

VYUŽITÍ DATA MININGOVÝCH METOD PŘI ZPRACOVÁNÍ DAT Z DEMOGRAFICKÝCH ŠETŘENÍ

Abstrakt

Cílem předkládané práce bylo popsat a následně demonstrovat na modelové úloze principy procesu dolování znalostí z databází, často označovaného jako data mining (DM). V teoretické části práce jsou popsány vybrané metodiky, na základě kterých se postupuje při DM procesu a dále jsou zjednodušeně popsány principy vybraných DM technik. V druhé části práce je pak realizována DM úloha, ve které se postupuje dle metodiky CRISP-DM. Jako modelová data pro tuto úlohu jsou vybrána data z výběrového šetření American Community Survey. Praktická část práce je rozdělena na dvě části. V první části je vyhotovena klasifikační úloha, jejíž cílem je zjistit, zda lze využít vybrané DM techniky k řešení problematiky chybějících údajů ve statistických šetřeních. Úspěšnost klasifikace a následné predikce hodnot u vybraných atributů se pohybovala v intervalu 55–80 %. Druhá část praktické části práce je pak zaměřena na hledání zajímavých znalostí ve vybraných datech pomocí asociačních pravidel a metody GUHA.

Klíčová slova: data mining, dolování znalostí z databází, statistická šetření, chybějící hodnoty, klasifikace, asociační pravidla, metoda GUHA, ACS