

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	<b>Porovnání reprezentací souborů logů pro metody detekce anomálií</b>
<b>Jméno autora:</b>	<b>Martin Hubal</b>
<b>Typ práce:</b>	bakalářská
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Vedoucí práce:</b>	Ing. Jan Drchal, PhD.
<b>Pracoviště vedoucího práce:</b>	Katedra počítačů

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>průměrně náročné</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Tato bakalářská práce přímo navazuje na předchozí diplomovou práci. Cílem bylo ověřit, zda jsou metody detekce anomálií nad logy, postavené na sémantických reprezentacích slov (fastText), v porovnání se základními přístupy založenými na bag-of-words reprezentací, robustnější z hlediska zpracování dosud neviděných zpráv. Student musel zvládnout poměrně složité specializované architektury neuronových sítí pro zpracování sekvenčních dat. Experimenty prováděné na clusteru jsou výpočetně náročné, proto bylo třeba dobře plánovat.	

<b>Splnění zadání</b>	<b>splněno</b>
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání bylo splněno bez výhrad.	

<b>Aktivita a samostatnost při zpracování práce</b>	<b>B - velmi dobře</b>
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Student byl po celou dobu aktivní a na všechny konzultace byl vždy připraven. Drobnou výtka mám k úplné finální fázi, kdy už jsem z časových důvodů nebyl schopen redigovat text.	

<b>Odborná úroveň</b>	<b>A - výborně</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
K odborné úrovni nemám námitek. Práce měla zejména experimentální charakter a experimenty byly navrženy, provedeny i vyhodnoceny správně. Původní hypotéza o vyšší robustnosti sémantických metod prokázána nebyla, to ale není chybou studenta. Problémovou se ukázala být použitá datová sada HDFS, která ve výsledku obsahovala pouze 48 rozdílných šablon logů, což vedlo k přeučování konvolučního modelu. Student se snažil různými metodami toto přeučování omezit, bohužel se to nepodařilo. Alternativní zdroje dat v dostatečném rozsahu bohužel neexistují.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>D - uspokojivě</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána v českém jazyce s občasnými překlepy. Text je však bohužel na mnoha místech stylisticky neobratný a v některých částech špatně strukturovaný – student v několika případech používá termíny, které ještě nebyly zavedeny. Nejvíce tím trpí popis velmi zásadní experimentální metodologie. Příkladem mohou být sekce popisující hyperparametry porovnávaných metod s matoucí poznámkou, že jsou jejich hodnoty nastavovány náhodně. Až mnohem dále v textu je zmínka o tom, že pro každý experiment bylo generováno sto nastavení a následně vybráno to nejlepší. Zde zase chybí informace o tom, z jaké množiny byly hodnoty hyperparametrů vybírány. Nejsem rovněž příliš spokojen s rozsahem úvodní části – popis použitých metod je příliš stručný a ve výsledku není zřejmá motivace pro analýzu metod založených na sémantice.	

**Výběr zdrojů, korektnost citací**

**A - výborně**

*Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

Zdroje jsou citovány správně.

**Další komentáře a hodnocení**

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

Výstupy práce poukazují na to, že v současné době není k dispozici žádná rozumně použitelná datová sada pro trénování modelů strojového učení na komplexnějších úlohách spojených s analýzou logů.

**III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.*

S přístupem studenta i výstupy jeho práce jsem celkově spokojen. Jediné závažnější výhrady mám k textu. Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Datum: 2.6.2022

Podpis: