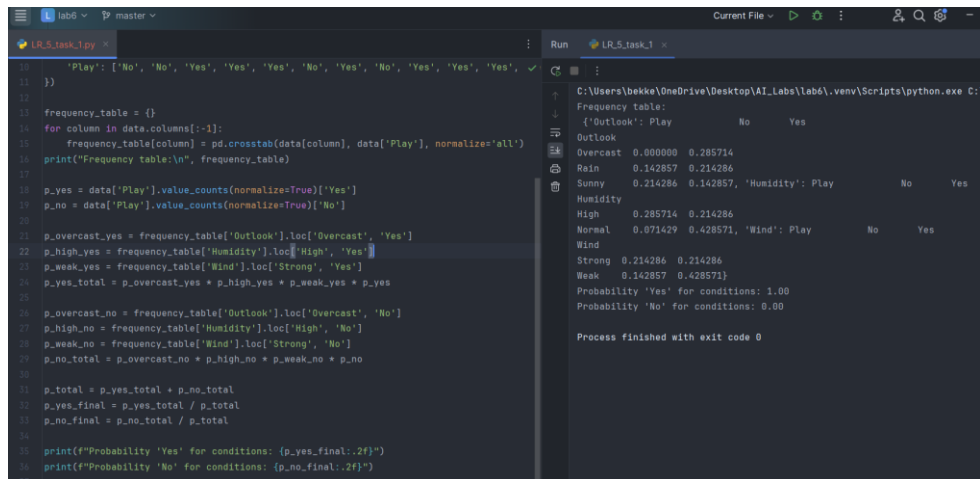


ЛАБОРАТОРНА РОБОТА № 6

Мета роботи: набути навичок працювати з даними і опонувати роботу у Python з використанням теореми Байєса

Завдання 1. Використовуючи данні з пункту 2 визначити відбудеться матч при наступних погодних умовах чи ні: Розрахунки провести з використанням Python



```
10 'Play': ['No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'No']
11 }
12
13 frequency_table = {}
14 for column in data.columns[1:]:
15     frequency_table[column] = pd.crosstab(data[column], data['Play'], normalize='all')
16 print("Frequency table:\n", frequency_table)
17
18 p_yes = data['Play'].value_counts(normalize=True)['Yes']
19 p_no = data['Play'].value_counts(normalize=True)['No']
20
21 p_overcast_yes = frequency_table['Outlook'].loc['Overcast', 'Yes']
22 p_high_yes = frequency_table['Humidity'].loc['High', 'Yes']
23 p_weak_yes = frequency_table['Wind'].loc['Strong', 'Yes']
24 p_yes_total = p_overcast_yes * p_high_yes * p_weak_yes * p_yes
25
26 p_overcast_no = frequency_table['Outlook'].loc['Overcast', 'No']
27 p_high_no = frequency_table['Humidity'].loc['High', 'No']
28 p_weak_no = frequency_table['Wind'].loc['Strong', 'No']
29 p_no_total = p_overcast_no * p_high_no * p_weak_no * p_no
30
31 p_total = p_yes_total + p_no_total
32 p_yes_final = p_yes_total / p_total
33 p_no_final = p_no_total / p_total
34
35 print(f"Probability 'Yes' for conditions: {p_yes_final:.2f}")
36 print(f"Probability 'No' for conditions: {p_no_final:.2f}")
```

Frequency table:

Outlook	Play	No	Yes
Overcast		0.080000	0.285714
Rain		0.142857	0.214286
Sunny		0.214286	0.142857
Humidity			
High		0.285714	0.214286
Normal		0.071429	0.428571
Wind			
Strong		0.214286	0.214286
Weak		0.142857	0.428571
Probability 'Yes' for conditions: 1.00			
Probability 'No' for conditions: 0.00			

Process finished with exit code 0

Рис 1. Код та результат його виконання

Висновок: відповідно до розрахунків, модель спрогнозувала, що гра відбудеться 100% за наступних умов:

