



POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Sharon Moscato

Název práce: Dolování dat z vybrané datové sady pro predikci podvodných pracovních nabídek

Autor posudku: Ing. Patrik Urbaník

Cíl práce: Cílem této diplomové práce je prozkoumat problematiku dobývání znalostí z databází a demonstrovat aplikaci predikčního modelu na reálných datech. Práce se zaměřuje na analýzu datového souboru obsahujícího pracovní nabídky, které jsou rozděleny do dvou kategorií: podvodné a pravdivé. Hlavním cílem je vytvořit predikční model, který bude schopen identifikovat podvodné nabídky práce.

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 type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input checked="" type="checkbox"/>	<input 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

Anti-plagiátorská kontrola eVŠKP identifikovala celkovou podobnost: 1 %.

Dílní připomínky a náměty:

Mám pouze pár připomínek, které se týkají menších kosmetických nedostatků:

- Zdroje jsou součástí popisků obrázků a tabulek, což se promítlo do ne moc hezky vypadajícího seznamu obrázků a tabulek. Doporučuji zdroje psát na samostatný řádek.
- Rozdělení práce na teoretickou a praktickou část pomocí nadpisů první úrovně není příliš šikovné, jelikož se tím znemožní použití dalších nadpisů první úrovně v celé dané části. Celý obsah je pak zpracován jen pomocí 2. a 3. úrovně nadpisů.

- Důsledkem toho vznikl nedostatečný prostor pro členění obsahu, který se nejvíce promítl v části 4.3.3 *Fáze 3: Příprava dat* a dále, kde jsou vnořeny další 3 úrovně „nadpisů“, ale z jejich formátování není na první pohled jasné jejich pořadí.

Celkové posouzení práce a zdůvodnění výsledné známky:

Předložená diplomová práce se zabývá problematikou dolování znalostí za účelem tvorby predikčního modelu pro identifikaci podvodných pracovních nabídek.

Práce je klasicky členěna na teoretickou a praktickou část. V teoretické části studentka seznamuje čtenáře se základy řešené problematiky, tzn. představení a vysvětlení relevantních pojmů a metod a také velmi podrobně popisuje metodiku pro dolování dat CRISP-DM. Popisované metody a samotnou CRISP-DM metodiku následně aplikuje při tvorbě vlastního modelu v praktické části.

V úvodu praktické části studentka popisuje použité nástroje, vybraná prostředí a zvolený datový soubor, který byl použit jako základ pro klasifikaci pracovních nabídek. Praktická část následně obsahuje zpracované jednotlivé fáze podle metodiky CRISP-DM. Především kapitoly/fáze 5.4 *Rozbor datového souboru* a 5.5 *Příprava dat* velmi detailně popisují, charakterizují a rozebírají použitý datový soubor a jednotlivé proměnné/sloupce.

Pro vlastní modelování byly implementovány dvě metody, a to Logistická regrese a Náhodný les. V závěru praktické části byly výstupy modelování řádně interpretovány, porovnány a zdůvodněny.

Struktura práce je přehledná, kapitoly na sebe logicky navazují a až na drobnosti, zmíněné v předchozí části, je formální zpracování povedené. Terminologie a odbornost je také na dobré úrovni. Veškeré pojmy a metody jsou v práci popsány velmi detailně a srozumitelně. Díky implementaci metodiky CRISP-DM má praktická část jasný řád. Za velké plus považuji i to, že v práci byly použity dvě různé metody a jejich výstupy porovnány, jelikož logistická regrese se ukázala jako nevhodná metoda pro řešený problém. Kladně hodnotím i výběr použité literatury, která se z velké části sestává z odborných článků.

Celkově práci hodnotím jako velmi zdařilou.

Otázky k obhajobě:

-

Práci doporučuji k obhajobě.

Navržená výsledná známka: A

V Hradci Králové, dne 3. září 2023

podpis