

ดิษฐ์พล มั่นธรรม 2553: การประยุกต์อัลกอริทึมต้นไม้ตัดสินใจกับการวินิจฉัยโรกระบบทางเดิน
หายใจ ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ) สาขาวิชาสถิติ ภาควิชาสถิติ อาจารย์ที่ปรึกษา
วิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์ลีลี อิงศรีสว่าง, Ph.D. 126 หน้า

การศึกษานี้มีวัตถุประสงค์ 1. เพื่อประยุกต์วิธีสืบค้นความรู้จากฐานข้อมูล โดยใช้อัลกอริทึมต้นไม้
ตัดสินใจกับการวินิจฉัยโรกระบบทางเดินหายใจ ในการจำแนกผู้ป่วย 3 โรค คือ โรคการติดเชื้อทางเดินหายใจ
ส่วนบนแบบเฉียบพลัน โรคไซนัสอักเสบเฉียบพลัน และโรคปอดอักเสบ ของโรงพยาบาลพระนครศรีอยุธยา
และ 2. เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมต้นไม้ตัดสินใจ 3 อัลกอริทึม ได้แก่ ID3 C4.5 และ CART
ในการจำแนกหรือคัดกรองผู้ป่วยทั้ง 3 โรค ข้อมูลที่ใช้ศึกษาเป็นข้อมูลเวชระเบียนผู้ป่วยนอกโรกระบบทางเดิน
หายใจ ของโรงพยาบาลพระนครศรีอยุธยา ระหว่างปี พ.ศ.2547 - 2549 จำนวน 7,327 คน โดยตัวแปรที่นำมา
พิจารณาประกอบด้วย อายุ อุณหภูมิร่างกาย เขตที่อยู่อาศัย อาชีพ อาการต่างๆ เช่น มีน้ำมูก มีไข้ คัดจมูก
แน่นหน้าอก ปวดกระบอกตา ปวดศีรษะ หายใจเหนื่อยหอบ และไอ เป็นต้น วิธีการศึกษา ทำการสืบค้น
ความรู้จากฐานข้อมูลผู้ป่วยโรกระบบทางเดินหายใจด้วยอัลกอริทึม ID3 C4.5 และ CART และเปรียบเทียบ
ประสิทธิภาพของอัลกอริทึมทั้งสาม โดยมีการคัดเลือกตัวแปรที่มีความสำคัญต่อการจำแนกผู้ป่วยแต่ละโรค
ด้วยอัลกอริทึม Best First Greedy Stepwise และ Genetic Search แล้วจึงนำตัวแปรที่ได้มาผ่านอัลกอริทึม
ต้นไม้ตัดสินใจข้างต้น เพื่อจำแนกผู้ป่วยทั้ง 3 โรค ด้วยระบบการเรียนรู้และตรวจสอบความถูกต้องของการ
จำแนกผู้ป่วยแต่ละโรคด้วยการแบ่งข้อมูลออกเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบตามหลัก Cross-
validation และ Percentage split

ผลการศึกษา วิธีการสืบค้นความรู้สำหรับจำแนกผู้ป่วยโรคการติดเชื้อทางเดินหายใจส่วนบนแบบ
เฉียบพลัน พบว่าการใช้ตัวแปรที่คัดเลือกได้เพียง 7 ตัวแปร กับอัตราส่วนข้อมูลฝึกสอนต่อข้อมูลทดสอบ
70 : 30 อัลกอริทึม C4.5 ให้ผลค่าวัดประสิทธิภาพสูงสุด คือให้ค่าความถูกต้อง 92.32% ค่าความไว 92.40%
ค่าพยากรณ์บวก 95.63% ค่าพยากรณ์ลบ 90.18% และค่า AUC 96.20% ส่วนการสืบค้นความรู้สำหรับ
จำแนกผู้ป่วยโรคปอดอักเสบ พบว่าการใช้ตัวแปรที่คัดเลือกได้เพียง 8 ตัว และใช้อัตราส่วนข้อมูลฝึกสอนต่อ
ข้อมูลทดสอบ 70 : 30 อัลกอริทึม C4.5 ให้ค่าวัดประสิทธิภาพสูงสุด คือให้ค่าความถูกต้อง 94.70% ค่าความ
ไว 93.42% ค่าพยากรณ์บวก 83.24% ค่าพยากรณ์ลบ 98.16% และค่า AUC 92.20% และการสืบค้นความรู้
สำหรับจำแนกผู้ป่วยโรคไซนัสอักเสบเฉียบพลัน พบว่าตัวแปรที่คัดเลือกได้เพียง 7 ตัว และใช้อัตราส่วนข้อมูล
ฝึกสอนต่อข้อมูลทดสอบ 50:50 อัลกอริทึม CART ให้ค่าวัดประสิทธิภาพสูงสุด คือให้ค่าความถูกต้อง 94.69%
ค่าความไว 100% ค่าพยากรณ์บวก 81.57% ค่าพยากรณ์ลบ 100% และค่า AUC 94.70% ซึ่งผลลัพธ์ที่ได้จาก
การประยุกต์อัลกอริทึมต้นไม้ตัดสินใจ สามารถนำไปเป็นแนวทางสนับสนุนการคัดกรองผู้ป่วยโรกระบบ
ทางเดินหายใจในเบื้องต้นเพื่อการวินิจฉัยต่อไป

ลายมือชื่อนิติ

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก