

# 肥料種類預測

大數據分析與應用 期末報告



*Nurturing Growth,  
Harvesting Prosperity!*

Presented by

4110029009 王致雅

4110029017 戴君芮

指導老師：蔡孟勳教授



# Contents

使用資料集

02

資料集 | 欄位介紹

SDGs

05

SDG 2 | 12 | 13 | 15

研究動機 & 目的

06

優化農業生產和肥料管理

魚骨圖

08

使用肥料的因素和策略

研究方法

10

EDA、特徵選取、建模

文獻探討

27

合理化施肥的意義與措施

結果與討論

28

選擇肥料最大的關鍵

結論&參考資料

29

針對性補充缺少元素的肥料

名稱 : Fertilizer Prediction

資料量 : 共99筆資料

來源 : <https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction>

欄位 : 共9種特徵，包括6種數值型，3種類別型

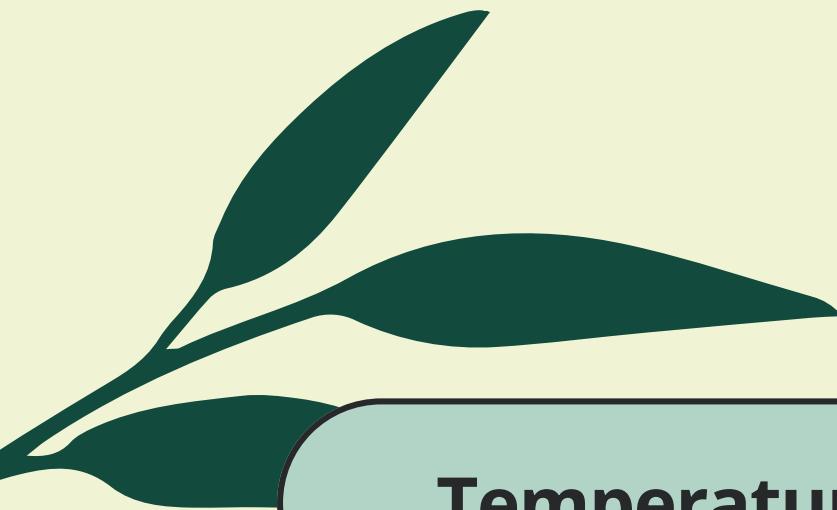
數值型

Temparature | Humidity | Moisture | Nitrogen | Potassium | Phosphorous

類別型

Soil Type | Crop Type | Fertilizer Name

# Introduction



# Numeric features

**Temperature**

土壤的溫度，分布於25°C~38°C

**Humidity**

土壤相對濕度，土壤中空氣中的水分含量的百分比，分佈於50%~72%

**Moisture**

土壤濕度（土壤水分含量），分布於25%~65%

**Nitrogen**

土壤中的氮素含量，分布於 4ppm~42ppm

**Potassium**

土壤中的鉀含量，分布於 0ppm~19ppm

**Phosphorous**

土壤中的磷含量，分布於 0ppm~42ppm



# Categorical features

## Soil Type

土壤的種類，共5種:

['Black' 'Clayey' 'Loamy' 'Red' 'Sandy']

## Crop Type

種植的作物種類，共11種:

['Barley' 'Cotton' 'Ground Nuts' 'Maize' 'Millets'  
'Oil seeds' 'Paddy' 'Pulses' 'Sugarcane' 'Tobacco' 'Wheat']

## Fertilizer Name

使用的肥料種類，共7種:

['10-26-26' '14-35-14' '17-17-17' '20-20' '28-28' 'DAP' 'Urea']

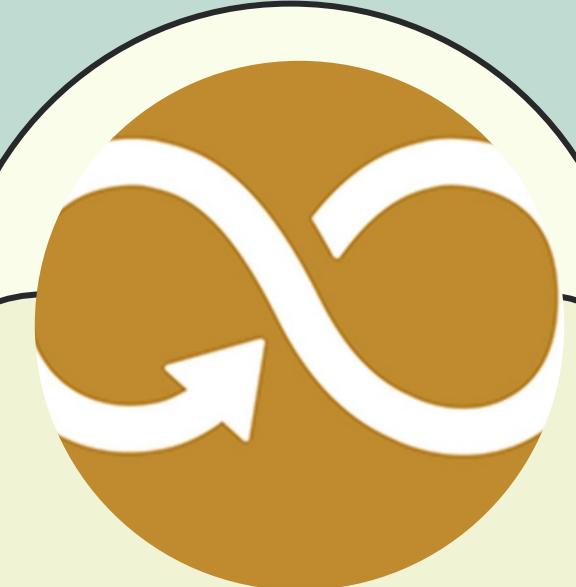
# SDGs

••• A PATH TO ACHIEVE •••



## 目標 2

消除飢餓：  
確保糧食安全，消除飢餓，促進永續農業



## 目標 12

責任消費及生產：  
促進綠色經濟，確保永續消費及生產模式



## 目標 13

氣候行動：  
完備減緩調適行動，以因應氣候變遷及其影響



## 目標 15

保育陸域生態：  
保育及永續利用陸域生態系，確保生物多樣性並防止土地劣化

## MOTIVATION \*

# Fertilizer Management

- 1. 提高農業生產力：**農業是全球糧食供應的關鍵，因此提高農業生產力是減少糧食短缺的重要因素。透過肥料預測，可以優化肥料使用，從而提高農作物產量。
- 2. 減少農業成本：**不正確的肥料使用可能導致資源浪費和不必要的成本。通過精確的肥料預測，可以節省資金，減少浪費。
- 3. 提高可持續性：**適當的肥料管理有助於減少對環境的不良影響，提高農業的可持續性。



# Objectives

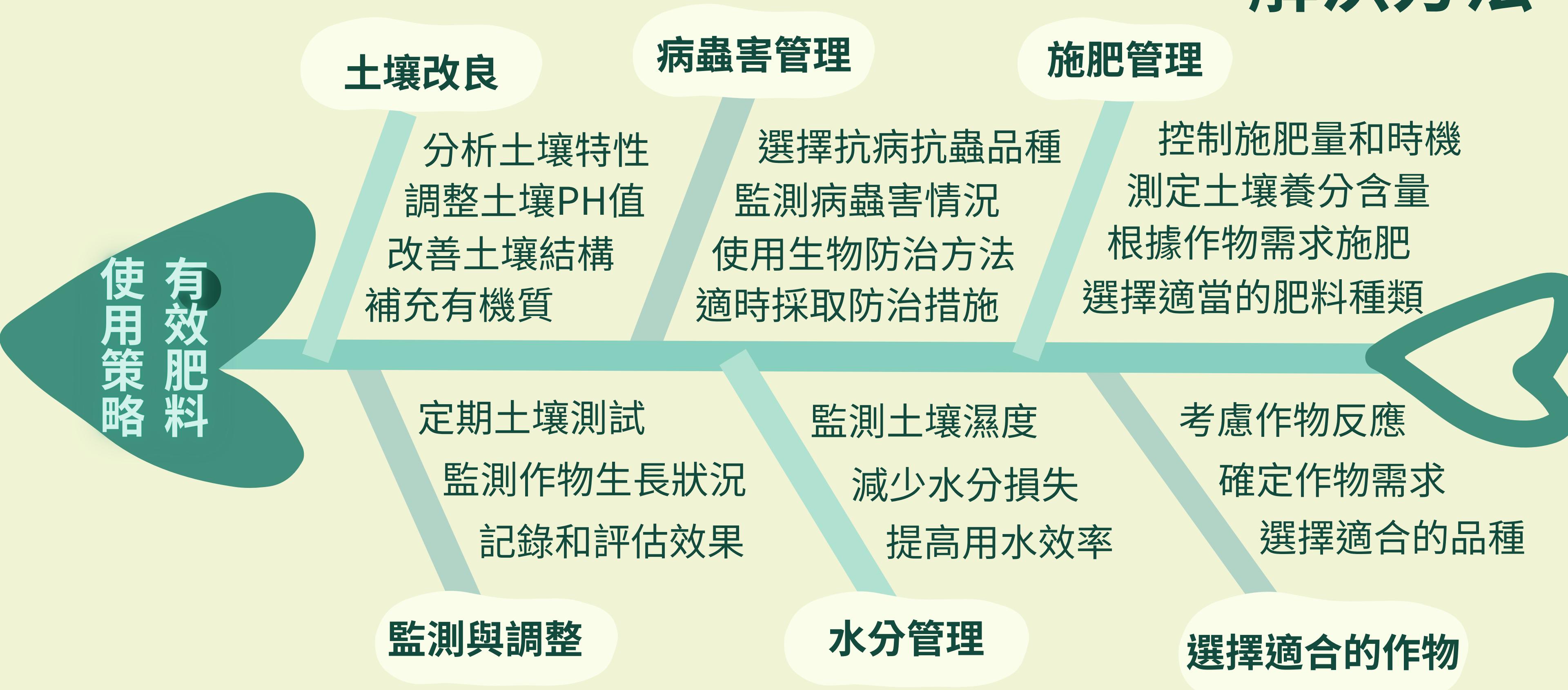
- 開發預測模型**：建立機器學習模型，使用土壤特徵（溫度、濕度、濕度、氮素、鉀、磷等）以及土壤類型、作物類型等信息，以預測最適合的肥料種類。
- 優化肥料使用**：幫助農民和農業專業人員選擇適當的肥料，從而提高農作物的生長和產量，同時減少浪費。
- 提高農業效益**：通過肥料預測，提高農業效益，實現更有效的農業生產，滿足不斷增長的人口需求。



