

การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัด เพื่อพัฒนาศักยภาพของนักศึกษา

APPLIED DATA MINING TECHNIQUES TO ANALYZE PRE-TEST AND SKILL-TEST FOR INCREASED STUDENT POTENTIAL

สุธีรา วงศ์อนันทรัพย์* ต๋องใจ แยมพกา อรวรรณ มุสิกะ

Suteera Vongansup, Tongjai Yampaka and Orawan Moosika

สาขาวิชาวิทยาการคอมพิวเตอร์

คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ

suteera_vo@cpc.ac.th โทร.0-2692-2360-4 ต่อ 406

บทคัดย่อ

สถาบันการศึกษามีหน้าที่ในการพัฒนาบัณฑิตให้มีศักยภาพตามความต้องการของตลาด และมีความรู้ความสามารถตามความถนัดของบัณฑิต งานวิจัยนี้นำผลการทำแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัดของกลุ่มตัวอย่างมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูลผลการวิเคราะห์แบบประเมินความรู้ก่อนเรียน ผลลัพธ์ที่ได้สามารถจำแนกนักศึกษา โดยใช้เทคนิคการจัดกลุ่มด้วยอัลกอริทึม K-mean กำหนดการจัดกลุ่มนักศึกษาได้เป็น 2 กลุ่ม กลุ่มนักศึกษาที่ด้อย (W) และกลุ่มนักศึกษาที่เด่น (S) จากการทำนายของชุดข้อมูลสอนให้ความถูกต้องเฉลี่ย 80.90 ผลการวิเคราะห์แบบทดสอบความถนัด ผลลัพธ์ที่ได้จำแนกผู้ตอบแบบสอบถามออกเป็นสาขาวิชา 4 สาขา ได้แก่ สาขาการเขียนโปรแกรม (P) สาขาเครือข่ายคอมพิวเตอร์ (N) สาขาฐานข้อมูล (D) และสาขาการผลิตซอฟต์แวร์ (S) สามารถสร้างเป็นต้นไม้ตัดสินใจ ทำให้ได้กฎความสัมพันธ์ของต้นแบบ 7 กฎ จากการทำนายของชุดข้อมูลสอนให้ร้อยละความถูกต้องเฉลี่ย 79.42 ผลจากงานวิจัยนี้สามารถนำแบบรูปที่ได้ไปใช้ในการทำนายคุณสมบัติของนักศึกษา เพื่อเป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น ลดความเสี่ยงของนักศึกษาที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์ และได้แบบทดสอบวัดความถนัดเฉพาะสาขาทางวิทยาการคอมพิวเตอร์ เพื่อพัฒนาศักยภาพของนักศึกษาตามความถนัดมีโอกาสได้งานทำหลังจบการศึกษามากขึ้น

คำสำคัญ: เหมืองข้อมูล, การประเมินความรู้ก่อนเรียน, การทดสอบความถนัด

Abstract

Educational institutions play important roles to graduates development and knowledge and ability improvement to meet market's need. The purposes of this research were pre-test on computer science and proficiency test. The test results were analyzed by data mining. As for pre-test assessment analysis, the results could distinguish students by using K-mean algorithm. The 27 students who did the test were divided into 2 groups. The Cluster 0 was 13 weak students (W); 48 percent. The Cluster 1 was 14 strong students (S); 52 percent. The average correctness was 80.90 percent. From the proficient test analysis, the results divided students into 4 majors; Programming (P), Computer Network (N), Database (D), and Software Engineering (S). The decision tree could be done, and also the 7 model relational rules. The average correctness was 79.42 percent. The assessment

results would be used for managing students and would be used to promote students' proficiencies, analysis, or data arranging.

Keywords: Data Mining, Pre Test, Skill Test

1. บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

การศึกษาระดับอุดมศึกษามุ่งเน้นให้นักศึกษาเรียนรู้ในศาสตร์เฉพาะ ซึ่งต่างจากระบบการศึกษาขั้นพื้นฐานที่ให้นักเรียนมีความรู้ในสายสามัญ รวมถึงรูปแบบการเรียนการสอนที่ต่างจากเดิม นักศึกษามีพื้นฐานความรู้ที่ต่างกัน สภาพแวดล้อมที่แปลกใหม่ จากสาเหตุดังกล่าวมีผลกระทบต่อการทำงานของนักศึกษา ส่งผลให้นักศึกษาในชั้นปีที่ 1 มีผลการเรียนที่อยู่ในเกณฑ์ไม่ดีและยังส่งผลกระทบต่อปรัยวิชาที่มีความต่อเนื่องในปีถัดไป เมื่อผลการเรียนที่อยู่ในเกณฑ์ที่ไม่ดีสะสมส่งผลให้นักศึกษามีความเสี่ยงที่จะไม่สำเร็จการศึกษาและต้องออกกลางคัน

