



POSUDEK VEDOUCÍHO BAKALÁŘSKÉ PRÁCE

Jméno studenta: Tomáš Brzek

Název práce: Data Science: Principy, technologie a znalosti

Autor posudku: prof. RNDr. Hana Skalská, CSc.

Cíl práce: Sestavení přehledu a porovnání definic, metod a technologií pro tuto oblast.

Pokusit se detekovat prospektivní znalosti, nezbytné pro mezioborovou komunikaci při návrhu aplikace tohoto typu. Ideálně pohledem studenta AI a jeho současného obsahu Bc studia na FIM.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)			
	A	C	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Dílicí připomínky a náměty:

Struktura práce: Kapitola 4 (Big data) má jen dvě strany, mohla být zařazená jako podkapitola v kapitole 3. V závěru (před kapitolou 5) by bylo vhodné začlenit stručné shrnutí pojmů i hlavních myšlenek, na tomto základu vysvětlit (zdůvodnit) koncepci a stručně obsah následujících částí práce).

Na straně 27 (tabulka dole) vytvoření `data.frame` se týká datového typu, ostatní příkazy jsou funkce, bylo by dobré oddělit popisy těchto různorodých typů. Nepřesná terminologie (například na str. 27: Které jsou „primitivní“ funkce v R?). Datové typy jsou popsány velmi stručně, mohly být zmíněny všechny typy, které jsou v práci použité.

K vlastní aplikaci: Chybí odkaz na zdroj dat. K příkladu I: Načítání dat ze schránky je nešťastná volba, před vlastní aplikací mohly být popsány aspoň některé varianty čtení dat ze souboru. V příkladu chybí definování formátu, obecně neplatí, že je automaticky číselný. Vysvětlení výsledků simulace výběrů je nesprávně formulované (str. 36, poslední odstavec – „opakování počtu pokusů tolikrát, než

se stane extrém“ – nelze, počet je Vámi zadaný, nebo „25 % vzorků je vyšších než 140300“ – nelze – jak může být vzorek vyšší než hodnota?). Pro opakovatelnost příkladu by bylo vhodné volit seed generátoru. Rozdíl mezi výsledky příkazů *summary* a *quantile* může být vysvětlen i jinými důvody. Funkce *MyMod* (str. 34) v práci asi není popsána, jenom je použita.

Příklad II. Je poměrně zdařilý a používá pokročilejší funkce a příkazy.

V závěru kapitoly mohly být přehledně shrnuté a stručně popsané použité funkce a příkazy.

V celé práci, zejména v částech na str. 1 – 30 nelze přehlédnout neobratné formulace, drobné překlepy, nesprávnou interpunkci, používání expresivních, ne odborných, nebo popularizačních výrazů a řadu čistě gramatických prohřešků. Na tyto problémy adresně upozorňuji v tištěné verzi práce. Tyto chyby většinou nemusely vzniknout, student byl včas upozorněn.

Celkové posouzení práce a zdůvodnění výsledné známky:

Student se v práci zabývá problematikou, se kterou se nesetkal během studia, vychází převážně z cizojazyčné literatury a pracoval se zájmem o danou oblast. Rozšířil si svůj pohled na problematiku analýzy dat. Obsah i zaměření práce pravidelně konzultoval, snažil se o získání nadhledu nad problematikou.

Otázky k obhajobě:

Domníváte se, že začlenění R do výuky statistiky by bylo pro studenty přínosné? Jaký model pro kombinaci R s jiným předmětem byste preferoval?

Práci doporučuji k obhajobě.

Navržená výsledná známka: B - výborně-velmi dobře

V Hradci Králové, dne 14. května 2017

podpis