



Hodnocení vedoucího závěrečné práce

Vedoucí práce:	Mgr. Vojtěch Rybář
Student:	Mikuláš Kočí
Název práce:	Srovnání interpretovatelných modelů s post-hoc vysvětlitelnými black-box modely
Obor / specializace:	Umělá inteligence 2021
Vytvořeno dne:	11. června 2024

Hodnotící kritéria

1. Splnění zadání

[1] zadání splněno

► [2] zadání splněno s menšími výhradami

[3] zadání splněno s většími výhradami

[4] zadání nesplněno

Práce měla za cíl provést rešerši interpretovatelných modelů strojového učení a porovnání jejich úspěšnosti oproti výkonnému black-box modelu. Včetně porovnání dostupných informací o tom, na základě čeho se model rozhodoval, kde u black-box modelu měly být využity metody pro post-hoc vysvětlitelnost, tedy např. metoda SHAP.

Zadání bylo splněno s těmito dvěma menšími výhradami:

1) Pro evaluaci byly využity pouze datasety s malým počtem příznaků, což limituje reálnou využitelnost získaných výsledků.

2) Mohlo být věnováno více pozornosti porovnání vzhledů poskytnutých interpretovatelnými metodami a black-box modelem s aplikovanou post-hoc vysvětlitelnou metodou.

2. Písemná část práce

65 / 100 (D)

Text je dobře strukturovaný, kapitoly na sebe logicky navazují. Práce je však příliš stručná na to, aby dostatečně vysvětlila použité metody, datasety, motivaci některých kroků a prezentovaných obrázků.

Vzhledem k tomu, že hlavním cílem bylo porovnat výkonnost modelů na různých datasetech, mohly být výsledky rovnou prezentovány po datasetech a ne po metodách, jak je v textu práce. Podkapitola věnující se porovnání modelů by mohla být přehledněji strukturována a u porovnání všech modelů uváděno více metrik než jen F1 skóre.

Práce obsahuje menší množství typografických chyb (konsistentně např. chybějící mezera před citací) a překlepů. Reference jsou relevantní a správně využívány.

3. Nepísemná část, přílohy

85 /100 (B)

Příložený kód obsahuje postup tréninku a evaluace rozmanitých přístupů a metod z mnoha různých softwarových balíčků.

Souhrnný Jupyter notebook obsahující kód, kterým byly získány výsledky v písemné části, neobsahuje žádný doprovodný text. Nelze jej tak bez písemné práce brát jako samonosný návod, jak metody aplikovat na jiných problémech.

4. Hodnocení výsledků, jejich využitelnost

89 /100 (B)

Získané výsledky ukazují, že při menším množství příznaků dokáží interpretovatelné metody svou přesností konkurovat state-of-the-art black-box modelům. Zároveň ukazují, že jak specifické interpretovatelné metody trénovat.

Tím ukazují směr, jak aplikovat metody strojového učení v oblastech, kde je kvůli potenciálně závažným důsledkům špatné predikce, potřeba vhléd, jak metoda vytváří své predikce.

5. Aktivita studenta

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

6. Samostatnost studenta

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

Celkové hodnocení

75 /100 (C)

Práce popisuje reprezentativní kategorie interpretovatelných modelů a ukazuje, že v případě menšího množství příznaků dosahují stejné přesnosti jako state-of-the-art modely. Kvalitu práce snižuje příliš stručná textová část.

Instrukce

Splnění zadání

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

Posudte, 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 posudte, 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 posudte schopnost studenta samostatně 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.