

Лабораторна робота №6

НАЇВНИЙ БАЙЄС В PYTHON

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

Хід роботи:

Завдання 6.1 Ретельно опрацювати теоретичні відомості: (Опрацьовано)

- теорему Байєса;
- які типи наївного байєсівського класифікатора є;
- де використовується Наївний Байєс.

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

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

| Варіант | Умова |
|---------|--|
| 7 | Outlook = Overcast, Humidity = High, Wind = Strong |

Лістинг коду:

```
class NaiveBayes:
    def __init__(self, likelihoods, class_probs):
        self.likelihoods = likelihoods
        self.class_probs = class_probs

    def calculate_probability(self, features):
        """Обчислює ймовірність кожного класу для заданих ознак."""
        probs = {}
        for class_name, class_prob in self.class_probs.items():
            prob = class_prob
            for feature, value in features.items():
                prob *= self.likelihoods[feature][value][class_name]
            probs[class_name] = prob
        return probs

    def normalize_probabilities(self, probs):
        """Нормалізує ймовірності."""
        total = sum(probs.values())
        return {class_name: prob/total for class_name, prob in probs.items() }
```

| | | | | | | | | | | | |
|-----------|------|-------------|--------|------|--|----------------------|--|------|---|--------|--|
| | | | | | ДУ «Житомирська політехніка».24.121.07.000 – Лр6 | | | | | | |
| | | | | | Звіт з лабораторної роботи | Лім. | | Арк. | | Аркуші | |
| Змн. | Арк. | № докум. | Підпис | Дата | | | | 1 | 5 | | |
| Розроб. | | Волков О.М. | | | | ФІКТ Гр. ІПЗ-21-5[2] | | | | | |
| Перевір. | | Іванов Д.А. | | | | | | | | | |
| Керівник | | | | | | | | | | | |
| Н. контр. | | | | | | | | | | | |
| Зав. каф. | | | | | | | | | | | |

```
# Дані
likelihoods = {
    "Outlook": {"Overcast": {"Yes": 4/10, "No": 0/4}},
    "Humidity": {"High": {"Yes": 3/9, "No": 4/5}},
    "Wind": {"Strong": {"Yes": 3/9, "No": 3/5}}
}

class_probs = {"Yes": 10/14, "No": 4/14}

# Умова
features = {"Outlook": "Overcast", "Humidity": "High", "Wind": "Strong"}

# Обчислення
model = NaiveBayes(likelihoods, class_probs)
probs = model.calculate_probability(features)
normalized_probs = model.normalize_probabilities(probs)

# Виведення результату
print("Ймовірності:")
for class_name, prob in normalized_probs.items():
    print(f"    {class_name}: {prob}")
```

Результат виконання

```
"D:\ЖДТУ\1 семестр\Системи штучного
Ймовірності:
    Yes: 1.0
    No: 0.0

Process finished with exit code 0
```

Рисунок 1 - Результат виконання програми

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

Лістинг коду:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import BayesianRidge
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import StandardScaler

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

# Перевірка на пропущені значення
df['price'] = df['price'].fillna(df['price'].mean()) # Заповнюємо пропущені ціни середнім значенням

# Перетворення дати в тип datetime
df['start_date'] = pd.to_datetime(df['start_date'])
```

| | | | | | | |
|------|------|-------------|--------|------|--|------|
| | | Волков О.М. | | | ДУ «Житомирська політехніка».24.121.07.000 – Лр7 | Арк. |
| | | Іванов Д.А. | | | | 2 |
| Змн. | Арк. | № докум. | Підпис | Дата | | |

```

df['insert_date'] = pd.to_datetime(df['insert_date'])

# Створення нової колонки 'days_to_start' для кількості днів до поїздки
df['days_to_start'] = (df['start_date'] - df['insert_date']).dt.days

# Вибір лише числових та категоріальних змінних для аналізу
df_encoded = pd.get_dummies(df, columns=['train_type', 'train_class', 'fare'])

# Вибір фіч та цільової змінної (ціна)
X = df_encoded.drop(['insert_date', 'origin', 'destination', 'start_date',
                    'end_date', 'price'], axis=1)
y = df_encoded['price']

# Масштабування даних (особливо важливо для регресії)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Розбиття даних на навчальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,
                                                    random_state=42)

# Створення моделі байєсівської регресії
model = BayesianRidge()

# Навчання моделі
model.fit(X_train, y_train)

# Прогнозування
y_pred = model.predict(X_test)

# Оцінка моделі
mse = mean_squared_error(y_test, y_pred)
r2 = model.score(X_test, y_test)

# Виведення результатів
print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')

# Побудова графіків
# 1. Графік розсіяння між реальними і передбаченими цінами
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, color='blue', alpha=0.5)
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='red',
         linestyle='--') # Лінія ідеального прогнозу
plt.title('Real vs Predicted Prices')
plt.xlabel('Real Prices')
plt.ylabel('Predicted Prices')
plt.show()

