathematics
  - We will focus on fundamental Machine Learning concepts
  - We will mostly discuss independent (iid) data
- Next semester, we will cover (even) more advanced topics (IN2323)
  - Generative models
  - Robustness
  - Sequential data
  - Graphs & networks

→ These are the core research topics of our group :-)

# What this course is not about

- This is **not** a pure Deep Learning course
  - look at IN2346, IN2349 instead
- This is **not** a course about Big Data (Hadoop, etc.)
  - look at IN2326 instead
- This is **not** an applied Data Science / Business Analytics course
  - look at IN2028, IN2339 instead

# Recommended reading

Our official reading recommendation:

- Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Berlin, New York, 2006 (free, online version available).

but we also like:

- Kevin Murphy, *Machine Learning: A probabilistic perspective*. MIT Press, 2012.

# What's next?

Brush up on your linear algebra, calculus, and probability theory knowledge.

Read

- <http://cs229.stanford.edu/section/cs229-linalg.pdf>
- <http://cs229.stanford.edu/summer2020/cs229-prob.pdf>
- Bishop [ch. 1.2.0 - 1.2.3, 2.1 - 2.3.0]
- Solve the math refresher (exercise sheet 1)