ในส่วนของนักศึกษาในชั้นปีอื่นก็ประสบกับปัญหาการเพิ่มความรู้และพัฒนาศักยภาพตามความถนัดของตนเอง เรียนไปโดยไม่รู้ตัวตนเองมีความถนัดทางด้านใด มีข้อเด่น หรือมีข้อด้อยในสาขาวิชาอะไรบ้างงานวิจัยนี้จึงได้ทำการสร้างแบบทดสอบประเมินความรู้ก่อนเรียน และประเมินความถนัดของนักศึกษา และนำผลการทำแบบทดสอบของนักศึกษามาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูล ซึ่งเทคนิคการทำเหมืองข้อมูลนี้เป็นเทคนิคที่ได้รับการยอมรับในการค้นหาสารสนเทศหรือความรู้ที่ซ่อนอยู่ในสารสนเทศเพื่อนำมาใช้ประโยชน์ในการตัดสินใจ งานวิจัยนี้มีกลุ่มตัวอย่างอยู่ 2 กลุ่ม ดังนี้

ตัวอย่างกลุ่มที่ 1 นักศึกษาชั้นปีที่ 1 ทำแบบทดสอบความรู้พื้นฐานรายวิชาทางคอมพิวเตอร์ที่นักศึกษาจะต้องเรียนในหลักสูตรวิทยาการคอมพิวเตอร์ ผลการทดสอบในนักศึกษากลุ่มนี้จะนำมาจัดกลุ่มเพื่อคัดกรองเด็กที่ได้ผลการทดสอบน้อยในรายวิชาใดจะมีแผนในการปรับพื้นฐานเด็กในกลุ่มนี้ ส่วนกลุ่มเด็กที่ได้ผลการทดสอบโดดเด่นในรายวิชาใดจะมีแผนในการส่งเสริมนักศึกษาที่มีโดดเด่นในรายวิชานั้นให้นักศึกษาเข้าประกวดผลงาน การสอบแข่งขัน ที่ตรงกับรายวิชาที่มีความโดดเด่น

ตัวอย่างกลุ่มที่ 2 ผู้ประกอบอาชีพทางคอมพิวเตอร์ ผลการทดสอบจะนำมาวิเคราะห์แยกกลุ่มออกตามความถนัด 4 สาขาดังนี้ สาขาฐานข้อมูล (Database) สาขาการเขียนโปรแกรม (Programming) สาขาเครือข่ายคอมพิวเตอร์ (Computer Network) และสาขาด้านการผลิตซอฟต์แวร์ (Software Engineering) โดยผลการทดสอบจะได้ต้นแบบที่จะใช้ทำนายคุณสมบัติของนักศึกษาในชั้นปีที่ 4 เพื่อหาความถนัดของนักศึกษาแต่ละคน แต่ละกลุ่มจะได้รับการเพิ่มศักยภาพตามสาขาที่ตนถนัด โดยมีแผนในการส่งนักศึกษาเข้าอบรมและสอบวัดมาตรฐานความรู้ในแต่ละสาขาเพื่อให้นักศึกษาได้มีใบประกาศนียบัตรใช้สำหรับนำไปสมัครงาน

ผลจากการทำแบบทดสอบของกลุ่มตัวอย่างจะนำมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อวิเคราะห์และประเมินความรู้ ความถนัดของนักศึกษารววิเคราะห์ข้อมูลผลการทำแบบทดสอบนักศึกษาชั้นปีที่ 1 จะใช้เทคนิคการจัดกลุ่มด้วยอัลกอริธึม K-mean เพื่อแยกนักศึกษาออกเป็นกลุ่มที่ด้อยและกลุ่มที่มีข้อเด่นการวิเคราะห์ข้อมูลผลการทำแบบทดสอบผู้ประกอบอาชีพทางคอมพิวเตอร์ สำหรับแบ่งสาขาตามความถนัด 4 ด้าน ผลที่ได้จะเป็นต้นแบบเพื่อใช้ทำนายข้อมูลนักศึกษาในชั้นปีที่ 4 เพื่อเพิ่มศักยภาพตามสาขาที่นักศึกษาถนัด สนับสนุนให้นักศึกษาได้รับการอบรมและสอบวัดมาตรฐานความรู้ตามสาขานั้นๆ

1.2 การทบทวนวรรณกรรมที่เกี่ยวข้อง

ที่ผ่านมาได้มีงานวิจัยที่ใช้เทคนิคการทำเหมืองข้อมูลในการวิเคราะห์ข้อมูลเกี่ยวกับการศึกษา เช่น ระบบวิเคราะห์ข้อมูลนักเรียนเชิงบูรณาการโดยใช้เทคโนโลยีเหมืองข้อมูลกรณีศึกษา โรงเรียนท่าทรายวิทยา (ศุภภัค, 2553) ได้นำเทคโนโลยี

เหมืองข้อมูลมาจัดกลุ่มนักเรียนตามระดับผลการเรียนเพื่อให้ครู-อาจารย์นำข้อมูลมาคัดกรองนักเรียน เพื่อความสะดวกในการวางแผนดูแลช่วยเหลือนักเรียนกลุ่มที่มีปัญหา และส่งเสริมนักเรียนที่มีความสามารถพิเศษ

การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาในระดับอุดมศึกษา (ชุตินา, 2010) เพื่อสร้างตัวแบบสำหรับปัจจัยที่มีผลต่อการตัดสินใจเลือกสาขาวิชาเรียนของนักศึกษาโดยการใช้เทคนิคการทำเหมืองข้อมูลด้วยวิธี K-fold cross validation พบว่าสามารถบอกตัวแปรสำคัญที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษาในระดับอุดมศึกษาได้

การประยุกต์ใช้เอพี-โกรธกับงานแนะแนวการศึกษาต่อในระดับอุดมศึกษา (ณรงค์ศักดิ์, 2011) งานวิจัยนี้ได้ประยุกต์กฎความสัมพันธ์ซึ่งเป็นหนึ่งในเทคนิคการทำเหมืองข้อมูลเพื่อช่วยแนะแนวการศึกษาให้กับนักศึกษาให้กับนักเรียนที่จะศึกษาต่อในระดับอุดมศึกษาหากความสัมพันธ์และสร้างต้นแบบมาใช้ทำนาย เพื่อแนะแนวการศึกษาต่อของนักเรียนที่กำลังจะเรียนต่อในสถาบันอุดมศึกษาได้เหมาะสมกับผลการเรียนของตนเอง

ประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพโดยอัตโนมัติด้วยเทคนิคเหมืองข้อมูล (ชัชชฎา, 2556) งานวิจัยนี้ใช้ข้อมูลภาวะมีงานทำของบัณฑิต และข้อมูลระเบียบประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษาซึ่งผลการทดลองนี้สามารถนำไปช่วยในการพิจารณาปรับปรุงการเรียนการสอน และช่วยวางแผนการเรียนให้กับนักศึกษาได้

จากงานวิจัยที่ผ่านมาเป็นการหาความสัมพันธ์หรือกฎที่ซ่อนอยู่โดยใช้เทคนิคการทำเหมืองข้อมูลมาวิเคราะห์ข้อมูลทางการศึกษาเพื่อใช้สนับสนุนการตัดสินใจ ไม่ว่าจะเป็นการแยกกลุ่มเพื่อดูแลและติดตามผลการเรียน หรือการสร้างต้นแบบใช้ในการทำนายคุณลักษณะของนักศึกษาเพื่อใช้ประกอบในการเลือกเรียนต่อในระดับอุดมศึกษา แต่งานวิจัยดังกล่าวยังไม่มีการสร้างต้นแบบสำหรับการแยกกลุ่มนักศึกษาตามความถนัดทางวิชาวิทยาการคอมพิวเตอร์ ทางคณะวิจัยจึงได้นำแนวคิดการทำเหมืองข้อมูลมาใช้ในการแยกกลุ่มเด็กตามความรู้ และตามความถนัดทางวิชาทางวิทยาการคอมพิวเตอร์

1.3 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อนำผลการทดสอบมาใช้ในการวางแผนในการเรียนการสอน และการดูแลนักศึกษาได้ตรงเป้าหมายมากขึ้น ช่วยลดความเสี่ยงที่นักศึกษาจะย้ายสาขา มีผลการเรียนตกต่ำ ส่งผลให้ไม่สำเร็จการศึกษา ผลการทดสอบมาพัฒนาศักยภาพของนักศึกษาตามความถนัดและส่งเสริมให้นักศึกษาได้รับการอบรมเพิ่มเติม พร้อมทั้งกลุ่มนักศึกษาที่สอบวัดมาตรฐานผ่านและมีประกาศนียบัตรเฉพาะสาขาไปสมัครงานมีโอกาที่จะได้งานทำสูงขึ้น

การทดสอบวิเคราะห์และประเมินความรู้ ความถนัดของนักศึกษาจะสามารถเป็นแนวทางในการดูแลนักศึกษาบนพื้นฐานข้อมูลที่มาจากการทดสอบและประเมินนักศึกษา ทำให้อาจารย์ผู้สอนและอาจารย์ที่ปรึกษาวางแผนในการเรียนการสอน และการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น ลดความเสี่ยงของนักศึกษาที่จะไม่สำเร็จการศึกษา ในส่วนการพัฒนาศักยภาพของนักศึกษาตามความถนัดและส่งเสริมให้นักศึกษาได้รับการอบรมเพิ่มเติม พร้อมทั้งกลุ่มนักศึกษาที่สอบวัดมาตรฐานผ่านและมีประกาศนียบัตรเฉพาะทาง ไปสมัครงานทำให้นักศึกษาที่เรียนจบมีโอกาที่จะได้งานทำในภาวะที่มีการแข่งขันที่สูง

2. วิธีการทดลองหรือวิธีการศึกษา

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำผลการทำแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัดแยกสาขาวิชา การวิเคราะห์ข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูลเพื่อหาแบบรูปของข้อมูล และนำแบบรูปดังกล่าวมาใช้ในการดูแลนักศึกษาและพัฒนาศักยภาพนักศึกษาแบ่งออกเป็น 2 ส่วนดังนี้

2.1 กลุ่มประชากรทดสอบ

กลุ่มที่ 1 นักศึกษาชั้นปีที่ 1 ทำแบบทดสอบความรู้พื้นฐานรายวิชาทางคอมพิวเตอร์ที่นักศึกษาจะต้องเรียนในหลักสูตรวิทยาการคอมพิวเตอร์ ผลการทดสอบในนักศึกษากลุ่มนี้จะนำมาจัดกลุ่มเพื่อคัดกรองเด็กที่ได้ผลการทดสอบน้อยในรายวิชาใดจะมีแผนในการปรับพื้นฐานเด็กในกลุ่มนี้ ส่วนกลุ่มเด็กที่ได้ผลการทดสอบโดดเด่นในรายวิชาใด จะมีแผนในการส่งเสริมนักศึกษาที่มีโดดเด่นในรายวิชานั้นให้นักศึกษาเข้าประกวดผลงาน การสอบแข่งขัน ที่ตรงกับรายวิชาที่มีความโดดเด่น

กลุ่มที่ 2 ผู้ประกอบอาชีพทางคอมพิวเตอร์ ผลการทดสอบจะนำมาวิเคราะห์แยกกลุ่มออกตามความถนัด 4 สาขาดังนี้ สาขาฐานข้อมูล (Database) สาขาการเขียนโปรแกรม (Programming) สาขาเครือข่ายคอมพิวเตอร์ (Computer Network) และสาขาด้านการผลิตซอฟต์แวร์ (Software Engineering) โดยผลการทดสอบจะได้ต้นแบบที่จะใช้ทำนายคุณสมบัติของนักศึกษาในชั้นปีที่ 4 เพื่อหาความถนัดของนักศึกษาแต่ละคน แต่ละกลุ่มจะได้รับการเพิ่มศักยภาพตามสาขาที่ตนถนัด โดยมีแผนในการส่งนักศึกษาเข้าอบรมและสอบวัดมาตรฐานความรู้ในแต่ละสาขาเพื่อให้นักศึกษาได้มีใบประกาศนียบัตรใช้สำหรับนำไปสมัครงาน

2.2 การวิเคราะห์ข้อมูลผลการทดสอบประเมินความรู้ก่อนเรียน

จะใช้เทคนิคการจัดกลุ่มข้อมูลด้วยอัลกอริธึม K-mean ในการจัดกลุ่มนักศึกษาที่ทำแบบทดสอบประเมินความรู้ก่อนเรียนเป็นเทคนิคที่ใช้จำแนกหรือแบ่ง Case (หมายถึง คน สัตว์ สิ่งของหรือ องค์กร ฯลฯ) หรือแบ่งตัวแปรออกเป็นกลุ่มย่อยๆ ตั้งแต่ 2 กลุ่มขึ้นไป โดยหลักการจัดกลุ่มคือ Case ที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือคล้ายกัน ส่วน Case ที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน ในงานวิจัยนี้ใช้ K-mean ในการจัดกลุ่มนักศึกษาออกเป็น 2 กลุ่มคือ กลุ่มเด่น (S) และกลุ่มด้อย (W)

ผลจากการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA โดยใช้เทคนิคการจัดกลุ่มด้วยอัลกอริธึม K-mean โดยกำหนดการจัดกลุ่มนักศึกษาเป็น 2 กลุ่มจากนักศึกษาที่ทำแบบทดสอบ 27 คน

Cluster 0 คือกลุ่มนักศึกษาที่ด้อย (W) มี 13 คน คิดเป็นร้อยละ 48

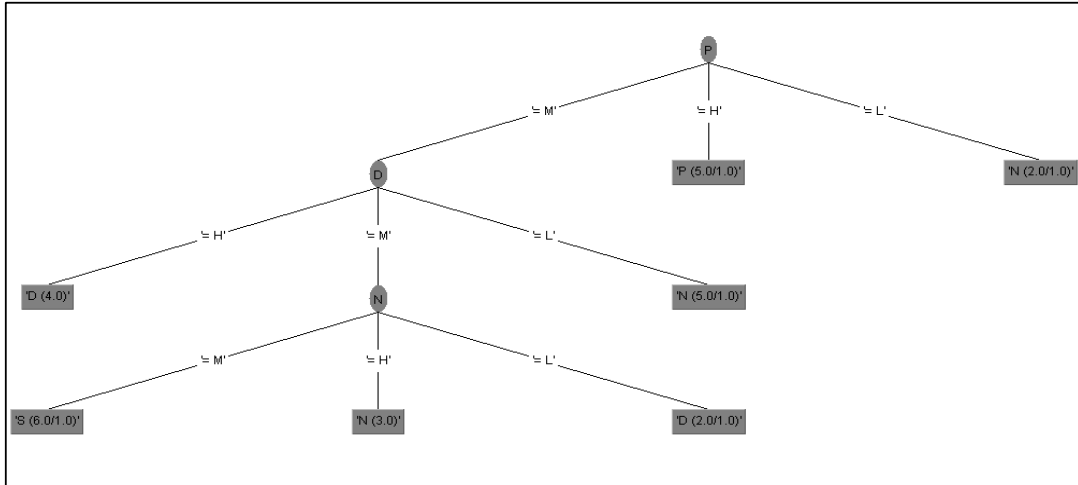
Cluster 1 คือกลุ่มนักศึกษาที่เด่น (S) มี 14 คน คิดเป็นร้อยละ 52

และจากข้อมูลข้างต้นนำข้อมูลนักศึกษากลุ่มเด่น หรือกลุ่มด้อยมาวิเคราะห์หาความสัมพันธ์โดยใช้เทคนิค Naive Bayes แยกตามประเภทการรับเข้าศึกษา ตามวุฒิการศึกษาที่จบ และตามเกรดเฉลี่ยเดิมก่อนเข้าเพื่อวิเคราะห์ความสัมพันธ์ระหว่างผลการศึกษาก่อนเข้าเรียนในระดับปริญญาตรี

