รายงานวิจัยฉบับสมบูรณ์

การวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี หลักสูตรสารสนเทศศึกษา ด้วยเทคนิคการทำเหมืองข้อมูลทางการศึกษา

An Analysis of Academic Achievement from Undergraduate Students for Information Studies Major using Educational Data Mining Techniques

อาจารย์ ดร.ณิชพน โชติจันทรกุล

โครงการวิจัยประเภทงบประมาณเงินรายได้มหาวิทยาลัย เงินรายได้ส่วนงาน มหาวิทยาลัยบูรพา

ประจำปีงบประมาณ พ.ศ. 2566

คณะมนุษยศาสตร์และสังคมศาสตร์

มหาวิทยาลัยบูรพา

2568

สัญญาเลขที่ HUSO 07/2566

รายงานวิจัยฉบับสมบูรณ์
การวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี หลักสูตรสารสนเทศศึกษา ด้วยเทคนิคการทำเหมืองข้อมูลทางการศึกษา
An Analysis of Academic Achievement from Undergraduate Students for Information Studies Major using Educational Data Mining Techniques

อาจารย์ ดร.ณิชพน โชติจันทรกุล
คณะมนุษยศาสตร์และสังคมศาสตร์

บทคัดย่อ

การทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตก่อนสิ้นสุดการศึกษา โดยเฉพาะในช่วงสามปีแรกของการศึกษา จะประโยชน์ในการออกแบบแนวทางสนับสนุนการศึกษาสำหรับช่วงเวลาที่เหลือให้กับนิสิตได้อย่างเหมาะสม ส่งผลให้นิสิตมีความสำเร็จด้านการศึกษาที่สูงขึ้น กระบวนการสกัดความรู้ลักษณะนี้โดยใช้ข้อมูลของนิสิตจัดเป็นศาสตร์แขนงหนึ่งของการทำเหมืองข้อมูลที่เรียกว่า การทำเหมืองข้อมูลทางการศึกษา (Educational Data Mining: EDM) งานวิจัยนี้เป็นการวิเคราะห์ปัจจัยที่เกี่ยวข้องในการส่งผลต่อผลสัมฤทธิ์ทางการศึกษาของนิสิตที่สำเร็จการศึกษาสาขาวิชาสารสนเทศศึกษา ผลลัพธ์ที่ได้นอกจากเพื่อพัฒนาศักยภาพให้กับนิสิตที่มีผลการเรียนในระดับต่าง ๆ แล้ว ยังเป็นข้อมูลสำหรับการปรับปรุงหลักสูตรใหม่ด้วยการค้นหารูปแบบที่เกี่ยวข้องกับรายวิชาที่เรียนอีกด้วย ชุดข้อมูลจะถูกนำไปใช้กับโมเดลการจำแนกประเภทเพื่อจัดกลุ่มนิสิตออกเป็นสี่กลุ่มเป้าหมาย ได้แก่ มีผลการเรียนดีเด่น ดีมาก ดี และพอใช้ งานวิจัยนี้ได้เลือกวิธีการทำเหมืองข้อมูลห้าวิธี ประกอบไปด้วย วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (K-Nearest Neighbor: KNN) วิธีแบบเบส์อย่างง่าย (Naïve Bayes: NB) วิธีต้นไม้ตัดสินใจ (Decision Tree: DT) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และวิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) ชุดข้อมูลถูกแบ่งออกเป็นสามกลุ่ม ได้แก่ ชุดข้อมูลด้านประชากร ชุดข้อมูลด้านเกรดของแต่ละรายวิชา และชุดข้อมูลด้านเกรดเฉลี่ยของแต่ละภาคการศึกษา งานวิจัยนี้ยังเกี่ยวข้องกับชุดข้อมูลที่ไม่สมดุลกันของกลุ่มตัวอย่างจำนวน 275 ชุด จึงได้มีการปรับให้สมดุลกันโดยใช้เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (Synthetic Minority Over-sampling Technique: SMOTE) และเปรียบเทียบผลลัพธ์กับชุดข้อมูลดั้งเดิม ส่งผลให้กลุ่มตัวอย่างมีจำนวนเพิ่มขึ้นเป็น 400 ชุด ผลการวิจัยแสดงให้เห็นว่าเกรดเฉลี่ยของแต่ละภาคการศึกษาในช่วงสามปีแรกมีค่าความแม่นยำสูงถึงร้อยละ 90.5 โดยใช้ชุดข้อมูลที่ปรับสมดุลแล้วประยุกต์เข้ากับโมเดลแบบเบส์อย่างง่าย ส่วนชุดข้อมูลด้านเกรดของแต่ละรายวิชาก็ให้ผลลัพธ์ที่ดีเช่นกัน ในขณะที่ชุดข้อมูลด้านประชากรมีอิทธิพลต่อการทำนายน้อยที่สุด งานวิจัยนี้เป็นการค้นหาความรู้ที่เป็นประโยชน์ในการสนับสนุนกลยุทธ์เพื่อการศึกษาให้นิสิตมีผลการเรียนที่ดีขึ้น อีกทั้งยังเป็นข้อมูลสำหรับปรับปรุงหลักสูตรสารสนเทศศึกษาในอนาคตได้อีกด้วย

Abstract

Predicting students' academic performance prior to completing their studies, particularly within the initial three years, proves advantageous for designing appropriate educational support strategies throughout the remainder of their academic journey. This contributes to higher academic success for students. This type of knowledge extraction, using student data, is a branch of data mining called Educational Data Mining (EDM). This research analyzes factors affecting the academic achievement of students graduating in the Information Studies Program. The results not only aim to enhance the potential of students with various academic performance levels but also provide insights for curriculum improvement by identifying patterns related to the studied courses. The dataset was applied to classification models to categorize students into four target groups: Excellent, Very good, Good, and Fair performance. Five data mining methods were selected for this research, including K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The dataset was divided into three groups: demographic dataset, grades of individual courses, and semester GPA dataset. This research also dealt with an imbalanced dataset of 275 samples, which was balanced using the Synthetic Minority Over-sampling Technique (SMOTE). The results were then compared with the original dataset, increasing the sample size to 400 samples. The findings indicate that the GPA of each semester during the first three years provides a high accuracy rate of 90.5% when using the balanced dataset applied to the Naïve Bayes model. The dataset containing individual course grades also produced good results, while the demographic dataset had the least influence on prediction. This research provides valuable knowledge to support educational strategies for improving student academic performance and serves as a basis for future improvements to the Information Studies curriculum.

กิตติกรรมประกาศ

โครงการวิจัยประเภทงบประมาณเงินรายได้มหาวิทยาลัย เงินรายได้ส่วนงาน มหาวิทยาลัยบูรพา ประจำปีงบประมาณ พ.ศ. 2666 เลขที่สัญญา Huso 07/2566

โครงการวิจัยนี้สำเร็จลุล่วงได้ด้วยการสนับสนุนจากภาควิชาสารสนเทศศึกษา และคณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา ที่อำนวยความสะดวกในด้านการประสานงานในทุกขั้นตอนของกระบวนการทำวิจัย รวมถึงกองทะเบียนและประมวลผลการศึกษา มหาวิทยาลัยบูรพา ที่ได้ให้ความอนุเคราะห์ข้อมูลของนิสิตภาควิชาสารสนเทศศึกษารหัส 59 ถึง 62 ซึ่งเป็นข้อมูลนำเข้าสำคัญในการทำวิจัย

ขอขอบพระคุณ คณะกรรมการผู้อนุมัติให้ทุนโครงการวิจัย และผู้ทรงคุณวุฒิที่พิจารณาข้อเสนอโครงการวิจัย ที่ได้กรุณาให้คำแนะนำในการแก้ไขปรับปรุงให้งานวิจัยมีคุณภาพมากยิ่งขึ้น โครงการวิจัยนี้ไม่อาจจะสำเร็จได้หากไม่ได้รับความกรุณาจากทุกท่าน

ท้ายนี้ขอขอบพระคุณ บิดา มารดา ญาติสนิท และมิตรสหายทุกคน ที่เป็นกำลังใจ คอยดูแลห่วงใยและให้การสนับสนุนแก่ผู้วิจัยในทุก ๆ ด้านเสมอมา จนผู้วิจัยสามารถทำโครงงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

อาจารย์ ดร.ณิชพน โชติจันทรกุล

1 กันยายน 2568

สารบัญ

หน้า

สารบัญ (ต่อ)

สารบัญภาพ

หน้า

สารบัญตาราง

หน้า

สารบัญตาราง (ต่อ)

หน้า

สารบัญตาราง (ต่อ)

หน้า

บทที่ 1
บทนำ

ความสำคัญและที่มาของปัญหา

เนื่องจากการเปลี่ยนแปลงของกระบวนการทางวิชาการจากเอกสารไปอยู่ในรูปแบบของดิจิทัลเกือบทั้งหมด สถาบันการศึกษาจึงได้ผลิตข้อมูลจำนวนมหาศาลที่เกี่ยวข้องกับนักเรียนนักศึกษาในแบบฟอร์มอิเล็กทรอนิกส์ การแปลงชุดข้อมูลจำนวนมหาศาลนี้ให้ไปเป็นความรู้จึงมีความสำคัญในการช่วยให้ ครู อาจารย์ ผู้บริหาร หรือผู้กำหนดนโยบาย นำความรู้ที่สืบค้นได้มาวิเคราะห์เพื่อประกอบการตัดสินใจในเรื่องต่าง ๆ นอกจากนี้ยังช่วยในด้านการปรับปรุงคุณภาพของกระบวนการการศึกษาให้ดีขึ้นอีกด้วย ความรู้ที่ถูกสกัดออกมานี้เป็นผลมาจากกระบวนการทำเหมืองข้อมูล ซึ่งประกอบไปด้วยขั้นตอนวิธีและเทคนิคมากมาย การทำเหมืองข้อมูลยังได้นำไปประยุกต์ใช้กับงานในด้านต่าง ๆ เช่น ด้านการแพทย์ ด้านการศึกษา ด้านการตลาด ด้านอุตสาหกรรม ด้านการเงิน ด้านความปลอดภัยบนระบบคอมพิวเตอร์ เป็นต้น สำหรับการประยุกต์ใช้วิธีการขุดข้อมูลจากชุดข้อมูลทางการศึกษาเรียกว่า การทำเหมืองข้อมูลทางการศึกษา Baker and Yacef (2009) เสนอการทำเหมืองข้อมูลทางการศึกษามา 5 ประเภท ได้แก่ การทำนาย การจัดกลุ่ม การทำเหมืองข้อมูลเชิงสัมพันธ์ การค้นพบ และการกลั่นกรองข้อมูลเพื่อการตัดสินใจของมนุษย์ งานในปัจจุบันนิยมผสมผสานสามแนวทางเข้าด้วยกัน ได้แก่ การทำนาย การจัดกลุ่ม และการกลั่นกรองข้อมูลเพื่อการตัดสินใจของมนุษย์

ความสำเร็จทางการศึกษาของนิสิตในสถาบันการศึกษาเป็นตัวชี้วัดสำคัญในการจัดการเรียนการสอนให้มีคุณภาพ การจัดประเภทของนิสิตในระยะแรกที่เริ่มต้นศึกษาในระดับมหาวิทยาลัยเป็นกลยุทธ์ที่มีประสิทธิภาพและมีประโยชน์อย่างมากในการลดโอกาสการพ้นสภาพของนิสิต อีกทั้งยังส่งเสริมผลสัมฤทธิ์ทางการศึกษาให้ดียิ่งขึ้น รวมถึงเพื่อการจัดสรรค์ทรัพยากรในสถาบันอุดมศึกษาให้เอื้อประโยชน์ต่อการเรียนรู้มากที่สุด

ในอดีต การตรวจหานิสิตที่อยู่ในกลุ่มเสี่ยงตั้งแต่เนิ่น ๆ ควบคู่ไปกับการกำหนดมาตรการป้องกันการพ้นสภาพของนิสิตทำได้ยากมาก จนกระทั่งความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ทางด้านปัญญาประดิษฐ์ (Artificial Intelligence: AI) โดยเฉพาะการทำเหมืองข้อมูล (Data Mining: DM) และการเรียนรู้ของเครื่อง (Machine Learning: ML) ส่งผลให้มีการใช้เทคนิคการทำนาย (Prediction) การจัดกลุ่ม (Clustering) และการค้นหาความสัมพันธ์ (Association) อย่างกว้างขวาง มีงานวิจัยมากมายที่นำเทคนิคเหล่านี้มาทำการวิเคราะห์หาความรู้ จัดประเภท และความสัมพันธ์ของข้อมูล เพื่อนำมาช่วยสนับสนุนการตัดสินใจในการวางแผนการศึกษาได้อย่างมีประสิทธิภาพสูงสุด

เทคนิคการทำเหมืองข้อมูลถูกนำไปใช้กับงานด้านต่าง ๆ มากมาย เช่น ธุรกิจด้านโทรคมนาคมได้นำวิธีการทำเหมืองข้อมูลเพื่อปรับปรุงการบริการลูกค้าโดยเน้นการค้นหารูปแบบพฤติกรรมของลูกค้า งานด้านการธนาคารหรือประกันภัยสามารถใช้การทำเหมืองข้อมูลเพื่อแก้ปัญหาการจัดการความเสี่ยงหรือการออกจาก

บัญชีของลูกค้าได้ การบริการด้านการศึกษาสามารถใช้อัลกอริทึมการทำเหมืองข้อมูลเพื่อค้นหาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของผู้เรียนและออกแบบนโยบายที่สนับสนุนผู้เรียนได้ งานด้านการผลิตสามารถใช้การทำเหมืองข้อมูลเพื่อทำนายปรากฏการณ์ต่าง ๆ ที่อาจเกิดขึ้นในขั้นตอนของการผลิตและหาวิธีแก้ไขไว้แต่เนิ่น ๆ ธุรกิจค้าปลีกสามารถใช้ข้อมูลลูกค้าขนาดใหญ่สำหรับวิเคราะห์พฤติกรรมการซื้อของลูกค้าทำให้สามารถทำนายยอดขายและเสนอบริการส่งเสริมการขาย (Sale Promotion) ที่สร้างความพึงพอใจให้กับลูกค้าได้ ด้านการแพทย์และสาธารณะสุขก็ยังมีการใช้เทคนิคการทำเหมืองข้อมูลในการแก้ปัญหาต่าง ๆ ไม่ว่าจะเป็นการทำนายโรคจากอาการของผู้ป่วย หรือการวิเคราะห์ปัจจัยเสี่ยงที่จะเกิดโรคร้ายแรงได้อย่างแม่นยำ จะเห็นได้ว่า การนำเทคนิคการทำเหมืองข้อมูลมาใช้ประโยชน์มีขอบเขตที่กว้างขว้าง และมีส่วนในการตัดสินใจวางแผนนโยบายทั้งในระดับบริหารและระดับปฏิบัติการได้อย่างมีประสิทธิภาพ

การทำเหมืองข้อมูลทางการศึกษาเป็นหัวข้อที่เกิดขึ้นใหม่สืบเนื่องมาจากการพัฒนาระบบการจัดการฐานข้อมูลทางการศึกษา (Educational Database Management System) ซึ่งเกี่ยวข้องกับการพัฒนาวิธีการสำรวจข้อมูลขนาดใหญ่ที่ไม่ซ้ำกับใครและมีปริมาณของข้อมูลที่มากขึ้นเรื่อย ๆ ซึ่งข้อมูลเหล่านั้นได้มาจากสถาบันการศึกษาในระดับต่าง ๆ โดยนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้กับชุดข้อมูลของผู้เรียนเพื่อทำความเข้าใจผู้เรียนให้ดีขึ้น การวิเคราะห์ข้อมูลสามารถนำไปใช้ในลักษณะต่าง ๆ ทั้งเรียนรู้จากสภาพแวดล้อมแบบโต้ตอบของผู้เรียน การเรียนรู้ร่วมกันโดยใช้อุปกรณ์คอมพิวเตอร์หรือสื่ออิเล็กทรอนิกส์ ไปจนถึงการบริหารจัดการโรงเรียนและมหาวิทยาลัย แม้ว่าโดยทั่วไปการวัดผลสัมฤทธิ์ทางการศึกษาจะนิยมใช้เกรดและเกรดเฉลี่ย แต่ข้อมูลที่เกี่ยวข้องกับสถานะภาพทางสังคมก็อาจมีผลเช่นกัน

ดังนั้นจึงสรุปได้ว่า การทำเหมืองข้อมูลทางการศึกษา (Educational Data Mining: EDM) คือการประยุกต์ใช้เทคนิคและวิธีการทำเหมืองข้อมูลกับข้อมูลทางการศึกษาของผู้เรียน ทั้งข้อมูลด้านวิชาการ (Academic) และข้อมูลด้านประชากร (Demographic) เพื่อวัตถุประสงค์ในการบริหารจัดการหลักสูตรทั้งในระดับภาควิชา คณะ และมหาวิทยาลัย ผ่านกระบวนการของการวิจัยอย่างเป็นรูปธรรม

ในปัจจุบัน หลักสูตรสารสนเทศศึกษา ภาควิชาสารสนเทศศึกษา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา ยังไม่มีระบบการกำกับติดตามหรือการเทียบเคียงสมรรถนะของผู้เรียนที่เป็นไปตามมาตรฐานของเครือข่ายการประกันคุณภาพมหาวิทยาลัยอาเซียน (AUN-QA version 4) ทำให้การรวบรวมข้อมูลเพื่อวิเคราะห์ผลการดำเนินงานหลักสูตรในแง่มุมต่าง ๆ โดยเฉพาะใน AUN-QA criterion 8 Output and Outcome ทำได้ยาก และการวิเคราะห์หาความสัมพันธ์ระหว่างผลการเรียนในแต่ละภาคการศึกษา ข้อมูลด้านประชากรของนิสิต และผลสัมฤทธิ์ทางการศึกษาของภาควิชาสารสนเทศศึกษา ก็เป็นเรื่องใหม่ที่ยังไม่มีการศึกษามาก่อน ดังนั้น ผู้วิจัยจึงสนใจค้นหาคำตอบที่เกี่ยวกับประเด็นด้านความสัมพันธ์เหล่านี้ เพื่อเป็นข้อมูลสำหรับการออกแบบหลักสูตรสารสนเทศศึกษาในอนาคตต่อไป