# 影響使用肥料的因素



# 解決方法



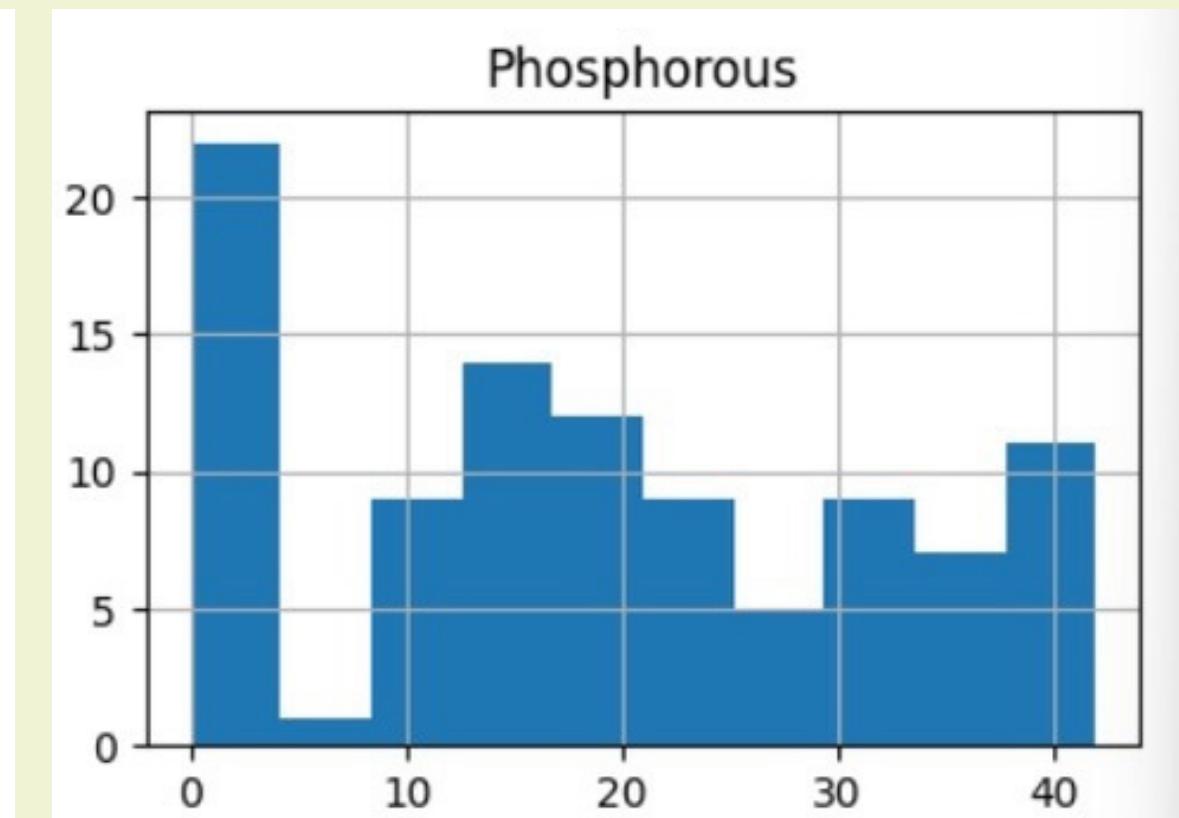
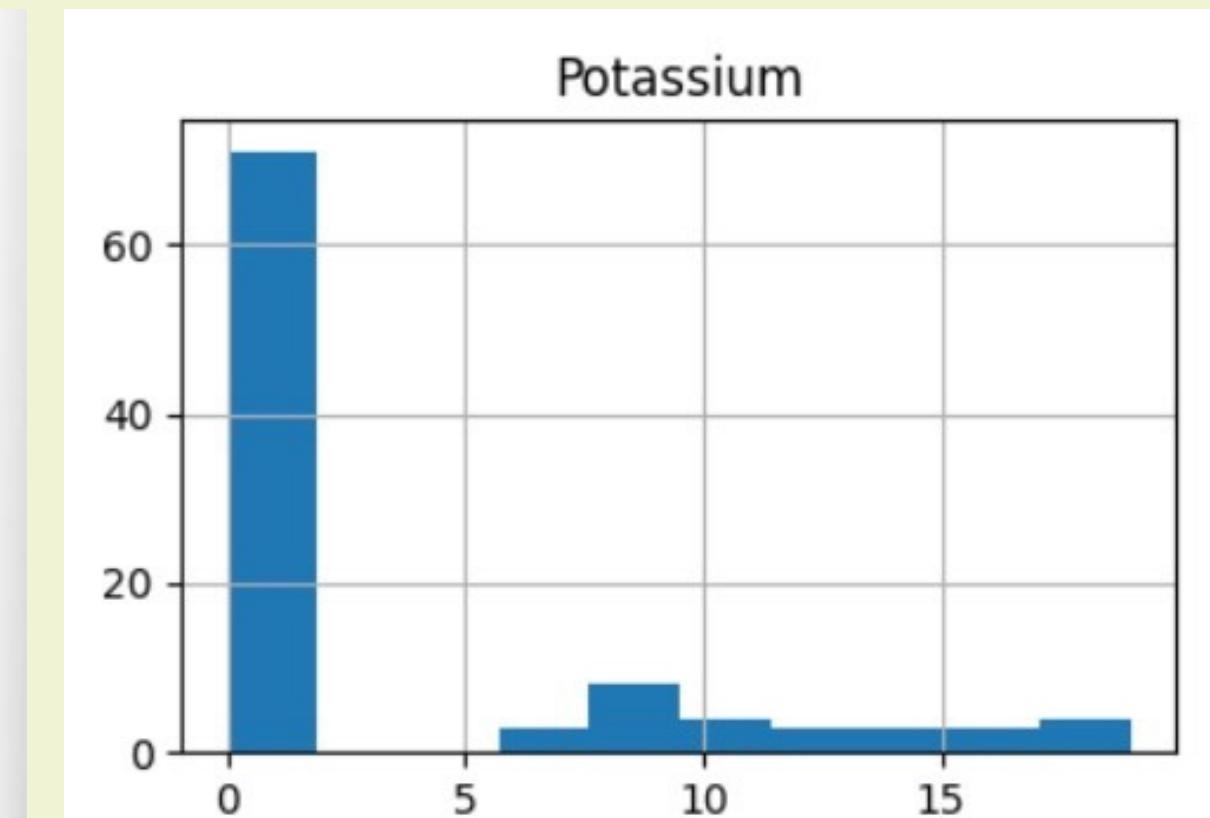
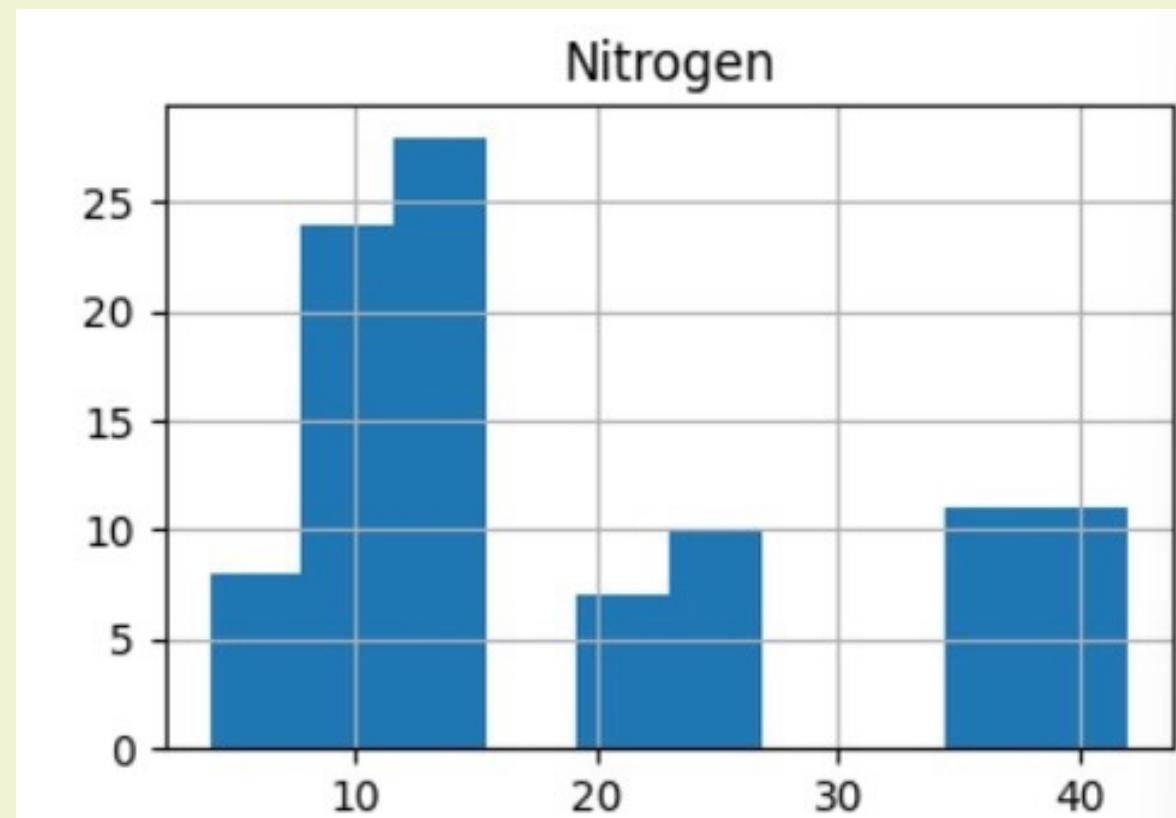
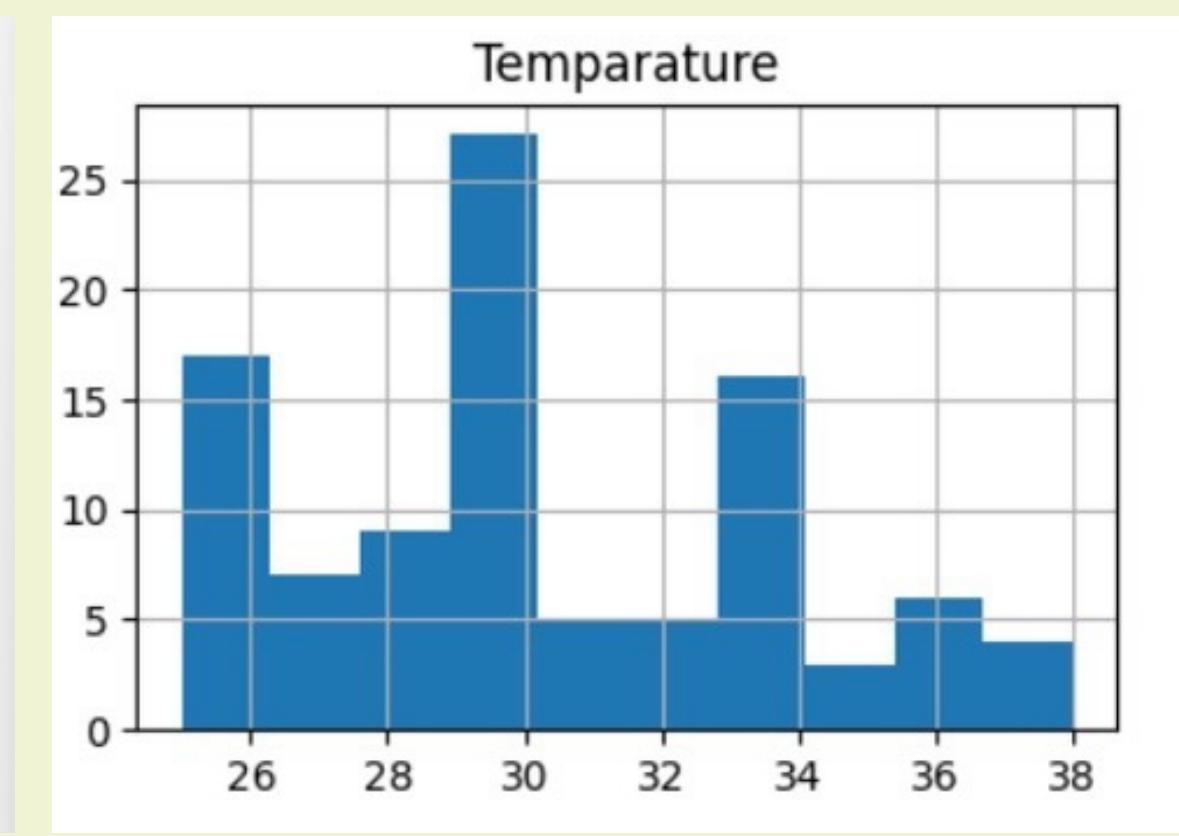
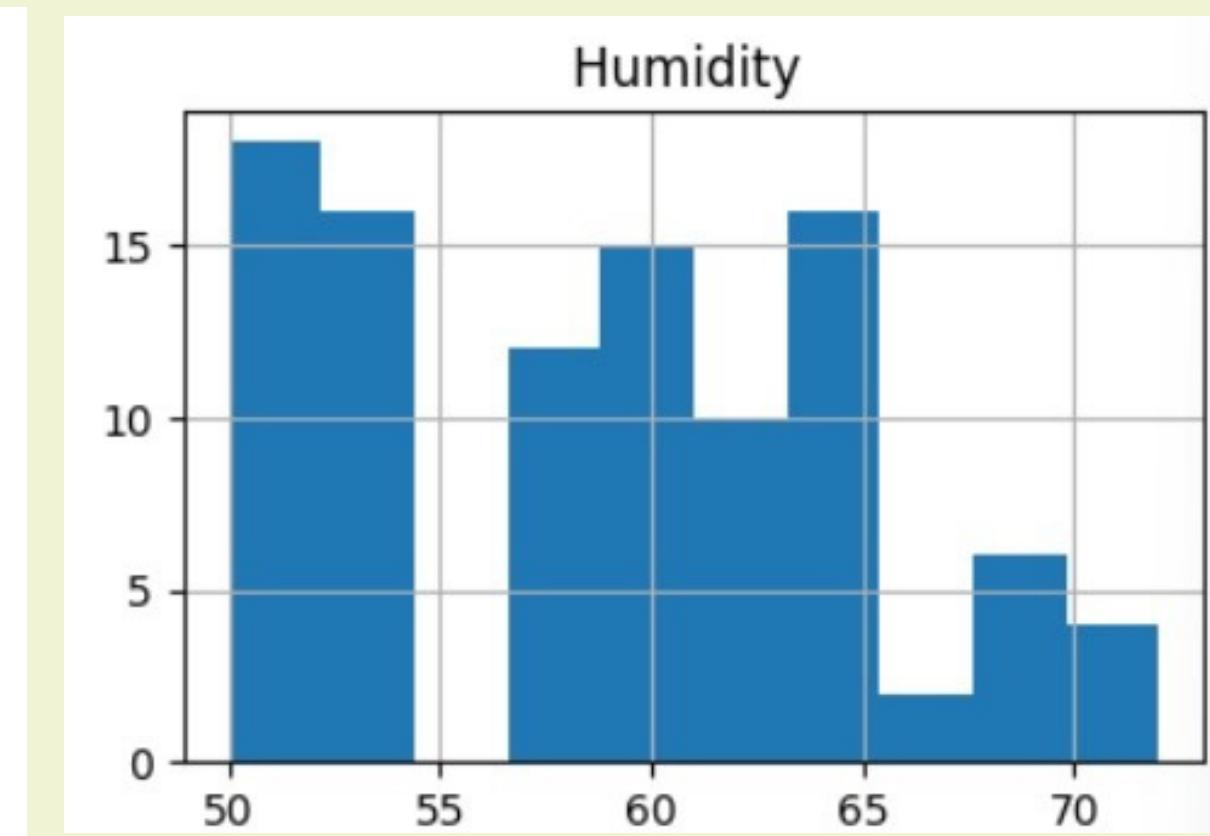
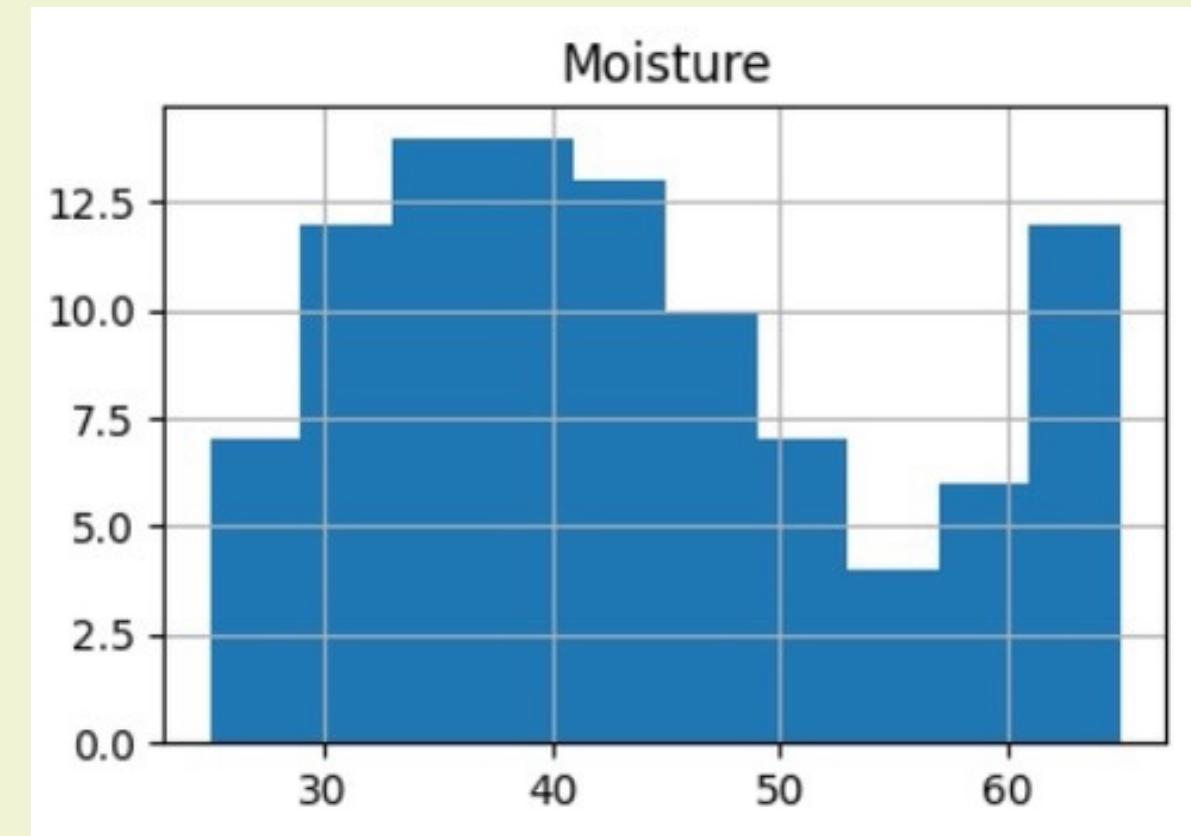
- **理解資料集：**
  - 深入了解數據的背景和特性。
  - 確保對數據中的缺失值、離群值和不一致性進行處理。
- **探索性數據分析 (EDA)：**
  - 使用統計和視覺化工具進行EDA。
  - 理解特徵之間的相關性。
- **相關性分析：**
  - 計算特徵之間的相關性，了解它們對目標變量的影響。
  - 可以使用相關矩陣、熱圖等方法。
- **SelectKBest：**
  - 利用統計檢驗方法，保留得分最高的K個特徵。
- **建模：**
  - 在決策樹、隨機森林等模型中，查看特徵的重要性分數，進行特徵選擇。

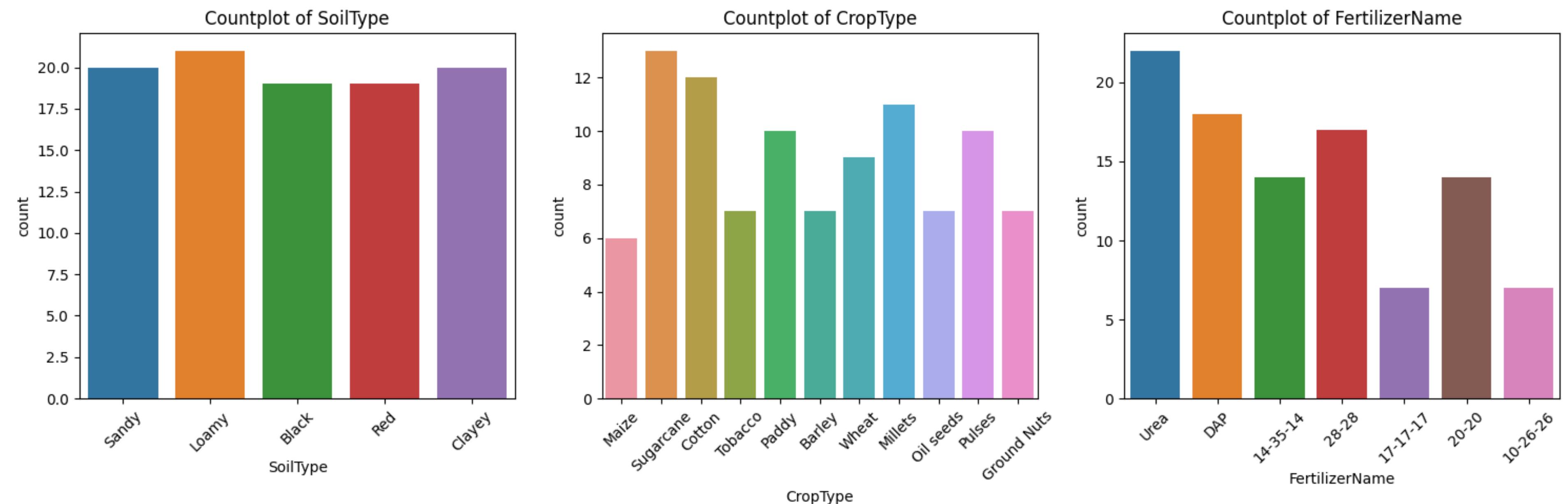
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   Temparature     99 non-null    int64  
 1   Humidity        99 non-null    int64  
 2   Moisture         99 non-null    int64  
 3   Soil Type       99 non-null    object 
 4   Crop Type       99 non-null    object 
 5   Nitrogen         99 non-null    int64  
 6   Potassium        99 non-null    int64  
 7   Phosphorous      99 non-null    int64  
 8   Fertilizer Name 99 non-null    object 
dtypes: int64(6), object(3)
memory usage: 7.1+ KB

```

		Temparature	Humidity	Moisture	Nitrogen	Potassium	Phosphorous
	<b>count</b>	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000
	<b>mean</b>	30.282828	59.151515	43.181818	18.909091	3.383838	18.606061
	<b>std</b>	3.502304	5.840331	11.271568	11.599693	5.814667	13.476978
	<b>min</b>	25.000000	50.000000	25.000000	4.000000	0.000000	0.000000
	<b>25%</b>	28.000000	54.000000	34.000000	10.000000	0.000000	9.000000
	<b>50%</b>	30.000000	60.000000	41.000000	13.000000	0.000000	19.000000
	<b>75%</b>	33.000000	64.000000	50.500000	24.000000	7.500000	30.000000
	<b>max</b>	38.000000	72.000000	65.000000	42.000000	19.000000	42.000000





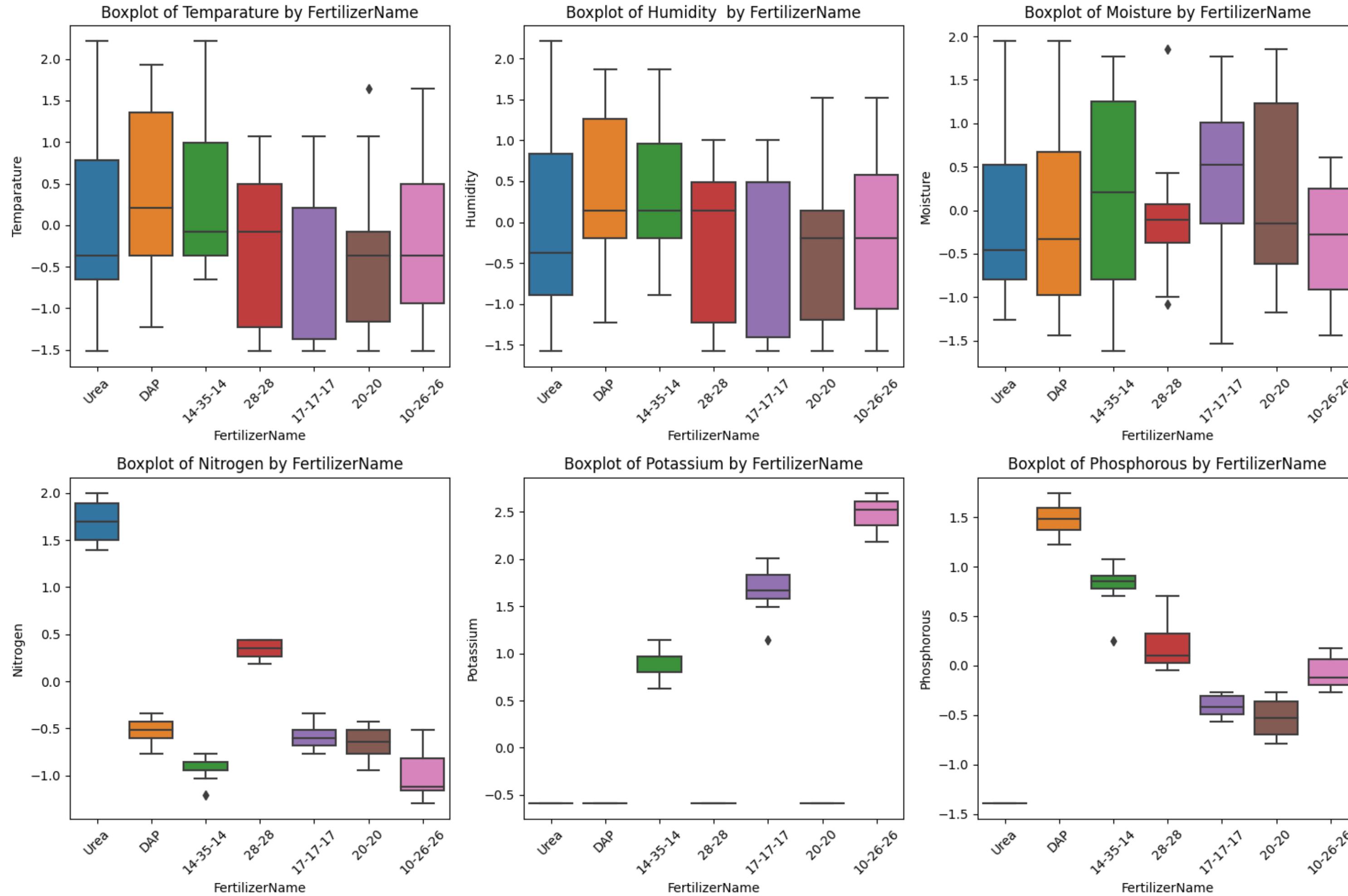
Loamy	21
Sandy	20
Clayey	20
Black	19
Red	19

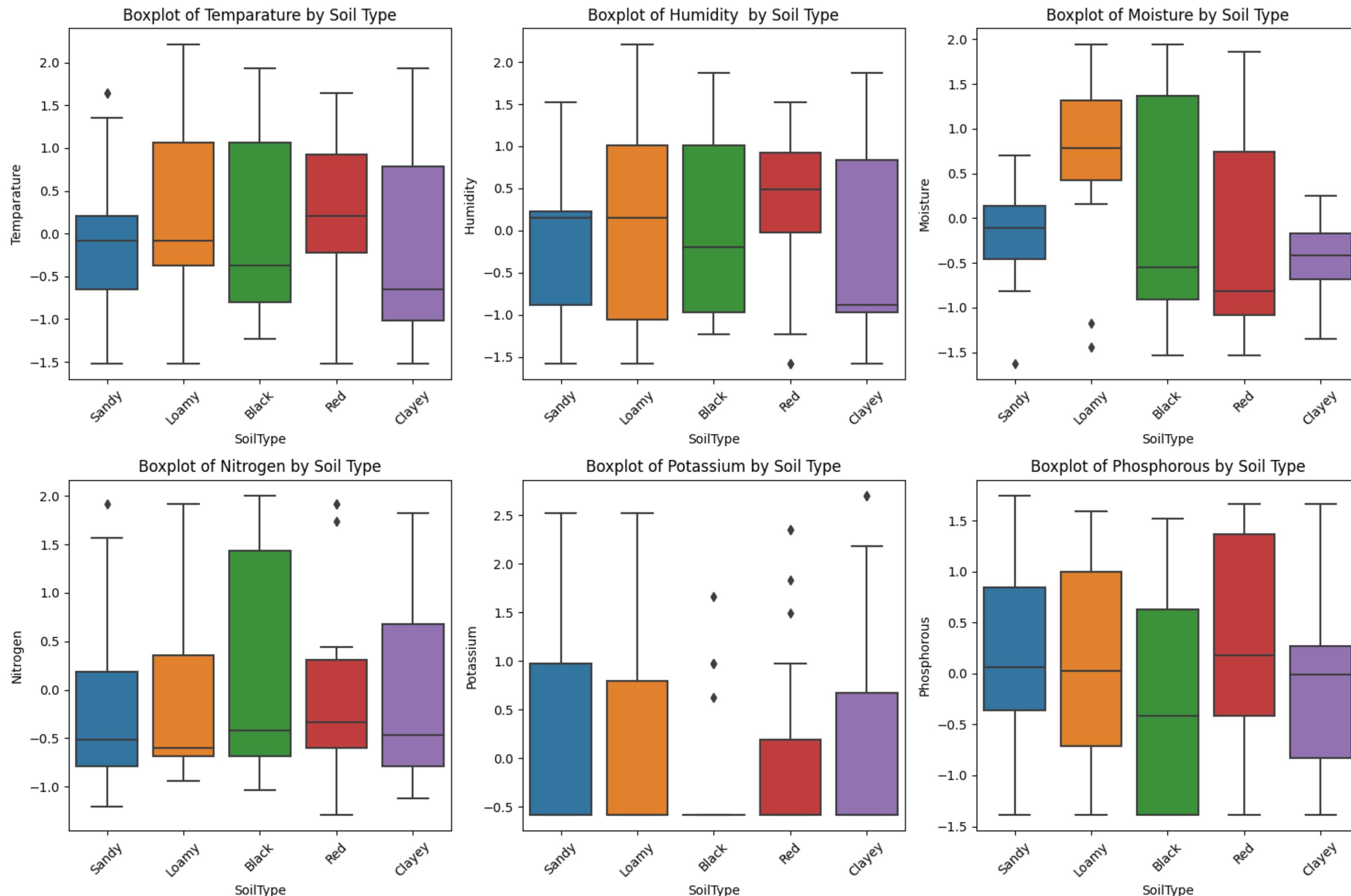
Sugarcane	13
Cotton	12
Millets	11
Paddy	10
Pulses	10
Wheat	9
Tobacco	7
Barley	7
Oil seeds	7
Ground Nuts	7
Maize	6

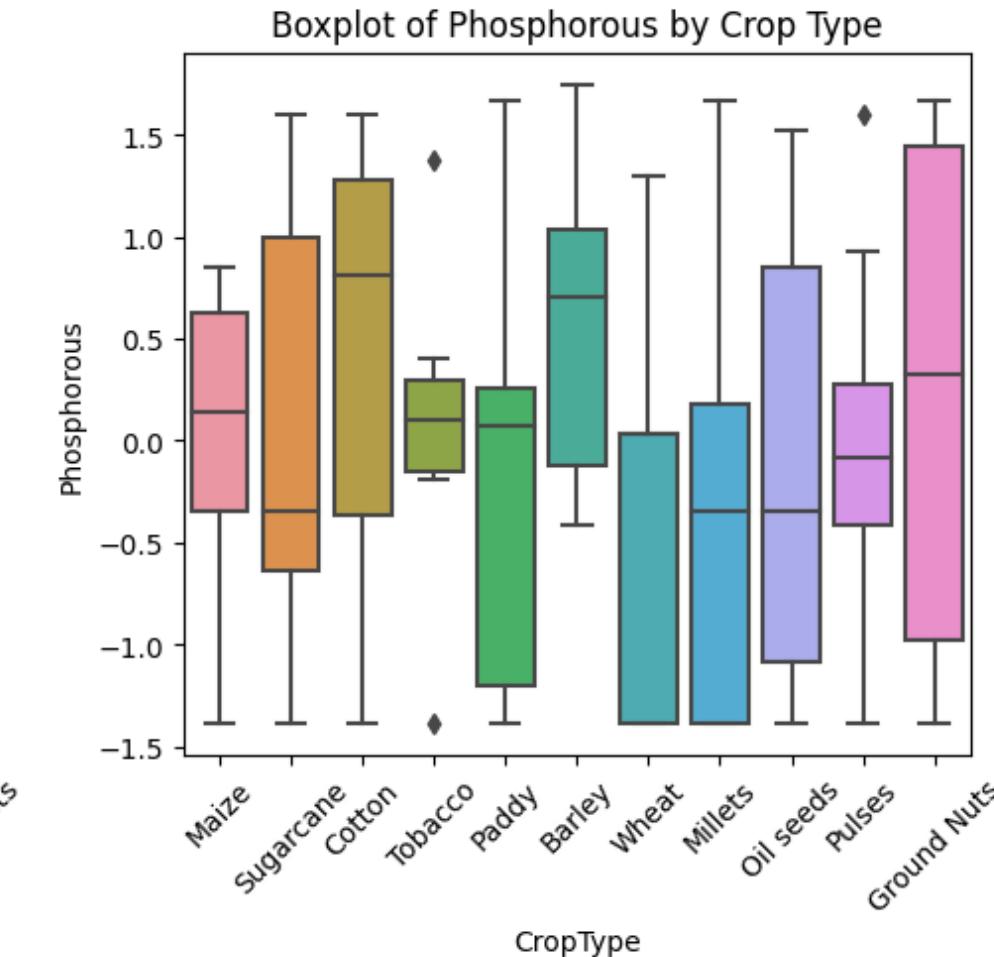
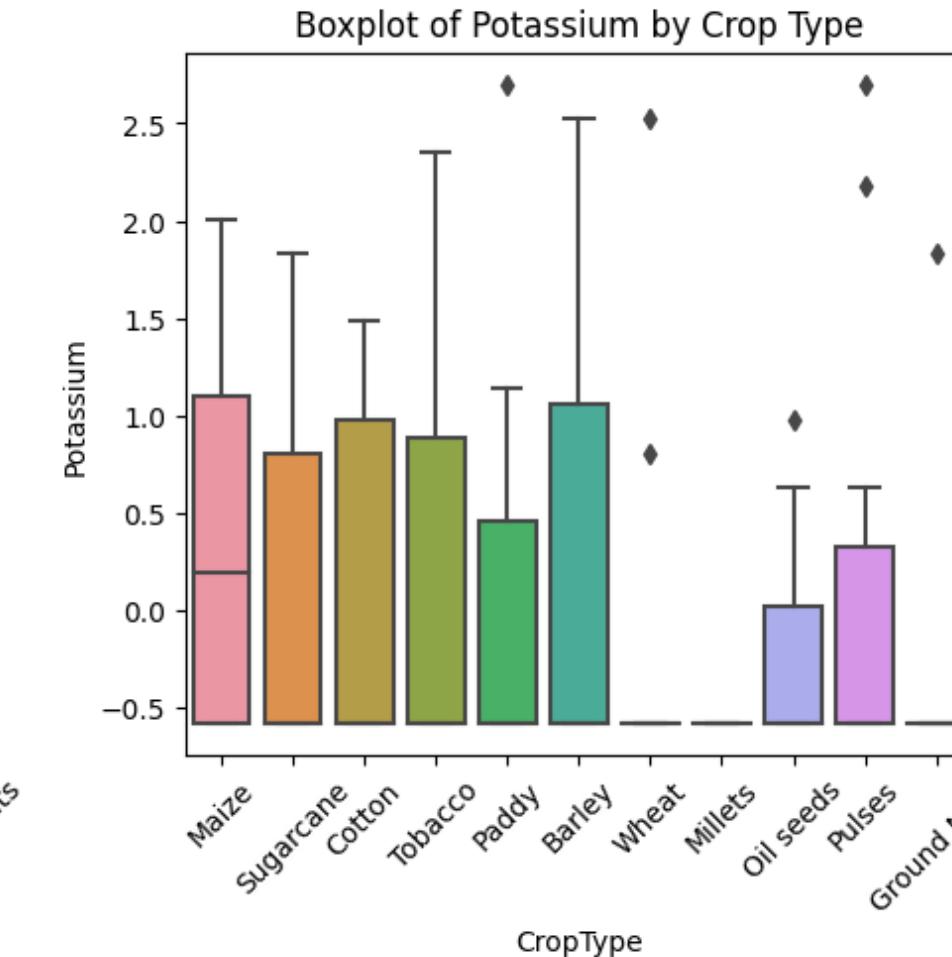
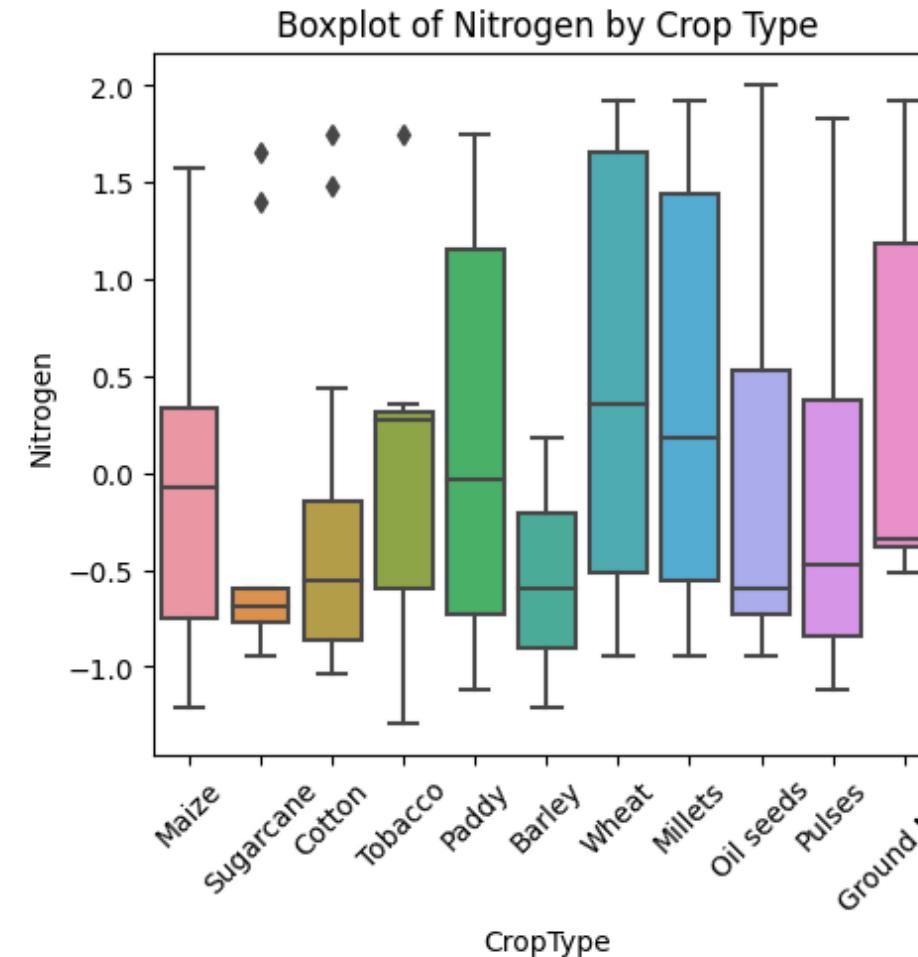
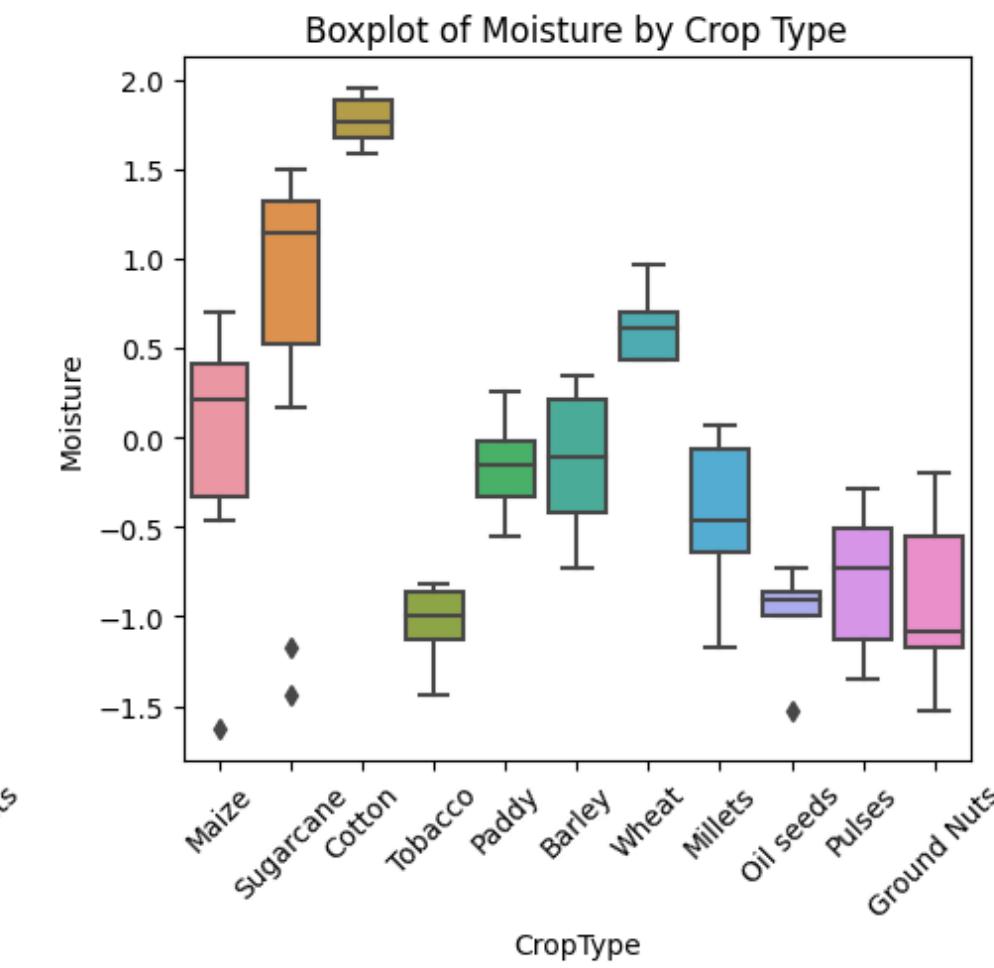
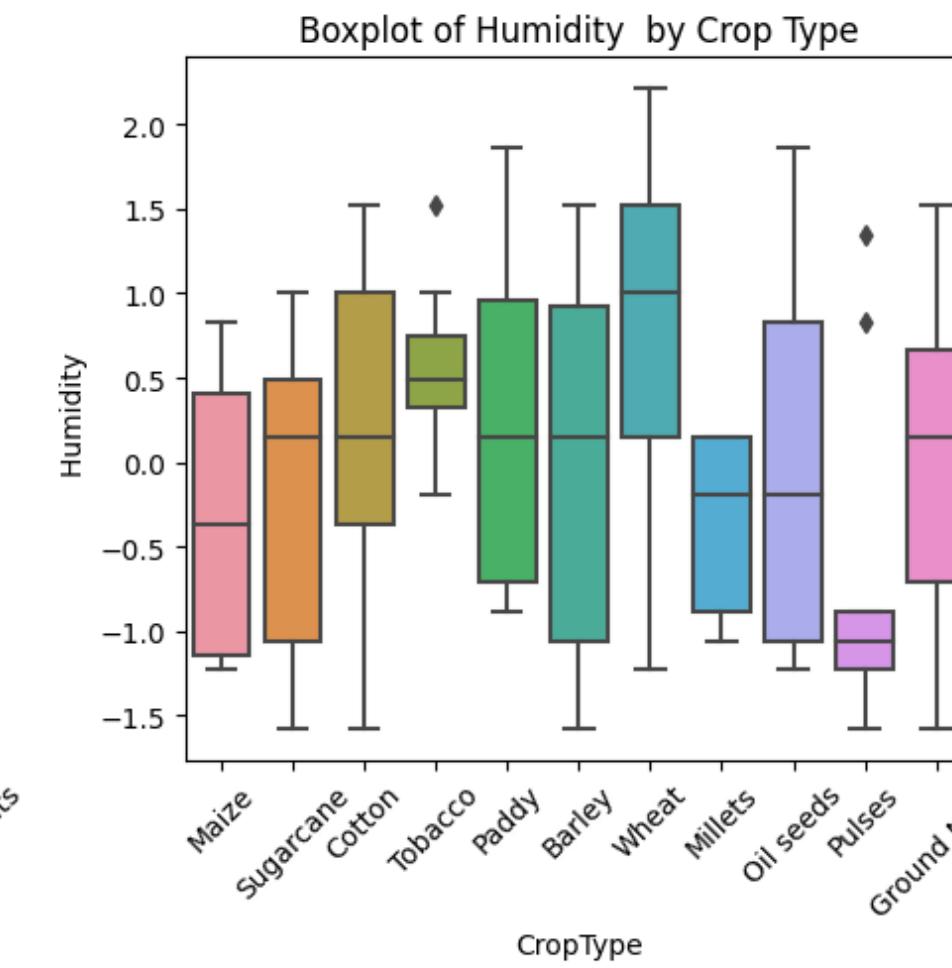
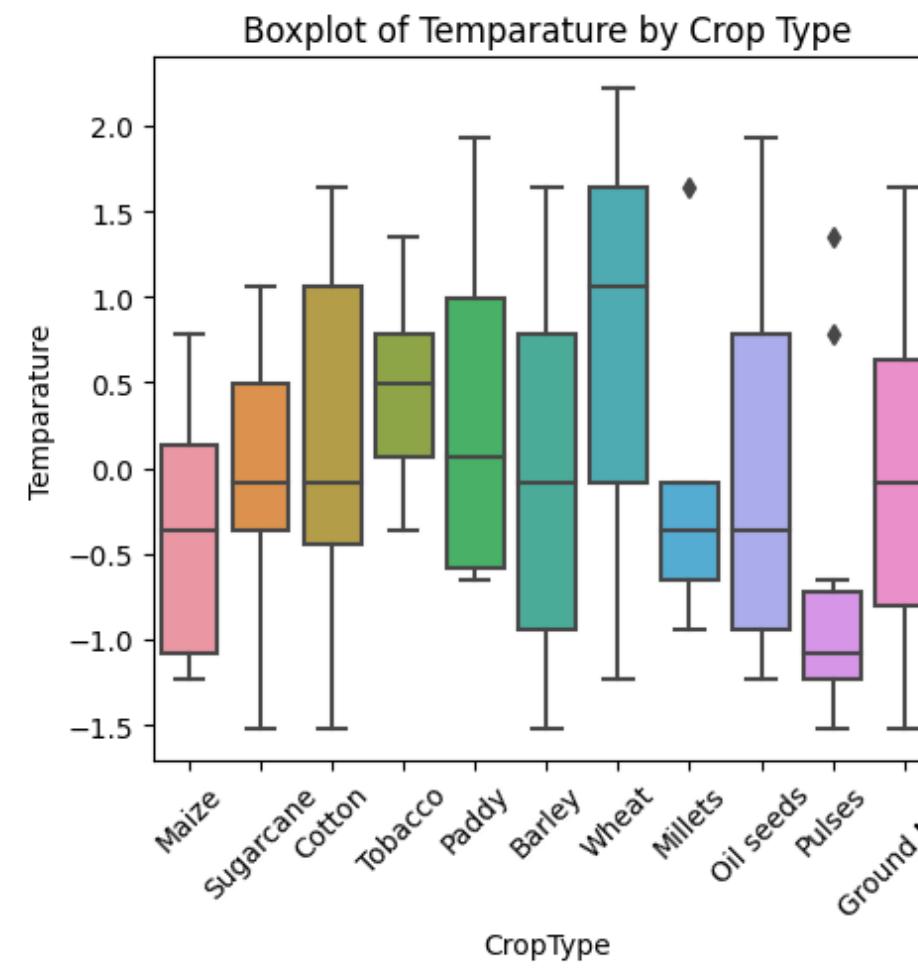
Urea	22
DAP	18
28-28	17
14-35-14	14
20-20	14
17-17-17	7
10-26-26	7

## 連續型特徵標準化

	Temparature	Humidity	Moisture	Nitrogen	Potassium	Phosphorous
0	-1.229084	-1.230737	-0.462064	1.567539	-0.584910	-1.387607
1	-0.368145	-1.230737	0.162128	-0.598658	-0.584910	1.297209
