

# ЛАБОРАТОРНА РОБОТА

## №1

**Тема:** Методи побудови правила класифікації

**Інтелектуальний аналіз  
даних**

Виконав  
Студент групи  
КН-216  
Коренга В. В.

# ЗМІСТ

1. Що таке Data Mining
2. 4 задачі Data Mining
3. Класифікація
4. Методи класифікації
5. Метод найближчого сусіда

# ЩО ТАКЕ DATA MINING

*Data Mining* – набір методів, алгоритмів та засобів необхідних для знаходження в даних корисної і доступної інформації, необхідних для прийняття рішень в різноманітних сферах людської діяльності.



# 4 ЗАДАЧІ DATA MINING



Побудова асоціативних  
правил



Класифікац  
ія



Кластеризац  
ія



Регресі  
я

# КЛАСИФІКАЦІЯ

**Класифікація** – віднесення об'єктів, процесів чи явищ до наперед визначених класів на основі значень їх характеристик (властивостей)

## **Задача**

**класифікації**  
Знайти значення залежної змінної від значень характеристик об'єкта

# МЕТОДИ КЛАСИФІКАЦІЇ

- Класифікація за допомогою дерев рішень;
- Байєсова (“наївна”) класифікація;
- Класифікація за допомогою штучних нейронних мереж;
- Класифікація методом опорних векторів;
- Статистичні методи, зокрема, лінійна регресія;
- Класифікація за допомогою методу найближчого сусіда;
- Класифікація CBR-методом;
- Класифікація за допомогою генетичних алгоритмів.
- Класифікація за допомогою дерев рішень

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Метод побудований на знаходженні найкоротшої відстані між об'єктами. Для цього використовують формулу Евклідової відстані:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Відповідно до методу потрібно знайти відстань від усіх вже класифікованих об'єктів до об'єкту, який потрібно класифікувати. Хто найближче – тому присвоюється той самий клас

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Сформуємо наступну вибірку даних

| Ім'я      | Кредитна історія | Місячний дохід | Кредит видано |
|-----------|------------------|----------------|---------------|
| Олександр | Хороша           | 6000           | Відхилено     |
| Владислав | Відмінна         | 10000          | Одобрено      |
| Олена     | Задовільна       | 12000          | Відхилено     |
| Володимир | Погана           | 11000          | Відхилено     |
| Сергій    | Хороша           | 8000           | Одобрено      |
| Анастасія | Хороша           | 9000           | Одобрено      |
| Анатолій  | Задовільна       | 15000          | Одобрено      |
| Емілія    | Погана           | 7000           | Відхилено     |
| Тимофій   | Задовільна       | 13000          | Одобрено      |
| Назар     | Задовільна       | 14000          | ?             |



# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Переведемо стовбець «Кредитна історія» у чисельний формат

| Ім'я      | Кредитна історія | Місячний дохід | Кредит видано |
|-----------|------------------|----------------|---------------|
| Олександр | 4                | 6000           | Відхилено     |
| Владислав | 5                | 10000          | Одобрено      |
| Олена     | 3                | 12000          | Відхилено     |
| Володимир | 2                | 11000          | Відхилено     |
| Сергій    | 4                | 8000           | Одобрено      |
| Анастасія | 4                | 9000           | Одобрено      |
| Анатолій  | 3                | 15000          | Одобрено      |
| Емілія    | 2                | 7000           | Відхилено     |
| Тимофій   | 3                | 13000          | Одобрено      |
| Назар     | 3                | 14000          | ?             |

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Використаємо формулу Евклідової відстані для визначення відстані від шуканого об'єкта до всіх інших. Для класифікації нам потрібні 2 дві властивості – кредитна історія та місячний дохід

$$1 \quad \sqrt{(3 - 4)^2 + (14000 - 6000)^2} = 8000.00006$$

$$2 \quad \sqrt{(3 - 5)^2 + (14000 - 10000)^2} = 4000.0005$$

$$3 \quad \sqrt{(3 - 3)^2 + (14000 - 12000)^2} = 2000$$

...

$$9 \quad \sqrt{(3 - 3)^2 + (14000 - 13000)^2} = 1000$$

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Запишемо відстані між даними і визначимо найближчих сусідів

| Ім'я      | Відстань   | Кредит видано |
|-----------|------------|---------------|
| Олександр | 8000.00006 | Відхилено     |
| Владислав | 4000.0005  | Одобрено      |
| Олена     | 2000       | Відхилено     |
| Володимир | 3000.00017 | Відхилено     |
| Сергій    | 6000.00008 | Одобрено      |
| Анастасія | 5000.0001  | Одобрено      |
| Анатолій  | 1000       | Одобрено      |
| Емілія    | 7000.00007 | Відхилено     |
| Тимофій   | 1000       | Одобрено      |
| Назар     | 0          | ?             |

