FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Simona Fornůsek, Ph.D.
Student:	Bc. Matej Hulák
Název práce:	Klasifikace síťového provozu pomocí strojového učení
Obor / specializace:	Počítačová bezpečnost
Vytvořeno dne:	29. května 2022

Hodnotící kritéria

1. Splnění zadání

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

Dle zadání se měl autor seznámit s principy monitorování síťového provozu pomocí rozšířených síťových toků a s metodami strojového učení používanými pro jejich analýzu a dále provést experimenty s různými modely strojového učení, dále porovnat modely strojového učení pracující nad rozšířenými síťovými toky s modely využívajícími pouze základní toky a implementovat vybranou metodu jako modul do systému NEMEA - všechny body ze zadání jsou v práci obsaženy a splněny.

2. Písemná část práce 60 / 100 (D)

Text práce se dá rozdělit na dvě části - první část tvoří rešerše, druhou poté popis implementace a experimentů. Zatímco druhá část je poměrně obsáhlá a celkem detailně popisuje provedené experimenty, k první části práce mám několik výtek. Zejména kapitola 2.4. "Klasifikačné metódy strojového učenia" je poměrně strohá, a určitě by zasloužila víc pozornosti, a to jak detailem a formálním popisem algoritmů, tak i třeba rozšířením rešerše o praktickém využití, vhodnosti a úspěšností jednotlivých algoritmů pro různé účely v rámci analýzy síťových dat. Práci chybí i detailnější rešerše stávajících řešení.

3. Nepísemná část, přílohy 90 / 100 (A)

Nepísemnou část práce tvoří zejména zdrojové kódy, což považuji za adekvátní. Kód je přehledný a dobře čitelný, orientaci usnadňují komentáře.

4. Hodnocení výsledků, jejich využitelnost 90 /100 (A)

Využití metod strojového učení pro analýzu síťových dat je aktuálním tématem, a jelikož implementační část práce ma formu modulu do již stávajícího systému NEMEA, zajisté najde praktického využití. Celkově by práci ale prospěla hlubší rešerše a detailnější srovnání s již publikovanými výsledky.

Celkové hodnocení 80 /100 (B)

Práci doporučuji k obhajobě a vzhledem k slabší rešeršní části práce se kloním k hodnocení stupněm B.

Otázky k obhajobě

Jaké jiné metody, kromě strojového učení se dnes využívají ke klasifikaci síťového provozu a v čem spočívají jejich případné nedostatky?
Jaké metody strojového učení využívají stávající publikované práce ke klasifikaci síťového provozu?

Instrukce

Splnění zadání

Posuďte, 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, posuďte 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 posuďte, 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posuďte, 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.