

## บทที่ 3

### วิธีการดำเนินงานวิจัย

การดำเนินงานวิจัยแบ่งออกเป็น 2 ส่วนคือ ส่วนของการเตรียมต้นไม้มัดตสันใจและการรวมข้อมูลที่ได้จากต้นไม้มัดตสันใจ (Compromiser) และส่วนของรูปแบบการวัดผลและการประเมินผล

#### 3.1 โครงสร้างของระบบ

โครงสร้างของระบบในการรวมต้นไม้มัดตสันใจด้วยการลงคะแนนโดยพิจารณาภูมิร่วมกันกับกฎที่อยู่ใกล้เคียงแบ่งเป็น 2 ส่วนคือ ส่วนการเตรียมการจำแนกหรือเตรียมต้นไม้มัดตสันใจและส่วนของการเรียนรู้ของต้นไม้มัดตสันใจ (Compromiser)

##### 3.1.1 ส่วนการเตรียมการจำแนกหรือต้นไม้มัดตสันใจ

ขั้นตอนการเตรียมดังนี้

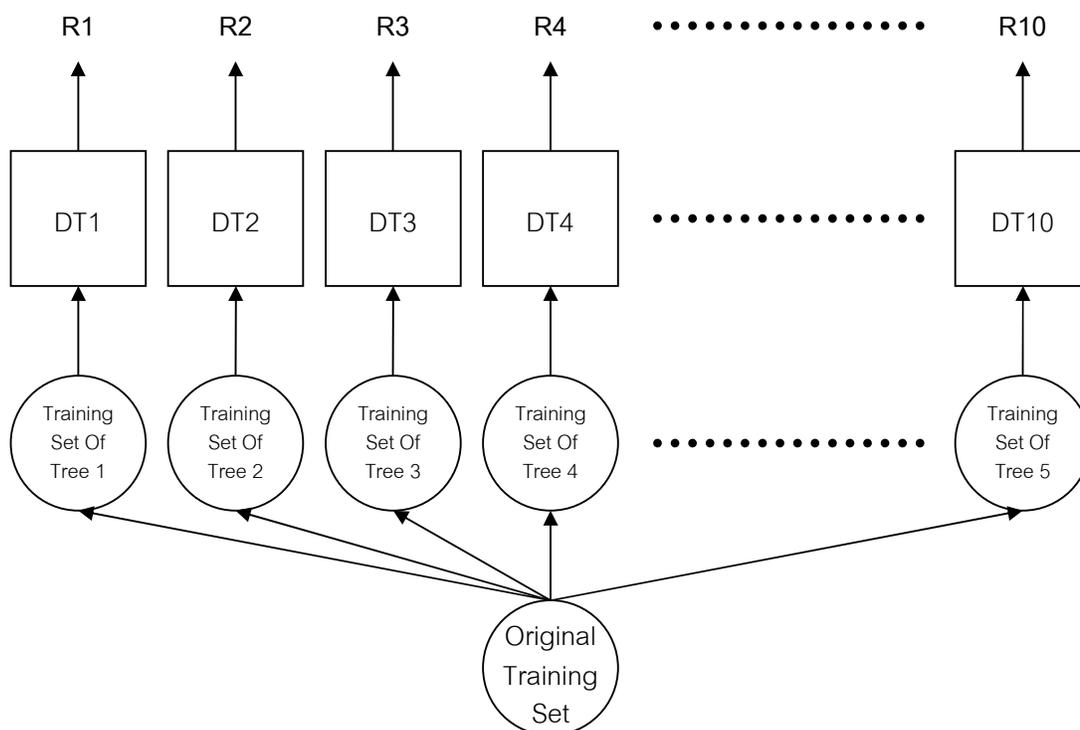
1 สุ่มเลือกชุดข้อมูลจากชุดข้อมูลตั้งต้นโดยวิธีบูตสเตรป ชุดข้อมูลที่สุ่มมีขนาดข้อมูลเท่ากับชุดข้อมูลตั้งต้น เนื่องจากการสุ่มชุดข้อมูลนี้ทำให้บางข้อมูลปรากฏมากกว่าหนึ่งครั้งในขณะที่บางข้อมูลอาจไม่ปรากฏเลย

2 ทำการแบ่งชุดข้อมูลแบบไขว้ข้ามสิบลุ่ม (10-Fold Cross Validation) โดยแบ่งชุดข้อมูลจำนวนเป็น 10 ส่วน นำข้อมูล 9 ส่วนมาสร้างเป็นต้นไม้มัดตสันใจหรือการจำแนกและนำข้อมูลที่เหลืออีก 1 ส่วนมาทำเป็นข้อมูลอินพุต (Input) เพื่อใช้เป็นชุดทดสอบ (Test Set)

3 คำนวณหาค่าความใกล้เคียงของกฎแต่ละกฎกับกฎทุกกฎ

4 ทำข้อ 1 ถึงข้อ 3 ซ้ำจนกระทั่งได้จำนวนต้นไม้มัดตสันใจตามที่กำหนด โดยในวิทยานิพนธ์นี้ทดลองโดยใช้ต้นไม้มัดตสันใจจำนวน 10 ต้น ดังภาพที่ 3.1

ภาพที่ 3.1 : การเตรียมการจำแนกในแต่ละการจำแนกและผลลัพธ์



โดย  $DT_n$  คือ ต้นไม้ตัดสินใจต้นที่  $n$  ;  $R_n$  คือ กฎที่ได้จากต้นไม้ตัดสินใจต้นที่  $n$

### 3.1.2 ส่วนของการเรียนรู้ของต้นไม้ตัดสินใจ

ขั้นตอนการการเรียนรู้ดังนี้

1 สร้างกลุ่มของแต่ละกฎด้วยการหาสมาชิกให้กับกฎนั้นๆ โดยกฎที่นำมาเป็นสมาชิกนั้นมีค่าความใกล้เคียงมากกว่าหรือเท่ากับ 0.6

2 ทำการรวมการจำแนกด้วยการลงคะแนนแบบทั่วไป โดยวิธานิพนธ์ฉบับนี้ได้ทดลองสองวิธีใหม่คือ

ก การลงคะแนนไปโดยกฎ<sup>+</sup>

ใช้วิธีลงคะแนนแบบทั่วไปร่วมกับกฎที่อยู่ใกล้เคียงหรือกฎที่เป็นสมาชิก คือเมื่อได้กฎที่จำแนกข้อมูลทดสอบแล้วจะลงคะแนนกับคำตอบของกฎนั้นโดยพิจารณาพร้อมกับคะแนนที่ได้จากกฎที่อยู่ใกล้เคียงด้วย

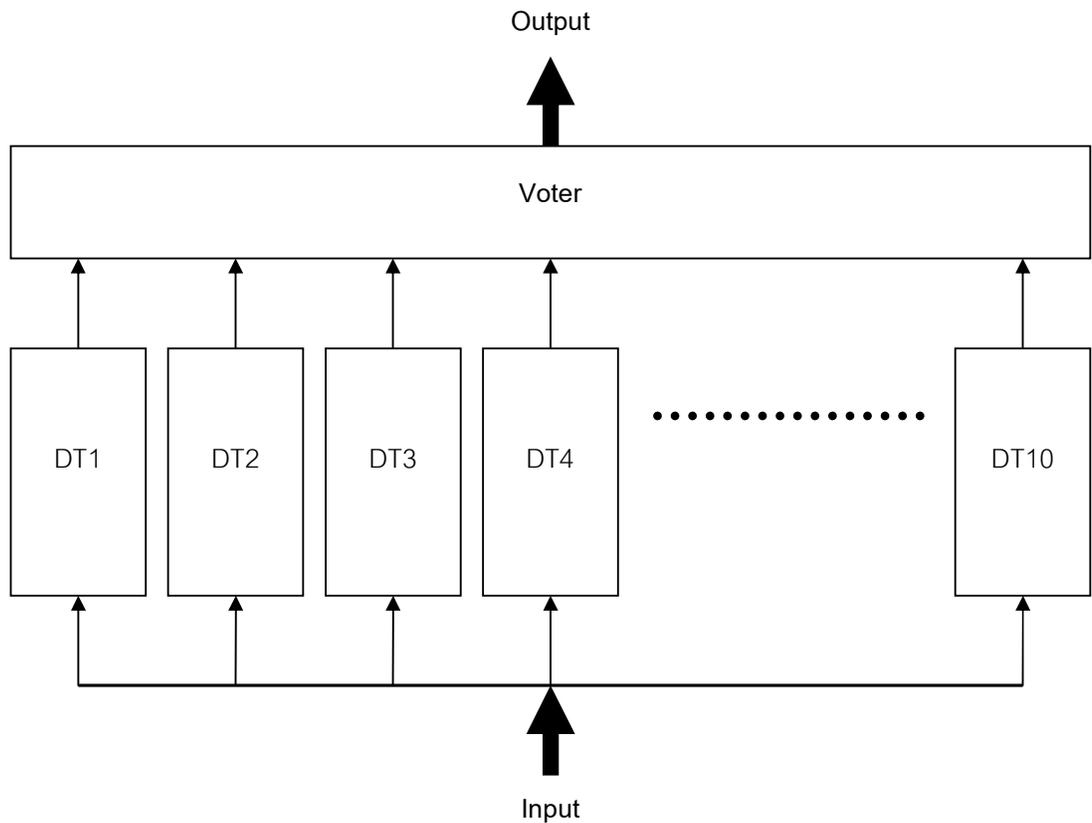
ข การลงคะแนนโดยตัวอย่างเรียนรู้<sup>+</sup>

