Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Dominik Smrž
Název práce	Feature extraction from Android application packages and its usage in machine learning for malware classification
Rok odevzdání	2017
Studijní program	Informatika **Studijní obor** Obecná informatika

Autor posudku	Mgr. Martin Bálek **Role** Vedoucí
Pracoviště	IÚUK

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

K celé práci	lepší	OK	horší	nevyhovuje
Obtížnost zadání	☐	☒	☐	☐
Splnění zadání	☐	☒	☐	☐
Rozsah práce *... textová i implementační část, zohlednění náročnosti*	☐	☒	☒	☐

Komentář Práce je svým zaměřením především aplikační. Na datech z reálného provozu ukazuje jednotlivé fáze přípravy dat a následné vyhodnocení použitelnosti několika algoritmů strojového učení. Autor v ní osvědčil schopnost vytvoření celého řetězce zpracování dat (od samotných apk souborů až ke změření kvality výsledného modelu). Přestože jednotlivé kroky se mohou jevit jako evidentní, jedná se o vcelku komplexní úlohu. Tuto komplexitu se autorovi podařilo zpřehlednit jen částečně (v hodnocení textové části popisuji především tyto nedostatky práce), proto navrhuji hodnocení v horší části stupně velmi dobře.

Textová část práce	lepší	OK	horší	nevyhovuje
Formální úprava *... jazyková úroveň, typografická úroveň, citace*	☐	☒	☐	☐
Struktura textu *... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu*	☐	☒	☒	☐
Analýza	☐	☒	☐	☐
Vývojová dokumentace	☐	☒	☒	☐
Uživatelská dokumentace	☐	☐	☐	☒

Komentář Práce je psaná v anglickém jazyce, který by mohl být lépe zvládnutý, ale jeho kvalita neztěžuje celkovou srozumitelnost práce. Citace technické specifikace [7] a netechnických zdrojů [1,2,8] nejsou dostatečně vypovídající, všechny ostatní citace odborných publikací jsou ovšem v pořádku.
Celkově se text snaží čtenáře provést jednotlivými kroky, které jsou potřebné k tomu, aby se podařilo navrhnout vhodné příznaky (features) pro klasifikaci malware/clean programů pro Android. Všechny tyto kroky jsou pro daný problém relevantní a fakticky správné. Autor průběžně popisuje, co ho k jednotlivým krokům či heuristikám vede, ale tato průběžnost způsobuje celkově menší přehlednost práce. Struktuře by prospělo výraznější oddělení popisu technických detailů od jejich komentářů ve vztahu k detekci malware. Výsledné porování voleb množiny příznaků a algoritmů strojového učení (kapitoly 5 a 6) bohužel není prezentováno dostatečně přehledně (grafy a tabulky), přestože při podrobném čtení je zřejmé, co autor vzhodnotil jako nejlepší řešení.
Za nedostatek považuji, že součástí práce nejsou (např. v elektronické příloze) přiložena trénovací a testovací data spolu s (byť jednoduchým) programem na jejich analýzu (např. ve formě skriptů v pythonu, které byly použity pro získání prezentovaných výsledků) a krátkým popisem, což bych v tomto případě považoval za postačující uživatelskou dokumentaci.

Implementační část práce	lepší	OK	horší	nevyhovuje
Kvalita návrhu *... architektura, struktury a algoritmy, použité technologie*	☐	☐	☐	☐
Kvalita zpracování *... jmenné konvence, formátování, komentáře, testování*	☐	☐	☐	☐
Stabilita implementace	☐	☐	☐	☐

Komentář Práce obsahuje pouze okrajové zmínky o implementační části - použité algoritmy strojového učení jsou součástí standardních knihoven (v tomto případě knihovna scikit pro Python). Jádro práce spočívá ve způsobu výběru jednotlivých příznaků a porovnání jejich vhodnosti pro problém klasifikace nikoliv v konkrétní implementaci. Implementace samotné extrakce těchto příznaků je vcelku přímočará (vhodně upravený parser apk souborů), ale přestože může skrývat mnohá technická úskalí, nepovažuji ji za nutnou součást práce.

Celkové hodnocení Velmi dobře
Práci navrhuji na zvláštní ocenění Ne

Datum 28. srpna 2017 **Podpis**