FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Magda Friedjungová, Ph.D.
Student:	Danila Makulov
Název práce:	Metody pro vysvětlení predikcí metod strojového učení
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	5. února 2024

Hodnotící kritéria

1. Splnění zadání

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

2. Písemná část práce 78 / 100 (C)

Písemná část práce je psána v anglickém jazyce srozumitelnou formou. Student používá relevantní zdroje (i když poněkud oblíbenou referencí je [2]).

Ke struktuře a obsahu mám několik připomínek. Práce je členěna do 5 kapitol. Členění je víceméně logické, až na kapitolu 3, kdy bych uvítala její rozdělení do více kapitol. Její struktura mi přijde trochu nepřehledná - měly by se zde nacházet experimenty, tedy popis jednotlivých implementací, návrh experimentů, popis dat. Student opět, sic stručně, představuje zvolené metody a pokládá si 3 cíle experimentů. Následně se zaobírá hned druhým cílem, pro který prezentuje výsledky. Navíc se zaobírá primárně disagreement problémem, který je sice v zadání zmíněn v bodě 3, ale dle mého by mu měl v praktické části předcházet bod 2 - samotná aplikace metod vysvětlitelnosti. To následuje později, v kontextu disagreement problému, což je zajímavé pojetí. Jsou zde smíchané experimenty, výsledky a diskuze. Student jako modely volí CatBoost a Random Forest, přičemž v teoretické části se zaobírá klasickou lineární regresí a rozhodovacím stromem - sekce 2.5 - zde by se alespoň v úvodním textu hodilo zmínit i logistickou regresi a kNN. Zároveň by se mi líbilo, kdyby v kapitole 3 byla zmínka, zda někdo na zvolené datasety již použil metody vysvětlitelnosti. Některé sekce obsahují pouze krátký odstavec textu, přestože je teoretická část poměrně rozsáhlá. V 2.10 bych uvítala rozsáhlejší popis metod. Student se vyhnul definici pojmů "interpretability" a "explainability", což pro jeho práci nevadí, ale šlo by o odstavec textu, ref. https://doi.org/10. 1016/j.inffus. 2023.101805. Nejsou mi příliš jasné formulace typu „It can be used for tabular, text and

image data." Student v úvodní části říká, že modely mají na vstupu vždy tabulární data, na která lze převést i obrázky apod. V textu se vyskytuje několik překlepů. Někdy jsou doslovné citace kurzívou, někdy ne. Některé názvy kapitol a sekcí by mohly být výstižnější. Např. 2 Literatura Review spíše odpovídá Methods Overview. U obrázků 3.6 apod. by se hodila opačná barevná škála - čím vyšší korelace, tím tmavší barva. Líbí se mi, že se student věnuje i „disagreement" problému. Ale přišlo by mi zajímavé zohlednit více článků - těch, které vysvětlitelnost používají a těch, které se věnují zmíněnému problému. Sekce 2.8 na několika stránkách tlumočí poznatky pouze jednoho článku. Student také prezentuje metody a experimenty pro obrazová data, ale nepřijde mi, že by této části věnoval stejné úsilí jako tabulárním datům. V daném podání to působí trochu nesourodě.

3. Nepísemná část, přílohy

85 /100 (B)

Chybí readme s popisem, co jednotlivé složky obsahují a jak dané modely, resp. experimenty, spustit. Přiložené jupyter notebooky jsou nicméně funkční a obsahují několik komentářů. Student zvolil vhodné technologie, které by mohly být v písemné práci prezentovány ucelenějším způsobem.

4. Hodnocení výsledků, jejich využitelnost

88 /100 (B)

Student zmapoval aktuální metody XAI primárně pro tabulární data a velkou část primárně praktické práce věnoval "disagreement" problému. V práci prezentuje několik zajímavých poznatků, i když se většinou jedná o srovnání dvou metod na dvou datasetech (k sekci Recommendations bych proto byla trochu skeptická). Práce může poskytnout dobrý základ pro zorientování se v problematice XAI a rozšíření prezentovaných experimentů.

Celkové hodnocení

87 /100 (B)

Student se poměrně dobře zorientoval ve specifické oblasti strojového učení a své znalosti doložil několika zajímavými experimenty.

Otázky k obhajobě

Stručně komisi prosím seznamte s dalšími možnostmi, jak vaše experimenty rozšířit. Na co byste se zaměřil, abyste podpořil poznatky shrnuté v kapitole 4?

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.

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.