



## POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

**Jméno studenta:** Bc. Lukáš Zasadil

**Název práce:** Data mining a freeware

**Autor posudku:** Hana Skalská

**Cíl práce:** Vyhledat a sestavit přehled reprezentantů komerčních a freeware produktů pro data mining. Popsat požadované funkcionality, porovnat vybrané produkty z této oblasti podle zvolených kritérií.

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

### Vyjádření k výsledku anti-plagiátorské kontroly

Práce byla vyhodnocena v Odevzdej.cz (celková podobnost 0%).

### Dílicí připomínky a náměty:

Struktura práce neodpovídá Metodickým pokynům (cíl práce a metodika až na str. 26).

V citované literatuře převažují webové zdroje, případně starší literatura.

Vysvětlení základních pojmů (data mining - DM, big data, typy dat, aplikační oblast) je neuspořádané až zavádějící (nesprávná interpretace, odkazy na nevhodné webové zdroje, matoucí je popis souvislosti mezi počátky DM a Bayesovým pojetím pravděpodobnosti).

Charakteristiky datových zdrojů, formátů a datových typů nejsou popsány systematicky, nesprávně je vysvětlený termín six sigma, přehled analytických metod převážně omezený na výčet z literatury roku 2003, apod.

Chybí jednotící myšlenky, které by představily studovanou problematiku.

Zdroje obrázků jsou uvedené jen v příloze seznamu obrázků, nikoliv u obrázků.

**Praktická část** je nevhodně vložena do kapitoly Metodologie.

Metodika práce je popsána jen zběžně a nepřesně (například software bude představený cituji: „ze všech úhlů“, hodnotící kritéria uvedená až u výsledků.).

### **Celkové posouzení práce a zdůvodnění výsledné známky:**

Výsledkem práce jsou volba a popis softwaru Weka, Rapid Miner a R. Ukázky a výpočty jsou provedeny pro málo rozsáhlá a jednoduchá data Iris, tedy ne pro typickou úlohu DM. Specifická zvolená hlediska pro porovnání softwaru jsou uvedena až v kapitole výsledků.

**Výsledky porovnání softwaru a jejich rozbor** jsou jedním z cílů práce. Definovaná kritéria jsou vyhodnocena metodou AHP pomocí softwaru Criteria Decision Plus. Výsledné porovnání vychází z posouzení jedním uživatelem a nelze ho zobecnit

Student prokázal určitou schopnost navrhnout postup a popis řešení zadané problematiky, zmíněné slabší stránky práce jsou vyjádřené v celkovém hodnocení.

Práci doporučuji k obhajobě v oboru Informační management, hodnocení D.

### **Otázky k obhajobě:**

Seznam reprezentantů vybraných k porovnání vycházel z průzkumu roku 2013, ve kterém se zjišťovat nejčastěji užívaný software v reálných aplikacích.

Změnily se (podle novějších průzkumů) do současnosti tyto preference? Jak byste případné změny charakterizoval?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: D**

**V Hradci Králové, dne 6. května 2021**

---

**podpis**