2	1.066752	1.006492	1.678023	-1.031898	0.970777	0.849740
3	0.492793	0.490209	-0.818745	0.267821	-0.584910	0.103958
4	-0.655125	-0.886548	0.251298	1.394244	-0.584910	-1.387607
...	...	...	...	...	...	...
94	-1.516063	-1.574926	-0.997086	0.441117	-0.584910	0.029379
95	-0.081166	0.146020	-1.442938	-1.291841	2.353610	-0.119777
96	2.214670	2.211154	0.697150	1.740835	-0.584910	-1.387607
97	1.640711	0.146020	-0.016213	-0.338715	-0.584910	1.670100
98	-0.368145	-0.198169	1.232171	-0.598658	-0.584910	-0.641825







## 類別型特徵轉換 : LabelEncoder

	Encoded
Original	
Black	0
Clayey	1
Loamy	2
Red	3
Sandy	4

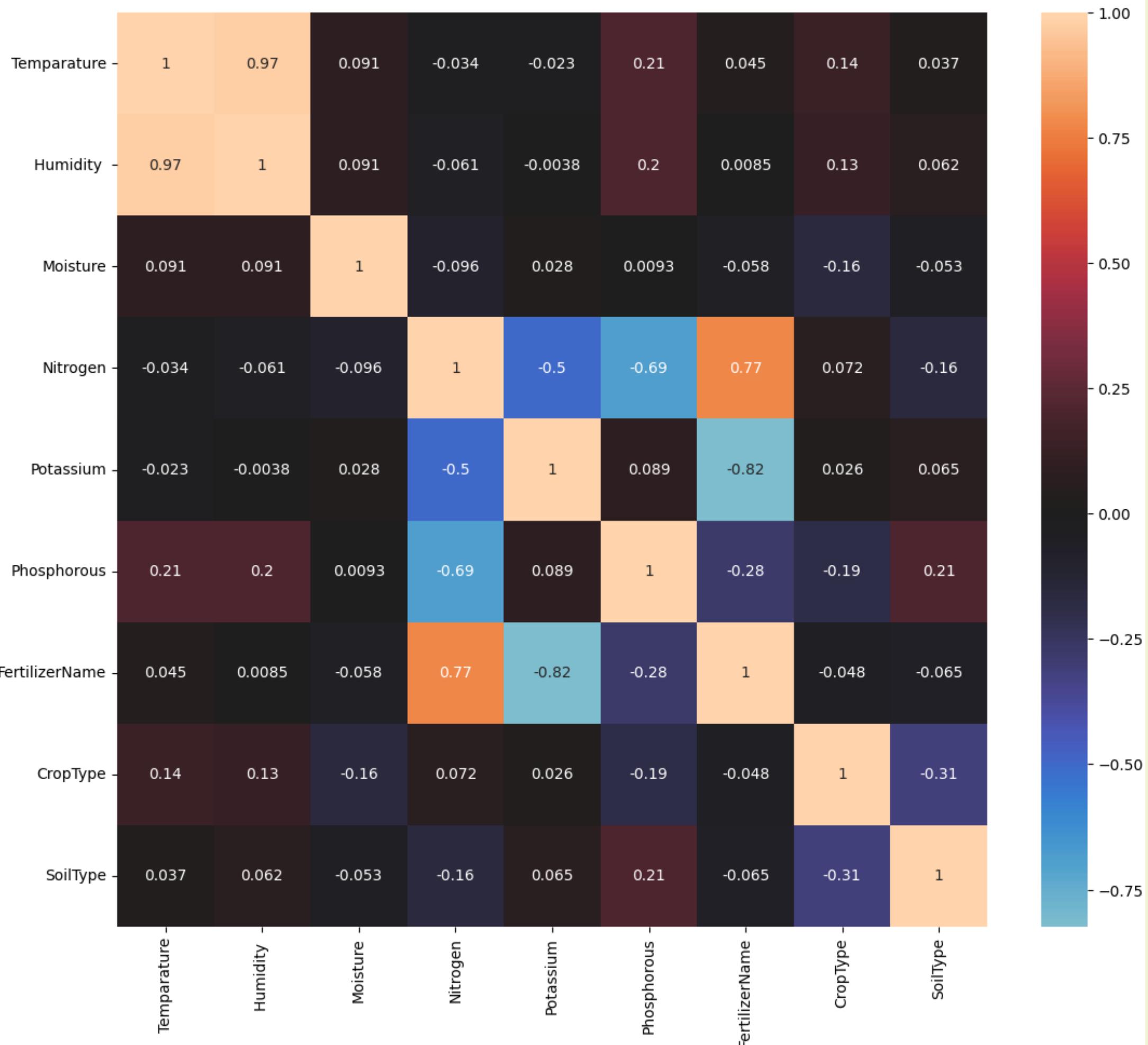
soil type

	Encoded
Original	
Barley	0
Cotton	1
Ground Nuts	2
Maize	3
Millets	4
Oil seeds	5
Paddy	6
Pulses	7
Sugarcane	8
Tobacco	9
Wheat	10

crop type

	Encoded
Original	
10-26-26	0
14-35-14	1
17-17-17	2
20-20	3
28-28	4
DAP	5
Urea	6

fertilizer name



1.溫度、濕度  
2.肥料種類、氮磷鉀



## 特徵選取SelectKBest

k=6

```
Index(['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous', 'CropType',
       'SoilType'],
      dtype='object')
```

k=5

```
Index(['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous', 'SoilType'],
      dtype='object')
```

k=4

```
Index(['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous'],
      dtype='object')
```

k=3

```
Index(['Nitrogen', 'Potassium', 'Phosphorous'],
      dtype='object')
```

## 未標準化特徵

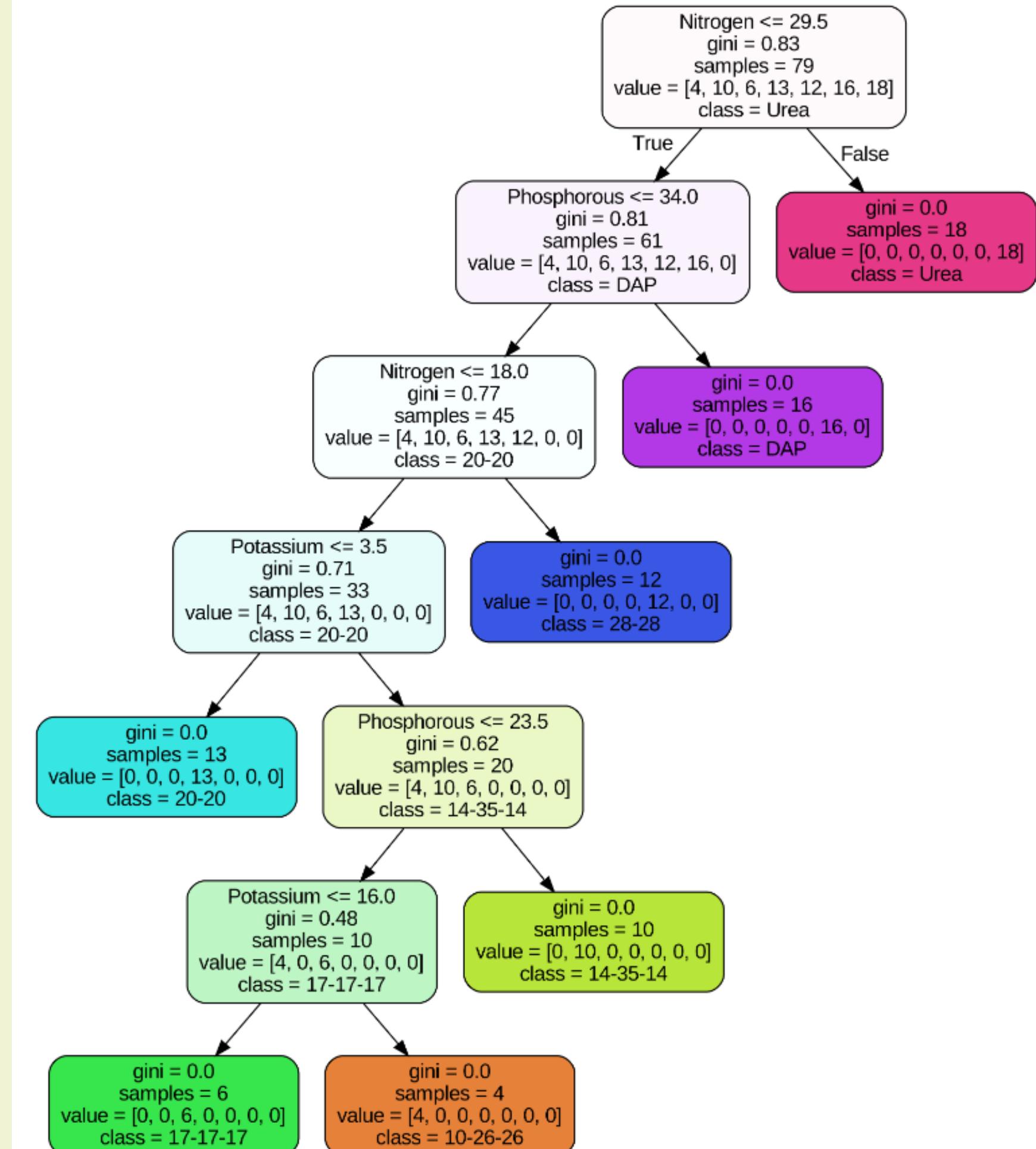
dct

	precision	recall	f1-score	support
10-26-26	1.00	0.67	0.80	3
14-35-14	1.00	0.75	0.86	4
17-17-17	0.33	1.00	0.50	1
20-20	1.00	1.00	1.00	1
28-28	1.00	1.00	1.00	5
DAP	1.00	1.00	1.00	2
Urea	1.00	1.00	1.00	4
accuracy			0.90	20
macro avg	0.90	0.92	0.88	20
weighted avg	0.97	0.90	0.92	20

第一分支：土壤氮含量 nitrogen

第二分支：土壤磷含量 phosphorous

第三分支：土壤鉀含量 potassium

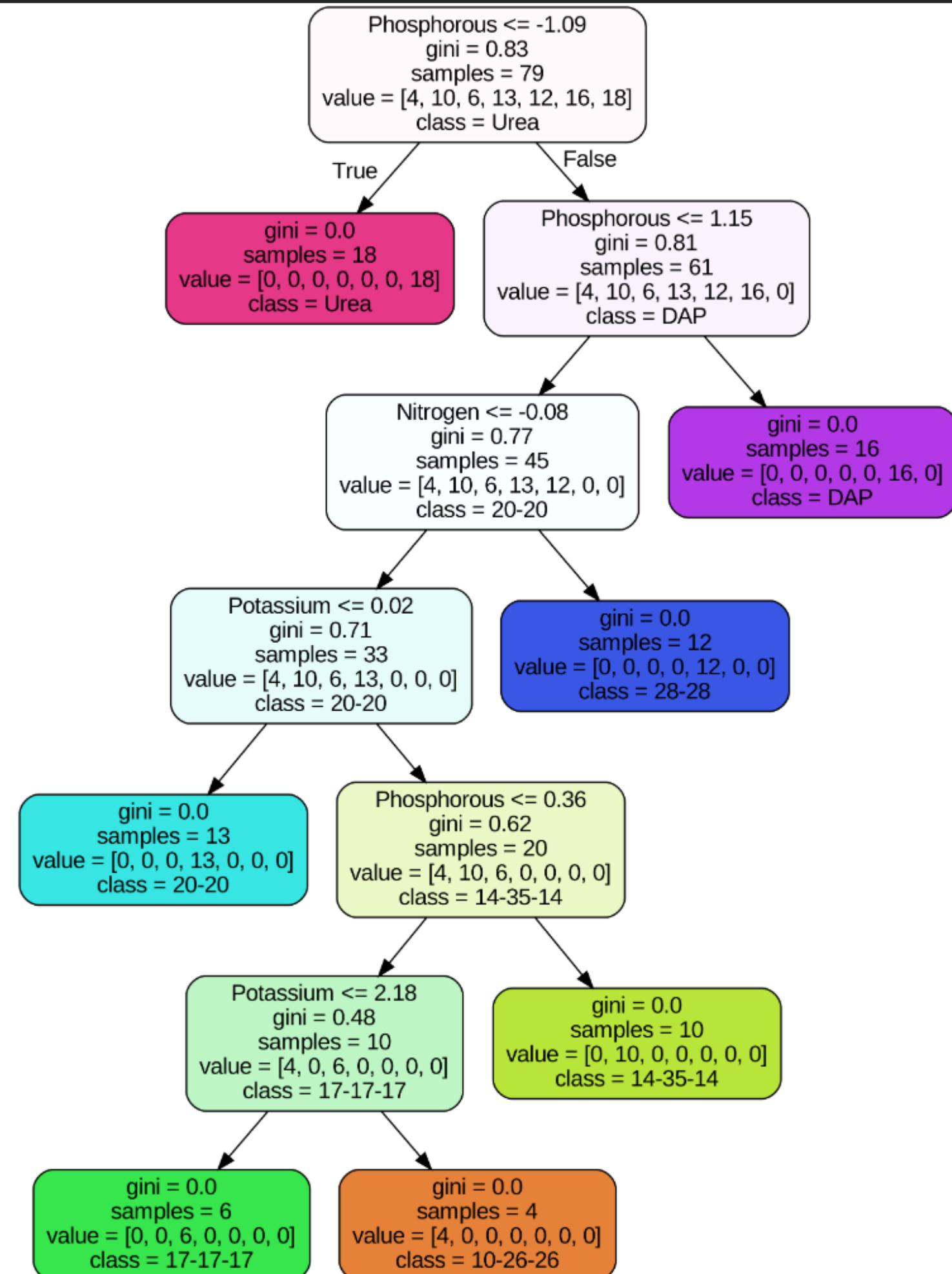


標準化後

dct

	precision	recall	f1-score	support
10-26-26	1.00	1.00	1.00	3
14-35-14	1.00	0.75	0.86	4
17-17-17	0.50	1.00	0.67	1
20-20	1.00	1.00	1.00	1
28-28	1.00	1.00	1.00	5
DAP	1.00	1.00	1.00	2
Urea	1.00	1.00	1.00	4
accuracy			0.95	20
macro avg	0.93	0.96	0.93	20
weighted avg	0.97	0.95	0.95	20

第一分支：土壤磷含量 phosphorous  
 第二分支：土壤氮含量 nitrogen  
 第三分支：土壤鉀含量 potassium



dct

rfc

XGB

## 使用全部特徵

	precision	recall	f1-score	support
10-26-26	1.00	1.00	1.00	3
14-35-14	1.00	0.75	0.86	4
17-17-17	0.50	1.00	0.67	1
20-20	1.00	1.00	1.00	1
28-28	1.00	1.00	1.00	5
DAP	1.00	1.00	1.00	2
Urea	1.00	1.00	1.00	4
accuracy			0.95	20
macro avg	0.93	0.96	0.93	20
weighted avg	0.97	0.95	0.95	20

	precision	recall	f1-score	support
0	0	1.00	0.33	3
1	1	0.67	1.00	4
2	2	1.00	1.00	1
3	3	1.00	1.00	1
4	4	1.00	1.00	5
5	5	1.00	1.00	2
6	6	1.00	1.00	4
accuracy			0.90	20
macro avg	0.95	0.90	0.90	20
weighted avg	0.93	0.90	0.89	20

	precision	recall	f1-score	support
0	0	1.00	0.67	3
1	1	0.80	1.00	4
2	2	1.00	1.00	1
3	3	1.00	1.00	1
4	4	1.00	1.00	5
5	5	1.00	1.00	2
6	6	1.00	1.00	4
accuracy			0.95	20
macro avg	0.97	0.95	0.96	20
weighted avg	0.96	0.95	0.95	20

## 使用 氮磷鉀

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3
1	1.00	0.75	0.86	4
2	0.50	1.00	0.67	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy			0.95	20
macro avg	0.93	0.96	0.93	20
weighted avg	0.97	0.95	0.95	20

	precision	recall	f1-score	support
0	0	1.00	1.00	3
1	1	1.00	1.00	4
2	2	1.00	1.00	1
3	3	1.00	1.00	1
4	4	1.00	1.00	5
5	5	1.00	1.00	2
6	6	1.00	1.00	4
accuracy			1.00	20
macro avg	1.00	1.00	1.00	20
weighted avg	1.00	1.00	1.00	20

	precision	recall	f1-score	support
0	0	1.00	0.67	3
1	1	0.80	1.00	4
2	2	1.00	1.00	1
3	3	1.00	1.00	1
4	4	1.00	1.00	5
5	5	1.00	1.00	2
6	6	1.00	1.00	4
accuracy			0.95	20
macro avg	0.97	0.95	0.96	20
weighted avg	0.96	0.95	0.95	20

knn

## 使用全部特徵

	precision	recall	f1-score	support
0	0.67	0.67	0.67	3
1	1.00	0.50	0.67	4
2	0.00	0.00	0.00	1
3	0.25	1.00	0.40	1
4	0.50	0.20	0.29	5
5	0.33	0.50	0.40	2
6	0.80	1.00	0.89	4
accuracy			0.55	20
macro avg	0.51	0.55	0.47	20
weighted avg	0.63	0.55	0.54	20

