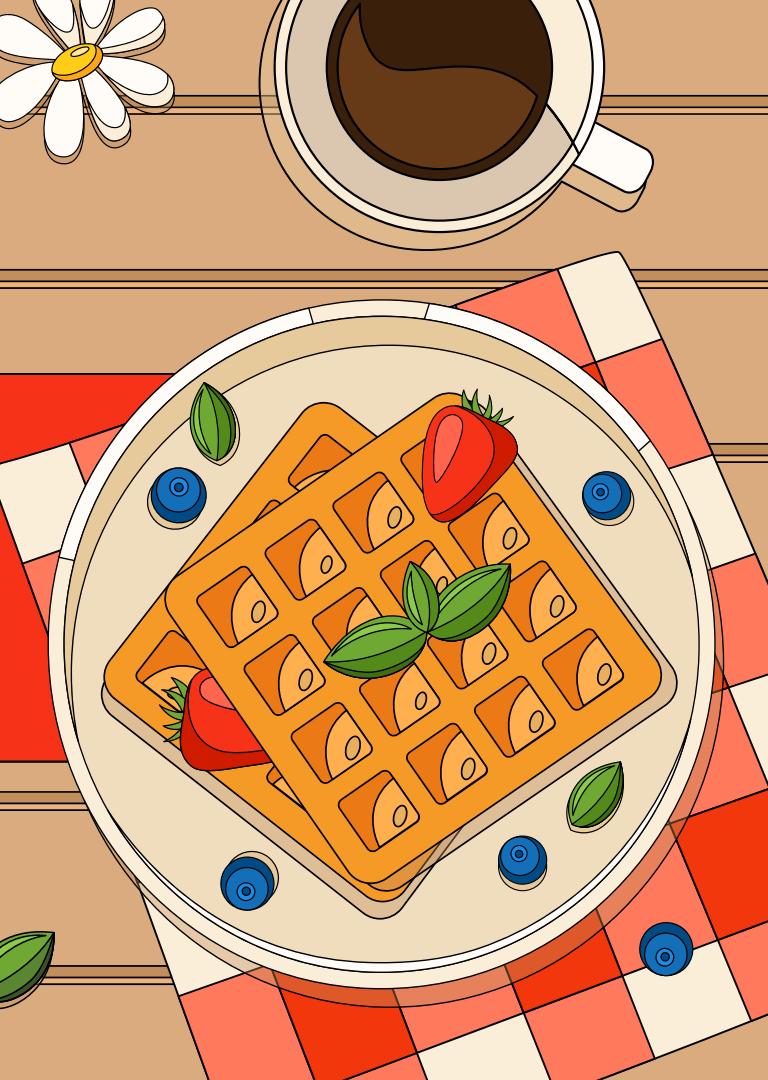


## **Assignment 3 - BDC**

**Phân tích, đánh giá và  
triển khai mô hình tối  
Ưu cho dịch vụ kinh  
doanh coffee**

**Let's get started!!!**



# Table of contents

1.

## Cơ sở lý thuyết

Lý thuyết cơ bản áp dụng vào mô hình.

3.

## Đánh giá

Đánh giá đề tài.

2.

## Mô hình

Quá trình huấn luyện mô hình.

4.

## Web

Trang web trình bày sản phẩm.

# 1.

## Cơ sở lý thuyết

Sau đây là giới thiệu cơ bản về những module chính được dùng cho mô hình.

# Cơ sở lý thuyết

**Scikit-learning** hay **sklearn** là thư viện học máy phổ biến nhất trong cộng đồng khoa học dữ liệu.

Và trong phần một này, chúng ta sẽ cùng tìm hiểu những module chính trong **sklearn** xuất hiện trong phần thực hành project.

# About Scikit-learning

**Scikit-learning** hay **sklearn** là thư viện học máy phổ biến nhất trong cộng đồng khoa học dữ liệu. Được viết bằng Python, **scikit-learning** cung cấp các công cụ hiệu quả và dễ sử dụng để xử lý dữ liệu đến triển khai các mô hình học máy.

**Scikit-learn** phổ biến chính vì sự dễ sử dụng của thư viện này, ta có thể dễ dàng xây dựng một mô hình học máy phổ biến thông qua một số câu lệnh **sklearn**.

