

การใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา Using Data Mining Techniques for Analyzing Factors that Influence Students' Library Use

ภัทรพงศ์ พงศ์ภัทรกานต์^{1,2} วิชัย พัวรุ่งโรจน์^{1,3} คมยुทธ ไชยวงษ์^{3,4}

สุชาดา พรหมโคตร^{1,3,4} ปาริชาติ แสงระชัย^{1,5}

¹สำนักวิทยบริการและเทคโนโลยีสารสนเทศ

²สาขาวิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์

³สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

⁴สาขาวิชาวิศวกรรมไฟฟ้าและอิเล็กทรอนิกส์ คณะเทคโนโลยีอุตสาหกรรม

⁵สาขาวิชาบรรณารักษศาสตร์และสารนิเทศศาสตร์ คณะมนุษยศาสตร์และสังคมศาสตร์

มหาวิทยาลัยราชภัฏเลย

e-mail: pathapong@lru.ac.th

บทคัดย่อ

งานวิจัยนี้นำเสนอการทดสอบวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา มหาวิทยาลัยราชภัฏเลย โดยใช้ข้อมูลการเข้าใช้บริการผ่านประตูอัตโนมัติในช่วงเดือนกุมภาพันธ์ถึงตุลาคม 2559 ที่มี 9 ปัจจัยพื้นฐาน คือ วันที่เข้าใช้บริการ ช่วงเวลา เพศ คณะ ชั้นปี จังหวัดที่เกิด หมู่เลือด จำนวนพี่น้อง และเกรดเฉลี่ยสะสม จำนวน 79,953 ชุดข้อมูล ทำการประมวลผลด้วยอัลกอริทึม C5.0, Neural Network และ CART เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของการคัดแยกข้อมูล ผลการศึกษา พบว่า อัลกอริทึม C5.0 ให้ค่าความถูกต้อง 97.78 % และใช้ระยะเวลาในการประมวลผลน้อยกว่าอัลกอริทึมที่นำมาเปรียบเทียบ ผลจากการวิเคราะห์ด้วยอัลกอริทึม C5.0 พบว่ามี 3 ปัจจัยที่มีอิทธิพลต่อการใช้บริการห้องสมุดของนักศึกษาที่ส่งผลตามคณะ คือ เกรดเฉลี่ยสะสม มีอิทธิพลสูงสุทธ้อยละ 93.8 เพศ มีอิทธิพลร้อยละ 6.0 และช่วงเวลา มีอิทธิพลร้อยละ 0.2 ซึ่งนำมาสร้างความสัมพันธ์ได้ 21 ระดับ ซึ่งเป็นแนวทางในการประชาสัมพันธ์และส่งเสริมนักศึกษาเข้ามาใช้บริการห้องสมุดผ่านคณะที่สังกัดได้ โดยเฉพาะเกรดเฉลี่ยมีผลอย่างมากในการเข้ามาใช้บริการ นักศึกษาที่มีเกรดสูงมีแนวโน้มการเข้าใช้ห้องสมุดมากกว่านักศึกษาที่มีเกรดต่ำ ดังนั้น ห้องสมุดควรเน้นไปที่การเปิดบริการหรือเชิญชวนให้นักศึกษาที่มีเกรดน้อยเข้าห้องสมุดมากขึ้น ห้องสมุดควรคิดกิจกรรมส่งเสริมใหม่เพิ่มมากขึ้นเพื่อให้นักศึกษามีความสนใจในการเข้าใช้ห้องสมุด

คำสำคัญ:

เหมืองข้อมูล, การจำแนก, วิเคราะห์ปัจจัย, การใช้บริการห้องสมุด

Abstract

This research sought to examine factors that influence students to use library at LoeiRajabhat University. The data for analysis comprised records of undergraduate students' library daily entry which were collected between February and October 2016 and related students' data records. Nine factors were chosen and brought into account including date entry, period of time in the day, sex, faculty, current year of study, place of birth, blood type, number of siblings, and GPA. Totally, 79953 records of data were brought into analysis by using C5.0, Neural Network and CART algorithms. By comparing the accuracy results from these three algorithms, C5.0 algorithm yielded the highest accuracy (97.78 %) and took the least amount of time for calculation. The results of C5.0 analysis indicated that there were 3 key factors that highly influenced students in library use. The highest influence was GPA at 93.8%, the second was sex at 6% and the last one was period of time in the day at 0.02%. These factors developed 21 levels of relationship that can be used for improving library services and influencing students to use library through the faculties they attend. Students with higher GPA were likely to use library more than those with less GPA. This suggests that the library should focus more on students with low GPA and influence them to use library. New activities should be initiated to attract more students to visit the library.

Keyword:

Data Mining, Classification, Factors Analysis, Library Use

บทนำ

การเรียนรู้ถือเป็นเรื่องสำคัญในสังคมปัจจุบัน เนื่องจากความเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็ว จากความก้าวหน้าทางเทคโนโลยี โดยเฉพาะทางเทคโนโลยีสารสนเทศและการสื่อสาร ดังนั้นการเรียนรู้ให้เท่าทันและศึกษาหาความรู้ให้กว้างขึ้นสามารถค้นหาข้อมูลได้จากอินเทอร์เน็ต และมีแหล่งข้อมูลที่สำคัญแห่งหนึ่งคือห้องสมุดที่สามารถเป็นแหล่งเรียนรู้ด้านต่าง ๆ ด้วยสารนิเทศ ทรัพยากรสารสนเทศในรูปแบบต่าง ๆ เช่น หนังสือ วารสาร หนังสือพิมพ์ เป็นต้น ในปัจจุบันห้องสมุดทำหน้าที่จัดเก็บรวบรวม จัดระบบเพื่อให้บริการสื่อสารนิเทศ ตลอดจนเทคโนโลยีคอมพิวเตอร์และเทคโนโลยีทางการสื่อสาร รวมถึงเครื่องมือในการค้นหาและดำเนินการให้บริการแก่ผู้ใช้ห้องสมุดเพื่อให้เกิดประโยชน์สูงสุด โดยวัตถุประสงค์หลักทั่วไปของห้องสมุดประกอบด้วย 5 ประการ คือ เพื่อการศึกษา เพื่อให้ความรู้และข่าวสาร เพื่อใช้ในการค้นคว้า เพื่อจรรโลงใจมนุษย์ และเพื่อนันทนาการ และยังเป็นส่วนหนึ่งในการส่งเสริมการเรียนการสอนการศึกษาแก่สถาบันการศึกษาการศึกษาในปัจจุบันมุ่งเน้นให้ผู้เรียนสามารถศึกษาและค้นคว้าหาความรู้ต่างๆ ด้วยตัวเองจากแหล่งทรัพยากรสารสนเทศ มาประกอบและประยุกต์กับความรู้ที่ได้จากการเรียนในห้องเรียน (ไพฑูริย์ สีนลารัตน์, 2557) กล่าวว่า การเรียนในระดับอุดมศึกษา ซึ่งเป็นการศึกษาเพื่อสร้างองค์ความรู้ใหม่ ผู้เรียนต้องทำการศึกษาค้นคว้าด้วยตัวเองมากกว่าที่เรียนรู้รับสิ่งต่างๆ ภายในห้องเรียน ซึ่งการจัด

การศึกษาจำเป็นต้องมีห้องสมุดเป็นแหล่งเรียนรู้เพื่อช่วยเพิ่มศักยภาพและคุณภาพของนักศึกษา เพราะห้องสมุดเป็นแหล่งรวบรวม ตำราวิชาการหลากหลายชนิดอำนวยความสะดวกในการศึกษาค้นคว้าหาความรู้ ผู้ใช้บริการสามารถเลือกสรรตามความต้องการ ซึ่งจะส่งผลให้เป็นผู้ที่มีความรู้ ความสามารถ ทนต่อเหตุการณ์ เป็นสถานที่ซึ่งมีผู้ให้บริการเต็มใจให้บริการโดยไม่เลือกปฏิบัติ ปลุกฝังนิสัยรักการอ่านและเป็นแหล่งสารสนเทศที่กว้างไกลเพื่อรับใช้สังคม

การให้บริการของห้องสมุดเป็นสิ่งจำเป็นมากในปัจจุบัน เมื่อห้องสมุดให้บริการตรงตามความต้องการของนักศึกษามากขึ้นเท่าใดนักศึกษาก็จะเข้าใช้บริการมากขึ้นเท่านั้น Spahr (2015) กล่าวว่า นักศึกษาระดับปริญญาตรีมีความต้องการสื่อสารสนเทศที่มากขึ้น รวมทั้งการเข้าใช้งานระบบฐานข้อมูล เมื่อห้องสมุดสามารถตอบสนองความต้องการของนักศึกษา จะส่งผลทำให้มีปริมาณผู้เข้าใช้บริการมากขึ้น จากงานวิจัยของ Murray, Ireland และ Hackathorn (2016) และ Herman (2016) พบว่า การเข้าใช้ห้องสมุดของนักศึกษาระดับปริญญาตรี ส่งผลโดยตรงต่ออัตราการคงอยู่ของนักศึกษาระดับชั้นปีที่ 1 และ 2 ของมหาวิทยาลัย และในทำนองเดียวกัน Soria, Fansen และ Nackerud (2013) พบว่า การใช้งานห้องสมุดของนักศึกษาระดับปริญญาตรีส่งผลต่อคะแนนสอบที่ดีขึ้นมากกว่านักศึกษาที่ไม่ได้เข้าใช้ห้องสมุด และทำให้อัตราการคงอยู่ของนักศึกษาดีขึ้น

