



# Posudek oponenta závěrečné práce

Oponent práce:	Ing. Karel Klouda, Ph.D.
Student:	Vojtěch Müller
Název práce:	Heuristika pro automatický výběr distribuce v Open Source H2O AutoML
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	12. června 2023

## Hodnotící kritéria

### 1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Všechny body zadání byly splněny. Zadání dovovalo poměrně velkou volnost v tom, jak k němu přistoupit. Autor si vybral spíše snazší cestu, i když i ta byla co se týče experimentů poměrně pracná. Jednoduchost byla vynucena i požadavky na rychlost, které vedly k nutnosti konstrukce jednoduchého modelu (stromu) založeného na rychle získatelných příznacích.

### 2. Písemná část práce

70/100 (C)

Text práce je dobře strukturovaný a čtenář se z něho dozví vše potřebné. Někdy je to ale pro čtenáře zbytečně náročné. V práci se vyskytují časté formální chyby. Např. nejsou někdy vysvětleny pojmy (nebo jsou vysvětleny nedostatečně) či značení ve vzorcích - příkladem je např. sekce 3.1.1.2: co je Exponential family, co je index  $t$ , atp. Další matoucí prohrěšek je to, že na očíslované sekce, tabulky, obrázky a rovnice je často odkazováno pouze číslem a čtenář tak neví, jestli jde např. o obrázek či sekci. A i když je např. slovo "section" použito, je nesprávně s malým písmenem. Očíslovány jsou i rovnice, na které není nikdy odkazováno.

Angličtina je srozumitelná, ale často jsou v textu nezvyklé formulace a někdy text spíše připomíná úryvek z emailu kolegovi (např. úvod do sekce 5.4). Toto je nepříjemné hlavně v kapitole 5 s vyhodnocením výsledků, kde jsem si občas nebyl jistý, zda rozumím textu správně.

### 3. Nepísemná část, přílohy

78/100 (C)

Sada příložených skriptů má formu několika skriptů a Jupyter Notebooků v jazyce Python, obsahuje také vytvořené datasety a Docker file pro zprovoznění H2O frameworku. Skripty i notebooky jsou decentně okomentované. Samotná implementace výsledné heuristiky je poměrně triviální.

### 4. Hodnocení výsledků, jejich využitelnost

78/100 (C)

Práce ukazuje, že automatické nastavení distribuce pro vybrané modely ve frameworku H2O může vést k lepším výsledkům. Experimenty působí ale jen jako první krok a k solidnějším závěrům a asi i lepším výsledkům by bylo dobré ještě provést další experimenty.

### Celkové hodnocení

73/100 (C)

Autor prokázal, že je schopen provést rozsáhlý experiment, který dává rozumné řešení netriviálního problému. Text práce i samotné experimenty by ale mohly být kvalitnější a dotaženější. Navrhuji práci hodnotit jako dobrou, tedy známkou C.

### Otázky k obhajobě

- 1) I v práci píšete, že některá pravidla ve finální heuristice jsou zbytečná. Nešel by tedy použít nějaký algoritmus na prořezávání stromů, který by snížil počet pravidel?
- 2) Mohl byste prosím přesně vysvětlit, co byla testovací data, ze kterých byl vykreslen obrázek 5.7 na str. 32?

## **Instrukce**

### **Splnění zadání**

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

### **Písemná část práce**

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

### **Nepísemná část, přílohy**

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

### **Hodnocení výsledků, jejich využitelnost**

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

### **Celkové hodnocení**

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.