

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Martin Adam

**Název práce** Strojové učení pro monitorování počítačových clusterů

**Rok odevzdání** 2020

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

**Autor posudku** Mgr. Martin Pilát, Ph.D. **Role** vedoucí

**Pracoviště** KTIML MFF UK

## Text posudku:

Autor se ve své práci zabývá možnostmi využití strojového učení v monitorování počítačových clusterů. Konkrétně se zaměřuje na metody, které na základě metrik reportovaných každým ze strojů v clusteru detekují anomálie, které mohou být způsobené nějakým chybovým stavem. K tomuto účelu využívá cluster MONIT v CERNu, kde nasimuloval několik běžných problémů, které mohou nastat a následně navrhl algoritmy, které umí tyto chyby detekovat.

Práce je rozdělena do pěti kapitol (plus úvod a závěr). V první kapitole student popisuje nástroje, které se běžně používají k monitorování počítačových clusterů, cluster MONIT, na kterém provádí experimenty, a související práce o využití strojového učení při monitorování clusterů. Druhá kapitola potom obsahuje podrobný popis problému a cíle práce.

Samotný přínos práce začíná v kapitole třetí, kde student popisuje, jakým způsobem nasbíral data potřebná pro další testování metod. Vzhledem k tomu, že chyby se clusteru nestávají příliš často, bylo potřeba je uměle vytvářet. Student tímto způsobem vytvořil dataset obsahující 26 anomálií čtyř různých typů. Dataset obsahuje několik desítek různých metrik poskytovaných jednotlivými stroji v clusteru. Takový dataset předtím neexistoval a sám o sobě tak představuje významný přínos pro budoucí práce v této oblasti. Dataset je zveřejněn v příloze práce. Tato kapitola také obsahuje stručnou analýzu a vizualizaci vytvořeného datasetu.

Ve čtvrté kapitole Martin Adam představuje navržený modul pro detekci anomálií. Zkoumá zde použití jak metod pro učení bez učitele, tak metod pro učení s učitelem. Ve druhém zmíněném případě se metody snaží předpovídat hodnoty jednotlivých metrik a chyba (anomálie) je detekována v případě, že se jejich predikce výrazně liší od skutečné hodnoty.

Navržené metody jsou otestovány v kapitole páté, kde se student zabývá i tím, jak často bude jako anomálie hlášeno něco, co ve skutečnosti není chyba. Ukazuje se, že pro snížení počtu těchto falešně pozitivních předpovědí stačí, když se vyfiltrují pouze takové situace, kdy předpověď modelu je špatná dva kroky po sobě. Výsledný algoritmus je schopný detekovat 70 % anomálií a nemá (na

testovacích datech) žádné falešně pozitivní případy. Navržený algoritmus se tak dá považovat za úspěšný a představuje zajímavý přínos k monitorování clusterů.

Celá práce je napsána velmi dobrou angličtinou a je snadno srozumitelná. Práce je psána poměrně stručně, obsahuje ale všechny podstatné informace. Jediné, co by se práci dalo vytknout, je to, že autor testoval relativně malé množství metod pro strojové učení. Na druhou stranu metody byly zvoleny tak, aby je bylo možné i prakticky využít, tj. přednost byla dána metodám, které jsou implementované (nebo je lze snadno implementovat) v nástrojích na sledování clusterů. Tento přístup zaměřený na praktické využití je ostatně vidět v celé práci. Kvalitu práce lze demonstrovat i tím, že její výsledky byly prezentovány již na dvou ročních konferencích CHEP (Computing in High-Energy and Nuclear Physics, roky 2018 a 2019), z CHEP 2018 již vyšel článek ve sborníku, z CHEP 2019 je článek v přípravě.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

V Praze dne 23. ledna 2020

Podpis: