

DATA MINING ЯК НАУКА, ХАРАКТЕРИСТИКА ОСНОВНИХ ВІДМІННОСТЕЙ СПОСОБІВ ВИДОБУТКУ ДАНИХ

Лопачев Д. В.

*студент 1 курсу магістратури
факультету кібербезпеки та інформаційних технологій
Національного університету «Одеська юридична академія»*

Термін Data Mining більш поширений у повсякденному житті, але іноді його плутають з Big Data. Якщо термін «великі дані» відноситься до постійно зростаючих і оброблених даних, то інтелектуальний аналіз даних – це процес глибокого занурення в ці дані для отримання критичних знань.

Термін «інтелект даних» визначається як процес виявлення в необроблених даних раніше невідомих, важливих, практичних і добре зрозумілих пояснень знань, необхідних для прийняття рішень у різних сферах людської діяльності [1].

Використовуючи програмне забезпечення для пошуку закономірностей у великих пакетах даних, підприємства можуть розробляти маркетингові стратегії, керувати кредитним ризиком, виявляти шахрайство, фільтрувати спам і навіть виявляти настрої користувачів.

Ринок систем аналізу даних зростає. Цьому сприяє діяльність великих корпорацій: SAS, IBM, Microsoft, Oracle та ін. [2]

Існує багато програм, які можуть виконувати завдання аналізу даних. Наприклад, SAS Enterprise Miner, Microsoft Analysis Services, SAS Customer Intelligence 360, SAS Credit Scoring, Board, Sas Revenue Optimization і RapidMiner [3].

Інтелектуальний аналіз даних покладається на ефективний збір, зберігання та комп'ютерну обробку даних. Інтелектуальний аналіз даних вважається окремою дисципліною в галузі науки про дані.

Термін «інтелектуальний аналіз даних» з'явився в академічних журналах ще в 1970 році, але по-справжньому він не прижився до появи Інтернету в 1990-х роках. У той час компаніям потрібно було аналізувати величезні обсяги різнорідних даних, щоб знайти важливі закономірності та навчитися передбачати поведінку клієнтів. Виявляється, що традиційні статистичні моделі не впораються з цим завданням [1].

Останні тенденції аналізу даних включають розвиток методів елементного аналізу віртуальної та доповненої реальності, їх інтеграцію з системами баз даних, вилучення біологічних даних для медичних інновацій, вебмайнінг (аналіз даних в Інтернеті), аналіз даних у реальному часі та захист конфіденційності у діяльності з вилучення даних.

Основною проблемою виявлення шаблонів даних є час, необхідний для пошуку в масиві інформації. Відомі методи або штучно обмежують цей пошук, або

будують ціле дерево рішень, що знижує ефективність пошуку. Вирішення цієї проблеми залишається основною метою розробників продуктів інтелекту даних.

Моделі аналізу даних використовуються для багатьох типів завдань:

- Прогноз: оцінка продажів, прогноз навантаження або простою сервера;
- Ризик і ймовірність: вибір правильних клієнтів для цільових розсилок, визначення балансу сценаріїв ризику, визначення ймовірностей за діагнозом чи іншими результатами;
- Послідовність пошуку: аналіз та вибір клієнтів під час здійснення ними покупок і прогноз їхньої поведінки
- Групування: розподіл клієнтів або події на кластери, аналіз та прогноз загальних характеристик цих кластерів [2].

Аналіз даних має кілька етапів. Перший – формулювання проблеми. Цей крок включає аналіз вимог бізнесу, визначення проблемних областей, оцінку показників моделі та визначення завдань для проекту аналізу. Другий етап – підготовка даних, їх агрегація та очищення. Ця робота передбачає видалення непотрібних даних та визначення джерела найбільш точних даних для створення таблиць аналізу. дослідження даних.

Наступні етапи – дослідження даних, побудова моделі, підтвердження та дослідження цієї моделі, розгортання та оновлення моделей. Після того, як модель запущена, їй потрібно принаймні оновити нові дані, а потім повторно обробити їх.

Видобуток даних в основному використовується в таких споживчих галузях, як роздрібна торгівля, фінанси, телекомунікації, маркетинг, страхування, виробництво, соціологія, медицина та рекомендаційні системи, такі як голосовий помічник Amazon Alexa і Siri від Apple [3].

Список використаних джерел:

1. Data mining in education, Cristobal Romero and Sebastian Ventura. URL: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1075>
2. Можливості використання технологій Data Mining при створенні бізнес-проекту, Бондар Є.О. 2019. URL: http://nubip.edu.ua/sites/default/files/zbirnik_10.pdf#page=46
3. Технологія застосування методів Data Mining при розробці рекомендаційних експертних систем для підбору стилю, Литовченко О.В. 2020. URL: <https://dspace.nau.edu.ua/handle/NAU/45203>

Науковий керівник: доцент Лобода Ю.Г.