

А. Ю. ІВАНИЦЬКА, канд. техн. наук, КНУ ім. Тараса Шевченко, Київ,

Д. Є. ІВАНОВ, д-р. техн. наук, доц., ПІММ НАНУ, Слов'янськ,

Л. В. ЗУБИК, канд. пед. наук, доц., КНУ ім. Тараса Шевченко, Київ

МОДЕЛЬ ПРОГНОЗУВАННЯ ПОВЕДІНКИ ПОКУПЦЯ НА ОСНОВІ ТЕХНОЛОГІЙ АНАЛІЗУ ВЕЛИКИХ ДАНИХ

Виконано аналіз наявних методів та моделей формування рекомендацій для потенційного покупця у мережових інформаційних системах з метою розробки ефективних модулів підбору реклами. Обґрунтовано ефективність використання технологій машинного навчання для аналізу вподобань користувачів на основі обробки даних про покупки, здійснені користувачами з подібним профілем. Запропоновано модель формування рекомендацій, основу на технології машинного навчання, перевірено її роботу на тестових наборах великих даних та виконано оцінку адекватності моделі на основі середньоквадратичного відхилення. Іл.: 3. Бібліогр.: 10 назв.

Ключові слова: прогнозування поведінки; технології аналізу великих даних; модулі підбору реклами.

Постановка проблеми. Нині постійно виникає необхідність обробки й зберігання даних великих обсягів та складної структури. Звичайному користувачеві досить важко обробляти таку кількість інформації, тому веб-сайти та великі медіа-ресурси використовують автоматизовані системи у вигляді окремих програмних модулів, так звані рекомендаційні системи (РС) із розвиненим математичним апаратом (машинне навчання, технології Big Data, інтернет речей) обробки даних для швидкого пошуку потрібної інформації. Відомо, що РС – це одні з найпотужніших систем машинного навчання, які інтернет-магазини впроваджують для збільшення продажів, їх постійно вдосконалюють та розвивають. Проблема підбору реклами на основі вподобань користувачів є актуальною й важливою для підприємців малого та середнього бізнесу, де кошти на рекламну діяльність зазвичай обмежені.

Проведено дослідження стану розвитку РС та наявних методів і моделей обробки даних, засобів штучного інтелекту, що формують рекомендації та прогнозують покупки і зазначені їх характеристики. Наявні системи підбору реклами на основі вподобань користувачів в залежності від способу роботи із даними поділяються на чотири види, а саме контент системи, системи з фільтрацією на основі знань, системи на основі колаборативної фільтрації (КФ), гібридні системи [1]. Сучасні методи та моделі формування рекомендацій, мають свої переваги та

недоліки, до останніх слід віднести проблеми точності прогнозу, швидкості обробки даних та недостатньої кількості й якості вхідних даних для стабільного функціонування РС [2].

Тому для вирішення проблеми підбору реклами, що відповідає інтересам користувача, необхідно створити модель, яка на підставі зібраних даних пропонує суміжні та супутні товари клієнту, використовуючи методи колаборативної фільтрації та матричної факторизації. Це дозволяє клієнту вибрати товар на сайті, що використовує такий підхід, та повертатись за подальшими покупками, а власникам бізнесу підвищити дохід.

Аналіз літератури. Сучасні РС виконують пропозиції продажу або прогноз вибору клієнтом товарів, послуг та розваг на основі методів та моделей різної складності обробки вхідних даних. Для побудови рейтингу фільмів використовуються оцінки, алгоритм подібності косинусів (cosine similarity) та отримують середньозважений рейтинг (weighted average of ratings) [3]. Автори [4] використовують метод формування рекомендацій на основі знань, який, поєднує в собі методи обмежень та подібних об'єктів, використовує метод багатокритеріального аналізу, що дає змогу надавати рекомендації навіть при суперечливості вимог користувача.

У дослідженнях [3 – 5] використано категоріальну, мішану і числову кластеризацію даних та виконано модифікацію методу кластеризації розповсюдження схожості, за рахунок додавання властивості нечіткості у моделі підготовки цільових пропозицій на основі вподобань користувачів, та діяльності користувачів у соціальних мережах. Комбіновані методи КФ даних [5] дозволяють обчислити потрібний рейтинг у результаті використання фільтрації на основі подібності покупця та товару. Рейтинг, що необхідно визначити, представляє собою зважену суму оцінок, отриманих в результаті застосування цих методів.

Під час використання колаборативної фільтрації даних на основі моделей, формуються рекомендаційні моделі, що реалізовані за допомогою таких підходів, як байєсівська мережа, кластерний аналіз, моделі Маркова, низькорангова матрична апроксимація [1, 6 – 9]. У дослідженнях [10] представлені РС та огляд моделей побудови рекомендацій на основі різних методів, які можуть бути використані для прогнозування результатів успішності студентів, таких як КФ k -найближчих сусідів, МФ та факторизація з упередженою матрицею, та прогнозування успішності учнів. В результаті аналізу існуючих методів виявлено, що математичний апарат дотичний до конкретної задачі, а показники точності – для певних даних, тому в даній роботі

запропоновано оцінити підходи до обробки даних на основі технологій machine learning і розробити нову модель формування рекомендацій.

Мета дослідження – розробка моделі формування рекомендацій на основі технологій машинного навчання для підвищення точності надання рекомендацій на базі вподобань користувачів

Для досягнення зазначеної мети слід вирішити наступні завдання:

- провести аналіз методів та моделей формування рекомендацій у наявних системах;
- виділити етапи побудови моделі формування рекомендацій на основі технологій machine learning;
- оцінити адекватність моделі на тестових даних.

Матеріали дослідження. Відповідно до практичного використання підходи до побудови РС мають певні переваги та недоліки, що впливають на вибір методу побудови РС. Проте, методи КФ не складні у реалізації, використовуються для різних предметних областей та мають середні показники для інших проблем. Після аналізу цих методів для вирішення задачі дослідження обрано фільтрацію методом КФ за подібністю продукту (item-based collaborative filtering (CF)) та матричну факторизацію.

КФ на основі предметів можна реалізувати в три етапи [2]:

1. Побудова матриці товарів для визначення ступеня подібності між ними.
2. Обчислення нових пропонованих товарів на основі обраного критерію подібності.
3. Пропонування продукції в системі верхнього рівня.

Отже, набір елементів розглядається в залежності від того, які саме оцінив користувач, та обчислюється подібність на конкретний елемент i , а потім вибирається k найбільш подібні товари $\{i_1, i_2, \dots, i_k\}$. Під час цього процесу необхідно обчислити відповідні подібності $\{s_{i1}, s_{i2}, \dots, s_{ik}\}$. Потім, коли знайдені найбільш подібні товари, прогноз розраховується шляхом взяття середньозваженої суми оцінки користувачів щодо подібних елементів.

Наступним важливим кроком у КФ на основі подібності продуктів полягає в обчисленні подібності між двома елементами i та j та у ізолюванні користувачів, які оцінили обидва ці елементи, а далі необхідно використати обраний метод обчислення подібності для визначення подібності $S_{i,j}$.

Після аналізу [9 – 10] зрозуміло, що використовують такі методи обчислення подібності: на основі косинусів, коефіцієнта кореляції Пірсона, коефіцієнта Жаккара та подібність на основі скоригованих

косинусів. Визначення подібності на основі базової косинусної міри при фільтрації даних на базі продуктів має недолік: у шкалі оцінок між різними користувачами не враховуються загальний рейтинг та пошук подібності у парах наборів, що можуть відповідати одному користувачу. Скоригована подібність косинусів компенсує цей недолік, віднімаючи середнє значення u -го користувача для кожного із загальних рейтингів. Тобто відбувається нормалізація даних за елементами. Для реалізації обраного методу подібності між елементами i та j обчислено таким чином:

$$ccsim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \underline{R}_u)(R_{u,j} - \underline{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \underline{R}_u)^2} \times \sqrt{\sum_{u \in U} (R_{u,j} - \underline{R}_u)^2}}, \quad (1)$$

де \underline{R}_u – середнє значення рейтингу u -го користувача.

Метод матричної факторизації полягає в тому, що вподобання користувача можуть визначатися невеликою кількістю прихованих факторів (характеристик). Тоді простір характеристик, у якому можна представити як користувачів, так і продукти, а взаємодії між ними можна отримати за допомогою скалярного добутку між відповідними векторами у цьому просторі. Треба знайти значення матриць U (користувачів) і V (продуктів) таким чином, щоб різниця квадратних помилок між точками продукту та відомим рейтингом (оцінкою) у матриці користувачів наближається до мінімуму, який можна відобразити наступним чином

$$\min(u, v) = \sum_{(i,j) \in k} (R_{u,j} - v_j \times u_i)^2. \quad (2)$$

Для відстеження якості рекомендацій запропонованих методів можна обрати метрики: середньоквадратична похибка (RMSE), середня абсолютна помилка (MAE), нормованим значенням середньої абсолютної похибки (NMAE). За даними навчальної частини здійснюється на основі кореляції оцінок, а за даними тестової частини здійснюється перевірка точності. Необхідно оцінити адекватність моделі на основі середньоквадратичного відхилення за виразом

$$RMSE = \sqrt{\frac{1}{p} \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - \hat{r}_{i,j})^2}, \quad (3)$$

де $r_{i,j}$ – дійсне значення оцінки предмету j користувачем i ;

$\hat{r}_{i,j}$ – прогнозоване значення оцінки предмету j користувачем i ;

p – загальна кількість отриманих оцінок в текстовій частині;

r_{\max} і r_{\min} – максимальне і мінімальне значення оцінки в тестовій матриці.

Етапи побудови моделі формування рекомендацій

Після аналізу методів та моделей фільтрації даних необхідно сформувати етапи побудови моделі формування рекомендацій. У дослідженні було використано дані Amazon.com, що описують низку продуктів індустрії краси. Інформація представлена рядками: Користувач, Продукт, Рейтинг, Часова мітка (UserId, ProductId, Rating, Timestamp).

Для формування рекомендацій пропонуються наступні етапи побудови моделі:

1. Зчитування вхідних даних.
2. Виявлення подібності користувачів на базі їх оцінок та персональних даних.
3. Формування переліку продуктів, що отримали більше оцінок ніж порогове значення.
4. Розрахунок подібності продуктів за алгоритмом кореляції на основі матричної факторизації.
5. Побудова й відображення рекомендацій користувачу.

На першому етапі для обчислення подібності користувачів, за випадковим збігом інтересів на основі оцінок, було прийнято використовувати формулу визначення подібності на основі базової косинусної міри (1).

Порогове значення "80" для фільтрування продуктів отримано у результаті практичного експерименту, адже для продуктів, що отримали менше оцінок, складно побудувати кореляційну матрицю.

Для розрахунку подібності продуктів на основі матричної факторизації застосовується формула (2), а формування матриці для кореляції між всіма відібраними продуктами відбувається наступним чином:

$$C = \frac{X^T X}{N}, \quad (4)$$

де X – матриця, у i -тому рядку котрої є оцінки товарів, які оцінив i -тий користувач;

N – кількість користувачів.

Слід зауважити, що матриця X попередньо нормалізована по стовпцях за виразом:

$$x_j^{new} = \frac{x_j - \text{mean}(x_j)}{\text{std}(x_j)}, \quad (5)$$

де x_j – j -та колонка матриці X ;

$mean(x_j)$ – середнє значення j -тої колонки матриці X ;

$std(x_j)$ – стандартне відхилення j -тої колонки матриці X .

Після цього для набору продуктів, що містяться у кореляційній таблиці використано метод Пірсона. Суть методу полягає в тому, що рейтинг присвоюється товару на основі оцінок інших подібних товарів, які вже включені покупцем у замовлення. Комбінація методів КФ на основі продукту та МФ даних проста в реалізації, тому впровадження моделі у підсистему підбору реклами займає менше часових та фінансових видатків.

В результаті відбувається формування рекомендації та демонстрація зображення супутніх або подібних товарів кінцевому користувачу.

Оцінка адекватності запропонованої моделі на тестових даних.

У файлі `ratings_Beauty.csv`, з яким працює модель формування рекомендацій, міститься близько 2 023 000 рядків вхідних даних (36 унікальних клієнтів, більше 3700 продуктів на сайті). Після виявлення та видалення порожніх значень, можна побудувати гістограму з кількістю продуктів згрупованих за рейтингом, подану на рис. 1.

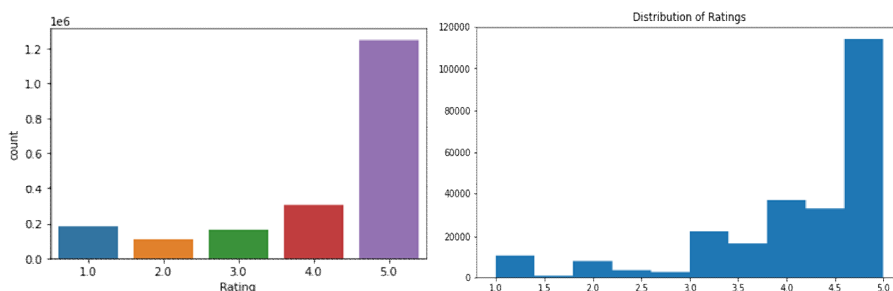


Рис. 1. Гістограма розподілу середніх значень оцінок для продуктів

Продукти отримують оцінки, потім для всіх користувачів розраховується рейтинг цього продукту. Потім розраховуються статистичні параметри (середнє значення для відповідних продуктів, `std`, `min` тощо).

Для застосування КФ необхідно обрати лише ті продукти, які оцінені не менше ніж 80 користувачами. Далі необхідно побудувати матрицю (використовуючи функцію `pd.pivot_table`), виконати нормалізацію та кореляцію даних.

Для перевірки роботи моделі формування рекомендацій можна виділити два продукти – сонцезахисний крем та тонер.

Якщо відсортувати результуючі дані, то можна отримати таблицю, що демонструє перші п'ять продуктів, що мають рейтинг більше 80

користувачів. За цією таблицею можна з'ясувати, які товари корелюють із першим продуктом, що записав у систему користувач. Тобто, із B009FKNGGQ краще всього корелюють продукти з наступними ID_PRODUCT: B009FKNGRA, B002MO3CG6, B008U2Y9BQ, B00AO379NE, B00604MSSC.

Модель формування рекомендацій дає можливість обрати серед продуктів з подібними оцінками, що дозволяє користувачу швидше визначитись із покупкою.

Для оцінки точності використано holdout метод, де стек даних поділено на два набори – навчальний (2/3 даних близько 130 0000 рядків) і набір тестів (1/3 даних – 600 000 рядків таблиці). Показник RMSE отримано для кількох методів підбору реклами, а саме CF, MF та CF+MF. Після класифікації сформовано розбиття на дані з самими високими оцінками користувачів і з падінням мінімального значення помилки (Err) і мінімальною кількістю епох. Комбінація методів КФ та МФ на базі вподобань користувачів мають менші значення середньоквадратичного відхилення ($RMSE = 0,0864$) між прогнозованою оцінкою та дійсною оцінкою продукту та швидкості навчання серед обраних для аналізу. Це свідчить про точність та швидкість обробки даних на основі тестового набору.

Реалізовану модель формування рекомендацій на основі технології машинного навчання – комбінації двох методів КФ та МФ на базі вподобань користувачів було використано, як можливість ставити оцінки товарів, при тестуванні.