ดังนั้น การศึกษาพฤติกรรมและปัจจัยของนักศึกษาต่อการเข้าใช้ห้องสมุด เป็นสิ่งสำคัญที่ต้องดำเนินการเพื่อห้องสมุดปรับปรุงการบริการในทุกๆ ด้านให้สามารถดึงดูดความสนใจของนักศึกษาในการเข้าใช้ห้องสมุด การบริหารทรัพยากรให้เหมาะสม หรือเพิ่มเติมสิ่งอำนวยความสะดวก เพื่อนำไปสู่ความสำเร็จในการศึกษาของนักศึกษาระดับปริญญาตรี ผู้วิจัยสนใจศึกษาพฤติกรรมของการเข้าใช้บริการห้องสมุด โดยการนำช่วงเวลาของการให้บริการของห้องสมุดที่เปิดให้บริการเวลาตั้งแต่ 8.00 น. ถึง 18.30 น. ทุกวันทำการของห้องสมุดมหาวิทยาลัย ตั้งแต่วันจันทร์ถึงวันศุกร์ยกเว้นวันหยุดเสาร์อาทิตย์และวันหยุดนักขัตฤกษ์ จึงได้ทำการวิเคราะห์ด้วยการจำแนกกลุ่มของผู้เข้าใช้บริการห้องสมุดมหาวิทยาลัยเพื่อวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษาด้วยเทคนิคเหมืองข้อมูล

วัตถุประสงค์

เพื่อเปรียบเทียบประสิทธิภาพการคัดแยกข้อมูลของอัลกอริทึม C5.0, Neural Network และ CART และวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษาให้เกิดความถูกต้องในการคัดแยกข้อมูล

ขอบเขตของการวิจัย

การวิจัยครั้งนี้ใช้ชุดข้อมูลการเข้าใช้บริการห้องสมุดของมหาวิทยาลัยราชภัฏเลย ผ่านประตูอัตโนมัติด้วยบัตร RFID ของนักศึกษาระดับปริญญาตรี ภาคปกติ ใช้อัตรนักศึกษาที่เข้าเรียน ตั้งแต่ปีการศึกษา 2555 – 2558 ที่เข้าใช้บริการตั้งแต่วันที่ 1 กุมภาพันธ์ ถึง 31 ตุลาคม พ.ศ. 2559 จำนวนรวม 79,953 ครั้ง

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

อัลกอริทึม C5.0

เป็นอัลกอริทึมที่ใช้เป็นตัวคัดแยกในรูปแบบของการสร้างต้นไม้ ซึ่งพัฒนามาจากอัลกอริทึม C4.5 ที่ทำการเพิ่มความถูกต้องของตัวคัดแยก ด้วยการวิธี Boosting ซึ่งเป็นการปรับค่าน้ำหนักของตัวอย่างที่เป็นข้อมูลมาทดสอบ โดยข้อมูลใดมีความแม่นยำมากอยู่แล้ว จะลดค่าน้ำหนักลงและไปเพิ่มค่าน้ำหนักให้ข้อมูลที่มีความแม่นยำน้อยซึ่งได้รวมฟังก์ชันต่าง ๆ เช่น ค่าความแปรปรวนของการคัดแยกผิด (Variable Misclassification Costs) ซึ่งจะมีการแยกค่าสูญหาย การคัดแยกของกลุ่มข้อมูลทำให้ลดการเกิดความแปรปรวนของการคัดแยกผิดได้ ในอัลกอริทึม C5.0 ยังให้

ความสำคัญในการกระจายน้ำหนักบนกลุ่มข้อมูล เพิ่มการรองรับชนิดข้อมูลใหม่ ๆ ซึ่งทำให้การประยุกต์ใช้งานสามารถทำได้ง่ายขึ้น (อุบลวรรณ กิจคณะ และคณะ, 2553)

ลำดับขั้นตอนการทำงานของอัลกอริทึม C5.0 มีดังนี้ เตรียมข้อมูลที่จะศึกษา โดยการจัดเรียงข้อมูลเป็นแถวและคอลัมน์ ซึ่งแถวเป็นค่าของข้อมูลที่น่ามาพิจารณาและคอลัมน์เป็นชื่อคุณสมบัติของข้อมูลนั้น ๆ สำหรับแต่ละคุณสมบัติ หาค่า Information Gain เพื่อเป็นการจัดลำดับความสำคัญ จากนั้นสร้างโหนดราก (Root node) จากค่า Information Gain มากที่สุด และสร้างกิ่งจากโหนดรากให้เป็นโหนดลูก (Children node) เมื่อมาสามารถสร้างกิ่งได้จะได้โหนดสุดท้ายเป็นผลลัพธ์ หรือ โหนดใบ (Leaf node)

การสร้างต้นไม้ด้วยการใช้ Gain Ratio วัดค่าความไม่บริสุทธิ์ของข้อมูล (Entropy) ดังสมการ (Ron Kohavi & Ross Quinlan, 1999)

$$E(S) = - \sum_{c=1}^n p(S_c) \times \log_2 p(S_c) \quad \text{--- (1)}$$

จากสมการ ค่า c_1 ถึง c_n เป็นกลุ่มข้อมูลย่อยของ S และ $p(S_c)$ คือ ความน่าจะเป็นของกลุ่มข้อมูล C ที่เกิดขึ้นใน S โดยการวัดค่า Information Gain เพื่อเป็นการสร้างลำดับ ทำได้ด้วยสมการ

$$Gain(S, V) = E(S) - \sum_{v \in \text{value}(V)} \frac{|S_v|}{|S|} \times E(S_v) \quad \text{--- (2)}$$

ค่า Gain ที่ได้มีค่าอคติ (Bias) ที่เป็นส่วนหนึ่งในการคัดแยกกลุ่มข้อมูล การลดค่าอคติ สามารถทำได้โดยการคำนวณในสมการ

$$SplitInfo(S, V) = \sum_{i=1}^m - \frac{S_i}{S} \times \log_2 \frac{S_i}{S} \quad \text{--- (3)}$$

การหาค่า Gain Ratio เป็นการแบ่งค่า Gain ด้วยค่าการ Split Info ดังสมการ

$$GainRatio(S, V) = \frac{Gain(S, V)}{SplitInfo(S, V)} \quad \text{--- (4)}$$

อัลกอริทึมโครงข่ายประสาทเทียม (Artificial Neural Network)

แบบจำลองทางคณิตศาสตร์ สำหรับการประมวลผลสารสนเทศ เพื่อจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ ให้มีความสามารถในการเรียนรู้ การจัดจำรูปแบบ (Pattern Recognition) และการอุปมาความรู้ (Knowledge Deduction) เช่นเดียวกับความสามารถที่มีในสมองของมนุษย์ โดยสามารถนำมาใช้ในการพยากรณ์ (Forecasting) ซึ่งเป็นการคาดคะเนลักษณะต่าง ๆ หรือเป็นการประเมินสิ่งที่จะเกิดขึ้นในอนาคตด้วยการคาดการณ์ล่วงหน้า โดยใช้ข้อมูลที่ผ่านมาในอดีต การพยากรณ์จึงมีบทบาทสำคัญในทุก ๆ ภาคส่วนที่จะช่วยให้ได้ข้อมูลในอนาคต เพื่อประกอบการวางแผนสนับสนุนการตัดสินใจในเรื่องต่างๆ (ภัทร์พงศ์ พงศ์ภัทรกานต์, 2552)

อัลกอริทึม CART

การวิเคราะห์อัลกอริทึม CART เป็นสถิติอนพาราเมตริก ที่ใช้ในการวิเคราะห์เพื่อแบ่งกลุ่มข้อมูลออกเป็นกลุ่มๆ โดยที่ตัวแปรอิสระเป็นตัวแปรเชิงกลุ่มหรือตัวแปรเชิงปริมาณก็ได้และเมื่อตัวแปรตามเป็นตัวแปรเชิงกลุ่มจะใช้เทคนิคการแบ่งกลุ่ม Classification Tree แต่ถ้าตัวแปรตามเป็นตัวแปรเชิงปริมาณจะใช้เทคนิคการแบ่งกลุ่ม Regression Tree ซึ่งทั้ง 2 กรณีนี้ มีขั้นตอนการวิเคราะห์ 4 ขั้นตอน คือ Tree Building, Stopping, Tree Pruning และ Optimal Tree Selection (ภัทร์พงศ์ พงศ์ภัทรกานต์, 2552)

ขั้นตอนและวิธีการดำเนินงาน

ข้อมูลการใช้บริการห้องสมุดของนักศึกษา มีลักษณะของข้อมูลที่ใช้ในการทดสอบจำนวน 9 ปัจจัยพื้นฐาน คือ วันที่เข้าใช้บริการ ช่วงเวลา เพศ คณะ ชั้นปี จังหวัดที่เกิด หมู่เลือด จำนวนพี่น้อง และเกรดเฉลี่ยสะสม มีจำนวนรวม 79,953 ชุดข้อมูล ลักษณะข้อมูลเป็นแบบแบ่งกลุ่มที่ถูกนำไปใช้ในการคัดแยก โดยมีวิธีการทดลอง ดังนี้

ผู้วิจัยได้นำอัลกอริทึม C5.0, Neural Network และ CART โดยใช้ชุดข้อมูลการเข้าใช้บริการห้องสมุดของมหาวิทยาลัยราชภัฏเลย ผ่านประตูอัตโนมัติด้วยบัตร RFID ของนักศึกษาระดับปริญญาตรี ภาคปกติ ใช้รหัสนักศึกษาที่เข้าเรียน ตั้งแต่ปีการศึกษา 2555 – 2558 ที่เข้าใช้บริการตั้งแต่วันที่ 1 กุมภาพันธ์ ถึง 31 ตุลาคม พ.ศ. 2559 จำนวนรวม 79,953 ชุดข้อมูลรวมนักศึกษาที่เข้าใช้ซ้ำแต่คัดเลือกเฉพาะที่เข้าออกไม่น้อยกว่า 30 นาที โดยมีปัจจัยพื้นฐาน 9 ปัจจัย โดยขั้นตอนการทำงานประกอบไปด้วย 3 ขั้นตอนหลักดังนี้