วัตถุประสงค์ของการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูลเพื่อศึกษาผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี ภาควิชาสารสนเทศศึกษา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา ที่ซึ่งมีเป้าหมายสำคัญคือ คาดการณ์ผลการเรียนของนิสิตเมื่อสิ้นสุดหลักสูตรในช่วง 3 ปีแรกของการศึกษา โดยจะเน้นไปที่ 2 ปีแรกเป็นสำคัญ นอกจากนี้ยังศึกษาเกี่ยวกับคุณลักษณะหรือแอตทริบิวต์ (Attribute) ที่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการศึกษาในภาพรวม ผลการศึกษานี้จะช่วยให้คณาจารย์สามารถบริหารจัดการการเรียนการสอนโดยให้คำแนะนำหรือให้ความช่วยเหลือแก่นิสิตที่คาดว่าจะมีผลการเรียนต่ำ เพื่อให้นิสิตมีโอกาสพัฒนาตนเองและได้รับผลการเรียนที่ดีขึ้นเมื่อสิ้นสุดหลักสูตร 4 ปี วัตถุประสงค์ที่ได้จากงานวิจัยมีดังนี้

1. ศึกษาและพัฒนาโมเดลสำหรับทำนายผลการสัมฤทธิ์ทางการเรียนของนิสิตเมื่อสำเร็จการศึกษาแล้ว โดยใช้ข้อมูลก่อนสำเร็จการศึกษา
2. ศึกษาความสัมพันธ์ของกลุ่มรายวิชาต่าง ๆ และคุณลักษณะด้านประชากร ว่ามีผลต่อผลสัมฤทธิ์ทางการศึกษาเมื่อสิ้นสุดการศึกษาแล้วอย่างไร

ขอบเขตการวิจัย

ขอบเขตงานวิจัยจะประกอบไปด้วยชุดข้อมูลด้านประชากร เกรดในแต่ละรายวิชา และเกรดเฉลี่ยของแต่ละภาคการศึกษา ของนิสิตภาควิชาสารสนเทศศึกษาที่สำเร็จการศึกษาแล้ว โดยนับตั้งแต่นิสิตรหัส 59 ถึง 62 ที่สำเร็จการศึกษาในปี พ.ศ. 2562 ถึง 2565 จำนวน 275 คน ซึ่งทั้งหมดเป็นนิสิตที่สำเร็จการศึกษาในหลักสูตรศิลปศาสตรบัณฑิต สาขาวิชาสารสนเทศศึกษา (หลักสูตรปรับปรุง ปี พ.ศ. 2559) ในส่วนของโมเดลการทำเหมืองข้อมูลจะใช้อัลกอริทึมการจำแนกประเภท 5 วิธี ได้แก่ การค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (K-Nearest Neighbor: KNN) วิธีแบบเบส์อย่างง่าย (Naïve Bayes: NB) ต้นไม้ตัดสินใจ (Decision Tree: DT) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และ โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) ส่วนการจัดลำดับความสำคัญของแอตทริบิวต์จะใช้เทคนิคการคัดเลือกคุณลักษณะด้วยค่าเกนความรู้ (Information Gain: IG) ผลลัพธ์ที่ได้สามารถนำไปใช้ปรับปรุงหลักสูตรในอนาคต (หลักสูตรปรับปรุง ปี พ.ศ. 2569) เพื่อออกแบบการเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น

ประโยชน์ที่คาดว่าจะได้รับ

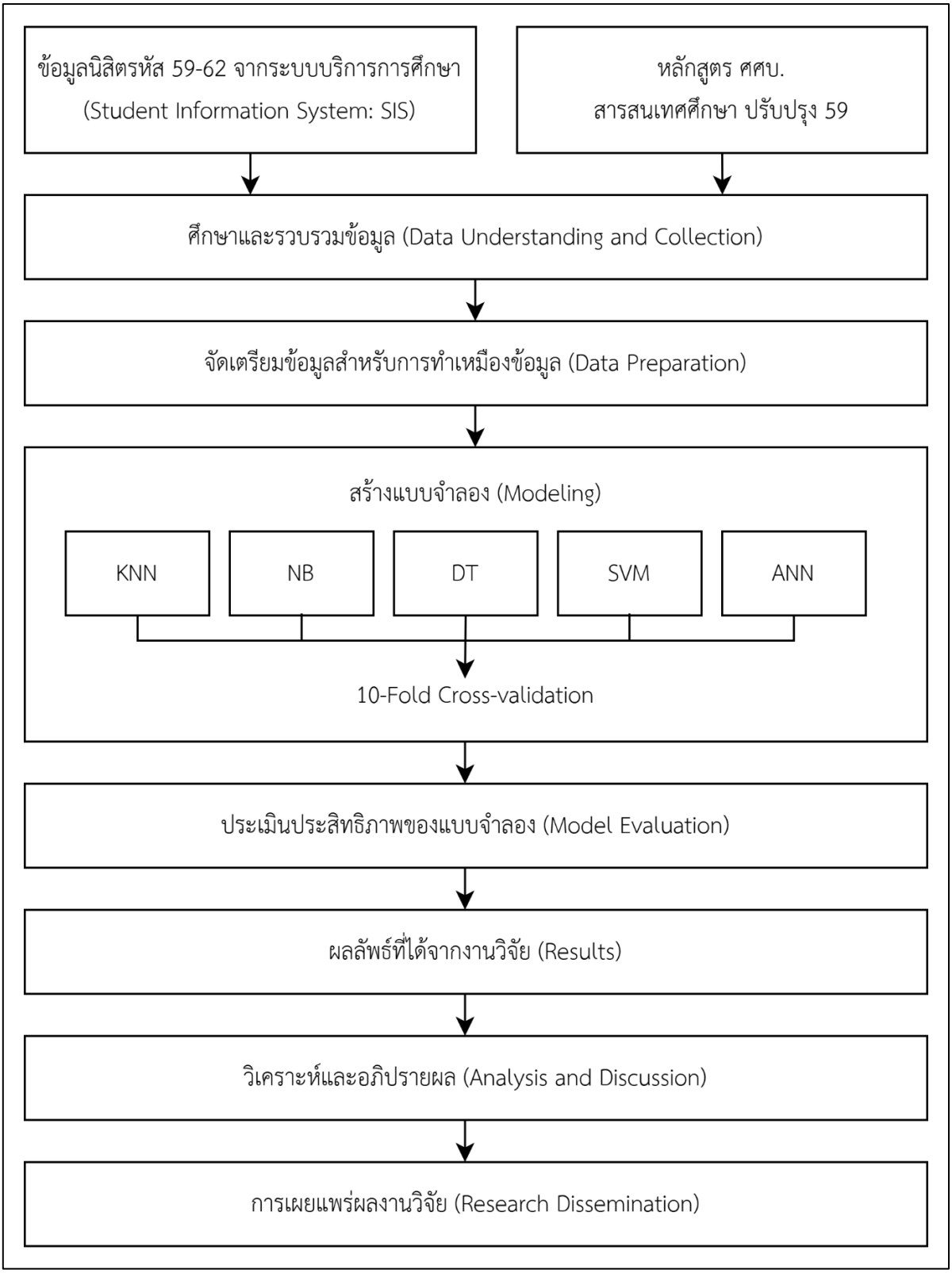
ประโยชน์ที่คาดว่าจะได้รับสามารถระบุเป็นข้อ ๆ ได้ดังต่อไปนี้

1. สามารถใช้เป็นแนวทางในการปรับปรุงหลักสูตรของภาควิชาสารสนเทศศึกษา ทั้งในด้านการกำหนดนโยบาย วางแผนงาน และการสนับสนุน เพื่อส่งเสริมให้นิสิตสำเร็จการศึกษาด้วยเกรดเฉลี่ยสะสมสูงที่สุด
2. ช่วยให้เข้าใจปรากฏการณ์ที่เกิดขึ้นในหลักสูตรสารสนเทศศึกษา เพื่อปรับปรุงผลสัมฤทธิ์ทางการศึกษาของนิสิตให้ดีขึ้นและมีอัตราการพ้นสภาพน้อยลง
3. เพื่อปรับปรุงรายวิชาของหลักสูตรทั้งรายวิชาในกลุ่มสารสนเทศศาสตร์และกลุ่มเทคโนโลยีสารสนเทศให้มีความเหมาะสมมากยิ่งขึ้น
4. เพื่อนำเสนอข้อมูลผลการวิเคราะห์หลักสูตร ซึ่งสามารถใช้สนับสนุนงานด้านประกันคุณภาพการศึกษาของหลักสูตรสารสนเทศศึกษาได้
5. เพื่อเผยแพร่ในวารสารวิชาการระดับชาติหรือระดับนานาชาติ
6. เพื่อเป็นแนวทางการวิเคราะห์ข้อมูลทางการศึกษาสำหรับภาควิชาอื่นที่สนใจนำไปใช้พัฒนางานวิจัยเพื่อพัฒนาหลักสูตรของตนต่อไป

กรอบแนวคิดงานวิจัย

งานวิจัยนี้จะทำการวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี ที่สำเร็จการศึกษาแล้วของหลักสูตรศิลปศาสตร์บัณฑิต ภาควิชาสารสนเทศศึกษา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา โดยรวบรวมข้อมูลของนิสิตที่สำเร็จการศึกษาแล้วจำนวน 275 คน ผลลัพธ์ที่ได้สามารถนำไปใช้ปรับปรุงหลักสูตรในอนาคต (หลักสูตรปรับปรุง ปี พ.ศ. 2569) และเพื่อออกแบบการเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น ซึ่งประกอบไปด้วยประเด็นต่าง ๆ ดังนี้

1. งานวิจัยนี้จะทำการสร้างโมเดลสำหรับการจำแนกประเภทหลายตัวเพื่อทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตตอนเรียนจบมหาวิทยาลัยตั้งแต่ช่วงครึ่งแรกของการศึกษา การวิเคราะห์จะพิจารณาจากข้อมูลก่อนเข้าเรียนและผลการเรียนใน 3 ปีแรก โดยพิจารณาคุณลักษณะทางประชากรและรายวิชาร่วมกัน
2. งานวิจัยนี้จะช่วยให้ทราบว่า คุณลักษณะหรือแอตทริบิวต์ใดที่มีผลต่อการเรียนในภาพรวม เพื่อให้สามารถช่วยเหลือนิสิตที่มีความเสี่ยงจะได้เกรดเฉลี่ยสะสมต่ำ หรือกระตุ้นให้นิสิตมีความพยายามที่จะทำเกรดในรายวิชาที่เหลือให้ดีขึ้น การค้นหาความรู้ที่ซ่อนอยู่นี้ต้องใช้อัลกอริทึมการทำเหมืองข้อมูลหลายอัลกอริทึมเพื่อวิเคราะห์ว่า อัลกอริทึมใดจะเหมาะสมที่สุด ดังนั้นจึงต้องมีการเปรียบเทียบผลลัพธ์ที่ได้จากหลากหลายอัลกอริทึมของการจำแนกประเภท (Classification)
3. งานวิจัยนี้ จะศึกษากลุ่มรายวิชาด้านสารสนเทศศาสตร์ (Information Science: IS) และด้านเทคโนโลยีสารสนเทศ (Information Technology: IT) ว่ากลุ่มทั้งสองนี้ส่งผลต่อผลสัมฤทธิ์ทางการเรียนอย่างไร



ภาพ 1-1

กรอบแนวคิดงานวิจัย

สำหรับกรอบแนวคิดงานวิจัยได้นำเสนอด้วยภาพ 1-1 โดยเริ่มจากนำเข้าข้อมูลที่มีอยู่แล้วในระบบบริการการศึกษาของมหาวิทยาลัยบูรพา และข้อมูลของหลักสูตรสารสนเทศศึกษาฉบับปรับปรุงปี 59 จากนั้นทำการศึกษาและรวบรวมข้อมูลก่อนเข้ากระบวนการทำเหมืองข้อมูล ขั้นตอนต่อไปจะเป็นการเตรียมความพร้อมของชุดข้อมูล เช่น จัดระเบียบชุดข้อมูลใหม่ แก้ไขข้อมูลที่ผิดพลาด หรือ แปลงข้อมูลให้อยู่ในรูปแบบที่ใช้สำหรับการสร้างโมเดลการทำนาย ต่อมาเป็นการสร้างโมเดลหรือแบบจำลองที่ใช้ในการทำนายด้วยอัลกอริทึม 5 อัลกอริทึม ดังที่ได้กล่าวไว้แล้วข้างต้น เมื่อได้โมเดลมาแล้วจึงทำการประเมินประสิทธิของโมเดลด้วยมาตรวัดที่เป็นมาตรฐานโดยอาศัยเมทริกซ์ความสับสน (Confusion Matrix) โดยจะพิจารณาจากค่าความแม่นยำ (Accuracy) ของโมเดลเป็นหลัก ผลลัพธ์ที่ได้จะนำมาวิเคราะห์และอภิปรายสรุปในลำดับต่อไป

บทที่ 2
ทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

แนวคิดและทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล คือ กระบวนการค้นหาความสัมพันธ์ รูปแบบ และแนวโน้มใหม่ ๆ โดยใช้ข้อมูลจำนวนมากที่เก็บไว้ในคลังข้อมูล (Data Warehouse: DW) แล้วใช้วิธีการทางคณิตศาสตร์และสถิติในการวิเคราะห์ข้อมูล การทำเหมืองข้อมูลช่วยให้สามารถสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database: KDD) โดยนำข้อมูลที่มีอยู่มาผ่านกระบวนการ แล้วดึงความรู้หรือสิ่งสำคัญที่ซ่อนอยู่ภายในชุดข้อมูลเหล่านั้นออกมา เพื่อใช้ในการวิเคราะห์หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้นได้อย่างแม่นยำ

การวิเคราะห์ที่ใช้เทคนิคการทำเหมืองข้อมูลสามารถจัดประเภท ได้ดังนี้:

1. วิธีการวิเคราะห์ทางสถิติแบบดั้งเดิม (Classical Statistics Methods) เช่น การวิเคราะห์การถดถอย (Regression Analysis) การวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis) และการวิเคราะห์แบ่งกลุ่ม (Cluster Analysis)
2. ปัญญาประดิษฐ์ เช่น ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithms: GA) การคำนวณด้วยประสาท (Neural Computing) และตรรกศาสตร์คลุมเครือ (Fuzzy Logic)
3. การเรียนรู้ของเครื่อง เช่น โครงข่ายประสาท (Neural Network: NN) การเรียนรู้เชิงสัญลักษณ์ (Symbolic Learning) และวิธีหาคำตอบที่เหมาะสมแบบฝูง (Swarm Optimization: SO) วิธีนี้ประกอบด้วยการผสมผสานระหว่างวิธีการทางสถิติขั้นสูงและการวิเคราะห์พฤติกรรมด้วยปัญญาประดิษฐ์ เทคนิคเหล่านี้สามารถนำมาใช้ประโยชน์ในด้านต่าง ๆ เพื่อวัตถุประสงค์ที่แตกต่างกัน เช่น การสกัดรูปแบบ การทำนายพฤติกรรม หรือการอธิบายแนวโน้มของชุดข้อมูลบางอย่าง

การทำเหมืองข้อมูลได้นำเทคนิคที่ใช้คอมพิวเตอร์ช่วยในการวิเคราะห์เพื่อประมวลผล โดยการสกัดข้อมูล (Extract Data) จากฐานข้อมูลขนาดใหญ่ (Big Data) เพื่อให้ได้สารสนเทศที่มีประโยชน์ และสามารถนำไปใช้ได้อย่างเหมาะสม องค์ความรู้ที่ค้นพบทั้งรูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลเป้าหมาย (Target Data) คือผลลัพธ์ที่ได้จากการแปลงข้อมูลดิบไปเป็นความรู้เชิงปฏิบัติ ซึ่งหน่วยงานสามารถนำความรู้นี้ไปแก้ไขปัญหา วิเคราะห์ผลกระทบในอนาคต และยังช่วยสนับสนุนการตัดสินใจในเรื่องต่าง ๆ เพื่อประโยชน์สูงสุดของหน่วยงาน

กระบวนการที่เป็นมาตรฐานอุตสาหกรรมสำหรับการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining: CRISP-DM) พัฒนาในปี 1996 โดยบริษัท DaimlerChrysler, SPSS และ NCR กระบวนการ CRISP-DM ดังแสดงในภาพ 2-1 เป็นกระบวนการที่มีประโยชน์สำหรับทำให้การทำเหมืองข้อมูลมีความเหมาะสมและเป็นมาตรฐาน โดยที่ CRISP-DM ประกอบไปด้วยวงจรชีวิต 6 ขั้นตอน (Phase) ดังนี้

1. ขั้นตอนการทำความเข้าใจธุรกิจ (Business Understanding) – เน้นที่การทำความเข้าใจวัตถุประสงค์และความจำเป็นของการดำเนินงาน ตระหนักถึงเป้าหมาย ข้อจำกัด และขอบเขตของข้อมูลที่จะนำมาวิเคราะห์ รวมถึงการจัดเตรียมกลยุทธ์เบื้องต้น

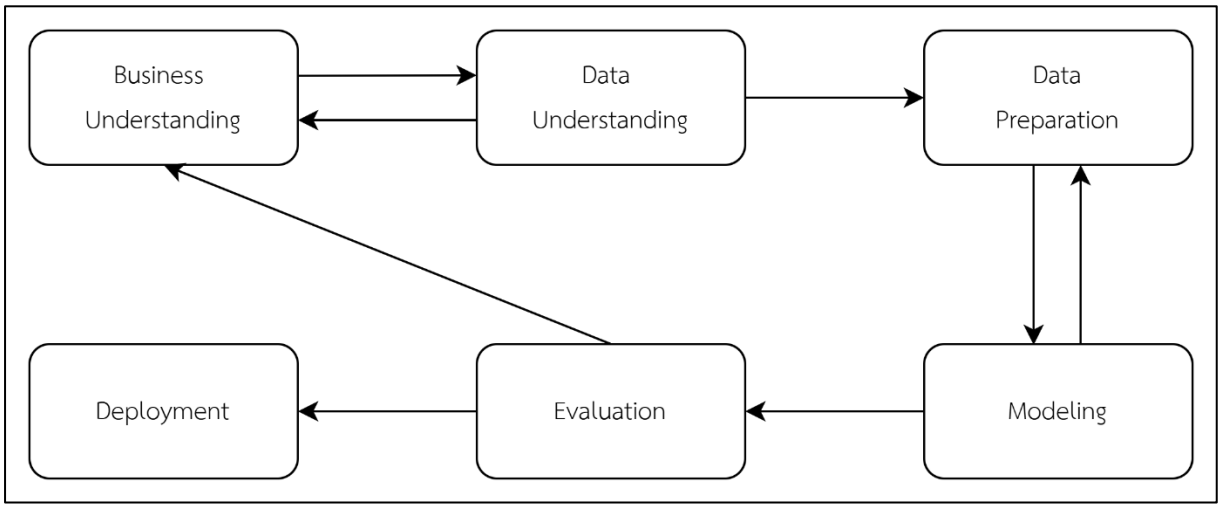
2. ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding) – เป็นขั้นตอนของการเก็บรวบรวมข้อมูลที่เกี่ยวข้อง ประเมินคุณภาพของข้อมูล คัดเลือกข้อมูลที่สนใจและมีความถูกต้อง เพื่อนำมาใช้ในทางปฏิบัติ

3. ขั้นตอนการจัดเตรียมข้อมูล (Data Preparation) – จัดเตรียมข้อมูลจากข้อมูลดิบ (Raw Data) เลือกข้อมูลที่ต้องการวิเคราะห์ แปลงข้อมูลให้เหมาะสม และจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับเครื่องมือสร้างตัวแบบ (Model) ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะคุณภาพของงานที่ได้ขึ้นอยู่กับคุณภาพของข้อมูลที่จัดเตรียมนี้ การเตรียมข้อมูลประกอบด้วยขั้นตอนย่อยหลัก คือ การคัดเลือกข้อมูล (Selection) การกลั่นกรองข้อมูล (Cleansing) และการแปลงรูปแบบของข้อมูล (Transformation)

4. ขั้นตอนการสร้างตัวแบบ (Modeling) – เป็นการสร้างตัวแบบจากข้อมูลที่ได้มาจากขั้นตอนก่อนหน้า ขั้นตอนนี้จะเลือกใช้เทคนิคการสร้างตัวแบบที่เหมาะสม จัดทำตัวแบบให้เป็นมาตรฐาน และทดสอบผลลัพธ์ที่หลากหลายสำหรับปัญหาเดียวกัน เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

5. ขั้นตอนการประเมินตัวแบบ (Evaluation) – เป็นการประเมินคุณภาพของตัวแบบที่ได้มา หาตัวแบบที่บรรลุวัตถุประสงค์หรือตรงกับเป้าหมายที่ตั้งไว้ จัดเป็นขั้นตอนปรับปรุงให้ได้ตัวแบบที่ดีที่สุดก่อนนำไปใช้จริง

6. ขั้นตอนการใช้งาน (Deployment) – เป็นการนำตัวแบบที่สร้างขึ้นไปใช้งาน โดยแปลงแนวคิดที่ได้ให้เป็นสารสนเทศเพื่อให้ง่ายต่อการเข้าใจ และสามารถนำไปใช้ประโยชน์ได้ รวมถึงการติดตามประเมินผลเพื่อพัฒนาปรับปรุงให้กระบวนการทำเหมืองข้อมูลสมบูรณ์ยิ่งขึ้นต่อไป

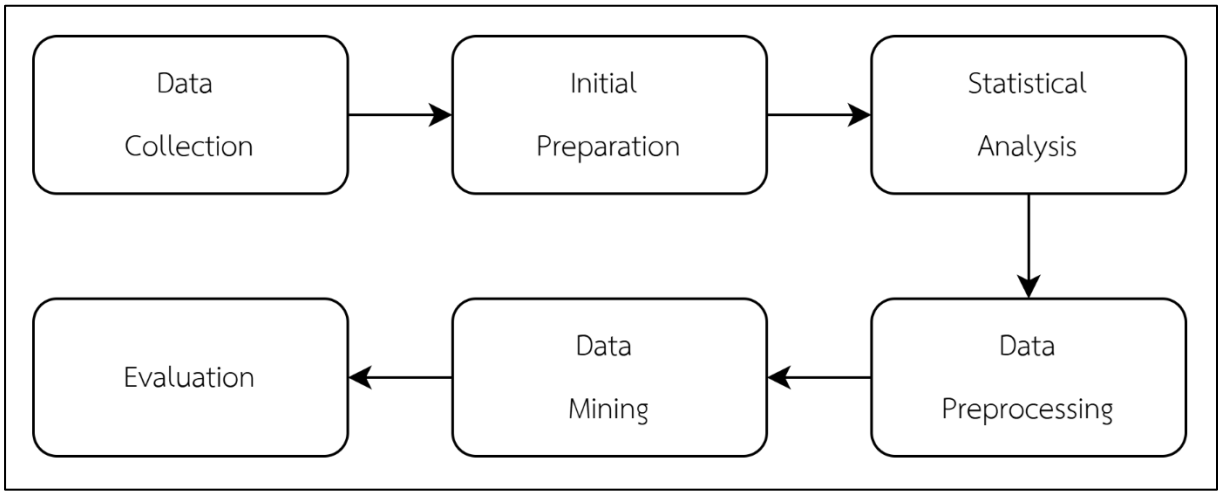


ภาพ 2-1

กระบวนการ CRISP-DM

การทำเหมืองข้อมูลทางการศึกษา (Educational Data Mining: EDM)

การทำเหมืองข้อมูลทางการศึกษาประกอบด้วย 6 ขั้นตอนหลัก (Alyahyan & Dustegor, 2020) ซึ่งแสดงไว้ในภาพ 2-2 ดังนี้



ภาพ 2-2

กระบวนการทำเหมืองข้อมูลทางการศึกษา

1. การรวบรวมข้อมูล (Data Collection)

การรวบรวมข้อมูล เป็นการดึงข้อมูลออกมาจากหลาย ๆ แหล่ง ซึ่งประกอบไปด้วยข้อมูลก่อนเข้ามหาวิทยาลัย ข้อมูลด้านประชากร ข้อมูลที่เกี่ยวข้องกับสภาพแวดล้อมของผู้เรียน รวมถึงข้อมูลทางด้านจิตวิทยา โดยปกติข้อมูลที่ใช้ในการทำเหมืองมักจะได้มาจากระบบสารสนเทศนักศึกษา (Student Information System: SIS) ของมหาวิทยาลัยนั้น ๆ แต่ก็มีข้อมูลของผู้เรียนบางส่วนที่ได้มาจากการทำแบบสำรวจของผู้เรียนเอง ซึ่งข้อมูลทั้งหมดจะถูกรวบรวมนำไปจัดเตรียมให้มีความพร้อมสำหรับการวิเคราะห์ในขั้นต่อไป

2. การเตรียมข้อมูลเบื้องต้น (Data Initial Preparation)

การเตรียมข้อมูลเบื้องต้น เป็นขั้นตอนที่สำคัญมากเนื่องจากรูปแบบดั้งเดิมของข้อมูล (ข้อมูลดิบ) โดยปกติแล้วข้อมูลดิบมักไม่มีความพร้อมสำหรับการวิเคราะห์และการสร้างโมเดล เนื่องจากชุดข้อมูลที่ได้มาส่วนมากมาจากการรวมตารางในระบบสารสนเทศชนิดต่าง ๆ ซึ่งเป็นไปได้ว่าอาจมีข้อมูลขาดหายไป ข้อมูลไม่สอดคล้องกัน ข้อมูลไม่ถูกต้อง หรือข้อมูลมีความซ้ำซ้อนกัน เพราะฉะนั้นข้อมูลดิบจึงจำเป็นต้องผ่านกระบวนการจัดเตรียมข้อมูลให้เหมาะสม ประกอบไปด้วย 1) การเลือกข้อมูล (Selection) 2) การทำความสะอาดข้อมูล (Cleaning) และ 3) การได้มาของตัวแปรใหม่ (Derivation of new variables) ขั้นตอนนี้มีความสำคัญอย่างมากและมักจะใช้เวลาดำเนินการนานที่สุด

การเลือกข้อมูล ประกอบด้วยการเลือกในแนวตั้ง (แอตทริบิวต์/ตัวแปร) และการเลือกในแนวนอน (กรณีตัวอย่าง/ระเบียน) (Pérez et al., 2015) ผลลัพธ์ที่ได้จากการเลือกข้อมูลทำให้ชุดข้อมูลมีจำนวนตัวแปรลดลงและง่ายต่อการทำความเข้าใจ

การทำความสะอาดข้อมูล เป็นการแก้ไขข้อมูลที่ไม่สอดคล้องกัน (Inconsistent) มีความผิดปกติเพราะสัญญาณรบกวน (Noise) และอาจเป็นข้อมูลที่สูญหาย (Missing) สำหรับข้อมูลที่ไม่มีการจัดเก็บค่าในตารางจะถือว่าเป็นข้อมูลที่สูญหาย ส่วนข้อมูลมีความผิดปกติเนื่องด้วยมีความแตกต่างจากข้อมูลอื่นเกินกว่าระยะปกติของชุดข้อมูลนั้น ข้อมูลสูญหายและมีค่าที่ผิดปกติเกิดขึ้นได้เสมอ ดังนั้นจำเป็นต้องจัดการกับชุดข้อมูลเหล่านี้โดยไม่ลดทอนคุณภาพของการทำนาย สองเทคนิคที่นิยมใช้จัดการปัญหานี้คือ 1. การลบแอตทริบิวต์หรือระเบียนข้อมูลที่มีค่าขาดหายไปเป็นจำนวนมาก 2. การแทนค่าด้วยวิธีการบางอย่าง เช่น ค่ามัธยฐาน (Median) ค่าเฉลี่ย (Mean) ค่าคงที่ที่เหมาะสม (Constant) หรือค่าที่เลือกมาแบบสุ่ม (Random)

3. การวิเคราะห์ทางสถิติ (Statistical Analysis)

การวิเคราะห์ทางสถิติเป็นการวิเคราะห์เบื้องต้น โดยมีวัตถุประสงค์เพื่อการแสดงภาพรวมของชุดข้อมูลซึ่งจะช่วยให้ผู้วิจัยเข้าใจข้อมูลได้ดียิ่งขึ้นเพื่อนำไปสู่ขั้นตอนต่อไปที่ต้องอาศัยเทคนิคการทำเหมืองข้อมูลและอัลกอริทึมที่มีซับซ้อน ตัวอย่างค่าสถิติที่สนใจได้แก่ ความถี่ (Frequency) ฐานนิยม (Mode) มัธยฐาน (Median) ค่าเฉลี่ย (Mean) ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าความแปรปรวน (Variance) พิสัย (Range) ค่าสหสัมพันธ์ (Correlation) เป็นต้น

4. การประมวลผลข้อมูลล่วงหน้า (Data Preprocessing)

การประมวลผลข้อมูลล่วงหน้าประกอบไปด้วย 2 ขั้นตอนหลัก คือ การแปลงรูปแบบของข้อมูล และการเลือกคุณสมบัติหรือตัวแปร (Feature Selection: FS) การแปลงรูปแบบของข้อมูลหรือเรียกสั้น ๆ ว่า การแปลงข้อมูล เป็นกระบวนการที่จำเป็นในการขจัดความแตกต่างในชุดข้อมูลเพื่อให้การทำเหมืองข้อมูลได้ผลลัพธ์ที่เหมาะสมด้วยวิธีการทำให้เป็นบรรทัดฐาน (Normalization) หรือการทำให้เป็นมาตรฐาน (Standardization)

การเลือกคุณลักษณะหรือตัวแปรมีเป้าหมายเพื่อเลือกชุดข้อมูลย่อยจากตัวแปรอินพุตทั้งหมด โดยเลือกเฉพาะตัวแปรที่มีประสิทธิภาพและลดจำนวนตัวแปรที่ไม่เกี่ยวข้องออกไป ทั้งนี้ก็เพื่อรักษาผลลัพธ์ของการทำนายให้มีความแม่นยำ อีกทั้งยังช่วยลดเวลาในการคำนวณให้น้อยลงอีกด้วย ส่งผลให้ประสิทธิภาพในภาพรวมของโมเดลดีขึ้น

5. การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลเป็นขั้นตอนการสร้างโมเดลสำหรับการทำนายผลสัมฤทธิ์ทางการเรียนโดยใช้ฟังก์ชันการเรียนรู้แบบมีผู้สอน (Supervised Learning) เพื่อการประมาณค่าที่คาดหวังของตัวแปรตาม (Dependent Variables) ตามคุณลักษณะของตัวแปรอิสระ (Independent Variables) นอกจากโมเดลเชิง

ทำนาย (Predictive Model) การทำเหมืองข้อมูลยังมุ่งเน้นไปที่โมเดลเชิงพรรณนา (Descriptive Model) เพื่อใช้สร้างรูปแบบที่อธิบายโครงสร้างพื้นฐานของความสัมพันธ์และความเชื่อมโยงกันระหว่างข้อมูลที่ขุดได้ โดยใช้ฟังก์ชันการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ตัวอย่างของโมเดลเชิงทำนายได้แก่ การจำแนกประเภท และการถดถอย (Regression) ในขณะที่การจัดกลุ่ม และการสร้างกฎความสัมพันธ์ (Association Rules) จัดเป็นการสร้างโมเดลเชิงพรรณนา

การจำแนกประเภทเป็นวิธีที่ได้รับความนิยมมากที่สุด รองลงมาคือการถดถอยและการจัดกลุ่ม เทคนิคการจำแนกประเภทที่ใช้บ่อยที่สุดคือ เครือข่ายแบบเบส์ (Bayesian Networks) โครงข่ายประสาท และต้นไม้ตัดสินใจ สำหรับเทคนิคการถดถอยที่ใช้ทั่วไปคือ การถดถอยเชิงเส้น (Linear Regression) และการถดถอยแบบลอจิสติก (Logistic Regression) สำหรับอัลกอริทึมการจัดกลุ่มที่ได้รับความนิยมคือ โครงข่ายประสาท อัลกอริทึมเคมีน (K-means) การจัดกลุ่มแบบคลุมเครือ (Fuzzy Clustering) และการวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis)

สำหรับเครื่องมือในการทำเหมืองข้อมูล โปรแกรม WEKA เป็นเครื่องมือที่มีการนำไปใช้งานมากที่สุดสำหรับการสร้างโมเดลเชิงทำนาย (Jayaprakash, 2018) เนื่องจาก WEKA เป็นเครื่องมือที่มีฟังก์ชันการทำงานตอบโจทย์การทำเหมืองข้อมูลได้แทบทุกประเภท ไม่ว่าจะเป็นการประมวลผลข้อมูลล่วงหน้า การจำแนกประเภท การสร้างกฎความสัมพันธ์ การถดถอย และการสร้างภาพนามธรรม (Visualization) ตลอดจนความเป็นมิตรกับผู้ใช้ โดยเฉพาะผู้ใช้ที่ไม่ถนัดด้านการเขียนโปรแกรมคอมพิวเตอร์หรือแม้กระทั่งการทำเหมืองข้อมูล นอกจากนี้ยังมีโปรแกรม RapidMiner ที่ได้รับความนิยมรองลงมาและมีการนำไปใช้อย่างกว้างขวางเช่นกัน

6. การประเมินผลลัพธ์ (Result Evaluation)

เนื่องจากกระบวนการทำเหมืองข้อมูลจะมีการสร้างโมเดลหลาย ๆ แบบ เพื่อการประเมินและเลือกโมเดลที่เหมาะสมที่สุดมาสรุปเป็นผลลัพธ์จากงานวิจัย สำหรับการประเมินประสิทธิภาพของอัลกอริทึมการจำแนกประเภทจะนิยมใช้เมทริกซ์ความสับสน (Confusion Matrix) ซึ่งเป็นตารางที่ใช้เพื่อกำหนดประสิทธิภาพของอัลกอริทึมการจำแนกประเภท เมทริกซ์ความสับสนแสดงภาพนามธรรมและสรุปประสิทธิภาพของอัลกอริทึมการจำแนกประเภทดังแสดงในตาราง 2-1

ตาราง 2-1

เมทริกซ์ความสับสน (Confusion Matrix)

Observation	Actually Positive	Actually Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

1. **True Positive (TP)** = ทำนายตรงกับสิ่งที่เกิดขึ้น กรณีที่ทำนายว่าจริงและที่เกิดขึ้นก็คือจริง
2. **True Negative (TN)** = ทำนายตรงกับสิ่งที่เกิดขึ้น กรณีที่ทำนายว่าไม่จริงและที่เกิดขึ้นก็ไม่จริง
3. **False Positive (FP)** = ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือทำนายว่าจริงแต่สิ่งที่เกิดขึ้นคือไม่จริง
4. **False Negative (FN)** = ทำนายไม่ตรงกับที่เกิดขึ้น คือทำนายว่าไม่จริงแต่สิ่งที่เกิดขึ้นคือจริง

โดย TP, TN, FP และ FN ในตารางจะแทนด้วยค่าความถี่ ซึ่งสามารถนำมาคำนวณหาค่าของมาตรวัดชนิดต่าง ๆ ได้ เช่น ค่าความแม่นยำ หรือ ค่าความเที่ยง เป็นต้น เราใช้เมทริกซ์ความสับสนนี้สำหรับประเมินประสิทธิภาพการทำนายด้วยโมเดลที่เราสร้างขึ้นในรูปแบบของมาตรวัด ดังที่ระบุไว้ในตาราง 2-2

ตาราง 2-2

มาตรวัดประสิทธิภาพ

มาตรวัด	สมการ
ค่าความแม่นยำ (Accuracy)	(TPs + TNs) / (TPs+TNs+FPs + FNs)
ค่าความเที่ยงตรง (Precision)	TPs / (TPs + FPs)
ค่าการระลึก (Recall)	TPs/(TPs+FNs)
ค่าเอฟ (F-Measure)	2 x (Precision x Recall)/(Precision + Recall)

