

Краснюк М.Т.
доцент
Оніщенко Д. С.
студентка
Донєв Д. Р.
студент

*Київський національний економічний
університет імені Вадима Гетьмана
djdondan@gmail.com
diana29.oni2@gmail.com*

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В РЕКЛАМІ

Проблематика: У сучасному цифровому середовищі онлайн-реклама є неодмінною складовою ефективною маркетинговою стратегії для бізнесу будь-якого масштабу. У той час як рекламодавці інвестують значні кошти в цифрову рекламу, важливо мати інструменти, що допомагають передбачити ймовірність успіху рекламної кампанії. Конкуренція між рекламодавцями є вкрай жорсткою. Кожна компанія намагається залучити увагу інтернет-користувачів і перетворити цю увагу на кліки, продажі та конверсії.

Вступ: Бізнес інвестує мільйони в рекламу. Провідним сегментом рекламної індустрії сьогодні є реклама в Інтернеті. Онлайн-ігри, мобільні додатки, соціальні мережі, платформи електронної комерції, пошукові портали, онлайн-відео, банери та інші цифрові медіа - це лише кілька прикладів багатьох цифрових медіа, які використовуються в онлайн-рекламі для охоплення потенційних клієнтів. Інтелектуальний аналіз даних докорінно змінив підхід бізнесу до маркетингу, надаючи розуміння поведінки клієнтів, визначаючи нові можливості та удосконалюючи маркетингові стратегії. Основними компонентами інтелектуального аналізу даних, які підтримують інтернет-маркетинг, є збір, об'єднання, аналіз і відповідне відображення даних бізнес-аналітики. Після інтелектуального аналізу даних та аналітики бізнес-аналітика допомагає визначити реальну модель поведінки споживачів. Таким чином, акцент на персоналізації та кастомізації контенту в багатьох цифрових каналах є одним із головних наслідків впливу Big Data, в будь-якій формі, на цифровий маркетинг.

Теза 1: Перш за все, важливо розуміти поведінку споживачів, уподобань і моделей купівлі за допомогою аналізу даних, що дозволяє компаніям адаптувати маркетингові кампанії, розробляти нові продукти, оптимізувати стратегії продажів і підвищити рівень задоволеності клієнтів. Серед ключових підходів, які розглядаються в процесі інтелектуального аналізу даних є: ідентифікація проблеми, підготовка даних, розробка моделі, використання моделі та моніторинг моделі. Використовують низку методів для класифікації та регресійних моделей, включаючи дерева рішень, штучні нейронні мережі, еволюційні алгоритми та правила асоціацій. Для того, щоб знайти приховані закономірності та ідеї з різних джерел даних, таких як документи, веб-матеріали та мультимедійні файли, досліджуються складні підходи до інтелектуального аналізу даних, такі як аналіз тексту, веб-аналіз та мультимедійний аналіз.

Крім того, розглядаються такі тактики CRM (customer relationship management), як готовність покупця, вигоди, поведінка, привід, психографія життя, демографія, життєвий цикл, географія, лояльність, продукт, прибутковість, взаємодія та задоволеність для сегментації. Акцент зроблено на ідеї аналізу споживчої цінності та важливості розуміння сприйнятої цінності, загальних переваг, витрат і суб'єктивних поглядів споживачів.

Теза 2: Останніми роками спостерігається великий інтерес до впровадження методів персоналізованого маркетингу та інструментів інтелектуального аналізу даних у цифрову компанію як стратегію змішаних методів, що включала як якісні, так і кількісні методи

дослідження. З точки зору якісних методів, наприклад, у одному з досліджень проводили глибинні інтерв'ю. Кількісні методи використовувалися для збору даних про поведінку споживачів з різних онлайн-джерел, включаючи електронну комерцію Shopee, Instagram, TikTok, WhatsApp та Instagram. Кількісний аналіз цих даних проводиться за допомогою таких методів, як аналіз асоціацій, який виявив взаємозв'язки між продуктами та моделями покупок, системи рекомендацій, які відповідають індивідуальним уподобанням клієнтів, та кластерний аналіз, який виявив групи клієнтів зі схожими уподобаннями та поведінкою. Використовуючи ці методи, дослідники змогли отримати всебічне розуміння того, як споживачі ставляться до індивідуальної маркетингової тактики і як вони поведуться в інтернет-магазинах.

Теза 3: Прогнозування CTR (коефіцієнт кліків) - це розрахунок ймовірності того, що користувач натисне на рекламне оголошення. Цей показник дуже важливий в системах онлайн-реклами. Згідно з традиційною парадигмою оплати за клік, рекламодавці отримують гроші лише тоді, коли глядачі натискають на їхні оголошення, а отже, вищий CTR може збільшити прибутки рекламодавців. Використовуючи сучасні підходи машинного навчання на основі великих обсягів даних, можна отримати більш точний прогноз CTR, що дозволить підвищити ефективність рекламних кампаній і знизити витрати на рекламу. Підходи, які дозволяють видавцям вибирати, яке рекламне зображення показувати конкретному користувачеві на веб-сторінці, щоб максимізувати CTR, користуються високим попитом через необхідність максимізації доходу для видавців. Тому в більшості систем онлайн-реклами основним обов'язком розміщення реклами є передбачення CTR. На прикладі однієї з найбільших рекламних платформ в Інтернеті, Google AdWords, можна продемонструвати, наскільки важливим є показник CTR для систем управління рекламними кампаніями. Моделі дерева рішень, моделі глибокого навчання, моделі машин факторизації та багатовимірні статистичні моделі - це чотири сучасні моделі, доступні для прогнозування CTR. Два найпоширеніші в літературі багатовимірні статистичні методи прогнозування CTR реклами - це логістична регресія та поліноми другого ступеня. Одним з методів, який найчастіше використовується для прогнозування CTR за багатьма факторами, є логістична регресія (LR). Він може повертати число від 0 до 1, що вказує на ймовірність кліка, а також лінійну суміш декількох атрибутів, яка відображає кореляцію між якостями та етикеткою. Однак LR не може врахувати ефекти взаємодії або нелінійні кореляції. У літературі LR часто використовують як еталон для оцінки ефективності рекомендованих моделей прогнозування CTR. Дослідники відзначають, що факторизовані параметри та машини факторизації (МФ) можуть допомогти в оцінці моделі на розріджених даних для прогнозування CTR. МФ та їхні розширення часто використовуються для прогнозування кліків у дослідницькій літературі з реклами. На конкурсі KDDCup 2012 вони продемонстрували, що МФ можуть точно оцінити взаємодію характеристик при прогнозуванні CTR реклами. У 2015 році було представлено стратегію онлайн-навчання для прогнозування CTR, яка використовує алгоритм Follow-TheRegularized-Leader (FTRL-проксимальний) у МФ. Хоча моделі низького порядку прості у створенні, вони не дуже добре справляються з прогнозуванням, особливо в складних рекламних системах. AUC показує, що моделі прогнозування CTR на основі LSTM та RNN перевершують лінійні (наприклад, LR та наївні байєсівські) та нелінійні (наприклад, нейронні мережі та випадкові ліси) моделі. Однак, оскільки LSTM навчаються послідовно, вони вимагають більше часу і пам'яті на етапі навчання. Деревоподібні моделі використовуються в широкому спектрі обчислювальних рекламних додатків. Gradient Boosting Decision Tree (GBDT) і XGBoost засновані на машині градієнтного бустінгу (GBM), яка продемонструвала великий успіх у прогнозуванні CTR. Для масової реклами у Facebook автори пропонують конкатенацію GBDT та LR для прогнозування CTR, де вихід GBDT слугує входом LR. В результаті GBDT значно підвищує продуктивність LR за рахунок автономного вибору ознак.

Теза 4: Ще один із підходів до вивчення задоволеності та незадоволеності користувачів, використовуючи аналіз текстів. Для цього можна метод латентного семантичного аналізу (LSA), який дозволяє витягти приховані семантичні зв'язки слів та фраз з корпусу документів.

У процесі обробки даних вилучаються зайві слова, об'єднуються пов'язані фрази та визначаються змінні, що вказують на задоволеність і незадоволеність користувачів. Результати дослідження підтверджують важливість тих чи інших елементів, а також допомагають визначити пропозицій чи послуг, які впливають на задоволеність чи незадоволеність користувачів. Далі ці знання використовують для створення фраз на креативах чи квізах, що збільшує зацікавленість користувачів, що в свою чергу збільшує дохід компанії. Оскільки текстова аналітика є відкритою структурою, вона може надавати більш точні звіти про клієнтський досвід, ніж звичайні методи опитування, такі як анкетування або інтерв'ю.

Аналіз великих даних про поведінку та задоволеність клієнтів значно полегшується завдяки застосуванню LSA та текстової аналітики.

Теза 5: Розглянемо приклад коли інтелектуальний аналіз даних в рекламі покращив роботу компанії порівняно з конкурентами. Ретельний аналіз особливої та креативної маркетингової тактики Tesla відрізняє її від традиційних методів, що використовуються в автомобільній індустрії для продажу та реклами. Цифровий маркетинг є основною стратегією, оскільки він не вимагає традиційних витрат на рекламу. Соціальні мережі, такі як Facebook, Instagram, Twitter і YouTube, є основними майданчиками для просування продуктів і бренду Tesla. Тут також проводиться вся реклама та продажі компанії. До 70% продажів компанії, за даними дослідників, здійснюються онлайн.

Працівники компанії проаналізували багато даних про покупців і їхню реакцію на маркетингові ініціативи компанії, використовуючи методи інтелектуального аналізу даних. Зокрема, споживачі були згруповані на основі їхніх інтересів та поведінкових тенденцій за допомогою алгоритмів кластеризації. Це дало змогу визначити, які елементи маркетингових зусиль приваблюють цільову аудиторію та сприяють збільшенню продажів і впізнаваності бренду.

