Received: Apr 15, 2021 **Revised:** May 20, 2021 **Accepted:** Jun 5, 2021

การพยากรณ์สถานะการเตรียมความพร้อมก่อนเข้าเรียนสำหรับวิชาหลักการโปรแกรมด้วยเทคนิคการทำเหมืองข้อมูล

FORECASTING THE PREPARATORY STATUS FOR THE PRINCIPLES OF PROGRAMMING USING DATA MINING TECHNIQUES

นันทวัน นาคอร่าม[1], สุนิพันธ์ ศรีสุพจนานนท์[2] ศิโรรัตน์ สายปัญญา[3]
[1]คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี [2]คณะบริหารธุรกิจ มหาวิทยาลัยธนบุรี
[3]สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
Nantawan Nakaram[1], Suniphan Srisuphotnanont[2], Sirorat Saypunya[3]
[1]Faculty of Science and Technology, Thonburi University [2]Faculty of Business Administration, Thonburi University, [3]Faculty of Science, Rajabhat University
E-mail: Nantawan_it@thonburi-u.ac.th

บทคัดย่อ

ปัจจุบันสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ได้เปิดสอนในรายวิชาหลักการโปรแกรม ซึ่งถือเป็นวิชาพื้นฐานสำคัญของหลักสูตรที่นักศึกษาสามารถนำความรู้ไปประยุกต์ใช้ต่อในรายวิชาที่สูงขึ้น โดยสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี ได้เปิดรับสมัครนักศึกษาที่จบการศึกษาระดับชั้นประกาศนียบัตรวิชาชีพชั้นสูง ทั้งในสาขาเทคโนโลยีสารสนเทศ สาขาคอมพิวเตอร์ธุรกิจ และสาขาอิเล็กทรอนิกส์ ทำให้พื้นฐานการเขียนโปรแกรมของนักศึกษาแต่ละคนแตกต่างกันส่งผลให้เกิดปัญหาในการเรียนรายวิชาหลักการโปรแกรมในชั้นเรียน ผู้วิจัยจึงมีแนวคิดในการพัฒนาวิธีการพยากรณ์ถึงความจำเป็นของนักศึกษาที่จะต้องได้รับการเตรียมความพร้อมก่อนเข้าเรียน โดยมีวัตถุประสงค์เพื่อ 1) สังเคราะห์โมเดลสำหรับการพยากรณ์ความจำเป็น (สถานะ) ของการเตรียมความพร้อมก่อนเรียน และ 2) เปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลของโมเดลด้วยเทคนิค Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine (SVM) โดยรวบรวมข้อมูลนักศึกษาชั้นปีที่ 1 ระหว่างปีการศึกษา 2559 ถึง 2563 มีปัจจัยที่เกี่ยวข้อง 7 แอททริบิวต์ และจำนวน 376 ชุดข้อมูล เมื่อทำการวิเคราะห์ค่าน้ำหนักของแอททริบิวต์พบว่ามีปัจจัยที่เกี่ยวข้องในการพยากรณ์สถานะของการเตรียมความพร้อมก่อนเรียน จำนวน 7 ปัจจัย และนำปัจจัยที่ได้มาทำการสร้างโมเดลทดสอบผลลัพธ์ด้วยวิธี 10-Fold Cross Validation และวัดประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) เพื่อหาวิธีการที่มีความถูกต้องมากที่สุด ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลพบว่าโมเดลที่สร้างด้วยเทคนิค Support Vector Machine (SVM) มีประสิทธิภาพสูงสุดมีค่าความถูกต้อง 95.58% และมีปัจจัยที่เกี่ยวข้องสูงสุด 5 อันดับ ได้แก่ ผลการทดสอบก่อนเรียน สาขาวิชาที่จบ ผลการเรียนก่อนจบการศึกษา ผลการเรียนในรายวิชาการเขียนโปรแกรมเบื้องต้น และผลการเรียนในรายวิชาวิชาคณิตศาสตร์ ตามลำดับ

คำสำคัญ: การทำเหมืองข้อมูล; การเตรียมความพร้อมก่อนเรียน; การพยากรณ์สถานะการเตรียมความพร้อม; หลักการโปรแกรม

Abstract

At present, Information Technology Program in the Faculty of Science and Technology at Thonburi University was taught in the course of the principle of programming. This course is an important foundation course in which students can apply their knowledge to higher courses in the field of Information Technology, Faculty of Science and Technology. It is accepting applications for students who

have graduated with a diploma in Information Technology, Business Computing, and Electronics. This results in the programming fundamentals of each student being different, being the cause of learning the principle of programming in the classroom. Therefore, the researcher has an idea to develop a method to forecast the need for students to prepare before entering to the class. The objectives of the research were 1) to synthesize the model for forecasting the necessity (status) of preschool preparation and 2) to compare the efficiency of classification of the model by Decision Tree, Naïve Bayes, Neural Network, and Support Vector Machine techniques (SVM). The collecting data of first-year students during the academic year 2016 to 2020, there were 7 attributes and 376 data sets. The weight analysis of the attribute, it was found that there were 7 factors related to forecasting the status of pre-school preparation. The results were tested with a 10-fold cross-validation method and the performance was measured with accuracy to determine the most accurate method. The results of the comparison of the efficiency of data classification revealed that the model created using the Support Vector Machine (SVM) technique had the highest efficiency with 95.58% accuracy and the most 5 relevant factors were pre-test results, graduated field, Pre-graduation results, grades in introductory programming courses, and grades in mathematics courses respectively.

