กิตติพงศ์ ชมบุญ : เทคนิคการค้นหาคลาสที่ค้นพบได้ยากสำหรับข้อมูลที่ขนาดแตกต่างกันมาก (RARE CLASS DISCOVERY TECHNIQUES FOR HIGHLY IMBALANCED DATA) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.กิตติศักดิ์ เกิดประสพ, 76 หน้า.

การค้นหาคลาสที่ค้นพบได้ยากเป็นงานหนึ่งของการทำเหมืองข้อมูลที่น่าสนใจ เพราะสามารถช่วยเหลือในการผลิตของภาคอุตสาหกรรมด้วยการค้นหารูปแบบการผลิตที่นำไปสู่เหตุการณ์ที่ผลผลิตเสียหายซึ่งมักจดเป็นเหตุการณ์ที่เกิดขึ้นไม่บ่อยจึงเรียกว่าคลาสที่ค้นพบได้ยาก ประโยชน์ที่ได้คือทำให้ทราบถึงสิ่งที่ควรปรับปรุงหรือแนวทางบำรุงรักษาเครื่องจักร เพื่อให้ได้ผลผลิตสมบูรณ์มากขึ้น แต่ในการค้นหาคลาสที่ค้นพบได้ยากนั้นสามารถทำได้ยาก เนื่องจากคุณสมบัติเด่นของคลาสที่มีอยู่น้อยจะถูกคุณสมบัติเด่นของคลาสส่วนใหญ่บดบัง ซึ่งถ้าหากทำการจำแนกข้อมูลด้วยวิธีการปกติแล้วจะไม่สามารถค้นพบคลาสที่ค้นพบได้ยากเลย ดังนั้นงานวิจัยนี้จึงนำเสนอเทคนิคการค้นหาคลาสที่ค้นพบได้ยาก

ในงานวิจัยนี้ได้นำเสนอแนวทางการแก้ปัญหาในการค้นหาคลาสที่ค้นพบได้ยาก ด้วยการคัดเลือกคอลัมน์ด้วยค่าสหสัมพันธ์ และเพิ่มจำนวนคลาสที่ค้นพบได้ยากด้วยวิธีการสุ่มเกิน เพื่อให้คลาสที่ค้นพบได้ยากมีจำนวนใกล้เคียงกับคลาสส่วนใหญ่ทำให้สามารถสร้างรูปแบบเพื่อการจำแนกคลาสที่ค้นพบได้ยากได้

มหาวิทยาลัยเทคโนโลยีสุรนารี

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2555 ลายมือชื่ออาจารย์ที่ปรึกษา_____

KITTIPONG CHOMBOON : RARE CLASS DISCOVERY TECHNIQUES FOR HIGHLY IMBALANCED DATA. THESIS ADVISOR : ASSOC. PROF. KITTISAK KERDPRASOP, Ph.D., 76 PP.

DATA MINING/CLASSIFICATION/RARE CLASS/OVER-SAMPLING

Rare class discovery is one of interesting data mining tasks that can support manufacturing industries by discovering process patterns that can lead to fault products. Faulty process is rarely occurred; it is thus called rare class. The benefit of rare class discovery is the knowledge regarding to the improvement of process or the tool maintenance policy. Rare class discovery is however a difficult task because its rarely occurrences are always dominant by the much larger group of frequently occurred events.

This research therefore proposes techniques to solve the rare class discovery problem. Our main techniques are feature selection with the correlation analysis and then the increase of minority class, which is rare event, with the over-sampling method. With the proposed techniques, patterns of rare class can finally be discovered.

School of Computer Engineering Student's Signature_____

Academic Year 2012 Advisor's Signature_____