Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษา หลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา

Analysis of Factors Influencing the Dismissal of Students Using Data Mining Techniques Case Study: Computer Science Program and Information Technology Program of Yala Rajabhat University

ซอและ เกปัน (Solaeh Kepan)*
พิมลพรรณ ลีลาภัทรพันธุ์ (Pimonpun Leelapatarapun)**
อัจฉราพร ยกขุน (Atcharaporn Yokkhun) ***

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา หลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา ซึ่งใช้ข้อมูลของนักศึกษาที่กำลังศึกษาและสำเร็จการศึกษาแล้วระหว่างปีการศึกษา 2555 – 2558 เป็นข้อมูลจากหลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 97 ระเบียน และหลักสูตรเทคโนโลยีสารสนเทศ จำนวน 202 ระเบียน ประกอบด้วย 26 คุณลักษณะ โดยใช้เทคนิคเหมืองข้อมูล ประกอบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (Back Propagation Neural Network : BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลการทดสอบ พบว่า ข้อมูลภูมิหลังไม่ใช่ปัจจัยในการทำนายการพ้นสภาพของนักศึกษาทั้งสองหลักสูตร นอกจากนี้ทั้งสองหลักสูตรยังมีปัจจัยที่แตกต่างกัน โดยปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาหลักสูตรเทคโนโลยีสารสนเทศ มี 3 ปัจจัย ได้แก่ ผลการเรียนรายวิชาพื้นฐานทางฟิสิกส์ ผลการเรียนรายวิชาแพลตฟอร์มเทคโนโลยี และ ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1 สำหรับหลักสูตรวิทยาการคอมพิวเตอร์มีเพียง 1 ปัจจัย ได้แก่ ผลการเรียนรายวิชาโครงสร้างข้อมูล สำหรับตัวแบบพยากรณ์ที่ใช้เทคนิค SVM ให้ความแม่นยำในการพยากรณ์สูงสุดที่ร้อยละ 97.75 โดยมีการประเมินค่าความแม่นตรงเท่ากับ 0.920 ค่าความระลึกเท่ากับ 0.993 และประสิทธิภาพโดยรวม เท่ากับ 0.926 ตัวแบบพยากรณ์ที่ใช้เทคนิค BP-NN ให้ความแม่นยำในการพยากรณ์ที่ร้อยละ 97.50 โดยมีการประเมินค่าความแม่นตรงเท่ากับ 0.905 ค่าความระลึกเท่ากับ 0.986 และประสิทธิภาพโดยรวมเท่ากับ 0.941 และตัวแบบพยากรณ์ที่ใช้เทคนิคต้นไม้ตัดสินใจให้ความแม่นยำในการพยากรณ์ที่ร้อยละ 97.00 โดยมี

* หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร มหาวิทยาลัยราชภัฏยะลา
B.Sc Information Technology , Yala Rajabhat University. E-Mail : solaeh.k@yru.ac.th
** หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร มหาวิทยาลัยราชภัฏยะลา
B.Sc Information Technology , Yala Rajabhat University. E-Mail : pimonpun.l@yru.ac.th
*** หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร มหาวิทยาลัยราชภัฏยะลา
B.Sc Information Technology , Yala Rajabhat University. E-Mail : atcharaporn.y@yru.ac.th

การประเมินค่าความแม่นตรงเท่ากับ 0.975 ค่าความระลึกเท่ากับ 0.986 และประสิทธิภาพโดยรวมเท่ากับ 0.980 ทั้งนี้ค่า RMSE ของเทคนิคต้นไม้ตัดสินใจมีค่าน้อยสุด ตามด้วยเทคนิค SVM และ BP-NN เท่ากับ 0.0383 0.0512 และ 0.0785 ตามลำดับ ผลของงานวิจัยนี้จะช่วยเป็นแนวทางให้นักศึกษาสามารถวางแผนการเรียนได้ เช่น การเพิ่มถอนรายวิชาให้เหมาะสมกับศักยภาพตนเอง และ การพัฒนาตนเองในการเรียนวิชาต่างๆ ที่มีความสำคัญ

คำสำคัญ : การพ้นสภาพการเป็นนักศึกษา การทำเหมืองข้อมูล การคัดเลือกคุณลักษณะ ตัวแบบพยากรณ์

Abstract

The objective of this research was to analyze the factors influencing the dismissal of students from Computer Science program and Information Technology program of Yala Rajabhat University using students and graduates' data during academic year 2012 – 2015. There were 97 records from Computer Science program and 202 records from Information Technology program those consisted of 26 features. Data mining techniques like Decision tree technique, Back Propagation Neural Network (BP-NN), and Support Vector Machine (SVM) were employed to propose forecasting models. The models were tested and compared using a 10-fold Cross Validation. The results revealed that biography was not an influencing factor on the students of both programs. Furthermore, the influencing factors of each program were different. For Information Technology program, the factors influencing the dismissal of the students were Physics grade points, Platform Technology grade points, and the first semester grade point average of the second academic year while the Data Structure grade points was the only one factor influencing the dismissal of Computer Science program students. The forecast model using the SVM technique had the highest forecasting accuracy at 97.75 percent while the value of precision, recall and F-Measure were 0.920, 0.993, and 0.926. The forecast model using the BP-NN technique had the highest forecasting accuracy at 97.50 percent while the value of precision, recall and F-Measure were 0.905, 0.986, and 0.941. The forecast model using the Decision tree technique had the highest forecasting accuracy at 97.00 percent with the precision value at 0.975, recall at 0.986 and F-Measure at 0.980. The RMSE value of Decision tree technique was the smallest, following by SVM and BPNN were at 0.0383, 0.0512, and 0.0785, respectively. The results of this research will guide the students to manage their academic plan for example, proper registration adjustment and self-development in important subjects learning.

