

AUTOML: VISÃO GERAL

Aprendizagem de Máquina Automatizado

Automated Machine Learning: State-of-The-Art and Open Challenges

Radwa Elshaw
University of Tartu, Estonia
radwa.elshaw@ut.ee

Mohamed Maher
University of Tartu, Estonia
mohamed.abdelrahman@ut.ee

Sherif Sakr
University of Tartu, Estonia
sherif.sakr@ut.ee

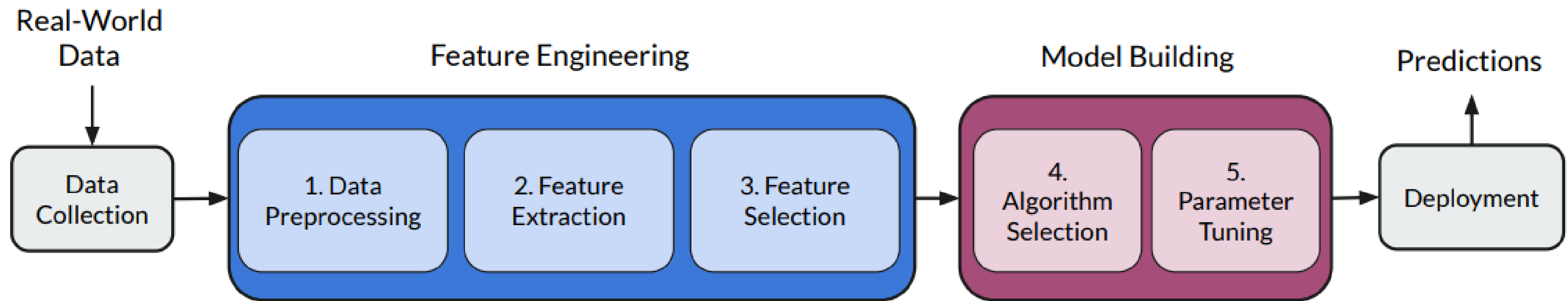


Table 1: Summary of the Main Features of Centralized AutoML Frameworks

| | Release Date | Core Language | Training Framework | Optimization Technique | ML Task | Meta Learning | User Interface | Automatic Feature Extraction | Open Source |
|--------------------------|--------------|---------------|-------------------------------|--|---|---------------|----------------|------------------------------|-------------|
| AutoWeka | 2013 | Java | Weka | Bayesian optimization | Single-label classification regression | × | ✓ | ✓ | ✓ |
| AutoSklearn | 2015 | Python | scikit-learn, | Bayesian optimization | Single-label classification regression | ✓ | × | ✓ | ✓ |
| TPOT | 2016 | Python | scikit-learn | Genetic Algorithm | Single-label classification regression | × | × | ✓ | ✓ |
| SmartML | 2019 | R | mlr, RWeka & other R packages | Bayesian optimization | Single-label classification | ✓ | ✓ | × | ✓ |
| Auto-MEKA _{GGP} | 2018 | Java | Meka | Grammar-based genetic algorithm | Multi-label classification | ✓ | × | × | ✓ |
| Recipe | 2017 | Python | scikit-learn | Grammar-based genetic algorithm | Single-label classification | ✓ | × | ✓ | ✓ |
| MLPlan | 2018 | Java | Weka and scikit-learn | Hierarchical Task Planning | Single-label classification | × | × | ✓ | ✓ |
| Hyperopt-sklearn | 2014 | Python | scikit-learn | Bayesian Optimization & Random Search | Single-label classification regression | × | × | ✓ | ✓ |
| Autosstacker | 2018 | - | - | Genetic Algorithm | Single-label classification | × | × | ✓ | × |
| VDS | 2019 | - | - | cost-based Multi-Armed Bandits and Bayesian Optimization | Single-label classification regression image classification audio classification graph matching | ✓ | ✓ | ✓ | × |
| AlphaDSM | 2018 | - | - | Reinforcement learning | Single-label classification regression | ✓ | × | ✓ | × |
| OBOE | 2019 | Python | scikit-learn | collaborative filtering | Single-label classification | ✓ | × | × | ✓ |
| PMF | 2018 | Python | scikit-learn | collaborative filtering & Bayesian optimization | Single-label classification | ✓ | × | ✓ | ✓ |

