Introduction to Data Science with Python

Würzburg, 24.02.2026

Arnim Bleier
arnim.bleier@gesis.org
http://gesis.org/person/arnim.bleier

Outline

- What is Machine Learning?
- Categories of Machine learning
  - Supervised learning
  - Unsupervised learning
  - Others are semi-supervised learning methods, reinforcement learning, recommender systems, ...
- A closer look at supervised learning
- A closer look at unsupervised learning
- Introduction to scikit-learn
- Working with data

What is Machine Learning?

- "Field of study that gives computers the ability to learn without being explicitly programmed" Arthur Samuel(1959)
- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." Tom Mitchell (1998)

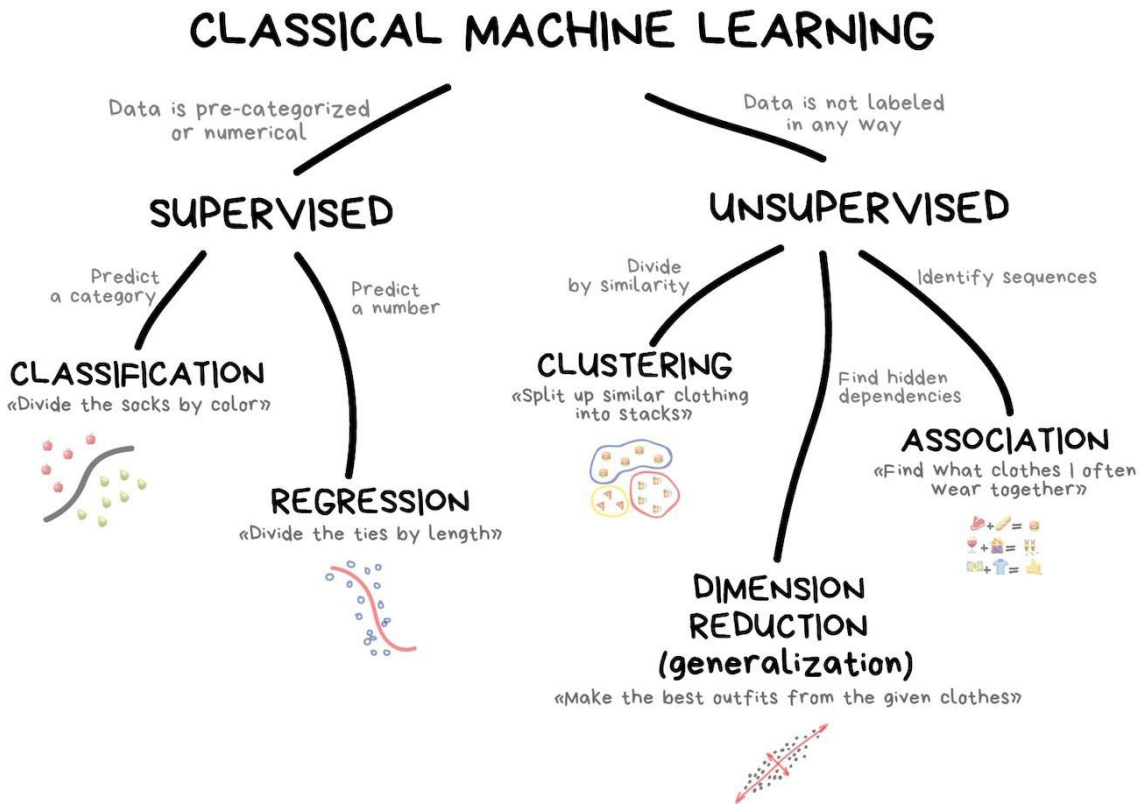
"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Your Mail programm "observes" which mails you do or do not classify as spam, and uses these observations to learn how to better filter spam messages. What is the task in this setting?

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

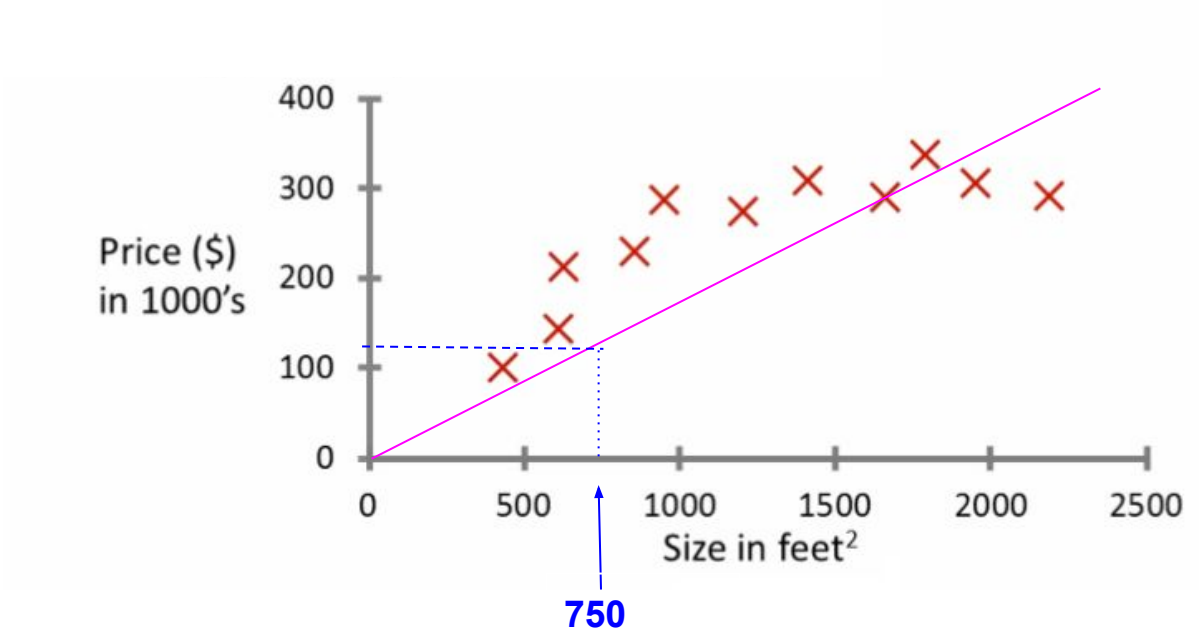
Your Mail programm "observes" which mails you do or do not classify as spam, and uses these observations to learn how to better filter spam messages. What is the task in this setting?

Classifying mails into the categories “spam” or “not spam”



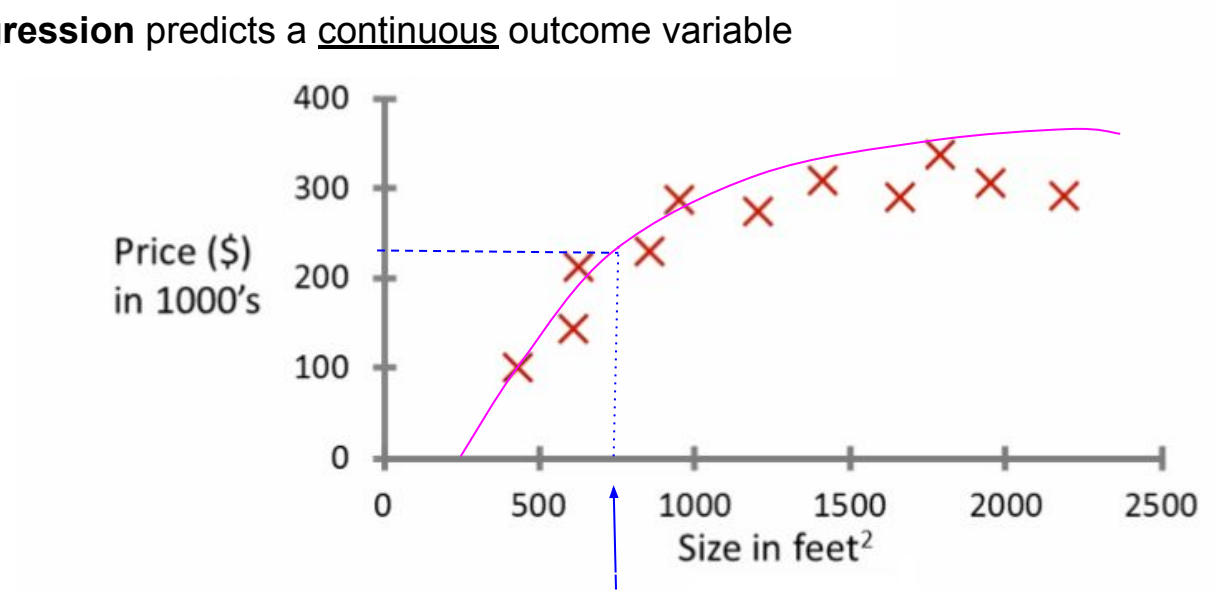
https://vas3k.com/blog/machine_learning/

