

## Варіант 1.

1. Метод опорних векторів для лінійно роздільного випадку: сутність методу. Який клас використовується для навчання за цим методом в scikit-learn, його основні параметри?

**Р 12.** Знайти правило класифікації за алгоритмом 1R для наступного набору даних:

№ клієнта	збереження	інші активи	кредитний ризик (y) – цільова змінна
1	середні	високі	низький
2	низькі	низькі	високий
3	високі	середні	середній
4	середні	середні	середній
5	високі	високі	низький
6	низькі	високі	низький
7	високі	низькі	низький
8	середні	низькі	високий

## Варіант 2.

1. Наївний байесівський класифікатор (Naive Bayes). Який клас використовується в scikit-learn для навчання моделі Naive Bayes? Описати його основні параметри.

**Р 13.** Застосувати алгоритм 1R для класифікації наступного набору даних:

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дош
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні

### Варіант 3.

1. Нелінійне узагальнення методу опорних векторів. Означення ядра і приклади.

**Р 14.** Знайти правило класифікації за алгоритмом 1R для наступного набору даних:

<b>Вік&gt;35</b>	<b>Дохід&gt;20</b>	<b>Нерухомість</b>	<b>Освіта</b>	<b>Кредит (y) – цільова змінна</b>
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

#### Варіант 4.

1. Як побудувати і навчити модель дерева рішень для класифікації в scikit-learn Python? Атрибути вузла дерева. Міра забрудненості Джині.

**Б 115.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: Куріння=Так, Кашель=Сильний =Так, Візит до Азії =Ні, Віддишка=Ні для наступної навчальної вибірки:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 5.

1. Проблема перенавчання (зверхчутливості) дерев рішень. Алгоритм C4.5 вибору змінної розбиття. Модифікований C4.5 для випадку неперервних атрибутів.

**Б 116.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: Куріння=Так, Кашель=Сильний=Ні, Візит до Азії=Так, Віддишка=Середня для наступної навчальної вибірки:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 6.

1. PR-крива та ROC-крива для оцінювання якості класифікації. Як вони будуються в scikit-learn?

**Б 117.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: Куріння=Ні, Кашель=Сильний=Ні, Візит до Азії=Так, Віддишка=Так для наступної навчальної вибірки:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 7.

1. Порівняння розв'язків за методом опорних векторів для лінійно роздільного і нероздільного випадків. Переваги і обмеження методу. Як визначається параметр  $C$  ?

**Б 118.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: Куріння=Ні, Кашель=Сильний=Так, Візит до Азії=Так, Віддишка=Середня для наступної навчальної вибірки:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 8.

1. Алгоритм розбиття побудови дерев рішень. Властивості алгоритму розбиття.

**Б 119.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: збереження=низькі, інші активи= високі, дохід=100 для наступної навчальної вибірки:

№ клієнта	збереження	інші активи	річний дохід	кредитний ризик (y) – цільова змінна
1	середні	високі	75	низький
2	низькі	низькі	50	високий
3	високі	середні	25	середній
4	середні	середні	50	середній
5	середні	середні	100	низький
6	високі	високі	25	низький
7	низькі	низькі	25	високий
8	середні	середні	75	середній
10	низькі	високі	25	високий
11	середні	високі	25	середній
12	високі	низькі	50	низький
13	високі	середні	40	низький
14	середні	низькі	60	високий
15	середні	високі	100	низький
16	низькі	високі	125	низький
17	середні	середні	25	високий
18	високі	низькі	150	низький
19	високі	середні	20	низький
20	середні	низькі	160	високий



## Варіант 9.

1. Метод опорних векторів для лінійно роздільного випадку. Сутність і обґрунтування методу.

**Б 120.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: збереження=середні, інші активи= високі, дохід=75 для наступної навчальної вибірки:

№ клієнта	збереження	інші активи	річний дохід	кредитний ризик (y) – цільова змінна
2	низькі	низькі	50	високий
3	високі	середні	25	середній
4	середні	середні	50	середній
5	середні	середні	100	низький
6	високі	високі	25	низький
7	низькі	низькі	25	високий
8	середні	середні	75	середній
9	низькі	високі	100	низький
10	низькі	високі	25	високий
11	середні	високі	25	середній
12	високі	низькі	50	низький
13	високі	середні	40	низький
14	середні	низькі	60	високий
15	середні	високі	100	низький
16	низькі	високі	125	низький
17	середні	середні	25	високий
18	високі	низькі	150	низький
19	високі	середні	20	низький
20	середні	низькі	160	високий

## Варіант 10.

1. Крива перевірки і крива навчання. Які висновки можна отримати на їх основі (пояснити на прикладі). За допомогою яких функцій scikit-learn будуються ці криві?

