Міністерство освіти та науки України

Харківський національний університет радіоелектроніки

Кафедра Інформатики

Звіт

з лабораторної роботи №3

з дисципліни «Інтелектуальний аналіз даних»

Виконав: Перевірив:

Ст.гр. ІТІНФ-20-1 проф. кафедри Інформатики

Самченко С. О. Гороховатський В. О.

Харків 2022

Лабораторна робота №3

**Мета роботи:**

Освоїти практичні засоби побудови правил класифікації та використання дерев рішень для задач ІАД.

**Завдання:**

Побудувати два дерева рішень за допомогою алгоритму ID3, для двох вибірок заданих власноруч.

**Виконання:**

*Перше дерево:*

Вибірка:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Outlook** | **Temperature, °С** | **Humidity** | **Wind** | **Play Tennis** |
| 1 | Sunny | 29 | 85 | – | No |
| 2 | Sunny | 27 | 90 | + | No |
| 3 | Overcast | 28 | 78 | – | Yes |
| 4 | Rain | 21 | 96 | – | Yes |
| 5 | Rain | 20 | 80 | – | Yes |
| 6 | Rain | 18 | 70 | + | No |
| 7 | Overcast | 18 | 65 | + | Yes |
| 8 | Sunny | 22 | 95 | – | No |
| 9 | Sunny | 21 | 70 | – | Yes |
| 10 | Rain | 24 | 80 | – | Yes |
| 11 | Sunny | 24 | 70 | + | Yes |
| 12 | Overcast | 22 | 90 | + | Yes |
| 13 | Overcast | 27 | 75 | – | Yes |
| 14 | Rain | 22 | 80 | + | No |

У виборці D 14 прикладів: [9+ ; 5-]

Рисунок 3.1 — Перша сторінка виконання завдання

Entary(D) =

Розрахунок виграш інформації для кожного елемента

*Для атрибута Outlook:*

Entropy (OutlookSunny) = – 2/5\* log2 (2/5) – 3/5\* log2 (3/5) = 0.971

Entropy (OutlookOvercast) = – 4/4\* log2 (4/4) = 0

Entropy (OutlookRain) = – 3/5\* log2 (3/5) – 2/5\* log2 (2/5) = 0.971

Gain (S, Outlook) = 0.94 – 5/14\*0,971 – 4/14\*0 – 5/14\*0,971 = 0.246

*Для атрибута Temperature:*

Entropy (Temperature29) = – 1/1 \* log2 (1/1) = 0

Entropy (Temperature28) = – 1/1 \* log2 (1/1) = 0

Entropy (Temperature27) = – 1/2 \* log2 (1/2) – 1/2 \* log2 (1/2) = 1

Entropy (Temperature24) = – 2/2 \* log2 (2/2) = 0

Entropy (Temperature22) = – 1/3 \* log2 (1/3) – 1/3 \* log2 (1/3) = 0,92

Entropy (Temperature21) = – 2/2 \* log2 (2/2) = 0

Entropy (Temperature20) = – 1/1 \* log2 (1/1) = 0

Entropy (Temperature18) = – 1/2 \* log2 (1/2) – 1/2 \* log2 (1/2) = 1

Gain (S, Temperature) = 0.94 – 1/14 \* 3 \* 0 – 2/14 \* 2 \* 0 – 2/14 \* 2 \* 1 – 3/14 \* 0,92 = 0.457

