



ชื่อ กิตติพัฒน์ ภิรมย์กิจ

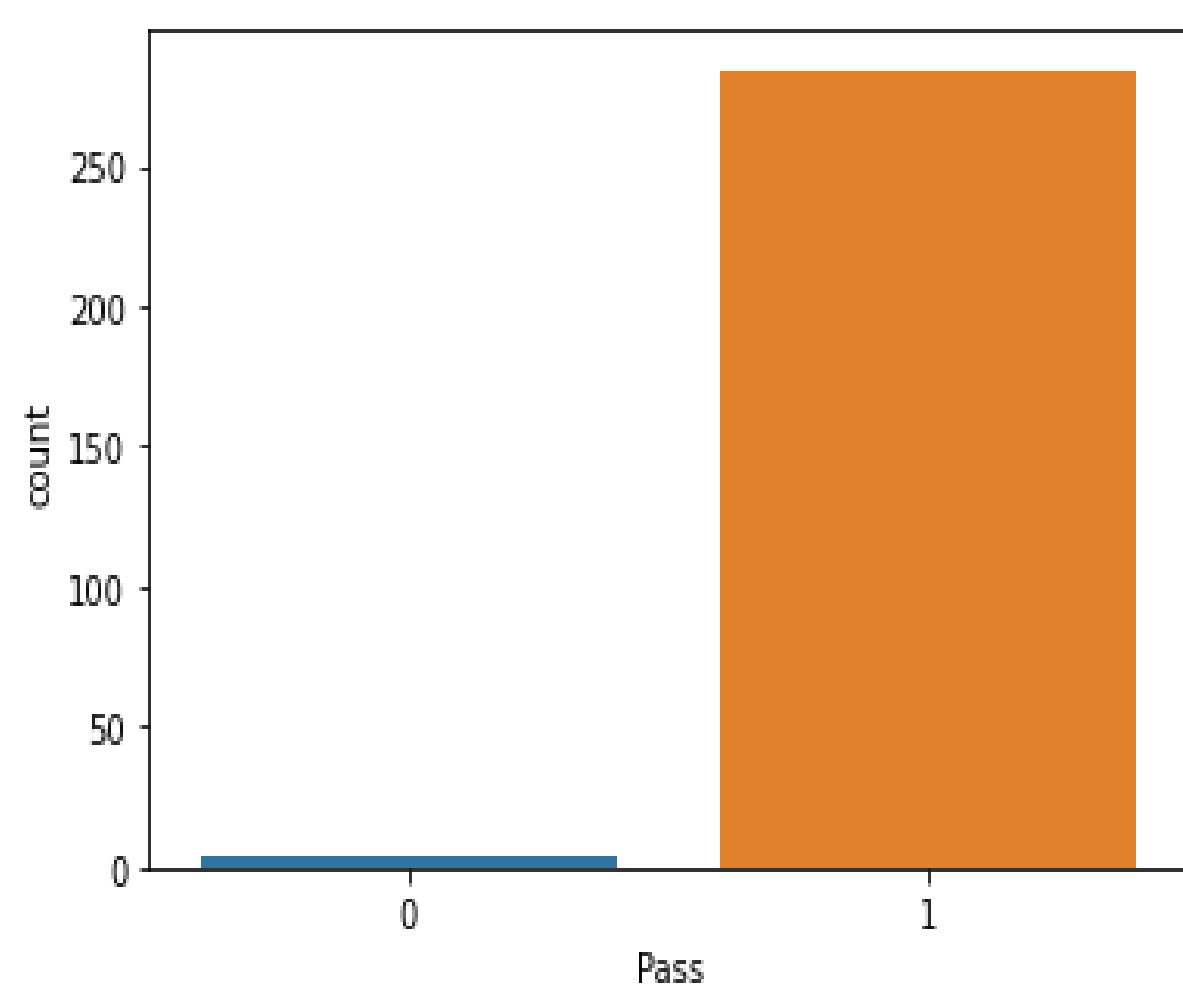
### บทคัดย่อ

การทำนายด้วยวิธีต้นไม้ตัดสินใจ เป็นวิธีการทางที่ช่วยให้วิเคราะห์ได้อย่างรวดเร็ว โดยในการศึกษาครั้งนี้ใช้โมเดลต้นไม้ตัดสินใจเพื่อทำนายคุณภาพน้ำหลังการบำบัดจากศูนย์บำบัดน้ำรวมในกรุงเทพมหานครผ่านค่ามาตรฐานคุณภาพน้ำทั้งหรือไม่ จากการศึกษาพบว่าแบบจำลองสามารถทำนายได้แม่นยำถึงร้อยละ 98.70