['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous', 'SoilType']

	precision	recall	f1-score	support
0	1.00	1.00	0.67	3
1	1.00	1.00	0.75	4
2	0.50	1.00	0.67	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	0.67	1.00	0.80	2
6	1.00	1.00	1.00	4
accuracy			0.90	20
macro avg	0.88	0.92	0.87	20
weighted avg	0.94	0.90	0.90	20

['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous']

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3
1	1.00	1.00	1.00	4
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy			1.00	20
macro avg	1.00	1.00	1.00	20
weighted avg	1.00	1.00	1.00	20

['Nitrogen', 'Potassium', 'Phosphorous']

	precision	recall	f1-score	support
0	1.00	0.67	0.80	3
1	1.00	1.00	1.00	4
2	0.50	1.00	0.67	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy			0.95	20
macro avg	0.93	0.95	0.92	20
weighted avg	0.97	0.95	0.95	20

## SVM

## 使用全部特徵

	precision	recall	f1-score	support
0	0.00	0.00	0.00	3
1	0.25	0.25	0.25	4
2	0.00	0.00	0.00	1
3	0.20	1.00	0.33	1
4	0.00	0.00	0.00	5
5	0.20	0.50	0.29	2
6	1.00	1.00	1.00	4
accuracy			0.35	20
macro avg	0.24	0.39	0.27	20
weighted avg	0.28	0.35	0.30	20

## ['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous', 'SoilType']

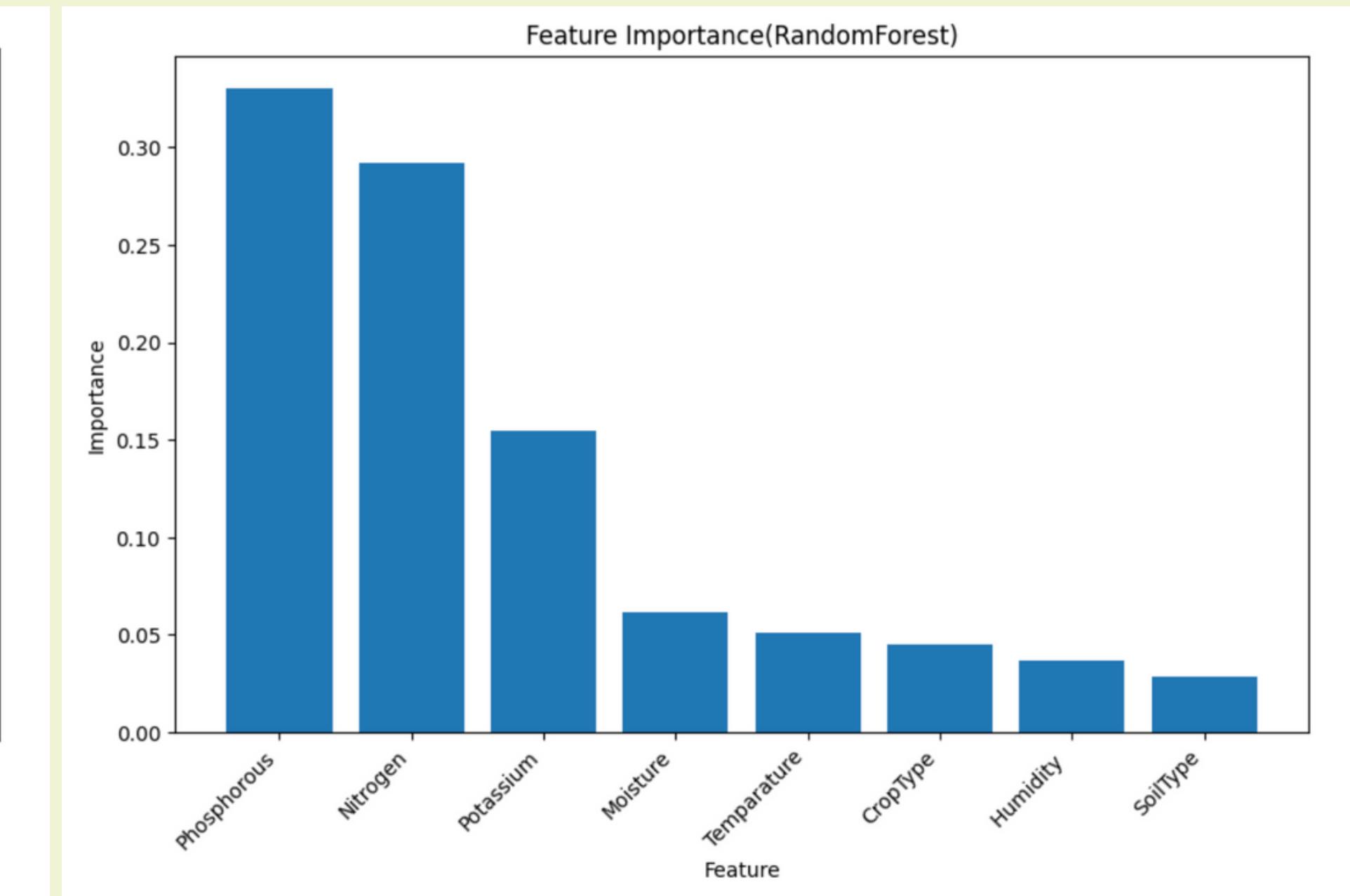
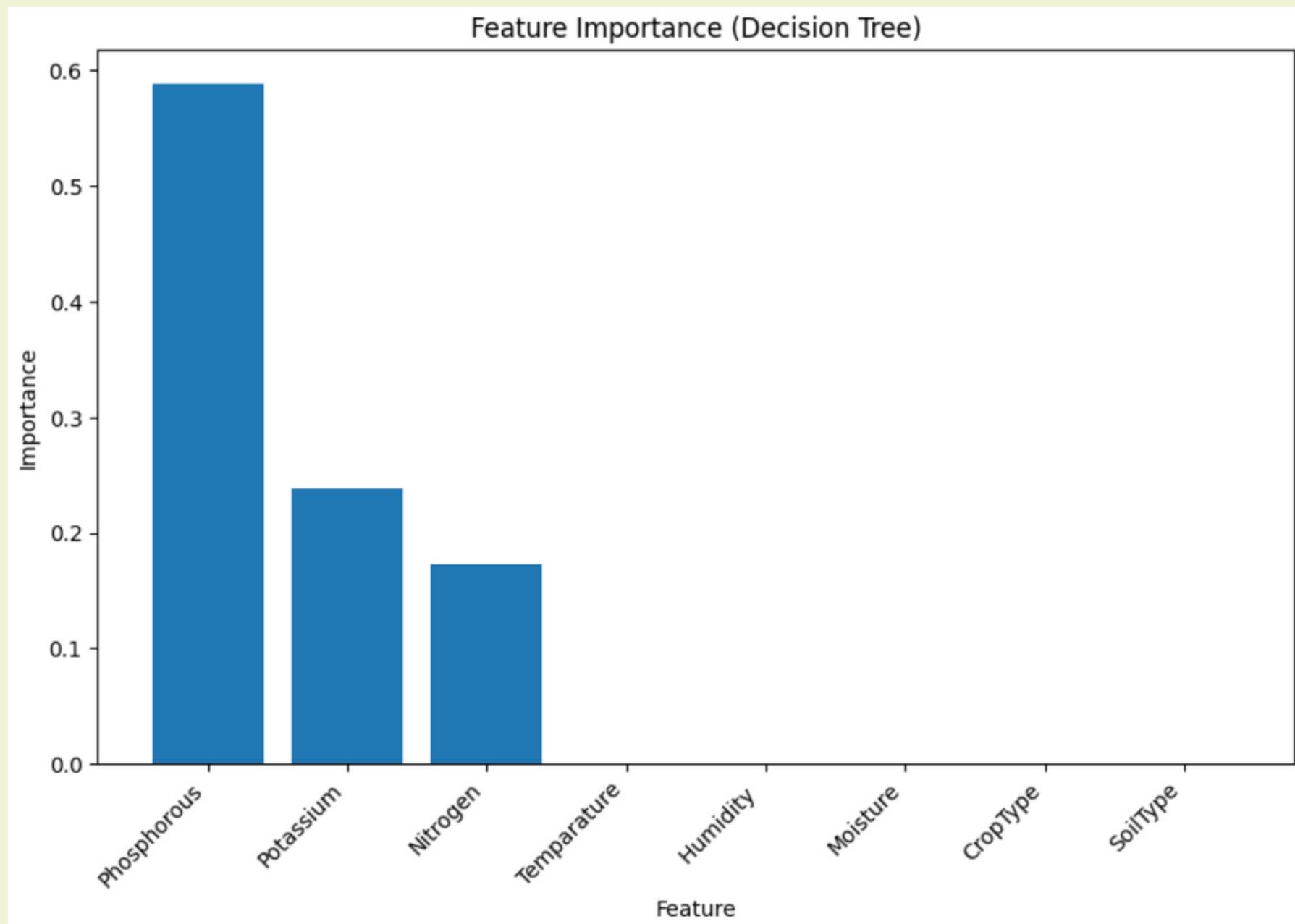
	precision	recall	f1-score	support
0	0.00	0.00	0.00	3
1	1.00	1.00	1.00	4
2	0.25	1.00	0.40	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy				20
macro avg	0.75	0.86	0.77	20
weighted avg	0.81	0.85	0.82	20

## ['Temparature', 'Nitrogen', 'Potassium', 'Phosphorous']

	precision	recall	f1-score	support
0	1.00	0.33	0.50	3
1	1.00	1.00	1.00	4
2	0.33	1.00	0.50	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy			0.90	20
macro avg	0.90	0.90	0.86	20
weighted avg	0.97	0.90	0.90	20

## ['Nitrogen', 'Potassium', 'Phosphorous']

	precision	recall	f1-score	support
0	1.00	0.67	0.80	3
1	1.00	1.00	1.00	4
2	0.50	1.00	0.67	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	4
accuracy			0.95	20
macro avg	0.93	0.95	0.92	20
weighted avg	0.97	0.95	0.95	20



## 何謂合理化施肥？

合理化施肥就是要農民依據農作物品種的生長特性，配合土壤條件與特性，肥料的種類、特性與施入土壤後對作物的生長反應，作為農田施用用量及肥料施用方法的依據，如此才可避免農田肥料施用過多，除了浪費肥料資源外直接影響農作物產品品質與產量，甚至會造成土壤品質劣化，及污染環境如地下水受氮肥污染，使硝酸鹽( $\text{NO}_3^-$ )濃度過高等不良效果。

## 合理化施肥原則

作物的生長過程所吸收必需養分是否吸收充足，是受到最少養分所限制，因此若不補充作物所缺乏要素，即使其他的要素均已供給量充足，作物也無法發揮其生長效果，所以施肥時要判斷土壤中是那一種要素最缺乏，針對最缺的要素施肥才能得到效果。

--->肥料選擇最大的因素是 土壤缺乏的要素

## 肥料的使用

原本以為溫度、濕度或種植作物會是影響最大的因素

經過對資料集的研究

**土壤中「氮」 「磷」 「鉀」 的含量才是選擇肥料最大的關鍵，**

也如同文獻中所說：

肥料選擇最大的因素是「土壤缺乏的要素」



透過深入研究相關資料集，我們發現肥料的使用實際上涉及到更多複雜的考量。

在肥料的使用過程中，首先需要仔細選擇種植的作物。每種作物對土壤的需求各異，因此必須根據目標作物的特性做出明智的選擇。其次，土壤的選擇同樣至關重要。不同的土壤類型擁有獨特的特性，直接影響著肥料的吸收和效果。

根據土壤的特性，有針對性地施用補充缺少元素的肥料。這種個性化的方法確保了肥料的有效利用，最大程度地提高了農作物的生長和產量。

農業生產的整體性需要綜合考慮作物、土壤和肥料之間的相互作用。透過精準的選擇和使用肥料，可以實現更有效的農業生產，同時促進可持續的農業發展。

文獻：<https://www.tcdares.gov.tw/ws.php?id=1733>

資料集：<https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction>

**王致雅**

使用資料集、研究動機、魚骨圖、建模、結論

**戴君芮**

SDGs、魚骨圖、EDA、結果與討論

Thank You