Keywords : Student Dismissal, Data mining, Feature Selection, Forecasting Models

Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

บทนำ

จากสถานการณ์ในปัจจุบันของการรับนักศึกษาเข้าศึกษาในหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศของมหาวิทยาลัยราชภัฏยะลา พบว่ามีนักศึกษาต้องพ้นสภาพโดยไม่สามารถสำเร็จการศึกษาได้ตั้งแต่ปีการศึกษา 2554 ถึง 2558 ตามตารางที่ 1 และตารางที่ 2

ตารางที่ 1 แสดงจำนวนนักศึกษาที่พ้นสภาพของหลักสูตรวิทยาการคอมพิวเตอร์

ปีการศึกษาที่เข้า	รายงานตัว	พ้นสภาพ	ร้อยละ
2554	82	19	23.17
2555	58	21	36.21
2556	54	14	25.93
2557	34	5	14.71
2558	41	10	24.39

ตารางที่ 2 แสดงจำนวนนักศึกษาที่พ้นสภาพของหลักสูตรเทคโนโลยีสารสนเทศ

ปีการศึกษาที่เข้า	รายงานตัว	พ้นสภาพ	ร้อยละ
2554	49	13	26.53
2555	91	24	26.37
2556	44	14	31.82
2557	33	5	15.15
2558	41	14	34.15

จากตารางที่ 1 และ ตารางที่ 2 จะเห็นว่านักศึกษาพ้นสภาพการเป็นนักศึกษาในนักศึกษาหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศโดยเฉลี่ยเกินร้อยละ 20 ของในแต่ละปีการศึกษาที่รับนักศึกษาเข้ามา จึงส่งผลกระทบต่อความคุ้มค่าของการจัดการศึกษาทั้งในส่วนของมหาวิทยาลัยรวมไปถึงนักศึกษาและผู้ปกครอง

วัตถุประสงค์การวิจัย

เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูลกรณีศึกษาหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา

งานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ทำการค้นคว้างานวิจัยและผลงานวิชาการต่าง ๆ ที่เกี่ยวข้อง ดังนี้

งานวิจัยที่เกี่ยวข้องกับเทคนิคเหมืองข้อมูล เช่น การจำแนกกลุ่มข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล (ประพัฒน์ พรมน้ำอ่าง วสุวรรธน์ พงศ์ขจร และ นิเวศ จิระวิชิตชัย, 2559) ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิค k-NN และเทคนิค Naïve Bayes

การระบุตัวผู้เขียนข้อความออนไลน์ภาษาไทยด้วยซัพพอร์ตเวกเตอร์แมชชีนและต้นไม้ตัดสินใจ (รังสิพรรณ มฤคทัต, 2558)

การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา (เสกสรร วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และ ดวงดาว วิชาดากุล, 2558) ใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree)

การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน (พรรณิภา บุตรเอก และ สุรเดช บุญลือ, 2557) ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-NN)

ระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจ (ชิดชนก ศรีชัยวงค์ ไพศาล ตระกูลสุข และ สุรเดช บุญลือ, 2557)

การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์ (นิภาพร ชนะมาร และ พรรณี สิทธิเดช, 2557) ใช้เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM)

ดังนั้นผู้วิจัยจึงเล็งเห็นความสำคัญในการศึกษาปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา หลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ เพื่อจะเป็นแนวทางในการแก้ปัญหาการพ้นสภาพของนักศึกษา โดยใช้เทคนิคการทำเหมืองข้อมูลในการพยากรณ์

การทำเหมืองข้อมูล (Data Mining)

ทฤษฎีในการใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) เป็นเทคนิคเพื่อค้นหารูปแบบ (Pattern) ของความรู้จากฐานข้อมูลจำนวนมหาศาลโดยอัตโนมัติ เพื่อนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้หรือสิ่งสำคัญออกมาเพื่อใช้ในการวิเคราะห์ข้อมูล แยกประเภท จำแนกรูปแบบและความสัมพันธ์ของข้อมูล จากคลังข้อมูลและนำสารสนเทศที่ได้ไปใช้ในการตัดสินใจ

โดยรูปแบบการทำเหมืองข้อมูลสามารถจำแนกออกเป็น 3 ประเภท (วิภาวรรณ บัวทอง, 2559) ดังนี้

1. กฎความสัมพันธ์ (Association rule)

เป็นการค้นหาความสัมพันธ์ของระหว่างสิ่งของต่างๆ เพื่อนำไปใช้ในการวิเคราะห์หรือทำนายปรากฏการณ์ต่างๆ เช่น การวิเคราะห์ข้อมูลการขายสินค้าจากระบบ ณ จุดขาย (POS) แล้วพิจารณาข้อมูลสินค้าที่ลูกค้ามักจะซื้อพร้อมกัน โดยสามารถใช้ในการจัดเรียงสินค้าต่อไป

2. การจำแนกประเภทข้อมูล (Classification)

เป็นการหากฎเพื่อระบุคุณสมบัติของวัตถุ เช่น การจัดกลุ่มนักเรียน เป็น กลุ่มดี กลุ่มปานกลาง กลุ่มอ่อน โดยพิจารณาจากประวัติและผลการเรียน เพื่อประกอบการจัดกิจกรรม หรือการแบ่งกลุ่มลูกค้าว่าเป็นผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติการกู้เงิน

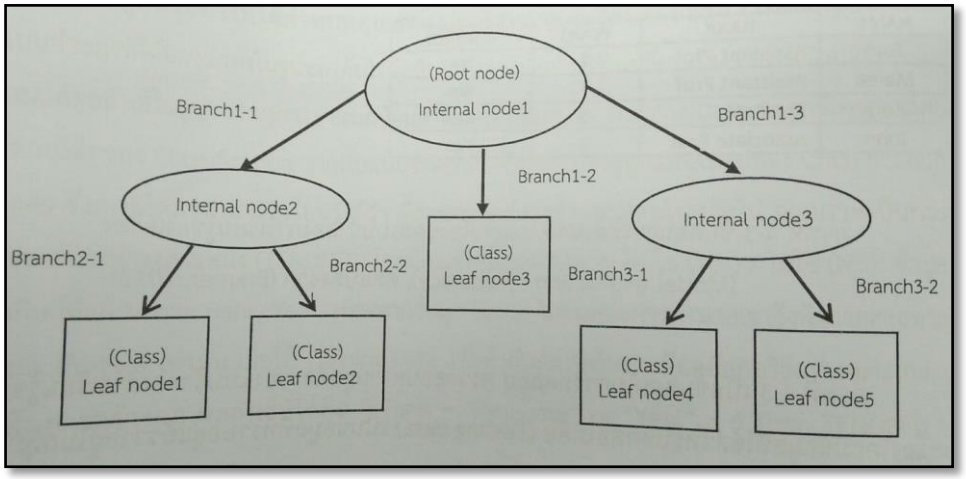
Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