ในทางปฏิบัติ เราสามารถสร้างเมทริกซ์ความสับสนให้มีมิติได้มากกว่า 2 × 2 มิติ โดยมีรูปแบบเป็น N × N มิติก็ได้ ทั้งนี้ขึ้นอยู่กับจำนวนประเภทของข้อมูล (Class) ที่กำหนดไว้ในชุดข้อมูลเรียนรู้ (Learning Dataset)

การทบทวนวรรณกรรม

การทำเหมืองข้อมูลทางการศึกษา มีบทบาทสำคัญในการค้นพบรูปแบบของความรู้เกี่ยวกับปรากฏการณ์ทางการศึกษาและกระบวนการเรียนรู้ (Anoopkumar & Rahman, 2016) รวมถึงการทำความเข้าใจประสิทธิภาพด้านการศึกษาของผู้เรียน (Baker & Yacef, 2009) ด้วยเหตุนี้ การทำเหมืองข้อมูลได้ถูกนำมาใช้ในการทำนายผลสัมฤทธิ์ทางการเรียนในด้านต่าง ๆ มากมาย เช่น ประสิทธิภาพของการเรียน (Xing, 2019), การรักษาผู้เรียนไม่ให้พ้นสภาพ (Parker, Hogan, Eastabrook, Oke & Wood, 2006), ความสำเร็จของการเรียน (Richard-Eaglin, 2017), ความพึงพอใจของการเรียน (Alqurashi, 2019), และอัตราการออกกลางคัน (Pérez, Castellanos & Correal, 2018)

การคาดคะเนผลการศึกษาของนิสิตตั้งแต่เนิ่น ๆ สามารถช่วยให้ผู้บริหารและคณาจารย์ตัดสินใจดำเนินการปรับปรุงหลักสูตรในประเด็นที่มีความสำคัญต่อการเรียนรู้ภายในเวลาที่เหมาะสม และยังช่วยให้สามารถวางแผนการฝึกอบรมพิเศษที่เหมาะสมเพื่อปรับปรุงอัตราความสำเร็จของนิสิตได้อีกด้วย มีงานวิจัย

มากมายที่ใช้เทคนิคการทำเหมืองข้อมูลทางการศึกษาเพื่อทำนายความสำเร็จของนิสิต โดยสามารถกำหนดเป้าหมายของงานวิจัยได้หลายระดับดังนี้

1. ระดับปริญญา: ทำนายความสำเร็จของนิสิตว่าใช้เวลาในการศึกษานานเท่าไหร่จึงจะได้รับปริญญา งานวิจัยลักษณนี้จะใช้เทคนิคการจำแนกประเภท เช่น ผ่านหรือตก (Pass or Fail) และเทคนิคการถดถอย เช่น หาค่าผลการเรียนเฉลี่ยสะสม (GPAX) ซึ่งปกติแล้วจะนิยมใช้เกรดเฉลี่ยของ 2 ปีแรกเพื่อประมาณการค่าของเกรดเฉลี่ยสะสมในปีถัดไปจนกระทั่งจบการศึกษา

2. ระดับชั้นปี: เป็นการคาดการณ์ผลสัมฤทธิ์ทางการศึกษาของนิสิตภายในสิ้นปีการศึกษานั้น ๆ

3. ระดับรายวิชา: เป็นการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตในรายวิชานั้น ๆ จากงานวิจัยของ Garg (2018) ระบุว่า ความแม่นยำของการทำนายในระดับปริญญาและระดับชั้นปีอยู่ที่ 62% ถึง 89% ในขณะที่ การทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตในระดับรายวิชาให้ความแม่นยำมากกว่า 89% ซึ่งถือได้ว่าในระดับรายวิชาให้ความแม่นยำดีที่สุด

4. ระดับการสอบ: เป็นการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตในการสอบสำหรับรายวิชาใดวิชาหนึ่ง ซึ่งการทำนายในระดับนี้ไม่เป็นที่นิยมเพราะเป็นการทำนายเพียงการสอบครั้งเดียว ซึ่งไม่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการศึกษาในภาพรวมอย่างมีนัยสำคัญ

แอตทริบิวต์ที่ส่งผลต่อการศึกษาและนิยมนำมาใช้ในการทำเหมืองข้อมูลทางการศึกษาประกอบไปด้วย ผลสัมฤทธิ์ทางการศึกษาก่อนเข้ามหาวิทยาลัย (Prior-academic Achievement) กลุ่มลักษณะทางประชากรของผู้เรียน (Student Demographics) กิจกรรมเกี่ยวกับอีเลิร์นนิง (E-learning Activity) คุณลักษณะทางจิตวิทยา (Psychological Attributes) และสภาพแวดล้อมในการเรียน (Learning Environment) บทความของ Shahiri, Husain and Rashid (2015) ระบุว่า มี 69% ของงานวิจัยทางด้านนี้ที่ใช้ผลสัมฤทธิ์ทางการศึกษาก่อนเข้ามหาวิทยาลัยและกลุ่มลักษณะทางประชากรของผู้เรียนในการทำเหมืองข้อมูลทางการศึกษา ส่วนปัจจัยทั่วไปที่นิยมนำมาใช้ในการทำนายผลสัมฤทธิ์ทางการศึกษาของผู้เรียนมากที่สุด คือ ผลการประเมินภายใน (Internal Assessment) และผลการเรียนเฉลี่ยสะสม Almarabeh (2017) ระบุว่าข้อมูลก่อนเข้ามหาวิทยาลัยและระหว่างที่อยู่ในมหาวิทยาลัยมีผลต่อการทำนายผลสัมฤทธิ์ทางการศึกษา สำหรับข้อมูลระหว่างที่ศึกษาอยู่ในมหาวิทยาลัยทั้งเกรดที่นิสิตได้รับแล้วในแต่ละภาคการศึกษาและเกรดเฉลี่ยสะสมล้วนเป็นองค์ประกอบสำคัญในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตตลอดหลักสูตร

ปรียารัตน์ นาคสุวรรณ์ และกิดาการ สายธนู (2555) ได้นำเทคนิคการจำแนกประเภทด้วยข้อมูลทางสถิติและโครงข่ายประสาทมาเปรียบเทียบกัน เพื่อทำนายผลสัมฤทธิ์ทางการเรียนในรายวิชาสถิติเบื้องต้น ของมหาวิทยาลัยบูรพา ผลการวิจัยพบว่า ตัวแปรอิสระมีความสำคัญในการกำหนดความสำเร็จของนิสิตซึ่งประกอบไปด้วย เพศ คณะ ผลสัมฤทธิ์ทางการเรียนในรายวิชาแคลคูลัส การลงทะเบียนในรายวิชาสถิติเบื้องต้นเป็นครั้งแรก เกรดเฉลี่ยสะสมก่อนเข้ามหาวิทยาลัย และเกรดเฉลี่ยสะสมปัจจุบัน โดยแบ่งกลุ่มเป็น 1. ได้เกรด W และ F 2. ได้เกรด D และ 3. มีผลสัมฤทธิ์ตั้งแต่ D+ ขึ้นไป การเปรียบเทียบพบว่า เทคนิคโครงข่ายประสาทเทียมให้ค่าความแม่นยำสูงกว่าการวิเคราะห์ด้วยค่าทางสถิติ

นนทวัฒน์ ทวีชาติ, อรยา เพ็งประจญ, วิไลรัตน์ ยาทองไชย และชูศักดิ์ ยาทองไชย (2563) พัฒนาระบบการทำนายการพ้นสภาพของนิสิตระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ โดยใช้เทคนิคการจำแนกประเภทด้วยต้นไม้ตัดสินใจบนอัลกอริทึม J48 ตัวแปรที่ใช้ประกอบไปด้วย สาขาวิชาที่เรียน เกรดเฉลี่ยสะสม 6 ภาคการศึกษา เกรดเฉลี่ยจากโรงเรียนมัธยม หลักสูตรที่จบจากโรงเรียนมัธยม ขนาดของโรงเรียน และสถานะทุนกู้ยืมเพื่อการศึกษา ผลการวิจัยพบว่า โมเดลที่พัฒนาขึ้นสามารถทำนายการพ้นสภาพได้อย่างแม่นยำ โดยมีค่าความแม่นยำ ค่าความเที่ยงตรง และค่าการระลึก เกินร้อยละ 95 ในทุกมาตรวัด และมีความพึงพอใจจากการใช้งานระบบจากอาจารย์และนิสิตในระดับที่มาก

วันทนี ประจวบศุภกิจ, ธัญญาภรณ์ บุญยัง และสุกัญญา บุญศรี (2564) พัฒนาตัวแบบด้วยเทคนิคการทำเหมืองข้อมูลเพื่อทำนายผลสัมฤทธิ์ทางการศึกษาจากพฤติกรรมการใช้สมาร์ทโฟนของนิสิตระดับปริญญาตรี หลักสูตรทางคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือและมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี โดยใช้ชุดข้อมูลจำนวน 623 ตัวอย่าง กับตัวแปรจำนวน 19 ตัวแปร ตัวแบบถูกสร้างขึ้นด้วยอัลกอริทึม 5 ชนิด ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด วิธีโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-layer Perceptron) วิธีการเรียนรู้แบบเบส์อย่างง่าย (Naïve Bayes: NB) และวิธีการสุ่มป่าไม้ (Random Forest: RF) ผลการวิจัยพบว่า วิธีการสุ่มป่าไม้มีความแม่นยำมากที่สุด โดยที่ปัจจัยพฤติกรรมส่วนบุคคลและพฤติกรรมการเล่นสมาร์ทโฟนที่แตกต่างกันส่งผลต่อผลสัมฤทธิ์ทางการศึกษา เพราะฉะนั้นจึงทำให้เกิดผลลัพธ์ที่แตกต่างกันด้วย

Raheela, Agathe, Syed and Najmi (2017) ได้นำเทคนิคการทำเหมืองข้อมูลมาศึกษาผลสัมฤทธิ์ทางการศึกษาของนิสิตระดับปริญญาตรีจำนวน 210 คน โดยสร้างตัวแบบเพื่อทำนายผลการเรียนที่นิสิตได้รับเมื่อสิ้นสุดหลักสูตร และยังหาความเกี่ยวข้องของผลการเรียนกับความก้าวหน้าที่เกิดขึ้นในระหว่างที่ยังคงศึกษาอยู่ ตัวแปรที่ใช้เป็นตัวแปรที่เกี่ยวข้องกับคะแนนหรือเกรดเท่านั้น อัลกอริทึมต้นไม้ตัดสินใจถูกนำมาใช้สำหรับการสร้างตัวแบบเพื่อจำแนกประเภทของชุดข้อมูลด้วยเกณฑ์สารสนเทศ 4 ตัว ประกอบไปด้วย Information Gain, Gini Index, ค่าความแม่นยำ และ Gain Ratio ในส่วนของการแบ่งกลุ่มข้อมูลผู้วิจัยได้เลือกอัลกอริทึมเอกซ์มีน (X-means) กับระยะห่างแบบยุคลิด (Euclidean Distance) และใช้เกณฑ์สารสนเทศของเบส์ (Bayesian Information Criterion: BIC) เป็นมาตรวัดในการแบ่งจำนวนกลุ่มที่ดีที่สุด ผู้วิจัยได้แบ่งนิสิตออกเป็นกลุ่มที่มีผลการเรียนดีและผลการเรียนต่ำ การประมวลผลด้านการทำเหมืองข้อมูลใช้โปรแกรม RapidMiner ในทุกขั้นตอน ผลการวิจัยพบว่า การคาดการณ์ว่านิสิตจะถูกจัดอยู่ในกลุ่มใดสามารถช่วยสนับสนุนการเรียนรู้ของนิสิตในกลุ่มที่มีผลการเรียนต่ำได้แต่เนิ่น ๆ อีกทั้งยังให้คำแนะนำหรือโอกาสบนเส้นทางการศึกษากับนิสิตที่อยู่ในกลุ่มที่มีผลการเรียนดีได้อีกด้วย

Migueis, Freitas, Garcia and Silva (2018) ใช้อัลกอริทึมการจำแนกประเภทในการสร้างโมเดลสองขั้นตอน (Two-stage Model) เพื่อทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตด้วยชุดข้อมูลนิสิตจำนวน 2459 คน โดยอาศัยข้อมูลที่ได้ในปีแรกของการศึกษา ในงานวิจัยได้ประยุกต์ใช้อัลกอริทึมต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน วิธีแบบเบส์อย่างง่าย เทคนิคการสุ่มป่าไม้ เทคนิคต้นไม้แบ็กกิ้ง (Bagged Tree) และเทคนิคต้นไม้แบบเพิ่มระดับ (Boosted Tree) ผลการวิจัยพบว่า ผลลัพธ์ที่ได้จากโมเดลมีค่าความแม่นยำสูงกว่า 95%

และเทคนิคการสุ่มป่าให้ค่าความแม่นยำดีที่สุด จะเห็นได้ว่าโมเดลสามารถนำไปใช้ในการวางแผนการเรียนทั้งในด้านส่งเสริมหรือสนับสนุนการศึกษาและป้องกันการพ้นสภาพของนิสิตได้อย่างมีประสิทธิภาพ

Baashar et al. (2022) สำรวจและวิเคราะห์วรรณกรรมใหม่ ๆ ที่เกี่ยวข้องกับการนำอัลกอริทึมโครงข่ายประสาทเทียมไปใช้ทำนายผลสัมฤทธิ์ทางการศึกษาของผู้เรียนในระดับมหาวิทยาลัย ผลการสำรวจพบว่า เทคนิคโครงข่ายประสาทเทียมนิยมนำไปใช้ผสมผสานกับการวิเคราะห์ข้อมูลและการทำเหมืองข้อมูลในประเภทอื่น ๆ เพื่อค้นหารูปแบบและประเมินผลสัมฤทธิ์ทางวิชาการ (Academic Achievement) งานวิจัยระบุว่า การศึกษาทางด้านนี้มุ่งเน้นไปที่ระดับมหาวิทยาลัยเป็นส่วนใหญ่เนื่องด้วยมีการเก็บข้อมูลของผู้เรียนในระดับนี้อยู่เป็นจำนวนมาก อีกทั้งผู้วิจัยที่มีทักษะความรู้ในการทำวิจัยด้านการทำเหมืองข้อมูลทางการศึกษามักเป็นบุคลากรในระดับมหาวิทยาลัย จึงไม่แปลกที่ชุดข้อมูลส่วนใหญ่มาจากนิสิตนักศึกษาเป็นหลัก ผลลัพธ์อีกประการหนึ่งคือ โครงข่ายประสาทเทียมมักให้ผลลัพธ์ของการประเมินประสิทธิภาพโมเดลด้วยค่าความแม่นยำสูงกว่าอัลกอริทึมชนิดอื่น นอกจากนี้ตัวแปรที่นิยมใช้คือเกรดเฉลี่ยสะสม ส่วนตัวแปรอื่น ๆ มีการนำมาใช้อยู่บ้างแต่ก็ไม่มีผลกระทบต่อประสิทธิภาพของโมเดลมากนัก และในปีเดียวกัน Baashar et al. (2022) ศึกษาผลสัมฤทธิ์ทางการศึกษาของนิสิตระดับปริญญาโทจำนวน 635 คนจากหลากหลายคณะ เช่น คณะบริหารธุรกิจ คณะวิศวกรรมศาสตร์ คณะเทคโนโลยีสารสนเทศ คณะการจัดการ เป็นต้น โดยการสร้างโมเดลเพื่อการทำนายด้วยอัลกอริทึมการเรียนรู้ของเครื่อง 6 อัลกอริทึม เป้าหมายของงานวิจัยจะมุ่งเน้นไปที่เกรดเฉลี่ยสะสมของนิสิตเป็นหลัก ประสิทธิภาพของโมเดลวัดด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) ระหว่างเกรดเฉลี่ยสะสมที่ได้จากการทำนายและเกรดเฉลี่ยสะสมจริง ผลการทดลองระบุว่า อัลกอริทึมโครงข่ายประสาทเทียมมีประสิทธิภาพดีที่สุดเพราะมีค่าความคลาดเคลื่อนต่ำที่สุดเมื่อเปรียบเทียบกับ 5 อัลกอริทึมที่เหลือ

Aldowash, Al-Samarraie and Fauzy (2019) สำรวจและสังเคราะห์วรรณกรรมที่เกี่ยวข้องกับการทำเหมืองข้อมูลทางการศึกษาและการวิเคราะห์การเรียนรู้ (Learning Analytics: LA) จากบทความวิจัยในศตวรรษที่ 21 จำนวน 402 บทความ พบว่า เทคนิคการทำเหมืองข้อมูลทางการศึกษาและการวิเคราะห์การเรียนรู้สามารถช่วยแก้ปัญหาที่เกี่ยวข้องกับการเรียนรู้ที่แตกต่างกันได้ เทคนิคการทำเหมืองข้อมูลที่มีการนำมาประยุกต์ใช้ได้แก่ การจำแนกประเภท (26.23%) การจัดกลุ่ม (21.25%) การทำเหมืองข้อมูลแบบรูปภาพ (15%) การใช้ค่าทางสถิติ (14.25%) การค้นหากฎความสัมพันธ์ (14%) การถดถอย (10.25%) การค้นหารูปแบบแบบลำดับ (6.5%) เป็นต้น การนำเทคนิคเหล่านี้มาใช้ส่งผลให้เกิดการพัฒนากลยุทธ์ในการเรียนรู้ของผู้เรียน นอกจากนี้ในบทความยังแนะนำเทคนิคที่เหมาะสมให้กับนักวิจัยสำหรับการทำวิจัยในกรณีต่าง ๆ อีกทั้งยังช่วยแนะนำเครื่องมือและข้อมูลที่จำเป็นเพื่อปรับปรุงระบบการศึกษาในระดับมหาวิทยาลัยให้ดียิ่งขึ้น

จากการทบทวนวรรณกรรมข้างต้นจะเห็นว่า การทำนายผลสัมฤทธิ์ทางการศึกษาของผู้เรียนสามารถทำได้อย่างแม่นยำ และหากใช้วิธีจำแนกประเภทที่มีจำนวนคลาสไม่มาก เช่น ทำนายว่าผ่านหรือตก ทำนายว่านิสิตอยู่ในกลุ่มผลการเรียนดีหรือไม่ดี เป็นต้น จะยิ่งให้ความแม่นยำที่สูงขึ้นไปอีก ประสิทธิภาพของโมเดลขึ้นอยู่กับหลายปัจจัย อาทิเช่น อัลกอริทึมที่เลือกใช้สร้างโมเดล ตัวแปรที่ใช้ในชุดข้อมูล หรือจำนวนชุดข้อมูลล้วนมีผลต่อการทำเหมืองข้อมูลทางการศึกษาทั้งสิ้น แอตทริบิวต์ที่เลือกจากข้อมูลส่วนบุคคลของผู้เรียนไม่ว่า

