

## АНАЛІЗ АНКЕТ ВІДВІДУВАЧІВ МЕРЕЖІ ТРЕНАЖЕРНИХ ЗАЛІВ

В сучасному світі людство все більше уваги приділяє питанням здоров'я та фізичного розвитку. В даних умовах відкривається багато нових відділень різних мереж фітнес клубів, тренажерних зал, оздоровчих центрів, тощо. З ростом кількості відвідувачів мережі, виникає необхідність ведення статистики, аналізу наявних даних про тренування з метою покращення якості послуг, відкриття нових відділень та створення нових абонементів на основі найпопулярніших послуг.

Актуальність теми обумовлена необхідністю побудови набору різних моделей інтелектуального аналізу даних з використанням сучасних методів аналізу даних, що дозволять підвищити якість послуг, які надаються мережею, простежити залежності та прогрес в тренуваннях відвідувачів, а також надати зручний засіб для аналізу даних відвідувачів мережі.

Сховище даних, що призначене для інтелектуального аналізу даних анкет відвідувачів тренажерної зали має тип «зірка». В таблиці фактів містяться дані про тренування відвідувачів. В якості вимірів маємо 4 таблиці, що зберігають особисті дані відвідувачів з анкет, часові проміжки тренувань, дані про обслуговуючий персонал мережі та дані про послуги, які надаються.

Для того, щоб за наявності анкети відвідувача, що містить такі дані як вік, стать, сімейне становище, факт наявності роботи та машини, кількість дітей, визначити, чи замовить даний відвідувач послуги персонального тренера, необхідно провести побудову правил для класифікації. Оскільки для класифікації необхідно провести аналіз одразу декількох незалежних змінних, таку класифікацію дозволяє виконати метод Naïve Bayes.

Ймовірність того, що відвідувач  $i_j$  відноситься до класу  $c_r$  (клас відвідувачів, що замовляють послуги персонального тренера) можна позначити як:

$$P(y = c_r) \quad (1)$$

Відповідно, ймовірність що відвідувач замовить послуги персонального тренера можна виразити як:

$$P(y = c_r | E) = \frac{P(E | y = c_r) \times P(y = c_r)}{P(E)} \quad (2)$$

Для кожного з правил за формулою Байеса визначається його вірогідність. Якщо врахувати, що всі змінні є незалежними одне від одного, то зможемо виразити дану ймовірність через добуток ймовірностей:

$$P(\text{тренер} = \text{так} | E) = P(\text{вік} = 25 | \text{тренер} = \text{так}) \times P(\text{стать} = \text{жіноча} | \text{тренер} = \text{так}) \times P(\text{кількість дітей} = 0 | \text{тренер} = \text{так}) \times P(\text{робота} = \text{так} | \text{тренер} = \text{так}) \times P(\text{машина} = \text{так} | \text{тренер} = \text{так}) \times P(\text{тренер} = \text{так}) \div P(E)$$

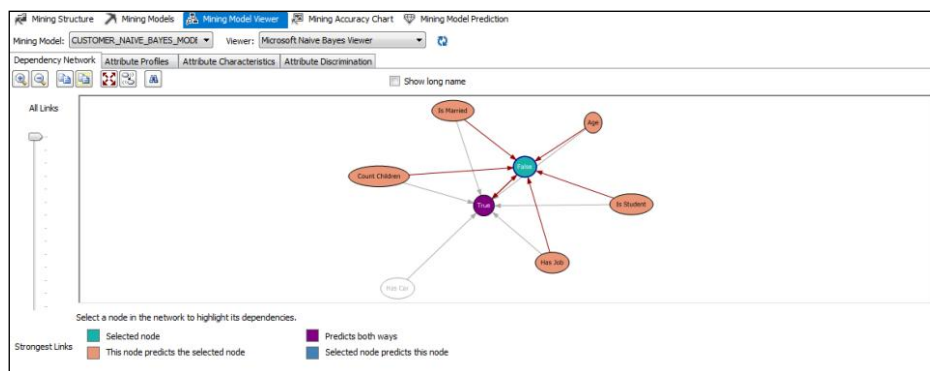


Рис. 1. Діаграма залежності факту, чи замовить відвідувач послуги персонального тренера від різних параметрів

В результаті, було отримано модель інтелектуального аналізу даних, що дозволяє класифікувати нового відвідувача та спрогнозувати, чи замовить даний відвідувач послуги персонального тренера, використовуючи простий та ефективний метод анкетування, враховуючи 6 незалежних параметрів.