2, 7, 12

Outlook = Overcast
Humidity = High
Wind = Strong

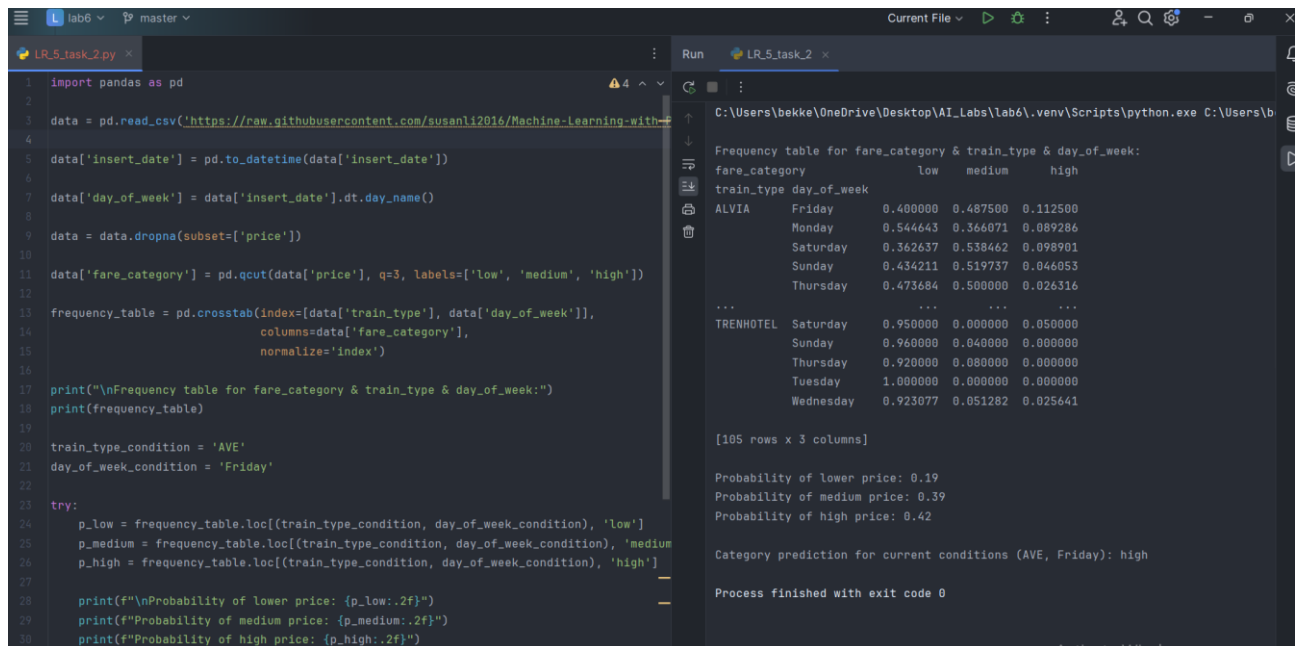
Перспектива = Похмуро
Вологість = Висока
Вітер = Сильний

Що ж вірним відповідно до таблиці правдоподібності:

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Рис 2. Таблиця правдоподібності

Завдання 2. Застосуєте методи байєсівського аналізу до набору даних про ціни на квитки на іспанські високошвидкісні залізниці.



```
1 import pandas as pd
2
3 data = pd.read_csv('https://raw.githubusercontent.com/susanli2016/Machine-Learning-with-Python/master/data/airbnb_data.csv')
4
5 data['insert_date'] = pd.to_datetime(data['insert_date'])
6
7 data['day_of_week'] = data['insert_date'].dt.day_name()
8
9 data = data.dropna(subset=['price'])
10
11 data['fare_category'] = pd.qcut(data['price'], q=3, labels=['low', 'medium', 'high'])
12
13 frequency_table = pd.crosstab(index=[data['train_type'], data['day_of_week']],
14                               columns=data['fare_category'],
15                               normalize='index')
16
17 print("\nFrequency table for fare_category & train_type & day_of_week:")
18 print(frequency_table)
19
20 train_type_condition = 'AVE'
21 day_of_week_condition = 'Friday'
22
23 try:
24     p_low = frequency_table.loc[(train_type_condition, day_of_week_condition), 'low']
25     p_medium = frequency_table.loc[(train_type_condition, day_of_week_condition), 'medium']
26     p_high = frequency_table.loc[(train_type_condition, day_of_week_condition), 'high']
27
28     print(f"\nProbability of lower price: {p_low:.2f}")
29     print(f"\nProbability of medium price: {p_medium:.2f}")
30     print(f"\nProbability of high price: {p_high:.2f}")
```

Frequency table for fare_category & train_type & day_of_week:

fare_category	low	medium	high
train_type day_of_week			
ALVIA Friday	0.400000	0.487500	0.112500
Monday	0.544643	0.366071	0.089286
Saturday	0.362637	0.538462	0.098901
Sunday	0.434211	0.519737	0.046053
Thursday	0.473684	0.500000	0.026316
...
TRENHOTEL Saturday	0.950000	0.000000	0.050000
Sunday	0.960000	0.040000	0.000000
Thursday	0.920000	0.080000	0.000000
Tuesday	1.000000	0.000000	0.000000
Wednesday	0.923077	0.051282	0.025641

[105 rows x 3 columns]

Probability of lower price: 0.19
Probability of medium price: 0.39
Probability of high price: 0.42

Category prediction for current conditions (AVE, Friday): high

Process finished with exit code 0

Рис 3. Код та результати розрахунків

На даному етапі вдалося проаналізувати структуру даних, які було завантажено через посилання, а також структуру набору цих даних. Для подальшого прогнозування, було розбито ціни на три категорії, потім створена таблиця правдоподібності і, нарешті, були розраховані ймовірності для заданих умов, а також були отримані метрики / оцінки якості.

Висновок: на даній лабораторній роботі я навчився використовувати теорему Баєса для прогнозування даних, а також закріпив знання з попередніх лабораторних робіт

Github: [link](#)