1) ขั้นตอนการเตรียมข้อมูล (Data Preprocessing) เมื่อได้ข้อมูลแล้ว ให้เตรียมข้อมูลเพื่อให้พร้อมที่จะนำไปทำการคัดกรองภายใต้หลักการดำเนินงานของเหมืองข้อมูลซึ่งแบ่งเป็นขั้นต่าง ๆ คือ การทำข้อมูลให้สมบูรณ์ (Data Cleaning) การคัดเลือกข้อมูล (Data Selection) และการปรับเปลี่ยนรูปแบบข้อมูล (Data Transformation) ให้มีความเหมาะสม

2) ขั้นตอนการหากฎจำแนก (Classification Rules) การจำแนกประเภทข้อมูลเป็นเทคนิคในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่างๆ โดยทำการสำรวจรายการในฐานข้อมูล เพื่อแยกแยะให้อยู่ในหมวดที่ได้กำหนดไว้ล่วงหน้า เทคนิคที่ใช้ในการจำแนกประเภทข้อมูลมีมากมายแต่มีเทคนิคที่นิยมกันมาก คือ เทคนิคการใช้ต้นไม้ตัดสินใจ ซึ่งเป็นเครื่องมือที่ช่วยกำหนดขอบเขตของปัญหาและช่วยสร้างทางเลือกที่เป็นไปได้ในการแก้ปัญหา ซึ่งอัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจที่นิยม คือ C5.0 เป็นต้น

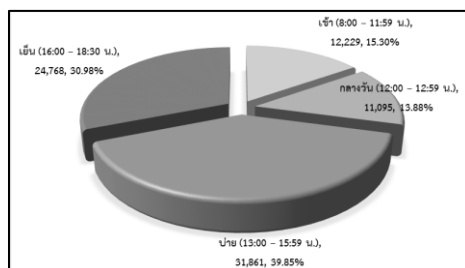
3) ขั้นตอนการคัดกรองแอททริบิวต์ (Attribute Filtering) ทำการคัดกรองแอททริบิวต์เพื่อให้ได้แอททริบิวต์ที่จำเป็นที่สุด สำหรับการจำแนกประเภทข้อมูล มีหลักการในการทำงานคือ การใช้เทคนิคการจำแนกข้อมูล โดยใช้ อัลกอริทึม C5.0, Neural Network และ CART เพื่อหาประสิทธิภาพในการจำแนกข้อมูลและการสร้างความสัมพันธ์ของแอททริบิวต์ได้ โดยใช้ปัจจัยที่ใช้ในการทดสอบ ดังตารางที่ 1

ตารางที่ 1 แสดงปัจจัยหรือแอททริบิวต์ที่ใช้การทดสอบ

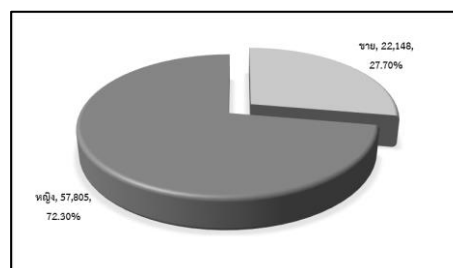
ที่	ปัจจัยหรือแอททริบิวต์	สถานะ
1	วันที่เข้าใช้บริการ (InDate)	None
2	ช่วงเวลา (IN)	Input
3	เพศ (Sex)	Input
4	คณะ (Faculty)	Output
5	ชั้นปี (Level)	Input
6	จังหวัดที่เกิด (Province)	Input
7	หมู่เลือด (Blood)	Input
8	จำนวนพี่น้อง (Family)	Input
9	เกรดเฉลี่ยสะสม (GPA)	Input

ผลการศึกษา อภิปรายผล และข้อเสนอแนะ

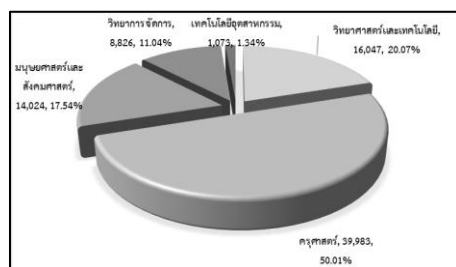
ผลการศึกษา ได้ข้อมูลสถิติพื้นฐานของจำนวนครั้งที่มาใช้บริการห้องสมุดของนักศึกษา จำแนกตามปัจจัย ดังภาพที่ 1-6



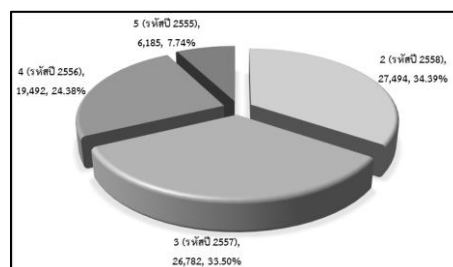
ภาพที่ 1 จำแนกตามช่วงเวลา (IN)



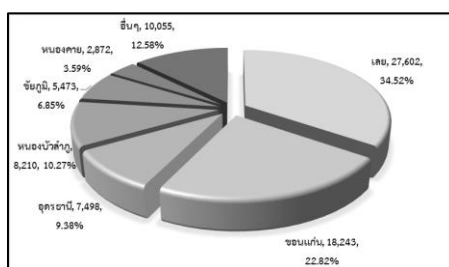
ภาพที่ 2 จำแนกตามเพศ (Sex)



