

Лабораторна робота №6

Наївний Байєс в Python

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

Посилання на гіт: <https://github.com/IPZ213mmv/Lab6>

Завдання 3. Використовую данні з пункту 2 визначити відбудеться матч при наступних погодних умовах чи ні: Розрахунки провести з використанням Python.

2, 7, 12	Outlook = Overcast Humidity = High Wind = Strong	Перспектива = Похмуро Вологість = Висока Вітер = Сильний
----------	--	--

Лістинг програми:

```
# Частотні таблиці з даними
frequency_table = {
    'Outlook': {
        'Yes': {'Sunny': 3, 'Overcast': 4, 'Rain': 2},
        'No': {'Sunny': 2, 'Overcast': 0, 'Rain': 3}
    },
    'Humidity': {
        'Yes': {'High': 3, 'Normal': 6},
        'No': {'High': 4, 'Normal': 1}
    },
    'Wind': {
        'Yes': {'Weak': 6, 'Strong': 3},
        'No': {'Weak': 2, 'Strong': 3}
    }
}

# Ймовірності для цільової змінної
P_Yes = 9 / 14
P_No = 5 / 14

def calculate_probability(attribute, value, target):
    """
    Розрахунок умовної ймовірності P(Attribute=value|Target)
    """
    return frequency_table[attribute][target][value] /
sum(frequency_table[attribute][target].values())

# Задані умови
conditions = {
    'Outlook': 'Overcast',
    'Humidity': 'High',
    'Wind': 'Strong'
}

# Обчислення ймовірностей для Yes і No
P_Yes_given_conditions = (
    calculate_probability('Outlook', conditions['Outlook'], 'Yes') *
    calculate_probability('Humidity', conditions['Humidity'], 'Yes') *
    calculate_probability('Wind', conditions['Wind'], 'Yes') *
    P_Yes
)
```

					Житомирська політехніка.24.121.12.000 – Лр6							
Змн.	Арк.	№ докум.	Підпис	Дата								
Розроб.		Маліновський М.В.			Звіт з Лабораторної роботи 6			Літ.	Арк.	Аркушів		
Перевір.		Голенко М.Ю.								1	3	
Керівник								ФІКТ Гр. ІПЗ-21-3				
Н. контр.												
Зав. каф.												

```

P_No_given_conditions = (
    calculate_probability('Outlook', conditions['Outlook'], 'No') *
    calculate_probability('Humidity', conditions['Humidity'], 'No') *
    calculate_probability('Wind', conditions['Wind'], 'No') *
    P_No
)

# Нормалізація
P_Total = P_Yes_given_conditions + P_No_given_conditions
P_Yes_final = P_Yes_given_conditions / P_Total
P_No_final = P_No_given_conditions / P_Total

print("Ймовірність гри (Yes):", round(P_Yes_final, 4))
print("Ймовірність гри (No):", round(P_No_final, 4))

# Висновок
if P_Yes_final > P_No_final:
    print("Матч відбудеться.")
else:
    print("Матч не відбудеться.")

```

Результат:

```

Ймовірність гри (Yes): 1.0
Ймовірність гри (No): 0.0
Матч відбудеться.

```

Завдання 4. Застосуєте методи байєсівського аналізу до набору даних про ціни на квитки на іспанські високошвидкісні залізниці.

Лістинг програми:

```

# Завантаження та ознайомлення з даними
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report

# Завантаження даних із локального файлу
file_path = 'renfe_small.csv' # Змініть на шлях до вашого локального файлу
data = pd.read_csv(file_path)

# Перегляд перших рядків та інформації про дані
print(data.head())
print(data.info())

# Попередня обробка даних
# Видалення рядків з пропущеними значеннями
data = data.dropna()

# Перетворення категоріальних змінних на числові
categorical_columns = ['insert_date', 'origin', 'destination', 'train_type',
                        'fare', 'ticket_type']
data = pd.get_dummies(data, columns=categorical_columns, drop_first=True)

# Розділення даних на навчальну та тестову вибірки
# Припустимо, що цільова змінна - 'fare'
X = data.drop('fare', axis=1)
y = data['fare']

# Розділення на навчальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,

```

		Маліновський М.В.			Житомирська політехніка.24.121.12.000 – Лр6	Арк.
		Голенко М.Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		2

```

random_state=42)

# Навчання наївного байєсівського класифікатора
# Ініціалізація та навчання моделі
model = GaussianNB()
model.fit(X_train, y_train)

# Прогнозування на тестових даних
y_pred = model.predict(X_test)

# Оцінка точності моделі
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

print(f'Точність моделі: {accuracy}')
print('Звіт про класифікацію:')
print(report)

```

Результат:

	insert_date	origin	...	train_class	fare
0	2019-04-22 08:00:25	MADRID	...	Turista	Flexible
1	2019-04-22 10:03:24	MADRID	...	Turista	Promo
2	2019-04-25 19:19:46	MADRID	...	Turista	Promo
3	2019-04-24 06:21:57	SEVILLA	...	Turista con enlace	Promo +
4	2019-04-19 21:13:55	VALENCIA	...	Turista	Promo

		Маліновський М.В.			Житомирська політехніка.24.121.12.000 – Лр6	Арк.
		Голенко М.Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		3