Програмна реалізація моделі формування рекомендацій на мові Python та її впровадження. Власник сайту може підключити рекомендаційну модель (підсистему підбору реклами на основі вподобань користувачів) до наявної платформи на основі запропонованої архітектури, що представлена на рис. 2.



Рис. 2. Запропонована архітектура підсистеми підбору реклами

Як видно з рис. 2, спочатку потрібно отримати вхідні дані для попередньої обробки, це можуть бути дані із різних джерел JSON, із системи управління базами даних (СУБД) – (MS SQL, My SQL, PostgreSQL), таблиці Excel та файли формату *.csv. Потім необхідно організувати завантаження даних про клієнта/клієнтів у систему або до СУБД системи.

Далі в системі відбувається формування рекомендацій за обраними алгоритмами, а саме метод колаборативної фільтрації за продуктами з алгоритмом подібності косинуса та матрична факторизація з алгоритмом кореляції Пірсона. Після цього можна виконувати збереження даних у файл або до актуальної СУБД. Завантажувати дані можна із певною періодичністю 12 годин (зазвичай обирають період 24 години), тоді дані замовника підсистеми копіюються у хмару.

Перед тим, як відбудеться впровадження моделі необхідно реалізувати сервер для видачі та відображення рекомендацій кінцевому користувачу.

Основними об'єктами підсистеми підбору реклами є користувачі сайту, товари та їх оцінки. Далі необхідно визначити функціональність та поведінку системи з вхідними та вихідними даними за допомогою узагальненої схеми, яка представлена на рис. 3 та дозволяє описати підсистему, що розробляється.



Рис. 3. Інтерпретація даних, з якими взаємодіє підсистема

Як видно з рис. 3, в якості вхідних даних системи можна використовувати не тільки дані про конкретного користувача, а також загальні відомості про поведінку користувачів на сайті, кількісні та якісні характеристики продуктів та контекст. Крім того, у разі використання рекомендацій для інтернет-магазинів більшість доступних даних – дані,

які відповідають можливості відвідування веб-сайту у прихованій формі (недосконалі дані, приховані персональні дані користувача тощо).

Програмний модуль був розроблений на мові програмування Python (у редакторі Jupyter Notebook) з використанням сервісу Anaconda та мови розробки інтерфейсу користувача XAML.

Структура проекту на Python має послідовний характер, а переглянути результати можна в файлі Colab_filt.html (браузер довільний). При оцінці адекватності моделі були розглянуті результати програмного рішення для кожного пункту алгоритму моделі формування рекомендацій.

Висновки. За результатами виконаного дослідження розроблена модель формування рекомендацій на основі вподобань користувачів, що використовує технології машинного навчання. При цьому було вирішено наступні завдання:

1) Виконано аналіз наявних методів та моделей обробки даних, засобів штучного інтелекту, що використовуються для побудови систем підбору реклами і зазначені їх характеристики, що дозволило обрати методи колаборативної фільтрації та матричної факторизації для розробки моделі формування рекомендацій.

2) Запропоновано структуру та розроблено модель формування рекомендацій на основі комбінації відомих методів обробки даних. Практична реалізація моделі здійснена мовою програмування Python.

3) Перевірено роботу моделі на тестових даних значного обсягу та виконано оцінку адекватності моделі на основі середньоквадратичного відхилення, яка підтвердила точність запропонованої моделі не менш ніж 91,39%, що вказує на достовірність отриманих результатів.

Список літератури:

