DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

The Role of Datasets in AI Platforms: Examples from AI4EU and TRUSTS

27.05.2021

Dr. Dominik Kowald
Know-Center GmbH

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION

DECSIS ubiwhere valuedate.io

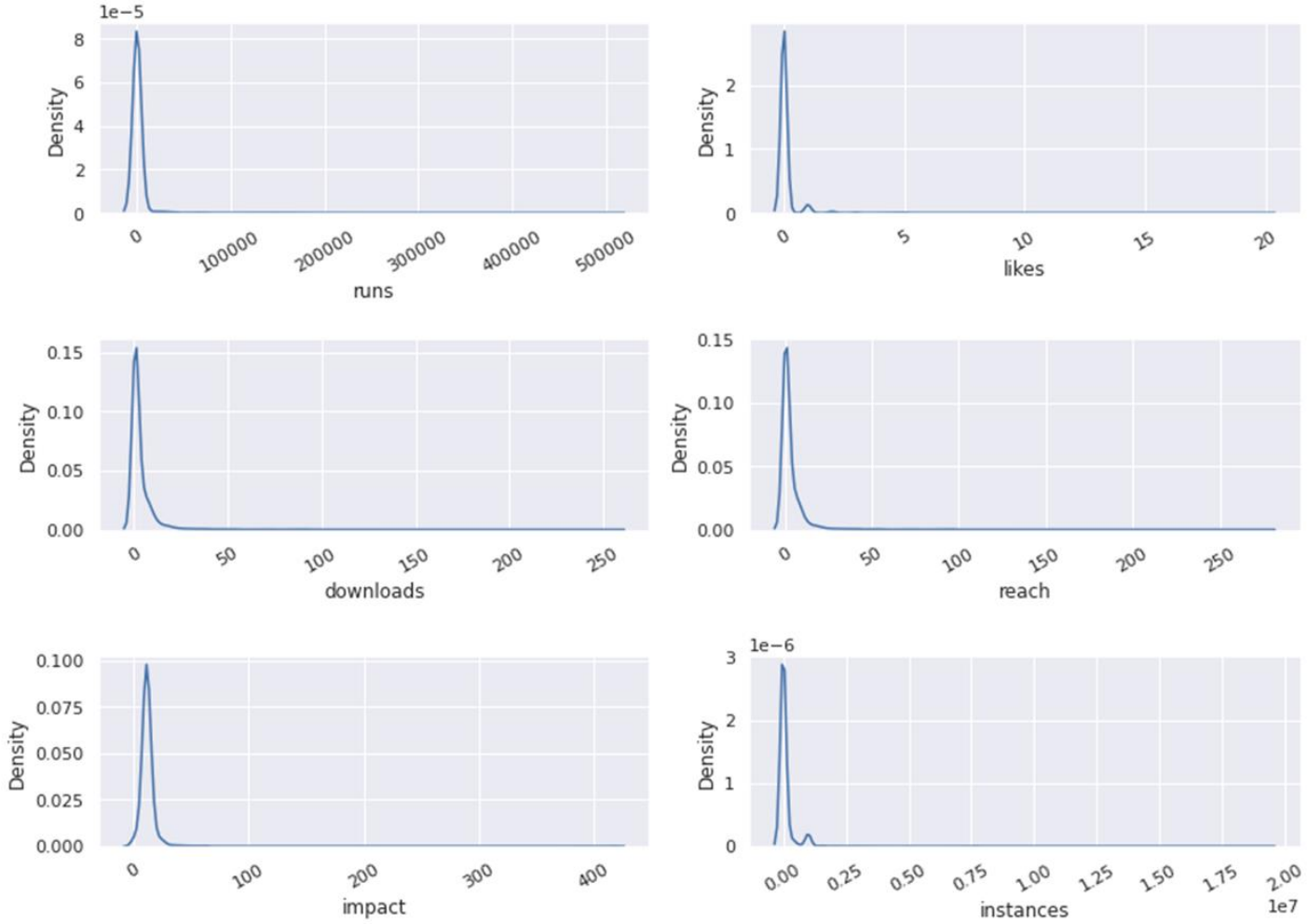
The Role of Datasets in AI Platforms

- Data is the “fuel” for AI models
- Key: Make use of datasets as easy as possible!
- 2 examples (work-in-progress)
  - Integration of open and high-quality datasets into AI4EU portal: https://www.ai4eu.eu/
  - Recommendation of datasets in TRUSTS platform: https://www.trusts-data.eu/

AI4EU: Integration of OpenML Datasets

- Smooth integration of external data sources into the platform
- Example - Open machine learning platform OpenML: https://www.openml.org/
- Technical
  - AcumosAI open-source framework offers „DataBroker“ concept to query data from OpenML API and feeds it to AI models
  - Challenge: determine high-quality datasets

