

Posudek vedoucího na bakalářskou práci

Název práce: Metoda CHAID a logistické regrese

Autor práce: Jaroslav Novák

Vedoucí: Ing. Adam Čabla

Oponent: Ing. Martin Matějka

Předkládaná bakalářská práce si klade za cíl představit základy metod z názvu a porovnat je z hlediska jejich výstupů a možné interpretace a to pomocí praktického příkladu na reálných datech. Téma se tedy týká problematiky analýzy závislosti, kdy vysvětlovanou proměnnou je proměnná kategoriální, zde konkrétněji binární. Obě metody se již ve vypracovaných tématech bakalářských prací vyskytly, autor bohužel zapomíná konkrétně svou práci odlišit. Její téma i cíl pokládám obecně za vhodné, obě metody se při praktickém použití totiž vzájemně doplňují.

Autor popisuje obě metody spolu s metodami hodnocení logistické regrese a klasifikační schopnosti v prvních dvou kapitolách velmi pečlivě, místy skoro až zdlouhavě. Jazyková skladba je místy kostrbatá, ale nakonec se vždy lze dopátrat toho, co chtěl autor přesně říci. Tomu dopomáhají i vhodné ilustrace.

Třetí kapitola obsahuje řešený příklad na datech z Českého sociálněvědního datového archivu Akademie věd ČR. Příklad je zvolený vhodně, protože dává vyniknout podobnostem i odlišnostem obou metod. Autor pečlivě popisuje zvolený postup včetně všech použitých nastavení. Text obsahuje všechny zásadní výstupy ve formě tabulek a grafů a tyto jsou správně popsány a podrobně vysvětleny v textu. Opět některé pasáže působí až zdlouhavě. Autor v této části dokládá zvládnutí postupu výpočtu obou metod ve statistickém softwaru i téměř úplné pochopení (viz otázka 2) jejich výstupů.

V závěru pak shrnuje zjištěné výhody, nevýhody a odlišnosti obou metod, kdy dochází k závěru, že je vhodné je užívat spolu.

Celková úprava práce působí pečlivým dojmem až na pár drobností (třeba text ve vzorci 2.1.4, automatické upozornění Wordu na str. 25, či občas chybějící indexy parametrů v popisu logistické regrese).

Autor pracuje s přiměřeně obsáhlou literaturou a pečlivě ji v textu odkazuje.

Celkově práci pokládám za zdařilou a doporučuji ji k obhajobě a v případě uspokojivého zodpovězení dotazů ji navrhuji hodnotit známkou **výborně**.

Dotazy k obhajobě:

1. V čem se vaše práce odlišuje od jiných prací na podobná témata zpracovaných na VŠE?
2. V komentáři si všímáte toho, že metoda CHAID nezachytí vliv vzdělání na tendenci ke spoření, který je přitom dle logistické regrese velmi silný. Čím si to vysvětlujete?
3. Ve výstupu logistické regrese máte množství kategorií vysvětlujících proměnných, které se statisticky významně neliší od zvolené referenční kategorie. Můžete popsat postup, jak je eliminovat v případě nominální a v případě ordinální škály?