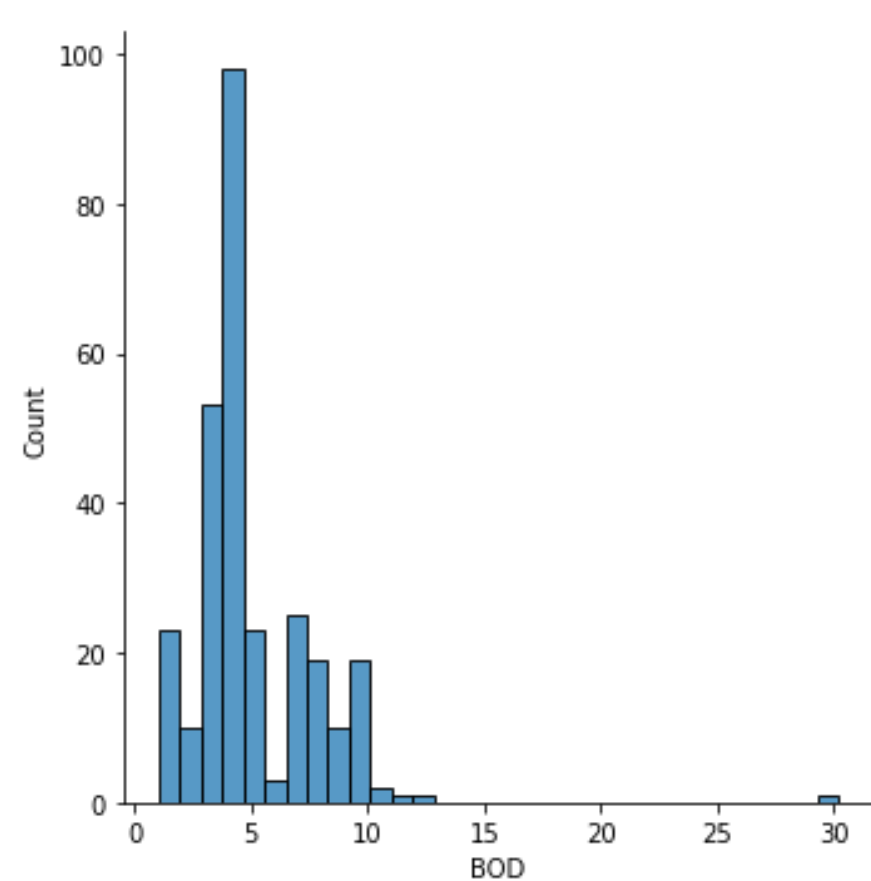
คำสำคัญ — ต้นไม้ตัดสินใจ, คุณภาพน้ำ, Decision Tree, Water quality

### บทนำ

น้ำเป็นทรัพยากรที่สำคัญ ในปัจจุบันมีการใช้น้ำเพื่อกิจกรรมต่างๆ ซึ่งน้ำจากกิจกรรมต่างๆทั้งอุปโภคและบริโภคนั้นเมื่อปล่อยลงแหล่งน้ำสาธารณะ อาจทำให้แหล่งน้ำเน่าเสียเนื่องจากความสกปรกของน้ำ จึงต้องมีการปรับปรุงคุณภาพของน้ำก่อนการระบายน้ำทิ้ง การวิเคราะห์ผ่านห้องปฏิบัติการทางวิทยาศาสตร์ต้องใช้เวลาในการวิเคราะห์มาก จึงใช้การเรียนรู้ของเครื่องเพื่อใช้ในการทำนายคุณภาพของน้ำ โดยในโครงงานนี้จึงได้ทำการศึกษาและทำนายคุณภาพน้ำด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ข้อมูลจากตัวชี้วัดพัฒนาเมือง มีจำนวนพารามิเตอร์ทั้งหมด 6 ตัว ได้แก่ pH BOD SS pH TN TP และ Oil and Grease



รูปที่ 1. แผนภูมิแสดงค่าน้ำที่ผ่านและไม่ผ่าน



รูปที่ 2. แผนภูมิการกระจายตัวของ BOD

### วิธีดำเนินการวิจัย

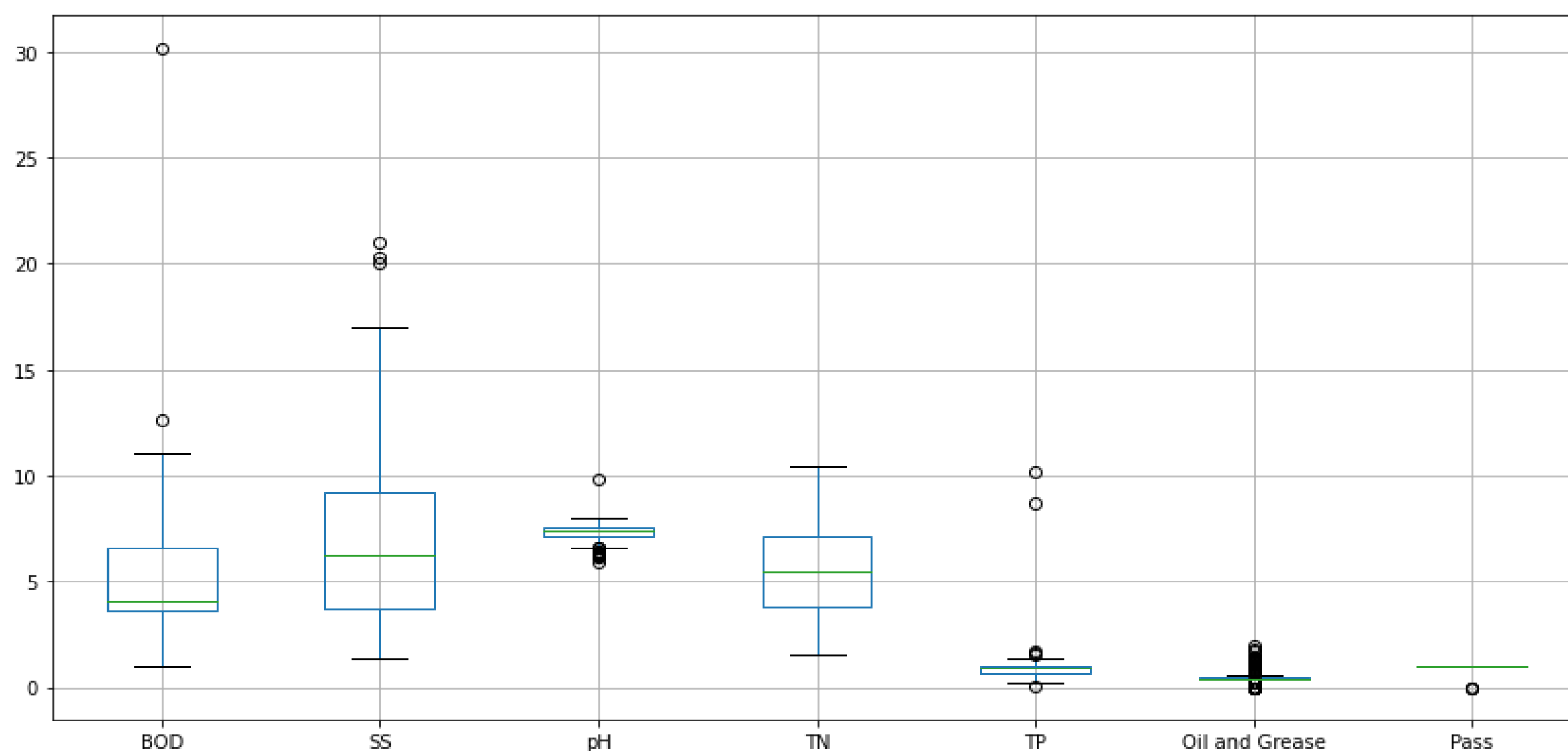
ในการดำเนินการวิจัยใช้ข้อมูลคุณภาพน้ำหลังการบำบัดของโรงควบคุมคุณภาพน้ำในพื้นที่กรุงเทพมหานครโดยเทียบกับค่ามาตรฐาน(ตารางที่ 1) เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูล คือ Google Colab มีขั้นตอนการดำเนินการดังนี้ การเตรียมข้อมูล การตรวจสอบข้อมูลเบื้องต้น ตรวจสอบรูปร่างของข้อมูล หาข้อมูลที่ไม่สามารถระบุค่าได้ สร้างตารางอธิบายชุดข้อมูล และเติมข้อมูลที่ไม่สามารถระบุค่าได้ด้วยค่าเฉลี่ยของพารามิเตอร์นั้น (ตารางที่ 2) สร้างแผนภูมิแสดงค่าน้ำที่ผ่านมาตรฐาน (รูปที่ 1) และแผนภูมิการกระจายตัวของ BOD (รูปที่ 2) การสอนโมเดล โดยใช้โมเดลต้นไม้ตัดสินใจ โดยใช้สองพารามิเตอร์ คือ ข้อมูลจริงและข้อมูลที่คาดการณ์ โดยได้ความแม่นยำอยู่ที่ 100% จากนั้นตรวจสอบความแม่นยำ และกำหนดโครงสร้างของโมเดลต้นไม้ตัดสินใจ โดยใช้ GridSearchCV เพื่อปรับแต่งโมเดล