2.3 การวิเคราะห์ข้อมูลผลการทำแบบทดสอบความถนัด

จะใช้เทคนิคการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจการจำแนกประเภทข้อมูลคือกระบวนการสร้างโมเดลจำแนกประเภทข้อมูล (Data Classification Model) เพื่อทำนายกลุ่มของข้อมูลใหม่ (Unseen data) ตัวอย่างของกลุ่ม เช่น กลุ่มผู้ที่มีความถนัดด้านเขียนโปรแกรม กลุ่มผู้มีความถนัดด้านฐานข้อมูล กลุ่มที่มีความถนัดด้านเครือข่าย และกลุ่มที่มีความถนัดด้านซอฟต์แวร์ ในที่นี้คำว่ากลุ่มจะเรียกว่า Class ของข้อมูล ซึ่งใน Class เดียวกันนั้นจะต้องมีข้อมูลที่มีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ใน Class ที่แตกต่างกัน

ผลลัพธ์ที่เกิดขึ้นจากการนำข้อมูลผลการทำแบบทดสอบวัดความถนัดซึ่งสามารถจำแนกผู้ตอบแบบสอบถามออกเป็นสาขาวิชา 4 สาขา ได้แก่ สาขาการเขียนโปรแกรม (P) สาขาเครือข่ายคอมพิวเตอร์ (N) สาขาฐานข้อมูล (D) และสาขาการผลิตซอฟต์แวร์ (S) สามารถสร้างเป็นต้นไม้ตัดสินใจได้ดังนี้



ภาพที่ 1 ต้นไม้ตัดสินใจเพื่อทำนายความถนัดแยกสาขา

การสร้างโมเดลจำแนกประเภทข้อมูล จะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่มคือกลุ่มข้อมูลเรียนรู้ (Training set) เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบ (Test set) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูล หลังจากได้ตัวแบบแล้วจะทำการทดสอบตัวแบบ การทดสอบและตรวจสอบความถูกต้องของตัวแบบที่สร้างขึ้นโดยใช้ข้อมูลชุดที่สอง คือ ชุดข้อมูลตรวจสอบ(Validation)ในการทดสอบตัวแบบตัวแบบที่สร้างขึ้นถ้าค่าของข้อมูลตรวจสอบมากกว่าหรือเท่ากับค่าข้อมูลที่ใช้ในการสร้างตัวแบบ แสดงว่าตัวแบบมีความน่าเชื่อถือดังจะแสดงรายละเอียดสรุปผลการทดลอง

3. ผลการศึกษาและอภิปรายผล

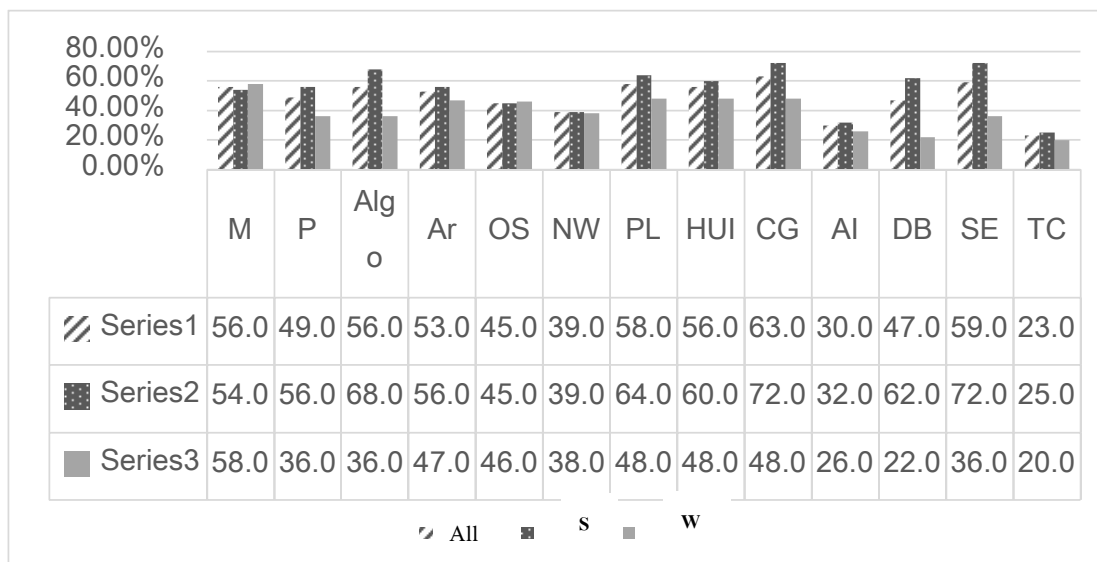
หลังจากที่ได้ศึกษาทฤษฎีที่ใช้ในงานวิจัย ดำเนินการวิจัยตามวัตถุประสงค์แล้ว จากนั้นจะนำผลการดำเนินงานวิจัยมาวิเคราะห์ตามหัวข้อต่อไปนี้

