



Hodnocení vedoucího závěrečné práce

Vedoucí práce: Ing. Karel Klouda, Ph.D.
Student: Bc. Filip Kojan
Název práce: Vytěžování informací z textů inzerátů pro prodeje automobilů
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 19. května 2021

Hodnotící kritéria

1. Splnění zadání

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

Zadání bylo splněno ve všech bodech. Jednalo se o aplikaci metod strojového učení na reálná data a to dokonce v rámci spolupráce s firmou.⁹²

2. Písemná část práce 92 /100 (A)

Práce je přehledně a srozumitelně sepsána, z textu je jasně pochopitelné, co, jak a proč bylo uděláno a jaké jsou výsledky. Stručně jsou popsány i méně obvyklé koncepty a metody strojového učení, které byly v práci využity.

3. Nepísemná část, přílohy 90 /100 (A)

Nepísemná část práce má formu skriptů v jazyce Python většinou v podobě Jupyter notebooků. Tyto skripty postupně provádějí všechny fáze práce s daty, od jejich zpracování (příp. i získání) až po aplikaci a vyhodnocení různých modelů. Tato forma zpracování je poměrně standardní a odpovídá přesně povaze práce, která měla zejména ověřit, zda daný problém je řešitelný a jakých výsledků je možné dosáhnout. Jupyter notebooky by mohly obsahovat více komentářů a být více přehledné, díky struktuře souborů je ale celkem jasné, co obsahují.

4. Hodnocení výsledků, jejich využitelnost 98 /100 (A)

Podívám-li se na práci jako na experiment, který měl zodpovědět na otázku, zda lze spolehlivě strojově získávat strukturalizované informace z textového inzerátu, musím jej hodnotit jako úspěšný a velmi poctivě provedený. Nejvíce na práci pana Kojana oceňuji

právě to, jak problém a vstupní data zanalyzoval a dle analýzy následně provedl experimenty. Kvůli různorodosti datových sad z jednotlivých zdrojů nebylo nastavení experimentů zcela přímočaré. Také velmi oceňuji pestrost přístupů, které autor volil podle povahy problému: Od použití regulárních výrazů či jednoduchých modelů po složitější neuronové sítě a gradient boosting. Výsledkem je pak jasný návod, kdy stačí jednoduchý přístup a kdy je třeba využít sofistikovanější nástroje.

5. Aktivita studenta

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

Práci komplikovalo pomalé získání dat od spolupracující firmy. Autor si ovšem iniciativně poradil a menší datovou sadu, na které šlo ověřit základní postupy, si vytvořil sám. Když následně data dorazila, byl schopen velmi rychle začít se samotnými experimenty a díky tomu vše stihl a nedostal se do časové tísně. Celkově musím hodnotit aktivitu studenta jako výbornou.

6. Samostatnost studenta

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

Student byl zcela samostatný, pravidelné konzultace byly věnovány jen reportování postupu a spíše koncepčním rozhodnutím.

Celkové hodnocení

95 /100 (A)

Práci považuji za ukázkové řešení zadaného praktického problému s (rozumným) využitím strojového učení. Výsledky jasně ukazují, že získávání strukturovaných dat z textů inzerátů je možné a navíc dávají i jasný návod, v jakých případech se má použít jaké řešení a jaké lze čekat problémy a výsledky. Práci proto navrhuji hodnotit jako výbornou.

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. 26/2017, 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.