ตารางที่ 1. ตารางมาตรฐานคุณภาพน้ำทั้ง

Parameter	Standard value	Unit	Method
BOD	≤ 20	mg/L	Azide Modification
SS	≤ 30	mg/L	Glass Fibre Filter Disc
pH	5.5 - 9	-	pH Meter
TN	≤ 20	mg/L	kjeldahl
TP	≤ 2	mg/L	Ascorbic Acid
Oil and Grease	≤ 5	mg/L	Partition gravimetric method

### ผลการทดลอง

หลังจากทำการแสดงด้วย Box Plot (รูปที่ 3) แสดงให้เห็นถึงค่าผิดปกติของข้อมูล ซึ่งไม่มีพารามิเตอร์ตัวใดมีค่าผิดปกติมากอย่างมีนัยสำคัญ ผลของการสอนโมเดลต้นไม้ตัดสินใจและทำการตั้งค่าการค้นหาแบบกริด ผลปรากฏว่ามีเปอร์เซ็นต์การจำแนกที่มีความแม่นยำสูงถึงร้อยละ 98.70 (รูปภาพที่ 4)



รูปที่ 3. แผนภูมิแสดงข้อมูลของแต่ละพารามิเตอร์คุณภาพน้ำ

ตารางที่ 2. ค่าน้อยที่สุด ค่าเฉลี่ย และ ค่ามากที่สุดของพารามิเตอร์

Parameters	Min	Mean	Max
BOD	1.000000	4.969236	30.200000
SS	1.320000	6.689063	21.000000
pH	5.900000	7.287743	9.790000
TN	1.510000	5.489324	10.410000
TP	0.070000	0.907262	10.150000
Oil and Grease	0.000000	0.421184	2.000000

```
print(f"Best: {grid_search_dt.best_score_:.3f} using {grid_search_dt.best_params_}")
means = grid_search_dt.cv_results_['mean_test_score']
stds = grid_search_dt.cv_results_['std_test_score']
params = grid_search_dt.cv_results_['params']

for mean, stdev, param in zip(means, stds, params):
    print(f"{mean:.3f} ({stdev:.3f}) with: {param}")

print("Training Score:", grid_search_dt.score(X_train, Y_train)*100)
print("Testing Score:", grid_search_dt.score(X_test, Y_test)*100)
```

```
Best: 0.987 using {'criterion': 'gini', 'min_samples_split': 2, 'splitter': 'random'}
0.986 (0.006) with: {'criterion': 'gini', 'min_samples_split': 2, 'splitter': 'best'}
0.987 (0.004) with: {'criterion': 'gini', 'min_samples_split': 2, 'splitter': 'random'}
0.987 (0.004) with: {'criterion': 'gini', 'min_samples_split': 4, 'splitter': 'best'}
0.987 (0.004) with: {'criterion': 'gini', 'min_samples_split': 4, 'splitter': 'random'}
0.987 (0.004) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'best'}
0.987 (0.004) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'random'}
0.986 (0.006) with: {'criterion': 'gini', 'min_samples_split': 8, 'splitter': 'best'}
```

รูปที่ 4. ค่าพารามิเตอร์ที่กำหนดในโมเดลและความแม่นยำของโมเดลต้นไม้ตัดสินใจ

### สรุป

งานวิจัยนี้ เริ่มจากการวิเคราะห์ข้อมูลของพารามิเตอร์และค่ามาตรฐานคุณภาพน้ำทั้ง จากผลการวิจัยพบว่าแบบจำลองต้นไม้ตัดสินใจมีความแม่นยำสูง อย่างไรก็ตามแบบจำลองนี้ เป็นข้อมูลเฉพาะโรงควบคุมคุณภาพน้ำกรุงเทพมหานคร และจำนวนตัวอย่างค่อนข้างน้อย จึงมีข้อจำกัดในการทำนายได้

### เอกสารอ้างอิง

ภ.อาจอาษา, “การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล” ,มหาวิทยาลัยสารคาม, ตุลาคม 2564.  
ศ. อาริรัชกุล, “การทำนายคุณภาพน้ำของคลองในกรุงเทพมหานครโดยใช้เทคนิคการทำเหมืองข้อมูล”, มหาวิทยาลัยจุฬาลงกรณ์, 2555.  
ศ. อาริรัชกุล, “การเปรียบเทียบแบบจำลองระบบผสมโครงข่ายประสาทเทียมพืซชี่กับแบบจำลอง โครงข่ายใยประสาทประดิษฐ์เพื่อทำการประมาณค่าออกซิเจนที่ใช้ในการย่อยสลายสารอินทรีย์โดยแบคทีเรีย: กรณีศึกษาคลองแสนแสบ , มหาวิทยาลัยราชภัฏสวนสุนันทา, 2553.