Висновок: Підкресляю стратегічну цінність інтелектуального аналізу даних у рекламі, особливо в тому, що він допомагає компаніям залишатися конкурентоспроможними на швидко мінливому цифровому ринку, допомагаючи їм передбачати майбутні тенденції, приймати обґрунтовані рішення і так далі. Інтелектуального аналіз даних забезпечує адаптованість маркетингових зусил, поліпшення прогнозів продажів і зміцнення зв'язків з клієнтами, що в результаті сприяє процвітанню і довголіттю організацій.

Список використаних джерел

1. Yusnidar Y., Dudi Yudhikusuma, Sari F. Personalized Marketing Strategy in Digital Business Using Data Mining Approach. *International Journal Software Engineering and Computer Science (IJSECS)*. 2023. Vol. 3, no. 2. P. 137–143. URL: <https://doi.org/10.35870/ijsecs.v3i2.1515> (date of access: 08.04.2024)
2. Alghanam O. A., Al-Khatib S. N., Hiari M. O. Data Mining Model for Predicting Customer Purchase Behavior in E-Commerce Context. *International Journal of Advanced Computer Science and Applications*. 2022. Vol. 13, no. 2. URL: <https://doi.org/10.14569/ijacsa.2022.0130249> (date of access: 09.04.2024).
3. Digital Marketing Platforms and Customer Satisfaction: Identifying eWOM Using Big Data and Text Mining / F. Kitsios et al. *Applied Sciences*. 2021. Vol. 11, no. 17. P. 8032. URL: <https://doi.org/10.3390/app11178032> (date of access: 08.04.2024).

4. Data mining: Application of digital marketing in education / C. Molina Huerta et al. *Advances in Mobile Learning Educational Research*. 2023. Vol. 3, no. 1. P. 621–629. URL: <https://doi.org/10.25082/amler.2023.01.011> (date of access: 08.04.2024).
5. Unveiling the Success Behind Tesla's Digital Marketing Strategy / P. Pathak et al. *Emerging Technologies in Data Mining and Information Security*. Singapore, 2022. P. 251–260. URL: https://doi.org/10.1007/978-981-19-4193-1_24 (date of access: 08.04.2024).
6. Mahmut Tekin Mehmet Etlioğlu Özdal Koyuncuoğlu and Ertuğrul Tekin. Data Mining in Digital Marketing. *Proceedings of the International Symposium for Production Research, Necmettin Erbakan University, Konya, Turkey*. 2018. Vol. 24, no. 1. 44-61. URL: https://doi.org/10.1007/978-3-319-92267-6_4 (date of access: 08.04.2024).
7. Yang Y., Zhai P. Click-through rate prediction in online advertising: A literature review. *Information Processing & Management*. 2022. Vol. 59, no. 2. P. 102853. URL: <https://doi.org/10.1016/j.ipm.2021.102853> (date of access: 08.04.2024).
8. Intelligence and Neuroscience C. Retracted: Advertising Click-Through Rate Prediction Based on CNN-LSTM Neural Network. *Computational Intelligence and Neuroscience*. 2023. Vol. 2023. P. 1. URL: <https://doi.org/10.1155/2023/9849374> (date of access: 09.04.2024).
9. He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R., Bowers, S., & Candela, J. Q. (2014). Practical lessons from predicting clicks on ads at Facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (ADKDD '14)*. Association for Computing Machinery, New York, NY, USA, 1–9. URL: <https://doi.org/10.1145/2648584.2648589>. (Дата звернення 27.03.2024)
10. An H., Ren J. XGBDeepFM for CTR Predictions in Mobile Advertising Benefits from Ad Context. *Mathematical Problems in Engineering*. 2020. Vol. 2020. P. 1–7. URL: <https://doi.org/10.1155/2020/1747315> (date of access: 08.04.2024).
11. Chapelle O., Manavoglu E., Rosales R. Simple and Scalable Response Prediction for Display Advertising. *ACM Transactions on Intelligent Systems and Technology*. 2015. Vol. 5, no. 4. P. 1–34. URL: <https://doi.org/10.1145/2532128> (date of access: 08.04.2024)

Науковий керівник: Краснюк М. Т., к.е.н., доцент.

Куліда І.В., Дівеев Я.І.

студенти

Київський національний економічний
університет імені Вадима Гетьмана

ydivееv@ukr.net

ВИЯВЛЕННЯ ТА ФІЛЬТРАЦІЯ ЗЛОВМИСНОГО МЕРЕЖЕВОГО ТРАФІКУ НА ОСНОВІ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ

Через мережу Інтернет передається велика кількість особистих даних користувачів, які є вразливими до різних атак з боку внутрішніх і зовнішніх зловмисників. З розвитком інтернет-технологій мережеві атаки стають частішими, а способи атак різноманітнішими. Кібератаки можуть призвести до витоку даних і завдати величезних економічних втрат. Наприклад, Facebook втратив сотні мільйонів доларів через витік даних через хакерські атаки. Щоб протистояти дедалі складнішим мережевим атакам, все більше і більше методів