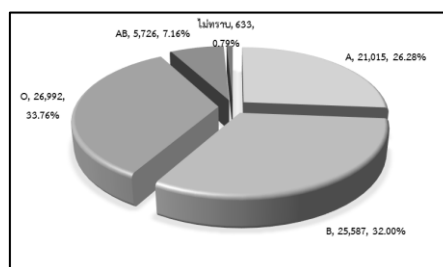
ภาพที่ 3 จำแนกตามคณะ (Faculty)



ภาพที่ 4 จำแนกตามชั้นปี (Level)



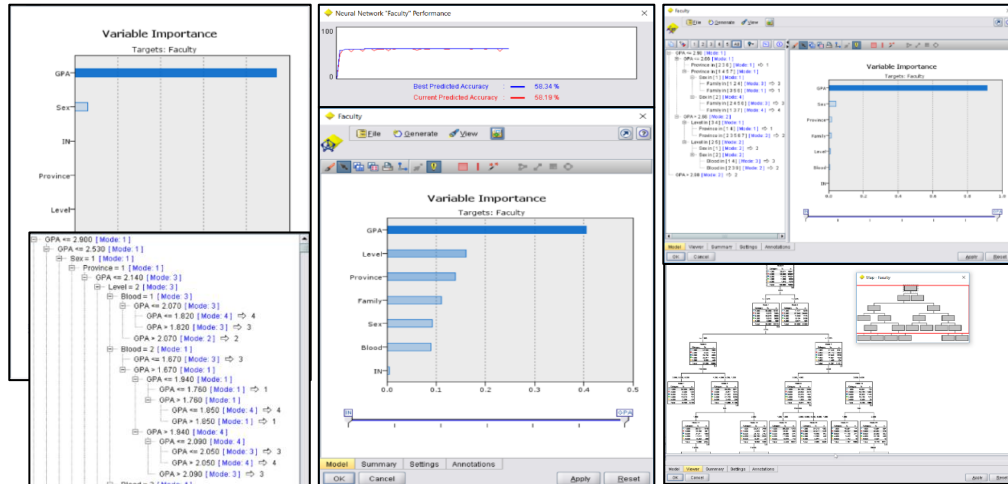
ภาพที่ 5 จำแนกตามจังหวัดที่เกิด (Province)



ภาพที่ 6 จำแนกตามหมู่เลือด (Blood)

จากภาพที่ 1 การจำแนกตามช่วงเวลา พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดในช่วงเวลาบ่าย 13:00-15:59 น. มากที่สุด คิดเป็นร้อยละ 39.85 จากภาพที่ 2 การจำแนกตามเพศ พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดเป็นเพศหญิงมากกว่าเพศชาย คิดเป็นร้อยละ 72.30 จากภาพที่ 3 การจำแนกตามคณะ พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดสังกัดคณะครุศาสตร์ มากที่สุด คิดเป็นร้อยละ 50.01 จากภาพที่ 4 การจำแนกตามชั้นปี พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดชั้นปีที่ 2 รหัสปี 2558 มากที่สุด คิดเป็นร้อยละ 34.39 จากภาพที่ 5 การจำแนกตามจังหวัดที่เกิด พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดเกิดที่จังหวัดเลย มากที่สุด คิดเป็นร้อยละ 34.52 และจากภาพที่ 6 การจำแนกตามหมู่เลือด พบว่า นักศึกษาเข้ามาใช้บริการห้องสมุดมีหมู่เลือดโอ (O) มากที่สุด คิดเป็นร้อยละ 33.76

ผู้วิจัยได้นำอัลกอริทึม C5.0, Neural Network และ CART มาทำการประมวลผลและจำแนกชุดข้อมูล นักศึกษาเข้ามาใช้บริการห้องสมุด ดังภาพที่ 7



ภาพที่ 7 การวิเคราะห์ชุดข้อมูลโดยใช้อัลกอริทึม C5.0, Neural Network และ CART

ผลการเปรียบเทียบประสิทธิภาพในการจำแนกชุดข้อมูลการให้บริการห้องสมุดของนักศึกษา ปรากฏผลดังตารางที่ 2

ตารางที่ 2 แสดงผลการเปรียบเทียบประสิทธิภาพในการจำแนกชุดข้อมูลการให้บริการห้องสมุดของนักศึกษา

อัลกอริทึม	ค่าความถูกต้อง (Accuracy)	ระยะเวลาในการประมวลผล
C5.0	97.78 %	2 วินาที
Neural Network	58.19 %	6 วินาที
CART	58.86 %	4 วินาที

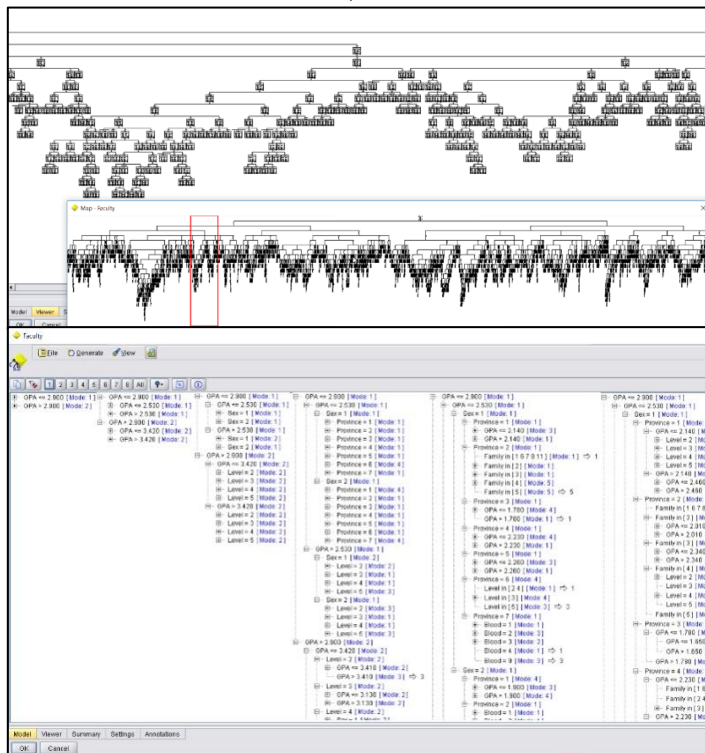
จากตารางที่ 2 พบว่าอัลกอริทึม C5.0 มีประสิทธิภาพในการจำแนกชุดข้อมูลการให้บริการห้องสมุดของนักศึกษาสูงที่สุด คือ มีค่าความถูกต้อง (Accuracy) ร้อยละ 97.78 โดยมีระยะเวลาในการประมวลผลชุดข้อมูลน้อยที่สุดที่ 2 วินาที ดังนั้น ผู้วิจัยจึงได้ทำการเลือกอัลกอริทึม C5.0 มาวิเคราะห์ปัจจัยในการให้บริการห้องสมุดของนักศึกษาต่อไป

ผลการวิเคราะห์ปัจจัยในการให้บริการห้องสมุดของนักศึกษา โดยใช้อัลกอริทึม C5.0 ซึ่งได้แสดงปัจจัยสำคัญจำนวน 3 ปัจจัย ดังตารางที่ 3

ตารางที่ 3 แสดงปัจจัยที่มีอิทธิพลต่อการให้บริการห้องสมุดของนักศึกษาที่ส่งผลตามคณะ

อันดับ	ปัจจัย	มีอิทธิพลร้อยละ
1	เกรดเฉลี่ยสะสม (GPA)	93.8
2	เพศ (Sex)	6.0
3	ช่วงเวลา (IN)	0.2
ไม่ส่งผล	ชั้นปี จังหวัดที่เกิด หมู่เลือด จำนวนพี่น้อง	ไม่มี

จากตารางที่ 3 สรุปได้ว่าเกรดเฉลี่ยสะสมมีอิทธิพลต่อการใช้บริการห้องสมุดของนักศึกษาสูงที่สุด คิดเป็นร้อยละ 93.8 และเพศ มีอิทธิพลต่อการใช้บริการห้องสมุดของนักศึกษา รองลงมาที่ร้อยละ 6.0



ภาพที่ 9 ความสัมพันธ์ของปัจจัยโดยใช้อัลกอริทึม C5.0

จากภาพที่ 8-9 แสดงความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับการใช้บริการห้องสมุดของนักศึกษาในรูปแบบต้นไม้ตัดสินใจ ซึ่งจะมีความลึก 21 ชั้น (Layer) ซึ่งสามารถยกตัวอย่างการรายงานผลที่ได้จากการทดลองมาวิเคราะห์เพื่อหาความสัมพันธ์ของปัจจัยที่ส่งผลต่อคณะ ดังตัวอย่าง

$GPA \leq 2.90 [Mode:1] \rightarrow GPA > 2.53 [Mode:1] \rightarrow Sex = 2 [Mode:1] \rightarrow INin [1 \ 2 \ 4] [Mode:4] \rightarrow 4$

แปลความหมายได้ว่าเกรดเฉลี่ยที่น้อยกว่าหรือเท่ากับ 2.90 และ มากกว่า 2.53 เพศหญิง (2) เข้าใช้บริการช่วงเวลาเช้า (1) เย็น (2) และเย็น (4) มีโอกาสเป็นนักศึกษาสังกัดคณะวิทยาการจัดการ (4) เป็นต้น

