



Hodnocení vedoucího závěrečné práce

Vedoucí práce: Ing. Magda Friedjungová, Ph.D.
Student: Bc. Michaela Kučerová
Název práce: Metody pro balancování binárních dat
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 1. 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

Zadání práce bylo splněno v celém rozsahu, v některých bodech i nad rámec zadání.

2. Písemná část práce

95 /100 (A)

Předložená práce je adekvátního rozsahu, je logicky členěna a jednotlivé kapitoly jsou informačně vyvážené. Studentka hned v první kapitole definuje problém a základní pojmy, které jsou potřeba k řešení zadané úlohy a čtenář se s nimi setká napříč celou prací. Od druhé kapitoly se teoretická část zabývá několika metodami pro řešení problému. Metody jsou popsány stručně, ale i tak je kapitola rozsáhlá. V některých případech studentka používá zkratku metody jako název podsekce (např. 2.3.1 TAEI), zde bych raději viděla celý název a zavedení zkratky až v samotném textu. Nicméně zpracování této kapitoly potvrzuje domněnku, že metod pro balancování dat existuje mnoho. Dále studentka představuje 10 metod, které zvolila pro experimenty na 9 datasetech a vlastní generativní metodu, kterou nazvala LIT-GAN. V práci jsou pečlivě popsána veškerá nastavení a jednotlivé kroky experimentů pro jejich opakovatelnost. Studentka využívá několika metrik a přináší rozsáhlé vyhodnocení včetně časové náročnosti jednotlivých algoritmů. Vše je přehledně strukturováno, v práci se vyskytují všechny potřebné reference. Práce je psána v anglickém jazyce, v některých místech by mohla být čtivější, ale to vnímám jako subjektivní a nepřiliš konstruktivní připomínku.

3. Nepísemná část, přílohy

98 /100 (A)

V praktické části práce studentka provedla rozsáhlé experimenty, kdy implementovala nebo využila stávajících implementací 10 metod pro balancování tříd, navrhla vlastní

metodu LIT-GAN a dané ověřila na 9 datasetech. Úspěšnost metod vyhodnotila 6 metrikami za využití 6 klasifikačních modelů (s/bez ladění jejich hyperparametrů). Studentka odevzdala všechny potřebné kódy spolu s readme, kde popisuje jejich spuštění. Předložené experimenty jsou opakovatelné, stálo by za zvážení jejich zveřejnění např. na GitHubu.

4. Hodnocení výsledků, jejich využitelnost

100/100 (A)

Předložená práce je vědecko-výzkumného typu. Přestože studentka navrhla vlastní metodu LIT-GAN (což v dané oblasti považuji za poněkud složité), její práce potvrdila již dosavadní výsledky, a to že neexistuje jedna nejlepší metoda pro řešení problému balancování, ale záleží na konkrétním datasetu a míře nevyvážení. Také bylo ukázáno, že si pořád velmi dobře vedou metody založené na SMOTE oproti novodobým metodám, které např. využívají neuronové sítě, a že v mnoha případech není nutné sahat po komplikovaném řešení. Mimo rozsáhlé experimenty studentka poskytuje i obsáhlou řešerši, na základě které lze usoudit, že metod pro balancování existuje celá řada a danou problematiku bychom tak mohli považovat za vytěženou. Přesto se studentka tohoto poznatku nezalekla a navrhla a experimentálně ověřila vlastní metodu, což vyžadovalo značné úsilí a pochopení dané problematiky a stávajících metod. Navržená metoda akceptuje různé interpolační metody, se kterými studentka provedla také několik experimentů. Na výsledky práce lze dále navázat, ať už dalšími experimenty, vylepšením navržené metody nebo aplikací statistických testů na vyhodnocení experimentů.

5. Aktivita studenta

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

Studentka pracovala aktivně, přicházela s vlastními nápady, dodržovala dohodnuté termíny a své postupy průběžně konzultovala.

6. Samostatnost studenta

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

Studentka pracovala samostatně, nezalekla se i obtížnějších úkonů a pozitivně se stavěla k řešení vyvstalých problémů.

Celkové hodnocení

99/100 (A)

Práci navrhuji hodnotit klasifikačním stupněm A, jelikož se jedná o kvalitní a rozsáhlou práci, která má pro danou oblast svůj přínos.

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.