จะเป็น อายุ เพศ ศาสนา สถานที่อยู่อาศัย ครอบครัว งาน หรือผลการเรียนเฉลี่ยสะสมที่ผ่านมา เพื่อทำนายผลสัมฤทธิ์ทางการศึกษาล้วนมีส่วนสำคัญต่อการทำนายอย่างถูกต้อง (Zimmermann, Brodersen, Heinimann and Buhmann, 2015) อย่างไรก็ตาม การทำนายด้วยวิธีการเหล่านี้ยังพบได้น้อยสำหรับชุดข้อมูลของผู้เรียนในหลักสูตรสารสนเทศศึกษาหรือสารสนเทศศาสตร์ เฉพาะฉะนั้น ผู้วิจัยจึงเห็นว่าการวิเคราะห์ข้อมูลชุดนี้จะได้รับองค์ความรู้ใหม่ ๆ ที่เป็นประโยชน์สำหรับการออกแบบหลักสูตรสารสนเทศศึกษาให้ดียิ่งขึ้น

บทที่ 3
วิธีการดำเนินการวิจัย

การวิเคราะห์ชุดข้อมูลที่มีความเกี่ยวข้องกับการศึกษาที่ผ่านมา นอกจากการวิเคราะห์ข้อมูลทางสถิติแล้ว ยังได้มีการนำเทคนิคการทำเหมืองข้อมูลทางการศึกษามาประยุกต์ใช้ โดยมุ่งเน้นไปที่การทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตเมื่อสำเร็จการศึกษาแล้ว การค้นคว้าวิจัยทางด้านนี้สามารถลงลึกในรายละเอียดที่แตกต่างกัน ตั้งแต่ระดับรายวิชาไปจนถึงระดับหลักสูตร ทั้งนี้ก็เพื่อการจัดกลุ่มเป้าหมายโดยพิจารณาจากกลุ่มผู้เรียนที่มีลักษณะหรือผลการเรียนคล้ายคลึงกัน ตัวอย่างเช่น กลุ่มนิสิตที่มีผลการเรียนผ่านหรือตกในรายวิชาใดวิชาหนึ่ง กลุ่มนิสิตที่มีระดับผลการเรียนที่ใกล้เคียงกันในแต่ละภาคการศึกษา หรือ กลุ่มนิสิตที่มีผลสัมฤทธิ์ทางการเรียนในระดับต่าง ๆ เมื่อสำเร็จการศึกษาแล้ว เป็นต้น ทั้งนี้ก็เพื่อวิเคราะห์รูปแบบหรือพฤติกรรมทั่วไปจากนิสิตทั้งหมด ไม่ว่าจะเป็นคุณลักษณะทางประชากรหรือผลการเรียนตลอดหลักสูตร และช่วยให้คณาจารย์ผู้รับผิดชอบหลักสูตรเข้าใจปัจจัยหรือคุณลักษณะต่าง ๆ ที่มีผลต่อการเรียนของนิสิต เพื่อออกแบบแนวทางในการสนับสนุนการศึกษาให้กับนิสิตที่มีผลการเรียนต่ำ หรือ ส่งเสริมให้นิสิตที่มีผลการเรียนดีพัฒนาต่อยอดให้มีความสำเร็จที่ดียิ่งขึ้น จะเห็นได้ว่า การทำเหมืองข้อมูลทางการศึกษานี้เป็นประโยชน์อย่างยิ่งในการพัฒนาหลักสูตรในอนาคต วิธีการดำเนินการวิจัยจะอิงหลักการทำเหมืองข้อมูลทางการศึกษาโดยมีรายละเอียดดังต่อไปนี้

การจัดเตรียมข้อมูล (Data Preparation)

ในการประเมินประสิทธิภาพของโมเดลการทำเหมืองข้อมูลจำเป็นต้องแน่ใจก่อนกว่าชุดข้อมูลที่นำมาเข้ากระบวนการมีคุณภาพเพียงพอ นั่นหมายถึง ข้อมูลมีความเหมาะสม สมบูรณ์ สม่ำเสมอ และถูกต้อง หากคุณภาพของข้อมูลไม่ดีอาจนำไปสู่ผลลัพธ์ที่ไม่ตรงกับความเป็นจริงหรือไม่ถูกต้องได้ ส่งผลให้เสียเวลาและทรัพยากรไปโดยเปล่าประโยชน์ ดังนั้น การจัดเตรียมข้อมูลจึงมีความสำคัญอย่างมากโดยจะอธิบายอย่างละเอียดในลำดับถัดไป

ปกติแล้วชุดข้อมูลที่นำมาใช้ทำเหมืองข้อมูลทางการศึกษาสามารถได้มาจากหลากหลายแหล่งข้อมูล แต่ที่ได้รับความนิยมใช้กันมากในปัจจุบันจะเป็นชุดข้อมูลที่ได้จากระบบสารสนเทศของผู้เรียน (Student Information System: SIS) ของสถาบันการศึกษานั้น ๆ ซึ่งข้อมูลที่อยู่ในระบบประกอบไปด้วย ข้อมูลเกรดในแต่ละรายวิชา ข้อมูลเกรดเฉลี่ยสะสม รวมถึงข้อมูลลักษณะทางประชากร เช่น อายุ เพศ เชื้อชาติ ศาสนา ที่อยู่ รายได้หรืออาชีพผู้ปกครอง วุฒิการศึกษาของผู้ปกครอง เป็นต้น นอกจากนี้ ชุดข้อมูลนำเข้ายังสามารถรวบรวมได้จากการทำแบบสำรวจหรือจากข้อมูลในระบบอีเลิร์นนิงอีกด้วย

สำหรับงานวิจัยนี้จะใช้ชุดข้อมูลที่มีอยู่แล้วในระบบบริการการศึกษาของมหาวิทยาลัยบูรพา โดยคัดเลือกคุณลักษณะหรือแอตทริบิวต์ที่น่าสนใจจากกลุ่มข้อมูลลักษณะทางประชากร เกรดเฉลี่ยก่อนเข้ารับการศึกษา (เกรดเฉลี่ยจากโรงเรียนมัธยม) กลุ่มข้อมูลเกรดในแต่ละรายวิชา เกรดเฉลี่ยของแต่ละภาคการศึกษา

ใน 3 ปีแรก และเกรดเฉลี่ยสะสมตอนสำเร็จการศึกษาแล้ว สำหรับกลุ่มข้อมูลเกรดในแต่ละรายวิชายังแยกรายวิชาออกได้เป็น 3 กลุ่มย่อยคือ 1. รายวิชาศึกษาทั่วไป (General Education: GE) 2. รายวิชาด้านสารสนเทศศาสตร์ (Information Science: IS) 3. รายวิชาด้านเทคโนโลยีสารสนเทศ (Information Technology: IT)

หลังจากคัดเลือกแอตทริบิวต์ที่จะนำมาทำการวิเคราะห์ได้แล้ว ลำดับต่อไปจะทำการตรวจสอบความถูกต้องและทำความสะอาดข้อมูลสำหรับข้อมูลที่มีความผิดปกติและข้อมูลที่สูญหาย เพื่อให้ได้ชุดข้อมูลที่มีความพร้อมสำหรับขั้นตอนการขุดเหมืองข้อมูล

ผลสัมฤทธิ์ทางการศึกษาเมื่อสำเร็จการศึกษาแล้วจะกำหนดให้เป็นประเภทของข้อมูลหรือคลาส ซึ่งอาจเรียกว่า ตัวแปรเป้าหมาย โดยแบ่งออกเป็น 4 กลุ่ม ดังแสดงในตาราง 3-1

ตาราง 3-1

ประเภทของข้อมูลเป้าหมาย

GPAX	Class
3.50 – 4.00	Excellent
3.00 – 3.49	Very Good
2.50 – 2.99	Good
2.00 – 2.49	Fair

ตัวแปรที่ใช้สำหรับการสร้างโมเดลพร้อมความหมายของตัวแปรแต่ละตัวได้กำหนดไว้ในตาราง 3-2 ตัวแปรลำดับที่ 1 ถึง 8 เป็นตัวแปรสำหรับชุดข้อมูลด้านประชากร โดยมีชนิดของตัวเลือกที่แตกต่างกัน ตัวแปรที่เป็นเกรดจะมีข้อมูลแบบไม่เป็นตัวเลขแต่เป็นเกรดตั้งแต่ A ถึง D ทั้งหมด 7 เกรด คือ {A, B+, B, C+, C, D+, D} ซึ่งอยู่ในลำดับที่ 9 ถึง 11 ในส่วนของตัวแปรชนิดตัวเลขที่เป็นเกรดเฉลี่ยของนิสิตจะแบ่งออกเป็น 4 กลุ่ม คือ {Excellent, Very Good, Good, Fair} ตามเกณฑ์ที่กำหนดไว้ในตาราง 3-1 ซึ่งเป็นตัวแปรลำดับที่ 1, 12 และ 13

ตัวแปรลำดับที่ 2, 3 และ 4 แบ่งกลุ่มประเภทได้ค่อนข้างชัดเจนอยู่แล้ว เริ่มจากจังหวัดที่อยู่ของนิสิตแบ่งตามภาคทั้งหมด 5 ภาค ได้แก่ เหนือ ใต้ ตะวันออกเฉียงเหนือ ตะวันตก และกลาง ส่วนเพศของนิสิตประกอบไปด้วยเพศชาย (Male) และเพศหญิง (Female) ส่วนข้อมูลด้านมีพี่น้องหรือไม่ แบ่งได้ว่ามี (Yes) และไม่มี (No)

ตัวแปรลำดับที่ 5 และ 7 แบ่งเป็นกลุ่มรายได้ของบิดามารดา ที่กำหนดให้มี 5 กลุ่ม คือ {High, Medium, Low, None, Undefine} ซึ่งตรงกับรายได้ {> 300,000 บาทต่อปี (> 25,000 บาทต่อเดือน), 150,000 - 300,000 บาทต่อปี (12,500 - 25,000 บาทต่อเดือน), < 150,000 บาทต่อปี (< 12,500 บาทต่อเดือน), ไม่มีรายได้, ไม่ระบุ} ตามลำดับ

ตัวแปรลำดับที่ 6 และ 8 แบ่งเป็นกลุ่มอาชีพของบิดามารดา ที่กำหนดให้มี 6 กลุ่ม คือ {Government Officer, State Enterprise, Private Employee, Personal Business, Agriculture, Undefine} ซึ่งตรงกับอาชีพ {รับราชการ/พนักงานราชการ/ลูกจ้างหน่วยงานราชการ, รัฐวิสาหกิจ, พนักงานหน่วยงานเอกชน/ลูกจ้างหน่วยงานเอกชน, ค้าขาย/ธุรกิจส่วนตัว/อาชีพอิสระ/รับจ้างอิสระแบบไม่ประจำ, เกษตร/ประมง, ไม่ระบุ} ตามลำดับ

ตาราง 3-2

แสดงตัวแปรที่ใช้ในการสร้างโมเดลพร้อมคำอธิบาย

ลำดับ	ตัวแปร	คำอธิบาย	ค่าที่เป็นไปได้
1	SGPA	เกรดเฉลี่ยจากโรงเรียนมัธยม	{Excellent, Very Good, Good, Fair}
2	Region	จังหวัดที่อยู่ของนิสิตแบ่งตามภาค	{North, South, Northeast, East, Central}
3	Gender	เพศของนิสิต	{Male, Female}
4	Sibling	มีพี่น้องหรือไม่	{Yes, No}
5	Father Income	รายได้บิดา	{High, Medium, Low, None, Undefine}
6	Father Occupation	อาชีพบิดา	{Government Officer, State Enterprise, Private Employee, Personal Business, Agriculture, Undefine}
7	Mother Income	รายได้มารดา	{High, Medium, Low, None, Undefine}
8	Mother Occupation	อาชีพมารดา	{Government Officer, State Enterprise, Private Employee, Personal Business, Agriculture, Undefine}
9	GESubjects	เกรดที่ได้รับในรายวิชาศึกษาทั่วไป	{A, B+, B, C+, C, D+, D}
10	ISSubjects	เกรดที่ได้รับในรายวิชาด้านสารสนเทศศาสตร์	{A, B+, B, C+, C, D+, D}
11	ITSubjects	เกรดที่ได้รับในรายวิชาด้านเทคโนโลยีสารสนเทศ	{A, B+, B, C+, C, D+, D}
12	GPA1-GPA3	เกรดเฉลี่ยของภาคการศึกษาที่ 1-6	{Excellent, Very Good, Good, Fair}
13	AGPA2-AGPA6	เกรดเฉลี่ยรวมของภาคการศึกษาที่ 2-6	{Excellent, Very Good, Good, Fair}

สำหรับตัวแปรในกลุ่มรายวิชาซึ่งอยู่ในลำดับที่ 9, 10 และ 11 มีรายวิชาอยู่เป็นจำนวนมาก ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็นหลากหลายแบบ ได้แก่ ชุดข้อมูลรายวิชาศึกษาทั่วไป ชุดข้อมูลรายวิชาเอกบังคับ ชุดข้อมูลรายวิชาเอกเลือก ชุดข้อมูลรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ ชุดข้อมูลรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ ชุดข้อมูลรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ และชุดข้อมูลรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ เพื่อสร้างโมเดลและทำการทดลองด้วยสภาพแวดล้วที่แตกต่างกัน สำหรับประเด็นเรื่องชุดข้อมูลได้อธิบายเพิ่มเติมไว้ในหัวข้อการพัฒนาโมเดลการทำเหมืองข้อมูลทางการศึกษา หน้าที่ 34

ถัดไปเป็นตารางข้อมูลของรายวิชาในแต่ละกลุ่ม ภายใต้หลักสูตรศิลปศาสตรบัณฑิต สาขาวิชาสารสนเทศศึกษา หลักสูตรปรับปรุง พ.ศ. 2559 โดยมีตัวอย่างชื่อรายวิชาในแต่ละกลุ่มดังนี้

หมวดรายวิชาศึกษาทั่วไป ได้แก่ ภาษาอังกฤษเพื่อการสื่อสาร ภาษาอังกฤษระดับมหาวิทยาลัย การเขียนภาษาอังกฤษเพื่อการสื่อสาร ทักษะการใช้ภาษาไทยเพื่อการสื่อสาร ชีวิตและสุขภาพ การออกกำลังการเพื่อคุณภาพชีวิต จิตวิทยาในการดำเนินชีวิตและการปรับตัว หลักเศรษฐกิจพอเพียงกับการพัฒนาสังคม ภัยพิบัติทางธรรมชาติ ก้าวทันนวัตกรรมทางวิทยาศาสตร์ ศิลปะและการคิดสร้างสรรค์ ก้าวทันสังคมดิจิทัลด้วยไอซีที โดยแสดงไว้ในตาราง 3-3

ตาราง 3-3

รายวิชาศึกษาทั่วไป

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	ภาษาอังกฤษเพื่อการสื่อสาร	English for Communication
2	ภาษาอังกฤษระดับมหาวิทยาลัย	Collegiate English
3	การเขียนภาษาอังกฤษเพื่อการสื่อสาร	English Writing for Communication
4	ทักษะการใช้ภาษาไทยเพื่อการสื่อสาร	Thai Language Skills for Communication
5	ชีวิตและสุขภาพ	Life and Health
6	การออกกำลังกายเพื่อคุณภาพชีวิต	Exercise for Quality of Life
7	จิตวิทยาในการดำเนินชีวิตและการปรับตัว	Psychology for Living and Adjustment
8	หลักเศรษฐกิจพอเพียงกับการพัฒนาสังคม	Sufficiency Economy and Social Development
9	ภัยพิบัติทางธรรมชาติ	Natural Disasters
10	ก้าวทันนวัตกรรมทางวิทยาศาสตร์	Contemporary Scientific Innovation
11	ศิลปะและการคิดสร้างสรรค์	Arts and Creativity
12	ก้าวทันสังคมดิจิทัลด้วยไอซีที	Moving Forward in a Digital Society with ICT

หมวดรายวิชาเอกบังคับทั้งหมดประกอบไปด้วย 18 รายวิชา และรายวิชาเอกเลือกทั้งหมดประกอบไปด้วย 10 รายวิชา ที่ซึ่งมีทั้งวิชาด้านสารสนเทศศาสตร์ และเทคโนโลยีสารสนเทศ ดังแสดงไว้ในตาราง 3-4 และตาราง 3-5

ตาราง 3-4

รายวิชาเอกบังคับทั้งหมด

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	การอ่านเพื่อวิชาชีพสารสนเทศ	Reading for Information Professional
2	เทคโนโลยีสารสนเทศ	Information Technology
3	สารสนเทศศาสตร์	Information Science
4	การพัฒนาทรัพยากรสารสนเทศ	Collection Development
5	บริการสารสนเทศและช่วยการค้นคว้า	Information and Reference Service
6	การนำเสนอและการฝึกอบรมในงานสารสนเทศ	Presentation and Training in Information Work
7	การจัดการเอกสารและสารสนเทศอิเล็กทรอนิกส์	Electronic Information and Record Management
8	การวิเคราะห์และออกแบบระบบสารสนเทศ	Information Systems Analysis and Design
9	การโปรแกรมในงานสารสนเทศ	Programming in Information Work
10	การจัดการสถาบันสารสนเทศ	Management of Information Institutes
11	การจัดระบบทรัพยากรสารสนเทศ	Organization of Information Resources
12	การทำรายการทรัพยากรสารสนเทศ	Cataloging of Information Resources
13	การจัดการฐานข้อมูลสำหรับงานสารสนเทศ	Database Management for Information Work
14	การออกแบบเว็บสำหรับงานสารสนเทศ	Web Design for Information Work
15	ระบบห้องสมุดอัตโนมัติ	Library Automation Systems
16	สัมมนาทิศทางและแนวโน้มทางสารสนเทศศาสตร์	Seminar on Current Issues and Trend in Information Science
17	การวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน	Library of Congress Classification
18	การวิจัยและสถิติทางสารสนเทศศึกษา	Research and Statistics in Information Studies

ตาราง 3-5

รายวิชาเอกเลือกทั้งหมด

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	ภาษาอังกฤษเพื่อวิชาชีพสารสนเทศ	English for Information Professional
2	การพิมพ์และธุรกิจการพิมพ์	Printing and Publishing Business
3	สารสนเทศทางวิทยาศาสตร์และเทคโนโลยี	Information in Science and Technology
4	การจัดการความรู้	Knowledge Management
5	การค้นคืนสารสนเทศอิเล็กทรอนิกส์	Electronic Information Retrieval
6	เทคโนโลยีมัลติมีเดีย	Multimedia Technology
7	การสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์	Data Communication and Computer Networks
8	การประยุกต์โปรแกรมกลุ่มโอเพนซอร์ส	Open Source Program Application
9	พฤติกรรมผู้ใช้และความต้องการสารสนเทศ	User Behavior and Information Needs
10	สารสนเทศภูมิปัญญาท้องถิ่น	Information of Local Wisdom

ตาราง 3-6

รายวิชาเอกบังคับด้านสารสนเทศศาสตร์

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	การอ่านเพื่อวิชาชีพสารสนเทศ	Reading for Information Professional
2	สารสนเทศศาสตร์	Information Science
3	การพัฒนาทรัพยากรสารสนเทศ	Collection Development
4	บริการสารสนเทศและช่วยการค้นคว้า	Information and Reference Service
5	การจัดการสถาบันสารสนเทศ	Management of Information Institutes
6	การจัดระบบทรัพยากรสารสนเทศ	Organization of Information Resources
7	การทำรายการทรัพยากรสารสนเทศ	Cataloging of Information Resources
8	ระบบห้องสมุดอัตโนมัติ	Library Automation Systems
9	การวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน	Library of Congress Classification

สำหรับรายวิชาด้านสารสนเทศศาสตร์ทั้งเอกบังคับ (9 วิชา) และเอกเลือก (5 วิชา) ได้จัดกลุ่มไว้ในตาราง 3-6 และ ตาราง 3-7 ซึ่งประกอบไปด้วยรายวิชา การอ่านเพื่อวิชาชีพสารสนเทศ สารสนเทศศาสตร์ การพัฒนาทรัพยากรสารสนเทศ บริการสารสนเทศและช่วยการค้นคว้า การจัดการสถาบันสารสนเทศ การจัดระบบทรัพยากรสารสนเทศ การทำรายการทรัพยากรสารสนเทศ ระบบห้องสมุดอัตโนมัติ การวิเคราะห์

หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน ภาษาอังกฤษเพื่อวิชาชีพสารสนเทศ การพิมพ์และธุรกิจการพิมพ์ การจัดการความรู้ พฤติกรรมผู้ใช้และความต้องการสารสนเทศ และ สารสนเทศภูมิปัญญาท้องถิ่น

ทั้งนี้ รายวิชาฝึกงาน นิสิตได้เกรดใกล้เคียงกัน มีจำนวนหน่วยกิตมากกว่ารายวิชาปกติ และถูกบรรจุอยู่ในภาคการศึกษาสุดท้าย จึงไม่นำมาเป็นส่วนหนึ่งของการสร้างโมเดลการทำนายในงานวิจัยนี้

ตาราง 3-7

รายวิชาเอกเลือกด้านสารสนเทศศาสตร์

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	ภาษาอังกฤษเพื่อวิชาชีพสารสนเทศ	English for Information Professional
2	การพิมพ์และธุรกิจการพิมพ์	Printing and Publishing Business
3	การจัดการความรู้	Knowledge Management
4	พฤติกรรมผู้ใช้และความต้องการสารสนเทศ	User Behavior and Information Needs
5	สารสนเทศภูมิปัญญาท้องถิ่น	Information of Local Wisdom

ตาราง 3-8

รายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	เทคโนโลยีสารสนเทศ	Information Technology
2	การนำเสนอและการฝึกอบรมในงานสารสนเทศ	Presentation and Training in Information Work
3	การจัดการเอกสารและสารสนเทศอิเล็กทรอนิกส์	Electronic Information and Record Management
4	การวิเคราะห์และออกแบบระบบสารสนเทศ	Information Systems Analysis and Design
5	การโปรแกรมในงานสารสนเทศ	Programming in Information Work
6	การจัดการฐานข้อมูลสำหรับงานสารสนเทศ	Database Management for Information Work
7	การออกแบบเว็บสำหรับงานสารสนเทศ	Web Design for Information Work
8	สัมมนาทิศทางและแนวโน้มทางสารสนเทศศาสตร์	Seminar on Current Issues and Trend in Information Science
9	การวิจัยและสถิติทางสารสนเทศศึกษา	Research and Statistics in Information Studies

สำหรับรายวิชาด้านเทคโนโลยีสารสนเทศทั้งเอกบังคับ (9 วิชา) และเอกเลือก (5 วิชา) ได้จัดกลุ่มไว้ในตาราง 3-8 และ ตาราง 3-9 ได้แก่ เทคโนโลยีสารสนเทศ การนำเสนอและการฝึกอบรมในงานสารสนเทศ การจัดการเอกสารและสารสนเทศอิเล็กทรอนิกส์ การวิเคราะห์และออกแบบระบบสารสนเทศ การโปรแกรมในงานสารสนเทศ การจัดการฐานข้อมูลสำหรับงานสารสนเทศ การออกแบบเว็บสำหรับงานสารสนเทศ สัมมนาทิศทางและแนวโน้มทางสารสนเทศศาสตร์ การวิจัยและสถิติทางสารสนเทศศึกษา สารสนเทศทางวิทยาศาสตร์และเทคโนโลยี การค้นคืนสารสนเทศอิเล็กทรอนิกส์ เทคโนโลยีมัลติมีเดีย การสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ และ การประยุกต์โปรแกรมกลุ่มโอเพนซอร์ส

ตาราง 3-9

รายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ

ลำดับ	ชื่อภาษาไทย	ชื่อภาษาอังกฤษ
1	สารสนเทศทางวิทยาศาสตร์และเทคโนโลยี	Information in Science and Technology
2	การค้นคืนสารสนเทศอิเล็กทรอนิกส์	Electronic Information Retrieval
3	เทคโนโลยีมัลติมีเดีย	Multimedia Technology
4	การสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์	Data Communication and Computer Networks
5	การประยุกต์โปรแกรมกลุ่มโอเพนซอร์ส	Open Source Program Application

สำหรับวิชาเอกเลือกบางรายวิชาอาจถูกแทนที่ด้วยรายวิชาเอกเลือกอื่นที่คล้ายคลึงกันซึ่งได้อธิบายไว้ในหัวข้อรายวิชาเอกเลือก (Elective Subjects Grade) หน้าที่ 46 ตัวอย่างของรายวิชาเอกเลือกที่ถูกแทนที่ เช่น สารสนเทศทางมนุษยศาสตร์และสังคมศาสตร์ การส่งเสริมการอ่าน สารสนเทศสำหรับเด็กและเยาวชน ธุรกิจสารสนเทศ บริการสารสนเทศเพื่อสังคมแห่งการเรียนรู้ จดหมายเหตุและพิพิธภัณฑ์ ห้องสมุดดิจิทัล การทำเหมืองข้อมูล เป็นต้น

จะเห็นได้ว่า รายวิชาด้านสารสนเทศศาสตร์มีจำนวนมากที่สุด ดังนั้น ผู้วิจัยจะทำการคัดเลือกเป็นบางรายวิชาเพื่อให้แต่ละกลุ่มรายวิชามีจำนวนเท่ากัน โดยพิจารณาจากลำดับความสำคัญและจำนวนนิสิตที่ลงทะเบียนเรียน ส่วนรายวิชาเอกบังคับจะให้น้ำหนักมากกว่าวิชาเอกเลือกเพราะถือว่าเป็นหัวใจของหลักสูตร สำหรับกระบวนการคัดเลือกข้อมูลจะอยู่ในขั้นตอนของการจัดเตรียมข้อมูลตามที่ได้กล่าวไว้แล้วข้างต้น

1. การคัดเลือกข้อมูล (Data Selection)

จำนวนมิติและระเบียนของข้อมูลมีความสำคัญอย่างมากสำหรับการวิเคราะห์ หากข้อมูลมีน้อยเกินไปหรือขาดความถูกต้องและสมบูรณ์ อาจส่งผลให้ทำนายผิดพลาดได้ ดังนั้นการคัดเลือกข้อมูลเป็นการลดมิติและระเบียนชองชุดข้อมูลให้มีความเหมาะสม การคัดเลือกข้อมูลตั้งต้นเป็นได้ทั้งแนวตั้งและแนวนอน โดยที่แนวตั้งจะเป็นการเลือกตัวแปรหรือแอตทริบิวต์ ส่วนแนวนอนจะเป็นการเลือกจำนวนระเบียนหรือเหตุการณ์ที่ได้

บันทึกไว้ งานวิจัยนี้มีข้อมูลของนิสิตจำนวน 275 คน เมื่อคัดเลือกเฉพาะข้อมูลที่มีความสมบูรณ์เพียงพอ พบว่าเหลือจำนวนนิสิตเท่ากับ 263 คน ซึ่งนับว่าลดลงจากชุดข้อมูลตั้งต้นไม่มากนัก จึงเพียงพอต่อการนำไปทำเหมืองข้อมูลในลำดับต่อไป สำหรับแอตทริบิวต์จะใช้ตัวแปรเกือบทั้งหมด ยกเว้นบางตัวแปรที่ไม่ให้ผลลัพธ์ที่ดีต่อโมเดลการทำนาย การคัดเลือกข้อมูลนอกจากจะเพิ่มประสิทธิให้กับการประเมินผลโมเดลแล้ว ยังช่วยให้การคำนวณมีความซับซ้อนน้อยลงอีกด้วย

2. การทำความสะอาดข้อมูล (Data Cleaning)

การได้มาซึ่งชุดข้อมูลจากแหล่งข้อมูลที่หลากหลายมักมีความไม่สอดคล้องกัน มีความเสียหาย และมีข้อมูลบางส่วนขาดหายไป บางครั้งอาจมีข้อมูลที่มีค่าห่างจากข้อมูลอื่นในชุดข้อมูลเดียวกันอย่างมาก หรือที่เรียกว่า ข้อมูลผิดปกติ จากงานวิจัยที่เกี่ยวข้องชี้ให้เห็นว่าความผิดปกติของข้อมูลในลักษณะดังกล่าวพบได้บ่อยในการทำเหมืองข้อมูลทางการศึกษา ดังนั้น ผู้วิจัยจริงต้องจัดการกับข้อมูลผิดปกติเหล่านี้โดยไม่ทำลายคุณภาพของโมเดลการทำนาย ซึ่งมีแนวทางในการจัดการอยู่ 2 วิธี คือ วิธีแรกจะทำการลบข้อมูลทั้งระเบียนหรือทั้งคอลัมน์ออกไปเลย ส่วนวิธีที่สองจะเป็นการแทนที่ข้อมูลด้วยค่าบางอย่าง เช่น ค่ามัธยฐาน ค่าเฉลี่ย ค่าคงที่ ค่าสุ่ม หรือค่าฐานนิยม เป็นต้น

งานวิจัยนี้จะจัดการกับปัญหาข้อมูลที่ผิดปกติหรือสูญหาย ด้วยการแทนค่าเฉลี่ย (Mean) สำหรับข้อมูลที่เป็นตัวเลข (Nominal) สำหรับข้อมูลที่ไม่ใช่ตัวเลข (Categorical) จะแทนด้วยค่าฐานนิยม (Mode) ในส่วนของเกรดแต่ละรายวิชา เมื่อนิสิตได้เกรด F มักจะลงทะเบียนเรียนใหม่ หรือหากเป็นวิชาเลือก ก็อาจเลือกไปเรียนวิชาอื่นแทน ดังนั้น กรณีได้เกรด F จะแทนด้วยเกรดใหม่ที่ได้รับ หรือถ้าไม่ได้ลงเรียนวิชาเดิมก็จะแทนด้วยเกรด D เพื่อให้ได้ชุดข้อมูลที่มีค่าของข้อมูลครบถ้วน ส่วนระเบียนที่ไม่สมบูรณ์ อาทิเช่น มีเกรด F หรือ W เป็นจำนวนมาก หรือรายวิชาแตกต่างจากผู้อื่นค่อนข้างมาก ระเบียนนั้นจะถูกตัดออกไป สำหรับกรณีศึกษาในงานวิจัยนี้ ข้อมูลดั้งเดิมที่ได้มาจากแหล่งข้อมูลมีความคลาดเคลื่อนหรือผิดปกติค่อนข้างน้อย ดังนั้น ชุดข้อมูลที่ใช้จึงมีความน่าเชื่อถือ และไม่จำเป็นต้องทำความสะอาดมากนัก

3. การประมวลผลข้อมูลล่วงหน้า (Data Preprocessing)

ขั้นตอนสุดท้ายก่อนการวิเคราะห์ข้อมูลและสร้างโมเดล คือ การประมวลผลข้อมูลล่วงหน้า ซึ่งประกอบด้วย การแปลงข้อมูล การจัดการกับชุดข้อมูลที่ไม่สมดุล และการคัดเลือกคุณลักษณะ การแปลงข้อมูลเป็นกระบวนการที่จำเป็นเพื่อกำจัดความแตกต่างในชุดข้อมูล เพื่อทำให้ข้อมูลมีความเหมาะสมมากยิ่งขึ้นสำหรับการทำเหมืองข้อมูล

สำหรับงานวิจัยนี้จะเน้นการแปลงข้อมูลแบบแบ่งกลุ่ม (Discretization) สำหรับข้อมูลที่ไม่ใช่ตัวเลข (Categorical) โดยเฉพาะข้อมูลที่เกี่ยวกับเกรดเฉลี่ยซึ่งได้แสดงไว้ในตาราง 3-3 ในส่วนของการคัดเลือกคุณลักษณะ (Feature Selection) ผู้วิจัยจะนำชุดข้อมูลที่ผ่านการแบ่งกลุ่มแล้วไปประยุกต์ใช้ในขั้นตอนของการทดสอบโมเดลในลำดับต่อไป

การแปลงข้อมูลโดยทั่วไปจะใช้เทคนิคการทำให้เป็นบรรทัดฐานของแอตทริบิวต์ที่เป็นตัวเลขเพื่อปรับขนาดของแต่ละแอตทริบิวต์ให้มีมาตราส่วนที่สอดคล้องกัน โดยแปลงข้อมูลให้อยู่ในช่วง 0 – 1 เท่านั้น สูตรการคำนวณหาค่าที่เป็นบรรทัดฐานแสดงด้วยสมการด้านล่าง

$$x_i' = \frac{x_i - min(x)}{max(x) - min(x)}$$

โดยที่ x_i' เป็นค่าที่ได้รับการปรับใหม่แล้ว, x_i คือค่าเริ่มต้น, ส่วน *min(x)* และ *max(x)* คือ ค่าต่ำสูงและสูงสุดของแอตทริบิวต์ตามลำดับ

การทำให้เป็นบรรทัดฐานจะช่วยปรับปรุงค่าความแม่นยำและประสิทธิภาพของการทำเหมืองด้วยอัลกอริทึมประเภทต่าง ๆ ให้ได้ผลลัพธ์ที่ดีกว่า (Shalabi, Shaaban & Kasasbeh, 2006) อย่างไรก็ดี สำหรับตัวแปรแบบที่ไม่ใช่ตัวเลข (Categorical Data) ให้คงไว้ซึ่งรูปแบบเดิม เนื่องจากให้ผลลัพธ์ของการสร้างโมเดลที่ดีอยู่แล้ว

อย่างไรก็ตาม ขั้นตอนการประมวลผลข้อมูลล่วงหน้าไม่ว่าจะเป็นการแปลงข้อมูลหรือการคัดเลือกคุณลักษณะ อาจนำไปสู่ผลลัพธ์ที่ไม่ดีเท่าที่ควร ดั้งนั้น ผู้วิจัยจึงจำเป็นต้องทำการทดสอบโมเดลด้วยค่าตัวแปรและพารามิเตอร์ (Parameter) ที่แตกต่างกัน แล้วค้นหาสภาพแวดล้อมที่ให้ผลลัพธ์ที่ดีที่สุดนั่นเอง

การวิเคราะห์ข้อมูลทางสถิติ (Statistical Analysis)

การวิเคาะห์ทางสถิติเบื้องต้นจะช่วยให้เข้าใจลักษณะของข้อมูลได้ดียิ่งขึ้น ค่าทางสถิติที่น่าสนใจได้แก่ ความถี่ ฐานนิยม มัธยฐาน ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าความแปรปรวน พิสัย ค่าสหสัมพันธ์ เป็นต้น ขั้นตอนนี้ช่วยให้สามารถประมวลผลข้อมูลล่วงหน้าเพื่อระบุค่าผิดปกติ กำหนดรูปแบบข้อมูลที่สูญหาย ศึกษาการกระจายตัวของตัวแปรต่าง ๆ และยังช่วยระบุความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรเป้าหมายได้อีกด้วย ทั้งหมดนี้ก็เพื่อช่วยให้การวางแผนการทำเหมืองข้อมูลเป็นไปอย่างมีประสิทธิภาพยิ่งขึ้น

การเลือกและสร้างโมเดล (Model Selection and Building)

โมเดลที่นิยมใช้ในแอปพลิเคชันการทำเหมืองข้อมูลทางการศึกษาแบ่งได้เป็น 2 ประเภทคือ การทำนายและการพรรณนา โมเดลเชิงทำนายใช้เพื่อคาดการณ์ผลลัพธ์ภายใต้การเรียนรู้แบบมีผู้สอน (Supervised Learning) ส่วนโมเดลเชิงพรรณนาใช้เพื่อสร้างรูปแบบที่อธิบายโครงสร้างของความสัมพันธ์และความเชื่อมโยงกันระหว่างข้อมูลซึ่งเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) อัลกอริทึมโมเดลเชิงทำนายที่จะนำมาใช้ในงานวิจัยนี้เป็นอัลกอริทึมที่ได้รับความนิยมอย่างมาก ประกอบไปด้วย วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (K-Nearest Neighbor: KNN) วิธีแบบเบส์อย่างง่าย (Naïve Bayes: NB) วิธี

ต้นไม้ตัดสินใจ (Decision Tree: DT) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และวิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) อัลกอริทึมดังกล่าวจัดเป็นอัลกอริทึมที่มีประสิทธิภาพในการทำนายผลลัพธ์จากข้อมูลที่มีอยู่ได้อย่างแม่นยำ จึงมีความเหมาะสมที่จะนำมาใช้กับงานวิจัยนี้เป็นอย่างยิ่ง สำหรับอัลกอริทึมโมเดลเชิงพรรณนาไม่อยู่ในขอบเขตของงานวิจัยนี้ อย่างไรก็ตามโมเดลเชิงทำนายก็เพียงพอต่อการวิเคราะห์ชุดข้อมูลตัวอย่างเพื่อวางแผนสนับสนุนนิสิตที่ต้องการการปรับปรุงผลสัมฤทธิ์ทางการศึกษาให้ดีขึ้น

กระบวนการของการเลือกโมเดลจะใช้วิธีเปรียบเทียบประสิทธิของโมเดลที่สร้างขึ้นด้วยอัลกอริทึมประเภทต่าง ๆ แล้วเลือกอัลกอริทึมที่ให้ผลลัพธ์ดีที่สุด ในขั้นตอนนี้จะใช้วิธีที่เรียกว่า การลองผิดลองถูก (Trial and Error) ซึ่งเป็นวิธีที่ง่ายและไม่ซับซ้อน อัลกอริทึมทั้งหมดจะถูกนำมาทดสอบหลายครั้งพร้อมกับปรับค่าพารามิเตอร์ (Parameter) ให้เหมาะสมจะกว่าจะได้ผลลัพธ์ที่ดีที่สุด

1. วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว (K-Nearest Neighbor: KNN)

เป็นวิธีการจำแนกประเภทของข้อมูลโดยการเปรียบเทียบข้อมูลกับข้อมูลตัวอย่างที่เก็บอยู่ในฐาน (Instance-based Learning) กล่าวคือ ข้อมูลเรียนรู้จะถูกจัดเก็บไว้เป็นตัวอย่างในฐานข้อมูล เมื่อเราต้องการจำแนกประเภทข้อมูลใหม่ อัลกอริทึมจะค้นหาตัวอย่างข้อมูลในฐานที่มีลักษณะใกล้เคียงข้อมูลใหม่มากที่สุด (เพื่อนบ้าน) จำนวน *K* ตัว มาใช้ร่วมกันตัดสินใจว่าข้อมูลใหม่นี้ควรจะเป็นประเภทอะไร หลักการที่ใช้คือ ถ้าข้อมูลใหม่ที่ต้องการทราบประเภทมีลักษณะใกล้เคียงกับข้อมูลใดในฐานข้อมูล ข้อมูลใหม่นั้น ก็ควรเป็นประเภทเดียวกันกับข้อมูลที่มีลักษณะใกล้เคียงกัน การตัดสินใจว่าข้อมูลใหม่จะเป็นประเภทใดอาจใช้วิธีที่เรียกว่า การลงมติเสียงข้างมาก (Majority Vote) กล่าวคือ ข้อมูลใหม่เป็นประเภทเดียวกับประเภทข้อมูลที่มีจำนวนมากที่สุดในข้อมูลเพื่อนบ้าน *K* ตัวนั้น การวัดความเหมือน (Similarity) ของข้อมูลใหม่และข้อมูลตัวอย่างสามารถใช้ตัววัดได้หลายตัวด้วยกัน

กรณีการวัดความเหมือนสำหรับตัวแบบเพื่อนบ้านที่ใกล้ที่สุดนิยมใช้ ระยะห่างแบบยุคลิด ซึ่งเป็นมาตรวัดระยะทางระหว่างข้อมูล โดยมองว่าข้อมูลเป็นจุดใน *n*-Dimensional space โดยที่ *n* คือ จำนวนตัวแปร หรือ Dimension ของข้อมูล การวัดนี้ใช้กับตัวแปรชนิดตัวเลข (Numerical) การคำนวณมีสูตรดังนี้

$$\text{Euclidean Distance } (X, Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$

โดยที่ X_i และ Y_i เป็นค่าของตัวแปรที่ *i* ของ *X* และ *Y* ตามลำดับ

การจำแนกประเภทข้อมูลโดยวิธี KNN จะสามารถให้ค่าความถูกต้องมากน้อยเพียงใดขึ้นอยู่กับหลายปัจจัย เช่น ความสมบูรณ์ของข้อมูลตัวอย่างที่ใช้เป็นตัวแทนของข้อมูลทั้งหมด หรือ ข้อมูลมี เสียงรบกวน (Noise) ปะปนมากน้อยเพียงใด เป็นต้น การเลือกค่า *K* ที่น้อยเกินไป เช่น *K* = 1 หรือ 2 อาจทำให้การจำแนก

ประเภทผิดพลาดได้เนื่องจากข้อมูลที่ใกล้สุดอาจเป็น Noise ก็ได้ ดังนั้นจึงควรกำหนดค่า *K* ให้เหมาะสม งานวิจัยนี้จะเลือกใช้ค่า *K* = 5

2. วิธีแบบเบส์อย่างง่าย (Naïve Bayes: NB)

วิธีแบบเบส์อย่างง่ายเป็นการจำแนกประเภทอย่างง่ายด้วยทฤษฎีความน่าจะเป็นโดยอาศัยทฤษฎีบทของเบส์ การจำแนกประเภทด้วยวิธีนี้เปรียบเสมือนการแบ่งกลุ่มที่ต้องการโดยใช้หลักของความน่าจะเป็น ซึ่งสามารถสรุปเป็นสูตรความน่าจะเป็นได้ดังนี้

$$P(E) = \frac{\text{จำนวนของเหตุการณ์ที่สนใจ}}{\text{จำนวนของทุกเหตุการณ์ที่เป็นไปได้}}$$

โดยที่ ค่าของความน่าจะเป็นมีค่าตั้งแต่ 0 ถึง 1

E = Event หรือ เหตุการณ์ที่สนใจ

0 หมายถึง ไม่สามารถเกิดเหตุการณ์ที่สนใจได้

1 หมายถึง สามารถเกิดเหตุการณ์ที่สนใจได้แน่นอน 100% และค่าที่ได้ยิ่งใกล้เลข 1 เท่าไหร่ แสดงว่า มีโอกาสเกิดเหตุการณ์ที่สนใจได้มากขึ้น

สรุปสมการของอัลกอริทึมวิธีแบบเบส์อย่างง่ายได้ดังนี้

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c)$$

โดยที่ *c* คือ ประเภท (Class)

x คือ แอตทริบิวต์ (Attribute, Variable, Feature)

P คือ ความน่าจะเป็น (Probability)

P(*c*|*x*) คือ ความน่าจะเป็นของข้อมูลที่มีแอตทริบิวต์ *x* จะเป็นคลาส *c* (Posterior Probability)

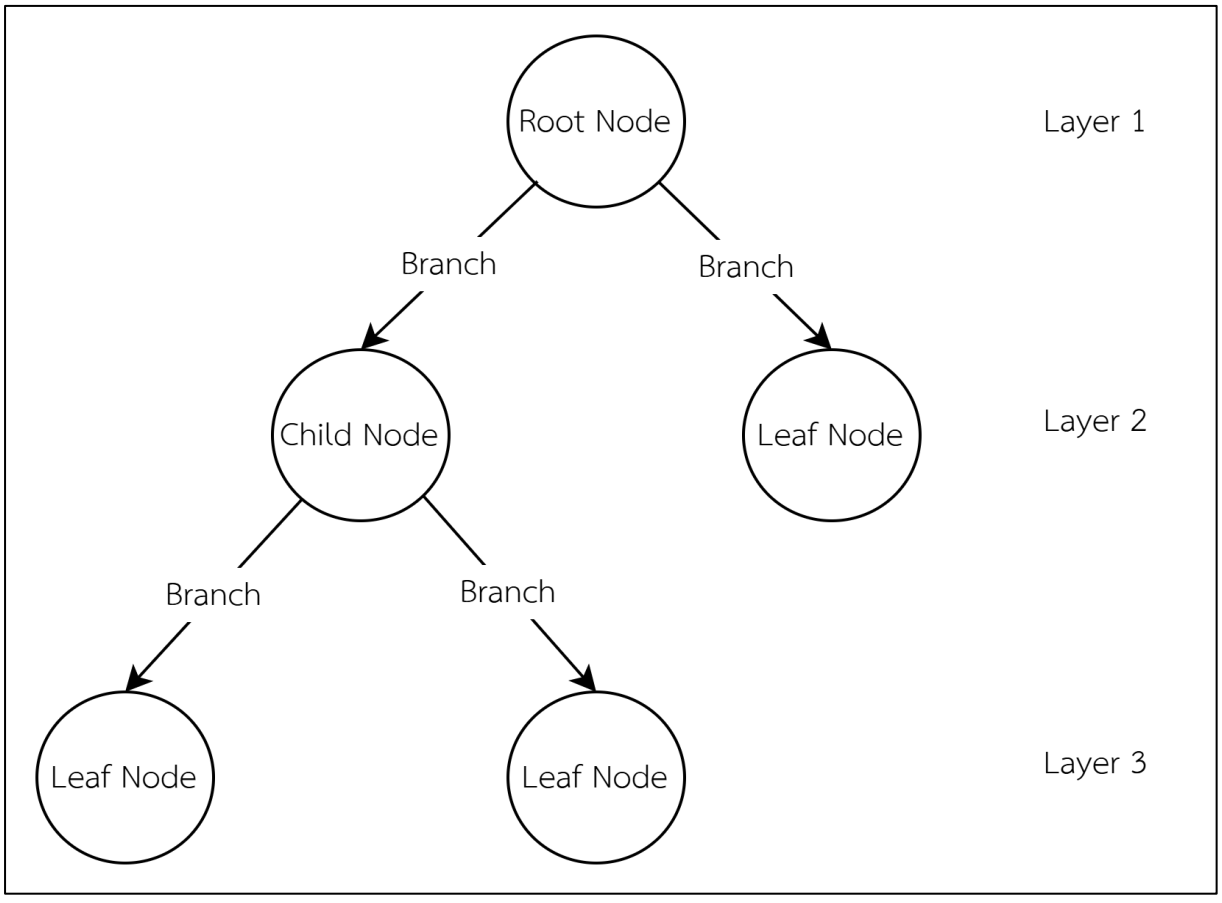
P(*x*|*c*) คือ ความน่าจะเป็นของข้อมูลที่มีแอตทริบิวต์ *x* และมีคลาส *c* (Likelihood)

P(*c*) คือ จำนวนคลาสที่อาจเกิดขึ้นต่อจำนวนคลาสทั้งหมด หรือ ความน่าจะเป็นของคลาส *c* (Class Prior Probability)

P(*x*) คือ ความน่าจะเป็นของจำนวนแอตทริบิวต์ทั้งหมดที่เป็นคลาส c (Predictor Prior Probability)

3. วิธีต้นไม้ตัดสินใจ (Decision Tree: DT)

วิธีต้นไม้ตัดสินใจเป็นเทคนิคการจำแนกประเภทตามแนวความคิดการแบ่งแยกและเอาชนะ (Divided and Conquer) หมายความว่า ในขั้นเริ่มต้นชุดข้อมูลจะถูกแบ่งออกเป็นส่วน ๆ ย่อยลงไปเรื่อย ๆ โดยพิจารณาจากค่าของตัวแปรประกอบไปด้วยการรวบรวมของโหนดการตัดสินใจ (Decision Node) มีการเชื่อมต่อกันด้วยกิ่งก้านต่าง ๆ (Branches) ขยายออกจากโหนดราก (Root Node) ไปจนถึงจุดสิ้นสุดที่โหนดใบ (Leaf node) แต่ละโหนดของต้นไม้ตัดสินใจประกอบด้วยเงื่อนไขที่ใช้ตัวแปรใดตัวแปรหนึ่งของข้อมูลในการตัดสินใจเลือกโหนดลูก (Child Node) โหนดใดโหนดหนึ่ง การตัดสินใจเริ่มจากโหนดรากของต้นไม้ และไล่ไปยังโหนดลูก จนถึงโหนดใบ ซึ่งจะบอกประเภทของข้อมูลนั้นได้ การสร้างต้นไม้ตัดสินใจจากข้อมูลเรียนรู้ต้องการได้ต้นไม้ที่ให้ความถูกต้องมากที่สุดในการจำแนกประเภทของข้อมูลเรียนรู้และข้อมูลทดสอบ รวมทั้งมีความลึกของต้นไม้ที่น้อยที่สุด เพื่อให้การตัดสินใจมีความรวดเร็ว เนื่องด้วยจำนวนครั้งที่ใช้ในการตัดสินใจจะมีน้อยครั้ง การเลือกตัวแปรและเงื่อนไขต้องคำนึงถึงการได้มาซึ่งต้นไม้ที่สามารถจำแนกประเภทของข้อมูลได้ถูกต้องมากที่สุด รวมทั้งมีความลึกของต้นไม้ที่น้อยที่สุดด้วย งานวิจัยนี้เลือกใช้อัลกอริทึม C4.5 สำหรับสร้างโมเดลต้นไม้ตัดสินใจ ภาพ 3-1 แสดงให้เห็นโครงสร้างทั่วไปของต้นไม้ตัดสินใจ



ภาพ 3-1

โครงสร้างโมเดลต้นไม้ตัดสินใจ

4. วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

ซัพพอร์ตเวกเตอร์แมชชีนเป็นอัลกอริทึมการเรียนรู้แบบมีผู้สอนที่สามารถนำมาใช้สร้างโมเดลการจำแนกประเภทของข้อมูลได้ ซัพพอร์ตเวกเตอร์แมชชีนนิยมใช้กับปัญหาที่มีข้อมูลขนาดไม่ใหญ่มากแต่มีคุณลักษณะจำนวนมาก หลักการทำงานของ ซัพพอร์ตเวกเตอร์แมชชีนจะสร้างเส้นแบ่งบนกราฟที่เรียกว่าไฮเปอร์เพลน (Hyperplane) ในการแบ่งประเภทของข้อมูลออกจากกัน จากนั้นคำนวณหาว่าไฮเปอร์เพลนใดใช้แยกประเภทได้ดีที่สุด การพิจารณาไฮเปอร์เพลนจะเลือกไฮเปอร์เพลนที่มีผลรวมของระยะห่างระหว่างเส้นไฮเปอร์เพลนกับเส้นตรงที่ลากผ่านจุดข้อมูลที่ใกล้ที่สุดและขนานกับเส้นไฮเปอร์เพลนของข้อมูลแต่ละกลุ่มมาก

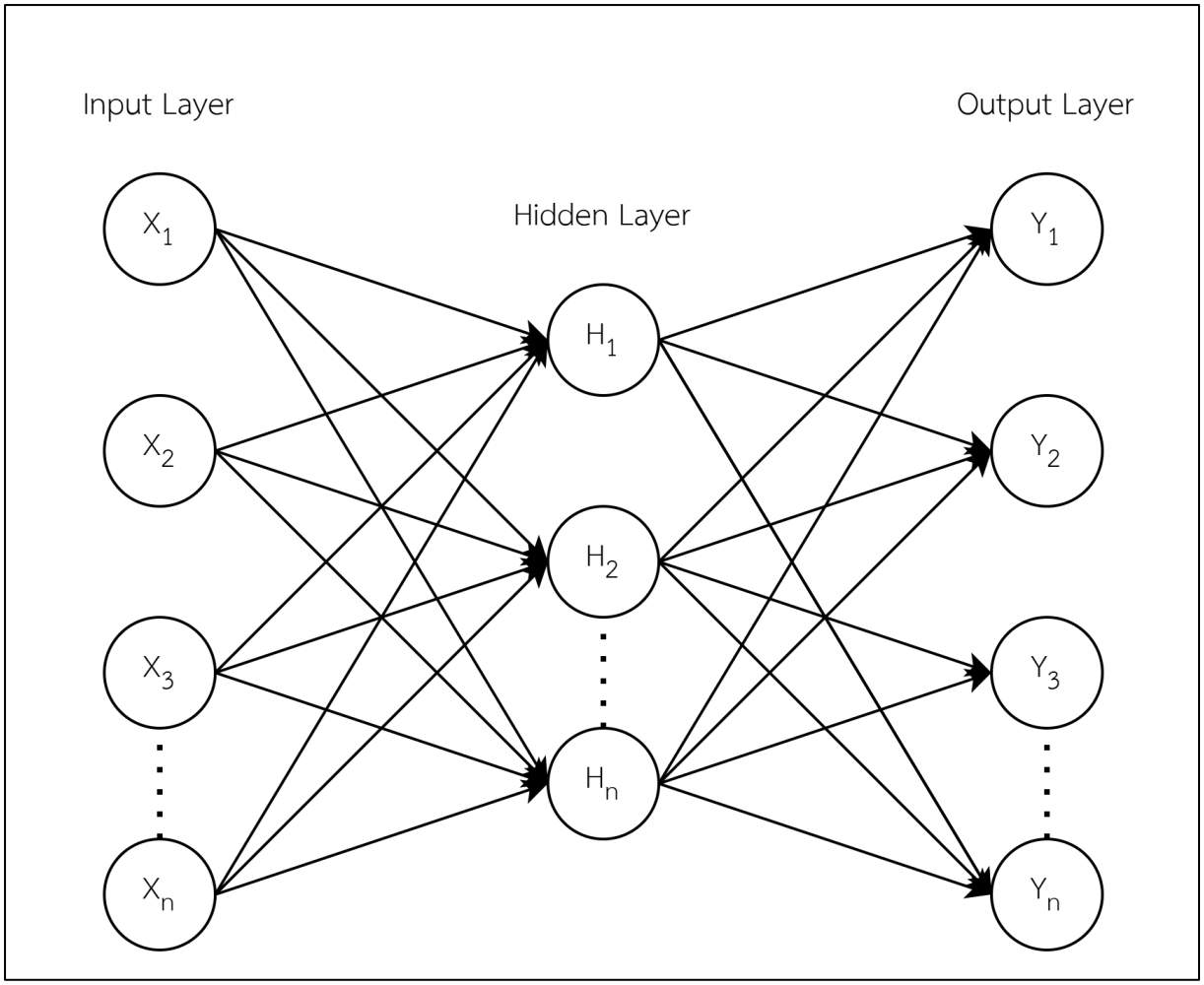
5. วิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

