Juraj Holas

Návrh projektu

## Téma:

Odhadovanie cien mobilných telefónov na základe ich parametrov a historickej ceny.

## Popis:

Projekt sa zameria na odhadovanie cien vybraných modelov mobilných telefónov (smartfónov) v slovenských internetových obchodoch. Skúmaná bude minimálna a priemerná cena jednotlivých modelov skrz niekoľko desiatok obchodov. Bude zisťovaný vzťah aktuálnej (resp. predpovedanej) ceny vzhľadom ku technickým parametrom telefónu, dátumu jeho uvedenia na trh, a počiatočnej predajnej cene.

## Zdroje dát:

Údaje potrebné pre tento projekt budú získavané z dvoch zdrojov:

1. Všetky technické parametre telefónov a dátumy uvedenia na trh budú čerpané z portálu gsmarena.com, kde sú verejne dostupné profily všetkých skúmaných modelov.
2. Údaje o cenách, najmä vývoj minimálnej a priemernej ceny produktu v slovenských internetových obchodoch, poskytne spoločnosť MINET s.r.o., prevádzkovateľ portálu najnakup.sk . Spoločnosť už vyjadrila súhlas na poskytnutie týchto údajov pre účely projektu.

V prípade potreby pribudnú ďalšie zdroje v priebehu práce na projekte.

Počet zahrnutých modelov telefónov bude cca 25, pričom budú zahŕňať modely ako vyššej, tak aj strenej a nižšej úrovne, modely rôznych výrobcov, a modely s rôznym dátumom prvého uvedenia na trh (v rozmedzí cca posledných dvoch rokov).

## Rozdelenie dát:

Údaje budú rozdelené na trénovaciu a testovaciu množinu podľa času – hodnoty minimálnej a priemernej ceny do určitého mesiaca budú slúžiť na trénovanie, všetky neskoršie na testovanie. Táto časová hranica bude stanovená približne na začiatok 4. kvartálu 2014, avšak pri príliš nových modeloch (napr. dátum uvedenia na svetový trh neskôr ako v 3. kvartáli 2014) bude táto hranica posunutá na neskoršie obdobie, aby boli nejaké dáta o týchto modeloch obsiahnuté aj v trénovacej množine.

## Vyhodnocovanie úspešnosti:

Po získaní požadovaného algoritmu z trénovacích dát sa spočíta jeho úspešnosť vzhľadom ku testovacím dátam. Chyba na každom odhadovanom atribúte (minimálna a priemerná cena) bude reprezentovaná ako percentuálna odchýlka od reálnej hodnoty tohto atribútu. Celková chyba bude následne vyhodnotená ako priemer chýb na jednotlivých atribútoch.