Posudek oponenta bakalářské práce

Student: Dubec Branislav
Téma: Analýza škodlivého šifrovaného síťového provozu (id 24215)
Oponent: Homoliak Ivan, Ing., Ph.D., UITS FIT VUT

1. Náročnost zadání **obtížnější zadání**

Zadanie bolo mierne obtiažnejšie z pohľadu bakalárskeho študijného programu a obsahovalo spracovanie a aplikáciu metód umelej inteligencie.

2. Splnění požadavků zadání **zadání splněno**

3. Rozsah technické zprávy **je v obvyklém rozmezí**

Práca obsahuje 38 strán latex-om vysádzaného textu, vyjmúc referencie a prílohy.

4. Prezentační úroveň předložené práce **80 b. (B)**

Práca je pre čitateľa pochopiteľná, jednotlivé kapitoly na seba logicky nadväzujú. Rozsahy a prehľadnosť väčšiny kapitol sú prípustné. Práca obsahuje mnohé časti preložené doslovne alebo skomolene z angličtiny a tým pádom je ten preklad často nesprávny: ako napr. distribúcia namiesto rozloženia pravdepodobnosti alebo hyperplán u SVM alebo algoritmy z dohľadom alebo vyškolený klasifikátor. Na druhej strane tiež zdôrazňujem, že mnohé z týchto termínov a oblastí podľa mojich informácií nie sú povinnou súčasťou bakalárskeho programu.

Pri experimentoch sa študent dopúšťa častej chyby a tou je použitie úspešnosti klasifikácie pri nevyváženom datasete. V týchto prípadoch je správne použiť metriky klasifikácie, ktoré sú odolné voči triednej nevyváženosti, ako napr. F1. V chybovej matici študent tiež neuvádza presnosť a odozvu klasifikácie. Dosiahnuté výsledky by si zaslúžili diskusiu a tiež návrh spôsobov ďalšej optimalizácie. Študent v závere spomína, že vyrovnaný počet inštancií v trénovacom datasete je dôležitý pre správne detekovanie testovacích dát, no nikde v texte sa tomuto problému nevenuje a rovnaký pomer týchto tried aj tak nie je zachovaný vo výslednom datasete. Chcel by som podotknúť, že v reálnej prevádzke je percento útokov len veľmi malé v porovnaní s normálnou prevádzkou. Preto reálnejšie datasety sa snažia zachovať tento pomer, resp. sa k nemu priblížiť. Druhou vecou je vyvážený pomer tried pri testovaní mnohých modelov, na čo sa pravdepodobne študent odkazuje (aj keď bez vysvetlenia).

5. Formální úprava technické zprávy **75 b. (C)**

Na strane 26 sú 2 obrázky plávajúce v strede strany. V práci sú poznámky pod čiarou použité nesprávne. Chýbajú bodky v popiskoch všetkých obrázkov u experimentov. Študent nepouživa tvrdé medzery pri citáciách. Práca obsahuje pravopisné chyby a preklepy, dokonca aj v poďakovaní práce u mena vedúceho aj konzultanta.

6. Práce s literaturou **75 b. (C)**

Literatúra neobsahuje niektoré dôležité články o detekcii anomálií a prienikov v sieťovej prevádzke (ako napr. Debar & Hacier). Namiesto toho, študent cituje články od autorov ako Ahmed, alebo Kumar & Sangwan, ktoré neprinášajú nič nové z pohľadu v práci zmienených citovaných taxonómií a definícií. Napríklad na strane 6 sú zmienené typy sieťových útokov definovaných autormi KDD CUP'99 datasetu, ale študent k nim cituje náhodnú prácu z ticícov, ktoré taxonómiu typov útokov len preberajú.

7. Realizační výstup **85 b. (B)**

Práca má zaujímavý realizačný výstup. No na dotiahnutie by bolo potrebné viac experimentov aj s prípadnými modifikáciami, napríklad využívajúcimi optimalizácie pomocou mriežky, prípadne genetických algoritmov.

8. Využitelnost výsledků

Výsledky práce môžu byť využité v naväzujúcej diplomovej práci alebo tiež v ďalších bakalárskych projektoch.

9. Otázky k obhajobě

Vysvetlite normálne rozloženie ako príklad parametrického modelu. Popíšte čo sú anomálie v zmysle tohto rozloženia.

10. Souhrnné hodnocení **85 b. velmi dobře (B)**

Práca je mierne obtiažnejšieho zadania. Zadanie bolo splnené vo všetkých bodoch. Študent volil relevantnú literatúru, no niektoré dôležité originálne zdroje chýbajú. Práca poskytuje aj výsledky, no tieto nie sú často detailne diskutované. Vzhľadom na náročnosť, prácu hodnotím celkovým stupňom B **(85 bodov)**.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 4. června 2021

Homoliak Ivan, Ing., Ph.D.
oponent