

กิจกรรมสัมมนาเสริม

ชุดวิชา 99711 วิทยาการข้อมูลและข้อมูลขนาดใหญ่

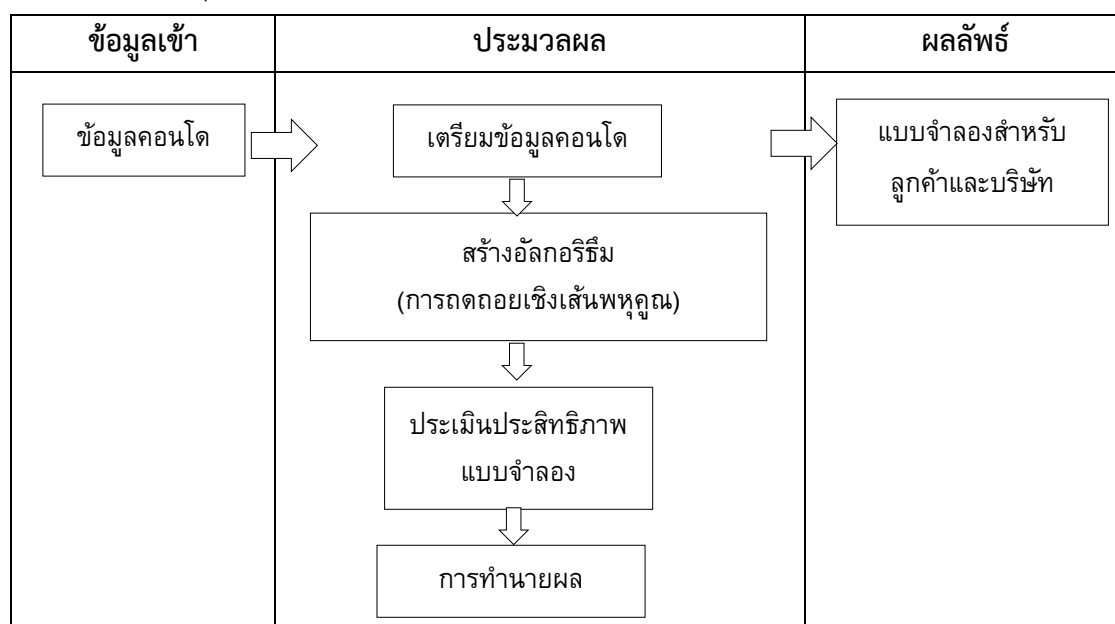
9-10 พฤศจิกายน 2567

รศ.ดร.วราภรณ์ ร่มสายหยุด (1/2567)

1) กรณีศึกษาการวิเคราะห์ราคาส่งหาริมทรัพย์โดยใช้การถดถอยเชิงเส้นพหุคูณ

ที่มาโจทย์ปัญหาวิเคราะห์ราคาคอนโดมิเนียมโดยใช้การถดถอยเชิงเส้นพหุคูณ เนื่องจากการกำหนดราคาส่งหาริมทรัพย์ โดยเฉพาะราคาคอนโดมิเนียม เกี่ยวข้องกับปัจจัยต่างๆ มากมาย เช่น ขนาดห้อง จำนวนห้องนอน ห้องน้ำ สถานที่ หรือทำเล ชั้นของห้องพัก และปัจจัยแวดล้อมอื่น ซึ่งปัจจัยต่างๆ เหล่านี้มีผลต่อการกำหนดราคาคอนโดมิเนียมที่ถูกต้อง เมื่อกำหนดราคาคอนโดมิเนียมที่ถูกต้อง จะส่งเสริมการซื้อขาย และลงทุน มีผลต่อความสำเร็จของธุรกิจอสังหาริมทรัพย์ ดังนั้นกรณีศึกษานี้มีวัตถุประสงค์หลักเพื่อวิเคราะห์ราคาคอนโดมิเนียมโดยใช้การถดถอยเชิงเส้นพหุคูณ สำหรับกำหนดราคาที่ต้อง โดยการนำข้อมูลคอนโดมิเนียม มาสร้างอัลกอริธึม ซึ่งการเลือกอัลกอริธึมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) พิจารณาจากข้อมูลปัจจัยต่างๆ ของคอนโดมิเนียม ทั้งขนาดห้อง จำนวนห้องนอน ห้องน้ำ สถานที่ หรือทำเล ชั้นของห้องพัก และปัจจัยแวดล้อมอื่น เป็นตัวแปรอิสระ (independence variable) หรือ แกน (X) และข้อมูลราคา เป็นตัวแปรตาม (dependence variable) หรือ แกน (Y) สำหรับสร้างแบบจำลองการทำนายราคาคอนโดมิเนียม ที่ถูกต้องและเหมาะสม