Supervised learning



Supervised learning

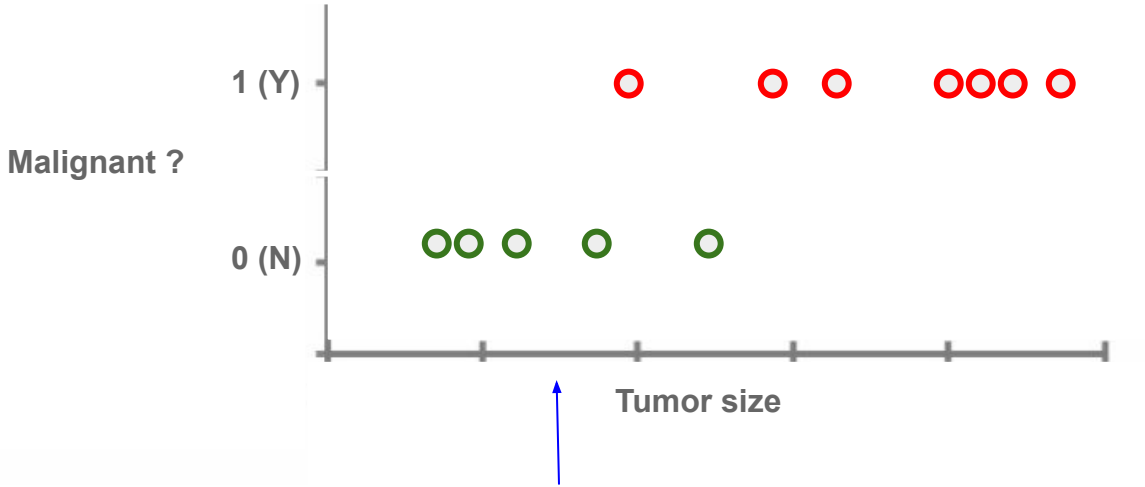
Regression predicts a continuous outcome variable



Requires training data

Supervised learning

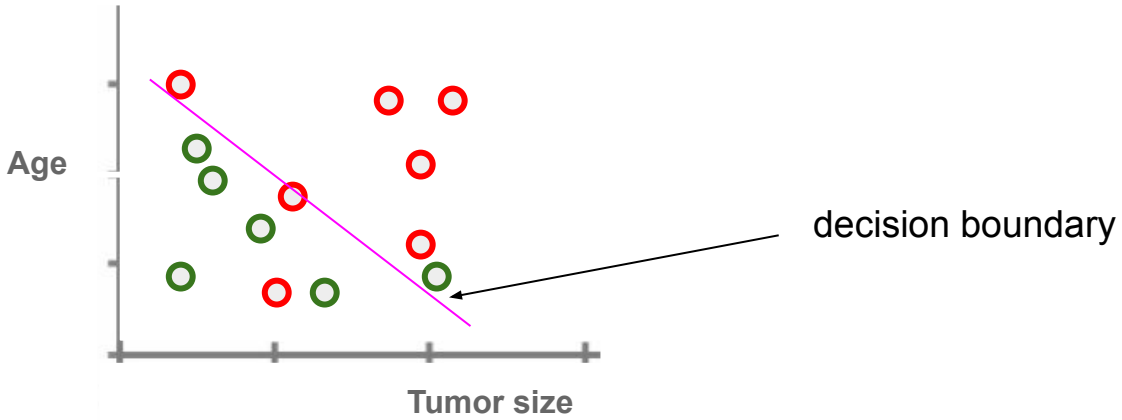
Classification predicts a discrete outcome variable