3.1 การวิเคราะห์ข้อมูลจากแบบทดสอบการประเมินความรู้ก่อนเรียน

การเตรียมข้อมูล โดยใช้กลุ่มตัวอย่างที่ทำแบบทดสอบเป็นนักศึกษาในสาขาวิชาวิทยาการคอมพิวเตอร์ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม วิทยาเขตจรัลพงษานุรักษ์ การสร้างตัวแบบที่ใช้ทำนายผล การทำเหมืองข้อมูลจะมีการแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนแรกจะแทนข้อมูลเรียนรู้ (Training set) ซึ่งจะใช้ข้อมูลชุดนี้สำหรับการสร้างตัวแบบและชุดหลังจะแทนข้อมูลตรวจสอบ (Validation set) ซึ่งใช้ข้อมูลชุดหลังสำหรับทดสอบตัวแบบที่สร้างได้ ในกระบวนการเลือกตัวอย่างจะถูกเลือกในโปรแกรมสำเร็จรูปที่ใช้วิเคราะห์ข้อมูลโดยงานวิจัยนี้ใช้โปรแกรม WEKA ในการวิเคราะห์ข้อมูล

ผลจากการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA โดยใช้เทคนิคการจัดกลุ่มด้วยอัลกอริธึม K-mean โดยกำหนดการจัดกลุ่มนักศึกษาเป็น 2 กลุ่มจากนักศึกษาที่ทำแบบทดสอบ 27 คน คือ กลุ่มนักศึกษาที่ด้อย (W) มี 13 คน คิดเป็นร้อยละ 48 และกลุ่มนักศึกษาที่เด่น (S) มี 14 คน คิดเป็นร้อยละ 52

สามารถนำข้อมูลมาจัดทำเป็นกราฟเพื่อวิเคราะห์ผลการจัดกลุ่มได้ดังนี้



ภาพที่ 2 กราฟผลการจัดกลุ่มของนักศึกษาแยกตาม 13 กลุ่มวิชา

จากกราฟแสดงให้เห็นผลการจัดกลุ่มของนักศึกษาแยกตาม 13 กลุ่มวิชา โดย Series 1 หมายถึงภาพรวม Series 2 หมายถึงนักศึกษากลุ่มเด่น (S) Series 3 หมายถึงนักศึกษากลุ่มด้อย (W) ข้อสังเกตจะเห็นว่านักศึกษากลุ่มเด่นจะได้คะแนนสูงกว่ากลุ่มด้อย ยกเว้นกลุ่มวิชาคณิตศาสตร์ และกลุ่มวิชาระบบปฏิบัติการ เมื่อพิจารณาถึงวุฒิการศึกษาและเกรดเฉลี่ยของแต่ละกลุ่มได้ผลลัพธ์ดังนี้กลุ่มที่วุฒิการศึกษาจากมัธยมศึกษาตอนปลาย จาก 22 คน ส่วนใหญ่จะอยู่ในกลุ่มด้อย 12 คน จาก 22 คน คิดเป็นร้อยละ 54 เนื่องจากไม่มีพื้นฐานทางวิชาแกน และกลุ่มวิชาเฉพาะทางด้านวิทยาการคอมพิวเตอร์ แต่จะมีคะแนนด้านคณิตศาสตร์ และทฤษฎีการคำนวณสูงกว่านักศึกษากลุ่มเด่นที่มาจากนักศึกษาที่รับเข้าด้วยวุฒิประกาศนียบัตรวิชาชีพ

แต่ขณะเดียวกันกลุ่มที่รับเข้าจากระดับประกาศนียบัตรวิชาชีพจาก 5 คน จะอยู่ในกลุ่มเด่น 4 คน คิดเป็นร้อยละ 80 จะมีคะแนนด้านวิชาแกน และกลุ่มวิชาเฉพาะทางด้านวิทยาการคอมพิวเตอร์สูงกว่านักศึกษาที่รับเข้าด้วยวุฒิการศึกษาจากมัธยมศึกษาตอนปลายแต่จะมีคะแนนด้านคณิตศาสตร์ และทฤษฎีการคำนวณต่ำ

3.2 การวิเคราะห์ข้อมูลจากแบบทดสอบวัดความถนัด

การเตรียมข้อมูล โดยใช้กลุ่มตัวอย่างที่ทำแบบทดสอบเป็นผู้ประกอบอาชีพสาขาวิชา 4 กลุ่ม ข้อมูลส่วนนี้จะเป็นการจำแนกประเภทข้อมูลโดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervise Learning) ผลการทดสอบจะต้องจำแนกคุณสมบัติของผู้ทำแบบทดสอบว่ามีความถนัดในด้านใด แล้วได้เป็นต้นแบบเพื่อใช้ทำนายข้อมูลใหม่ได้เทคนิคการจำแนกข้อมูลใช้ต้นไม้ตัดสินใจการจำแนกประเภทข้อมูลคือกระบวนการสร้างโมเดลจำแนกประเภทข้อมูล (Data Classification Model) เพื่อทำนายกลุ่มของข้อมูลใหม่ (Unseen data) ตัวอย่างของกลุ่ม เช่น กลุ่มผู้ที่มีความถนัดด้านเขียนโปรแกรม กลุ่มผู้มีความถนัดด้านฐานข้อมูล กลุ่มที่มีความถนัดด้านเครือข่าย และกลุ่มที่มีความถนัดด้านซอฟต์แวร์ ในที่นี้คำว่ากลุ่มจะเรียกว่า Class ของข้อมูล ซึ่งใน Class เดียวกันนั้นจะต้องมีข้อมูลที่มีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ใน Class ที่แตกต่างกัน

ผลลัพธ์ที่เกิดขึ้นจากการนำข้อมูลผลการทำแบบทดสอบวัดความถนัดซึ่งสามารถจำแนกผู้ตอบแบบสอบถามออกเป็นสาขาวิชา 4 สาขา ได้แก่ สาขาการเขียนโปรแกรม (P) สาขาเครือข่ายคอมพิวเตอร์ (N) สาขาฐานข้อมูล (D) และสาขาการผลิตซอฟต์แวร์ (S) ได้กฎความสัมพันธ์ของต้นแบบดังนี้

กฎข้อที่ 1 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับต่ำ จะมีความถนัดทางด้าน Network

กฎข้อที่ 2 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลาง จะมีความถนัดทางด้าน Programming

กฎข้อที่ 3 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลางและคะแนนความถนัดด้าน Database ได้ระดับสูงจะมีความถนัดทางด้าน Database

กฎข้อที่ 4 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลาง และคะแนนความถนัดด้าน Database ได้ระดับต่ำจะมีความถนัดทางด้าน Network

กฎข้อที่ 5 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลาง และคะแนนความถนัดด้าน Database ได้ระดับกลางและคะแนนความถนัดด้าน Network ได้ระดับกลางจะมีความถนัดทางด้าน Software

กฎข้อที่ 6 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลาง และคะแนนความถนัดด้าน Database ได้ระดับกลางและคะแนนความถนัดด้าน Network ได้ระดับสูงจะมีความถนัดทางด้าน Network

กฎข้อที่ 7 ถ้าผลคะแนนวันความถนัดด้าน Programming ได้ระดับกลาง และคะแนนความถนัดด้าน Database ได้ระดับกลางและคะแนนความถนัดด้าน Network ได้ระดับต่ำจะมีความถนัดทางด้าน Database

3.3 การตรวจสอบความถูกต้องของตัวแบบ

การสร้างตัวแบบของแบบทดสอบทั้งสองแบบจะใช้โปรแกรม WEKA ในการสร้างตัวแบบโดยแบบทดสอบประเมินความรู้ก่อนเรียนจะใช้เทคนิคการจัดกลุ่มโดยใช้อัลกอริทึม K-mean และการจำแนกข้อมูลจากแบบทดสอบวัดความถนัดใช้เทคนิคต้นไม้ตัดสินใจโดยใช้อัลกอริทึมการหาความสัมพันธ์ของข้อมูลด้วย J48 การตรวจสอบความถูกต้องของตัวแบบด้วยการแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลเรียนรู้ (Training set) เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบ (Test set) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูลหลังจากได้ตัวแบบแล้วจะทำการทดสอบตัวแบบการทดสอบและตรวจสอบความถูกต้องของตัวแบบที่สร้างขึ้นโดยใช้ข้อมูลชุดที่สองคือชุดข้อมูลตรวจสอบ (Validation) ในการทดสอบตัวแบบที่สร้างขึ้น ถ้าค่าของข้อมูลตรวจสอบมากกว่าหรือเท่ากับค่าข้อมูลที่ใช้ในการสร้างตัวแบบแสดงว่าตัวแบบมีความน่าเชื่อถือดังจะแสดงรายละเอียดดังนี้

ตารางที่ 3.1 ความถูกต้องของตัวแบบตามสัดส่วนการแบ่งข้อมูล 60:40

ประเภทแบบทดสอบ	ร้อยละความถูกต้องเฉลี่ย
แบบทดสอบประเมินความรู้ก่อนเรียน	80.90
แบบทดสอบวัดความถนัด	79.42

จากการวิเคราะห์ที่ผ่านมามีผลการดำเนินการวิจัยสามารถสร้างแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัด แบบทดสอบทั้งสองแบบผ่านการทดลองและวัดประสิทธิภาพของแบบทดสอบก่อนนำไปใช้ หลังจากนั้นจะนำแบบทดสอบไปใช้กับกลุ่มเป้าหมายที่วางไว้และนำผลการทำแบบทดสอบกับกลุ่มเป้าหมายมาทำนายเป็น 2 กลุ่มดังนี้

นักศึกษาชั้นปีที่ 1 เพื่อจัดกลุ่มนักศึกษาออกเป็น 2 กลุ่ม กลุ่มเด่น (S) และกลุ่มด้อย (W) หรือใช้เพื่อค้นหารายวิชาที่ด้อย และรายวิชาที่เด่น ในแต่ละกลุ่ม การวิเคราะห์ข้อมูลจะใช้เทคนิคการทำเหมืองข้อมูลการจัดกลุ่มของข้อมูล (Clustering) เพื่อหาคุณสมบัติของรายวิชา เช่น ถ้านักศึกษาทำแบบทดสอบได้คะแนนต่ำในกลุ่มรายวิชาการเขียนโปรแกรม นักศึกษากลุ่มนั้นควรได้รับการปรับปรุงพื้นฐานเพิ่มเติมในด้านการเขียนโปรแกรม เป็นต้น