# Scikit-learning / Sklearn

**Feature Selection**

Mutual\_info\_regression

**Model Selection**

train\_test\_split

**Tree**

DecisionTreeRegressor

# Scikit-learning / Sklearn

**Ensemble**

Random Forest

**Linear model**

Linear Regression

**Chỉ số đánh giá**

MSE

RMSE

R-Squared

# Feature Selection

Tên hàm

Mutual\_info\_regression

Chức năng

Hàm đo mức độ phụ thuộc lẫn nhau giữa hai biến ngẫu nhiên

Cú pháp

```
sklearn.feature_selection.mutual_info_regression(X, y)
```

# Model Selection

Tên hàm

train\_test\_split

Chức năng

Dùng để chia data thành 2 bộ dữ liệu để train và test

Cú pháp

```
sklearn.model_selection.train_te  
st_split(*arrays, test_size=None,  
train_size=None)
```

# Decision Tree

**Decision Tree** là một công cụ ra quyết định bằng cách sử dụng sơ đồ cây như flowchart hoặc là một mô hình cây quyết định với các kết quả có thể xảy ra. Thuật toán **Decision Tree** thuộc loại thuật toán học có giám sát. Nó có thể được dùng cho cả biến liên tục hay phân loại.

# Decision Tree Regression

**Decision Tree Regression:** xem xét các đặc trưng của đối tượng và đào tạo trong mô hình ở dạng cấu trúc cây để dự đoán dữ liệu trong tương lai nhằm tạo ra đầu ra là giá trị liên tục có ý nghĩa.

*sklearn.tree.DecisionTreeRegressor()*

# Ensemble

Ưu

Cải thiện độ chính xác  
của kết quả dự đoán.

Nhược

Khó để hiểu cách  
phân loại của  
**Ensemble**.



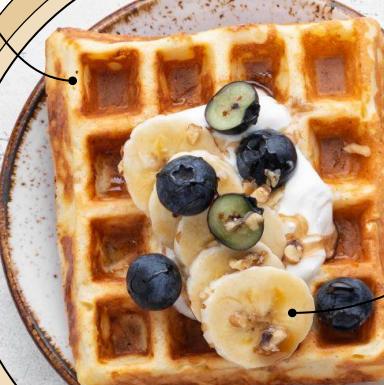
# Ensemble: Random Forest

## Kỹ thuật

là một kỹ thuật tổng hợp có khả năng thực hiện hồi quy và phân loại bằng việc sử dụng nhiều cây quyết định.

## Ý tưởng

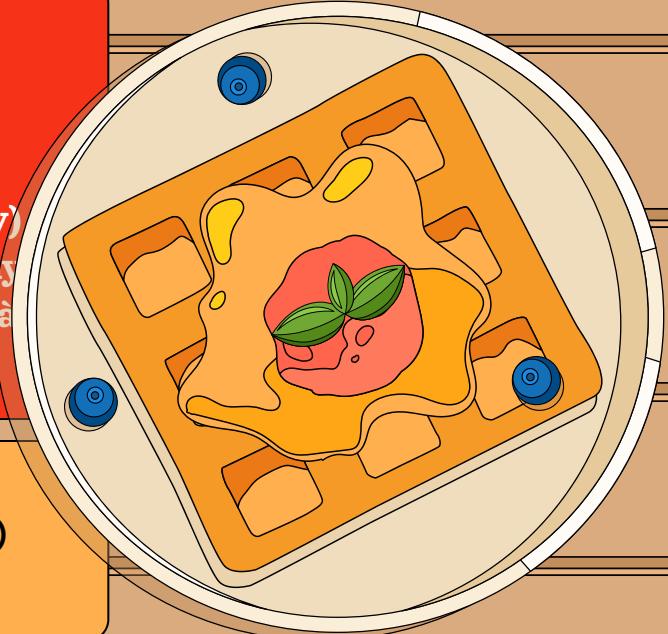
Kết hợp nhiều cây quyết định để xác định đầu ra cuối cùng thay vì dựa vào các cây quyết định riêng lẻ.



## ***Linear Regression-Hồi quy tuyến tính***

Thực hiện dự đoán một giá trị biến phụ thuộc ( $y$ ) dựa trên biến độc lập cho trước ( $x$ ). Kĩ thuật này nhằm tìm ra mối quan hệ tuyến tính giữa đầu vào và đầu ra  $y$ .

```
sklearn.linear_model.LinearRegression()
```



# Chỉ số đánh giá mô hình

MSE

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Mean Squared Error:  
Phương sai

RMSE

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Root Mean Squared  
Error: Độ lệch  
chuẩn

R-squared

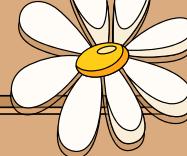
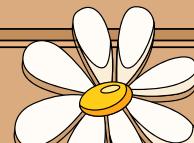
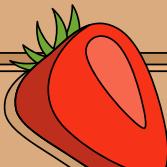
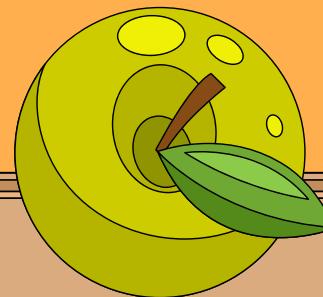
$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Hệ số tương quan R  
bình phương

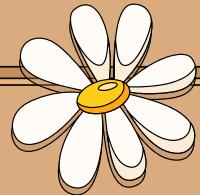
# Mô hình

2.

Sau đây là quá trình huấn luyện  
mô hình.



# Description



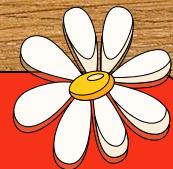
Dữ liệu được dùng để huấn luyện cho mô hình là một tập dữ liệu chứa sản lượng bán ra các mặt hàng thực phẩm của một cửa hàng có nhiều chi nhánh. Chi tiết của bộ dữ liệu như sau:

ID	Week	Center ID	Meal ID	Checkout Price	Base Price	Number Order
1379560.0	1.0	55.0	1885.0	136.83	152.29	177.0
1466964.0	1.0	55.0	1993.0	136.83	135.83	270.0
1346989.0	1.0	55.0	2539.0	134.86	135.86	189.0





# Feature Selection



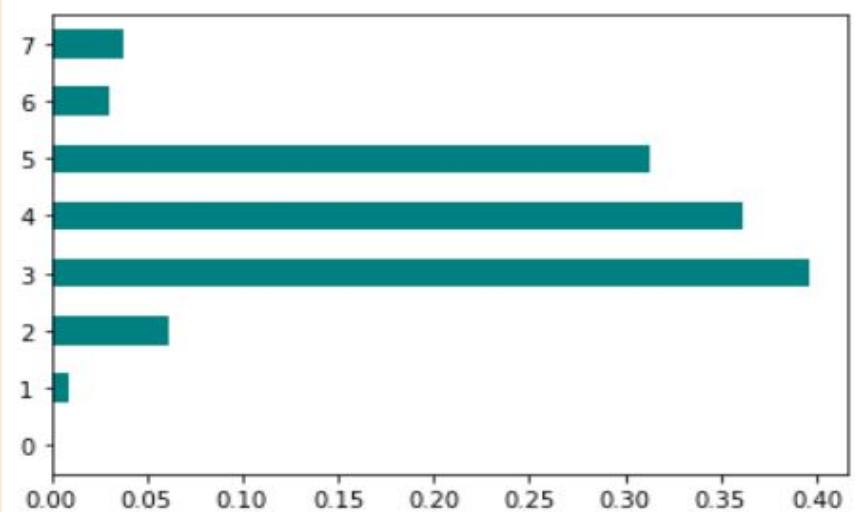
## Feature Selection

Ta sẽ sử dụng hàm **mutual\_info\_regression** để có thể thấy được mức độ liên quan giữa các nhân tố với biến mục tiêu và thực hiện thể hiện bằng đồ thị.

# Code

```
from sklearn.feature_selection import  
mutual_info_regression  
import matplotlib.pyplot as plt  
%matplotlib inline  
  
importances= mutual_info_regression(x_train,y_train)  
featureimp=pd.Series(importances)  
featureimp.plot(kind='barh',color="teal")  
plt.show()
```

# Results



**Remove**

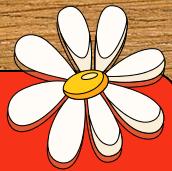
Loại 0,1,6,7 vì có chỉ số khá thấp.

**Select**

Chọn 2,3,4,5



# Training



# Train Model

**Chỉ số đánh giá**

MSE

RMSE

R-Squared

**Models**

Linear  
Regression

Decision  
Tree

Random  
Forest

**Mục đích**

Thực hiện huấn luyện  
trên 3 **mô hình** khác  
nhau và đánh giá dựa  
trên các **chỉ số đánh  
giá**.