Requires training data

Supervised learning

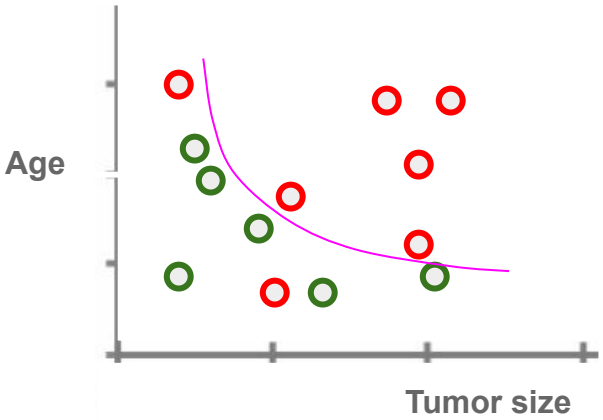
Classification predicts a discrete outcome variable



Requires training data

Supervised learning

Classification predicts a discrete outcome variable

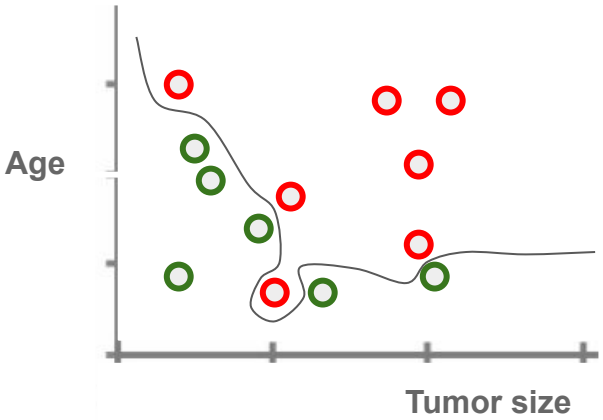


Model complexity

Requires training data

Supervised learning

Classification predicts a discrete outcome variable



Model complexity

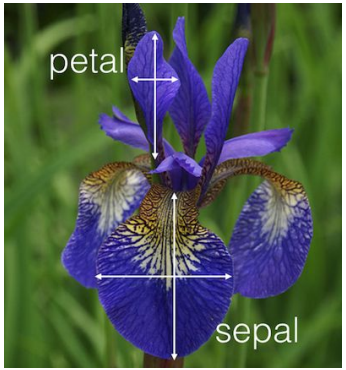
Requires training data

Supervised learning

The Iris flower data set

Features:

- sepal length
- sepal width
- petal length
- petal width

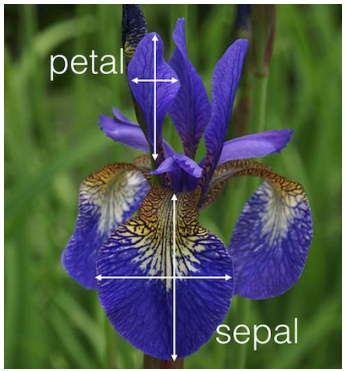


Supervised learning

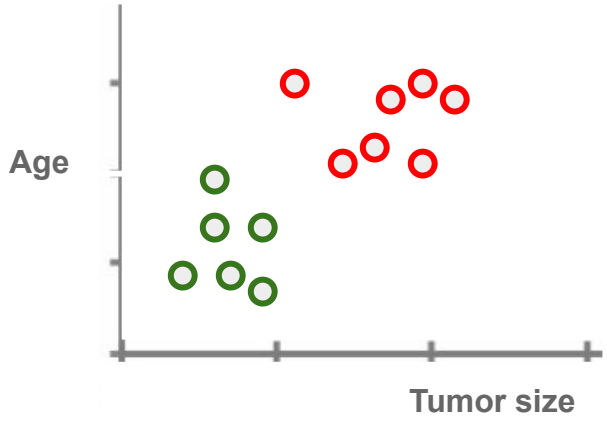
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Features:

- sepal length
- sepal width
- petal length
- petal width

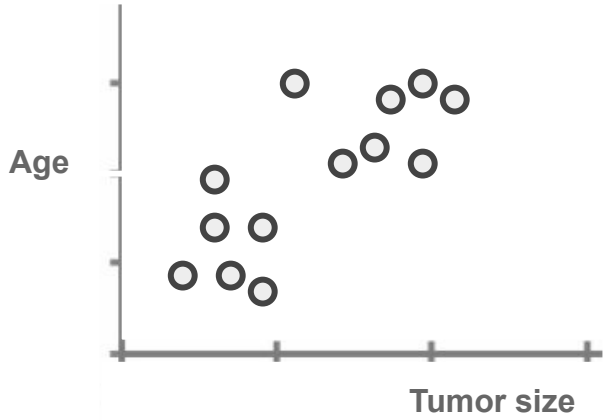


Supervised learning



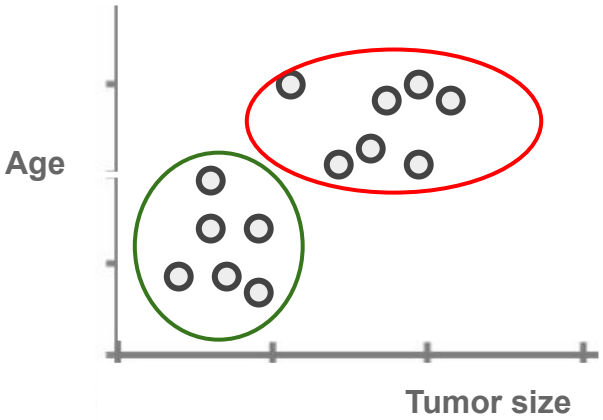
Requires training data

Unsupervised learning



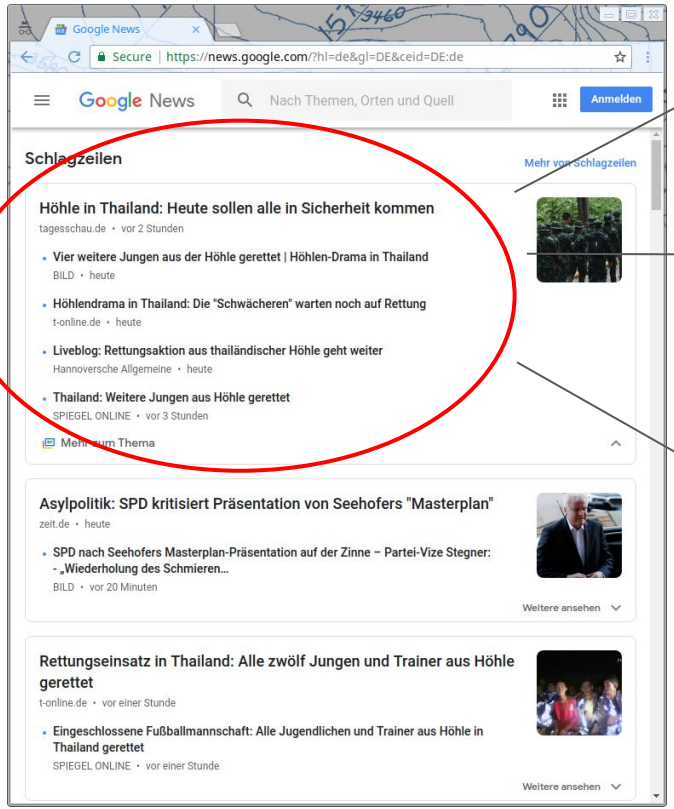
In the absence of training data

Unsupervised learning

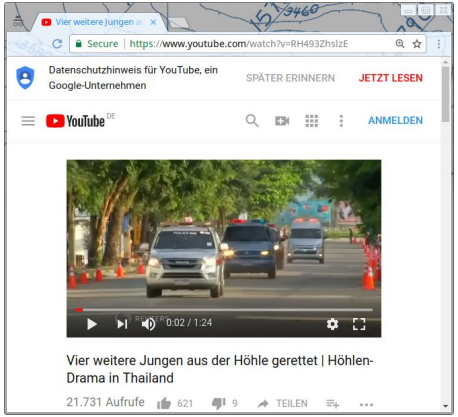


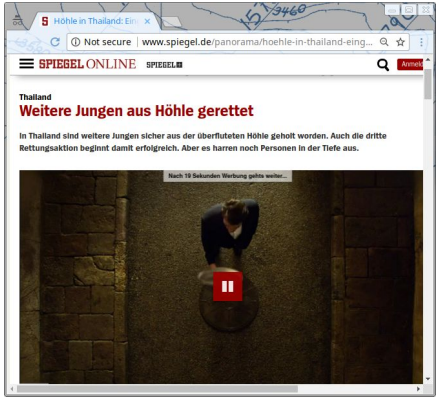
In the absence of training data

Example 1

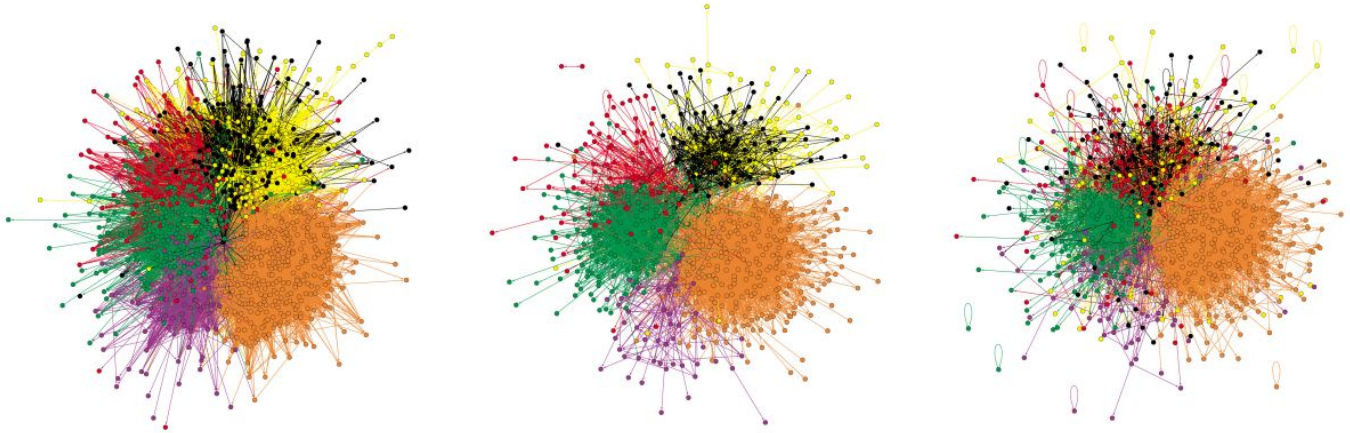

Google News
Secure | https://news.google.com/?hl=de&gl=DE&ceid=DE:de
Google News
Nach Themen, Orten und Quell
Anmelden
Schlagzeilen
Mehr von Schlagzeilen
Höhle in Thailand: Heute sollen alle in Sicherheit kommen
tagesschau.de • vor 2 Stunden
Vier weitere Jungen aus der Höhle gerettet | Höhlen-Drama in Thailand
BILD • heute
Höhlendrama in Thailand: Die "Schwächeren" warten noch auf Rettung
t-online.de • heute
Liveblog: Rettungsaktion aus thailändischer Höhle geht weiter
Hannoversche Allgemeine • heute