AutoML in The Wild: Obstacles, Workarounds, and Expectations

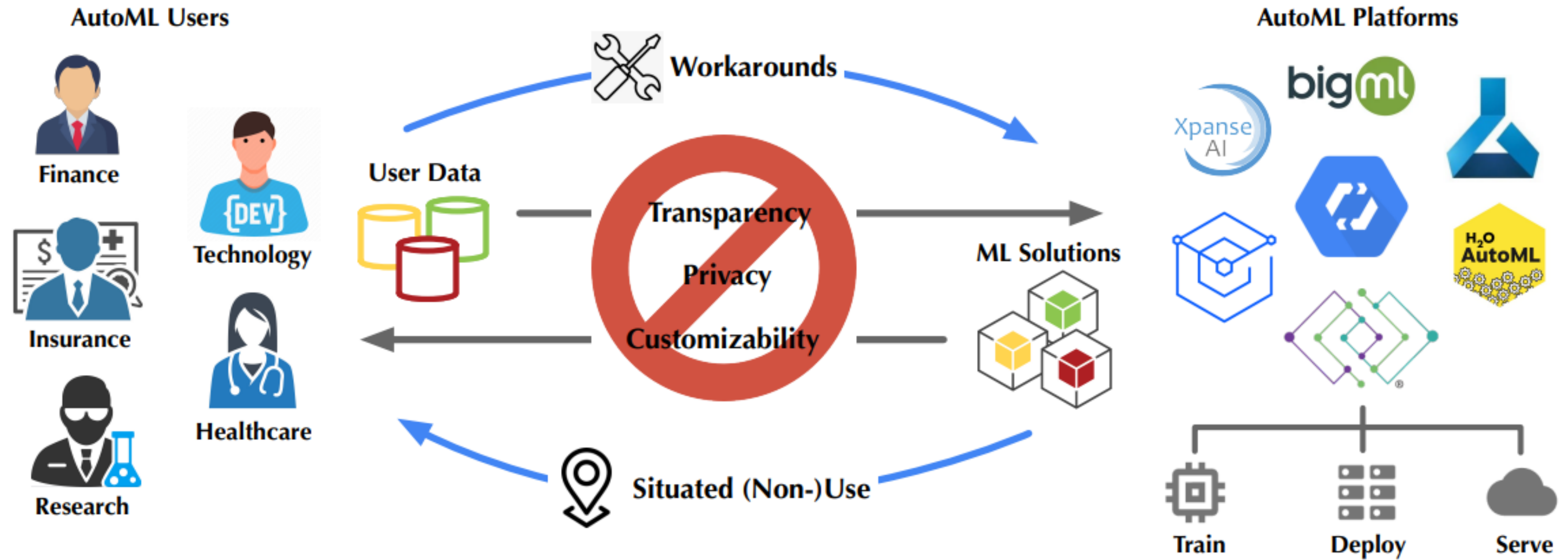
Yuan Sun
yws5055@psu.edu
Pennsylvania State University
University Park, USA

Qiurong Song
qzs5098@psu.edu
Pennsylvania State University
University Park, USA

Xininig Gui
xinninggui@psu.edu
Pennsylvania State University
University Park, USA

Fenglong Ma
fenglong@psu.edu
Pennsylvania State University
University Park, USA

Ting Wang
inbox.ting@gmail.com
Pennsylvania State University
University Park, USA



The autoML Jungle—An Overview

Yusuf Kirikkayis ✉

Ulm University, 89081 Ulm, Germany

Machine Learning Pipeline

Automated machine Learning Pipeline

Data Collection

Data Preparation

Feature Engineering

Model Selection

Hyperparameter Tuning

Model generation

Model Evaluation

| Tool\ Framework | Model Selection | ML Task | User Interface | Binding | Open Source |
|----------------------------|-----------------------------|--|-------------------|-------------------------------------|-------------|
| scikit-learn | Grid- & Random Search | Supervised- and Unsupervised learning | ✗ | Python | ✓ |
| automl | Random search | Supervised learning | ✗ | Python | ✓ |
| TPOT | Genetic algorithm | Supervised- and Unsupervised learning | ✗ | Python, R | ✓ |
| auto-sklearn | Random search | Supervised learning | ✗ | Python | ✓ |
| Hyperopt-Sklearn | Random search | Supervised learning | ✗ | Python | ✓ |
| H2O | Ensemble Selection | Supervised learning | ✓ | Python, JavaScript R and Tableau | ✓ |
| AlphaD3M | Monte Carlo Tree Search | Supervised learning | ✗ | Python | ✗ |
| Google Cloud AutoML Tables | Ensemble & Random Search | Supervised- and Unsupervised learning | ✓ | Python, R and Node.js | ✗ |
| Amazon SageMaker Autopilot | Random & Bayesian Search | Supervised- and Unsupervised learning | ✓ | Python, R and Node.js | ✗ |
| Azure ML Studio | Ensemble & Random Search | Supervised learning | ✓ | Python, R and Node.js | ✗ |