1. *Thanh-Nhan H.* Methods for building course recommendation systems / *H. Thanh-Nhan, H. Nguyen, N. Thai-Nghe* // Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016. – P. 163-168.
2. *Іваницька А.Ю.* Розробка моделі формування рекомендацій при побудові підсистеми підбору контекстної реклами / *А.Ю. Іваницька, Л.В. Зубик, Д.Є. Іванов* // Проблеми інформатики та моделювання (ПІМ-2021). Тезиси двадцять першої міжнародної науково-технічної конференції. – Харків: НТУ "ХПІ", 2021. – С. 26.
3. *Koren Y.* Factorization meets the neighborhood: a multifaceted collaborative filtering model / *Y. Koren* // International conference on Knowledge discovery and data mining: Proceedings of the 14th ACM SIGKDD, 2008. – ACM. – 2008. – P. 426-434.
4. *Item-Based Collaborative Filtering in Python.* The practice of making the item-based collaborative filtering in python. – Access mode: URL: <https://towardsdatascience.com/item-based-collaborative-filtering-in-python-91f747200fab> (Data: 16.09.2021).
5. *Lobur M.* The method of sequential clustering for predicting recommendations / *M. Lobur, M. Shvarts, Y. Stekh* // CAD in Machinery Design-Implementation and Education Problems:

Proc. of the XXV Polish-Ukrainian Conference, Bielsko-Biała, 20–21 October 2017. – 2017. – P. 19-20.

6. *Hernando A.* A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model / *A. Hernando, J. Bobadilla, F. Ortega* // *Knowledge-Based Systems*. – 2016. – Vol. 97. – P. 188-202.

7. *Onokoy L.* Modern approaches to building recommender systems for online stores / *L. Onokoy, J. Lavendels* // *Applied Computer Systems*. – 2019. – № 24. – P. 18-24.

8. *Steck H.* Training and testing of recommender systems on data missing not at random / *H. Steck* // *International conference on Knowledge discovery and data mining. Proceedings of the 16th ACM SIGKDD: 2010*. – ACM. – 2010. – P. 713–722.

9. *Koren Ye.* Matrix factorization techniques for recommender systems / *Ye. Koren, R. Bell, C. Volinsky* // *Computer*. – Vol. 42. – № 8. – 2009. – P. 30-37.

10. *Thai-Nghe N.* Recommender system for predicting student performance / *N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme* // *In Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010): Elsevier's Procedia CS, 2010*. – Vol. 1. – 2010. – P. 2811–2819.

References:

1. Thanh-Nhan, H., Nguyen, H., Thai-Nghe, N. (2016), "Methods for building course recommendation systems", *Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016*, pp. 163–168.

2. Ivanytska, A.Yu., Zubyk, L.V., Ivanov, D.E. (2021), "Development of the model for the formation of recommendations for the construction of a subsystem for the selection of contextual advertising", *Problems Informatics and Modeling (PIM-2021). Thesis of the twenty-first international scientific and technical conference*. - Kharkiv: NTU "KhPI", September 9-14, 2021, P.26.

3. Koren, Y. (2008), "Factorization meets the neighborhood: a multifaceted collaborative filtering model", *International conference on Knowledge discovery and data mining: Proceedings of the 14th ACM SIGKDD, 2008*, ACM, 2008, pp. 426–434.

4. Item-Based Collaborative Filtering in Python. The practice of making the item-based collaborative filtering in python: Access mode: URL: <https://towardsdatascience.com/item-based-collaborative-filtering-in-python-91f747200fab> (Data: 16.09.2021).

5. Lobur, M., Shvarts, M., Stekh, Y. (2017), "The method of sequential clustering for predicting recommendations", *CAD in Machinery Design-Implementation and Education Problems: Proc. of the XXV Polish-Ukrainian Conference: October 20–21, 2017, Bielsko-Biała, 2017*, pp. 19–20.

6. *Hernando, A., Bobadilla, J., Ortega, F.* (2017), "A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model", *Knowledge-Based Systems*, Vol. 97, pp. 188–202.

7. *Onokoy, L., Lavendels J.* (2019), "Modern approaches to building recommender systems for online stores", *Applied Computer Systems*, No. 24, pp. 18–24.

8. *Steck, H.* (2010), "Training and testing of recommender systems on data missing not at random", *International conference on Knowledge discovery and data mining: Proceedings of the 16th ACM SIGKDD: 2010*, ACM, 2010, pp. 713–722.

9. *Koren, Ye., Bell R., Volinsky, C.* (2009), "Matrix factorization techniques for recommender systems", *IEEE COMPUTER Published by the IEEE Computer Society*, Vol. 42, No. 8, 2009, pp. 30–37.