Thailand: Weitere Jungen aus Höhle gerettet
SPIEGEL ONLINE • vor 3 Stunden
Mehr zum Thema
Asylpolitik: SPD kritisiert Präsentation von Seehofers "Masterplan"
zeit.de • heute
SPD nach Seehofers Masterplan-Präsentation auf der Zinne – Partei-Vize Stegner: - „Wiederholung des Schmieren...
BILD • vor 20 Minuten
Weitere ansehen
Rettungseinsatz in Thailand: Alle zwölf Jungen und Trainer aus Höhle gerettet
t-online.de • vor einer Stunde
Eingeschlossene Fußballmannschaft: Alle Jugendlichen und Trainer aus Höhle in Thailand gerettet
SPIEGEL ONLINE • vor einer Stunde
Weitere ansehen


Secure | https://www.tagesschau....
ARD Navigation
ARD
tagesschau.de
Menü
Startseite
Ausland
Höhle in Thailand, Heute sollen alle in Sicherheit kommen
Höhle in Thailand
Heute sollen alle in Sicherheit kommen
Stand: 10.07.2018 13:21 Uhr
In Thailand ist eine neue Rettungsaktion aus der Höhle für die noch verbliebenen Jungen und ihren Trainer im Gange. Und es soll die letzte Aktion sein. Das Ziel ist, dass bis zum Abend alle in Sicherheit sind.