*Для атрибута Humidity:*

Entropy (Humidity96) = – 1/1\* log2 (1/1) = 0

Entropy (Humidity95) = – 1/1\* log2 (1/1) = 0

Entropy (Humidity90) = – 1/2\* log2 (1/2) – 1/2\* log2 (1/2) = 1

Entropy (Humidity85) = – 1/1\* log2 (1/1) = 0

Entropy (Humidity80) = – 2/3\* log2 (2/3) – 1/3\* log2 (1/3) = 0,92

Entropy (Humidity78) = – 1/1\* log2 (1/1) = 0

Entropy (Humidity75) = – 1/1\* log2 (1/1) = 0

Entropy (Humidity70) = – 2/3\* log2 (2/3) – 1/3\* log2 (1/3) = 0,92

Entropy (Humidity65) = – 1/1\* log2 (1/1) = 0

Gain (S, Humidity) = 0.94 – 1/14 \* 0,6 – 2/14 \* 1 – 3/14 \* 2 \* 0,92 = 0.403

*Для атрибута Wind:*

Entropy (*Wind* -) = – 6/8 \* log2 (6/8) – 6/8 \* log2 (6/8) = 0,81

Entropy (*Wind* +) = – 3/6 \* log2 (6/8) – 3/6 \* log2 (6/8) = 1

Gain (S, *Wind*) = 0,94 – 8/14 \* 0,81 \* – 6/14 \* 1 = 0,048

Рисунок 3.2 — Друга сторінка виконання завдання

Найбільший виграш інформації у Temperature

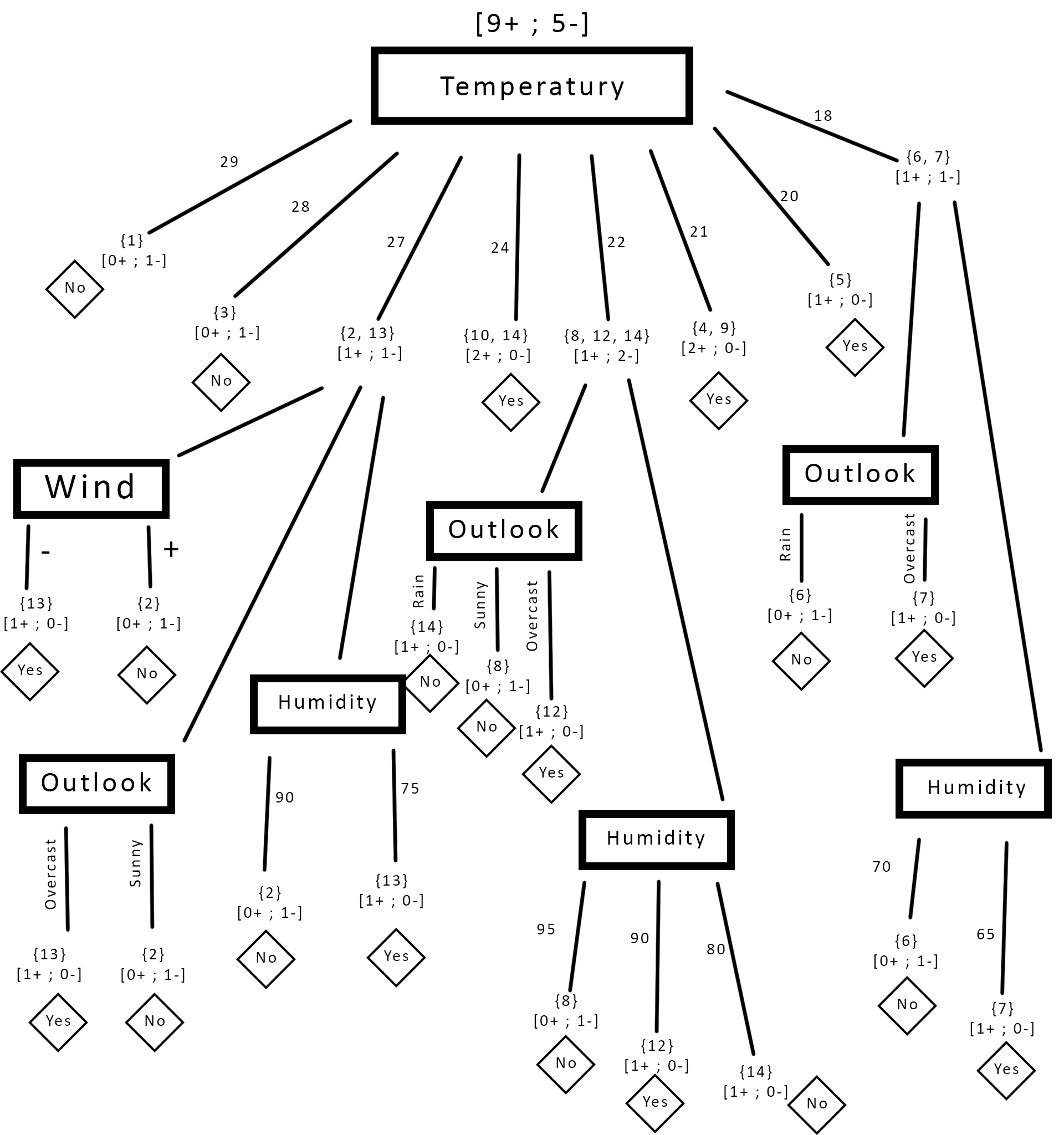
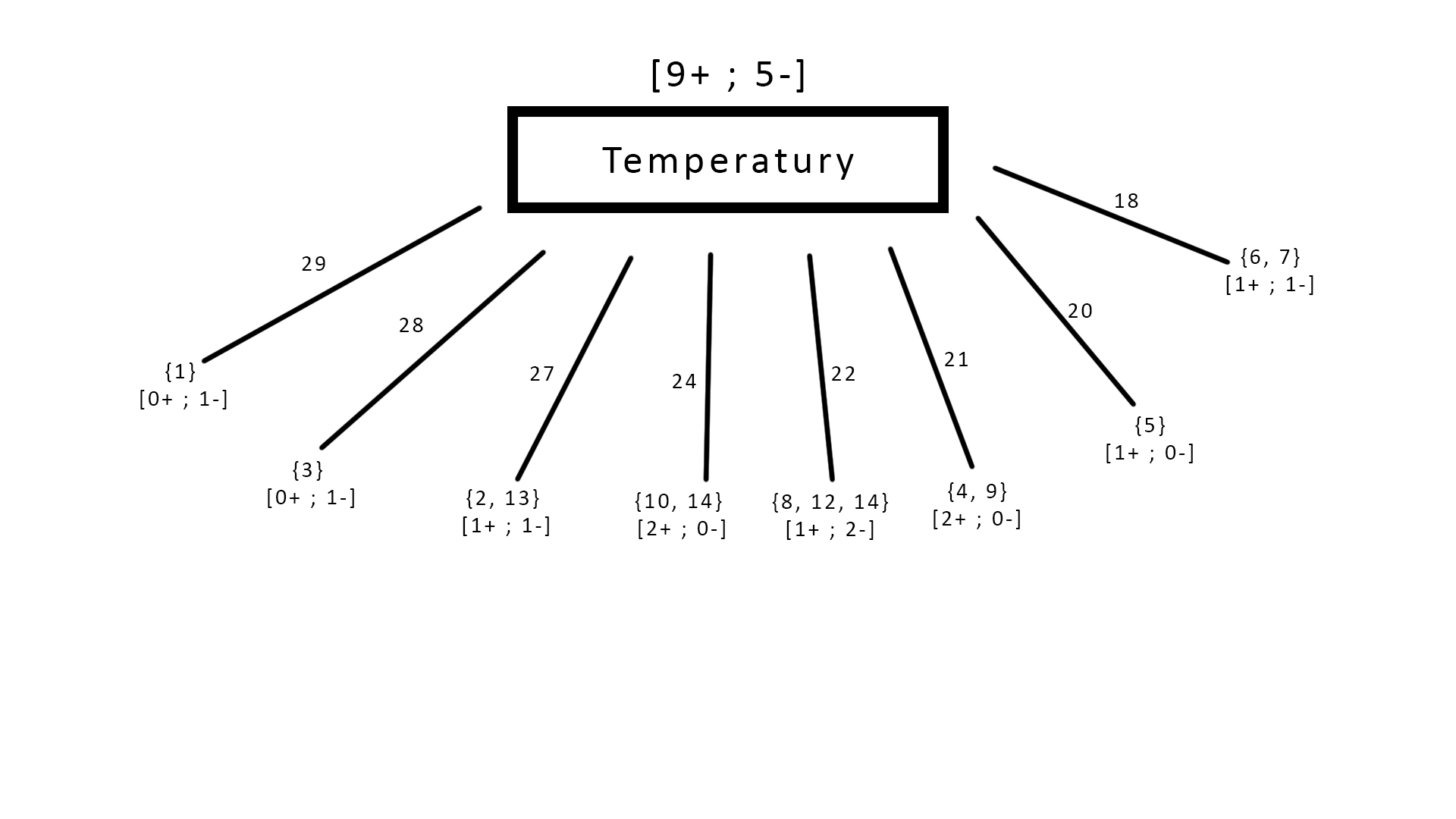


Рис. 1 – Виконання завдання

S22 = {8, 12, 14}

Gain(S22, Outlook) = 0,92 – 1/3 \* 0 – 1/3 \* 0 – 1/3 \* 0 = 0,92

Gain(S22, Humidity) = 0,92 – 1/3 \* 0 – 1/3 \* 0 – 1/3 \* 0 = 0,92

Gain(S22, Wind) = 0,92 – 1/3 \* 0 – 2/3 \* 1 = 0,253

S27 = {2, 13}

Gain(S27, Outlook) = 1 – 1/2 \* 0 – 1/2 \* 0 = 1

Gain(S27, Humidity) = 1 – 1/2 \* 0 – 1/2 \* 0 = 1

Gain(S27, Wind) = 1 – 1/2 \* 0 – 1/2 \* 0 = 1

S18 = {6, 7}

Gain(S18, Outlook) = 1 – 1/2 \* 0 – 1/2 \* 0 = 1

Gain(S18, Humidity) = 1 – 1/2 \* 0 – 1/2 \* 0 = 1

Gain(S18, Wind) = 1 – 2/2 \* 1 = 0

Рисунок 3.6 — Шоста сторінка виконання завдання

*Друге дерево:*

Вибірка:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Outlook** | **Temperature** | **Humidity** | **Wind** | **Play Tennis** |
| 1 | Sunny | Hot | High | Weak | No |
| 2 | Sunny | Hot | High | Strong | No |
| 3 | Overcast | Hot | High | Weak | Yes |
| 4 | Rain | Mild | High | Weak | Yes |
| 5 | Rain | Cool | Normal | Weak | Yes |
| 6 | Rain | Cool | Normal | Strong | No |
| 7 | Overcast | Cool | Normal | Strong | Yes |
| 8 | Sunny | Mild | High | Weak | No |
| 9 | Sunny | Cool | Normal | Weak | Yes |
| 10 | Rain | Mild | Normal | Weak | Yes |
| 11 | Sunny | Mild | Normal | Strong | Yes |
| 12 | Overcast | Mild | High | Strong | Yes |
| 13 | Overcast | Hot | Normal | Weak | Yes |
| 14 | Rain | Mild | High | Strong | No |

Крок 1.

