

Zadání

Pro účely tohoto projektu do předmětu Získávání znalostí z databází, byla vybrána data „databáze hodnocení hotelů.“ Jedná se o dataset reviews uživatelů Trip Advisoru.

Popis datasetu

Zmíněný dataset obsahuje následující atributy:

1. USER COUNTRY – stát, ze kterého daný uživatel pochází (polynomial)
2. NR. REVIEWS – celkový počet hodnocení, která uživatel napsal (integer)
3. NR. HOTEL REVIEWS – počet hodnocení hotelů, která uživatel napsal (integer)
4. HELPFUL VOTES – hlasy hodnotící nápomocnost napsaného hodnocení (integer)
5. SCORE – bodové hodnocení na škále 1 až 5, které uživatel hotelu udělil (integer)
6. PERIOD OF STAY – část sezóny, ve které daný uživatel hotel navštívil (polynomial)
7. TRAVELER TYPE – typ cestovatele (polynomial)
8. POOL – bazén součástí hotelu (binomial)
9. GYM – posilovna součástí hotelu (binomial)
10. TENNIS COURT – tenisový kurt součástí hotelu (binomial)
11. SPA – lázně součástí hotelu (binomial)
12. CASINO – kasino součástí hotelu (binomial)
13. FREE INTERNET – možnost internetového připojení zdarma (binomial)
14. HOTEL NAME – název hotelu (polynomial)
15. HOTEL STARS – počet hvězd hotelu (real)
16. NR. ROOMS – počet pokojů v hotelu (integer)
17. USER CONTINENT – kontinent, ze kterého uživatel pochází (polynomial)
18. MEMBER YEARS – počet let, kdy je uživatel členem (integer)
19. REVIEW MONTH – měsíc, kdy uživatel napsal hodnocení (polynomial)
20. REVIEW WEEKDAY – den v týdnu, kdy uživatel napsal hodnocení (polynomial)

Formulace úlohy

V rámci tohoto projektu bude řešena úloha, jejíž cílem bude z příslušného datasetu vydolovat data, která budou schopna pomoci hotelům oslovit cílovou skupinu. Použijeme sloupce, které budou charakterizovat uživatele, působnost, jejich zkušenosti a jejich hodnocení. Ve výsledku dostaneme data, která se budou dolovat z dat referenčních. Na základě těchto vydolovaných dat budou moci hotely odhadnout a příslušně přizpůsobit nabídku svých služeb, jako jsou například binární hodnoty v našich datech a podobně. A tým být více připravení nato, co jaký uživatel, z které země preferuje a potřebuje.