

Evolutionary optimization of machine learning pipelines

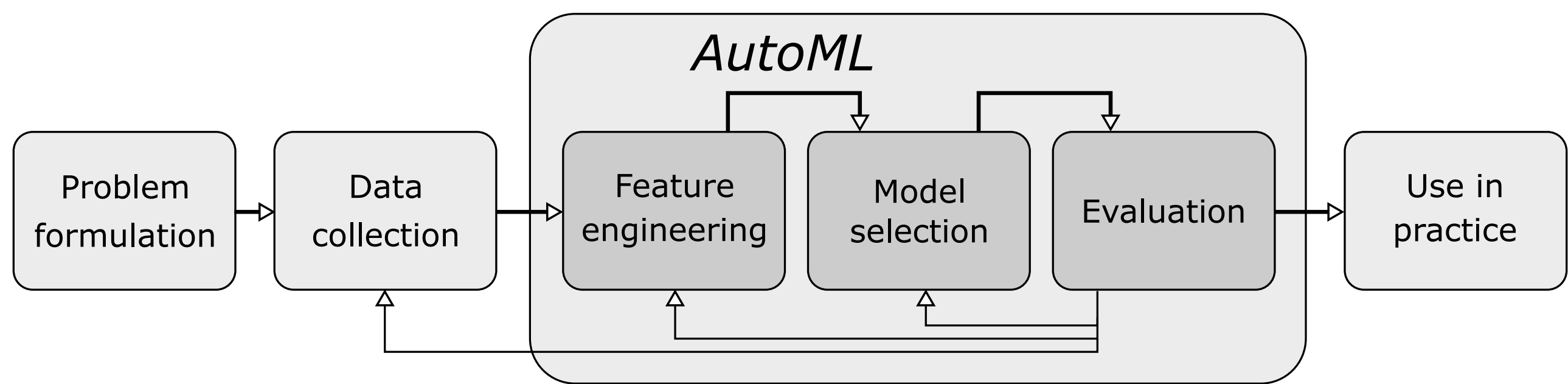
Gabriela Suchopárová

Abstract

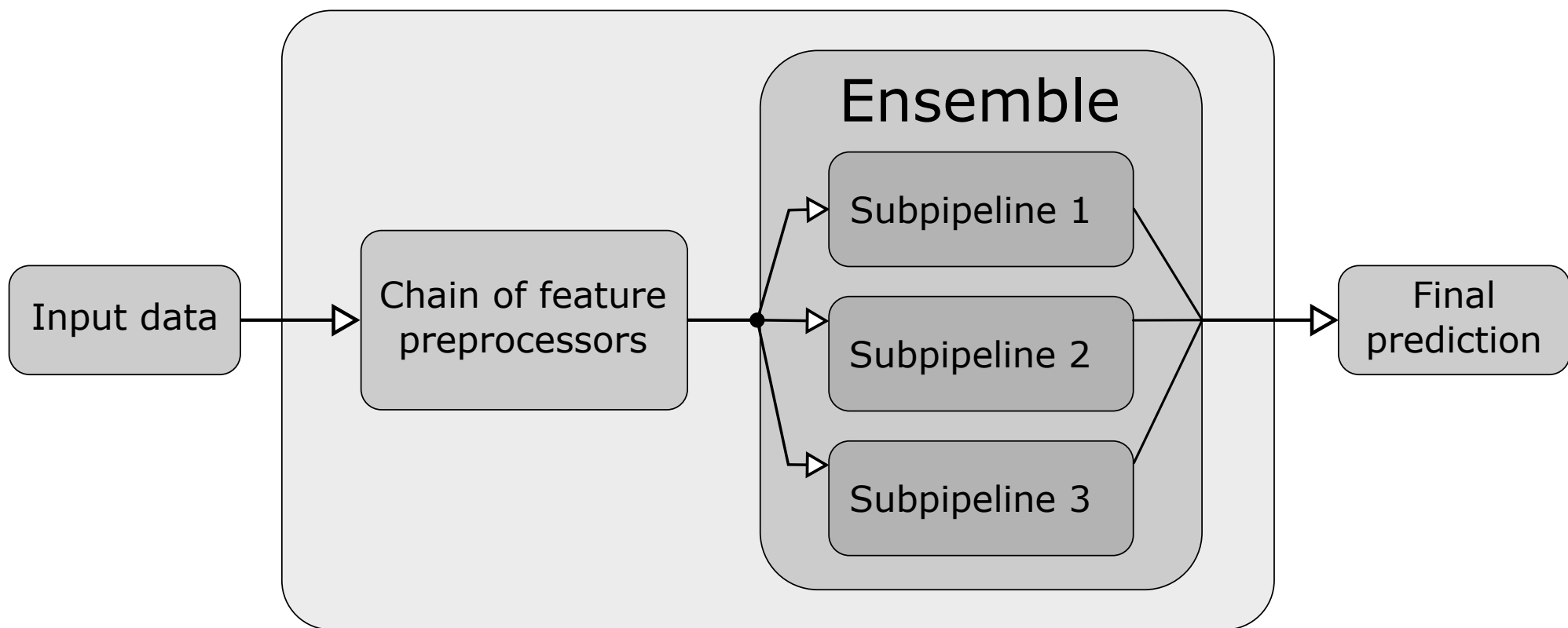
The subject of this work is the automated machine learning (AutoML), which is a field that aims to automatize the process of model selection for a given machine learning problem. We have developed a system that, for a given supervised learning task represented by a dataset, finds a suitable pipeline — combination of machine learning, ensembles and preprocessing methods. For the search we designed a special instance of the developmental genetic programming which enables us to encode directed acyclic graph pipelines into a tree representation. The system is implemented in the Python programming language and operates on top of the scikit-learn library. The performance of our solution was tested on 72 datasets of the OpenML-CC18 benchmark with very good results.