นักศึกษาชั้นปีที่ 4 เพื่อนักศึกษาเพื่อแบ่งสาขาตามความถนัด และนำมาสร้างต้นแบบโดยใช้เทคนิคการทำเหมืองข้อมูลการจำแนกประเภท (Classification) ด้วยการนำผลการทดสอบมาวิเคราะห์กรณีนี้นักศึกษาทำคะแนนได้สูง

ในด้านใดก็จะจัดกลุ่มให้นักศึกษามีความถนัดในด้านนั้น และนำไปเป็นต้นแบบใช้ในการทำนายความถนัดของนักศึกษาในรุ่นต่อ ๆ ไป

ผลลัพธ์ที่ได้จากงานวิจัยนี้จะได้แบบทดสอบที่ใช้ในการทดสอบความรู้ก่อนเรียนของนักศึกษาปี 1 เพื่อเป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น ลดความเสี่ยงของนักศึกษาที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์ และได้แบบทดสอบวัดความถนัดเฉพาะสาขาทางวิทยาการคอมพิวเตอร์ เพื่อพัฒนาศักยภาพของนักศึกษาตามความถนัดมีโอกาสดำเนินการทำหลังจบการศึกษามากขึ้น

4. สรุป

จากการวิเคราะห์ที่ผ่านมาผลการดำเนินการวิจัยสามารถสร้างต้นแบบการจัดกลุ่มนักศึกษาจากการทำเหมืองข้อมูลจากผลการทำแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัด ซึ่งแบบทดสอบทั้งสองแบบผ่านการทดลองและวัดประสิทธิภาพของแบบทดสอบก่อนนำไปใช้ การทดลองดำเนินการกับกลุ่มประชากรทดสอบผลลัพธ์ที่ได้เป็นต้นแบบซึ่งจะนำไปทำนายกลุ่มเป้าหมายที่วางไว้กลุ่มเป้าหมายมาทำนายเป็น 2 กลุ่มดังนี้

นักศึกษาชั้นปีที่ 1 เพื่อค้นหารายวิชาที่ด้อย และรายวิชาที่เด่น การวิเคราะห์ข้อมูลจะใช้เทคนิคการทำเหมืองข้อมูลการจัดกลุ่มของข้อมูล (Clustering) เพื่อจัดกลุ่มนักศึกษาตามคะแนนของรายวิชาและใช้ในการดูแลนักศึกษา เช่น ถ้านักศึกษาทำแบบทดสอบได้คะแนนต่ำในกลุ่มรายวิชาการเขียนโปรแกรม นักศึกษากลุ่มนั้นควรได้รับการปรับปรุงพื้นฐาน

เพิ่มเติมในด้านการเขียนโปรแกรม เป็นต้นนักศึกษาชั้นปีที่ 4 เพื่อนักศึกษาเพื่อแบ่งสาขาตามความถนัดโดยใช้กฎ 7 กฎที่ได้จากต้นแบบมาทำนายกับชุดข้อมูลใหม่เพื่อทำนายความถนัดของนักศึกษาแต่ละคนเพื่อพัฒนาศักยภาพของนักศึกษา หรือหาความน่าจะเป็นของอาชีพตามความถนัดของนักศึกษา

ผลลัพธ์ที่ได้จากงานวิจัยนี้จะใช้เป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น ลดความเสี่ยงของนักศึกษาที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์ และได้แบบทดสอบวัดความถนัดเฉพาะสาขาทางวิทยาการคอมพิวเตอร์ เพื่อพัฒนาศักยภาพของนักศึกษาตามความถนัดมีโอกาสดำเนินการทำหลังจบการศึกษาเพิ่มขึ้น

5. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนทุนวิจัยงบประมาณผลประโยชน์ 2558 จากคณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก

6. เอกสารอ้างอิง

ชัชชฎา วันดี, ช. ส. 2556. ประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพโดยอัตโนมัติด้วยเทคนิคเหมืองข้อมูล. การประชุมวิชาการมหาวิทยาลัยมหาสารคามวิจัย ครั้งที่ 9, 263-269.

ชุตติมา อุตมะมุณี, ช.ส. 2010. การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. Journal of Information Science and Technology Vol 1, ISSUE2, 39-48.

ญาใจ ลิ้มปิยกรณ์. 2553. การทำเหมืองข้อมูล. เอกสารคำสอน. จุฬาลงกรณ์มหาวิทยาลัย.

พิชิต ฤทธิ์จรูญ. 2556. หลักการวัดและประเมินผลการศึกษา. พิมพ์ครั้งที่ 4. แฮร์ส ออฟ เคอร์มีส์ : กรุงเทพฯ.

ศุภกัศ เจริญเลิศมงคล. 2553. ระบบวิเคราะห์ข้อมูลนักเรียนเชิงบูรณาการโดยใช้ เทคโนโลยีเหมืองข้อมูล กรณีศึกษา โรงเรียนท่าสายวิทยา. วิทยานิพนธ์. วิทยาศาสตร์มหาบัณฑิต. สาขาวิชาระบบสารสนเทศเพื่อการจัดการ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.