

Міністерство освіти і науки України
Національний університет «Запорізька політехніка»

кафедра програмних засобів

ЗВІТ

з лабораторної роботи № 1

з дисципліни «Теорія прийняття рішень» на тему:

**«ЕВРИСТИЧНИЙ АЛГОРИТМ КЛАСИФІКАЦІЇ З
ВИКОРИСТАННЯМ ДАНИХ НАВЧАЛЬНОГО
ЕКСПЕРИМЕНТУ»**

Виконав:

ст. гр. КНТ-113сп

Іван ЩЕДРОВСЬКИЙ

Прийняв:

доцент

Олена Подковаліхіна

1 Мета роботи

Вивчити один із евристичних алгоритмів індивідуального прогнозування – класифікацію за однією ознакою з використанням даних навчального експерименту; навчитися обирати оптимальне порогове значення ознаки.

2 Завдання до лабораторної роботи

2.1. Провести класифікацію резисторів, прийнявши в якості ознаки, що характеризує їхню стабільність, напругу фліккерного шуму $U_{ш}$, а в якості прогнозованого параметру – коефіцієнт старіння α .

Масив вихідних даних, отриманих у результаті навчального експерименту, приведено в таблиці 1.

Таблиця 1 – Дані навчального експерименту

№ тр-ра	$U_{ш}$	$\alpha \cdot 10^3$	№ тр-ра	$U_{ш}$	$\alpha \cdot 10^3$	№ тр-ра	$U_{ш}$	$\alpha \cdot 10^3$	№ тр-ра	$U_{ш}$	$\alpha \cdot 10^3$
1	2.5	0.07	16	8.4	0.18	31	6.5	0.18	46	6.3	0.16
2	3.8	0.07	17	3.7	0.05	32	5.7	0.12	47	8.6	0.31
3	3.8	0.09	18	8.5	0.28	33	4.7	0.2	48	13	0.34
4	3.8	0.15	19	5.5	0.1	34	8.5	0.22	49	10.2	0.26
5	6.3	0.15	20	4.3	0.08	35	5.5	0.11	50	12	0.35
6	6.3	0.18	21	11.1	0.3	36	4.6	0.11	51	5.7	0.17
7	4.6	0.1	22	5.3	0.08	37	8.8	0.3	52	10.3	0.30
8	8.6	0.28	23	5.5	0.15	38	8.5	0.27	53	5.5	0.1
9	4.6	0.15	24	10.1	0.24	39	5.8	0.16	54	3.4	0.03
10	7.7	0.23	25	5.7	0.15	40	7.6	0.22	55	7.9	0.24
11	6.5	0.22	26	4.9	0.17	41	3.7	0.04	56	4.9	0.18
12	10.2	0.27	27	10.3	0.27	42	6.5	0.17	57	10.3	0.28
13	4.6	0.18	28	7.9	0.24	43	6.7	0.18	58	4.6	0.09
14	8.4	0.26	29	2.7	0.06	44	11.3	0.31	59	8.4	0.27
15	8.5	0.27	30	2.3	0.05	45	10.3	0.29	60	8.5	0.29

Примітка: напругу шуму $U_{ш}$ приведено в умовних одиницях.

При виконання завдання значення $U_{ш}$ помножити на номер варіанту. Порогове значення прогнозованого параметру $U_{ш.кл}$ дорівнює номеру варіанту * 10. Граничне значення $\alpha * 10^3$ дорівнює: номер варіанту / 100. Також кожне значення $U_{ш}$ потрібно помножити на номер варіанту

2.2. Написати й налагодити програму, яка реалізує:

- побудову поля кореляції;
- визначення числа вірних і помилкових рішень про віднесення екземплярів до того або іншого класу;
- визначення ймовірностей помилкових і вірних рішень;
- протоколювання результатів розрахунків у вигляді таблиці .

2.3. За даними навчального експерименту програмно реалізувати побудову поля кореляції ознаки X і прогнозованого параметра Y .

