

ЛАБОРАТОРНА РОБОТА № 6

ДОСЛІДЖЕННЯ МЕТОДІВ АНСАМБЛЕВОГО НАВЧАННЯ.

Мета: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити методи ансамблів у машинному навчанні.

Хід роботи

Пункт2

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

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

Варіант	Умова	
1, 6, 11	Outlook = Overcast Humidity = High Wind = Weak	Перспектива = Похмуро Вологість = Висока Вітер = Слабкий

Лістинг програми:

```
# Частотні таблиці (попередньо обчислені з набору даних)
P_Yes = 10 / 14
P_No = 4 / 14

# Ймовірності для кожного атрибута для "Yes" та "No"
likelihoods = {
    'Outlook': {
        'Overcast': {'Yes': 4 / 10, 'No': 0 / 4},
        'Sunny': {'Yes': 3 / 10, 'No': 2 / 4},
        'Rain': {'Yes': 3 / 10, 'No': 2 / 4}
    },

```

					ДУ «Житомирська політехніка».24.122.06.000 – Лр6		
Змн.	Арк.	№ докум.	Підпис	Дата			
Розроб.		Кайданович Б.Р.			Звіт з лабораторної роботи		
Перевір.		Маєвський О.В.					
Керівник							
Н. контр.							
Зав. каф.					ФІКТ Гр. КН-21-1[1]		
					Лім.	Арк.	Аркушів
						1	

```

'Humidity': {
    'High': {'Yes': 3 / 9, 'No': 4 / 5},
    'Normal': {'Yes': 6 / 9, 'No': 1 / 5}
},
'Wind': {
    'Strong': {'Yes': 6 / 9, 'No': 2 / 5},
    'Weak': {'Yes': 3 / 9, 'No': 3 / 5}
}
}

def predict_match(outlook, humidity, wind):
    # Обчислення ймовірностей для "Yes"
    P_Yes_given_conditions = (
        likelihoods['Outlook'][outlook]['Yes'] *
        likelihoods['Humidity'][humidity]['Yes'] *
        likelihoods['Wind'][wind]['Yes'] *
        P_Yes
    )

    # Обчислення ймовірностей для "No"
    P_No_given_conditions = (
        likelihoods['Outlook'][outlook]['No'] *
        likelihoods['Humidity'][humidity]['No'] *
        likelihoods['Wind'][wind]['No'] *
        P_No
    )

    # Нормалізація значень
    P_Total = P_Yes_given_conditions + P_No_given_conditions
    P_Yes_normalized = (P_Yes_given_conditions / P_Total * 100) if P_Total != 0
else 0
    P_No_normalized = (P_No_given_conditions / P_Total * 100) if P_Total != 0 else 0

    return P_Yes_normalized, P_No_normalized

# Вхідні дані (можна змінювати для різних прогнозів)
outlook = "Overcast" # Перспектива
humidity = "High" # Вологість
wind = "Weak" # Вітер

# Результат прогнозування
P_Yes_result, P_No_result = predict_match(outlook, humidity, wind)
print(f"Ймовірність, що матч відбудеться (Yes): {P_Yes_result:.2f}%")
print(f"Ймовірність, що матч не відбудеться (No): {P_No_result:.2f}%")

```

Виконання програми:

```

Ймовірність, що матч відбудеться (Yes): 100.00%
Ймовірність, що матч не відбудеться (No): 0.00%

Process finished with exit code 0

```

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

– Вхідні дані: https://raw.githubusercontent.com/susanli2016/Machine-Learning-with-Python/master/data/renfe_small.csv

					ДУ «Житомирська політехніка».24.122.06.000 – Лр6	Арк.
		Маєвський О.В.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

Лістинг програми:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report

url = "https://raw.githubusercontent.com/susanli2016/Machine-Learning-with-Python/master/data/renfe_small.csv"
data = pd.read_csv(url)

print("Перші 5 рядків даних:")
print(data.head())
print("\nКолонки в наборі даних:", data.columns, "\n")

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

# Видаляємо текстові/дата-часові колонки, які не потрібні для моделі
data = data.drop(columns=['insert_date', 'origin', 'destination', 'start_date', 'end_date', 'train_class'])

# Список колонок, які хочемо закодувати
columns_to_encode = ['train_type', 'fare']

# Фільтруємо, залишаючи тільки ті, що наявні в даних
columns_to_encode = [col for col in columns_to_encode if col in data.columns]

# Перетворення категоріальних змінних на числові за допомогою One-Hot Encoding
data = pd.get_dummies(data, columns=columns_to_encode)

# Створимо категорії цін: низька (< 25), середня (25-50), висока (> 50)
data['price_category'] = pd.cut(data['price'], bins=[0, 25, 50, np.inf], labels=['Low', 'Medium', 'High'])

# Визначимо вхідні та вихідні змінні
X = data.drop(['price', 'price_category'], axis=1)
y = data['price_category']

# Розділимо дані на тренувальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

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

# Прогнозуємо на тестовій вибірці
y_pred = model.predict(X_test)

# Оцінюємо точності моделі
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
```

Використання програми:

					ДУ «Житомирська політехніка».24.122.06.000 – Лр6	Арк.
		Маєвський О.В.				3
Змн.	Арк.	№ докум.	Підпис	Дата		

Перші 5 рядків даних:

```
      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
```

[5 rows x 9 columns]

Колонки в наборі даних: Index(['insert_date', 'origin', 'destination', 'start_date', 'end_date', 'train_type', 'price', 'train_class', 'fare'],

Accuracy: 0.3630227439471754

Classification Report:

	precision	recall	f1-score	support
High	0.92	0.29	0.44	4446
Low	0.04	0.99	0.07	164
Medium	0.87	0.46	0.60	2205
accuracy			0.36	6815
macro avg	0.61	0.58	0.37	6815
weighted avg	0.88	0.36	0.49	6815

Посилання на ГітХаб: <https://github.com/KaidanovychBohdan/SystemOfAI>

Висновок: в ході виконання лабораторної роботи опрацював спеціалізовані бібліотеки та мову програмування Python дослідити методи ансамблів у машинному навчанні.

					ДУ «Житомирська політехніка».24.122.06.000 – Лр6	Арк.
		Маєвський О.В.				4
Змн.	Арк.	№ докум.	Підпис	Дата		