The Best of both Worlds: Challenges in Linking Provenance and Explainability in Distributed Machine Learning

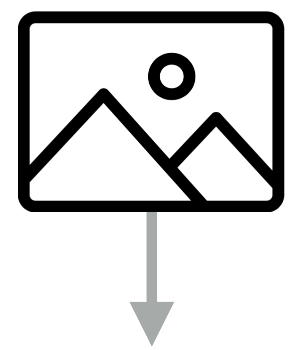
Stefanie Scherzinger, OTH Regensburg, Germany
Christin Seifert, University of Twente, Netherlands
Lena Wiese, Leibniz University Hannover, Hanover, Germany

@ICDCS, Dallas, TX, 2019-07-09

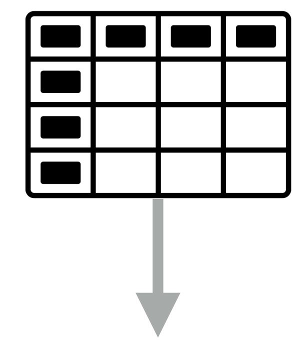
End-2-End Explanations

Explainable AI (ML)

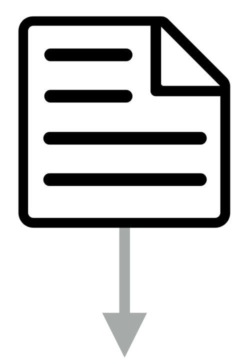
- Ubiquitous AI, Algorithmic Accountability [1]
- GDPR “right to explanation” [2]
- IEEE, ACM Code of Ethics: “Be fair and take action not to discriminate.” [3]



