

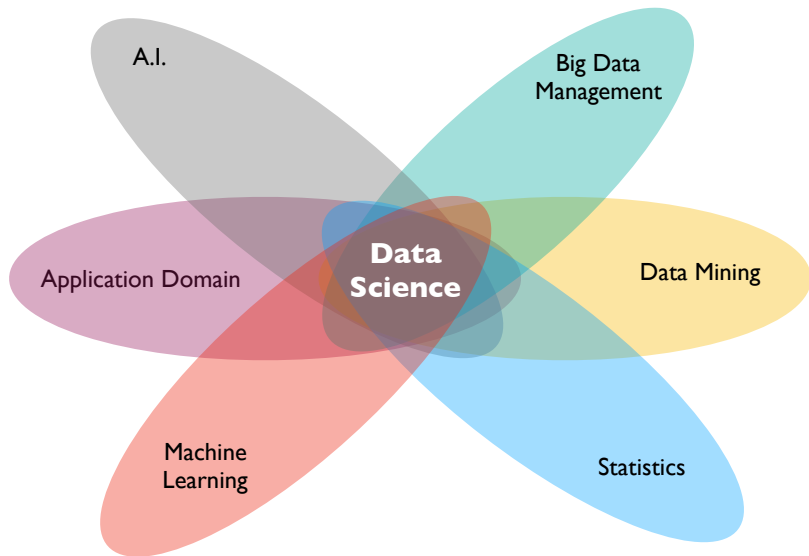
Data Science \neq Machine Learning

Prof. Dr. Jens Dittrich

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

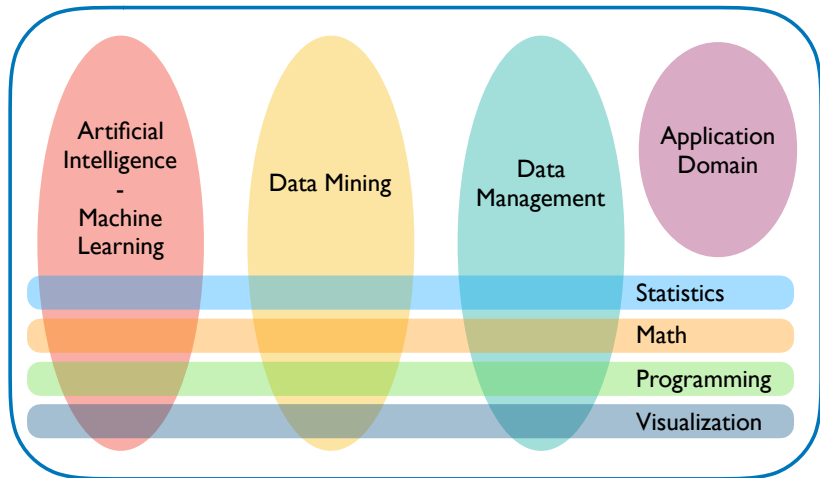
April 19, 2018

Data Science (*one possible View*)



Data Science (*another View*)

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

Data Science = Three Thirds

Definition (Data Science)

Data Science =

1/3 Artificial Intelligence (\supset Machine Learning) +

1/3 Data Mining +

1/3 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).

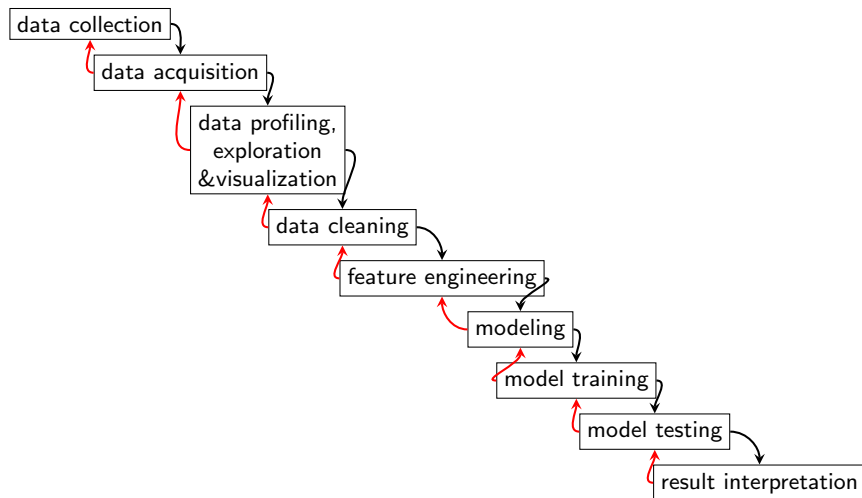
Opportunity!

In Data Science there is tremendous opportunity for **data management**.

Translates to:

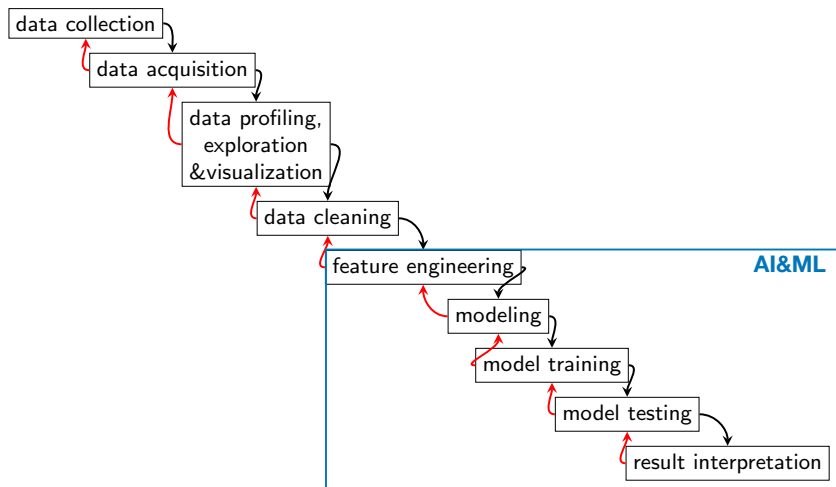
In Data Science there is tremendous opportunity for **us!**

The Data Science-Pipeline/Waterfall Model



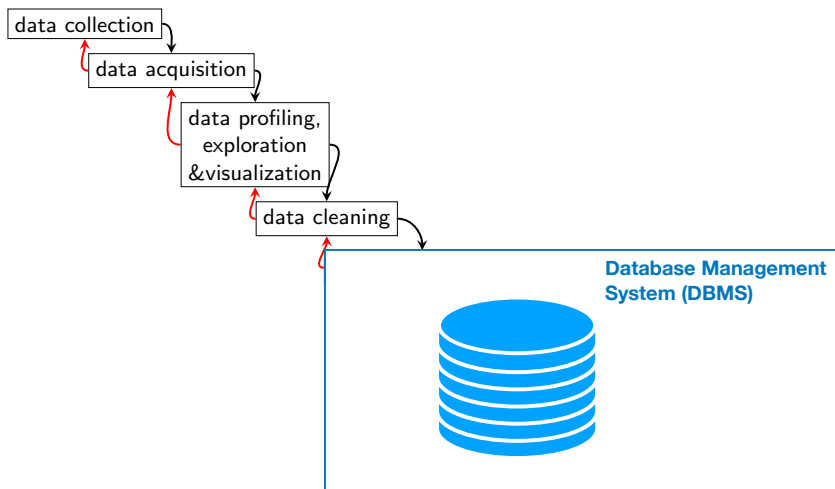
This is **at the same time** a process model **and** a dataflow.

AI/ML as the Goal



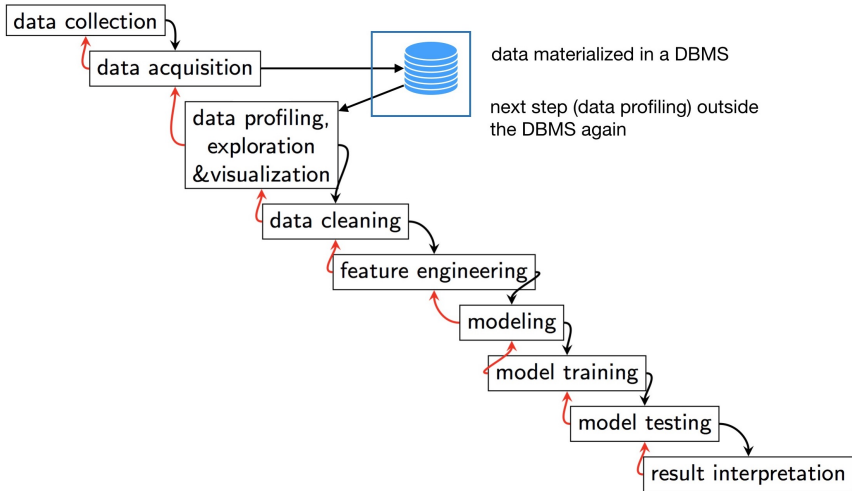
Data Collection through data cleaning has a single goal here:
enable AI&ML

Data Curation/OLAP as the Goal



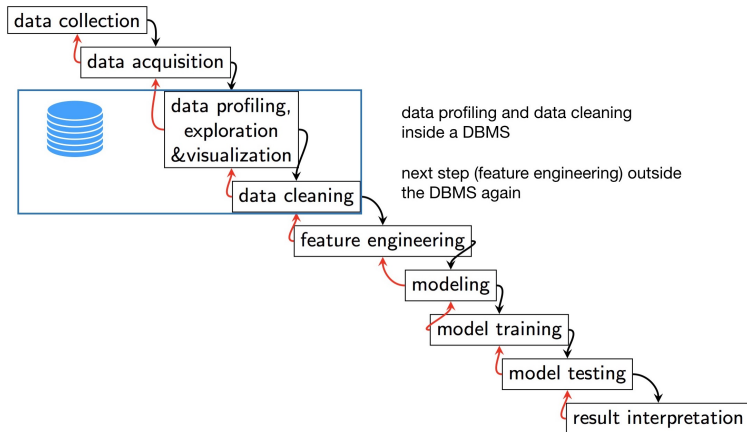
Data Collection through data cleaning has a single goal here:
enable data curation/OLAP

Alternative: DBMS as an Intermediate Tool



In principle, this is possible at any step.

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 =

1/3 Artificial Intelligence (\supset Machine Learning) +

1/3 Data Mining +

1/3 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).