

ASAP-DM: A Framework for Automatic Selection of Analytic Platforms for Data Mining

<u>Manuel Fritz</u>, Osama Muazzen, Michael Behringer, Holger Schwarz

> IPVS University of Stuttgart



University of Stuttgart

Motivation



University of Stuttgart

Pitfall 1: Different APIs lead to cumbersome code adjustments



Pitfall 1: Different APIs lead to cumbersome code adjustments



Pitfall 2: Lack of possibilities to call specific APs





Pitfall 2: Lack of possibilities to call specific APs



Pitfall 3: "What is the 'best' AP for my problem?"



Pitfall 3: "What is the "best" AP for my problem?"

Dataset

Data Mining Task

Analytic Platform









Pitfall 3: "What is the "best" AP for my problem?"

Evaluation Setup

Dataset characteristics

D

- Number of entities
- Number of features
- Filesize
- ...

Training

- Decision Tree: -
- k-Means: $k \in \{3, 6, 9, 12\}$
- FP-Growth: *minSup*, *minConf* ∈ {0.25, 0.5, 0.75, 0.99}

<u>Test</u>

- Decision Tree: -
- k-Means: $k \in \{5, 10, 15, 20\}$
- FP-Growth: *minSup*, *minConf* ∈ {0.3, 0.6, 0.9}





AP

M

Which *AP* performs best for *D*, *M*?



Pitfall 3: "What is the "best" AP for my problem?" Results



Runtime savings of up to 99.71 % (1483 seconds or almost 25 minutes) are possible!

Pitfall 3: "What is the "best" AP for my problem?"



Conclusion







Thank you!



Manuel Fritz

e-mail Manuel.Fritz@ipvs.uni-stuttgart.de phone +49 (0) 711 685-88238 fax +49 (0) 711 685-78238

University of Stuttgart Institute for Parallel and Distributed Systems Universitätsstraße 38, 70569 Stuttgart, Germany