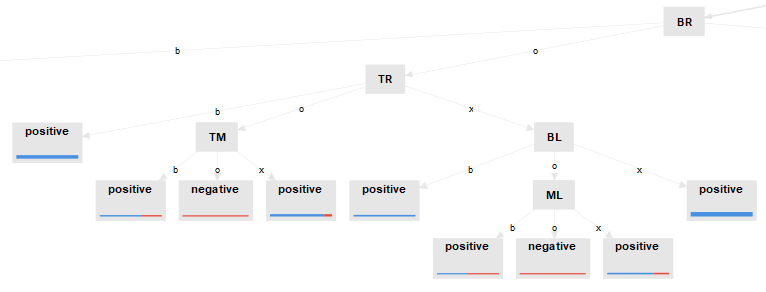
**Lab 7** **204453: Pattern Recognition**

**คำสั่ง:** ให้นักศึกษาดาวน์โหลดไฟล์ชื่อ Tic-Tac-Toe Endgame จาก UC Irvine Machine Learning Repository

จากนั้นให้ทำการทดลอง ต้นไม้การตัดสินใจ (Decision Tree) ดังนี้

1. แปลงไฟล์ข้อมูลให้เป็น .csv แล้วนำไฟล์เข้า RapidMiner Studio
2. แสดงผลรูปต้นไม้การตัดสินใจ (Decision Tree)
3. เปรียบเทียบ Accuracy กับ Decision Tree, Random Forest และ LightGBM

1. ไฟล์ Tic-Tac-Toe Endgame  
- รายละเอียดการรู้จำแบบ  
Tic-Tac-Toe Endgame เป็นไฟล์ชุดข้อมูลของผลการเล่นเกม Tic-Tac-Toe โดยเราต้องการที่จะทำนายว่าลักษณะของรูปเกมแต่ละแบบจะมีชั้นข้อมูลเป็น positive หรือเป็น negative แทนการชนะของฝ่าย x และ ฝ่าย o ตามลำดับ  
- ชั้นข้อมูล  
ชั้นข้อมูลมี 2 ชั้นข้อมูล คือ positive แทนการชนะของ x และ negative แทนการชนะของ o  
- ค่าทางสถิติ  
ชุดข้อมูล Tic-Tac-Toe มีจำนวนระเบียน 958 ระเบียน, มีจำนวนสดมภ์ 10 สดมภ์, มีจำนวนชั้นข้อมูล 2 ชั้นข้อมูล และ จำนวนข้อมูลในแต่ละชั้นข้อมูล คือ ชั้นข้อมูล positive มีจำนวนข้อมูล 626 ข้อมูล และชั้นข้อมูล negative มีจำนวนข้อมูล 332 ข้อมูล

2. แสดงผลรูปต้นไม้การตัดสินใจ (Decision Tree)  
โดยต้นไม้การตัดสินใจใช้ Parameters ดังต่อไปนี้ (criterion = gini\_index, maximal depth = 10, apply pruning = True)  


3. ผลการทดลองของตัวแบบ Decision Tree, Random Forest และ LightGBM  
- Model Accuracy on Dataset  
 Decision Tree’s Accuracy = 85.71%  
 Random Forest’s Accuracy = 98.85%  
 LightGBM’s Accuracy = 95.83%

- Model Accuracy by Confusion Matrix

**Decision Tree** usingParameters: (criterion = gini\_index, maximal depth = 10, apply pruning = True)

|  |  |  |  |
| --- | --- | --- | --- |
|  | *true positive* | *true negative* | *class precision* |
| *Pred. positive* | 571 | 82 | 87.44% |
| *Pred. negative* | 55 | 250 | 81.97% |
| *class recall* | 91.21% | 75.30% |  |

**Random Forest** usingParameters: (number of trees = 1000, criterion = gain ratio, maximal depth = 10)

|  |  |  |  |
| --- | --- | --- | --- |
|  | *true positive* | *true negative* | *class precision* |
| *Pred. positive* | 626 | 11 | 98.27% |
| *Pred. negative* | 0 | 321 | 100.00% |
| *class recall* | 100.00% | 96.69% |  |

**LightGBM** usingParameters: (number of trees = 100, reproducible = False, maximal depth = 5)

|  |  |  |  |
| --- | --- | --- | --- |
|  | *true positive* | *true negative* | *class precision* |
| *Pred. positive* | 612 | 26 | 95.92% |
| *Pred. negative* | 14 | 306 | 95.62% |
| *class recall* | 97.76% | 92.17% |  |

4. สรุปผลการทดลอง  
ความแม่นยำในการทำนายชั้นข้อมูลด้วยตัวแบบ Random Forest มีความแม่นยำสูงสุดเมื่อเทียบกับตัวแบบ Decision Tree หรือตัวแบบ LightGBM เนื่องจากตัวแบบ Random Forest เกิดจากการรวม Decision Tree หลาย ๆ Tree เข้าด้วยกัน เปรียบเสมือนกับการที่เราได้ทำข้อสอบซ้ำ ๆ หลายรอบ นั่นจึงทำให้ตัวแบบ Random Forest สามารถทำนายค่าได้ด้วยความแม่นยำที่สูงมาก ในขณะที่ตัวแบบ LightGBM ซึ่งเป็นการรวม Decision Tree หลาย ๆ Tree เช่นเดียวกันกับ Random Forest นั้น สามารถทำนายการจำแนกชั้นข้อมูลได้แม่นยำมากกว่าตัวแบบ Decision Tree ดังนั้นจึงสรุปได้ว่าการรวมต้นไม้หลาย ๆ ต้นเข้าด้วยกันจะสามารถทำนายได้แม่นยำกว่า การใช้ต้นไม้ต้นเดียวอย่างเช่นตัวแบบ Decision Tree ซึ่งพอเป็นต้นไม้ต้นเดียวแล้วจะทำให้ต้นไม้ต้องมีการแตกกิ่งหรือมีความลึกที่มากเกินไปจนอาจทำให้เกิด Overfitting ได้ ตัวแบบ Decision Tree ในการทดลองนี้จึงมีผลการทำนายที่ต่ำสุดเมื่อเทียบกับตัวแบบ Random Forest และ LightGBM

**หมายเหตุ:** รายงาน ให้ใช้รูปแบบของ Lab 5

**Tic-Tac-Toe Endgame Data Set**

https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame