

บทที่ 3 ขั้นตอนการทำงาน

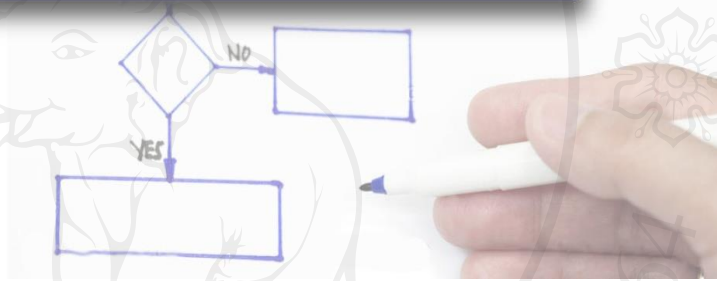
จากการศึกษาและค้นคว้าเพื่อนำไปสู่การพัฒนาระบบและทดลองเพื่อทดสอบระบบให้มีประสิทธิภาพในด้านของการจำแนกกลุ่มของคลาสจากข้อมูลที่นำเข้าสู่ระบบซึ่งนำไปสู่การสร้างต้นไม้มัดตัดสินใจให้มีผลที่ถูกต้องใกล้เคียงกับข้อมูลจริงที่ใช้ทดสอบมากที่สุด ดังรูปที่ 3.1 ที่แสดงให้เห็นถึงขั้นตอนการทำงานของระบบการสร้างต้นไม้มัดตัดสินใจที่มีประสิทธิภาพ



รูปที่ 3.1 แสดงถึงขั้นตอนการทำงานของระบบการสร้างต้นไม้มัดตัดสินใจที่มีประสิทธิภาพ

3.1 การเตรียมข้อมูล

ข้อมูลที่ใช้ในการทำงานของระบบได้นำชุดข้อมูล Iris Dataset ที่อยู่รูปแบบของไฟล์ .CSV และมีการจัดลำดับของฟิลด์โดยให้แอตทริบิวต์ที่เป็นคลาสอยู่ด้านขวาสุดของตารางชุดข้อมูล และด้านซ้ายจะเป็นข้อมูลที่ใช้เปรียบเทียบตามลำดับ มีการนำเข้าสู่ข้อมูลสู่ระบบที่เป็นเว็บแอปพลิเคชันผ่านปุ่ม Input จากนั้นระบบจะอัปโหลดข้อมูลสู่ระบบ



รูปที่ 3.2 ขั้นตอนในการอัปโหลดไฟล์ของชุดข้อมูลเข้าไปยังระบบ

3.2 การวิเคราะห์ข้อมูล

เมื่อนำข้อมูลเข้าสู่ระบบจะเข้าสู่การวิเคราะห์ข้อมูล ในการวิเคราะห์ส่วนนี้จะแบ่งการทำงานเป็น 2 ส่วนคือ

3.2.1 การวิเคราะห์ข้อมูลก่อนการประมวลผล (Pre-processing)

ขั้นตอนในส่วนนี้จะเป็นการวิเคราะห์ข้อมูลที่ผ่านการเตรียมให้อยู่ในรูปแบบของไฟล์ .CSV (จากขั้นตอนที่ 3.1) โดยมีรูปแบบการวิเคราะห์ข้อมูลทั้งหมดคือ

- Binding Depth แบบ 2 Bin
- Binding Depth แบบ 3 Bin
- Binding Mean แบบ 2 Bin
- Binding Mean แบบ 3 Bin
- Z-Score
- Max-min normalization

โดยผู้ใช้งานจะเลือกใช้รูปแบบที่กล่าวข้างต้นให้เหมาะสมกับข้อมูลแต่ละแบบเพื่อที่จะทำให้การประมวลผลมีประสิทธิภาพมากที่สุด

Development of Efficient Decision Tree Construction System

Set Data type
Set Data type is Set type of data for preprocessing.

	Petal width	Petal length	Sepal width	Sepal
1	2	14		33
2	24	56		31
3	23	51		31
4	2	10		36
5	20	52		30
6	19	51		27
7	13	45		28
8	16	47		33
9	17	45		25

Column :1 Max-min normalization Column :2 Binding Depth 3 Bin
Column :3 Z-Score Column :4 Binding Mean 3 Bin

Process Status

- Upload file
- Setting fields
- Data preprocessing
- Decision Tree Processing
- Test
- Download data test file
- Download data test Breast Tissue

Next

รูปที่ 3.3 แสดงตารางข้อมูลที่อ่านจากไฟล์ข้อมูล และขั้นตอนที่ผู้ใช้งานจะต้องมีการเลือกชนิดของการ Pre-Processing ในแต่ละคอลัมน์

3.2.2 การวิเคราะห์ข้อมูลในส่วน Decision tree

ขั้นตอนในส่วนนี้จะเป็นการนำข้อมูลที่ผ่านการวิเคราะห์ก่อนการประมวลผลมาแล้วมาทำการสร้างต้นไม้ตัดสินใจอย่างมีประสิทธิภาพ โดยที่การสร้างต้นไม้ตัดสินใจนั้นจะมีขั้นตอนต่างๆดังนี้

- การคำนวณหาค่าคาดการณ์ของกลุ่มการจำแนกข้อมูล
- การคำนวณหาข้อมูลคาดการณ์เพื่อหาโหนด (Node) แยกของแต่ละคุณลักษณะ
- การคำนวณค่าได้เปรียบเพื่อเปรียบเทียบลำดับความสำคัญของแต่ละคุณลักษณะข้อมูล

หลังจากที่ระบบได้ผ่านขั้นตอนดังกล่าวแล้วก็จะได้ต้นแบบของต้นไม้ตัดสินใจเพื่อที่จะนำไปเข้าสู่กระบวนการทดสอบเพื่อหาอัลกอริทึมของต้นไม้ตัดสินใจที่มีความผิดพลาดน้อยที่สุด

3.3 การทดสอบข้อมูล

การทดสอบข้อมูลจะเป็นการทดสอบ โดยการใช้เทคนิค Cross validation ในการทดสอบโครงสร้างของต้นไม้ตัดสินใจโดยมีการแบ่งข้อมูลเป็น 4 ชุดและมีการสลับลำดับแถวข้อมูลเพื่อให้ข้อมูลมีการกระจายตัวมากที่สุด

3.4 การแสดงผลข้อมูล

หลังจากที่ผ่านกระบวนการทดสอบเรียบร้อยแล้ว ระบบจะทำการประมวลผลและสร้างต้นไม้ตัดสินใจออกมาในรูปแบบอัลกอริทึมเงื่อนไข (IF ELSE) อีกทั้งมีการแสดงความถูกต้องเมื่อมีการทดสอบแบบ Cross validation ในลำดับข้อมูลที่ต่างกัน 4 ชุด

Development of Efficient Decision Tree Construction System

Decision Tree Summary

Data set.

Petal width	Petal length	Sepal width	Sepal
1	2	14	33
2	24	56	31
3	23	51	31
4	2	10	36
5	20	52	30
6	19	51	27
7	13	45	29
8	16	47	33
9	17	45	25

Data Pre-processing.

Petal width	Petal length	Sepal width	Sepal
9	over	midium	under
10	over	midium	over
11	under	low	over
12	over	high	under
13	under	low	over
14	under	low	over
15	under	midium	under
16	over	high	under
17	under	midium	under

Decision Tree Training set Testing.

Training Set1.

```

if(col1==low){ class = Setosa }
if(col1==high){ class = Verginica }
if(col1==midium){ class = Versicolor }

```

Correct Percent : 97.3684210526

Training Set2.

```

if(col1==low){ class = Setosa }
if(col1==high){ class = Verginica }
if(col1==midium){ class = Versicolor }

```

Correct Percent : 97.3684210526

Training Set3.

```

if(col1==low){ class = Setosa }
if(col1==high){ class = Verginica }
if(col1==midium){ class = Versicolor }

```

Correct Percent : 97.3684210526

Training Set4.

```

if(col1==low){ class = Setosa }
if(col1==high){ class = Verginica }
if(col1==midium){ class = Versicolor }

```

Correct Percent : 97.3684210526

Summary Correct Percent : 97.3684210526

รูปที่ 3.4 แสดงตารางข้อมูล อัลกอริทึมของต้นไม้ตัดสินใจ และเปอร์เซ็นต์ความถูกต้องของต้นไม้ตัดสินใจ