# Інструкція.

# Файл «scientific trees - учень»

## **Дерево рішень (регресії)**

### Теорія.

До завдань прийняття рішень в умовах ризику, відносяться завдання, вихідні дані в яких можна описати за допомогою імовірнісних розподілів. У подібних моделях термін ризик має цілком певний сенс: розглядається кілька станів, і ми можемо зробити припущення про ймовірності настання кожного можливого стану.

Якщо рішення приймається в умовах ризику, то вартості альтернатив зазвичай описуються імовірнісними розподілами. Тобто прибуток (витрати), пов'язаний з кожним альтернативним рішенням, є випадковою величиною (повернуть або не повернуть кредит: в одному випадку ми отримаємо прибуток, в іншому - збитки). В якості критерію прийняття рішення використовується очікуване значення вартості - математичне сподівання (М). Всі альтернативи порівнюються з точки зору максимізації очікуваного прибутку або мінімізації очікуваних витрат.

Чисельні значення доходів (результати) прораховуються, починаючи з кінця "гілок", поступово наближаючись до вихідного питання.

### Приклад дерева рішення (регресії) в Excel.

A screenshot of a computer

Description automatically generated

Рис.1

## Завдання.

### Аркуш «дерево рішень»

Умова і таблиця прогнозованих прибутків містяться на аркуші. Дерево побудувати за прикладом, наведеним вище (Рис.1). Додати лічильник ймовірності від 10% до 90%. Відповідь помістити в вказаних комірках.

A screenshot of a computer

Description automatically generated

Рис.2

## **Аркуш «дерево ймовірностей»**

### Теорія.

Дерева ймовірностей — метод візуального обчислення ймовірностей шляхом побудови послідовності подій. Вони показують усі можливі результати експерименту, наприклад результати підкидання монети двічі (показано нижче). Червона лінія відділяє перше сальто від другого сальто (Рис.3).

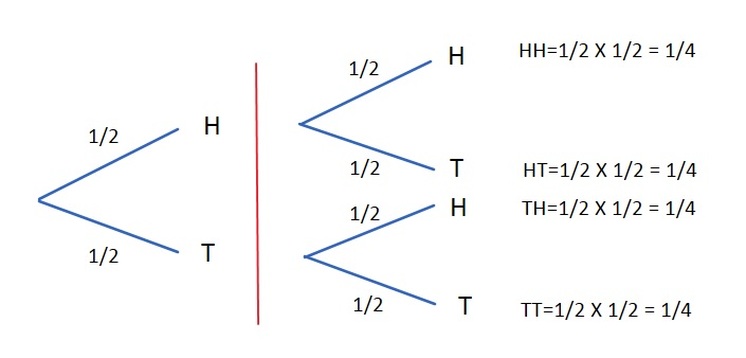


Рис.3

Кожна гілка дерева представляє можливий результат. Дроби вздовж гілок представляють імовірності цієї гілки. H = орел і T = решка. Це працює наступним чином:

у першому фліпі є два результати, H або T. Якщо результат H, тоді слідуйте за верхньою гілкою. Якщо ні, дотримуйтеся нижньої гілки. Друге перекидання призводить або до H, або до T, якщо H, то слідує за цією гілкою, або якщо T, слідує за цією гілкою. Ви множите дроби вздовж шляху чи гілки, якими ви йдете. Отже, HH = 1/2 x 1/2 або 1/4. Це означає, що шанс отримати HH становить 0,25 або 25%, або 1/4.

## Завдання

Розташувати на аркуші «дерево\_ймовірностей» дерево ймовірностей, що візуалізує результат підкидання двох гральних кубиків – ймовірність випадання пар чисел на верхній грані (парне-непарне). (Рис.4). Буквою П означено парне число, буквою Н – непарне число.