Keywords: Data mining; Preparation before study; Forecasting the status of pre-school preparation; Principle Programming Erg

บทนำ

ปัจจุบันเทคโนโลยีสารสนเทศได้เข้ามามีบทบาทอย่างมากต่อการดำรงชีวิตของมนุษย์ในสังคมยุคโลกาภิวัตน์ เทคโนโลยีสารสนเทศเป็นรากฐานสำคัญในการพัฒนาที่มีบทบาทต่อทุกสถาบันการศึกษา ผู้เรียนจำเป็นต้องแสวงหาความรู้และเรียนรู้อยู่ตลอดเวลา รู้จักการใช้เทคโนโลยีสารสนเทศและการสื่อสารเพื่อให้เกิดประโยชน์ต่อการพัฒนาตนเองและสังคมอย่างต่อเนื่อง ระบบการเรียนการสอนโดยใช้เทคโนโลยีสารสนเทศเป็นฐานนั้นช่วยพัฒนาผลสัมฤทธิ์ทางการเรียน และความสามารถในการใช้เทคโนโลยีสารสนเทศในการเรียนรู้ของผู้เรียนให้สูงขึ้น ซึ่งเป็นคุณลักษณะอันพึงประสงค์ของผู้เรียนในสังคมปัจจุบัน และถือเป็นสมรรถนะหลักของบุคคลในศตวรรษที่ 21

การเขียนโปรแกรมคอมพิวเตอร์ถือว่าเป็นการผสมผสานกันระหว่างศาสตร์ของวิทยาศาสตร์และคณิตศาสตร์เข้าด้วยกัน ผู้เขียนต้องเข้าใจขั้นตอนการแก้ไขปัญหาและรูปแบบของภาษาที่จะใช้เป็นอย่างดี จึงสามารถเขียนโปรแกรมที่ควบคุมเครื่องคอมพิวเตอร์ให้ทำงานได้ตามความต้องการ [3][4][7] โดยประกอบด้วย 5 ขั้นตอน ได้แก่ การวิเคราะห์ปัญหา การออกแบบโปรแกรม การเขียนโปรแกรม การทดสอบโปรแกรม และการจัดทำเอกสารประกอบการเขียนโปรแกรม [1] หลักการเขียนโปรแกรมคอมพิวเตอร์เป็นขั้นตอนจัดทำโปรแกรมที่ช่วยให้การเขียนโปรแกรมเป็นไปอย่างมีประสิทธิภาพ และได้ผลลัพธ์ตามวัตถุประสงค์ เนื่องจากแต่ละขั้นตอนจะช่วยให้เกิดความเป็นระเบียบ การเรียบเรียงแนวคิดมีความชัดเจนไม่สับสนและง่ายต่อการเขียนหรือพัฒนาโปรแกรม

คณะวิทยาศาสตร์และเทคโนโลยี สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยธนบุรี ได้เปิดสอนรายวิชาหลักการโปรแกรม ซึ่งถือเป็นความรู้พื้นฐานที่สำคัญสำหรับนักศึกษาสาขาเทคโนโลยีสารสนเทศที่จะต้องมีความรู้ความเข้าใจ และสามารถเขียนโปรแกรมเพื่อประยุกต์ใช้กับงานต่างๆ ทั้งภาคธุรกิจและอุตสาหกรรมได้ อีกทั้งความรู้เรื่องหลักการโปรแกรมจะถูกใช้เป็นพื้นฐานสำหรับการเรียนในรายวิชาที่สูงขึ้น เช่น การเขียนโปรแกรมบนเว็บ การพัฒนาโปรแกรมประยุกต์สำหรับอุปกรณ์เคลื่อนที่ การเขียนสคริปต์เพื่อควบคุมการทำงานของระบบเครือข่าย เป็นต้น โดยนักศึกษาที่ลงทะเบียนเรียนในรายวิชาดังกล่าว คือ นักศึกษาชั้นปีที่ 1 ที่จบการศึกษาในระดับประกาศนียบัตรวิชาชีพชั้นสูง จากสาขาวิชาเทคโนโลยีสารสนเทศ สาขาวิชาคอมพิวเตอร์ธุรกิจ และสาขาวิชาอิเล็กทรอนิกส์ ทำให้การเรียนในชั้นเรียนมีนักศึกษาที่มีระดับความรู้

พื้นฐานที่แตกต่างกันส่งผลกระทบต่อการเรียนที่นักศึกษาที่มีระดับความรู้พื้นฐานที่สูงกว่าจะต้องมารอผู้ที่มีระดับความรู้ที่ต่ำกว่าเพื่อให้มีพื้นฐานที่ใกล้เคียงกันจึงจะเปลี่ยนไปเรียนในหัวข้อถัดๆ ไป

จากปัญหาข้างต้น ผู้วิจัยจึงมีแนวคิดที่จะทำการวิเคราะห์ข้อมูลเพื่อพยากรณ์ความจำเป็น (สถานะ) ของการเตรียมความพร้อมก่อนเรียน และเปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลของโมเดลด้วยเทคนิค Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine (SVM) โดยพิจารณาจากปัจจัยที่เกี่ยวข้อง ได้แก่ ข้อมูลการเรียนในระดับชั้นประกาศนียบัตรวิชาชีพ และข้อมูลผลการทดสอบก่อนเรียนในรายวิชาหลักการโปรแกรม ที่ดำเนินการโดยสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี เพื่อนำข้อมูลที่ได้จากการพยากรณ์ไปใช้ประกอบการพัฒนาระบบสนับสนุนการตัดสินใจในการระบุว่านักศึกษาคนใดที่จะต้องได้รับการเตรียมความพร้อมก่อนเรียน

การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง

1 ปัจจัยที่ส่งผลต่อความสำเร็จการเรียนวิชาหลักการโปรแกรม

จากการทบทวนวรรณกรรมที่เกี่ยวข้องกับการพยากรณ์การประสบความสำเร็จทางการเรียนด้านคอมพิวเตอร์และการเรียนวิชาหลักการโปรแกรม พบว่าปัจจัยที่ส่งผลต่อการประสบความสำเร็จนั้นมีอยู่หลายปัจจัย สามารถสรุปได้ดังนี้

1) เพศ (Sex) [3][4] เป็นปัจจัยหนึ่งที่เกี่ยวข้องกับความสำเร็จที่ถูกกล่าวถึงในหลายงานวิจัย เนื่องจากเพศชายกับเพศหญิงจะมีลักษณะและความสามารถในการเรียนรู้ที่แตกต่างกัน
2) ผลการเรียน (Grade) [4][5][6] เป็นปัจจัยที่สะท้อนให้เห็นถึงประสิทธิภาพของผลการศึกษาที่ผ่านมา ทั้งในเรื่องความรับผิดชอบในการเรียนและความสามารถในการศึกษาเรียนรู้
3) พื้นฐานในการเขียนโปรแกรม (Programming Background) [3][4][5][6][9] เป็นปัจจัยหลักที่สะท้อนให้เห็นความรู้ความสามารถพื้นฐานทางด้านการเขียนโปรแกรมที่มีอยู่ ณ ปัจจุบันของผู้เรียน
4) พื้นฐานทางคณิตศาสตร์ (Mathematics Background) [3][4][6][9] สะท้อนให้เห็นว่าผู้ที่มีความรู้พื้นฐานทางคณิตศาสตร์ที่ดีจะมีทักษะในการคิดและเข้าใจถึงลำดับโครงสร้างในการทำงาน

จากการทบทวนวรรณกรรมข้างต้น ผู้วิจัยได้พิจารณาเลือกทั้ง 4 ปัจจัย เพิ่มเติมปัจจัยสาขาวิชาที่นักศึกษาจบจากระดับชั้นประกาศนียบัตรวิชาชีพชั้นสูง เนื่องจากแต่ละสาขาวิชาจะเน้นศาสตร์ความรู้และการประยุกต์ใช้ทางด้านคอมพิวเตอร์ที่แตกต่างกัน และปัจจัยสุดท้ายที่นำมาพิจารณาร่วม คือ ผลการทดสอบก่อนเรียนของวิชาหลักการโปรแกรมที่ดำเนินการโดยสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี

2 การทำเหมืองข้อมูล (Data Mining) เป็นเทคนิคการวิเคราะห์ข้อมูลเพื่อค้นหารูปแบบ (Pattern) ความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติและการเรียนรู้ของเครื่องคอมพิวเตอร์แล้วนำผลวิเคราะห์ที่ได้ไปใช้ประโยชน์เพื่อการตัดสินใจ