*Ентропія всієї таблиці:*

Entropy (D) = - p+ log2 p+ - p- log2 p- = - 9/14 \* log2 (9/14) - 5/14 \* log2 (5/14) = 0.94

*Для атрибута Outlook:*

Entropy (OutlookSunny) = – 2/5\* log2 (2/5) – 3/5\* log2 (3/5) = 0.971

Entropy (OutlookOvercast) = – 4/4\* log2 (4/4) = 0

Entropy (OutlookRain) = – 3/5\* log2 (3/5) – 2/5\* log2 (2/5) = 0.971

Gain (S, Outlook) = 0.94 – 5/14\*0,971 – 4/14\*0 – 5/14\*0,971 = 0.246

*Для атрибута Temperature:*

Gain (S, Temperature) = 0.94 – 4/14\* Entropy (TemperatureHot) – 6/14\* Entropy (TemperatureMild) – 4/14\* Entropy (TemperatureCool)

Entropy (TemperatureHot) = – 2/4\* log2 (2/4) – 2/4\* log2 (2/4) = 1

Entropy (TemperatureMild) = – 4/6\* log2 (4/6) – 2/6\* log2 (2/6) = 0.918

Entropy (TemperatureCool) = – 3/4\* log2 (3/4) – 1/4\* log2 (1/4) = 0.811

Gain (S, Temperature) = 0.94 – 4/14\*1 – 6/14\*0,918 – 4/14\*0,811 =0.029

*Для атрибута Humidity:*

Gain (S, Humidity) = 0.94 – 7/14\* Entropy (HumidityHigh) – 7/14\* Entropy (HumidityNormal)

Entropy (HumidityHigh) = – 3/7\* log2 (3/7) – 4/7\* log2 (4/7) = 0.985

Entropy (HumidityNormal) = – 6/7\* log2 (6/7) – 1/7\* log2 (1/7) = 0.592

Gain (D, Humidity) = 0.94 – (7/14)\*0.985 – (7/14)\*0.592 = 0.151

*Для атрибута Wind:*

Gain (S, Wind) = 0.94 – 8/14\* Entropy (WindWeak) – 6/14\* Entropy (WindStrong)

Entropy (WindWeak) = – 6/8\* log2 (6/8) – 2/8\* log2 (2/8) = 0.811

Entropy (WindStrong) = – 3/6\* log2 (3/6) – 3/6\* log2 (3/6) = 1

Gain (D, Wind) = 0.94 – (8/14)\*0.811 – (6/14)\*1.0 = 0.048

Максимальний виграш інформації у атрибута **Outlook = 0.246**, тому він попадає в корінь дерева рішень



Рис. 2 – Перший крок алгоритму ID3

Крок 2. Для підмножини прикладів із значенням атрибута Outlook = Sunny необхідно знайти виграш інформації.

SSunny= {1, 2, 8, 9, 11}

Gain (SSunny, Wind) = 0.97 – 2/5\*0 – 2/5\*1– 2/5\*0 = 0.019

Gain (SSunny, Temperature) = 0.97 – 2/5\*0 – 2/5\*1– 2/5\*0 = 0.57

Gain (SSunny, Humidity) = 0.97 – 3/5\*0 – 2/5\*0 = 0.97

Максимальний виграш інформації у атрибута Humidity, тому дерево рішень має наступний вигляд:



Рис. 3 – Другий крок алгоритму ID3

Крок 3. Для підмножини прикладів із значенням атрибута Outlook = Rain необхідно знайти виграш інформації.

SRain = {4, 5, 6, 10, 14}

Gain (SRain, Wind) = 0.97 – 3/5\*0 – 2/5\*0 = 0.97

Gain (SRain, Temperature) = 3/5\*0,92 – 2/5\*1 = 0.019

Gain (SRain, Humidity) = 3/5\*0,92 – 2/5\*1 = 0.019

Максимальний виграш інформації у атрибута Wind, тому дерево рішень має наступний вигляд:



Рис. 4 – Фінальний вигляд дерева

**Висновок:** Освоїв практичні засоби побудови правил класифікації та використання дерев рішень для задач ІАД. Побудував два дерева рішень для заданих вибірок.