**Б 121.** Знайти прогнозне значення цільової змінної за методом Naïve Bayes, якщо атрибути приймають наступні значення: **Вік>35**=Так, **Дохід>20**=Так, **Нерухомість** =Так, **Освіта** = Вища для наступної навчальної вибірки:

<b>Вік&gt;35</b>	<b>Дохід&gt;20</b>	<b>Нерухомість</b>	<b>Освіта</b>	<b>Кредит (y) – цільова змінна</b>
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 11.

1. Переваги і обмеження алгоритму розбиття. Регуляризація дерев рішень. Міри ефективності дерева рішень.

**Б 122.** Знайти прогнозне значення цільової змінної за методом Naive Bayes, якщо атрибути приймають наступні значення: **Вік>35**=Так, **Дохід>20**= Ні, **Нерухомість**=Ні, **Освіта** = Ні для наступної навчальної вибірки:

<b>Вік&gt;35</b>	<b>Дохід&gt;20</b>	<b>Нерухомість</b>	<b>Освіта</b>	<b>Кредит (y) – цільова змінна</b>
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 12.

1. Задачі машинного навчання (МН) з учителем і без учителя.

**Т 71. Чи є у пацієнта бронхіт?** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “Куріння” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

### Варіант 13.

1. Основи байесівської класифікації. Максимум апостеріорної імовірності. Штраф за помилку. Середній ризик.

**Т 72.** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “Дош” для наступного набору даних, при розрахунку Gain використати ентropійний критерій:

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дош
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні

## Варіант 14.

1. Методи зупинки побудови дерева рішень для класифікації. Як робиться зупинка побудови дерева класифікації в scikit-learn?

### Т 73. Чи є у пацієнта бронхіт?

Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h$  = “Кашель=Сильний” для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 15.

1. Оцінка ефективності класифікації: перехресна перевірка моделі. Решітчатий і рандомізований пошук. Навести назви відповідних функцій scikit-learn та їх основні параметри.

**Т 74.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h = \text{“Температура”}$  для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використовувати міру Джині.

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дош
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні

## Варіант 16.

1. Як визначаються метрики якості моделей класифікації: confusion matrix, accuracy, precision, recall, F1-score.

**Т 75. Чи є у пацієнта бронхіт?** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “ Візит до Азії ” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні



## Варіант 17.

1. Наївний байесівський класифікатор (алгоритм Naive Bayes). Який клас використовується в scikit-learn для навчання моделі Naive Bayes? Описати його основні параметри.

**Т 76. Чи є у пацієнта бронхіт?** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “**Віддишка**” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Куріння	Кашель=Сильний	Візит до Азії	Віддишка	Бронхіт (y) – цільова змінна
Так	Так	Так	Так	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Так	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Так	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Так	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Так	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Так	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Так	Ні
Ні	Ні	Ні	Ні	Ні

## Варіант 18.

1. Метод опорних векторів для лінійно роздільного випадку. Сутність і обґрунтування методу.

**Т 77.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h = \text{“Температура”}$  для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині.

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дощ
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні

## Варіант 19.

1. Постановка задачі та ідея методу опорних векторів для лінійно роздільного випадку. Який клас використовується для навчання лінійної моделі опорних векторів в scikit-learn? Описати його основні параметри.

**Т 78.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h = \text{"Вік} > 35"$  для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині.

Вік>35	Дохід>20	Нерухомість	Освіта	Кредит (y) – цільова змінна
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 20.

1. Проблема перенавчання (зверхчутливості) дерев рішень. Алгоритм C4.5 вибору змінної розбиття. Модифікований C4.5 для випадку неперервних атрибутів.

**Т 79.** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “Дохід>20” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Вік>35	Дохід>20	Нерухомість	Освіта	Кредит (у) – цільова змінна
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 21.

1. PR-крива та ROC-крива для оцінювання якості класифікації. Як вони будуються в scikit-learn?

**Т 80.** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “ Нерухомість ” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Вік>35	Дохід>20	Нерухомість	Освіта	Кредит (y) – цільова змінна
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 22.

1. Постановка задачі класифікації на два класи, поняття відступу, функціоналу помилок та функції втрат. Лінійний класифікатор.

**Т 81.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h = \text{“Освіта”}$  для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині.