3. การแบ่งกลุ่ม (clustering)

เป็นการแบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม ๆ โดยอัติโนมัติ เช่น การแบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกัน โดยพิจารณาจากลักษณะของอาการ เพื่อพิจารณาหาสาเหตุของโรค

เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

ทฤษฎีในการใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นการจำแนกข้อมูลจำนวนมากให้เป็นกลุ่มที่มีขนาดเล็กลง โดยใช้กฎในการตัดสินใจ และสามารถสร้างแบบจำลองได้จากกลุ่มตัวอย่างที่กำหนดไว้ล่วงหน้า โดยต้นไม้ตัดสินใจมีลักษณะคล้ายโครงสร้างต้นไม้ แต่ละโหนดมีกิ่งที่แสดงเงื่อนไขในการทดสอบ และโหนดปลายสุดแสดงถึงเป้าหมายที่กำหนดไว้ ดังภาพที่ 1



ภาพที่ 1 แนวความคิดองค์ประกอบของต้นไม้ตัดสินใจ
ที่มา : วิภาวรรณ บัวทอง (2559)

จากภาพที่ 1 ในการพิจารณาเกณฑ์ที่ช่วยตัดสินใจในการเลือก root node คือ ทดลองเลือกคุณลักษณะของแต่ละตัวมาทำหน้าที่ root node แล้วหาค่า Gain ratio โดยอาศัยการคำนวณของทฤษฎีสารสนเทศ และค่าสารสนเทศขึ้นอยู่กับความน่าจะเป็นของข้อมูลและจะเลือกลักษณะที่ให้ค่า Gain ratio มากที่สุดเป็น root node ดังสมการ (1) Info(T) เป็นฟังก์ชันที่ระบุปริมาณข้อมูลที่ต้องการเพื่อให้สามารถจำแนกคุณลักษณะที่ต้องการได้และสมการ (2) $info_x(T)$ เป็นฟังก์ชันที่ระบุปริมาณข้อมูลที่ต้องการเพื่อจำแนกคลาสของข้อมูลโดยใช้คุณลักษณะ x เป็นตัวตรวจสอบเพื่อแยกข้อมูลต่อไป (วิภาวรรณ บัวทอง, 2559)

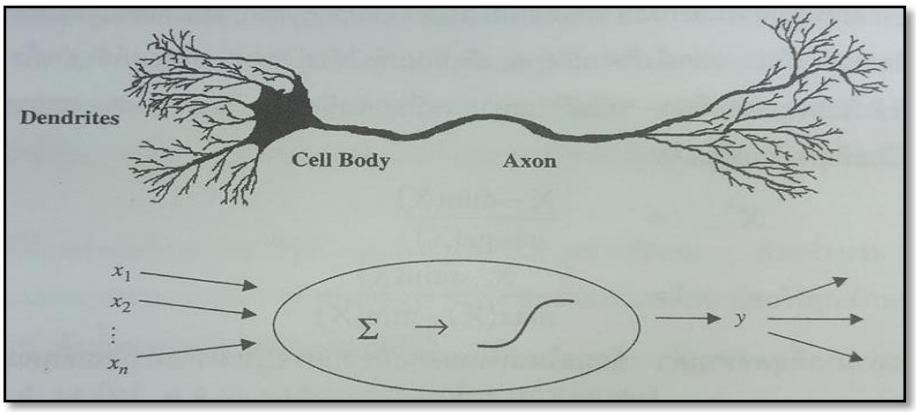
$$\text{info}(T) = -\sum_{j=1 \text{ to } k}[freq(C_j,T)/|T|] \; x \; log_2[freq(C_j,T)/|T|] \text{ bits} \quad (1)$$

$$\text{info}_x(T) = \sum_{i=1 \text{ to } n}(\frac{|T_i|}{|T_i|}) \; x \; info(T_i) \text{ bits} \quad (2)$$

เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-NN)

ทฤษฎีในการใช้เทคนิคโครงข่ายประสาทเทียมจะมีพื้นฐานมาจากนิวรอนที่เป็นของจริงในสมองมนุษย์ นิวรอนหรือเซลล์ประสาทจะมีหน้าที่หลักสองประการ คือ การคำนวณ และการส่งผลที่ได้จากการคำนวณไปยังอีกปลายหนึ่งของเซลล์ประสาทอย่างรวดเร็ว เพื่อให้สามารถส่งผลไปอย่างเซลล์อื่น ๆ ได้อย่างทันที ซึ่งทั้งหมดนี้จะมีการทำงานโดยอาศัยหลักการทางไฟฟ้า ซึ่งจะมีลักษณะคล้ายกับต้นไม้ที่ปราศจากใบ ที่แต่ละกิ่งและรากที่เชื่อมโยงกันด้วยลำต้น โดยมีการทำงานคือจะรับอินพุต (Input) จากแหล่งต่าง ๆ นำมารวมเข้าด้วยกัน และมีการทำงานในแบบไม่เป็นเส้นตรง (Non-Linear) และส่งเอาต์พุตสุดท้ายออกมา เซลล์ประสาทมีความหลากหลายอย่างมาก จะมีส่วนประกอบพื้นฐาน 4 ส่วนเหมือน ๆ กัน ส่วนประกอบนั้นจะเรียกตามชื่อทางชีววิทยาว่า เดนไดร์ท (Dendrites) โซม่า (Soma) แอ็ดซอน (Axon) และซินแนปส์ (Synapses)

เดนไดร์ทเป็นส่วนขยายหรือส่วนต่อของโซม่า ซึ่งมีลักษณะคล้ายขนซึ่งทำหน้าที่เหมือนเป็นช่องทางการนำเข้าของอินพุต และเดนไดร์ทจะทำการรับอินพุตผ่านซินแนปส์ของเซลล์ประสาทอื่น ๆ โดยโซม่าจะค่อยประมวลผลสัญญาณไฟฟ้าที่รับเข้ามาตลอดเวลา แล้วส่งผลการทำงานเป็นเอาต์พุตออกไปให้เซลล์ประสาทอื่น โดยผ่านทางแอ็ดซอนและซินแนปส์ ดังภาพที่ 2



ภาพที่ 2 ตัวแบบประสาทที่แท้จริงและตัวแบบประสาทเทียม
ที่มา : สายชล สินสมบูรณ์ทอง (2558)

โครงข่ายประสาทเทียมแบบย้อนกลับ (Back propagation neural network : BP-NN) ใช้ความคลาดเคลื่อนของการทำนาย (ค่าจริง – ค่าผลลัพธ์) สำหรับระเบียนหนึ่งๆ และจะจัดความคลาดเคลื่อนเป็นส่วน ๆ เพื่อการเชื่อมต่อต่างๆ โดยมีกฎการแพร่แบบย้อนกลับได้ดังนี้ (สายชล สินสมบูรณ์ทอง,2558)

$$W_{ij,new} = W_{ij,current} + \Delta w_{ij}$$

โดยที่ $\Delta w_{ij} = n\delta_j x_{ij}$

ซึ่งขึ้นอยู่กับโหนดว่าอยู่ในชั้นข้อมูลออกหรือชั้นซ่อน ดังสมการต่อไปนี้

$$\delta = output_j(1 - output_j)(actual_j - output_j)$$ สำหรับโหนดในชั้นข้อมูลออก

$$\delta = output_j(1 - output_j)\sum_{downstream} W_{jk}\,\delta_j$$ สำหรับโหนดในชั้นซ่อน

Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

เทคนิคซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)

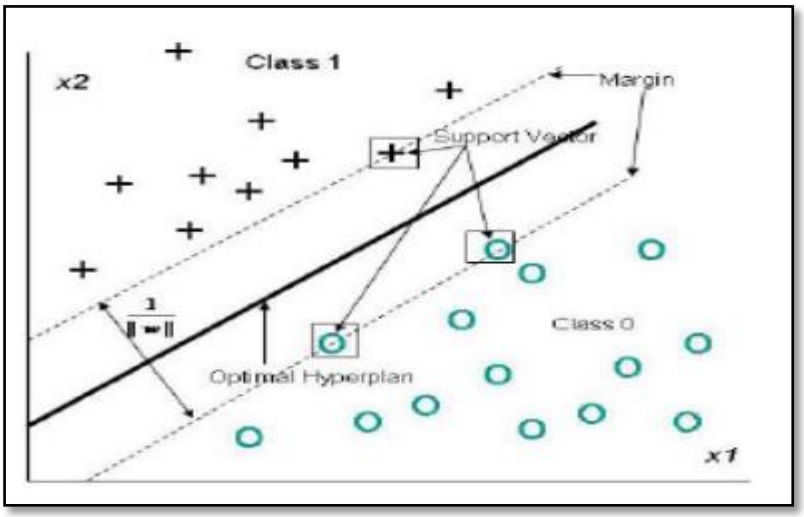
ทฤษฎีในการใช้เทคนิคซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) หลักการของวิธีการนี้ใช้เพื่อหาระนาบการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยใช้สมการเส้นตรงเพื่อแบ่งเขตข้อมูล 2 กลุ่มออกจากกัน ดังภาพที่ 3 โดยมีวัตถุประสงค์ที่จะพยายามที่จะทำการลดความผิดพลาดจากการทำนาย (Minimize error) พร้อมกับเพิ่มระยะแยกแยะให้มากที่สุด (Maximized Margin) ซึ่งต่างจากเทคนิคโดยทั่วไปเช่น โครงขายประสาทเทียม (Artificial Neural Network: ANN) ที่มุ่งเพียงทำให้ความผิดพลาดจากการทำนายให้ต่ำที่สุดเพียงอย่างเดียว โดยจะใช้ฟังก์ชั่นแม็ปข้อมูลจาก Input Space ไปยัง Feature Space และสร้างฟังก์ชั่นวัดความคล้ายที่เรียกว่าเคอร์เนลฟังก์ชั่น (Kernel Function) บน Feature Space เหมาะใช้สำหรับข้อมูลที่มีลักษณะมิติของข้อมูลที่มีปริมาณมาก โดยกำหนดให้ (xi ,yi),...,(xn,yn) เป็นตัวอย่างที่ใช้สำหรับการสอน n คือ จำนวนข้อมูล ตัวอย่าง m คือ จำนวนมิติข้อมูลเข้า และ y คือ ผลลัพธ์มีค่า + 1 หรือ -1 ดังสมการต่อไปนี้

(xi ,yi),...,(xn,yn) เมื่อ x ϵ Rm , yϵ {+1,-1}

สำหรับปัญหาเชิงเส้น มิติข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยระนาบตัดสินใจ ซึ่งคำนวณได้ดังสมการ และเมื่อ w คือ ค่าน้ำหนักและ b คือค่า bias สมการ ใช้สำหรับจำแนกประเภทของข้อมูล

(w*x)+b=0 โดย (w*x)+b>0 ถ้า yi = +1 และ (w*x)+b

อย่างไรก็ตามซัพพอร์ทเวกเตอร์แมชชีนมีเคอร์เนลฟังก์ชั่น (Kernel Function) ที่สามารถประยุกต์ใช้ในการแก้ปัญหาได้หลายวิธีโดยต้องเลือกเคอร์เนลให้เหมาะสมกับลักษณะข้อมูล (พรพล ธรรมรงค์รัตน์ ลัดดา ปรีชาวีรกุล และวิภาดา เวทย์ประสิทธ์, 2553)



ภาพที่ 3 ตัวอย่างการทำนายแบบซัพพอร์ทเวกเตอร์แมชชีน

ที่มา : พรพล ธรรมรงค์รัตน์ ลัดดา ปรีชาวีรกุล และวิภาดา เวทย์ประสิทธ์ (2553)

วิธีการวิธีวิจัย

วิธีวิจัยครั้งนี้แบ่งออกเป็น 3 ขั้นตอน 1) การจัดเตรียมข้อมูล 2) การสร้างตัวแบบพยากรณ์ และ 3) การวัดประสิทธิภาพของตัวแบบพยากรณ์

