Data Science ≠ Machine Learning

Prof. Dr. Jens Dittrich

bigdata.uni-saarland.de
daimond.ai
twitter.com/jensdittrich

April 19, 2018
Data Science (one possible View)

- Application Domain
- Machine Learning
- A.I.
- Big Data Management
- Data Mining
- Statistics

Data Science
The Data Science Cake

Ingredients:
- 50g statistics
- 120g linear algebra
- 200g programming
- 1kg visualisation
- 300g software engineering

Additional skills:
- creativity
- out of the box thinking
- grit
- team spirit

Artificial Intelligence/Machine Learning
Data Management
Data Mining
Application Domain

© istock.com sasilsolutions
Data Science = Three Thirds

**Definition (Data Science)**

Data Science =

\[
\frac{1}{3} \text{ Artificial Intelligence (⊃ Machine Learning)} + \\
\frac{1}{3} \text{ Data Mining} + \\
\frac{1}{3} \text{ Data Management.}
\]

**Preach and get involved!**

It is our job as a community to spread the word about this and get involved! Otherwise we will again witness the reinvention of the wheel (e.g. like in NoSQL).
Opportunity!

In Data Science there is tremendous opportunity for data management.

Translates to:

In Data Science there is tremendous opportunity for us!
The Data Science-Pipeline/Waterfall Model

This is **at the same time** a process model and a dataflow.
Data Collection through data cleaning has a single goal here: enable AI&ML
Data Curation/OLAP as the Goal

Data Collection through data cleaning has a single goal here: enable data curation/OLAP
Alternative: DBMS as an Intermediate Tool

In principle, this is possible at any step.

Prof. Dr. Jens Dittrich

Data Science ≠ Machine Learning
Alternative: DBMS as an Intermediate Tool

Examples: DeepDive, HoloClean,
other steps: MonetDB/Tensorflow marriage
any data transformation: relational algebra-style, e.g. Pandas, stateless DBMS, Spark/Flink, etc.
Summary: Data Science = Three Thirds

Definition (Data Science)

Data Science =

\[ \frac{1}{3} \text{ Artificial Intelligence (⊃ Machine Learning)} + \]
\[ \frac{1}{3} \text{ Data Mining} + \]
\[ \frac{1}{3} \text{ Data Management}. \]

Preach and get involved!

It is our job as a community to spread the word about this and get involved! Otherwise we will again witness the reinvention of the wheel (e.g. like in NoSQL).