Вік>35	Дохід>20	Нерухомість	Освіта	Кредит (y) – цільова змінна
Так	Так	Так	Вища	Так
Так	Так	Так	Середня	Так
Так	Так	Так	Ні	Ні
Так	Так	Ні	Вища	Так
Так	Так	Ні	Середня	Ні
Так	Так	Ні	Ні	Ні
Так	Ні	Так	Вища	Ні
Так	Ні	Так	Середня	Ні
Так	Ні	Так	Ні	Ні
Так	Ні	Ні	Ні	Ні
Ні	Так	Так	Вища	Так
Ні	Так	Так	Середня	Так
Ні	Так	Так	Ні	Так
Ні	Так	Ні	Вища	Так
Ні	Так	Ні	Середня	Ні
Ні	Так	Ні	Ні	Ні
Ні	Ні	Так	Вища	Так
Ні	Ні	Так	Середня	Ні
Ні	Ні	Ні	Вища	Ні

## Варіант 23.

1. Переваги і обмеження алгоритму розбиття. Регуляризація дерев рішень. Міри ефективності дерева рішень.

**Т 82.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h$  = “збереження” для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині.

№ клієнта	збереження	інші активи	річний дохід	кредитний ризик (y) – цільова змінна
1	середні	високі	75	низький
2	низькі	низькі	50	високий
3	високі	середні	25	середній
4	середні	середні	50	середній
5	середні	середні	100	низький
6	високі	високі	25	низький
7	низькі	низькі	25	високий
8	середні	середні	75	середній
9	низькі	високі	100	низький
10	низькі	високі	25	високий
11	середні	високі	25	середній
12	високі	низькі	50	низький
13	високі	середні	40	низький
14	середні	низькі	60	високий
15	середні	високі	100	низький
16	низькі	високі	125	низький
17	середні	середні	25	високий
18	високі	низькі	150	низький
19	високі	середні	20	низький
20	середні	низькі	160	високий

## Варіант 24.

1. Порівняння розв'язків за методом опорних векторів для лінійно роздільного і нероздільного випадків. Переваги і обмеження методу. Як визначається параметр  $C$  ?

**Т 83.** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “інші активи” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

№ клієнта	збереження	інші активи	річний дохід	кредитний ризик (y) – цільова змінна
1	середні	високі	75	низький
2	низькі	низькі	50	високий
3	високі	середні	25	середній
4	середні	середні	50	середній
5	середні	середні	100	низький
6	високі	високі	25	низький
7	низькі	низькі	25	високий
8	середні	середні	75	середній
9	низькі	високі	100	низький
10	низькі	високі	25	високий
11	середні	високі	25	середній
12	високі	низькі	50	низький
13	високі	середні	40	низький
14	середні	низькі	60	високий
15	середні	високі	100	низький
16	низькі	високі	125	низький
17	середні	середні	25	високий
18	високі	низькі	150	низький
19	високі	середні	20	низький
20	середні	низькі	160	високий



## Варіант 25.

1. Постановка задачі та ідея методу опорних векторів для лінійно роздільного випадку. Функціонал помилок.

**Т 84.** Розрахувати значення приросту інформації Gain та відношення приросту Gain-ratio для незалежної змінної “**Де грає**” для наступного набору даних, при розрахунку Gain використати ентропійний критерій:

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дош
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні

## Варіант 26.

1. Оцінка ефективності класифікації: перехресна перевірка (cross-validation) моделі. Навести назви відповідних функцій scikit-learn та їх основні параметри.

**Т 85.** Розрахувати значення величини  $G(T, x_h)$  в алгоритмі CART для незалежної змінної  $x_h = \text{“Суперник”}$  для наступного набору даних. У формулі для розрахунку  $G(T, x_h)$  в алгоритмі CART в якості  $G(T_i, V)$  використати міру Джині.

Перемога (y) – цільова змінна	Де грає	Суперник	Температура	Дощ
Так	Вдома	Вище	Норма	Ні
Так	Вдома	Нижче	Низька	Ні
Ні	Вдома	Вище	Низька	Ні
Ні	Вдома	Вище	Висока	Так
Ні	Вдома	Нижче	Висока	Так
Так	Вдома	Нижче	Норма	Так
Ні	Вдома	Вище	Висока	Ні
Так	Вдома	Нижче	Висока	Ні
Ні	В гостях	Нижче	Висока	Так
Так	В гостях	Вище	Низька	Так
Так	В гостях	Нижче	Низька	Так
Так	В гостях	Вище	Норма	Так
Ні	В гостях	Вище	Висока	Так
Так	В гостях	Нижче	Висока	Ні
Так	В гостях	Вище	Висока	Ні
Так	В гостях	Нижче	Норма	Ні
Ні	В гостях	Вище	Низька	Ні
Так	В гостях	Нижче	Низька	Ні