ขั้นตอนการวิเคราะห์ราคาคอนโดมิเนียมโดยใช้การถดถอยเชิงเส้นพหุคูณด้วยไพธอน ประกอบด้วย 3 กระบวนการทำงานหลัก ได้แก่การนำข้อมูลเข้า (input data) การประมวลผลข้อมูล (data processing) และการแสดงผลลัพธ์ (output) ดังภาพที่ 8.6



ภาพที่ 8.6 ภาพรวมการวิเคราะห์ราคาคอนโดมิเนียมโดยใช้การถดถอยเชิงเส้นพหุคูณ

ขั้นตอนการเขียนคำสั่ง รายละเอียดดังต่อไปนี้

1) นำเข้า library และแสดงข้อมูล

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn as sk
mycondo = pd.read_csv('/content/sample_data/Condo.csv')
mycondo.head(5)
```

ผลลัพธ์

	bedrooms	bathrooms	sqft_lot	floors	zone	price
0	3	1	1180	1.0	sathon	221900
1	3	2	2570	2.0	sathon	538000
2	2	1	770	1.0	sathon	180000
3	4	3	1960	1.0	sathon	604000
4	3	2	1680	1.0	sathon	510000

2) ตรวจสอบความสัมพันธ์ระหว่างข้อมูลกับคลาสเป้าหมาย

```
chkfeatures=[name for name in mycondo.columns if mycondo[name].dtype == 'object']
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
for i in list(chkfeatures):
    mycondo[i]=le.fit_transform(mycondo[i])
for x in chkfeatures:
    print(x," = ",mycondo[x].unique())
```

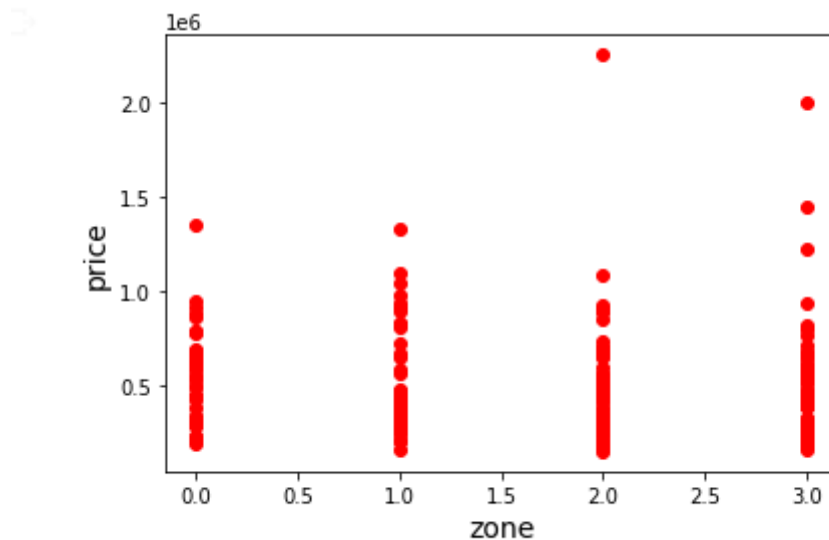
3) แสดงข้อมูล 10 รายการ

```
mycondo.head(10)
```

4) กราฟแสดงความสัมพันธ์ระหว่างข้อมูล

```
x = mycondo['zone']
y = mycondo['price']
plt.scatter(x,y,color='red')
plt.xlabel('zone', fontsize=14)
plt.ylabel('price', fontsize=14)
plt.show()
```

ผลลัพธ์



5) สร้างแบบจำลองด้วย Linear Regression

```
from sklearn.linear_model import LinearRegression
x = mycondo[['bedrooms', 'zone']]
y = mycondo['price']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, train_size=0.8, test_size=0.2, random_state=0)
modelRE = LinearRegression()
modelRE.fit(x_train, y_train)
```

6) การหาสัมประสิทธิ์ความสัมพันธ์ระหว่างข้อมูล

```
coeff1 = pd.DataFrame(modelRE.coef_, x.columns, columns=['Coefficient'])
coeff1
```

7) การหาค่าข้อมูลจริง และข้อมูลการทำนาย

```
y_predicted = modelRE.predict(x_test)
y_predicted
y_test
df = pd.DataFrame({'Actual': [y_test], 'Predicted': [y_predicted]})
print(df)
```

8) การแสดงค่า a, b1 และ b2

```
from sklearn.metrics import mean_squared_error
rmse = mean_squared_error(y_test, y_predicted)
print('Evaluation Result :\n-----')
print('The intercept is:', modelRE.intercept_)
print('The coefficient is:', modelRE.coef_)
```

```
print('The rmse is:',rmse)
```

9) การทำนายราคาคอนโด

```
X1 = 3
X2 = 2
def predictCondoPrice():
    a = xxxx
    b1 = xxx
    b2 = xxx
    error = 0
    Y = a + np.sum(b1*X1) + np.sum(b2*X2) + 0
    print (Y)
predictCondoPrice()
```

ราคาคอนโดควรจะเป็นเท่าไร.....