

УДК 004.89

Яцишин В.В. канд. техн. наук, доцент, Шуптарський В.В., Цісарук Д.А.

Тернопільський національний технічний університет імені Івана Пулюя

АЛГОРИТМИ МАШИННОГО НАВЧАННЯ ДЛЯ СЕГМЕНТАЦІЇ КОРИСТУВАЧІВ У МАРКЕТИНГОВИХ КОМП'ЮТЕРНИХ СИСТЕМАХ

Yatsyshyn V.V. PhD, Assoc. Prof., Shuptarskyi V.V., Tsisaruk D.A.

MACHINE LEARNING ALGORITHMS FOR USER SEGMENTATION IN MARKETING COMPUTER SYSTEMS

Найбільш поширеними і широко використовуваними методами сегментації, які застосовуються при побудові маркетингових систем є групування за ознаками і методи статистичного аналізу.

В основі методу групування лежить принцип послідовного розбиття множини об'єктів на підмножини (групи) з врахуванням найбільш важливих властивостей чи ознак. Одна з таких властивостей об'єкту виділяється як критерій на основі якого можна сформуванати деяку систему показників (тип продукції, виробник товару, потенційний споживач конкретного виду товару).

Побудова маркетингових комп'ютерних систем вимагає застосування сучасних методів і засобів, які повинні бути орієнтованими на вирішення задач аналізу поведінки користувачів, формування промоакцій, сегментації товарів, послуг і користувачів. Одними із таких трендових технологій для проведення маркетингових досліджень є методи машинного навчання, зокрема, кластеризація і класифікація користувачів і товарів.

У сфері машинного навчання кластеризація передбачає навчання без вчителя, тобто для цього типу алгоритму існує лише один набір вхідних даних без міток. Тому потрібно розв'язати задачу одержання інформації, не знаючи попередньо, яким буде вихід.

Кластеризація використовується в проектах для компаній, які хочуть виявити спільні властивості у своїх клієнтів, щоб застосувати сегментацію клієнтів, створити карти подорожей клієнтів або знайти групи та сформуванати рекомендовані набори товарів чи послуг.

Таким чином, якщо значний відсоток клієнтів мають певні спільні риси (вік, тип сім'ї тощо), компанія може рекомендувати проведення певної кампанії, послуги чи товару. Кластеризація також корисна для одержання загальних уявлень про інформацію якою оперує компанія.

З іншого боку, класифікація належить до алгоритмів навчання з вчителем, тобто наявний контроль за навчанням. Це означає, що вхідні дані мають мітки класів і наперед відомий можливий вихід алгоритму. Розрізняють бінарну класифікацію, яка розв'язує задачі з категорійними відповідями (наприклад, "так" і "ні"), і мультикласифікація, для задач, де потрібно знайти більше двох класів, відповідаючи на більш відкриті відповіді, такі як «чудово», «звичайний» і «недостатній». Класифікація використовується в багатьох галузях, наприклад у біології або в десятковій класифікації Дьюї для книг, при виявленні спаму в електронних листах та ін. Класифікація зазвичай використовується у фінансовому секторі. В епоху онлайн-транзакцій, коли використання готівки помітно зменшилося, необхідно визначити, чи безпечні переміщення за допомогою карток. Суб'єкти можуть класифікувати операції як правильні або шахрайські, використовуючи історичні дані про поведінку клієнтів, щоб дуже точно виявляти шахраїв.