

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการวิจัย

จากการดำเนินการวิจัยจะได้โปรแกรม C4.5 ที่สามารถทำงานได้บนระบบปฏิบัติการวินโดวส์ 95 ซึ่งประกอบด้วยโปรแกรมย่อย 3 โปรแกรม คือ

1. โปรแกรมจัดการข้อมูล
2. โปรแกรมสร้างต้นไม้ตัดสินใจ และการใช้งานต้นไม้ตัดสินใจ
3. โปรแกรมสร้างกฎการตัดสินใจ และการใช้งานกฎการตัดสินใจ

เมื่อนำโปรแกรมนี้ไปทดสอบกับข้อมูลตัวอย่างทดสอบ 8 ตัวอย่าง โดยใช้ค่าตัวเลือกที่เป็นค่าโดยปริยาย เพื่อเปรียบเทียบความถูกต้องของโปรแกรมใหม่ที่ทำงานบนระบบปฏิบัติการวินโดวส์กับโปรแกรมเดิมที่ทำงานบนระบบปฏิบัติการยูนิกซ์ พบว่าผลการทดสอบไม่มีความแตกต่างกันในเรื่องความถูกต้องของต้นไม้ตัดสินใจหรือกฎการตัดสินใจที่ได้ แต่จะแตกต่างกันในเรื่องความเร็วในการทำงาน ทั้งนี้เนื่องจากประสิทธิภาพของเครื่องที่แตกต่างกัน และเมื่อนำโปรแกรม C4.5 ที่ได้ไปทดสอบกับตัวอย่างการอนุมัติสินเชื่อของธนาคารพาณิชย์ ซึ่งประกอบด้วยลักษณะไม่แบ่งพวก 27 ลักษณะ แบ่งเป็นลักษณะที่มีค่าต่อเนื่อง 12 ลักษณะ และมีค่าไม่ต่อเนื่อง 15 ลักษณะ จำนวนทั้งหมด 4,000 ตัวอย่าง แบ่งเป็นตัวอย่างที่ใช้ในการสอน 3,600 ตัวอย่าง และเป็นตัวอย่างที่ใช้ในการทดสอบ 400 ตัวอย่าง โดยใช้วิธี เทน-เวย์ครอสวาไลเดชัน (ten-way cross validation) และทดลองปรับค่าตัวเลือกในการสร้างต้นไม้ตัดสินใจจำนวน 50 ชุด และทดลองปรับค่าตัวเลือกในการสร้างกฎการตัดสินใจอีกจำนวน 300 ชุด เพื่อทดสอบความเป็นไปได้ในการนำไปใช้งาน พบว่าโปรแกรมนี้สามารถสร้างต้นไม้ตัดสินใจที่มีค่าอัตราส่วนความผิดพลาดและขนาดของต้นไม้ต่าง ๆ กัน ขึ้นอยู่กับการตั้งค่าตัวเลือกในการเรียนรู้ โดยเมื่อตั้งค่าตัวเลือกเป็นค่าโดยปริยายคือ ค่า MINOBS เท่ากับ 2 และค่า CF เท่ากับ 25 % และใช้ค่ามาตรฐานอัตราส่วนเกณฑ์ในการสร้างต้นไม้ตัดสินใจ จะได้ต้นไม้ตัดสินใจอยู่ในระดับกลาง คือ ขนาดของต้นไม้ตัดสินใจที่ได้ไม่เล็กหรือใหญ่เกินไปและมีอัตราส่วนความผิดพลาดไม่สูงมาก แต่ถ้าต้องการต้นไม้ตัดสินใจที่มีขนาดเล็กก็สามารถทำได้โดยการตั้งค่าตัวเลือก MINOBS ให้มีค่าสูงขึ้น ทำให้ต้นไม้มีการแตกกิ่งน้อยลง ซึ่งจะมีผลทั้งต้นไม้ตัดสินใจก่อนการตัดแต่งและต้นไม้ตัดสินใจหลังจากการตัดแต่ง หรือใช้การตั้งค่าตัวเลือก CF ให้ต่ำลง ซึ่งจะทำให้มีการตัดแต่งกิ่งมากขึ้น แต่วิธีนี้จะมีผลเฉพาะต้นไม้ตัดสินใจหลังจากการตัดแต่งแล้วเท่านั้น ส่วนเมื่อทดสอบโปรแกรมนี้ในการสร้างกฎการตัดสินใจพบว่าค่าตัว

เลือกที่มีผลต่ออัตราส่วนความผิดพลาดมากที่สุดเมื่อทดสอบกับตัวอย่างสอนคือค่าความซ้ำ ซึ่งค่าความซ้ำที่สูงขึ้นจะทำให้ค่าอัตราส่วนความผิดพลาดลดลง โดยจะเริ่มคงที่เมื่อค่าความซ้ำมากกว่า 12 สาเหตุที่เป็นดังนี้เนื่องจากลักษณะไม่แบ่งพวกที่ใช้สอนมีลักษณะที่ซ้ำกันหลายลักษณะ ส่วนอัตราส่วนความผิดพลาดเมื่อทดสอบกับตัวอย่างทดสอบจะไม่แตกต่างกันมากนักในแต่ละการตั้งค่าตัวเลือก

จากผลการทดสอบอาจพอที่จะสรุปได้ว่าโปรแกรม C4.5 สามารถที่จะประยุกต์ใช้กับการอนุมิตินเชื่อได้ ถ้าอัตราส่วนความผิดพลาดอยู่ในระดับที่พอใจ ซึ่งก็ขึ้นอยู่กับธนาคารแต่ละแห่งว่าจะยอมรับความผิดพลาดที่ระดับเท่าใด โดยจากการทดสอบกับตัวอย่างนี้โดยการให้ค่าตัวเลือกที่เป็นค่าโดยปริยาย จะให้อัตราส่วนความผิดพลาดเมื่อทดสอบกับตัวอย่างทดสอบของตนไม่ตัดสินใจประมาณ 26.6 % และเมื่อทดสอบกับกฎการตัดสินใจจะผิดพลาดประมาณ 26.7 % ซึ่งค่าที่ได้ยังนับว่ามีความผิดพลาดสูง ดังนั้นจึงควรจะได้มีการปรับปรุงต้นไม้ตัดสินใจหรือกฎการตัดสินใจให้มีความถูกต้องมากขึ้น โดยสาเหตุที่ทำให้มีความผิดพลาดสูงอาจเนื่องมาจาก ลักษณะไม่แบ่งพวกที่ใช้ในการเรียนรู้ไม่ครอบคลุมถึงลักษณะบางลักษณะใช้ในการตัดสินใจหรือจำนวนตัวอย่างมีจำนวนน้อยเกินไป หรือตัวอย่างไม่ครอบคลุมการตัดสินใจในบางลักษณะ หรืออาจจะมาจากความผิดพลาดของข้อมูลตัวอย่างเอง ดังนั้นอาจจะต้องมีการปรับเปลี่ยนลักษณะหรือตัวอย่างที่ใช้ในการเรียนรู้แล้วทดลองเรียนรู้สร้างต้นไม้ตัดสินใจหรือกฎการตัดสินใจใหม่ เพื่อให้ได้ต้นไม้ตัดสินใจหรือกฎการตัดสินใจที่เหมาะสมในการนำไปใช้งานต่อไป

ปัญหาที่พบและข้อเสนอแนะ

จากการทดลองใช้โปรแกรมนี้พบว่ามีบางจุดที่ควรจะมีการเพิ่มเติม ได้แก่

1. ในโปรแกรมการจัดการข้อมูล การเปลี่ยนจอภาพจากการดูข้อมูลที่ละตัวอย่าง (Record View) ไปเป็นการดูข้อมูลแบบทั้งหมด (List View) จะเปลี่ยนได้ช้าถ้าจำนวนตัวอย่างมีมาก เนื่องจาก การเปลี่ยนแต่ละครั้งจะมีการสร้างข้อมูลในลิสต์วิวใหม่ทั้งหมด ดังนั้นควรจะใช้การเปิดหน้าต่าง ทั้ง 2 แบบพร้อมกันแต่สามารถจะสลับหน้าต่างไปมาได้โดยไม่ต้องสร้างใหม่
2. โปรแกรมยังขาดการแสดงผลของสถานะการทำงาน หรือการยกเลิกการทำงานกลางคันได้ เนื่องจากบางงาน เช่น การเรียนรู้เพื่อสร้างกฎการตัดสินใจจะใช้เวลาในการเรียนรู้แต่ละครั้ง เมื่อสั่งให้โปรแกรมทำงานแล้วจะไม่ทราบว่าโปรแกรมยังทำงานอยู่หรือไม่
3. โปรแกรมยังขาดการนำข้อมูลเข้าจากฐานข้อมูลโดยผ่าน ODBC (Open Database Connectivity) เพื่อดึงข้อมูลจากฐานข้อมูลเข้าสู่โปรแกรมการจัดการข้อมูลได้โดยตรง
4. ในการใช้งานต้นไม้ตัดสินใจหรือกฎการตัดสินใจ ยังขาดการถอยหลังกลับไปยังคำถามก่อนหน้า