1. ขั้นการจัดเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้ คือ ข้อมูลของนักศึกษาที่ศึกษาระดับปริญญาตรี หลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 97 ระเบียน ประกอบด้วย คุณสมบัติ 26 ตัวแปร และหลักสูตรเทคโนโลยีสารสนเทศ จำนวน 202 ระเบียน ประกอบด้วยคุณสมบัติ 26 ตัวแปร แบ่งเป็นตัวแปรอิสระ 25 ตัวแปร ได้แก่ ข้อมูลภูมิหลังต่าง ๆ และข้อมูลผลการเรียนของรายวิชาที่ศึกษาในแผนการเรียนชั้นปีที่หนึ่งและชั้นปีที่สอง ตัวแปรตามหรือตัวแปรพยากรณ์ คือ สถานะการพ้นสภาพหรือไม่พ้นสภาพของนักศึกษา โดยข้อมูลที่ไม่เกี่ยวข้องกับการทำนาย เช่น รหัสนักศึกษา ชื่อ นามสกุล ซึ่งได้ถูกคัดออก ดังแสดงในตารางที่ 3 และตารางที่ 4

2. การสร้างตัวแบบพยากรณ์

เป็นขั้นตอนของการออกแบบและสร้างตัวแบบพยากรณ์ทั้ง 3 ตัวแบบ หลังจากนั้นจะมีการคัดเลือกตัวแบบพยากรณ์ที่มีประสิทธิภาพความแม่นยำสูงสุด ระหว่างตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียม (Back Propagation Neuron Network: BP-NN) เพื่อนำตัวแบบที่ได้ไปใช้ในการพยากรณ์โอกาสพ้นสภาพของนักศึกษาต่อไป ซึ่งข้อมูลที่นำเข้าฝึกสอนและทดสอบจะมี 25 คุณลักษณะ ส่วนข้อมูลนำออกจะเป็นสถานะพ้นสภาพการศึกษาของนักศึกษา (พ้นสภาพ/ไม่พ้นสภาพ)

3. ขั้นการวัดประสิทธิภาพตัวแบบในการพยากรณ์

การวิจัยนี้ครั้งนี้ผู้วิจัยใช้การทดสอบแบบไขว้ทบ (K-fold cross Validation) แบบ 10 ส่วน แล้วทำการวัดประสิทธิภาพตัวแบบการพยากรณ์โดยใช้เกณฑ์การวัดประสิทธิภาพของตัวแบบรู้จำด้วยวิธี Predictive Modeling ซึ่งประกอบด้วยค่าความแม่นตรง (Precision) ค่าความระลึก (Recall) ค่าประสิทธิภาพโดยรวม (F-Measure) และค่าความแม่นยำ (Accuracy) ดังสมการ (1) - (4) และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) ดังสมการ (5) ตามลำดับ (เอกสิทธิ์ พัชรวงค์ศักดา, 2557)

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F-Measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y-Y)^2} \quad (5)$$

ตารางที่ 3 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์สำหรับหลักสูตรเทคโนโลยีสารสนเทศ

ลำดับ	ชื่อตัวแปร	รายละเอียด
1	S4112106	รายวิชาพื้นฐานการเขียนโปรแกรม
2	S4112143	รายวิชาพื้นฐานด้านเทคโนโลยีสารสนเทศ
3	S4112147	รายวิชาพื้นฐานทางฟิสิกส์
4	S4112148	รายวิชาคณิตศาสตร์สำหรับเทคโนโลยีสารสนเทศ
5	S4112144	รายวิชาแพลตฟอร์มเทคโนโลยี
6	S4112145	รายวิชาเทคโนโลยีเว็บ
7	S4112146	รายวิชาเทคโนโลยี เครื่องมือ และการประยุกต์ใช้งานเครือข่ายสังคมออนไลน์
8	S4112149	รายวิชาสถิติสำหรับเทคโนโลยีสารสนเทศ
9	S4112210	รายวิชาระบบฐานข้อมูล
10	GPA1	ผลการเรียนเฉลี่ยปีที่ 1 ภาคการศึกษาที่ 1
11	GPA2	ผลการเรียนเฉลี่ยปีที่ 1 ภาคการศึกษาที่ 2
12	KYS	ข้อมูลการกู้ยืมเงิน กยศ. ปีที่ 1
13	S4112213	รายวิชาระบบจัดการฐานข้อมูล
14	S4112214	รายวิชาการวิเคราะห์และออกแบบระบบ
15	S4112250	รายวิชาระบบสารสนเทศเพื่อการจัดการ
16	S4112251	รายวิชาเทคโนโลยีเครือข่าย
17	GPA21	ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1
18	GPA22	ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 2
19	KYS1	ข้อมูลการกู้ยืมเงิน กยศ. ปีที่ 2

ตารางที่ 3 (ต่อ)

ลำดับ	ชื่อตัวแปร	รายละเอียด
20	ACCGPA	เกรดเฉลี่ยสะสม
21	FA	อาชีพของบิดา
22	MA	อาชีพของมารดา
23	FA1	รายได้ของบิดา
24	MA1	รายได้ของมารดา
25	PARSTATUSNAME	สถานภาพของบิดา-มารดา
26	STT	สถานภาพของนักศึกษา

ตารางที่ 4 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์สำหรับหลักสูตรวิทยาการคอมพิวเตอร์

ลำดับ	ชื่อตัวแปร	รายละเอียด
1	S4100103	รายวิชาเทคโนโลยีสารสนเทศในชีวิตประจำวัน
2	S4109104	รายวิชาแคลคูลัส 1
3	S4112105	รายวิชาขั้นตอนวิธี
4	S4112147	รายวิชาพื้นฐานทางฟิสิกส์
5	S2105243	รายวิชาภาษาอังกฤษสำหรับวิทยาศาสตร์
6	S4109105	รายวิชาแคลคูลัส 2
7	S4112106	รายวิชาพื้นฐานการเขียนโปรแกรม
8	S4112108	รายวิชาคณิตศาสตร์ดิสครีตสำหรับคอมพิวเตอร์
9	GRADE1	ผลการเรียนเฉลี่ยปีที่ 1 ภาคการศึกษาที่ 1
10	GRADE2	ผลการเรียนเฉลี่ยปีที่ 1 ภาคการศึกษาที่ 2
11	KYS	ข้อมูลการกู้ยืมเงิน กยศ. ปีที่ 1
12	S4112104	รายวิชาวิธีการเชิงตัวเลขสำหรับวิทยาการคอมพิวเตอร์
13	S4112107	รายวิชาโครงสร้างข้อมูล
14	S4112210	รายวิชาระบบฐานข้อมูล
15	S4112211	รายวิชาการเขียนโปรแกรมคอมพิวเตอร์ขั้นสูง
16	S4112251	รายวิชาเทคโนโลยีเครือข่าย
17	GRADE21	ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1
18	GRADE22	ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 2
19	KYS1	ข้อมูลการกู้ยืมเงิน กยศ. ปีที่ 2
20	ACCGPA	เกรดเฉลี่ยสะสม
21	FA	อาชีพของบิดา

Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

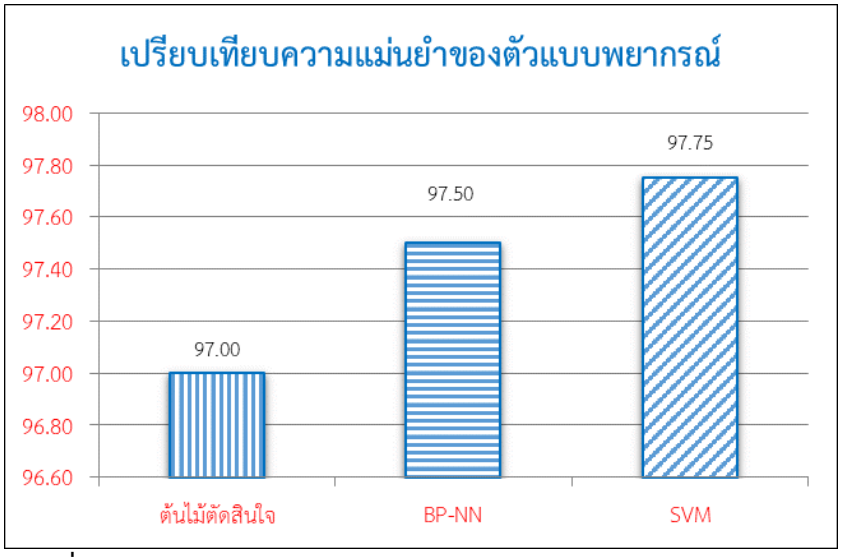
ตารางที่ 4 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์สำหรับหลักสูตรวิทยาการคอมพิวเตอร์ (ต่อ)

ลำดับ	ชื่อตัวแปร	รายละเอียด
22	MA	อาชีพของมารดา
23	FA1	รายได้ของบิดา
24	MA1	รายได้ของมารดา
25	PARSTATUSNAME	สถานภาพของบิดา-มารดา
26	STT	สถานภาพของนักศึกษา

ผลการวิจัย

จากการนำข้อมูลประวัติของนักศึกษาหลักสูตรวิทยาคอมพิวเตอร์ จำนวน 97 ระเบียน และหลักสูตรเทคโนโลยีสารสนเทศ จำนวน 202 ระเบียน ไปทดสอบกับตัวแบบพยากรณ์ 3 ตัว แบบที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียม (BP-NN)

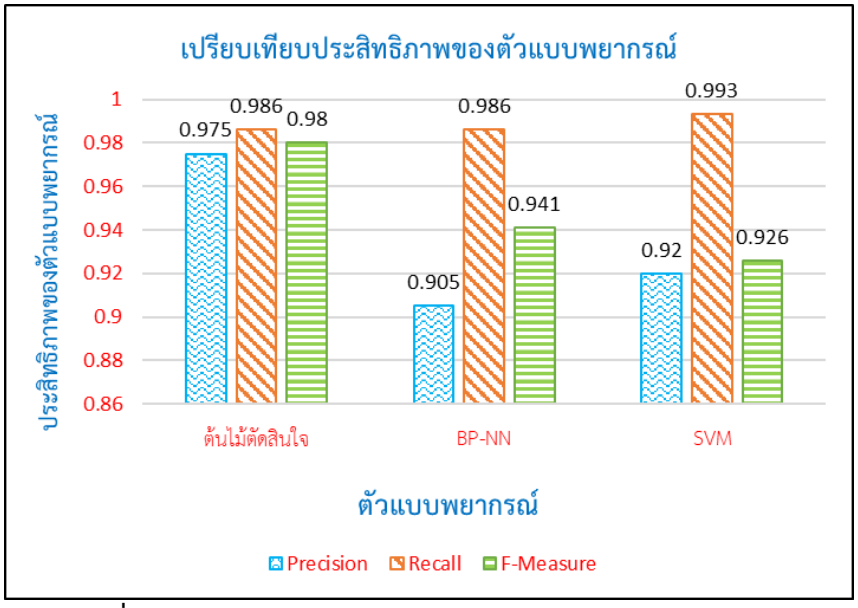
ผลการทดสอบค่าความแม่นยำของตัวแบบพยากรณ์ พบว่า ตัวแบบพยากรณ์ที่ใช้เทคนิค SVM ให้ความแม่นยำในการพยากรณ์สูงสุด ตามด้วยเทคนิค BP-NN และเทคนิคต้นไม้ตัดสินใจ ที่ร้อยละ 97.75 97.50 และ 97.00 ตามลำดับ ดังภาพที่ 4 และผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์โดยใช้ค่าความแม่นตรง ค่าความระลึก และค่าประสิทธิภาพโดยรวม พบว่า ตัวแบบพยากรณ์ต้นไม้ตัดสินใจให้ค่าความแม่นตรงเท่ากับ 0.975 ค่าความระลึกเท่ากับ 0.986 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.980 ตัวแบบพยากรณ์ BP-NN ให้ค่าความแม่นตรงเท่ากับ 0.905 ค่าความระลึกเท่ากับ 0.986 และค่าประสิทธิภาพ โดยรวมเท่ากับ 0.941 และตัวแบบพยากรณ์ SVM ให้ค่าความแม่นตรงเท่ากับ 0.920 ค่าความระลึกเท่ากับ 0.993 และค่าประสิทธิภาพ โดยรวมเท่ากับ 0.926 ดังตารางที่ 5 และภาพที่ 5



ภาพที่ 4 กราฟแสดงการเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์

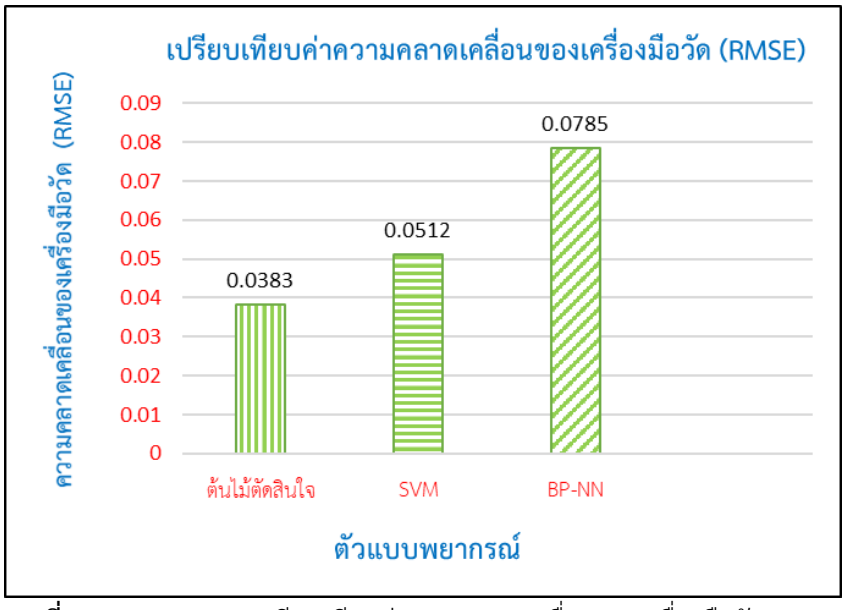
ตารางที่ 5 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

ตัวแบบพยากรณ์	Precision	Recall	F-Measure
SVM	0.920	0.993	0.926
BP-NN	0.905	0.986	0.941
ต้นไม้ตัดสินใจ	0.975	0.986	0.980



ภาพที่ 5 กราฟแสดงการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

นอกจากนี้ เมื่อทำการเปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัด (Root Mean Square Error : RMSE) ของเทคนิคต้นไม้ตัดสินใจมีค่าน้อยสุด ตามด้วยเทคนิค SVM และ BP-NN เท่ากับ 0.0383 0.0512 และ 0.0785 ตามลำดับ ดังภาพที่ 6



ภาพที่ 6 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัด (RMSE)

Veridian E-Journal, Science and Technology Silpakorn University
Volume 5 Number 4 July – August 2018 ISSN 2408 - 1248

ปัจจัยที่ส่งผลในการพยากรณ์การพ้นสภาพการเป็นนักศึกษาหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารเทศ ข้อมูลภูมิหลังไม่ใช่ข้อมูลสำคัญในการทำนายการพ้นสภาพของนักศึกษา ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาหลักสูตรเทคโนโลยีสารสนเทศ มี 3 ปัจจัย ได้แก่ ผลการเรียนรายวิชาพื้นฐานทางฟิสิกส์ ผลการเรียนรายวิชาแพลตฟอร์มเทคโนโลยี และ ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1 สำหรับหลักสูตรวิทยาการคอมพิวเตอร์มีเพียง 1 ปัจจัย ได้แก่ ผลการเรียนรายวิชาโครงสร้างข้อมูล ตามตารางที่ 6

ตารางที่ 6 คุณลักษณะที่เป็นปัจจัยในการพ้นสภาพ

ลำดับ	ชื่อตัวแปร	รายละเอียด	หลักสูตร
1	S4112147	รายวิชาพื้นฐานทางฟิสิกส์	เทคโนโลยีสารเทศ
2	S4112144	รายวิชาแพลตฟอร์มเทคโนโลยี	เทคโนโลยีสารเทศ
3	GPA21	ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1	เทคโนโลยีสารเทศ
4	S4112107	รายวิชาโครงสร้างข้อมูล	วิทยาการคอมพิวเตอร์

สรุปผลและอภิปรายผล

งานวิจัยนี้ได้เทคนิคเหมืองข้อมูลมาใช้ในการพยากรณ์หาปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา ประกอบด้วย เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (Back Propagation Neural Network :BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มาทำการสร้างตัวแบบพยากรณ์ สำหรับหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา ซึ่งใช้ข้อมูลของนักศึกษาที่กำลังศึกษาและสำเร็จการศึกษาแล้วระหว่างปีการศึกษา 2555 – 2558 ของหลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 97 ระเบียน และหลักสูตรเทคโนโลยีสารสนเทศ จำนวน 202 ระเบียน ประกอบด้วย 26 คุณลักษณะ โดยใช้เทคนิคเหมืองข้อมูล ประกอบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (Back Propagation Neural Network :BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลการทดสอบพบว่า ข้อมูลภูมิหลังไม่ใช่ข้อมูลสำคัญในการทำนายการพ้นสภาพของนักศึกษา โดยปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาหลักสูตรเทคโนโลยีสารสนเทศ มี 3 ปัจจัย ได้แก่ ผลการเรียนรายวิชาพื้นฐานทางฟิสิกส์ ผลการเรียนรายวิชาแพลตฟอร์มเทคโนโลยี และ ผลการเรียนเฉลี่ยปีที่ 2 ภาคการศึกษาที่ 1 สำหรับหลักสูตรวิทยาการคอมพิวเตอร์มีเพียง 1 ปัจจัย ได้แก่ ผลการเรียนรายวิชาโครงสร้างข้อมูล สำหรับตัวแบบพยากรณ์ที่ใช้เทคนิค SVMs ให้ความแม่นยำในการพยากรณ์สูงสุดที่ร้อยละ 97.75 โดยมีการประเมินค่าความแม่นตรงเท่ากับ 0.920 ค่าความระลึกเท่ากับ 0.993 และประสิทธิภาพโดยรวม เท่ากับ 0.926 ตัวแบบพยากรณ์ที่ใช้เทคนิค BP-NN ให้ความแม่นยำในการพยากรณ์ที่ร้อยละ 97.50 โดยมีการประเมินค่าความแม่นตรงเท่ากับ 0.905 ค่าความระลึกเท่ากับ 0.986 และประสิทธิภาพโดยรวมเท่ากับ 0.941 และ ตัวแบบพยากรณ์ที่ใช้เทคนิคต้นไม้ตัดสินใจให้ความแม่นยำในการพยากรณ์ที่ร้อยละ 97.00 โดยมีการประเมินค่าความแม่นตรงเท่ากับ 0.975 ค่าความระลึกเท่ากับ 0.986 และประสิทธิภาพโดยรวมเท่ากับ 0.980 ทั้งนี้ค่า RMSE ของเทคนิคต้นไม้ตัดสินใจมีค่าน้อยสุด ตามด้วยเทคนิค SVM และ BP-NN เท่ากับ 0.0383 0.0512

และ 0.0785 ตามลำดับ ซึ่งสอดคล้องกับผลการวิจัยของ ประพัฒน์ พรมน้ำอ่าง วสุวรรธน์ พงศ์ขจร และ นิเวศ จิระวิชิตชัย (2559) รังสิพรรณ มฤคทัต (2558) และงานวิจัยของ พรรณิภา บุตรเอก และ สุรเดช บุญลือ (2557)

จึงสรุปได้ว่า การใช้เทคนิคเหมืองข้อมูลแบบซัพพอร์ตเวกเตอร์แมชชีนเหมาะสมที่สุดกับการหาปัจจัยที่ใช้ในการสร้างตัวแบบพยากรณ์ในงานวิจัยนี้ เมื่อพิจารณาคุณลักษณะที่ส่งผลในการพยากรณ์ปัจจัยการพ้นสภาพการเป็นนักศึกษาหลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารเทศ ปัจจัยที่สำคัญมี 4 ปัจจัย คือ รายวิชาเนื้อหาเฉพาะของหลักสูตรและผลการเรียนเฉลี่ยในปีที่ 2 ภาคการศึกษาที่ 1 โดยข้อมูลภูมิหลังไม่ใช่ข้อมูลสำคัญในการทำนายการพ้นสภาพของนักศึกษา เนื่องจากสภาพทั่วไปของนักศึกษาในมหาวิทยาลัยราชภัฏยะลาเป็นนักศึกษาที่มีฐานะยากจนมากกว่าร้อยละ 80 ใช้เงินจากกองทุนเงินให้กู้ยืมเพื่อการศึกษามากกว่าร้อย 90 เป็นนักศึกษาในพื้นที่ 3 สามจังหวัดชายแดนภาคใต้ คือ ปัตตานี ยะลา และนราธิวาส

ข้อเสนอแนะในการวิจัยพบว่า การใช้เทคนิคเหมืองข้อมูลเป็นการค้นหาความรู้จากฐานข้อมูลที่มีอยู่มาพยากรณ์ปัจจัยการพ้นสภาพการเป็นนักศึกษา ซึ่งในการวิจัยครั้งต่อไปควรมีการเพิ่มปัจจัยอื่น ๆที่น่าจะมีผลต่อการเรียนของนักนักศึกษามาพิจารณาเป็นปัจจัยในการพยากรณ์และควรมีข้อมูลมากพอในการทำเหมืองข้อมูล อาจใช้เทคนิคเหมืองข้อมูลเทคนิคอื่น ๆ มาเปรียบเทียบเพื่อให้ได้ผลการพยากรณ์ที่มีประสิทธิภาพมากขึ้น

ผลของงานวิจัยนี้ให้ประโยชน์ในการวิเคราะห์ปัจจัยการพ้นสภาพการเป็นนักศึกษา โดยจะช่วยเป็นแนวทางให้นักศึกษาสามารถวางแผนการเรียนได้ เช่น การเพิ่มถอนรายวิชาให้เหมาะสมกับศักยภาพตนเอง และ การพัฒนาตนเองในการเรียนวิชาต่างๆ ที่มีความสำคัญ

เอกสารอ้างอิง

ภาษาไทย

ชิดชนก ศรีชัยวงศ์ ไพศาล ตระกูลสุข และสุรเดช บุญลือ.(2557). **ระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจ**. Veridian E-Journal Science and Technology Silpakorn University ปีที่ 1, ฉบับที่ 6 (พฤศจิกายน – ธันวาคม) : 1-14.

นิภาพร ชนะมาร และ พรรณี สิทธิเดช. 2557. **การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์**. วารสารมหาวิทยาลัยราชภัฏสกลนคร ปีที 6, ฉบับที่ 12 (กรกฎาคม – ธันวาคม) : 31-45.

ประพัฒน์ พรมน้ำอ่าง วสุวรรธน์ พงศ์ขจร และ นิเวศ จิระวิชิตชัย. 2559. **การจำแนกกลุ่มข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล**. วารสารวิทยาศาสตร์และเทคโนโลยี, มทร.ธัญบุรี ปีที่ 6, ฉบับที่ 1 (มกราคม - มิถุนายน) : 94-101.

พรพล ธรรมรงค์รัตน์ ลัดดา ปรีชาวีรกุล และวิภาดา เวทย์ประสิทธ์. 2553. **การจำแนกประเภทเว็บเพจโดยใช้ค่าความถี่เอกสารและซัพพอร์ตเวกเตอร์แมชชีน**. The 12th National Computer Science and Engineering Conference 2008.

พรรณิภา บุตรเอก และสุรเดช บุญลือ.(2557). **การพยากรณ์โอกาสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน**. Veridian E-Journal Science and Technology Silpakorn University ปีที่ 1, ฉบับที่ 6 (พฤศจิกายน – ธันวาคม) : 40-49.

รังสิพรรณ มฤคทัต. (2558). **การระบุตัวผู้เขียนข้อความออนไลน์ภาษาไทยด้วยซัพพอร์ตเวกเตอร์แมชชีนแลต้นไม้ตัดสินใจ**. วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 25, ฉบับที่ 1 (ม.ค. - เม.ย.) : 103-111.

วิภาวรรณ บัวทอง. 2559. **การทำเหมืองข้อมูล : Data Mining**. ภูเก็ต : มหาวิทยาลัยราชภัฏภูเก็ต.

สายชล สินสมบูรณ์ทอง. (2558). **การทำเหมืองข้อมูล (Data Mining)**. กรุงเทพฯ : จามจุรีโปรดักท์.

เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล. (2558). **การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์**. Veridian E-Journal Science and Technology Silpakorn University ปีที่ 1, ฉบับที่ 6 (พฤศจิกายน–ธันวาคม) : 1-17.

เอกสิทธิ์ พัชรวงค์ศักดา. (2557). **การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น**. กรุงเทพฯ : เอเชีย ดิจิตอล การพิมพ์.