FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Hodnocení vedoucího závěrečné práce

Student:	Bc. Jakub Sekera
Vedoucí: práce:	Mgr. Jakub Růžička
Název práce:	Aplikace strojového učení pro analýzu bezpečnostních auditních záznamů v kontextu GDPR
Obor:	Počítačová bezpečnost

Datum vytvoření: 6. 6. 2020

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 4:*
1. Splnění zadání	***1=zadání splněno,*** *2=zadání splněno s menšími výhradami,* *3=zadání splněno s většími výhradami,* *4=zadání nesplněno*

Popis kritéria:
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.

Komentář:

Byly splněny všechny body zadání - studium auditních bezpečnostních záznamů a požadavků GDPR, srovnání metod strojového učení v této doméně (vč. shrnutí současné praxe), implementace funkčního prototypu a jeho otestování. Vyzdvihuji výběr aktuálního tématu, využití strojového učení v oboru informační bezpečnosti, které překračuje obsah vyučovaný v rámci studijní specializace autora práce.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
2. Písemná část práce	*100 (A)*

Popis kritéria:
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. 26/2017, 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.

Komentář:

V relevantnosti jednotlivých kapitol, oddělení vlastních myšlenek od přejatých (jak v textu, tak v dokumentaci programového kódu) a logické struktuře práce neshledávám žádné nedostatky. V závěru každé kapitoly nechybí její shrnutí. Autor čerpá z velkého množství aktuální odborné článkové tvorby a online zdrojů, což je vzhledem k relativní "mladosti" tématu nezbytné. V úvodu do oblasti využití strojového učení pro analýzu log záznamů a pro inspiraci ohledně prováděných analýz jsou v textu využity a citovány i odborné knižní publikace. Oceňuji velký detail práce, její srozumitelnost a názornost, která však zároveň není na úkor odbornosti.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
3. Nepísemná část, přílohy	*100 (A)*

Popis kritéria:
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

Komentář:

Praktickým výstupem práce je funkční prototyp /skript detekující anomálie v bezpečnostních auditních záznamech. Skript byl vytvořen za využití programovacího jazyku Python a moderního frameworku TensorFlow (resp. pomocí nadstavby Keras). Použité nástroje jsou diskutovány v kapitole 5.3 "Použité programovacie jazyky, knižnice a nástroje". Autorův výzkumný proces je dobře zdokumentován a snadno replikovatelný díky prostředí Jupyter Notebook, v němž jsou detailně popsány jednotlivé kroky analýzy (příprava dat, výběr modelu, "ladění" modelu, interpretace výsledků atp.). Git repozitář obsahuje dokumentaci k práci s vytvořeným programovým kódem, v přílohách textu je k dispozici uživatelská příručka, které neopomíjí ani uživatele bez přístupu k datovému souboru, jež byl autorovi poskytnut pod NDA firmou Cisco.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Hodnocení výsledků, jejich využitelnost	*95 (A)*

Popis kritéria:
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.

Komentář:

Teoretická část práce poslouží jako velmi dobrý úvod do problematiky log managementu, GDPR a automatizování detekce podezřelých událostí nad rámec "hardcoded" expertních pravidel. Programový kód z praktické části lze využít v produkčním provozu, a sice jako experimentální podpůrný nástroj pro činnost bezpečnostního analytika. Byť je zjevné, jak na výstupy práce navázat a jak vzniklý programový kód upotřebit v praxi, je škoda, že nezbyl čas na otestování kódu v reálném provozu. Autor by tak dostal feedback nejen na věcnou správnost svého řešení (kterou pokrývají posudky a obhajoba práce), ale i zpětnou vazbu na užitečnost pro konkrétního uživatele /firmu. Vzhledem k ostatním kvalitám práce a jejímu již tak velkému rozsahu se jedná spíše o komentář, nikoliv zásadní výtku (to, co jinak výborné práci "chybí k dokonalosti"). Doporučuji Jakubovi tento krok učinit i po úspěšné obhajobě, jelikož se bude jednat o cennou zkušenost.

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 5:*
5. Aktivita a samostatnost studenta	5a: ***1=výborná aktivita,*** *2=velmi dobrá aktivita,* *3=průměrná aktivita,* *4=slabší, ale ještě dostatečná aktivita,* *5=nedostatečná aktivita* 5b: ***1=výborná samostatnost,*** *2=velmi dobrá samostatnost,* *3=průměrná samostatnost,* *4=slabší, ale ještě dostatečná samostatnost,* *5=nedostatečná samostatnost*

Popis kritéria:
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 (5a). Posuďte schopnost studenta samostatné tvůrčí práce (5b).

Komentář:

Proaktivní a velmi zodpovědný přístup. Důsledná příprava na konzultace a hlídání smluvených deadlines. Zapracování všech připomínek, případně věcná diskuze. Nad rámec těchto bodovitých poznámek vyzdvihuji flexibilní reakci na potíže s obstaráním kvalitního datasetu (který finálně poskytl Jakubův zaměstnavatel, firma Cisco), díky níž nijak nebyla dotčena kvalita práce.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Celkové hodnocení	*99 (A)*

Popis kritéria:
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.

Text hodnocení:

Navrhuji hodnocení klasifikačním stupněm A (výborně). Zajímavé a mezioborové téma analýzy bezpečnostních log záznamů pomocí strojového učení bylo zpracováno velmi zevrubně po teoretické stránce, s praktickým výstupem a včetně detailní dokumentace výzkumného procesu, jež umožňuje snadnou replikovatelnost a navázání na dosažené výsledky. Doporučuji zkušební komisi zvážit navržení této závěrečné práce na Cenu děkana za vynikající DP.

Podpis vedoucího práce: