

РЕФЕРАТ

Дисертацію виконано на 80 аркушах, вона містить 3 додатки та перелік посилань на використані джерела з 52 найменувань. У роботі наведено 16 рисунків та 4 таблиці.

Актуальність теми. Прибуток інтернет-магазинів в світі досягнув 1.18 трильйона доларів США у 2016 році [1]. З кожним роком кількість-інтернет магазинів збільшується, а конкуренція між ними посилюється. Провідні інтернет-магазини використовують алгоритми машинного навчання, щоб прогнозувати поведінку користувачів та, таким чином, збільшити кількість користувачів та покупок. Для створення таких алгоритмів інтернет-магазини публікують анонімізовані дані про історію поведінки користувачів на їх сайтах на платформах задач машинного навчання [2] та залучають незалежних фахівців для роботи з цими даними. Більшість алгоритмів спрямовані на те, щоб запропонувати користувачеві товари, які він ще не бачив, але які можуть його зацікавити, таким чином спонукаючи користувача купляти більше за один сеанс. Проте, такі алгоритми не розраховані на те, щоб полегшити користувачеві повторне використання сайту інтернет-магазину чи сприяти поверненню користувача на сайт для здійснення подальших покупок. Саме тому, є актуальним розроблення системи побудованої на основі алгоритмів машинного навчання, метою роботи якої є прогнозування поведінки користувача для полегшення його роботи з сайтом інтернет-магазину, що фокусується на товарах, які користувач вже купляв та може обрати повторно. Таких підхід забезпечить більшу кількість повернень користувачів до даного магазину, та, збільшення прибутків магазину в наслідок цього, отож в результатах роботи системи найбільше зацікавлені власники інтернет-магазинів.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно з планом науково-дослідних робіт кафедри прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Мета і задачі дослідження. Метою дисертаційної роботи є покращення

ергономічних властивості сервісу електронних покупок шляхом зменшення кількості операцій та витрат часу, необхідних для вибору та пошуку товарів на веб-сайті.

Для досягнення вказаної мети було розв'язано такі задачі:

- систематизувати існуючі методи афінитивного аналізу;
- систематизувати існуючі методи класифікації;
- провести порівняльний аналіз існуючих систем прогнозування покупок користувачів на основі статистичних даних;
- розробити компоненти системи прогнозування покупок користувачів на основі статистичних даних;
- розробити інтерфейси взаємодії компонентів системи прогнозування покупок користувачів на основі статистичних даних;
- отримати дані;
- виконати попередню обробку даних;
- розробити модель алгоритму прогнозування;
- розробити програмне забезпечення системи прогнозування покупок користувачів на основі статистичних даних;
- провести експериментальні дослідження з використанням різних методів класифікації;
- сформулювати рекомендації щодо подальшого використання та вдосконалення системи.

Об'єктом дослідження є задача аналізу ринкової корзини, афінитивний аналіз, асоціативні правила, алгоритми пошуку асоціативних правил: Apriori, Ectat, FP-growth, AprioriDP, AIS, OPUS search, data mining та поняття підтримки(support), достовірності(confidence), підйомної сили(lift), переконання(conviction), алгоритми машинного навчання, алгоритми класифікації, а саме, метод k-найближчих сусідів, метод випадковий ліс, наївний баєсів класифікатор, метод потенційних функцій, побудова штучних нейронних мереж, програмне забезпечення для обробки та зберігання великих обсягів даних(Big data), системи керування базами даних категорії NoSQL та SQL, а саме Cassandra та PostgreSQL, алгоритм опрацювання даних великих обсягів — MapReduce.

Предметом дослідження є алгоритм прогнозування повторних покупок

користувачів із застосуванням алгоритмів машинного навчання системи прогнозування покупок користувачів на основі статистичних даних.

Методи дослідження. Для розв'язання поставленої задачі використовувалися такі методи: методи афінитивного аналізу (для пошуку асоціативних правил); методи машинного навчання (для розроблення алгоритму прогнозування); методи теорії алгоритмів та програмування (для програмної реалізації розроблених компонентів), методи оптимізації (для вибору налаштувань системи для досягнення найкращих результатів прогнозування).

Наукова новизна одержаних результатів складається з наступних положень:

- уперше поставлено задачу прогнозування покупок користувачів, яка, на відміну від існуючих, фокусується на повторному виборі користувачем товару;
- удосконалено методи прогнозування вибору користувача, які, на відміну від існуючих, враховують частоту вибору користувачем товару, що дає змогу робити прогнозування для непопулярних товарів;
- запропоновано евристичні стратегії обробки даних для обробки наявних статистичних даних;
- удосконалено підходи до прогнозування для деяких даних, які відрізняються від існуючих тим, що працюють за розкладом, а не в режимі реального часу, що дає змогу зменшити навантаження на апаратне забезпечення системи.

Практичне значення одержаних результатів. Запропоновано систему, яку може бути використано для покращення ергономічних властивостей сайту інтернет-магазину шляхом інтеграції системи до сайту інтернет-магазину. Система робить прогнозування вибору користувача на наступному етапі здійснення покупок та повертає список товарів, які можуть зацікавити користувача для відображення їх на сайті інтернет-магазину. Систему можна налаштувати у відповідності з потребами інтернет-магазину. Розроблено алгоритм прогнозування для наявних статистичних даних.

Апробація результатів дисертації. Алгоритм попередньої обробки даних для прогнозування представлено на 10-й науково-технічній конференції магістрантів і аспірантів «Прикладна математика і комп'ютинг» (2018 р.)

Публікації. Частина матеріалів дисертації викладено в тезах 10-ї наукової

конференції магістрантів і аспірантів «Прикладна математика і комп'ютинг»
(2018 р.)

Ключові слова: машинне навчання, класифікація, афінитивний аналіз,
обробка даних, прогнозування.