Secure | https://www.youtube.com/watch?v=RH493ZhsIzE
Datenschutzhinweis für YouTube, ein Google-Unternehmen
SPÄTER ERINNERN
JETZT LESEN
YouTube
ANMELDEN
0:02 / 1:24
Vier weitere Jungen aus der Höhle gerettet | Höhlen-Drama in Thailand
21.731 Aufrufe
621
9
TEILEN


Not secure | www.spiegel.de/panorama/hoehle-in-thailand-eing...
SPIEGEL ONLINE
Thailand
Weitere Jungen aus Höhle gerettet
In Thailand sind weitere Jungen sicher aus der überfluteten Höhle geholt worden. Auch die dritte Rettungsaktion beginnt damit erfolgreich. Aber es harren noch Personen in der Tiefe aus.
Nach 10 Sekunden Werbung gehts weiter...

Example 2



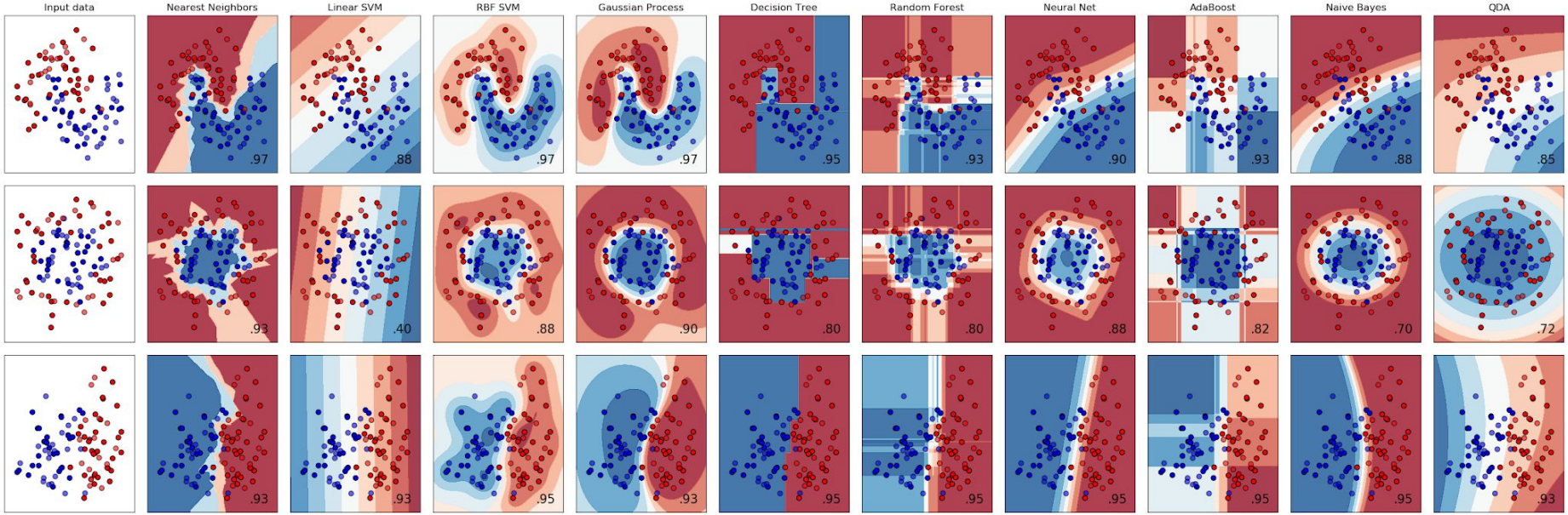
(a) Following (H=0.83) (b) Retweeting (H=0.90) (c) Mentioning (H=0.79)

Examples of online conversational practices on Twitter: Structures of the aggregate following, retweeting, and mentioning networks of German politicians from 9 weeks before to 4 weeks after the federal election 2013. The vertices in the networks correspond to user handles and are color-coded by party affiliation (colors given in Table 1). Arcs correspond to following/retweeting/mentioning relationships.

Unsupervised learning

Which of the following problems would you address using an supervised algorithm?

- ❏ Given emails labeled as spam / not spam,learn a classifier.
- ❏ Grouping news articles into sets covering the same story / event.
- ❏ Given a database of patients diagnosed with cancer learn to classify new patients into having malignant or non malignant cancer.



https://scikit-learn.org/

Sources

https://wingshore.wordpress.com/2014/11/01/what-is-supervised-learning/
https://www.wikipedia.org/
http://cs229.stanford.edu/syllabus.html
H. Lietz, C. Wagner, A. Bleier, and M. Strohmaier. When politicians talk: Assessing online conversational practices of political parties on twitter. In Proceedings of the International Conference on Weblogs and Social Media, ICWSM 2014,