ใช้วิธีลงคะแนนแบบทั่วไปร่วมกับกฎที่อยู่ใกล้เคียงเช่นกัน แต่จะลงคะแนนจากคำตอบของข้อมูลที่จำแนกได้ด้วยกฎที่จำแนกข้อมูลทดสอบได้ คือเมื่อได้กฎที่จำแนกข้อมูลทดสอบแล้วจะลงคะแนนกับคำตอบโดยคำตอบของข้อมูลที่จำแนกได้ด้วยกฎนั้น โดยพิจารณาร่วมกับคะแนนที่ได้จากคำตอบของข้อมูลที่จำแนกได้ด้วยกฎที่อยู่ใกล้เคียงด้วย

3 ใช้คำตอบที่มีคะแนนสูงสุดเป็นคำตอบสุดท้ายของการรวมการจำแนกแต่ละวิธี

4 ทำข้อ 1 ถึงข้อ 3 ซ้ำโดยใช้ค่าความใกล้เคียงมากกว่าหรือเท่ากับ 0.7, 0.8 และ 0.9 แทนค่า 0.6 ในข้อ 1 ตามลำดับ

ภาพที่ 3.2 : ส่วนของการลงคะแนนที่ได้ในแต่ละการจำแนก (Compromiser)



ภาพที่ 3.3 : อัลกอริทึมการลงคะแนนไปโดยกฎ<sup>+</sup>

```

MajorityRulePlus ()
Do While Not EndOfDicisionTree () {
  If ClassifiedByRule () {
    Do While Not EndOfClass {
      AmountOfVotes[ClassID] += NumberOfClass &_
      FromRuleMember[ClassID]
      MoveToNextClass ()
    }
  }
  MoveToNextDecisionTree
}
ReturnMaxVoteClass ()
END MajorityRulePlus

```

ภาพที่ 3.4 : อัลกอริทึมการลงคะแนนโดยตัวอย่างเรียนรู้<sup>+</sup>

```

MajorityClassPlus ()
Do While Not EndOfDicisionTree () {
  If ClassifiedByRule () {
    Do While Not EndOfClass {
      AmountOfVote += AmountOfClassFromTrainingSet &_
      WhereClassifiedByRuleThatClassifyNewExample
      MoveToNextClass ()
    }
  }
  MoveToNextDecisionTree
}
ReturnMaxVoteClass ()
End MajorityClassPlus

```

### 3.2 วิธีการวัดประสิทธิภาพ

ทำการเปรียบเทียบประสิทธิภาพโดยการวัดความถูกต้องของวิธีการลงคะแนนโดยกฎ<sup>+</sup> เปรียบเทียบกับวิธีแบ็กกิง และวิธีการลงคะแนนโดยตัวอย่างเรียนรู้<sup>+</sup> เปรียบเทียบกับวิธีการลงคะแนนแบบทั่วไปโดยตัวอย่างเรียนรู้