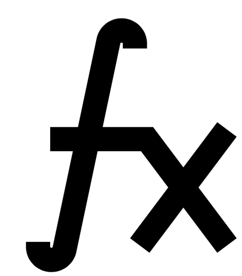
mountains, sunset



not creditworthy



positive

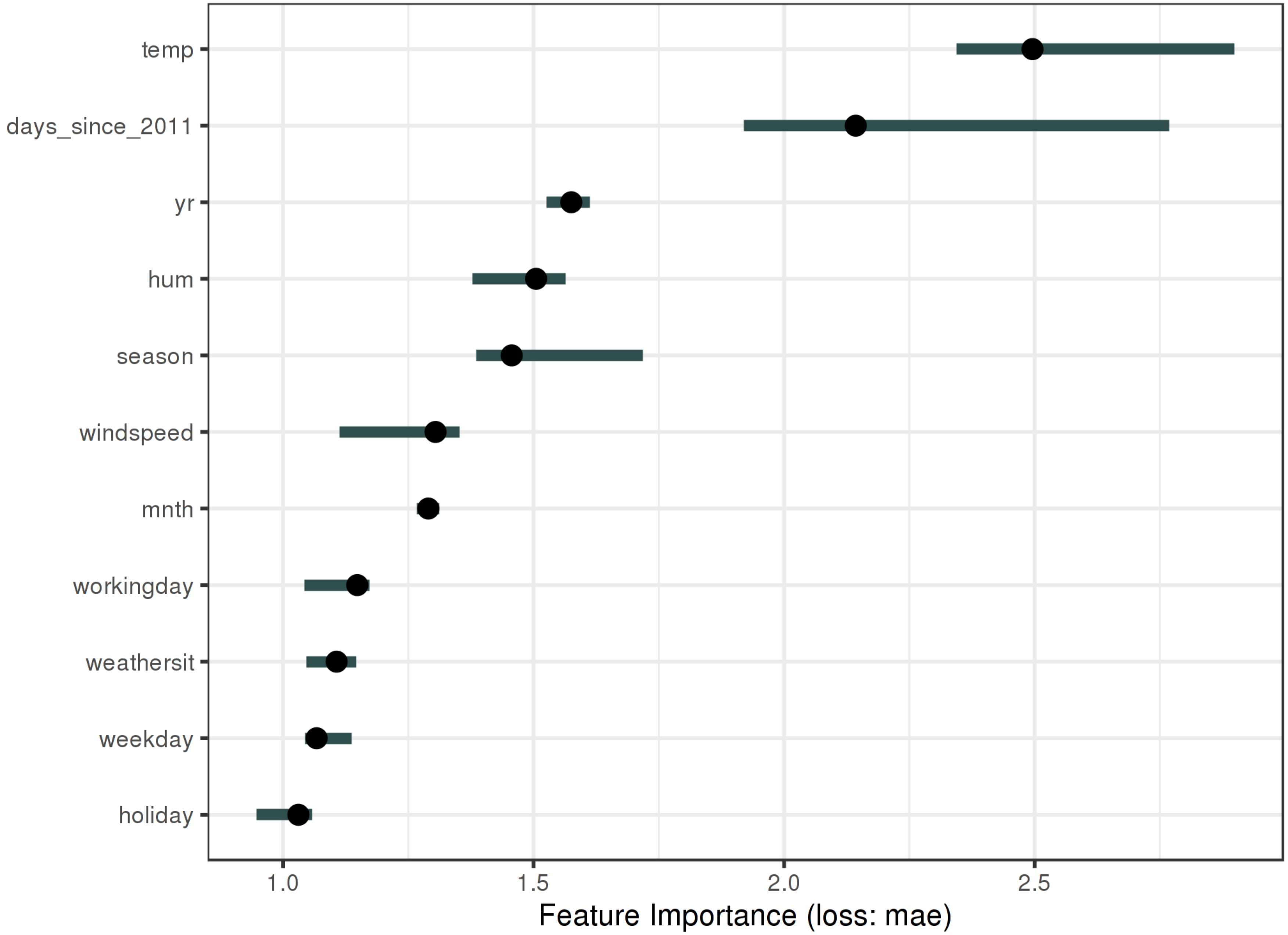


[1]N. Diakopoulos, “Accountability in algorithmic decision making,” *Commun. ACM*, vol. 59, no. 2, pp. 56–62, Jan. 2016.
[2] B. Goodman and S. Flaxman, “European Union regulations on algorithmic decision-making and a “right to explanation”,” *ArXiv e-prints*, Jun. 2016.
[3]https://www.acm.org/code-of-ethics

End-2-End Explanations

Feature Based Explanations

- predict number of rented bikes



source: https://christophm.github.io/interpretable-ml-book/feature-importance.html

End-2-End Explanations

Counterfactual Explanations

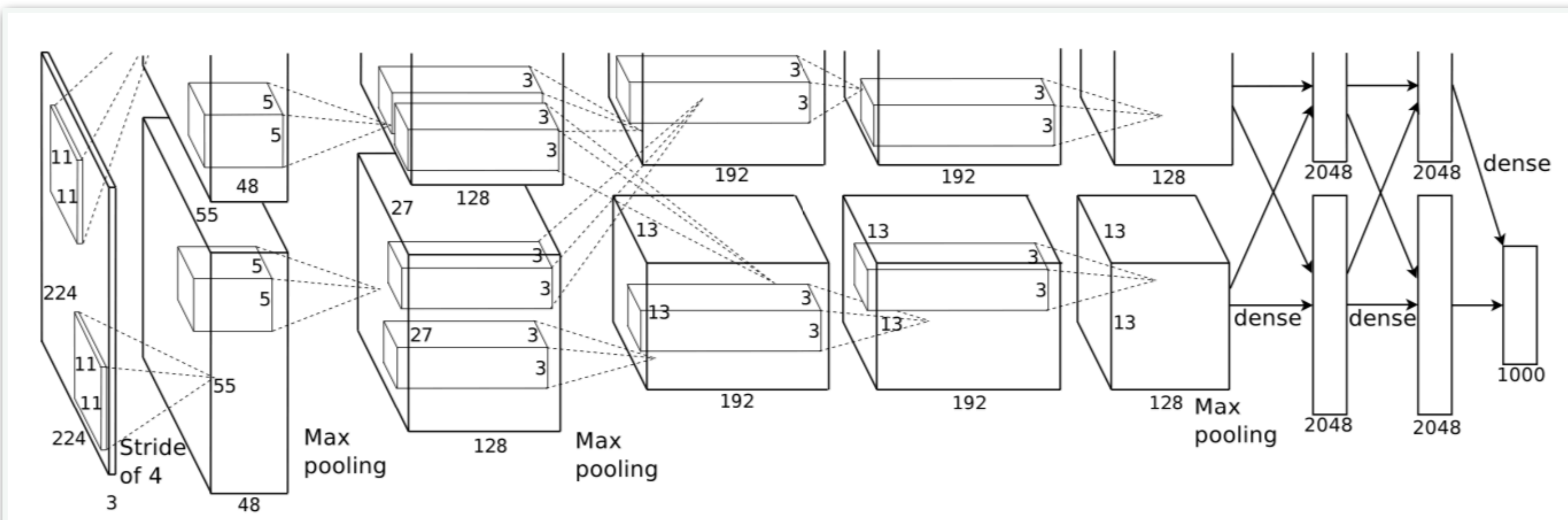
- predict a student's average grade of the first year at law school
- features: grade point average (GPA) prior to law school, race and law school entrance exam scores (SAT score)
- "What needs to be changed to get a score of "0" (average)?

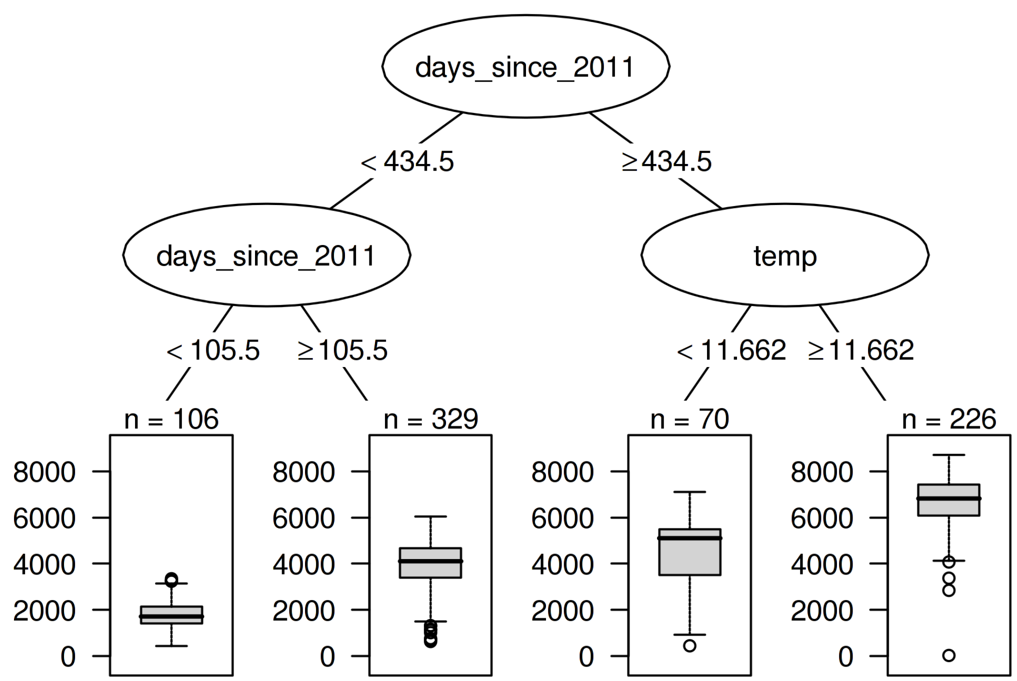
Score	GPA	LSAT	Race	GPA x'	LSAT x'	Race x'
0.17	3.1	39.0	0	3.1	34.0	0
0.54	3.7	48.0	0	3.7	32.4	0
-0.77	3.3	28.0	1	3.3	33.5	0
-0.83	2.4	28.5	1	2.4	35.8	0
-0.57	2.7	18.3	0	2.7	34.9	0

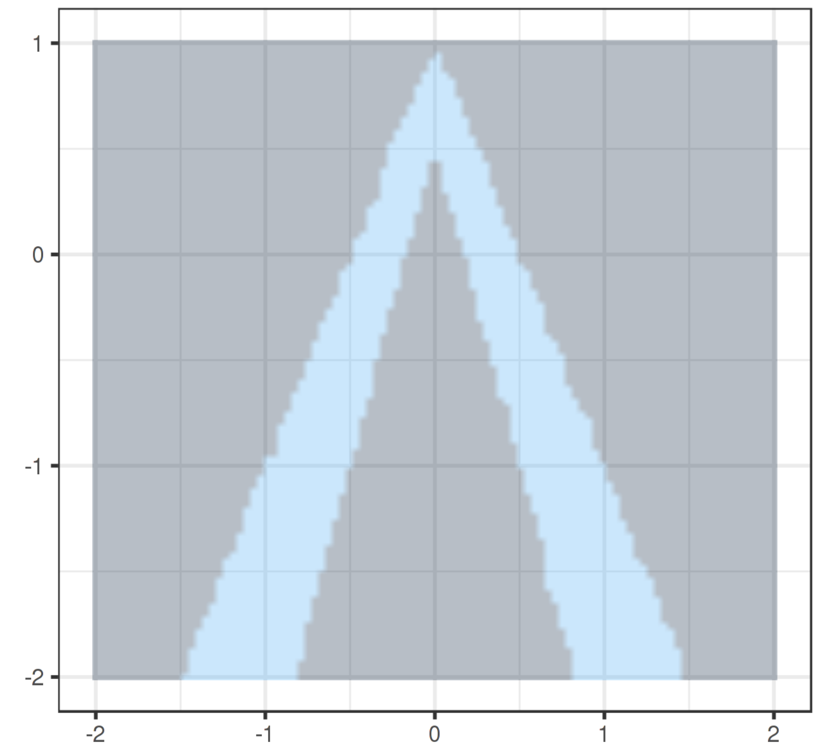
source: https://christophm.github.io/interpretable-ml-book/counterfactual.html#counterfactual

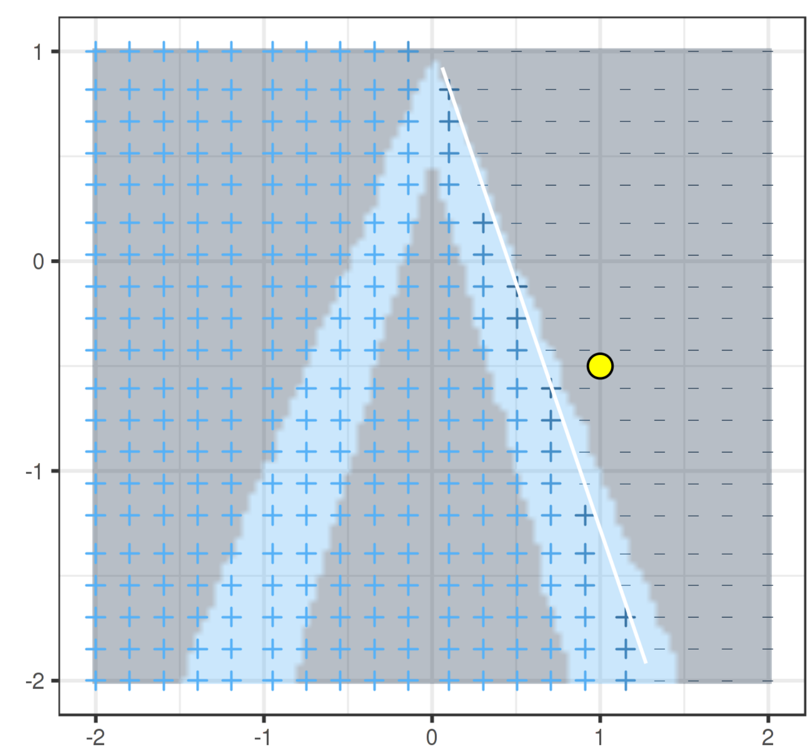
End-2-End Explanations

Model-based explanations









https://christophm.github.io/interpretable-ml-book/lime.html#lime-for-tabular-data

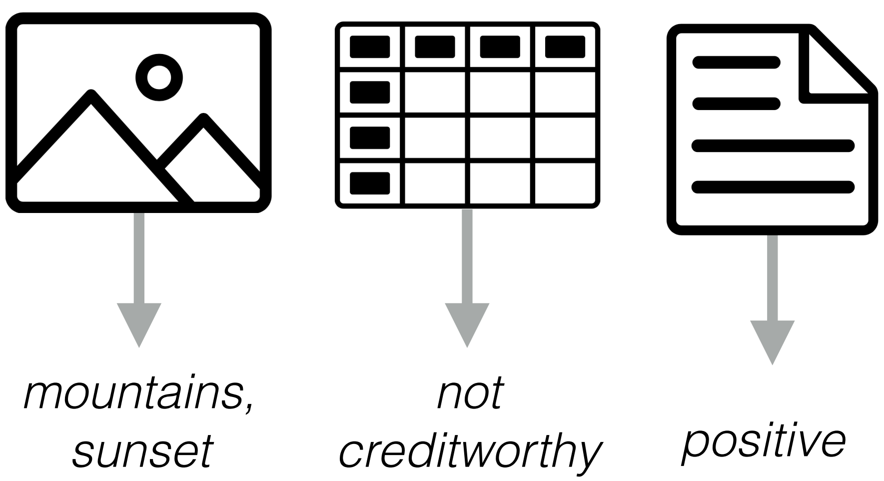
Krizhevsky, Alex. "ImageNet Classification with Deep Convolutional Neural Networks" (PDF). Retrieved 17 November 2013.

End-2-End Explanations

Explainable AI	Distributed processing

Explainable AI

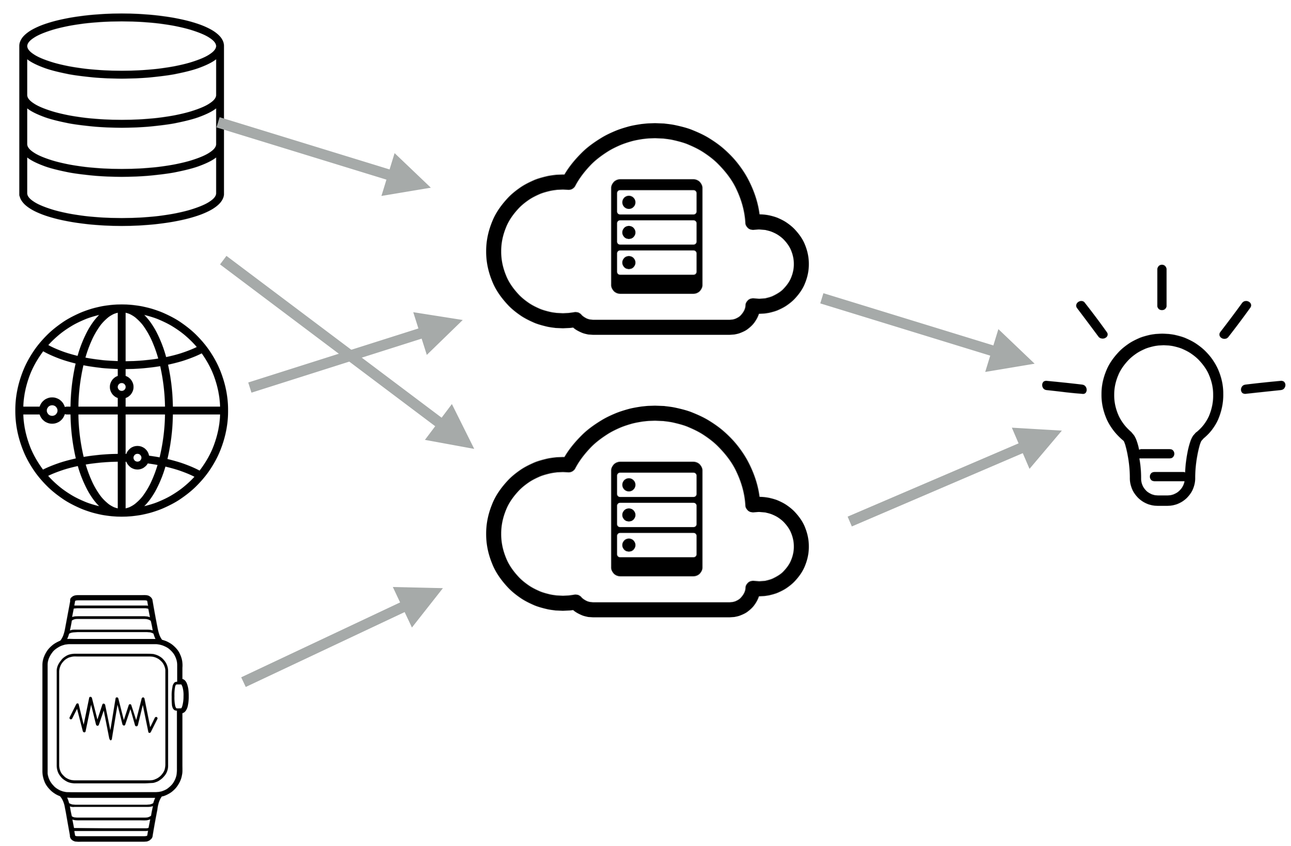
- Assumes clean data sets



explanations are not truthful

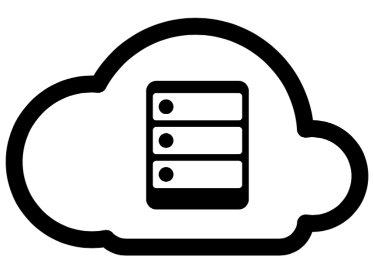
Distributed processing

- distributed data
  - distributed Megasets (Anja)
- distributed processing
  - on the edge (Marilyn), in the cloud, fog
- pre-processing

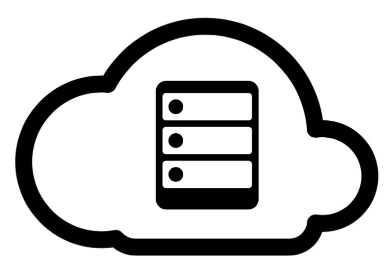


Illustrative Example

Name (N)	Age (A)	Pizzas (P)	Sport (S)	Fit (F)
		TRAINING DATA		
Amy	35	0	1	1
Bob	20	2	1	1
Charlie	32	2	0	0
Dave	*null*	5	*null*	"N"
Eve	24	*null*	1	"0"
Francis	35	0	1	1
Greg	20	0	1	1
Haley	32	2	0	0
		TEST DATA		
Zoe	40	7	1	?

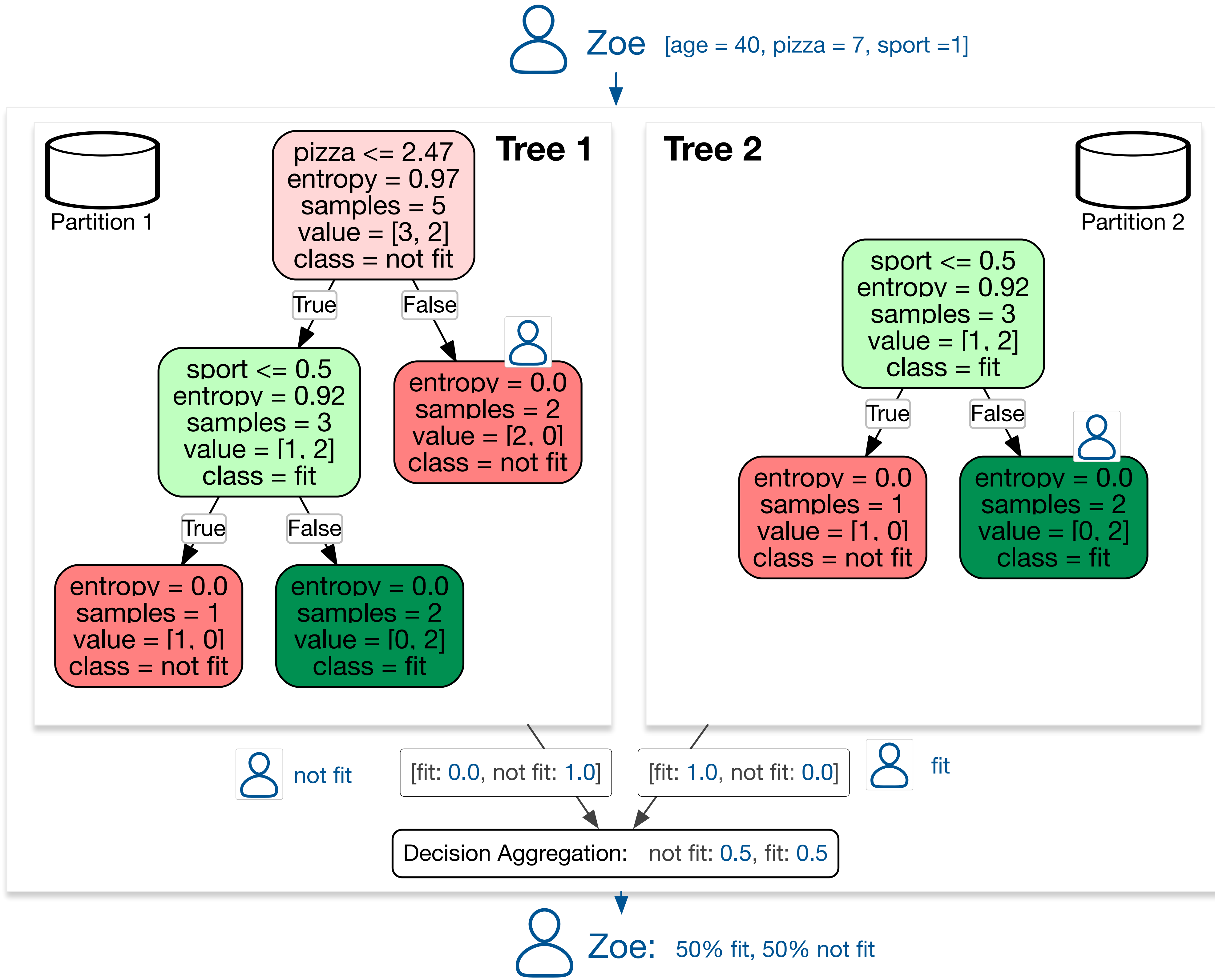


N	A	P	S	F
A	35.	0.	1.	1
B	20.	2.	1.	1
C	32.	2.	0.	0
D	23.04	5.	0.68	0
E	24.	2.94	1.	0



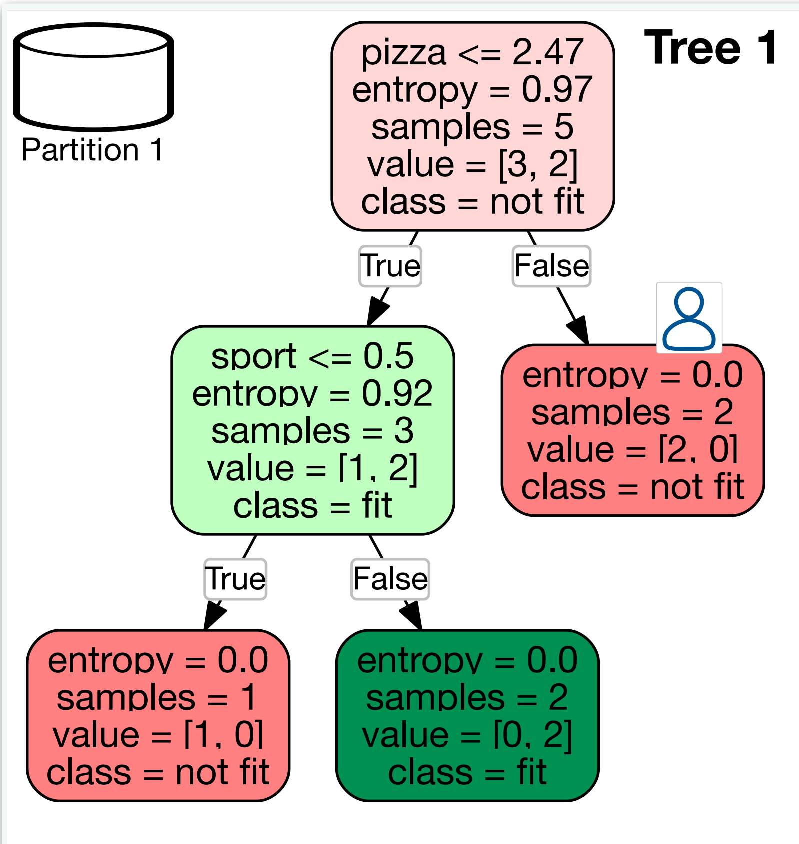
N	A	P	S	F
F	35.	0.	1.	1
G	20.	4.	1.	1
H	32.	2.	0.	0

Illustrative Example



how to make the decision? which to trust more? how to explain the decision?

Illustrative Example

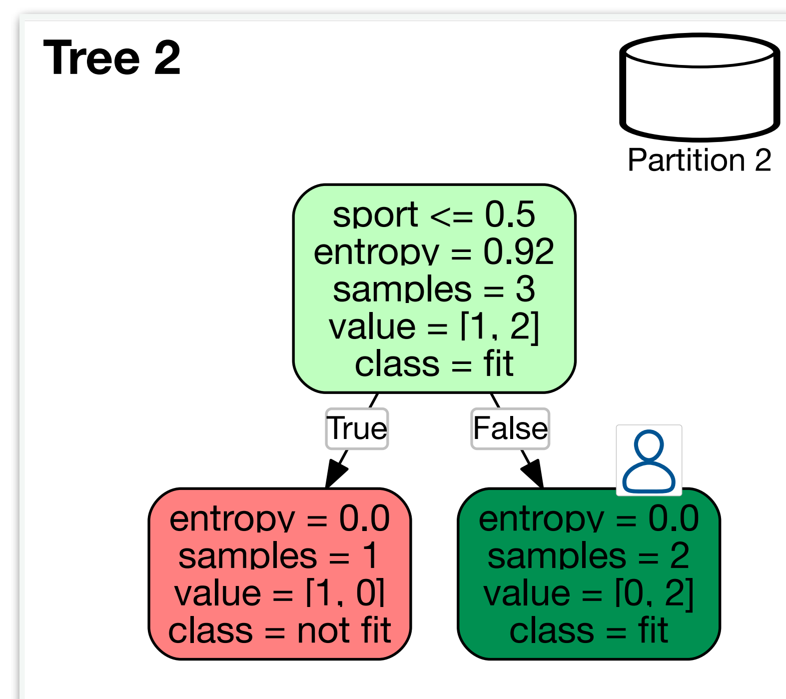


N	A	P	S	F
A	35.	0.	1.	1
B	20.	2.	1.	1
C	32.	2.	0.	0
D	23.04	5.	0.68	0
E	24.	2.94	1.	0

normalized and imputed data

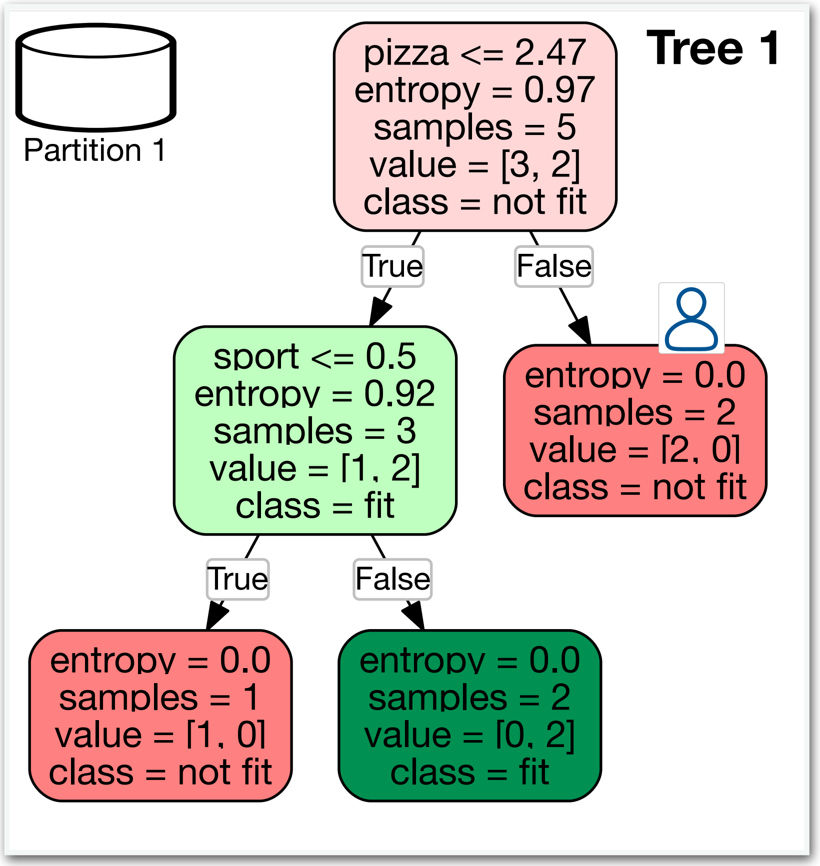
N	A	P	S	F
F	35.	0.	1.	1
G	20.	4.	1.	1
H	32.	2.	0.	0