A screenshot of a computer

Description automatically generated

Розмістити в лівому верхньому куті зображення верхніх граней кубиків. Це зображення випадкових чисел, що знаходяться в інтервалі від 1 до 6. Вони змінюються при натисканні F9. Два об’ємні зображення кубиків, що знаходяться нижче, мають на верхніх гранях такі самі зображення , а зображення лівої і правої грані залежать від верхньої. Відповідність значень граней надано в таблиці:

Рис.4

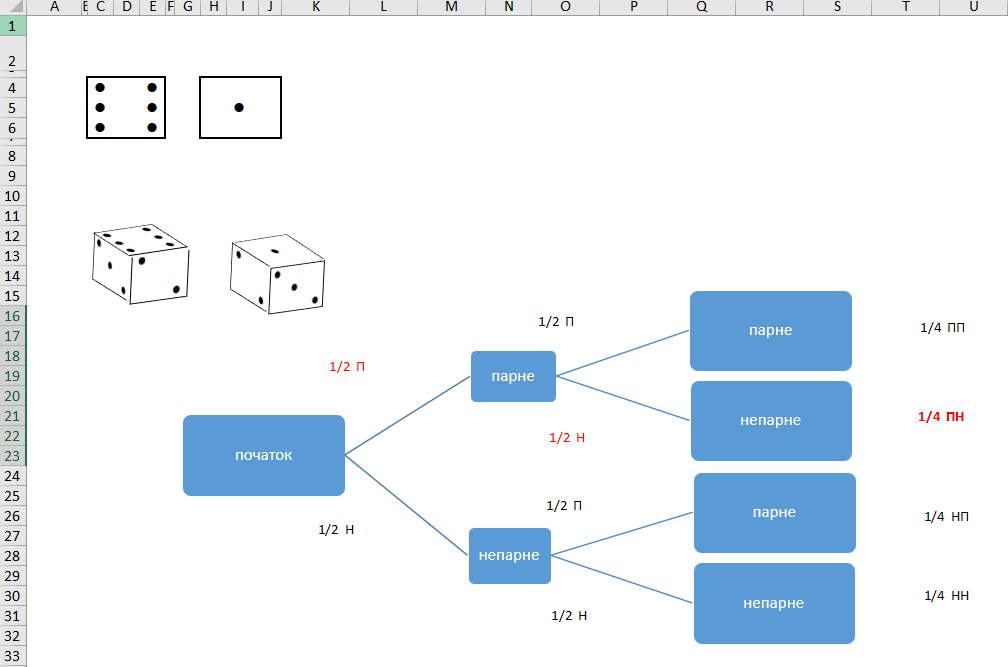
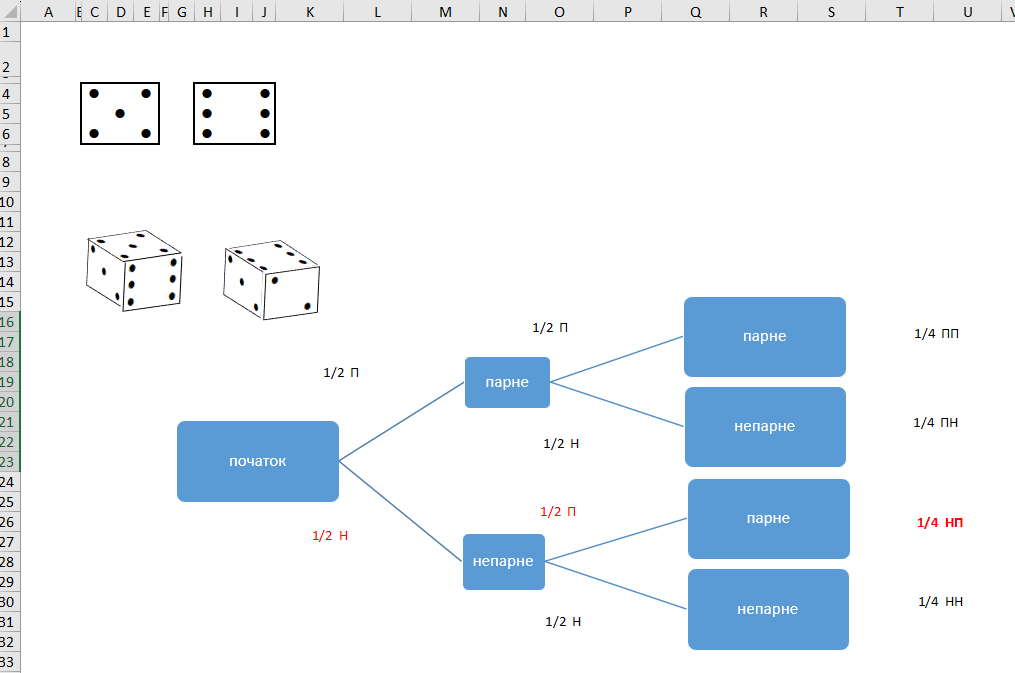
|  |  |  |
| --- | --- | --- |
| Відповідність граней кубика | | |
|  |  |  |
| Верхня грань | Ліва грань | Права грань |
| 1 | 2 | 3 |
| 2 | 3 | 6 |
| 3 | 1 | 5 |
| 4 | 5 | 6 |
| 5 | 3 | 6 |
| 6 | 3 | 2 |

