

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Rajat Sharma  
**Název práce** Evolutionary techniques in AutoML  
**Rok odevzdání** 2024  
**Studijní program** Informatika      **Studijní obor** Umělá inteligence

**Autor posudku** Roman Neruda  
**Pracoviště** ÚI AV ČR

**Role** Vedoucí

## Text posudku:

The thesis deals with an important problem from the machine learning area – the so-called AutoML approaches to automatically build and fine-tune machine learning algorithms. Author is focused on traditional machine learning techniques gathered in the scikit-learn library, with the aim to create data-dependent classifiers by means of search and optimization algorithms. The goal of the thesis was to design and implement evolutionary algorithm exploring the search space of pipelines representing complex machine learning workflows.

The structure of the work is the following, after a motivation in introduction, the preliminaries and related work sections describe the state of the art in the AutoML field with the focus on traditional methods. Several search methods, that are used later in the work, are also described, including hill climbing, simulated annealing and evolutionary algorithms. Chapter 3 represents the core of author's work. Here, the design of AutoML algorithm is described, starting with representation of possible solutions (individuals), and going through the sequence of search algorithms, from simple ones, such as exhaustive search, to evolutionary search. In Chapter 4, the proposed algorithms are experimentally verified on several benchmark data sets from scikit-learn and OpenML benchmark suit. The thesis is wrapped up in the Conclusions chapter, briefly stating the main results and possible future work, such as hyper-parameter tuning.

The main contributions of the thesis are:

1. The design of the AutoML algorithm is thorough and contains four different search methods including the base-line exhaustive search, two local searches, and the evolutionary algorithms. This allows for extensive performance comparisons between inherently different search approaches.
2. The experimental part tests these approaches on relatively wide range of classification problems, from “toy” data sets to the OpenML classification suite data. Especially the results on larger, real-life data are of interest and should probably be submitted back to the OpenML portal.
3. The text of the thesis is brief but clear and contains all the necessary information.

Among the weaker points of the thesis:

1. The related work part of the thesis should be more thorough and mention more

relevant work in the area.

2. Several parts of the algorithms proposed could be improved or extended, as already stated by the author in the conclusions. Personally, I think that the hyper-parameters tuning procedure would improve the results achieved and make the system more usable in practice. Based on the above evaluation, I consider the submitted text satisfying the requirements for the thesis, and recommend it for the defense.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

*Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).*

**Datum** 1. September 2024

**Podpis**