การนำไปใช้ประโยชน์

การศึกษาพฤติกรรมและปัจจัยของนักศึกษาต่อการใช้บริการห้องสมุด เพื่อนำข้อมูลจากการศึกษาไปใช้สำหรับการปรับปรุงการให้บริการทุกๆ ด้านภายในห้องสมุดทำให้เกิดการดึงดูดและความสนใจแก่นักศึกษาในการที่เข้าใช้บริการ ในงานวิจัยนี้ทำการศึกษาข้อมูลเบื้องต้นสำหรับการจัดเตรียมข้อมูลเพื่อนำไปวิเคราะห์ทางเหมืองข้อมูลด้วยอัลกอริทึม C5.0 เพื่อหาปัจจัยที่เกี่ยวข้องของผู้ใช้บริการห้องสมุดโดยจากการนำข้อมูลที่ได้นำมาวิเคราะห์และหาความสัมพันธ์ของปัจจัยที่เกี่ยวข้องซึ่งเป็นแนวทางในการประชาสัมพันธ์และส่งเสริมนักศึกษาเข้ามาใช้บริการห้องสมุดผ่านคณะที่สังกัดได้ โดยเฉพาะเกรดเฉลี่ยมีผลอย่างมากในการเข้ามาใช้บริการ นักศึกษาที่มีเกรดสูงมีแนวโน้มการเข้าใช้ห้องสมุดมากกว่านักศึกษาที่มีเกรดต่ำ ดังนั้น ห้องสมุดควรเน้นไปที่การเปิดบริการหรือเชิญชวนให้นักศึกษาที่มีเกรด

น้อยเข้าห้องสมุดมากขึ้น ห้องสมุดควรคิดกิจกรรมส่งเสริมใหม่เพิ่มมากขึ้นเพื่อให้ นักศึกษามีความสนใจในการเข้าใช้ ห้องสมุดแนวทางการศึกษาต่อในอนาคตจะนำข้อมูลนักศึกษาของทั้งมหาวิทยาลัยมาทำการวิเคราะห์เพื่อหาปัจจัยที่ ส่งผลต่อการไม่เข้ามาใช้บริการในห้องสมุดตลอดปี พ.ศ. 2559 เพื่อวางแผนในการประชาสัมพันธ์และกระตุ้นให้ นักศึกษาที่ไม่เข้ามาใช้บริการห้องสมุดได้เห็นประโยชน์ต่อไป

รายการอ้างอิง

- ธารรัตน์ อุดมชัยบรรเจิด และผุสดี บุญรอด. (2557). การสร้างแบบจำลองการพยากรณ์ราคากองทุนรวมโดยใช้ เหมืองข้อมูลและโครงข่ายประสาทเทียม. รายงานการประชุมระดับชาติ *Proceeding of The 10th National Conference on Computing and Information TechnologyNCCIT 2014*(น.890-895). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- ไพฑูรย์ สีนารัตน์. (2557). จำเป็นต้องปฏิวัติการศึกษาไทย (พิมพ์ครั้งที่ 4). กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ภัทรพงศ์ พงศ์ภัทรกานต์. (2552).การเปรียบเทียบการจำแนกข้อมูลของแบบจำลอง CART, SVM, C5.0 และแบบ ผลผสมผสานกัน. รายงานการประชุมระดับชาติ *Proceeding ofThe 5th National Conference on Computing and Information TechnologyNCCIT 2009*(น.1101-1106).กรุงเทพฯ: มหาวิทยาลัย เทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- อุบลวรรณ กิจคณะ, วนิดา แก่นอากาศ และวิโรจน์ ทวีปวรเดช. (2553). การเรียนรู้รูปแบบรหัสพันธุกรรมเพื่อจำแนก ชนิดของโรคมะเร็งเม็ดเลือดขาวด้วยอัลกอริทึม C5.0. *วารสารวิจัยมหาวิทยาลัยขอนแก่น*, 10(3),21-28.
- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer, Heidelberg.
- Azzalini, A.,&Scarpa, B. (2012).*Data Analysis and Data Mining: an Introduction*. Oxford University Press, New York.
- Bay, S. D.,&Pazzani, M. J. (2001).Detecting Group Differences: Mining Contrast Sets.*Data Mining and Knowledge Discovery*, 5, 213–246.
- Herman, S. W. (2016). Positive Correlation between Academic Library Services and High-Impact Practices for Student Retention.*Evidence Based Library & Information Practice*, 11(1), 85-87.
- Larose, D. T.,& Larose, C. D. (2015).*Data Mining and Predictive Analytics*. John Wiley & Sons, New Jersey.
- Murray, A., Ireland, A., &Hackathorn, J. (2016).The Value of Academic Libraries: Library Services as a Predictor of Student Retention. *College & Research Libraries*, 77(5), 631-642.
- Ron Kohavi,& Ross Quinlan. (1999). *Decision Tree Discovery*.Retrieved from <http://ai.stanford.edu/~ronnyk/treesHB.pdf>
- Soria, K. M., Fansen, J., &Nackerud, S. (2013). Library Use and Undergraduate Student Outcomes: New Evidence for Students' Retention and Academic Success. *Portal: Libraries &The Academy*, 13(2), 147-164.
- Spahr, K. (2015). Influences on Undergraduate Business Students' Perceptions about the Adequacy of Library Information Resources. *Behavioral & Social Sciences Librarian*, 34(4), 214-229.