# Results

	Linear Regression	Decision Tree	Random Forest
r2_score	0.1113	0.1649	0.1746
MSE	128151.8484	120422.5630	119015.7227
RMSE	357.983	347.0195	344.986

# Đánh giá

3.

Sau đây là ý nghĩa cũng như  
hướng phát triển của mô hình  
trong tương lai

## Ý nghĩa

Việc đưa ra dự đoán được nhu cầu của các mặt hàng nói chung và của các thực phẩm nói riêng tại các khu vực khác nhau sẽ đem lại lợi ích kinh tế cho các doanh nghiệp, khi đó sẽ cung ứng vừa đủ các mặt hàng đến từng khu vực, hạn chế việc di chuyển nhiều lần giữa các khu vực để cung cấp các mặt hàng, tiết kiệm được thời gian cũng như chi phí di chuyển.

# HƯỚNG PHÁT TRIỂN

Trong tương lai, có thể sẽ sử dụng thêm nhiều mô hình học kết hợp khác để xây dựng mô hình, đồng thời thu thập thêm nhiều dữ liệu có giá trị liên quan đến đề tài.

# Web

4.

Nhấn **vào đây** để đến trang web  
trình bày sản phẩm.

# Team 9

Lê Quốc Huy

Thành viên

Lưu Quốc Hưng  
Thịnh

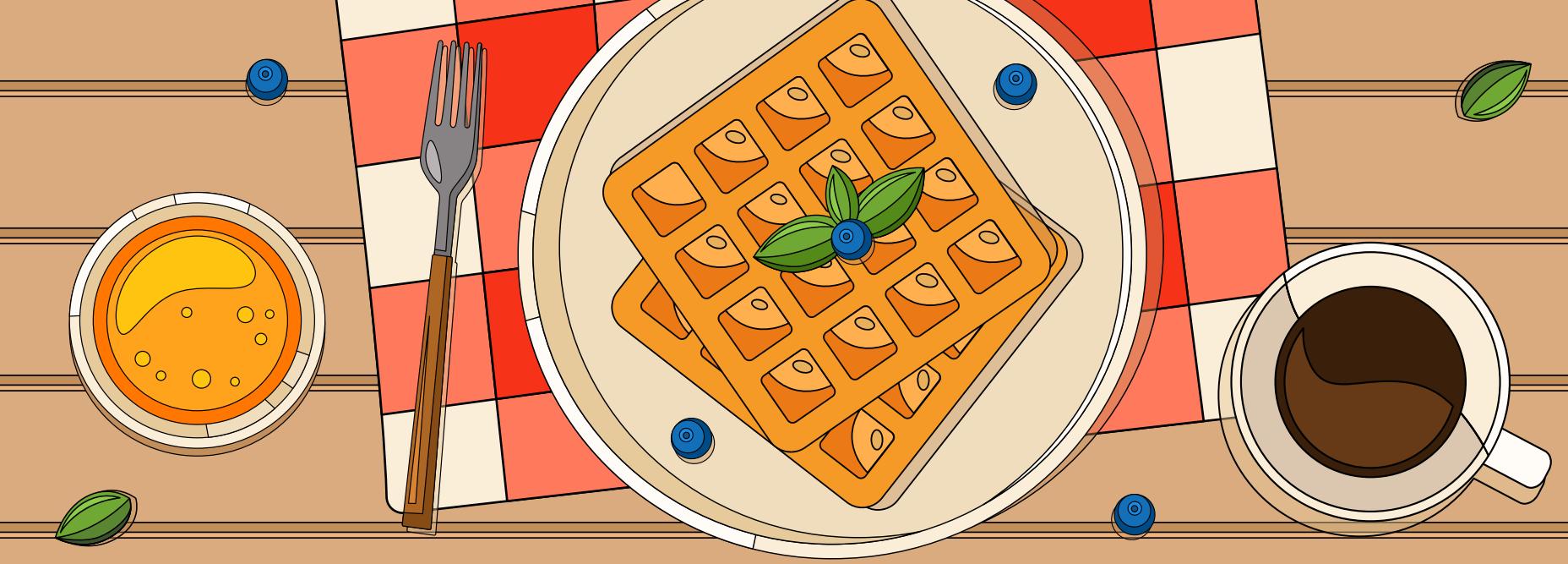
Thành viên

Đoàn Thảo Nhi

Trưởng  
nhóm

Trương Hoàng  
Nguyên Vũ

Thành viên



**Thank you for your attention!**