



ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ У ПОБУДОВІ СИСТЕМИ РЕКОМЕНДАЦІЙ НА ОСНОВІ ДАНИХ КОРИСТУВАЧІВ

Голубничий Д.Ю., к.т.н., доцент, кафедра ІС, ХНЕУ ім. С. Кузнеця
Кація І.О., студент, кафедра ІС, ХНЕУ ім. С. Кузнеця

У сучасному цифровому просторі найбільш ефективними системами вважаються ті, які можуть затримати увагу користувача на найбільш довгий час та протягом цього періоду утримання пропонувати контент, який максимально точно відповідатиме уподобанням користувача та змушуватиме робити вибори, вигідні як розробникам системи, так і зацікавленим третім особам.

Для побудови такої системи або застосування життєво необхідно розробити систему рекомендацій, яка аналізуватиме дані, надані користувачем під час використання системи у визначений розробниками спосіб, будь то споживання відеоконтенту у соцмережі, перегляд фільмів на стримінговому сервісі чи зберігання рецептів та опублікування відгуків на них, та збиратиме портрет користувача відповідно до заданих характеристик. Ця задача є нетривіальною і знаходження взаємозв'язків між зовсім різних між собою факторами може бути занадто комплексною та ресурсозатратною для звичайного аналітика, особливо якщо платформа, для якої будується система рекомендацій, має значну аудиторію, яка щоденно користується ресурсом.

Ця проблема зовсім не нова, і способи її вирішення активно вивчаються ще з кінця 1950-х років, адже ця задача є всього лише відгалудженням більшої системи, пов'язаної із необхідністю впорядковувати і аналізувати великі обсяги даних. Саме так з'явилась галузь, нині znana як ML, або машинне навчання (Machine Learning). Із розвитком цієї галузі виникли багато різних алгоритмів та методів, покликаних поліпшити існуючі результати класифікації та установлення зв'язків між різними параметрами у великих наборах даних. Для вибраної нами підзадачі, а саме побудова рекомендаційної системи, найкраще підходять наступні методи та алгоритми, а саме KNN алгоритм та кластеризація [1], а також три головні підходи, які виділяють при побудові таких систем. Вони включають в себе колаборативний, content-based та гідбридний підходи [2, 3].

KNN algorithm, він же алгоритм знаходження найближчого сусіда базується на припущенні, що схожі точки даних матимуть також схожі або однакові характеристики. Він широко використовується у задачах на класифікацію та регресію та ілюструється рисунком 1.

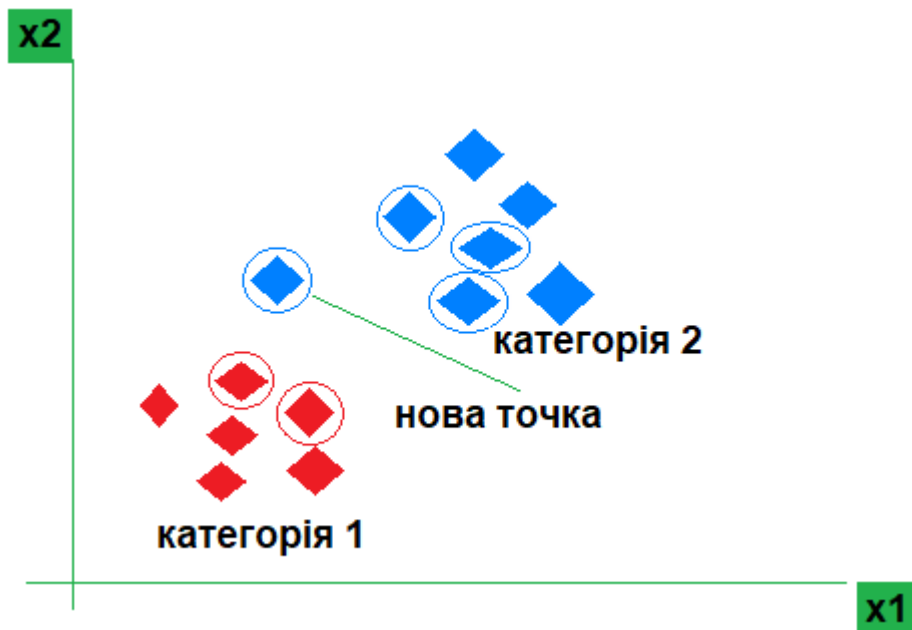


Рисунок 1 – Ілюстрація принципу роботи KNN алгоритму

Суть кластеризації полягає в тому, аби відсортувати точки даних таким чином, аби об'єкти у групах, які зветься кластерами, мали між собою більше спільного, ніж з точками даних з інших кластерів. Це один із основних методів аналізу даних, який використовується не лише у машинному навчанні, а і у інших сферах.

Три ж підходи, про які було згадано раніше, відрізняються лише об'єктом аналізу. Колаборативний підхід для побудови рекомендацій використовує інформацію про користувача, про його смаки та вибори, а потім за допомогою згаданих вище алгоритмів сортує користувачів у групи за визначеними характеристиками і пропонує конкретній групі користувачів один і той же сет даних.

Content-based підхід, натомість, концентрується на вивченні предметів, якими цікавиться користувач, та їхніх характеристик. Після збору інформації визначаються предмети, які мають характеристики, схожі на ті якими цікавився користувач, та пропонуються йому.

Гібридний підхід же збирає у собі два попередні підходи і дає найточніші рекомендації, концентруючись і на користувачах, і на предметах водночас.

Список літератури

1. Furtado, F., & Singh, A. (2020). Movie recommendation system using machine learning. *International journal of research in industrial engineering*, 9(1), 84-98. https://www.riejournal.com/article_106395_c6c0038-f1bf5d4c421bd552d0541d6be.pdf.
2. Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems. *Expert Systems with Applications*, 97(1), 205-227. <http://dx.doi.org/10.1016/j.eswa.2017.12.020>.
3. Грабовський, Є.М. (2017). Специфіка міждисциплінарного підходу стосовно створення системи підтримки електронного навчання у вищій школі нового покоління, *Системи обробки інформації*, 4(150), 206-209.