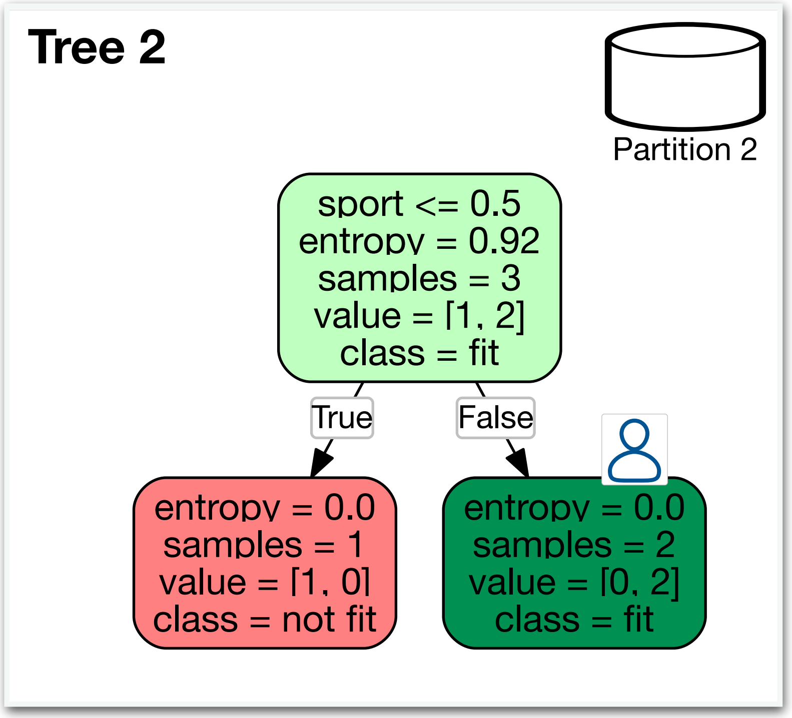
low sample size



Illustrative Example

Model-based explanation






```
IF (pizza <=2.47)
   AND (sport > 0.5)
THEN FIT
```

```
IF (sport > 0.5)
THEN FIT
```

Can I eat pizza?

[more types on explanations in the paper]

Challenges (and Solutions)

Access to Provenance Information

for truthful explanations, we would like to guarantee that all data processing steps are repeatable, and we also have all information on *model provenance*

Provenance information in machine learning libraries

Data Volume

provenance for ML algorithms adds to data volume

Efficient storage and querying of provenance information (e.g., HDFS)

Provenance Granularity

different levels of provenance are necessary (example: in first table data imputation needs to be tracked vs. second table had not imputed values)

Intelligent adaption of level of granularity for provenance data

Bias and Fairness

biased data distributions that do not follow the general trend; Simpson's paradox (e.g. model in table 1 is less accurate for females)

Bias-aware ML algorithms (statistical comparison across machines)

similar to reproducibility of experiments (Logan et al., ICDCS2019)

Challenges (and Solutions)

Provenance Visualization

provenance information needs to be accessible to humans (e.g. ML developers)

Solutions for high-dimensional data from VIS and HCI community

Variability and Lack of Standards

not clear yet which provenance data needs to be tracked; different data base standards, integration systems, ML libraries

Communities need to agree on data exchange format

Data Freshness

more recent data might be more important for ML models, in distributed setting stale data is more likely

Track provenance for data creation and modification

Data Protection and Privacy

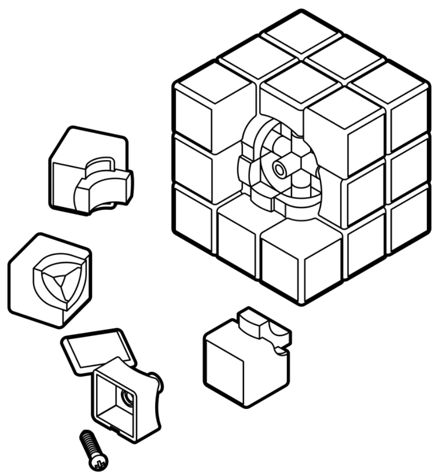
provenance tracking might be a data privacy breech for (some or all) nodes in a distributed setting

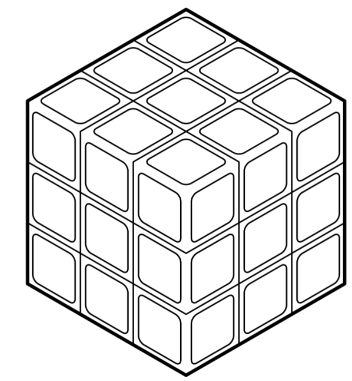
Trade-offs between anonymization and providing provenance data + ML explainability

Summary

Without knowledge about data and model provenance we are unable to truthfully explain and assess the trustworthiness of the resulting machine learning decision.

- **Access to Provenance Information**
- **Provenance Granularity**
- **Provenance Visualization**
- **Data Freshness**
- **Data Volume**
- **Bias and Fairness**
- **Variability and Lack of Standards**
- **Data Protection and Privacy**





Contact:

stefanie.scherzinger@oth-regensburg.de

c.seifert@utwente.nl

wiese@l3s.de