3 ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคในการจำแนกกฎโดยจะตรวจสอบข้อมูลและสร้างต้นไม้เพื่อการพยากรณ์ สำหรับโครงสร้างของต้นไม้ตัดสินใจจะมีการแตกแขนงไปตามเงื่อนไขหรือเส้นทางของกิ่งไม้ และข้อมูลที่คาดคะเนว่าจะเกิดขึ้นจะใช้กฎในรูปแบบ if-then rule ต้นไม้ตัดสินใจจะคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดมาเป็นโหนดราก (Root node) จากนั้นจะเลือกคุณสมบัติที่มีความสัมพันธ์ถัดไปเรื่อย จากการคำนวณ Information Gain (IG) โดยเลือกคุณสมบัติที่มีค่า IG สูงที่สุด คำนวณได้จากสมการ

$$IG(parent, child) = Entropy(parent) - [p(c1)xEntropy(c1) + p(c2)xEntropy(c2) + ...] \quad (1)$$

เมื่อ Entropy(c1) คือ $-p(c1) \log p(c1)$

p(c1) คือ ค่าความน่าจะเป็นของค่า c1

c คือ ปัจจัยแต่ละตัวที่เกี่ยวข้อง

ซึ่งค่า Entropy นี้จะใช้ในการวัดความแตกต่างกันของข้อมูล ถ้าข้อมูลมีความแตกต่างกันน้อยค่า Entropy จะมีค่าต่ำ ถ้าข้อมูลมีความแตกต่างกันมากค่า Entropy จะมีค่าสูง Decision Tree จะคำนวณค่า IG ของแต่ละแอททริบิวต์เทียบกับคลาสเพื่อหาแอททริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของโครงสร้างต้นไม้

4 Naïve Bayesian เป็นเทคนิคการทำเหมืองข้อมูลที่ใช้การคัดแยกประเภทข้อมูลที่อาศัยหลักการความน่าจะเป็น (Probability) ตามทฤษฎีของเบย์ เป็นขั้นตอนวิธีในการจำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ มีสมมติฐานว่าปริมาณของความสนใจขึ้นกับการกระจายความน่าจะเป็น ถือเป็นเทคนิคในการแก้ปัญหาที่สามารถคาดการณ์ผลลัพธ์ได้ โดยทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณลักษณะ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน โดยกำหนดให้ความน่าจะเป็นของข้อมูลดังสมการ

$$P(C \mid A) = P(A \mid C) \times P(C) / P(A) \qquad (2)$$

P(C|A) คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอททริบิวต์ A จะมีคลาส C

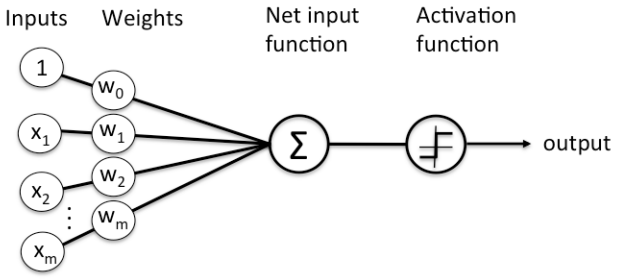
P(A|C) คือ ค่าความน่าจะเป็นที่ข้อมูล Training data ที่มีคลาส C และมีแอททริบิวต์ A โดยที่ A = a1∩a2...∩aM เมื่อ M คือ จำนวนของแอททริบิวต์ใน Training Data

P(C) คือ ความน่าจะเป็นของคลาส C

P(A) คือ ความน่าจะเป็นของคลาส A

5 Neural Network เป็นเทคนิคการทำเหมืองข้อมูลที่ใช้โมเดลทางคณิตศาสตร์มาประมวลผลสารสนเทศด้วยการคำนวณแบบ Connectionist โดยได้แนวคิดมาจากการจำลองการทำงานของเซลล์สมองมนุษย์ ที่แต่ละเซลล์ประสาทจะประกอบไปด้วยเดนไดรท์ (Dendrite) ซึ่งเป็นตัว input ของเซลล์ และแอกซอน (Axon) เป็นเสมือน output ของเซลล์ที่จะส่งกระแสประสาทไปยังเซลล์ตัวอื่น ถ้าสมองได้รับกระแสไฟฟ้ามากพอจะทำให้เซลล์ส่งต่อกระแสประสาทไปเรื่อยๆ

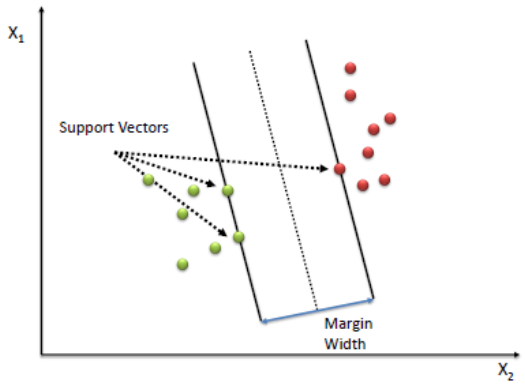
โครงสร้างของประสาทเทียมจะประกอบด้วย input และ output โดยแบ่งเป็นชั้นหรือ layer ซึ่งจะมีชั้นคั่นตรงกลางคือ hidden layer โดยโครงสร้างประสาทเทียมจะมีหน่วยย่อยเรียกว่า perceptron ซึ่งเทียบเท่าได้กับเซลล์สมองของมนุษย์หนึ่งเซลล์



ภาพที่ 1 Schematic of Rosenblatt's perceptron (Source: [7])

6 ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นอัลกอริทึมที่นำมาใช้ในการวิเคราะห์และจำแนกข้อมูล โดยอาศัยโมเดลทางคณิตศาสตร์เรื่องการหาสัมประสิทธิ์ของสมการ เพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นเลือกเส้นที่แบ่งข้อมูลได้ดีที่สุด

หลักการของอัลกอริทึมคือ การนำข้อมูลที่ใช้สอนมากระจายเป็นเวกเตอร์ในระนาบ หรือ สเปซ N มิติ (Feature Space) จากนั้นคำนวณหาเส้นไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวคเตอร์อินพุทออกเป็นประเภทต่างๆ โดยเส้นไฮเปอร์เพลนต้องเป็นเส้นที่มีค่า margin หรือระยะห่างระหว่างจุดของข้อมูลที่อยู่ใกล้กับไฮเปอร์เพลนทั้งสองด้าน คือ d+ และ d- มากที่สุด เพราะถ้าค่า margin น้อย แสดงว่าอาจมีจุดหนึ่งของไฮเปอร์เพลนที่อยู่ใกล้ข้อมูลของแต่ละคลาสมากเกินไป ทำให้ข้อมูลใหม่ที่อยู่ห่างออกไปเล็กน้อยเกิดการทำนายผิดพลาดได้ โดยข้อมูลที่อยู่บนขอบของ margin เรียกว่า support vector



ภาพที่ 2 Support Vector Machine (Source: [8])

7 การวัดประสิทธิภาพ

การวัดประสิทธิภาพของโมเดลการพยากรณ์ที่สร้างด้วยวิธีการ 10-Fold Cross Validation จะแบ่งข้อมูลออกหลายส่วน โดยการวิจัยนี้กำหนดให้มีการแบ่งข้อมูลออกเป็น 10 ส่วน แต่ละส่วนมีจำนวนข้อมูลเท่ากัน จากนั้นข้อมูลส่วนหนึ่งจะใช้สำหรับทดสอบประสิทธิภาพของโมเดลทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ ซึ่งปัจจัยที่ใช้ในการวัดประสิทธิภาพประกอบด้วย

- ค่าความถูกต้อง (Accuracy) คือ จำนวนข้อมูลที่ทำนายถูกทุกคลาส เป็นการวัดความถูกต้องของโมเดลโดยพิจารณารวมทุกกรณี
- ค่าความแม่นยำ (Precision) คือ ค่าที่ใช้พิจารณาสิ่งที่ทำนายออกมาแล้วว่าทายถูกกี่เปอร์เซ็นต์
- ค่าความระลึก (Recall) คือ จำนวนข้อมูลที่ทำนายถูก เป็นการวัดความถูกต้องของโมเดล
- ค่าความถ่วงดุล (F-measure) คือ ค่าเฉลี่ยของค่าความแม่นยำและค่าความระลึก
-

วิธีดำเนินการวิจัย

งานวิจัยนี้ได้เสนอวิธีการทำเหมืองข้อมูลโดยใช้โปรแกรม Rapid Miner Studio 9 โดยมีกระบวนการวิเคราะห์ข้อมูลด้วย Cross-Industry Standard Process for Data Mining: CRISP-DM ดังนี้

1 Business Understanding เป็นกระบวนการกำหนดวัตถุประสงค์สำหรับการวิเคราะห์เหมืองข้อมูลเพื่อพยากรณ์ถึงสถานะการเตรียมความพร้อมก่อนเข้าเรียนในรายวิชาหลักการโปรแกรมของนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ชั้นปีที่ 1 ที่จบการศึกษามาจากระดับชั้นประกาศนียบัตรวิชาชีพชั้นสูง โดยมีปัจจัยที่เกี่ยวข้อง คือ สาขาวิชาและโรงเรียนที่จบการศึกษา ผลการเรียนเมื่อจบการศึกษา เพศ ผลการเรียนในรายวิชาที่สะท้อนความสำเร็จในการเรียนวิชาหลักการโปรแกรม และผลการทดสอบก่อนเข้าเรียนวิชาหลักการโปรแกรมที่ดำเนินการโดยสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี

2 Data Understanding เป็นการรวบรวมข้อมูลของนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ชั้นปีที่ 1 ที่จบการศึกษาในระดับชั้นประกาศนียบัตรวิชาชีพชั้นสูง ตั้งแต่ปีการศึกษา 2559 ถึง 2563 ตามรอบระยะเวลาการดำเนินการของหลักสูตร ประกอบด้วยข้อมูล 2 ส่วน คือ 1) ข้อมูลที่ได้จากใบระเบียนแสดงผลการเรียนตามหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) ที่จะแสดงถึงเพศ สาขาวิชาและสถานศึกษาที่จบ ผลการเรียนเมื่อจบการศึกษา ผลการเรียนในรายวิชาคณิตศาสตร์ วิชาการเขียนโปรแกรมเบื้องต้น และวิชาเทคโนโลยีสารสนเทศเพื่อการจัดการอาชีพ และ 2) ข้อมูลผลการทดสอบก่อนเรียนในรายวิชาหลักการโปรแกรม ที่ทางสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ได้ทำการทดสอบนักศึกษาทุกคนก่อนเริ่มเรียนในรายวิชาดังกล่าว โดยจัดเก็บ Data set เป็น 2 ชุด คือ Training data set จำนวน 376 รายการ และ Test set จำนวน 53 รายการ รวมทั้งสิ้น 429 รายการ

3 Data Selection เป็นกระบวนการคัดเลือกเฉพาะข้อมูลที่เกี่ยวข้อง ประกอบด้วย 9 แอททริบิวต์ ดังตารางที่ 1

ตารางที่ 1 รายละเอียดแอททริบิวต์ที่ใช้สำหรับดำเนินงานวิจัย

No.	Attribute (Unit)	Type	Description
1	Sex	Integer	เพศ กำหนดให้ 1 แทนเพศชาย และ 2 แทนเพศหญิง
2	Major	Integer	สาขาวิชาที่เรียน กำหนดให้ 1 แทนสาขาวิชาเทคโนโลยีสารสนเทศ 2 แทนสาขาวิชาคอมพิวเตอร์ธุรกิจ 3 แทนสาขาวิชาอิเล็กทรอนิกส์
3	School	Integer	สถานศึกษา
4	Grade	Real	ผลการเรียนเมื่อจบการศึกษา
5	Math	Real	ผลการเรียนวิชาคณิตศาสตร์
6	Program	Real	ผลการเรียนวิชาการเขียนโปรแกรมเบื้องต้น
7	IT	Real	ผลการเรียนวิชาเทคโนโลยีสารสนเทศเพื่อการจัดการอาชีพ
8	Pretest	Integer	ผลการทดสอบก่อนเรียนในรายวิชาหลักการโปรแกรม
9	Preparation (0 or 1)	Binomial	สถานะการเตรียมความพร้อมก่อนเข้าเรียน กำหนดให้ 0 แทนสถานะของผู้ที่จะต้องเตรียมความพร้อม 1 แทนสถานะของผู้ที่ผ่านการเตรียมความพร้อม

จากรายละเอียดข้อมูลทั้งหมด 9 แอททริบิวต์ข้างต้น ผู้วิจัยได้ทำการพิจารณาพบว่า School เป็นปัจจัยละเอียดอ่อนที่ต่อการพยากรณ์และสรุปผลของข้อมูล และเนื่องด้วยหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูงของแต่ละสถานศึกษามีโครงสร้างและรายละเอียดที่เหมือนกันตามที่สำนักงานคณะกรรมการการอาชีวศึกษากำหนด ผู้วิจัยจึงได้ทำการตัดแอททริบิวต์ School ออกไป

4 Data Transformation เป็นการแปลงข้อมูลเพื่อให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ข้อมูลตามอัลกอริทึมของเหมืองข้อมูลที่เลือกใช้ และกำหนดให้ Preparation เป็นข้อมูลที่ใช้ทดสอบคำตอบ (Class) จะแบ่งเป็น 2 ประเภท คือ False หมายถึง สถานะของผู้ที่จะต้องเตรียมความพร้อม จำนวน 284 รายการ และ True หมายถึง สถานะของผู้ที่ผ่านการเตรียมความพร้อม จำนวน 92 รายการ ดังตารางที่ 2

ตารางที่ 2 จำนวนแอททริบิวต์แยกตามคลาส Occupancy

Index	Nominal value	Absolute count	Fraction
1	False	284	0.755
2	True	92	0.245

5 Modeling เป็นขั้นตอนการสร้างและทดสอบความถูกต้องของโมเดล วิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลทำการสังเคราะห์โมเดลจากข้อมูลที่มีอยู่ เพื่อพยากรณ์ถึงความถูกต้องในการระบุสถานะของการเตรียมความพร้อมก่อนเรียนในรายวิชาหลักการโปรแกรมที่ได้จากข้อมูลเพศ สาขาวิชา ผลการเรียนก่อนจบการศึกษา ผลการเรียนในรายวิชาคณิตศาสตร์ วิชาการเขียนโปรแกรมเบื้องต้น วิชาเทคโนโลยีสารสนเทศเพื่อการจัดการอาชีพ และผลการทดสอบก่อนเข้าเรียนวิชาหลักการโปรแกรมที่ดำเนินการโดยสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี การวิจัยครั้งนี้ผู้วิจัยได้วิเคราะห์ค่าน้ำหนักของแอททริบิวต์และพบว่าผลการทดสอบก่อนเรียน (Pretest) มีน้ำหนักสูงสุด รองลงมาคือ สาขาวิชาที่จบ (Major) ผลการเรียนก่อนจบการศึกษา (Grade) ผลการเรียนในรายวิชาการเขียนโปรแกรมเบื้องต้น (Program) และวิชาคณิตศาสตร์ (Math) ตามลำดับ ดังตารางที่ 3 และทำการสร้างโมเดลการทำนายด้วยเหมืองข้อมูล 4 เทคนิค คือ 1) Decision Tree 2) Naïve Bayes 3) Neural Network และ 4) Support Vector Machine (SVM) ทดสอบผลลัพธ์ด้วยวิธี 10-Fold

validation ในการวัดประสิทธิภาพของการจำแนกประเภทข้อมูล ได้แก่การหาค่าความถูกต้อง ความแม่นยำ ความระลึก และค่าความถ่วงดุล

ตารางที่ 3 Weighting to Data (IG)

Row No.	Attribute	Weight
1	Pretest	0.390
2	Major	0.304
3	Grade	0.252
4	Program	0.200
5	Math	0.162
6	IT	0.017
7	Sex	0.010

6 การประเมินผล (Evaluation) เป็นขั้นตอนของการนำผลลัพธ์ที่ได้จากกระบวนการพยากรณ์มาทำการทดสอบความถูกต้องแม่นยำ และความน่าเชื่อถือของโมเดล โดยงานวิจัยครั้งนี้ จะนำผลลัพธ์แท้จริงของข้อมูลที่ใช้ทดสอบ (Testing Data) มาเปรียบเทียบกับข้อมูลผลลัพธ์ที่ได้จากโมเดลการพยากรณ์ โดยการวัดประสิทธิภาพของข้อมูลจะอาศัย Confusion Matrix เพื่อคำนวณหาความแม่นยำ

ผลการวิจัย

1 ผลการวิเคราะห์ปัจจัยที่เกี่ยวข้องในการพยากรณ์ความถูกต้องในการระบุสถานะของการเตรียมความพร้อมก่อนเรียนในรายวิชาหลักการโปรแกรม สำหรับนักศึกษาชั้นปีที่ 1 สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ที่จบการศึกษามาจากระดับชั้นประกาศนียบัตรวิชาชีพชั้นสูง โดยการลดมิติข้อมูล (Attribute Selection) และการคำนวณค่าน้ำหนักของแอททริบิวต์พบว่ามีจำนวน 7 ปัจจัยที่เกี่ยวข้อง ได้แก่ ผลการทดสอบก่อนเรียน (Pretest) สาขาวิชาที่จบ (Major) ผลการเรียนก่อนจบการศึกษา (Grade) ผลการเรียนในรายวิชาการเขียนโปรแกรมเบื้องต้น (Program) ผลการเรียนวิชาคณิตศาสตร์ (Math) ผลการเรียนวิชาเทคโนโลยีสารสนเทศเพื่อการอาชีพ (IT) และเพศ (Sex) ตามลำดับ

2 ผลการสังเคราะห์และวัดประสิทธิภาพของโมเดล โดยการทดสอบจะใช้วิธี Cross Validation แบบวนรอบทั้งหมด 10 ครั้ง (10-fold Cross Validation) และแสดงค่าความถูกต้องแม่นยำ ดังตารางที่ 4

ตารางที่ 4 ผลการเปรียบเทียบค่าประสิทธิภาพของอัลกอริทึม

Algorithm	Accuracy	Precision	Recall	F-measure
Decision Tree	91.15%	85.71%	80.00%	82.76%
Naïve Bayes	90.27%	78.79%	86.67%	82.54%
Neural Network	93.81%	96.00%	80.00%	87.27%
Support Vector Machine (SVM)	95.58%	96.00%	85.71%	90.57%

จะเห็นได้ว่าโมเดลที่สร้างด้วยเทคนิค Support Vector Machine (SVM) มีค่าความถูกต้องสูงถึงร้อยละ 95.58 รองลงมาคือเทคนิค Neural Network มีค่าความถูกต้องร้อยละ 93.81 เทคนิค Decision Tree มีค่าความถูกต้องร้อยละ 91.15 และเทคนิค Naïve Bayes มีค่าความถูกต้องน้อยที่สุดคือร้อยละ 90.27

สรุปและอภิปรายผลการวิจัย

จากการรวบรวมข้อมูลการพยากรณ์ความถูกต้องในการระบุสถานะของการเตรียมความพร้อมก่อนเรียนในรายวิชาหลักการโปรแกรม สำหรับนักศึกษาชั้นปีที่ 1 สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัย

ธนบุรี ที่ได้จากข้อมูลในใบระเบียนแสดงผลการเรียนตามหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) และข้อมูลผลการทดสอบก่อนเรียนในรายวิชาหลักการโปรแกรม ที่ทางสาขาวิชาเทคโนโลยีสารสนเทศได้ทำการทดสอบนักศึกษาทุกคนก่อนเริ่มเรียนในรายวิชาดังกล่าว ตั้งแต่ปีการศึกษา 2559 ถึงปีการศึกษา 2563 จำนวน 376 รายการ พบสถานะของผู้ที่จะต้องเตรียมความพร้อม จำนวน 284 รายการ และสถานะของผู้ที่ผ่านการเตรียมความพร้อม จำนวน 92 รายการ และเมื่อทำการวิเคราะห์ค่าน้ำหนักของแอททริบิวต์พบว่าปัจจัยที่เกี่ยวข้องในการพยากรณ์สถานะการเตรียมความพร้อมก่อนเรียน 5 อันดับ ได้แก่ ผลการทดสอบก่อนเรียน สาขาวิชาที่จบ ผลการเรียนก่อนจบการศึกษา ผลการเรียนในรายวิชาการเขียนโปรแกรมเบื้องต้น และผลการเรียนในรายวิชาวิชาคณิตศาสตร์ ตามลำดับ จากผลการวิเคราะห์ผลการทดลองสรุปได้ว่าเทคนิคที่พยากรณ์ผลได้ถูกต้องแม่นยำมากที่สุดคือ เทคนิค Support Vector Machine (SVM) รองลงมาคือ เทคนิค Neural Network เทคนิค Decision Tree และเทคนิคที่ถูกต้องแม่นยำน้อยที่สุดคือ Naïve Bayes

จากผลการวิเคราะห์ข้อมูลข้างต้นทางผู้วิจัยจึงมีแนวคิดในการเลือกเทคนิค Support Vector Machine (SVM) และนำข้อมูลมาเป็นแนวทางในการพัฒนาระบบสนับสนุนการตัดสินใจในการระบุสถานะว่านักศึกษาคนใดจะต้องเข้ารับการอบรมเพื่อเตรียมความพร้อมเบื้องต้นในรายวิชาหลักการโปรแกรมให้กับนักศึกษาชั้นปีที่ 1 สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธนบุรี ซึ่งสามารถใช้ได้กับนักศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยธนบุรี ที่มีการเปิดสอนในรายวิชาการเขียนโปรแกรมคอมพิวเตอร์เบื้องต้นได้ เนื่องจากวิชาที่เกี่ยวข้องกับหลักการโปรแกรมคอมพิวเตอร์เป็นวิชาสำคัญที่มีความจำเป็นต่อนักศึกษาทั้ง 2 สาขาวิชา เพื่อใช้เป็นฐานความรู้ที่สามารถประยุกต์ได้กับรายวิชาที่สูงขึ้นไป

เอกสารอ้างอิง

[1] พนภาค ผิวเกลี้ยง และ มาเรียม นิลพันธุ์. (2014). การพัฒนาชุดทักษะการเขียนโปรแกรมสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5. *วารสารศิลปากรศึกษาศาสตร์วิจัย ปีที่ 6 ฉบับที่ 1 (มกราคม – มิถุนายน).* (pp. 80-90).

[2] A. M. Shahiri, W. Husain and N. A. Rashid. (2015). A Review on Predicting Student's Performance using Data Mining Techniques. *The Third Information Systems International Conference*. (pp. 414-422).

[3] B.C. Wilson and S. Shrock. (2001). Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors. *Thirty-second SIGCSE technical symposium on Computer Science Education* (pp. 184-188).

[4] B.C. Wilson. (2002). A Study of Factors Promoting Success in Computer Science including Gender Difference. *Computer Science Education*, *vol. 12, no. 1-2* (pp. 141-164). New York, USA.

[5] J. Allert. (2004). Learning style and factors contributing to success in an introductory computer science course. *IEEE International Conference on Advanced Learning Technologies.* (pp. 385-389).

[6] S. Bergin and R. Reilly. (2005). Programming: Factors that Influence Success. *ACM SIGCSE Bulletin*, *vol. 37*, (pp. 411-415).

[7] S. Raschka. (2015). *Single-Layer Neural Networks and Gradient Descent*. Retrieved 24 September 2020 from http://sebastianraschka. com/Articles/2015_singlelayer_neurons.html

[8] S. Sayad. *Support Vector Machine – Classification (SVM).* Retrieved 24 September 2020 from http://www.saedsayad.com/support_vector_machine.htm

[9] Y. ERDOGAN, E. AYDIN, and T. KABACA. (2007). Identifying Predictors of Programming Achievement. *6th WSEAS International Conference on Education and Educational Technology Conference* (pp. 100-115). Venice, Italy.