

Лабораторна робота №6

Тема: Наївний Байєс в Python

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

Хід роботи:

Завдання 6.1. Ретельно опрацювати теоретичні відомості.

Завдання 6.2. Ретельно розібрати приклад: прогнозування з використанням теореми Байєса.

Варіант	Умова	
5, 10, 15	Outlook = Rain Humidity = High Wind = Strong	Outlook = Дощ Вологість = Висока Вітер = Сильний

Загальні ймовірності (Priors):

- $P(\text{Yes}) = 9/14$
- $P(\text{No}) = 5/14$

Ймовірності для «Yes» (з відповідних таблиць):

- $P(\text{Outlook} = \text{Rain}|\text{Yes}) = 3/9$
- $P(\text{Humidity} = \text{High}|\text{Yes}) = 3/9$
- $P(\text{Wind} = \text{Strong}|\text{Yes}) = 3/9$

Ймовірності для «No» (з відповідних таблиць):

- $P(\text{Outlook} = \text{Rain}|\text{No}) = 2/5$
- $P(\text{Humidity} = \text{High}|\text{No}) = 4/5$
- $P(\text{Wind} = \text{Strong}|\text{No}) = 3/5$

$$P(\text{Yes}|\text{Rain, High, Strong}) = P(\text{Rain}|\text{Yes}) * P(\text{High}|\text{Yes}) * P(\text{Strong}|\text{Yes}) *$$

$$P(\text{Yes}) = 3/9 * 3/9 * 3/9 * 9/14 = 1/42 \approx 0,02381$$

					ЖИТОМИРСЬКА ПОЛІТЕХНІКА.25.121.15.000 – Лр.6					
Змн.	Арк.	№ докум.	Підпис	Дата	Звіт з лабораторної роботи №6			Літ.	Арк.	Аркушів
Розроб.		Кохан Т.О.								
Перевір.		Маєвський О. В.							1	4
Реценз.								ФІКТ, гр. ІПЗ-22-3		
Н. Контр.										
Зав.каф.		Вакалюк Т.А.								

$$P(\text{No}|\text{Rain, High, Strong}) = P(\text{Rain}|\text{No}) * P(\text{High}|\text{No}) * P(\text{Strong}|\text{No}) * P(\text{No})$$

$$= 2/5 * 4/5 * 3/5 * 5/14 = 12/175 \approx 0,06857$$

$$P(\text{Yes}) = 0,02381 / (0,02381 + 0,06857) = 0,2577$$

$$P(\text{No}) = 0,06857 / (0,02381 + 0,06857) = 0,7423$$

Ймовірність того, що матч не відбудеться (74,2%), вища за ймовірність того, що він відбудеться (25,8%).

Завдання 6.3. Використовую данні з пункту 2 визначити відбудеться матч при наступних погодних умовах чи ні: Розрахунки провести з використанням Python.

Лістинг програми:

```
p_rain_yes = 3/9
p_high_yes = 3/9
p_strong_yes = 3/9
p_rain_no = 2/5
p_high_no = 4/5
p_strong_no = 3/5
p_yes = 9/14
p_no = 5/14
score_yes = p_rain_yes * p_high_yes * p_strong_yes * p_yes
score_no = p_rain_no * p_high_no * p_strong_no * p_no
print(f"Показник для 'Yes': {score_yes:.5f}")
print(f"Показник для 'No': {score_no:.5f}")
total_score = score_yes + score_no
final_p_yes = score_yes / total_score
final_p_no = score_no / total_score
print("\n--- Результати прогнозу ---")
print(f"Ймовірність, що матч відбудеться ('Yes'): {final_p_yes:.1%}")
print(f"Ймовірність, що матч не відбудеться ('No'): {final_p_no:.1%}")
if final_p_yes > final_p_no:
    print("\nВисновок: Модель прогнозує, що матч відбудеться.")
else:
    print("\nВисновок: Модель прогнозує, що матч НЕ відбудеться.")
```

Результат роботи програми:

```
Показник для 'Yes': 0.02381
Показник для 'No': 0.06857

--- Результати прогнозу ---
Ймовірність, що матч відбудеться ('Yes'): 25.8%
Ймовірність, що матч не відбудеться ('No'): 74.2%

Висновок: Модель прогнозує, що матч НЕ відбудеться.

Process finished with exit code 0
```

					ЖИТОМИРСЬКА ПОЛІТЕХНІКА.25.121.15.000 – Лр.6	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		2

Завдання 6.4. Застосуєте методи байєсівського аналізу до набору даних про ціни на квитки на іспанські високошвидкісні залізниці.