2.4. Задавшись одним з можливих порогових значень ознаки $X_{кл}$, для заданого порогового значення прогнозованого параметра $Y_{гр}$ виділити області рішень за класами $K1$ і $K2$, а також області (ріш $K1/K2$), (ріш $K1/K1$), (ріш $K2/K2$) , (ріш $K2/K1$).

2.5. Для заданого в п.3.4 $X_{кл}$ знайти значення ймовірності прийняття помилкових рішень $R_{пом}$.

2.6. Для значень $X_{кл}$ у діапазоні від X_{min} до X_{max} визначити число вірних і помилкових рішень, що полягають у віднесенні екземплярів до класів $K1$ або $K2$, тобто n (ріш $K1/K1$), n (ріш $K2/K2$), n (ріш $K1/K2$), n (ріш $K2/K1$), а також загальне число рішень, прийнятих про віднесення екземплярів, відповідно, до класу $K1$ і $K2$ – n (ріш $K1$), n (ріш $K2$).

2.7. Для значень $X_{кл}$ у діапазоні від X_{min} до X_{max} знайти ймовірності помилкових і вірних рішень $P(K1/\text{ріш } K2)$, $P(K2/\text{ріш } K1)$, $P(\text{ріш } K1/K2)$, $P(\text{ріш } K2/K1)$, $R_{пом}$, $R_{вірн}$, а також ймовірності $P(\text{ріш } K1)$, $P(\text{ріш } K2)$.

2.8. Результати, отримані в п.п. 1.3.6 та 1.3.7, програмно представити у вигляді таблиці.

2.9. Представити результати, отримані у п.1.3.8, у графічному вигляді.

2.10. Проаналізувати отримані результати й вибрати оптимальне значення X (опт), при якому імовірність прийняття вірних рішень буде максимальною

3 Текст програми

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

GENERATE_EXCEL = False

data = pd.read_csv("data.csv", sep=";", decimal=',')
A_LIMIT = 0.28
N = len(data)

print("Min:", data['U'].min(), "Max:", data['U'].max())

thresholds = np.linspace(data['U'].min(), data['U'].max(), 100)
results = []

for u_kl in thresholds:
    k1_k1 = len(data[(data['U'] <= u_kl) & (data['a'] <= A_LIMIT)])
    k2_k2 = len(data[(data['U'] > u_kl) & (data['a'] > A_LIMIT)])
    k1_k2 = len(data[(data['U'] > u_kl) & (data['a'] <= A_LIMIT)]) # Producer risk count
    k2_k1 = len(data[(data['U'] <= u_kl) & (data['a'] > A_LIMIT)]) # Consumer risk count

    n_res_k1 = k1_k1 + k2_k1
    n_res_k2 = k2_k2 + k1_k2

    p_correct = (k1_k1 + k2_k2) / N
    p_error = (k1_k2 + k2_k1) / N

    p_k2_res_k1 = k2_k1 / n_res_k1 if n_res_k1 > 0 else 0
    p_k1_res_k2 = k1_k2 / n_res_k2 if n_res_k2 > 0 else 0
    p_res_k1_k2 = k2_k1 / len(data[data['a'] > A_LIMIT])
    p_res_k2_k1 = k1_k2 / len(data[data['a'] <= A_LIMIT])

    p_res_k1 = n_res_k1 / N
    p_res_k2 = n_res_k2 / N

    results.append({
        'U_kl': u_kl,
        'n(K1/K1)': k1_k1, 'n(K2/K2)': k2_k2, 'n(K1/K2)': k1_k2, 'n(K2/K1)': k2_k1,
        'n(resK1)': n_res_k1, 'n(resK2)': n_res_k2,
        'P_correct': p_correct, 'P_error': p_error,
        'P(K2/resK1)': p_k2_res_k1, 'P(K1/resK2)': p_k1_res_k2,
        'P(resK1/K2)': p_res_k1_k2, 'P(resK2/K1)': p_res_k2_k1,
        'P(resK1)': p_res_k1, 'P(resK2)': p_res_k2
    })

res_df = pd.DataFrame(results)
```

```

# Table
pd.set_option('display.max_columns', None)
pd.set_option('display.width', 1000)
table_transposed = res_df.set_index('U_kl').T.round(2)
print(table_transposed)

if GENERATE_EXCEL:
    table_transposed.to_excel("lb1_table.xlsx")