Workflows

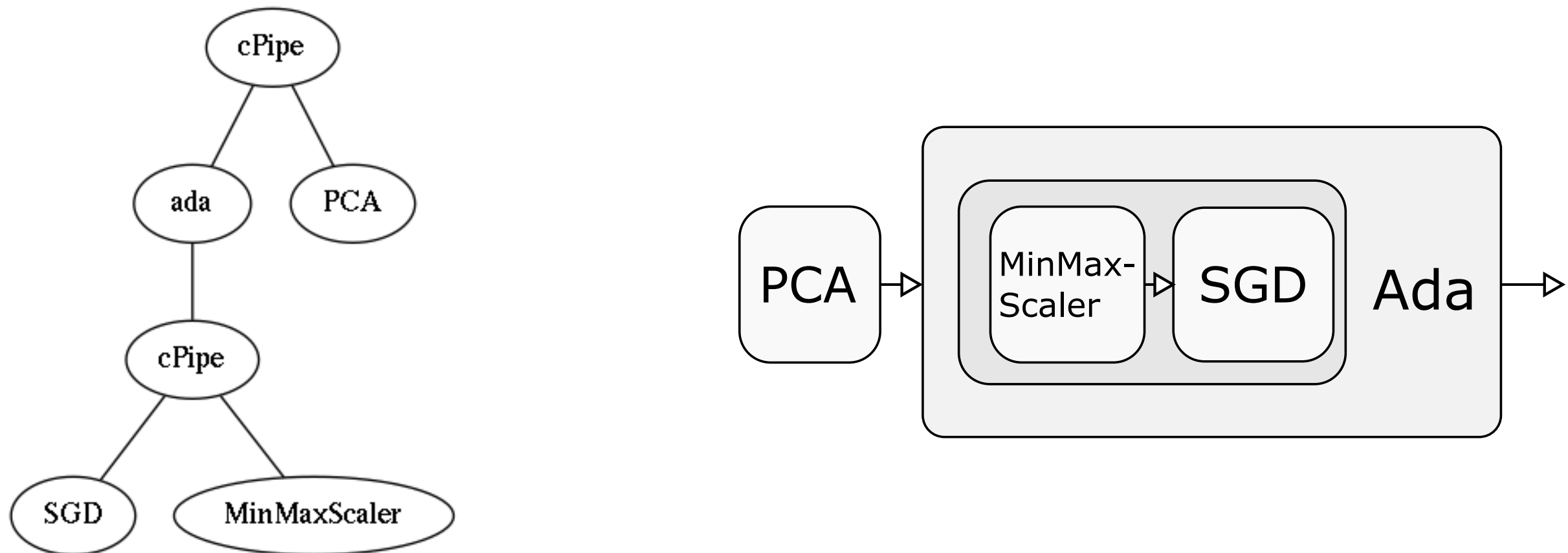
Here will be a general workflow description.



Existing systems focused only on...



Some Box



Here will be a description of pipeline-to-tree encoding.

OpenML-CC18