# 2. Гістограма різниці між реальними та передбаченими цінами
plt.figure(figsize=(10, 6))
plt.hist(y_test - y_pred, bins=20, color='green', edgecolor='black', alpha=0.7)
plt.title('Difference between Real and Predicted Prices')
plt.xlabel('Difference')
plt.ylabel('Frequency')
plt.show()

```

Результат виконання

| | | | | | | |
|------|------|-------------|--------|------|--|------|
| | | Волков О.М. | | | ДУ «Житомирська політехніка».24.121.07.000 – Лр7 | Арк. |
| | | Іванов Д.А. | | | | 3 |
| Змн. | Арк. | № докум. | Підпис | Дата | | |



Рисунок 2 - Результат виконання програми

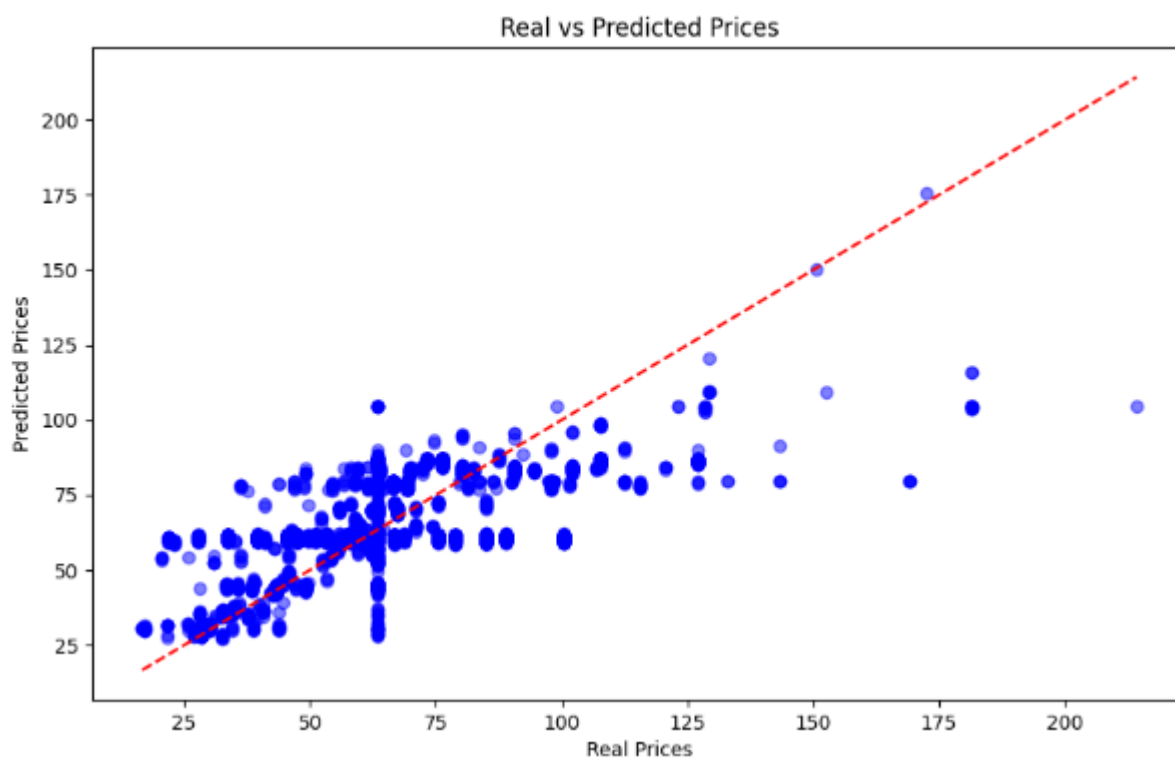


Рисунок 3 - Результат виконання програми

| | | | | | | |
|------|------|-------------|--------|------|--|------|
| | | Волков О.М. | | | ДУ «Житомирська політехніка».24.121.07.000 – Лр7 | Арк. |
| | | Іванов Д.А. | | | | 4 |
| Змн. | Арк. | № докум. | Підпис | Дата | | |

```
"D:\ЖДТУ\1 семестр\Системи штучного інт  
Mean Squared Error: 313.70792974700163  
R-squared: 0.4633947966707235  
  
Process finished with exit code 0
```

Рисунок 4 - Результат виконання програми

Висновок: набув практичних навичок роботи з даними і опонував роботу у Python з використанням теореми Байєса.

| | | | | | | |
|------|------|-------------|--------|------|--|------|
| | | Волков О.М. | | | ДУ «Житомирська політехніка».24.121.07.000 – Лр7 | Арк. |
| | | Іванов Д.А. | | | | 5 |
| Змн. | Арк. | № докум. | Підпис | Дата | | |