# Optimal
best = res_df.loc[res_df['P_correct'].idxmax()]
print(f"\nOptimal threshold: {best['U_kl']:.2f}, Max Accuracy: {best['P_correct']:.4f}")

# Visualization
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(10, 12))

ax1.plot(res_df['U_kl'], res_df['P_correct'], label='P correct', color='green', lw=2)
ax1.plot(res_df['U_kl'], res_df['P_error'], label='P error', color='red', lw=2)
ax1.axvline(x=best['U_kl'], color='blue', ls='--', label=f'Optimum: {best["U_kl"]:.2f}')
ax1.set_title('Main Decision Probabilities')
ax1.legend()
ax1.grid(True)

ax2.plot(res_df['U_kl'], res_df['P(K2/resK1)'], label='Consumer Risk P(K2/resK1)', color='orange')
ax2.plot(res_df['U_kl'], res_df['P(K1/resK2)'], label='Producer Risk P(K1/resK2)', color='purple')
ax2.set_title('Risk Probabilities')
ax2.set_xlabel('U_threshold')
ax2.legend()
ax2.grid(True)

plt.tight_layout()
plt.show()

```

4 Аналіз отриманих результатів

Дані навчального експерименту показані в таблиці 1

Мій варіант це 28

Тому, граничне значення для $a * 10^3 = \frac{28}{100} = 0.28$, а граничне значення для

$$U_{ш. кл} = 28 * 10 = 280$$

Далі всі напруги $U_{ш}$ з даних навчального експерименту, які показані в таблиці 1, були помножені на номер варіанту, щоб знайти початкові дані для аналізу. Для цього було використано Excel. Також з використанням Excel було виділено області рішень за класами та побудовано діаграму кореляції

Області рішення показані разом з перетвореними даними за варіантом спеціальними кольорами. Зеленим коли і напруга шуму, і коефіцієнт старіння більші за межові параметри. Жовтим коли тільки коефіцієнт більший, а шум ні. Синім коли тільки шум більший, а коефіцієнт ні. Це показано на рисунку 1