Benchmark and Survey of Automated Machine Learning Frameworks

Marc-André Zöller

MARC.ZOELLER@USU.COM

USU Software AG

Rüppurrer Str. 1, Karlsruhe, Germany

Marco F. Huber

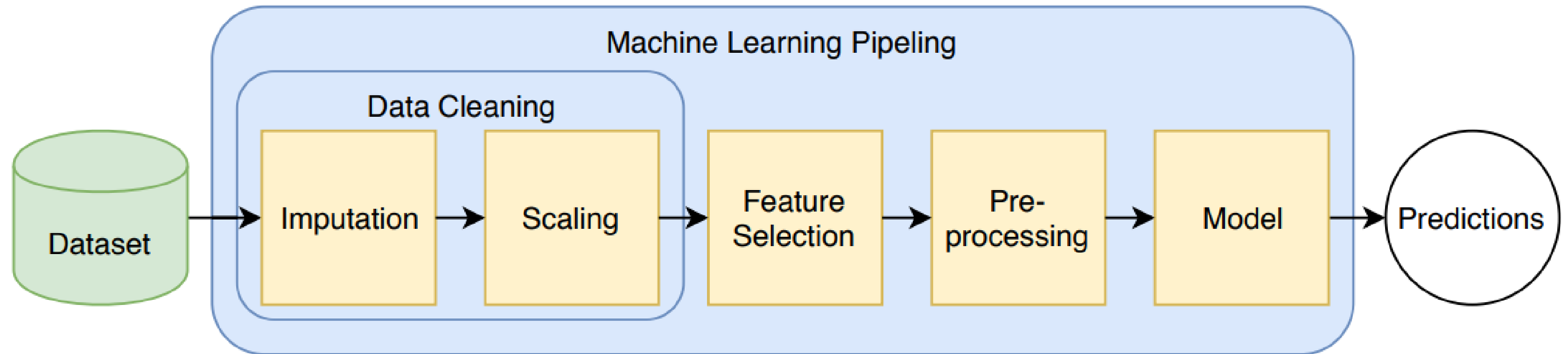
MARCO.HUBER@IEEE.ORG

Institute of Industrial Manufacturing and Management IFF,

University of Stuttgart, Allmandring 25, Stuttgart, Germany &

Fraunhofer Institute for Manufacturing Engineering and Automation IPA

Nobelstr. 12, Stuttgart, Germany

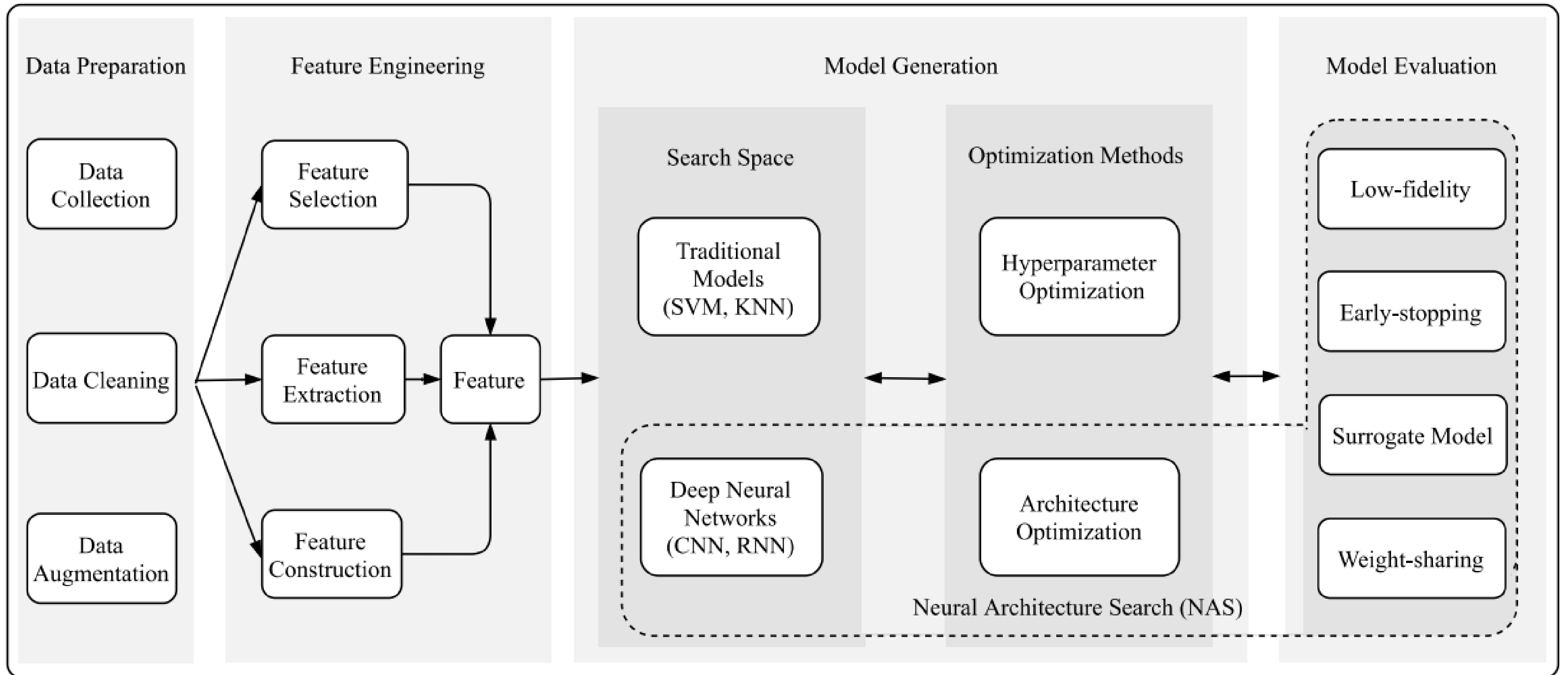


| Framework | CASH Solver | Structure | Ensem. | Cat. | Parallel | Time |
|---------------|---------------|-----------|--------|------|----------|------|
| DUMMY | – | Fixed | no | no | no | no |
| RANDOM FOREST | – | Fixed | no | no | no | no |
| TPOT | Genetic Prog. | Variable | no | no | Local | yes |
| HPSKLEARN | HYPEROPT | Fixed | no | yes | no | yes |
| AUTO-SKLEARN | SMAC | Fixed | yes | Enc. | Cluster | yes |
| RANDOM SEARCH | Random Search | Fixed | no | Enc. | Cluster | yes |
| ATM | BTB | Fixed | no | yes | Cluster | no |
| H2O AUTOML | Grid Search | Fixed | yes | yes | Cluster | yes |

AutoML: A survey of the state-of-the-art

Xin He, Kaiyong Zhao, Xiaowen Chu^{*}

Department of Computer Science, Hong Kong Baptist University, Hong Kong



AUTOML: FRAMEWORKS

Aprendizagem de Máquina Automatizado

Auto-Sklearn

- Linguagem de código: Python
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: scikit-learn
- Descrição: O Auto-Sklearn é uma biblioteca de AutoML baseada em scikit-learn. Ele realiza busca automática de hiperparâmetros e seleção de modelos para encontrar a melhor combinação para um determinado conjunto de dados.

TPOT

- Linguagem de código: Python
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: scikit-learn
- Descrição: O TPOT (Tree-based Pipeline Optimization Tool) é uma biblioteca de AutoML que utiliza algoritmos genéticos para otimizar pipelines de aprendizado de máquina. Ele automatiza a seleção de modelos, pré-processamento de dados, seleção de recursos e ajuste de hiperparâmetros.

Dummy

- Linguagem de código: Python
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: scikit-learn
