


Методи та стадії Data Mining





■ *Метод* – норма або правило, визначений шлях, спосіб, прийняття завдання теоретичного, практичного, пізнавального, управлінського характеру

■ *Алгоритм* – точний припис відносно послідовності дій (кроків), що перетворюють вихідні дані у шуканий результат



Класифікація стадій Data

Mining

- Виявлення закономірностей (вільний пошук)
- Використання виявлених закономірностей для передбачення невідомих значень (прогностичне моделювання)
- Аналіз виключень

Вільний пошук

На стадії вільного пошуку відбувається дослідження набору даних з метою пошуку прихованих закономірностей. Попередні гіпотези щодо вигляду закономірностей не визначаються.

Закономірність – суттєвий та постійно повторюваний взаємозв'язок, що визначає етапи та форми процесу становлення, розвитку різних явищ або процесів.



Вільний пошук

Дії вільного пошуку:

- виявлення закономірностей умовної логіки (conditional logic)
- виявлення закономірностей асоціативної логіки (associations and affinities)
- виявлення трендів та коливань (trends and variations)



Прогностичне моделювання

Дії прогностичного моделювання:

- ▶ передбачення невідомих значень (outcome prediction)
- ▶ прогнозування розвитку процесів (forecasting)



Аналіз виключень (forensic analysis)

Дії аналізу виключень:

- виявлення відхилень (deviation detection)



Класифікація методів Data Mining :

- ▶ за принципом роботи з вихідними навчальними даними
- ▶ за підходами до навчання математичних моделей

Методи Data Mining за принципом роботи з вихідними навчальними даними:

- безпосереднє використання даних, або збереження даних – *кластерний аналіз, метод найближчого сусіда, метод k-найближчого сусіда, міркування за аналогією*
- виявлення та використання формалізованих закономірностей, або дистиляція шаблонів – *логічні методи, методи візуалізації, методи крос-табуляції, методи, що ґрунтуються на рівняннях*

Методи Data Mining за підходами до навчання математичних моделей:

- статистичні методи – *дескриптивний аналіз та опис вихідних даних, аналіз зв'язків, багатовимірний статистичний аналіз, аналіз часових рядів*
- кібернетичні методи – *штучні нейронні мережі, еволюційне програмування, генетичні алгоритми, асоціативна пам'ять, нечітка логіка, дерева рішень, системи обробки експертних знань*