Лістинг програми:

```
import pandas as pd
import pymc as pm
import arviz as az
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

df = pd.read_csv('renfe_small.csv')
df_clean = df.dropna(subset=['price', 'train_type', 'fare']).copy()

df_clean = df_clean.sample(n=200, random_state=42)

df_clean['train_type_code'] = pd.Categorical(df_clean['train_type']).codes
df_clean['fare_code'] = pd.Categorical(df_clean['fare']).codes

train_type_labels = pd.Categorical(df_clean['train_type']).categories
fare_labels = pd.Categorical(df_clean['fare']).categories

print("Дані успішно підготовлені.")
print(f"Категорії типів поїздів: {list(train_type_labels)}")
print(f"Категорії тарифів: {list(fare_labels)}")

plt.figure(figsize=(14, 6))

plt.subplot(1, 2, 1)
sns.histplot(np.log1p(df_clean['price']), kde=True)
plt.title('Розподіл лог-цін на квитки')
plt.xlabel('log(Ціна + 1)')
plt.ylabel('Частота')

plt.subplot(1, 2, 2)
sns.boxplot(x='train_type', y=np.log1p(df_clean['price']), data=df_clean)
plt.title('Лог-ціна залежно від типу поїзда')
plt.xlabel('Тип поїзда')
plt.ylabel('log(Ціна + 1)')
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()

with pm.Model() as renfe_model:
    beta0 = pm.Normal('beta0', mu=0, sigma=10)
    beta_train_type = pm.Normal('beta_train_type', mu=0, sigma=5,
shape=len(train_type_labels))
    beta_fare = pm.Normal('beta_fare', mu=0, sigma=5, shape=len(fare_labels))
    sigma = pm.HalfNormal('sigma', sigma=5)

    mu = beta0 + beta_train_type[df_clean['train_type_code']] +
beta_fare[df_clean['fare_code']]

    price_obs = pm.Normal('price_obs', mu=mu, sigma=sigma,
observed=np.log1p(df_clean['price']))

    idata = pm.sample(500, tune=250, cores=1, random_seed=42, progressbar=True)
    idata.extend(pm.sample_posterior_predictive(idata, random_seed=42))

print("Модель успішно навчена.")
```

					ЖИТОМИРСЬКА ПОЛІТЕХНІКА.25.121.15.000 – Лр.6	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		3

```
summary = az.summary(idata, var_names=['beta0', 'beta_train_type', 'beta_fare',
'sigma'], hdi_prob=0.94)
print("\nЗведення параметрів моделі:")
print(summary)

az.plot_ppc(idata, num_pp_samples=100)
plt.title('Апостеріорна предиктивна перевірка (лог-ціни)')
plt.show()
```

Результат роботи програми:

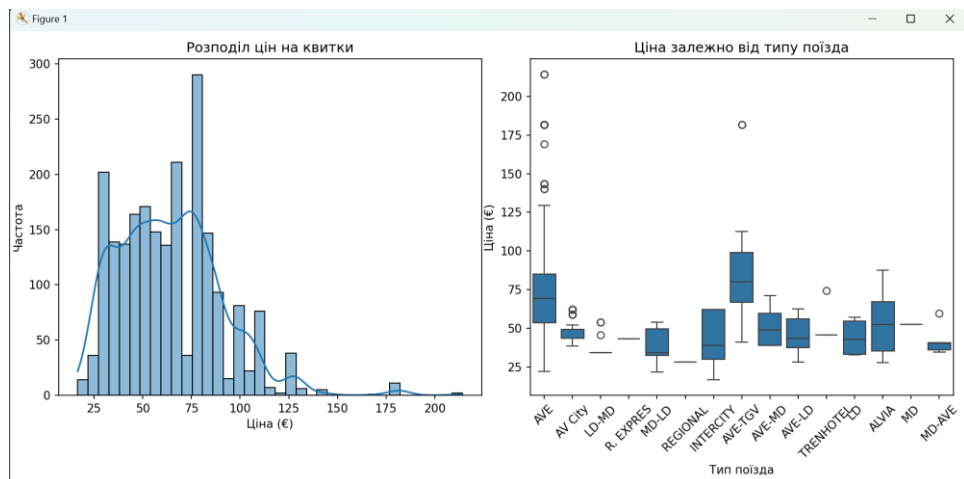


Рис. 6.1. Гістограма цін на квитки, діаграма розмаху цін за типами поїздів

```
Дані успішно підготовлені.
Категорії типів поїздів: ['ALVIA', 'AV City', 'AVE', 'AVE-LD', 'AVE-MD', 'AVE-TGV', 'INTERCITY', 'LD', 'LD-MD', 'MD', 'MD-AVE', 'MD-LD', 'R.
Категорії тарифів: ['Adulto ida', 'Flexible', 'Promo', 'Promo +']
Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Sequential sampling (2 chains in 1 job)
NUTS: [beta0, beta_train_type, beta_fare, alpha]
```

Рис. 6.2. Вивід у консоль

Висновок: під час виконання лабораторної роботи було отримано практичні навички роботи з набором даних, опановано побудову байєсівських моделей у Python, а також здійснено аналіз залежності цін на квитки від типу поїзда та тарифу з використанням теореми Байєса.

Посилання на git: <https://github.com/KokhanTetiana/AI-Systems>