



POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

- Jméno studenta:** Tomáš Brzek
- Název práce:** Data Science: Principy, technologie a znalosti.
- Autor posudku:** Tomáš Kozel
- Cíl práce:** Sestavení přehledu a porovnání definic, metod a technologií pro tuto oblast. Pokusit se detekovat prospektivní nezbytné znalosti. Ideálně pohledem studenta AI a jeho současného obsahu vzdělávání.

Povinná kritéria hodnocení práce	Stupeň hodnocení					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	x					
Vymezení cíle a jeho naplnění			x			
Zpracování teoretických aspektů tématu				x		
Zpracování praktických aspektů tématu		x				
Adekvátnost použitých metod			x			
Hloubka a správnost provedené analýzy				x		
Práce s literaturou				x		
Logická stavba a členění práce			x			
Jazyková a terminologická úroveň						x
Formální úprava a náležitosti práce					x	
Vlastní přínos studenta				x		
Využitelnost výsledků práce v teorii (praxi)			x			

Dílčí připomínky a náměty:

Práce obsahuje řadu pravopisných, stylistických a formálních chyb. Namátkou je zdokumentuji v následujících bodech:

- práce používá nestandardní formát - řádkování cca 1 a zarovnání textu výhradně doleva,
- místy je používána „ich forma“,
- časté chyby v přičestí minulém
 - s. 3 - „*prvky, hodnoty... by měli*“,
 - s. 6 - „*korporace si byli nuceni uvědomit...*“,
 - s. 9 - „*data odpovídaly*“,
 - s. 18 - tabulka, 2.ř. - „*Jaké komedie vyhráli ocenění*“,
- s. 7 - vyšínutí z vazby (v lepším případě) - „*Hlavními charakteristiky těchto dat ...*“
- s. 11 - „*...mají mnoho společného s poli, které jsou...*“
- není používáno zcela konzistentní odkazování na zdroje citací, např.
 - v úvodu je zmiňován magazín Harvard Business Review bez odkazu na číslo zdroje,
 - s. 7 - definice pojmu „*Big Data*“,
 - s. 23 - zmiňován blog R. A Meuchena bez čísla zdroje, apod.
- s. 35 - pod obrázkem 9.1.1. - zmatek v používání pojmů kraj a okres.

Celkové posouzení práce a zdůvodnění výsledné známky:

Téma práce hodnotím jako vysoce aktuální a užitečné. Na druhou stranu jde o téma, které je zjevně ještě pro studenta bakalářského studia vcelku náročné a pokud by bylo realizováno až v závěru magisterského studia, dokázal by se autor s tématem „poprat“ pravděpodobně lépe. Záslužná je jistě snaha o ucelený pohled na stále ještě mladou a ne zcela teoreticky usazenou problematiku. Aby však takový text poskytl komplexní pohled na věc, bylo by třeba jít v řadě aspektů více do hloubky. I zde ale platí výše uvedená poznámka o celkové náročnosti tématu pro úroveň aplikovaného informatika - bakaláře. Na textu je vidět snaha se s danou problematikou poprat a částečně se to daří. Velkým nedostatkem práce je ale jazyková stránka. Práce je proseta řadou pravopisných chyb, které studentovo snažení značně degradují. Pokud by formální a jazyková stránka práce byla v pořádku, bylo by možné ji hodnotit někde na hranici známky C. Díky naprosto nedostatečné závěrečné korektuře mi nezbývá než navrhnout v případě úspěšné obhajoby známku E.

Otázky k obhajobě:

1. Jaké další nástroje pro práci s velkými daty vedle R a Hadoop jste při zpracovávání podkladů pro práci poznal?
2. Které předměty studia na FIM (neomezujte se jen na bakalářské studium AI) byste považoval za klíčové pro přípravu na pozici Data Scientist ?

Práci doporučuji k obhajobě.

Navržená výsledná známka: E - dobře (v případě úspěšné obhajoby)

V Hradci Králové, dne 16. května 2017

podpis