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Маємо два найближчих сусіда

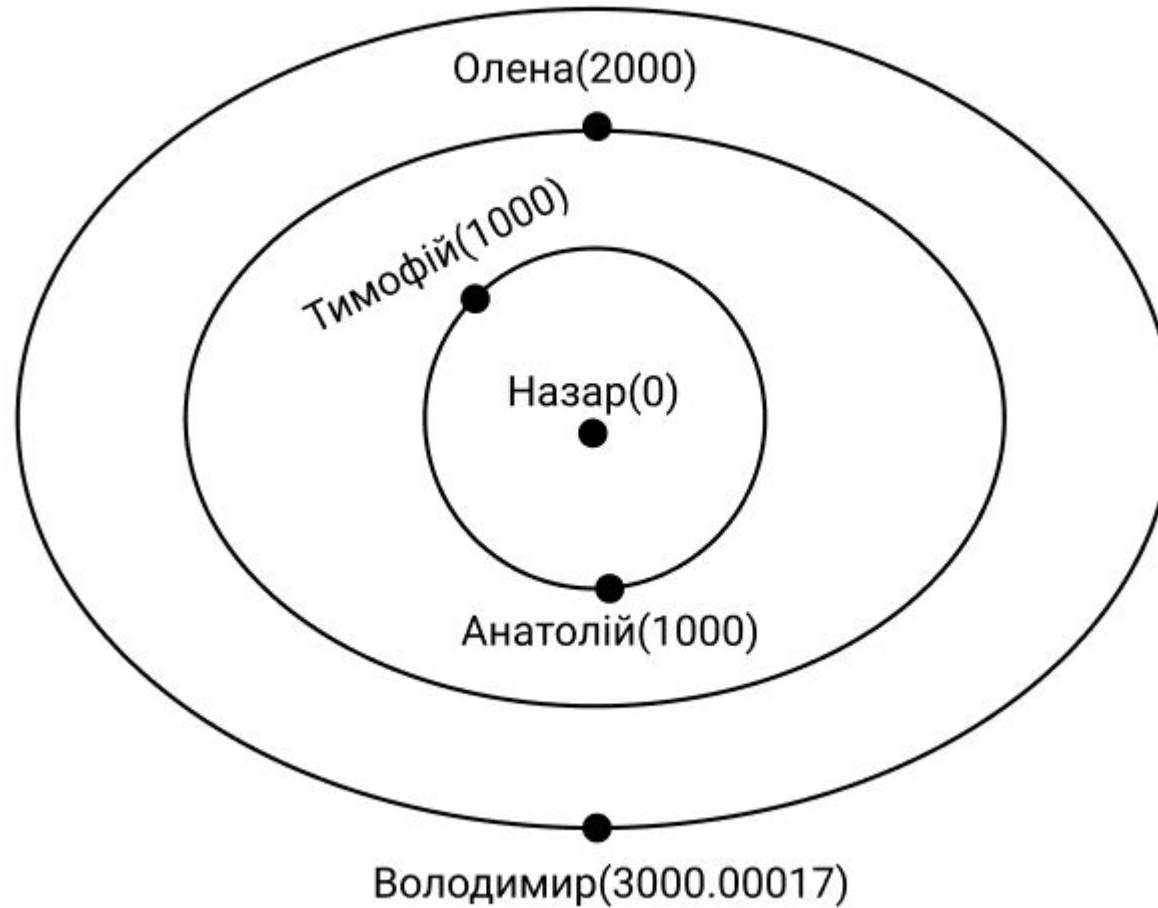
|          |      |          |
|----------|------|----------|
| Анатолій | 1000 | Одобрено |
| Тимофій  | 1000 | Одобрено |

Отже можемо видати кредит даному клієнту (віднести його до класу клієнтів з одобреними кредитами)

|       |   |          |
|-------|---|----------|
| Назар | 0 | Одобрено |
|-------|---|----------|

# МЕТОД НАЙБЛИЖЧОГО СУСІДА

Намалюємо графічне представлення вибірки з відстаннями



ДЯКУЮ ЗА  
УВАГУ

---