При натисканні F9 повинна відбуватися зміна всіх граней кубика.

Створити зображення верхніх граней кубика, що відображають випадкові числа від 1 до 6.

Створити два об’ємних зображення кубика за наведеними вище правилами.

Побудувати дерево ймовірності засобами MS Excel.



Вздовж гілок помістити значення ймовірності даної гілки з позначкою парності/непарності. Числа біля листів дерева – ймовірність випадання відповідної пари. Ймовірності стають червоного кольору в залежності від значень верхніх граней кубиків (Рис.5). На аркуші «грані» містяться зображення всіх граней кубика.

Рис.5

# **Аркуш «Дерево рішень (класифікації)»**

Теорія.

Дерево рішень — це контрольований алгоритм машинного навчання, який використовується як для регресії, так і для формулювання проблеми класифікації.

# ID3 [Ітеративний Dichotomiser3]

(Це найпопулярніший алгоритм, який використовується для побудови дерев.

ID3 розшифровується як Iterative Dichotomizer3 і отримав таку назву тому, що алгоритм ітеративно (повторно) дихотомізує (ділить) об’єкти на дві або більше груп на кожному кроці. ID3 – це алгоритм, винайдений Россом Квінланом, який використовується для створення дерева рішень із набору даних.

ID3 є основним алгоритмом для побудови дерева рішень. Він використовує жадібний пошук зверху вниз у просторі всіх можливих гілок без повернення назад. Цей алгоритм використовує приріст інформації та ентропію для побудови класифікаційного дерева рішень.

Кроки для створення дерева рішень

a) Візьміть увесь набір даних як вхідні дані.

b) Обчисліть ентропію цільової змінної, а також атрибутів предикторів.

c) Розрахуйте інформаційний приріст усіх атрибутів.

d) Виберіть атрибут із найвищим приростом інформації як кореневий вузол

e) Повторіть ту саму процедуру для кожної гілки, доки вузол прийняття рішень для кожної гілки не буде завершено.

## Приклад

На аркуші «Simpson» надано приклад побудови дерева класифікації визначення статі персонажів мультфільму «Сімпсони» за ознаками довжини волосся, ваги, віку.

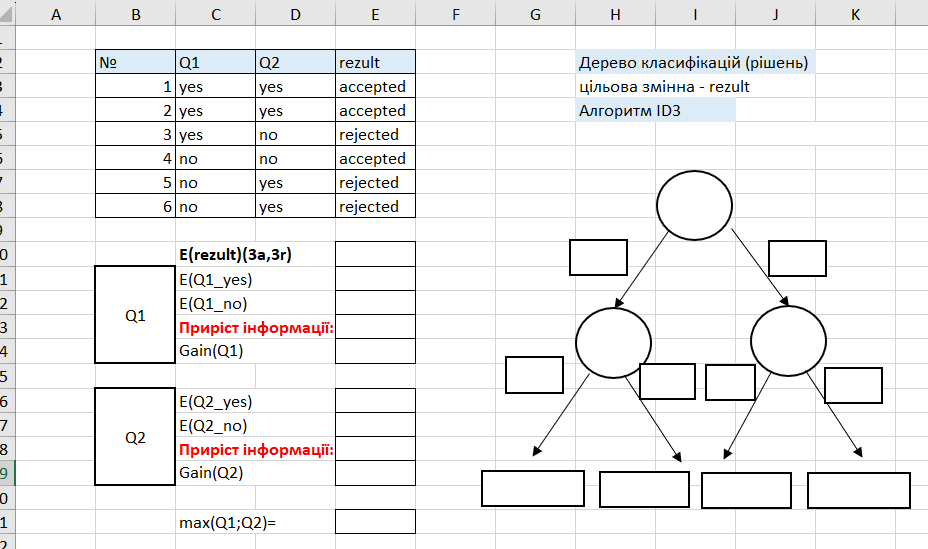
Класифікація відбувається на базі наданої таблиці.