AI4EU: OpenML Dataset Selection

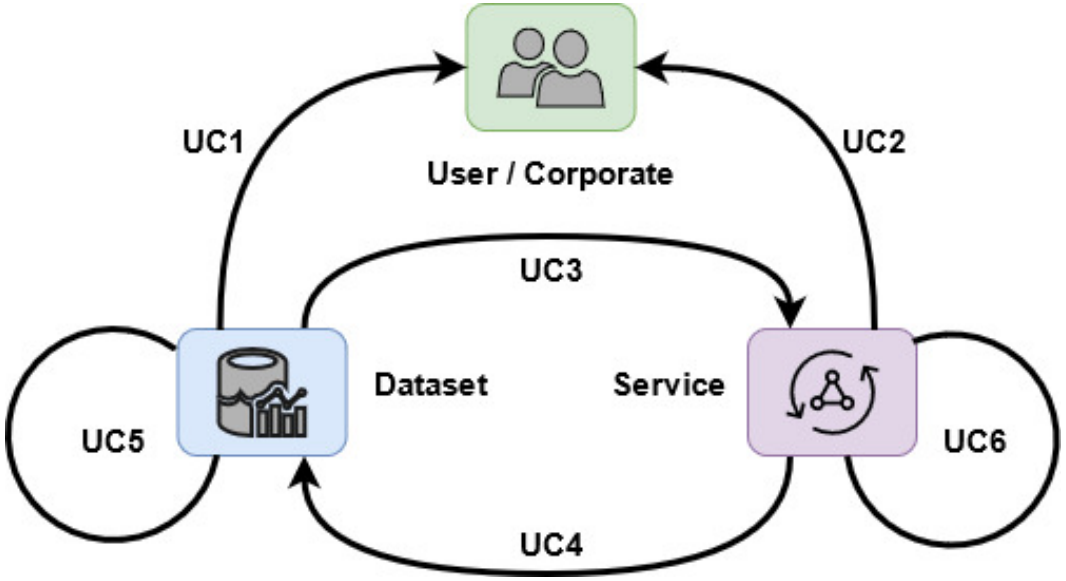


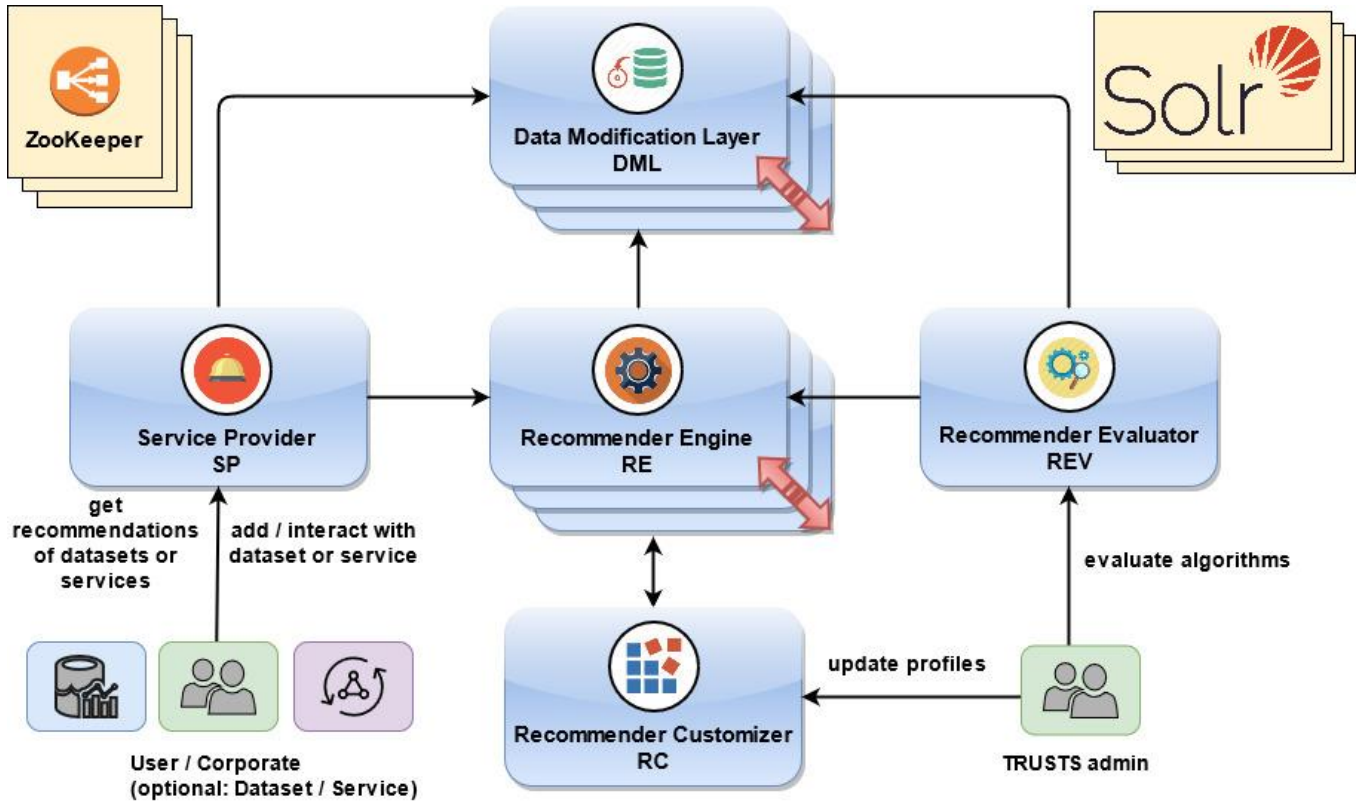
3,276 datasets

('impact' >= 20) | 'reach' >= 20)
& ('instances' >= 100)

155 high-quality datasets

TRUSTS: Recommending Datasets





ScaR framework: http://scar.know-center.tugraz.at/

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Thank you!

Take-home message:

easy (i) integration of, and (ii) access to datasets is key for the success/sustainability of AI platforms

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION

DECSIS ubiwhere valuedate.io