10. *Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A. and Schmidt-Thieme, L.* (2010), "Recommender system for predicting student performance", *In Proceedings of the 1st*

Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), Elsevier's Procedia CS, Vol. 1, 2010, pp. 2811–2819.

Статтю представив д-р техн. наук, проф. КНУ Сайко В.Г.

Надійшла (received) 10.12.2021

Ivanytska Anastasiya, PhD Tech
Taras Shevchenko National University of Kyiv
Bohdan Hawrylyshyn str. 24, Kyiv, 04116
Tel.: (095) 333-51-01, e-mail: a.titova.wk@gmail.com
ORCID ID: 0000-0002-4803-2090

Ivanov Dmitry, Dr.Sci.Tech, Ass. Professor,
Institute of Applied Mathematics and Mechanics,
str. Gen.Batyuka, 19, Slavyansk, 84100
Tel: (063) 559-51-90, e-mail: dmitry.ivanov.iamm@gmail.com
ORCID ID: 0000-0001-9956-6589

Zubyk Liudmyla, PhD Ped, Ass. Professor,
Taras Shevchenko National University of Kyiv
Bohdan Hawrylyshyn str. 24, Kyiv, 04116
Tel.: (066) 60-42-556, e-mail: l.v.zubyk@univ.kiev.ua
ORCID ID: 0000-0002-2087-5379

УДК 004.852

Модель прогнозування поведінки покупця на основі технологій аналізу великих даних / Іваницька А.Ю., Іванов Д.Є., Зубик Л.В. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2021. – № 2 (6). – С. 89 – 100.

Виконано аналіз наявних методів та моделей формування рекомендацій для потенційного покупця у мережевих інформаційних системах з метою розробки ефективних модулів підбору реклами. Обґрунтовано ефективність використання технологій машинного навчання для аналізу переваг користувачів на основі обробки даних про покупки, здійснені користувачами з подібним профілем. Запропоновано модель формування рекомендацій, основу на технології машинного навчання, перевірено її роботу на тестових наборах великих даних та виконано оцінку адекватності моделі на основі середньоквадратичного відхилення. Іл.: 3. Бібліогр.: 10 назв.

Ключові слова: прогнозування поведінки; технології аналізу великих даних; модулі підбору реклами.

УДК 004.852

Модель прогнозирования поведения покупателя на базе технологий анализа больших данных / Иваницкая А.Ю., Иванов Д.Е., Зубик Л.В. // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2021. – № 2 (6). – С. 89 – 100.

Выполнен анализ существующих методов и моделей формирования рекомендаций для потенциального покупателя в сетевых информационных системах с целью разработки эффективных модулей подбора рекламы. Обоснована эффективность использования технологий машинного обучения для анализа предпочтений пользователей на базе обработки данных о покупках, выполненных пользователями с подобным профилем. Предложена модель формирования рекомендаций, основанная на технологии машинного обучения, проверена ее работа на тестовых наборах больших данных и выполнена оценка адекватности модели на основе среднеквадратического отклонения. Ил.: 3. Библиогр.: 10 назв.

Ключевые слова: прогнозирование поведения; технологии анализа больших данных; модули подбора рекламы.

УДК 004.852

Buyer behavior forecasting model based on big data analysis technologies / Ivanytska A.Yu., Ivanov D.E., Zubyk L.V. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2021. – № 2 (6). – P. 89 – 100.

The analysis of the available methods and models of formation of recommendations for the potential buyer in network information systems for the purpose of development of effective modules of selection of advertising was executed. The efficiency of using machine learning technologies for the analysis of user preferences based on the processing of data on purchases made by users with a similar profile was substantiated. A model of recommendation formation based on machine learning technology was proposed, its work on test data sets was tested and the adequacy of the RMSE model was assessed. Figs.: 3. Refs.: 10 titles.

Keywords: behavior prediction; big data analysis technologies; advertising selection modules.