



POSUDEK VEDOUCÍHO BAKALÁŘSKÉ PRÁCE

Jméno studenta: David Illner

Název práce: Data mining a možnosti nekomerčního softwaru

Autor posudku: Hana Skalská

Cíl práce: Úloha data mining v procesu vyhledávání znalostí z dat (KDD) a možnosti nekomerčního softwaru (například Rattle, Orange nebo Weka).

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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input 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"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input 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"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input 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"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input 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"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Vyjádření k výsledku anti-plagiátorské kontroly

Celková podobnost: 0 % (k dokumentu nebyly nalezeny podobnosti v repozitáři Odevzdej.cz)

Dílčí připomínky a náměty:

K práci nemám připomínky zásadního charakteru. Student průběžně komunikoval a adekvátně reagoval na doporučení a připomínky k předkládaným verzím práce. Některé prohřešky se i přesto nepodařilo definitivně odfiltrovat, například: gramatické (i – y), formátování literatury a podobně.

Celkové posouzení práce a zdůvodnění výsledné známky:

Teoretická část práce vychází z rešerše k problematice. Vysvětluje pojem KDD a popisuje jednotlivé kroky tohoto procesu, porovnává metodiky SEMMA a CRISP-DM. Podrobněji se zabývá kroky přípravy dat pro analýzu, zejména postupy čištění dat, transformace dat

a explorace dat. Pro aplikaci student zvolil úlohu klasifikace, proto v teoretické části popisuje metody lineární diskriminační analýzy a logistické regrese.

Aplikace je vytvořena v prostředí Rattle, které poskytuje poměrně příznivé uživatelské rozhraní pro použití řady pokročilých balíčků programu R z oblasti klasifikace, explorace i vizualizace a zpřístupňuje tak R pro širší okruh uživatelů.

Jednotlivé kroky aplikace jsou logicky provázané. Pro zvolená data jsou pomocí Rattle popsány a realizované kroky procesu od přípravy dat až po zhodnocení kvality modelu metodou ROC křivky.

Student se tématu práce věnoval se zaujetím, obsah pravidelně konzultoval a ve finále prokázal zvládnutí problematiky, se kterou se z velké části nesetkal během bakalářského studia. Prokazuje schopnost nastudovat a aplikovat poznatky daného tématu v rozsahu, který odpovídá požadavkům na bakalářskou práci.

Přestože se v práci nepodařilo zcela odstranit některé prohřešky (formální, stylistické i gramatické), navrhuji výsledné hodnocení A.

Otázky k obhajobě:

Zdůvodněte svoji volbu Rattle oproti původně uvažovaným Orange a Weka.

Práci doporučuji k obhajobě.

Navržená výsledná známka: A

V Hradci Králové, dne 13. května 2020

podpis