### Завдання.

На аркуші «дерево рішень (класифікації)» міститься таблиця результатів опитування. Цільова змінна result приймає значення accepted, rejected.

Незалежні змінні (предиктори) - Q1,Q2. Вони приймають значення yes, no.

Обчислити ентропію цільової функції і атрибутів. Розрахувати інформаційний приріст кожного атрибута. Розташувати в вказаних місцях. Визначити кореневий вузол. Визначити внутрішні вузли.

 Заповнити корінь, вузли, листя в побудованому дереві значеннями. Додати написи на гілках (Рис.6)

.

Рис.6

## **. Аркуш «Дерево відрізків»**

### Теорія

Нехай задано деякий індексований масив даних ai, i=1,2,…,n

Дерево відрізків – це повне бінарне дерево, у якому кожна вершина відповідає за деякий відрізок у масиві.

Корінь дерева відповідає за весь масив, два його сини – за ліву та праву половини масиву і так далі. Кожен вузол дерева, що відповідає за більше ніж один елемент відрізка масиву, має двох синів, що відповідають за ліву та праву половини масиву. Листки дерева відповідають за окремі елементи масиву.

Для прикладу, побудуємо дерево відрізків для швидкого обчислення суми елементів масиву

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 7 | 6 | 5 | 4 | 2 | 8 |

Спочатку, створимо листки дерева, що є елементами масиву. При цьому потрібно врахувати, що оскільки дерево відрізків є повним бінарним деревом, то кількість листків має обов’язково бути степенем двійки. Це необхідно для його коректної побудови. Якщо розмір заданого вхідного масиву не є степенем двійки, то його необхідно доповнити фіктивними елементами що не вплинуть на коректну побудову дерева (тобто на значення функції на розширеному проміжку), до найближчого ступеня двійки. Наприклад, у нашому випадку, функція повертає суму на заданому проміжку, відповідно, необхідно доповнити масив нулями. У нашому випадку масив складається з восьми елементів, тому його розширення здійснювати не потрібно.

Рис.7

На наступному кроці створюємо батьківські вузли, у які записуємо суму лівого і правого синів

Рис.8

Заповнюємо наступний рівень дерева відрізків, як суму значень у вузлах передостаннього рівня (Рис.9).

Рис.10

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

L

R

Рис.9

Нарешті, остаточно, дерево відрізків набуде вигляду (Рис.10).

Пронумеруємо елементи і визначимо відрізок, причому позначимо перший елемент у відрізку, елементи якого потрібно скласти, літерою L, а останній — R (Рис.10).

Наприклад, потрібно обчислити суму з 10 по 14 елемент на дереві. L=10; R=14.

Алгоритм обчислення суми елементів відрізка за допомогою дерева відрізків надано в вигляді програмного коду.

У фрагменті програми, що реалізує обчислення суми елементів масиву, за допомогою дерева відрізків, t –масив, ans – змінна, що накопичує суму.

A math equations on a white background

Description automatically generated

## Завдання.

На аркуші «дерево відрізків» реалізувати алгоритм обчислення суми відрізка елементів масиву.A screenshot of a graph

Description automatically generated

Вхідними даними для обчислення являються номери елементів на дереві.

Рис.11

Побудувати дерево відрізків, як діаграму (Рис.11).

При зміні значень елементів масиву повинні змінюватися написи на діаграмі і, звісно, значення суми (Рис 12). При зміні вхідних даних – змінюється значення суми.

A screenshot of a graph

Description automatically generated

Рис.12

Підказка : вам пропонується заповнити таблицю, але якщо ви маєте інше рішення в межах Excel ( без VBA), будь ласка, маєте право його застосувати. Але ваше рішення повинно працювати з деревом відрізків.

**Вдалої роботи!**