

I. IDENTIFIKAČNÍ ÚDAJE

| | |
|-----------------------------------|---|
| Název práce: | Hledání častých podmnožin a multi-label klasifikace uživatelských odpovědí na téma wellbeingu |
| Jméno autora: | Lucie Borovičková |
| Typ práce: | bakalářská |
| Fakulta/ústav: | Fakulta elektrotechnická (FEL) |
| Katedra/ústav: | Katedra kybernetiky FEL ČVUT |
| Oponent práce: | Doc. Ing. Jiří Kléma, PhD. |
| Pracoviště oponenta práce: | Katedra počítačů FEL ČVUT |

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

| | |
|---|-------------------------|
| Zadání | průměrně náročné |
| <i>Hodnocení náročnosti zadání závěrečné práce.</i> | |
| Zadání předpokládá předzpracování dotazníků spokojenosti na pracovišti, které obsahují textové odpovědi i odpovědi vyjadřující spokojenost zaměstnanců na číselné škále. Následně jsou dotazníky klasifikovány dle klíčových témat, ke kterým se vyjadřují a úkolem je i zjistit, zda mezi výskytem témat existují statisticky významné souvislosti. Zadání je poměrně náročné pokud jde o rozsah řešených úloh, na druhou stranu, zpracovávaná data jsou na dnešní poměry (velmi pravděpodobně) malého rozsahu a řešení problému explicitně nevyžaduje použití složitějších metod. | |

| | |
|--|------------------------------------|
| Splnění zadání | splněno s menšími výhradami |
| <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i> | |
| Důležité body zadání byly splněny. Menší výhrady zahrnují: chybějící rešerše metod předzpracování textových dat, nesplnění volitelného bodu navrhujičního použití neuronových sítí ke klasifikaci témat. | |

| | |
|--|----------------|
| Zvolený postup řešení | správný |
| <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i> | |
| Studentka volí správná řešení jednotlivých úloh, všechna dílčí řešení však volí přímočará a jednodušší (TFIDF vektorizace velmi krátkých textových vstupů, neřešený vztah mezi vstupy různého typu, multi-label klasifikace pojatá jako sekvenční řešení binárních klasifikačních úloh, částečně manuální a ad hoc zpracování výsledných častých množin položek). Vzhledem k šíři zadání a tomu, že jde o bakalářskou práci, je to pochopitelné a v pořádku. Hodnocení vynikající by ale podle mého názoru některé takové pokusy předjímalo. | |

| | |
|--|------------------------|
| Odborná úroveň | B - velmi dobře |
| <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i> | |
| Pro odbornou úroveň práce platí stejná úvaha jako pro zvolený postup řešení, tj. ty postupy, které byly aplikovány, byly pravděpodobně aplikovány korektně (neprováděl jsem revizi kódu apod.). Úvahy uvedené v práci jsou správné a demonstrují, že řešení bakalářské práce přispělo k tomu, že se studentka seznámila s řadou často používaných postupů strojového učení a dolování dat. | |

| | |
|--|------------------|
| Formální a jazyková úroveň, rozsah práce | C - dobře |
| <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i> | |
| Práce je psána velmi dobrou angličtinou, k jazykové úrovni výhrady nemám. Rozsah práce je spíše nadstandardní. Za ne úplně ideální ale pokládám to, jak je práce strukturována, respektive jakým tématům je věnován jak velký prostor. Osobně bych nepokládal za tak důležitý například obecný rozbor hodnotících metrik, který má v práci 6 stran. Naproti tomu je nedostatečně diskutován rozsah dostupných dat. Doprovodný github patrně obsahuje jen náhled na vstupní soubory, kromě zadání jsem ale nikde nenašel jejich skutečnou velikost. Podobně by bylo vhodné mít k dispozici i další základní | |

statistiky (vyplněnost, délka textových odpovědí, základní přehled demografie, apod.). Bez těchto údajů si nelze udělat obrázek o složitosti a využitelnosti řešení.

Výběr zdrojů, korektnost citací

B - velmi dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Studentka pracuje spíše s tradičními zdroji, některé postupy jsou asi předurčeny dostupností v Pythonu (frequent itemsets + chí-kvadrát test). Reference mají až na výjimky jednotný formát ([Bla08]). Vlastní výsledky jsou od literatury přirozeně odděleny.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Jde o dobře odvedenou práci, která splňuje klíčové body zadání a může být základem prakticky využitelného pracovního postupu pro vyhodnocení zaměstnaneckých dotazníků. Rozsah odvedené práce odpovídá nárokům kladeným na bakalářskou práci. Práce má i jasně pojmenovatelné nedostatky, pro které slovo dobře doslova určuje první návrh mého konečného hodnocení. Pokud se u obhajoby prokáže, že jsem špatně odhadl skutečný rozsah vstupních dat nebo složitost prací, hodnocení by mohlo být i velmi dobré.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

1. Mohla byste komisi blíže seznámit s původem a rozsahem Vašich vstupních dat?
2. V úvodu kapitoly 2 píšete o důležitosti vysvětlitelnosti rozhodnutí klasifikátorů v NLP úlohách. Jak konkrétně se tato úvaha promítla do Vaší práce?
3. Pracujete s velmi krátkými texty. Zohlednila jste stručnost odpovědí při návrhu svého řešení? Mám na mysli specifika tzv. "short text classification"?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Datum: 26.5.2021

Podpis: