FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Hodnocení vedoucího závěrečné práce

Vedoucí práce:	Mgr. Olha Jurečková
Student:	Bc. Ihor Salov
Název práce:	Detekce malwaru pomocí vizualizačních technik
Obor / specializace:	Počítačová bezpečnost
Vytvořeno dne:	31. května 2024

Hodnotící kritéria

1. Splnění zadání

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

Zadání považuji za splněné jen částečně. Není splněn krok 4, ve kterém měly být výsledky získané studentem pomocí CNN porovnány s výsledky získanými pomocí jiných technik detekce malwaru. Student se také místo detekce malwaru věnoval klasifikaci malwaru do rodin, přičemž vůbec nepoužíval čisté (ne malwarové) soubory, což bylo požadováno v zadání. Na experimentální části student začal pracovat těsně před odevzdáním, proto tato část není dokončena. Nedostatečně je popsán zejména postup použití CNN, např. struktura CNN nebo hyperparameter tunning.

2. Písemná část práce 55 / 100 (E)

Student v práci podrobně popsal existující výzkum v oblasti detekce malwaru pomocí vizualizačních technik. Text práce obsahuje gramatické chyby, např. krátký abstrakt obsahuje několik gramatických chyb, dále v textu jsou chyby ve formatování a experimenty nejsou dostatečně popsány. Práce je střídavě psána v první osobě jednotného čísla a v první osobě množného čísla. Nepřesné nebo neformální vyjadřování, např. "Overall, the model performed very well.", "This is what this thesis will be about." Student zkopíroval několik obrázků (např. obr. 3.3) z knihy [53] a neuvedl to ve své práci. Práce obsahuje nadbytečné části, které do diplomové práce nepatří, např. kap. 4.1.1, 4.1.2 nebo 4.1.3, které se dále nepoužívají a mohly by být výrazně zredukovány.
Níže uvádím některé komentáře ke kap. 4 a 5.
Kapitola 4 obsahuje 4 a půl stránky popisů datových sad SOREL a SORRY, které však nebyly použity v experimentální části.
Student v kap. 4.3 uvádí, že vytvořil dataset SORRY, ale nespecifikuje přesně jak. Konkrétně není jasné, jak byly binární soubory převedeny na obrázky (byl použit

Algoritmus 1 ze strany 16?) a jak se student vypořádal s různými rozměry obrázků.
V kap. 5.1.2 se uvádí "To evaluate the effectiveness of my solution, I will compare performance metrics of CNN with Random Forest Classifier." toto srovnání však v práci chybí.
V kap. 5.2 se uvádí "For each model and each dataset I've measured the next set of metrics: accuracy, precision, recall and F1-score." ale tyto metriky jsou uvedeny pouze pro model MobileNetV2 a pro datovou sadu MaleVis.
Obrázek 5.4 ani tabulka 5.1 nejsou spomenuty v textu. Tabulka 5.1 je také navíc nekompletní - chybí měření u 4 z 5 modelů.

3. Nepísemná část, přílohy 40 /100 (F)

Student neodevzdal zdrojové kódy jako přílohu k práci. Po vyžádání mi je dodatečně zaslal. Experimentální část v nich nebyla dokončena.

4. Hodnocení výsledků, jejich využitelnost 30 /100 (F)

Text práce obsahuje poměrně podrobný popis algoritmu CNN, který může pomoci k lepšímu pochopení tohoto algoritmu a jeho následné implementaci. Využití výsledků práce nelze posoudit, protože experimentální část není dokončena.

5. Aktivita studenta

[1] výborná aktivita
[2] velmi dobrá aktivita
▸ **[3] průměrná aktivita**
[4] slabší, ale ještě dostatečná aktivita
[5] nedostatečná aktivita

Student nebyl moc aktivní a některé moje termíny odevzdání ignoroval, nebo se ozval někdy až o měsíc později.

6. Samostatnost studenta

[1] výborná samostatnost
[2] velmi dobrá samostatnost
▸ **[3] průměrná samostatnost**
[4] slabší, ale ještě dostatečná samostatnost
[5] nedostatečná samostatnost

Student samostatně nastudoval potřebnou teorii a část experimentů implementoval v jazyce Python.

Celkové hodnocení 47 /100 (F)

Student v práci podrobně popsal dosavadní výzkum v oblasti detekce malwaru pomocí vizualizačních technik a teorii kolem konvolučních neuronových sítí, které byly jádrem jeho práce. Student by měl experimentální část práce přepracovat a doplnit 4. bod ze zadání. Experimentální část práce není dokončena a tudíž není splněno zadání práce. Za těchto okolností práci hodnotím známkou F a nedoporučuji ji k obhajobě.

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.

Aktivita studenta

V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

Samostatnost studenta

V souvislosti s průběhem a výsledkem práce posuďte schopnost studenta samostatné tvůrčí práce.

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.