

VYSOKÉ UČENÍ TECHNICKÉ FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Získávání znalostí z databází Databáze restaurací - zadání

1 Databáze restaurací

Pro účely projektu byla vybrána data z databáze restaurací, která byla získána *Národním centrem pro výzkum a technologický rozvoj CENIDET* v Mexiku. Autory jsou *Rafael Ponce Medellín* a *Juan Gabriel González SernaTato*. Data byla použita v roce 2011 při výzkumu účinků příslušných kontextových prvků na výkon systému restaurací, kde bylo úkolem vygenerovat seznam top-restaurací podle zákaznických preferencí. Zároveň článek zmiňuje i problém vybírání důležitých atributů pro relevantní určení seznamu. [1]

1.1 Popis dat

Databáze obsahuje celkem 9 souborů ve formátu CSV. Z nich jsou v 5 souborech uloženy informace o jednotlivých restauracích, které lze spojit pomocí id restaurace, jež je obsaženo v každém z těchto 5 souborů. Dále tyto soubory obsahují následující informace o restauracích:

1. typ kuchyně
2. otevírací doba (hodina a den)
3. podporovaný typ platby
4. typ parkovacího místa
5. zeměpisné souřadnice, název, adresa, město, stát, země, fax, PSČ, prodej alkoholu, kuřácká oblast, dress code, přístupnost, výšku cen, web, atmosféra, franšíza, typ prostor, další služby

V dalších 3 souborech jsou pak obsaženy informace pouze o zákazníkovi, které jsou tentokrát propojeny pomocí id zákazníka. Kromě id zákazníka pak obsahují soubory tyto informace:

1. preferovaná kuchyně
2. preferovaný typ platby
3. zeměpisné souřadnice bydliště, kuřák, úroveň pití, preferovaný styl oblékání, atmosféra, způsob dopravy, rodinný stav, potomci, rok narození, zájmy, osobnost, náboženství, aktivity, oblíbená barva, hmotnost, rozpočet, výška

V posledním souboru jsou pak uloženy informace o hodnocení zákazníka dané restaurace. Kromě id restaurace a id zákazníka jsou zde atributy:

1. celkové hodnocení, hodnocení jídla a hodnocení služeb

1.2 Úloha

Úlohou projektu bude predikovat spokojenost zákazníka s restaurací na základě informací o restauraci a zákazníkovi, kde spokojenost bude vycházet z minulých hodnocení restaurací klienty. Nejlepší známka odpovídá hodnotě dvě a nejhorší nule. Nejprve bude probíhat analýza vhodných atributů pro predikci a následně budou vyzkoušeny a porovnány různé metody a postupy, ze kterých bude následně vybrána ta metoda s nejlepšími výsledky. Tato predikce může do budoucna pomoci zákazníkovi ke snadnějšímu vyhledávání restaurace, která by měla nejvíce uspokojit jeho potřeby.

Reference

- [1] *Blanca Vargas-Govea, Juan Gabriel González-Serna, Rafael Ponce-Medellín. Effects of relevant contextual features in the performance of a restaurant recommender system. In RecSys'11: Workshop on Context Aware Recommender Systems (CARS-2011), Chicago, IL, USA, October 23, 2011, Online: <http://ceur-ws.org/Vol-791/paper8.pdf>*