- Descrição: O Dummy é uma ferramenta simples de AutoML baseada na biblioteca scikit-learn. Ele fornece uma solução básica para problemas de classificação e regressão, gerando modelos "dummy" que podem ser usados como referência para avaliar o desempenho de outros modelos mais sofisticados.

AlphaD3M

- Linguagem de código: Python
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: D3M
- Descrição: O AlphaD3M é um framework de AutoML desenvolvido pela DARPA (Agência de Projetos de Pesquisa Avançada de Defesa). Ele visa automatizar o processo de desenvolvimento de pipelines de ciência de dados, incluindo etapas como pré-processamento de dados, seleção de algoritmos e ajuste de hiperparâmetros. O AlphaD3M é construído em cima do D3M, um ecossistema de software para aprendizado de máquina automatizado.

H2O

- Linguagem de código: Python, R, Java, Scala
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: H2O
- Descrição: O H2O é uma plataforma de AutoML de código aberto que oferece uma ampla gama de algoritmos de aprendizado de máquina. Ele possui uma interface intuitiva para treinar modelos e realizar tarefas como classificação e regressão. Além disso, o H2O permite o uso de GPUs para acelerar o processo de treinamento.

OBOE

- Linguagem de código: Python
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: scikit-learn, Optunity
- Descrição: O OBOE (Optimized Black-Box Experimenter) é um framework de AutoML que enfoca a otimização de hiperparâmetros. Ele usa técnicas de otimização para encontrar os melhores valores de hiperparâmetros para um modelo de aprendizado de máquina, tornando mais eficiente a busca por combinações ideais de parâmetros. O OBOE é construído sobre a biblioteca scikit-learn e usa o Optunity para a otimização.

AutoWeka

- Linguagem de código: Java
- Plataforma: Linux, macOS, Windows
- Bibliotecas principais: Weka
- Descrição: O AutoWeka é um framework de AutoML baseado no Weka, uma biblioteca popular de aprendizado de máquina em Java. Ele automatiza a seleção de algoritmos, ajuste de hiperparâmetros e avaliação de desempenho. O AutoWeka usa técnicas de busca em espaço de hiperparâmetros para encontrar a melhor configuração de modelo para um conjunto de dados.

FIM!