

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Bc. Jan Babušík

**Název práce** Detekce anomalit v log datech

**Rok odevzdání** 2021

**Studijní program** Informatika **Studijní obor** Umělá inteligence

**Autor posudku** Mgr. Marta Vomlelová Ph.D. **Role** vedoucí

**Pracoviště** Department of Theoretical Computer Science and Mathematical Logic

## Text posudku:

Práce se zabývá detekcí anomalit v log datech. Představuje dvě současné práce založené na hlubokých neuronových sítích, představuje tři datasety, na které následně aplikuje základní verzi LSTM modelu a porovnává jeho úspěšnost na jednotlivých datasetech, srovnává výsledky s úspěšností jednoduchých statistických modelů (PCA, N-gram).

Teoretická část práce obsahuje úvod do rekurentních neuronových sítí, popis algoritmů DeepLog a LogRobust. Z těchto algoritmů byly v následujících experimentech použity jen algoritmy založené na 'statických' částech logu (tzv. keys), proměnlivá data jako např. prodleva mezi jednotlivými klíči v bloku zůstala nevyužita.

Oceňuji přípravu vlastních dat až k naučení použitelného modelu. Přes přístup k databázi a spolupráci firmy se ukázalo jako nerealistické získat vlastní anotovaná data pro učení. Proto se student zaměřil na modely, které pro učení nepotřebují negativní příklady a modely na vlastních (HAVIT) datech ohodnocuje podle schopnosti rozlišit data vygenerovaná systémem od náhodně generovaných dat. Autor ukazuje, že ani veřejné datasety používané v odborných pracích nejsou pro učení ideální (jen 589 různých průběhů v 11 miliónech logů v HDFS datech, chybějící *instance\_id* u mnoha záznamů v OpenStack datech).

Kromě LSTM modelu autor zahrnul do porovnání i jednoduché modely, výsledky uvedl

v přehledných tabulkách a doprovodil grafy zdůvodňujícími volbu meta-parametrů.

Student prokázal schopnost samostatné odborné práce.

Představil algoritmy užití v nedávných článcích, připravil data pro učení modelu, modely naučil a porovnal výsledky jednotlivých modelů na veřejných i vlastních datech.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

V Praze dne 23. 8. 2021

Podpis