

ปัจจุบันปริมาณข้อมูลต่างๆ ได้เพิ่มขึ้นอย่างรวดเร็วทำให้เกิดการล้นของข้อมูล ซึ่งเมื่อเราต้องการที่จะหาข้อมูลที่สำคัญ จึงเป็นการยากที่จะหาและอาจจะได้ข้อมูลที่ไม่ตรงกับความต้องการ จึงจำเป็นที่จะต้องมีการจัดการกับข้อมูลที่ไม่จำเป็นออกไป ซึ่งการค้นพบความรู้ในฐานข้อมูล (Knowledge Discovery in Database - KDD) ก็เป็นวิธีการที่มีประสิทธิภาพอย่างหนึ่งที่จะจัดการกับข้อมูลเหล่านี้ได้ โดยข้อมูลที่พบได้ทั่วไปจะแบ่งเป็น 2 แบบ คือข้อมูลที่ต่อเนื่องกัน (Continuous data) และข้อมูลที่ไม่ต่อเนื่องกัน (Discrete data) ในงานวิจัยนี้ได้นำเสนอวิธีการจัดการกับข้อมูลที่ต่อเนื่อง ซึ่งจะต้องทำการแปลงข้อมูลที่ต่อเนื่องให้เป็นข้อมูลที่ไม่ต่อเนื่องก่อน โดยเสนอการแปลงข้อมูลที่ต่อเนื่อง โดยใช้ค่าความสัมพันธ์สูงสุดระหว่างกลุ่มกับคุณสมบัติมาใช้แบ่งช่วงเริ่มต้นแต่ช่วงที่ได้มีจำนวนมากจึงใช้ไค-สแควร์ในการรวมช่วงให้น้อยลง ก่อนใช้ทฤษฎีกราฟเซต (Rough Set Theory) ในการจัดการกับคุณสมบัติที่ไม่จำเป็นและทำการค้นพบกฎความรู้จากตารางการตัดสินใจในการทดสอบกับฐานข้อมูล Iris จากการทดสอบการแปลงข้อมูลที่ต่อเนื่องโดยใช้การปรับปรุงความสัมพันธ์ของกลุ่มคุณสมบัติสามารถแปลงค่าคุณสมบัติต่อเนื่องเป็นช่วงได้รวม 14 ช่วง มีค่าความสัมพันธ์เฉลี่ยระหว่างคุณสมบัติกับกลุ่มเท่ากับ 0.599 จำนวนกฎความรู้ที่ได้มีจำนวน 10 กฎพร้อมลดจำนวนกฎที่ขัดแย้งกันลงเหลือกฎเดียวและมีค่าความถูกต้องของค่ากลุ่มของข้อมูลเท่ากับ 97.33%

The amount of data were rapidly increased recently and caused the data overwhelming. That led to be difficult in searching the required data. The method of eliminating the redundant data was needed. One of the efficient methods was the Knowledge Discovery in Database (KDD). Generally data can separate in 2 cases. One is continuous data and the other is discrete data.

This research presented the method of handling continuous data that could be transform continuous data to discrete data. We presented Class Attribute Interdependence Maximization for specified number of intervals when initializing the discretization intervals and uses chi-square for reduced number of intervals. Then the rough sets theory was applied in eliminating the unnecessary attributes and the discovering performance of knowledge rules from that decision table was done. Iris Dataset was used to test the method. The testing results with CAIM+CHI2 had 14 total intervals, 0.599 of the average class-attribute interdependence redundancy, 10 rules with only one conflicted rule and 97.33% for accuracy of discretized data with class.