№ гр-ра	U _ш	$\alpha \cdot 10^3$	№ гр-ра	U _ш	$\alpha \cdot 10^3$	№ гр-ра	U _ш	$\alpha \cdot 10^3$	№ гр-ра	U _ш	$\alpha \cdot 10^3$
1	70	0,07	16	235,2	0,18	31	182	0,18	46	176,4	0,16
2	106,4	0,07	17	103,6	0,05	32	159,6	0,12	47	240,8	0,31
3	106,4	0,09	18	238	0,28	33	131,6	0,2	48	364	0,34
4	106,4	0,15	19	154	0,1	34	238	0,22	49	285,6	0,26
5	176,4	0,15	20	120,4	0,08	35	154	0,11	50	336	0,35
6	176,4	0,18	21	310,8	0,3	36	128,8	0,11	51	159,6	0,17
7	128,8	0,1	22	148,4	0,08	37	246,4	0,3	52	288,4	0,3
8	240,8	0,28	23	154	0,15	38	238	0,27	53	154	0,1
9	128,8	0,15	24	282,8	0,24	39	162,4	0,16	54	95,2	0,03
10	215,6	0,23	25	159,6	0,15	40	212,8	0,22	55	221,2	0,24
11	182	0,22	26	137,2	0,17	41	103,6	0,04	56	137,2	0,18
12	285,6	0,27	27	288,4	0,27	42	182	0,17	57	288,4	0,28
13	128,8	0,18	28	221,2	0,24	43	187,6	0,18	58	128,8	0,09
14	235,2	0,26	29	75,6	0,06	44	316,4	0,31	59	235,2	0,27
15	238	0,27	30	64,4	0,05	45	288,4	0,29	60	238	0,29

Рисунок 1 – Таблиця з перетвореними даними та областями рішення

Далі на основі цієї таблиці було побудовано діаграму кореляцій. В діаграмі всі кольори та їх значення відповідають кольорам з таблиці. Діаграма показана на рисунку 2

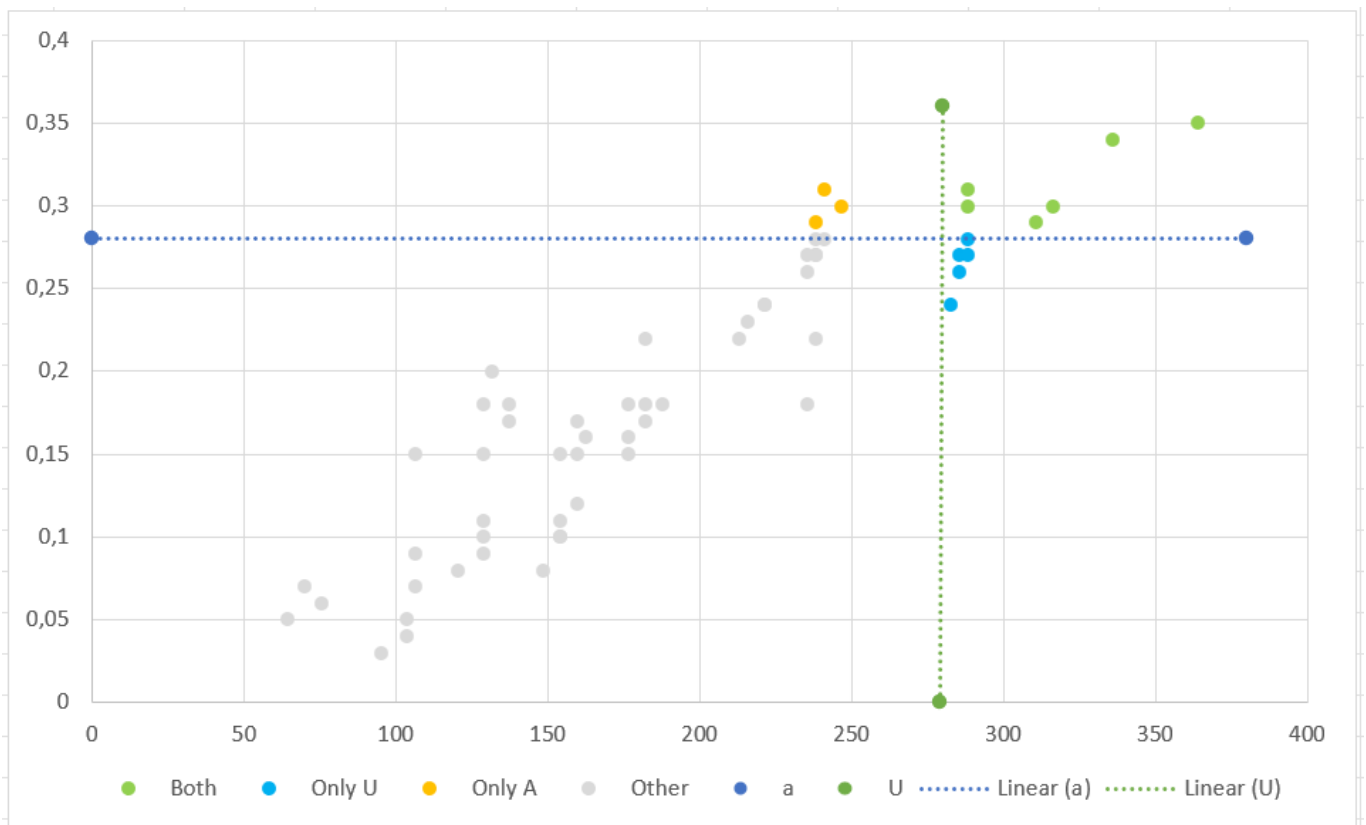


Рисунок 2 – Діаграма кореляцій

$n(K_1|розвK_1)$ – число вірних рішень про віднесення екземпляра фактично до класу K_1 до цього ж класу. Тобто $U_{ш} \leq U_{шкл}$ та $\alpha \leq \alpha_{гран}$. Такі елементи показані сірим на діаграмі та без кольору в таблиці. $n(K_1|розвK_1) = 46$

$n(K_2|розвK_2)$ – число вірних рішень про віднесення дефектних екземплярів до дефектних, класу K_2 . Тобто $U_{ш} > U_{шкл}$ та $\alpha > \alpha_{гран}$. Талі елементи показані зеленим на діаграмі та таблиці. $n(K_2|розвK_2) = 6$

$n(K_2|розвK_1)$ – число помилкових рішень, що полягає у віднесенні екземплярів класу K_2 до класу K_1 . Воно дорівнює числу екземплярів, у яких $U_{ш} \leq U_{шкл}$ та $\alpha > \alpha_{гран}$. Такі елементи показані жовтим на діаграмі та в таблиці. $n(K_2|розвK_1) = 3$

$n(K_1|розвK_2)$ - число помилкових рішень, що полягає у віднесенні екземплярів класу K_1 до класу K_2 . оно дорівнює числу екземплярів, у яких $U_{ш} > U_{шкл}$ та $\alpha \leq \alpha_{гран}$. Такі елементи позначені синім кольором на діаграмі та в таблиці. $n(K_1|розвK_2) = 5$

$$n(K_1|розвK_1) = 46$$

$$n(K_2|розвK_2) = 6$$

$$n(K_2|розвK_1) = 3$$

$$n(K_1|розвK_2) = 5$$

$n(K_1)$ та $n(K_2)$ – число екземплярів, що фактично належать до класу K_1 та K_2 відповідно. Їх сума дорівнює кількості екземплярів, що приймали участь в експерименті: $n(K_1) + n(K_2) = n$

$n(розвK_1)$ та $n(розвK_2)$ – загальне число рішень, що прийняті про віднесення екземплярів відповідно до класу K_1 та K_2 за прогнозом: $n(розвK_1) + n(розвK_2) = n$

$$n(K_1|розвK_1) + n(K_1|розвK_2) = n(K_1)$$

$$n(K_2|розвK_1) + n(K_2|розвK_2) = n(K_2)$$

$$n(K_1|розвK_1) + n(K_2|розвK_1) = n(розвK_1)$$

$$n(K_1|розвK_2) + n(K_2|розвK_2) = n(розвK_2)$$

$$n(K_1|розвK_1) + n(K_1|розвK_2) = n(K_1) = 46 + 5 = 51$$

$$n(K_2|розвK_1) + n(K_2|розвK_2) = n(K_2) = 3 + 6 = 9$$

$$n(K_1|розвK_1) + n(K_2|розвK_1) = n(розвK_1) = 46 + 3 = 49$$

$$n(K_1|розвK_2) + n(K_2|розвK_2) = n(розвK_2) = 5 + 6 = 11$$

Ризик споживача

$$P(K_2|розвK_1) = \frac{n(K_2|розвK_1)}{n(розвK_1)}$$

Ризик виробника

$$P(K_1|розвK_2) = \frac{n(K_1|розвK_2)}{n(розвK_2)}$$

Умовні ймовірності прийняття помилкових рішень

$$P(\text{розв}K1|K2) = \frac{n(K2|\text{розв}K1)}{n(K2)}$$

$$P(\text{розв}K2|K1) = \frac{n(K1|\text{розв}K2)}{n(K1)}$$

Умовні ймовірності належності екземпляра до класу K1 (ймовірність виявитися годним будь-якого, взятого навмання екземпляра):

$$P(K1) = \frac{n(K1)}{n}$$

та до класу K2 (ймовірність виявитися дефектним будь-якого, взятого навмання екземпляра):

$$P(K2) = \frac{n(K2)}{n}$$

Апріорні ймовірності прийняття рішень про віднесення екземпляру до класу K1

$$P(\text{розв}K1) = \frac{n(\text{розв}K1)}{n}$$

та до класу K2

$$P(\text{розв}K2) = \frac{n(\text{розв}K2)}{n}$$

Ймовірність помилки

$$P_{\text{пом}} = \frac{n(K2|\text{розв}K1) + n(K1|\text{розв}K2)}{n}$$

Та ймовірність прийняття правильних рішень (ефективність розпізнання)

$$P_{\text{вірн}} = 1 - P_{\text{пом}} = \frac{n(K1|\text{розв}K1) + n(K2|\text{розв}K2)}{n}$$

$$P(K2|розвK1) = \frac{n(K2|розвK1)}{n(розвK1)} = \frac{3}{49} = 0.0612$$

$$P(K1|розвK2) = \frac{n(K1|розвK2)}{n(розвK2)} = \frac{5}{11} = 0.4545$$

$$P(розвK1|K2) = \frac{n(K2|розвK1)}{n(K2)} = \frac{3}{9} = \frac{1}{3} = 0.3333$$

$$P(розвK2|K1) = \frac{n(K1|розвK2)}{n(K1)} = \frac{5}{51} = 0.098$$

$$P(K1) = \frac{n(K1)}{n} = \frac{51}{60} = 0.85$$

$$P(K2) = \frac{n(K2)}{n} = \frac{9}{60} = 0.15$$

$$P(розвK1) = \frac{n(розвK1)}{n} = \frac{49}{60} = 0.8166$$

$$P(розвK2) = \frac{n(розвK2)}{n} = \frac{11}{60} = 0.1833$$

$$P_{\text{пом}} = \frac{n(K2|розвK1) + n(K1|розвK2)}{n} = \frac{3 + 5}{60} = \frac{8}{60} = 0.1333$$

$$P_{\text{вірн}} = 1 - P_{\text{пом}} = \frac{n(K1|розвK1) + n(K2|розвK2)}{n} = \frac{46 + 6}{60} = \frac{52}{60} = 0.8666$$

Для значення Ушкл, обраного за варіантом значення помилки складає 13.3%. Ця частина роботи була виконана в excel та з використанням калькулятора

Для виконання наступної частини роботи буде написана програма на python.

Мінімальним значенням напруги є 64.4, максимальним 364.0. Цей діапазон значень було розбито на 10 частин для створення таблиці, та на 100 частин для створення графіку. В графіці можливо використовувати в тому числі і 10 частин, але при їх збільшені він буде більш детальнішим

Таблиця результатів виконання програми показана в таблиці 2

Таблиця 2 – Обчислені дані

Uш	64,4	97,68	130,98	164,26	197,55	230,84	264,13	297,42	330,71	364
n(K1/K1)	1	4	15	27	34	38	46	51	51	51
n(K2/K2)	9	9	9	9	9	9	6	4	2	0
n(K1/K2)	50	47	36	24	17	13	5	0	0	0
n(K2/K1)	0	0	0	0	0	0	3	5	7	9
n(resK1)	1	4	15	27	34	38	49	56	58	60
n(resK2)	59	56	45	33	26	22	11	4	2	0
P correct	0,17	0,22	0,4	0,6	0,72	0,78	0,87	0,92	0,88	0,85
P error	0,83	0,78	0,6	0,4	0,28	0,22	0,13	0,08	0,12	0,15
P(K2/resK1)	0	0	0	0	0	0	0,06	0,09	0,12	0,15
P(K1/resK2)	0,85	0,84	0,8	0,73	0,65	0,59	0,45	0	0	0
P(resK1/K2)	0	0	0	0	0	0	0,33	0,56	0,78	1
P(resK2/K1)	0,98	0,92	0,71	0,47	0,33	0,25	0,1	0	0	0
P(resK1)	0,02	0,07	0,25	0,45	0,57	0,63	0,82	0,93	0,97	1
P(resK2)	0,98	0,93	0,75	0,55	0,43	0,37	0,18	0,07	0,03	0

Графічне представлення даних для 100 частин показано на рисунку 3

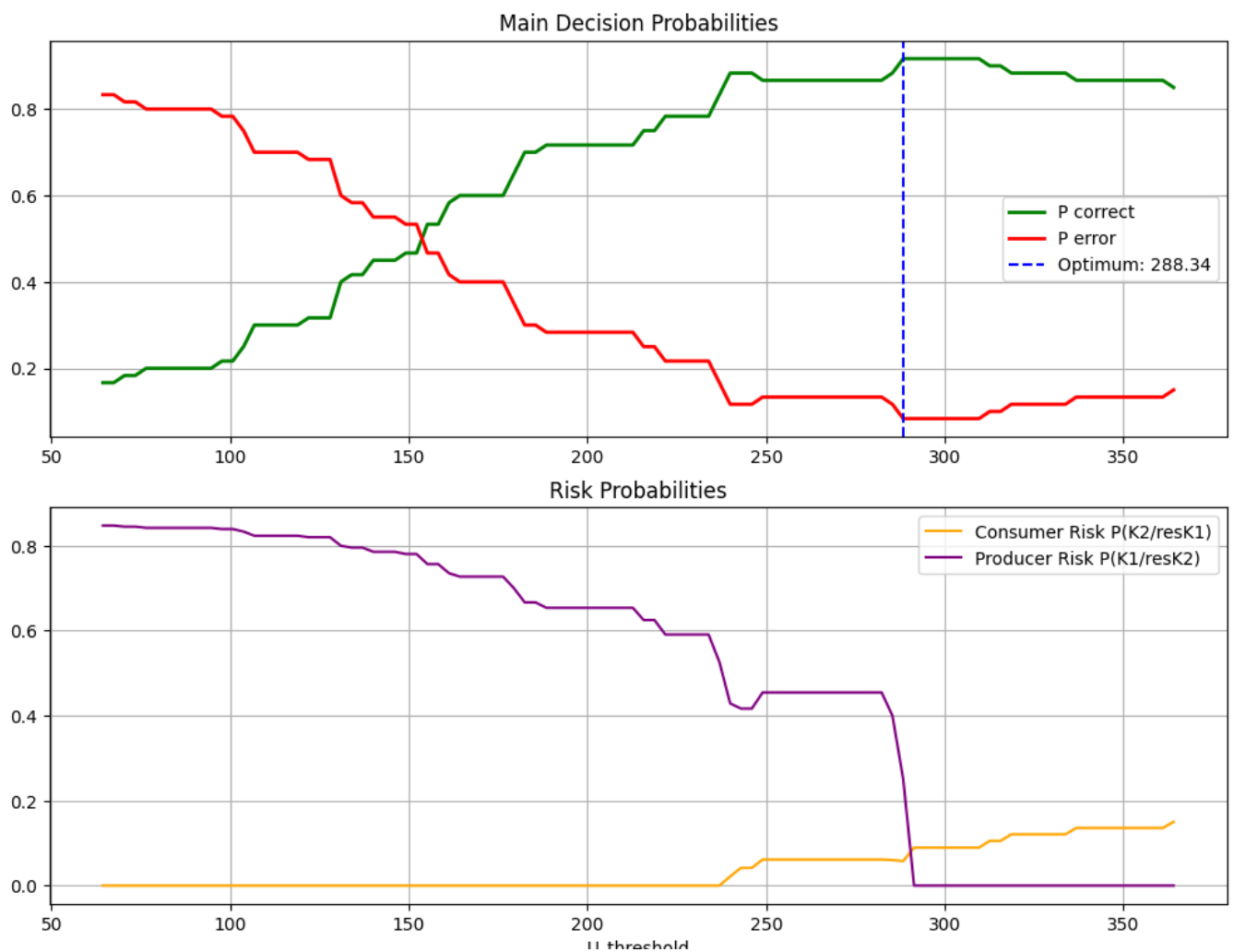


Рисунок 3 – Графічне представлення даних

Оптимальним значенням, при якому імовірність прийняття вірних рішень буде максимальною, є 288.34 при 10 частинах та 297.42 при 100 частинах. Вони обидва мають акуратність ($P_{correct}$) в 0.9167.

5 Висновки

Я вивчив один із евристичних алгоритмів індивідуального прогнозування – класифікацію за однією ознакою з використанням даних навчального експерименту; навчився обирати оптимальне порогове значення ознаки