โครงข่ายประสาทเทียมเป็นศาสตร์แขนงหนึ่งทางด้านปัญญาประดิษฐ์ มีรูปแบบโครงสร้างและการทำงานคล้ายกับสมองของสิ่งมีชีวิต ซึ่งมีการปรับเปลี่ยนตัวเองต่อการตอบสนองของข้อมูลรับเข้าตามกฎของการเรียนรู้ โครงข่ายประสาทเทียมที่ใช้ในการจำแนกประเภทข้อมูลเป็นเครือข่ายแบบข้อมูลป้อนไปข้างหน้า (Feed Forward Network) ที่ประกอบด้วยชั้นของเซลล์ประสาทหลายชั้น ส่งผลให้เกิดเป็นโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron: MLP) เซลล์ประสาทชั้นหนึ่งจะรับค่าผลลัพธ์จากเซลล์ประสาทที่อยู่ชั้นก่อนหน้า แล้วทำการคำนวณด้วยฟังก์ชันแบบไม่เชิงเส้น (Non-linear) และให้ผลลัพธ์ที่ส่งออกไปยังเซลล์ประสาททุกเซลล์ในชั้นถัดไปผ่านเส้นเชื่อมโยง

โครงข่ายประสาทเทียมประกอบไปด้วย ชั้นรับข้อมูล (Input Layer) ชั้นปกปิด (Hidden Layer) และชั้นส่งผลลัพธ์ (Output Layer) ดังที่แสดงไว้ในภาพ 3-2 จำนวนชั้นปกปิดของโครงข่ายประสาทเทียมและจำนวนเซลล์ที่อยู่ในชั้นปกปิดถือเป็นพารามิเตอร์ที่ต้องปรับค่าเพื่อให้ได้ผลลัพธ์ในการจำแนกประเภทข้อมูลที่มีความถูกต้องมากที่สุด สำหรับจำนวนเซลล์ของชั้นรับข้อมูลและชั้นส่งผลลัพธ์จะขึ้นอยู่กับจำนวนข้อมูลนำเข้าและจำนวนค่าของผลลัพธ์ รวมถึงวิธีการแทนค่าข้อมูลนำเข้าและค่าผลลัพธ์ด้วย

การเชื่อมโยงเซลล์ประสาทระหว่างชั้นจะมีลักษณะเชื่อมโยงแบบสมบูรณ์ (Fully Connected) กล่าวคือ เซลล์ประสาทแต่ละตัวจะส่งผลลัพธ์จากการคำนวณของตัวเองไปยังเซลล์ประสาททุกตัวในชั้นถัดไป เส้นเชื่อมโยงแต่ละเส้นระหว่างเซลล์ประสาทประกอบด้วย ค่าน้ำหนักซึ่งทำหน้าที่ขยายค่าของผลลัพธ์ที่ถูกส่งผ่านเส้นเชื่อมโยงของมันโดยการคูณด้วยค่าน้ำหนักนี้ ผลคูณที่ได้จะถูกส่งเข้าสู่เซลล์ประสาทในชั้นถัดไป

งานวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบหลายชั้นโดยใช้ค่าพารามิเตอร์มาตรฐานที่อยู่ในเครื่องมือการทำเหมืองข้อมูล รายละเอียดของเครื่องมือได้อธิบายไว้ในหัวข้อถัดไป



ภาพ 3-2

โครงสร้างโมเดลโครงข่ายประสาทเทียม

เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูล (Data Analysis Tools)

เครื่องมือที่สนับสนุนในการสร้างโมเดลการทำเหมืองข้อมูลทางการศึกษาสำหรับงานวิจัยมีอยู่มากมาย เครื่องมือแบบโอเพนซอร์ส (Open source) ที่นิยมใช้ในการวิเคราะห์ชุดข้อมูลมากที่สุดคือโปรแกรม WEKA โปรแกรมนี้สามารถใช้สำหรับการประมวลผลข้อมูลล่วงหน้า การจำแนกประเภท การสร้างกฎความสัมพันธ์ การจัดกลุ่ม และการวิเคราะห์การถดถอย อีกทั้งยังเป็นมิตรกับผู้ใช้ และที่สำคัญเป็นโปรแกรมที่เข้าถึงได้ง่าย สามารถดาวน์โหลดมาใช้โดยไม่มีค่าใช้จ่าย โปรแกรม WEKA รุ่น 3.8 เป็นรุ่นล่าสุดที่มีความเสถียร ดังนั้นรุ่น 3.8 จึงได้รับความนิยมมากกว่า

นอกจากโปรแกรม WEKA แล้ว ภาษาโปรแกรมอย่างไพธอน (Python) ก็นิยมถูกนำมาใช้กับงานออกแบบและพัฒนาโมเดลการทำเหมืองข้อมูลทางการศึกษาเช่นกัน เนื่องด้วยภาษาไพธอนมีคลังโปรแกรม (Program Library) ที่สนับสนุนการวิเคราะห์ข้อมูล อาทิเช่น Pandas, NumPy, Scikit-learn, TensorFlow และ Keras ซึ่งแต่ละโปรแกรมมีพารามิเตอร์ให้เลือกใช้มากมาย

วิธีที่พบได้บ่อยที่สุดในการประเมินโมเดลการทำเหมืองข้อมูลคือ การแบ่งข้อมูลออกเป็นสองชุดข้อมูลย่อย ได้แก่ ชุดข้อมูลฝึกสอน (Training Data) และชุดข้อมูลทดสอบ (Testing Data) ชุดข้อมูลฝึกสอนใช้เพื่อสร้างโมเดลการทำเหมืองข้อมูลด้วยอัลกอริทึมชนิดต่าง ๆ ในขณะที่ชุดข้อมูลทดสอบใช้เพื่อทดสอบหรือตรวจสอบความถูกต้องของโมเดล ด้วยวิธีนี้จะช่วยให้ปัญหาที่เรียกว่า Overfitting เกิดขึ้นได้ยาก นอกจากนี้ยังมีวิธีการสร้างโมเดลในรูปแบบอื่นอีก ตัวอย่างเช่น การสุ่มตัวอย่าง (Random Sampling) หรือ การสุ่มตัวอย่างแบบแบ่งชั้น (Stratified Sampling) แต่สำหรับงานวิจัยนี้จะใช้เทคนิคที่เรียกว่า การตรวจสอบแบบไขว้ (Cross-validation: CV)

การสร้างโมเดลโดยใช้เทคนิคการตรวจสอบแบบไขว้ จะแบ่งชุดข้อมูลออกเป็นส่วน ๆ ตามค่าของตัวแปร *k* ที่กำหนด งานวิจัยนี้จะแบ่งชุดข้อมูลออกเป็น 10 ซับเซต แล้วกำหนดให้ 1 ซับเซตเป็นชุดข้อมูลทดสอบ และอีก 9 ซับเซตที่เหลือจะใช้เป็นชุดข้อมูลฝึกสอน โดยเวียนสลับกันจำนวน 10 ครั้ง การประเมินลักษณะนี้เรียกว่า 10-Fold Cross-Validation (k-Fold CV) โดยกำหนดให้ค่า $k = 10$ การกำหนดค่า *k* นี้เป็นที่นิยมในงานวิจัยด้านการทำเหมืองข้อมูลทางการศึกษา ดังนั้นการสร้างโมเดลทำเหมืองข้อมูลในงานวิจัยนี้จึงเลือกกำหนดให้ค่า $k = 10$

การประเมินโมเดล (Model Evaluation)

การประเมินโมเดลเป็นกระบวนการสำคัญที่ทำให้ทราบประสิทธิภาพของโมเดลที่ถูกพัฒนาขึ้น และใช้เปรียบเทียบโมเดลหลายตัวเพื่อค้นหาโมเดลที่ดีที่สุดสำหรับการใช้งาน ทั้งนี้ ขึ้นอยู่กับประเภทและวัตถุประสงค์ของโมเดลการทำเหมืองข้อมูลในแต่ละงานวิจัย เราสามารถใช้มาตรวัดการประเมินที่แตกต่างกันเพื่อวัดประสิทธิภาพของโมเดลได้ ตัวอย่างเช่น หากโมเดลการทำเหมืองข้อมูลของเราเป็นโมเดลการจำแนกประเภท ซึ่งกำหนดประเภทให้กับชุดข้อมูลที่มีอยู่ เราสามารถใช้มาตรวัด เช่น ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) หรือค่าเอฟ (F-Measure) เป็นต้น ตัวชี้วัดเหล่านี้จะเปรียบเทียบประเภทที่เป็นเป้าหมายจริงกับกับประเภทที่คาดการณ์ไว้ เพื่อวัดประสิทธิภาพของโมเดลการทำเหมืองข้อมูลของเราว่าสามารถจัดประเภทข้อมูลได้อย่างถูกต้องเพียงใด

อีกวิธีการหนึ่ง หากโมเดลการทำเหมืองข้อมูลเป็นโมเดลการถดถอย (Regression) ซึ่งจะคาดการณ์ค่าตัวเลขสำหรับชุดข้อมูลที่มีอยู่ เราสามารถใช้มาตรวัด เช่น ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) ค่ารากของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error) หรือ ค่าอาร์กำลังสอง (R-Squared) เป็นต้น ตัวชี้วัดเหล่านี้จะเปรียบเทียบค่าที่คาดการณ์ไว้กับค่าจริง และวัดว่าโมเดลการทำเหมืองข้อมูลของเรามีความถูกต้องมากน้อยแค่ไหน และใช้มาตรวัดเหล่านั้นในการอธิบายความแปรผันของโมเดลดังกล่าว

สำหรับงานวิจัยนี้ เราใช้ตารางเมทริกซ์ความสับสน (Confusion Matrix) เพื่อเป็นเครื่องมือในการวัดประสิทธิภาพของผลลัพธ์ที่ได้จากงานวิจัย โดยจำแนกประเภทของชุดข้อมูลจำนวน N ประเภท ในรูปแบบตาราง $N \times N$ ซึ่งโมเดลทุกตัวที่สร้างขึ้นจะใช้ตารางเมทริกซ์ความสับสนเพื่อคำนวณหาตัวชี้วัดต่าง ๆ แล้ว

นำมาเปรียบเทียบกันเพื่อวิเคราะห์หาโมเดลที่เหมาะสมที่สุด ตัวชี้วัดที่นำมาพิจารณาประกอบไปด้วย ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ตัวชี้วัดเหล่านี้ถูกนำมาใช้ร่วมกันในงานวิจัยเพื่อประเมินคุณภาพของโมเดลจากอัลกอริทึมชนิดต่าง ๆ ดังที่ได้อธิบายไว้แล้วในหัวข้อการทำเหมืองข้อมูลทางการศึกษา (Educational Data Mining: EDM) หน้าที่ 9

การจัดการกับชุดข้อมูลที่ไม่สมดุล (Imbalanced Dataset)

ในโลกความแห่งความเป็นจริง ชุดข้อมูลที่ใช้สำหรับการทำเหมืองข้อมูลอาจมีการกระจายตัวของคลาสที่ไม่สมดุลกัน นั่นคือ บางคลาสมีจำนวนของระเบียนค่อนข้างมาก ในขณะที่บางคลาสอาจมีจำนวนของระเบียนที่ค่อนข้างน้อย ทำให้แต่ละคลาสมีจำนวนระเบียนแตกต่างกันอย่างมาก ส่งผลให้การพัฒนาโมเดลจำแนกประเภทบนชุดข้อมูลที่ไม่สมดุลกันในระดับที่รุนแรงอาจไม่สะท้อนผลลัพธ์ที่ถูกต้องได้ เนื่องจากคลาสคลาสที่มีสมาชิกเป็นส่วนน้อยจะมีประสิทธิภาพในการทำนายที่ต่ำ ดังนั้น จึงเกิดเป็นแนวทางต่าง ๆ ที่จะช่วยจัดการกับชุดข้อมูลที่ไม่สมดุลเหล่านั้น

วิธีหนึ่งในการจัดการกับชุดข้อมูลที่ไม่สมดุลคือ การสุ่มตัวอย่างคลาสที่มีสมาชิกเป็นส่วนน้อยให้มีมากขึ้น โดยการทำซ้ำตัวอย่างในคลาสชนกลุ่มน้อย แม้ว่าตัวอย่างเหล่านี้จะไม่เพิ่มข้อมูลใหม่ใด ๆ ให้กับโมเดลก็ตาม แต่เราสามารถสังเคราะห์ตัวอย่างขึ้นมาใหม่จากตัวอย่างที่มีอยู่แล้วทดแทนเพิ่มเติมได้ นี่คือวิธีการเพิ่มข้อมูลที่เรียกว่า เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (Synthetic Minority Oversampling Technique: SMOTE) ซึ่งเป็นอีกหนึ่งวิธีของการสุ่มตัวอย่างข้อมูลและแก้ไขการกระจายตัวของคลาสได้เป็นอย่างดี การสังเคราะห์กลุ่มตัวอย่างเพิ่มเติมจะใช้วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว โดยที่ผู้วิจัยสามารถกำหนดร้อยละของการสุ่มตัวอย่างและจำนวนเพื่อนบ้านได้อย่างอิสระ เมื่อได้จำนวนตัวอย่างหรือระเบียนที่สมดุลกันแล้ว ชุดข้อมูลนี้จึงถูกนำไปสร้างเป็นโมเดลจำแนกประเภทให้มีประสิทธิมากขึ้นต่อไป

รูปแบบของผลลัพธ์งานวิจัย (Results)

หลังจากนำชุดข้อมูลมาผ่านการวิเคราะห์และทดลองในรูปแบบที่แตกต่างกัน ทั้งค่าพารามิเตอร์และสภาพแวดล้อมประกอบ ผลลัพธ์ที่ได้จะถูกแสดงในรูปแบบของตารางที่ระบุค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ของโมเดลประเภทต่าง ๆ โดยประกอบไปด้วยบทวิเคราะห์และการอภิปรายผลอย่างละเอียด เพื่อนำเสนอผลลัพธ์ที่ได้จากการวิจัยนี้ ตารางแสดงผลจะชี้ให้เห็นประสิทธิในการจำแนกประเภทของอัลกอริทึมที่แตกต่างกันด้วยมาตรวัดที่ระบุไว้ข้างต้นเพื่อเปรียบเทียบประสิทธิภาพของโมเดลด้วยค่าเฉลี่ยของมาตรวัด นอกจากนี้ยังวิเคราะห์ความสำคัญของคุณลักษณะต่าง ๆ ในชุดข้อมูลว่ามีผลต่อการจำแนกประเภทหรือไม่ และมากน้อยเพียงใด ด้วยตารางที่แสดงให้เห็นระดับความสำคัญของตัวแปรที่มีผลกระทบต่อผลสัมฤทธิ์ทางการศึกษาเมื่อสิ้นสุดการศึกษาแล้ว

บทที่ 4
ผลการวิจัย

การพัฒนาโมเดลการทำเหมืองข้อมูลทางการศึกษา

การทดลองสำหรับงานวิจัยนี้เริ่มจาก การนำชุดข้อมูลที่ผ่านกระบวนการจัดเตรียมข้อมูลไว้แล้ว ซึ่งมีจำนวนทั้งหมด 263 ระเบียน จากทั้งหมด 275 ระเบียน โดยแปลงจากไฟล์นามสกุล .xlsx ไปเป็นนามสกุล .csv เพื่อเตรียมพร้อมก่อนการอัปโหลดชุดข้อมูลในตาราง 4-1 เข้าสู่โปรแกรม Weka โดยกำหนดรูปแบบการสร้างโมเดลด้วยเทคนิคการตรวจสอบแบบไขว้ (k-Fold Cross Validation) ที่ซึ่ง *k* มีค่าเป็น 10 นั่นคือ 10-Fold Cross Validation ในส่วนของโมเดลที่สร้างจากอัลกอริทึมการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว (KNN) สำหรับงานวิจัยนี้ เราได้กำหนดให้ *K* มีค่าเป็น 5

ตาราง 4-1

ชุดข้อมูลที่ใช้ในการวิเคราะห์

Dataset	Description
Student Background	ชุดข้อมูลด้านประชากรของนิสิต
GE Subjects Grade	ชุดข้อมูลเกรดของรายวิชาศึกษาทั่วไป
Core Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกบังคับ
IS Core Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกบังคับด้านสารสนเทศศาสตร์
IT Core Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ
Elective Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกเลือก
IS Elective Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกเลือกด้านสารสนเทศศาสตร์
IT Elective Subjects Grade	ชุดข้อมูลเกรดของรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ
All Subjects Grade	ชุดข้อมูลเกรดของรายวิชาทั้งหมด
Early Years GPA	ชุดข้อมูลด้านเกรดเฉลี่ยในแต่ละภาคการศึกษาและเกรดเฉลี่ยรวม

ในการวิเคราะห์หาค่าสถิติเบื้องต้นเราได้พิจารณาจำนวนประเภทที่เป็นเป้าหมาย (คลาส) ทั้งหมดมี 4 ประเภท เพื่อแสดงจำนวนนิสิตในแต่ละประเภทว่ามีกี่คน โดยระบุไว้ในตาราง 4-2 ซึ่งจะเห็นได้ว่า จำนวนของนิสิตที่อยู่ในคลาส Very Good และ Good มีจำนวนมากที่สุด และมีจำนวนใกล้เคียงกัน ส่วนคลาส Excellent และ Fair มีจำนวนน้อย และน้อยที่สุดตามลำดับ ซึ่งอาจมีผลกับการวิเคราะห์อยู่บ้าง เนื่องจากจำนวนของระเบียนข้อมูลในแต่ละคลาสไม่สมดุลกัน (Imbalance Dataset) อย่างไรก็ตาม การทดลองจะยังคงใช้ชุดข้อมูลที่เป็นข้อเท็จจริงนี้โดยมองข้ามความไม่สมดุลกัน และเปรียบเทียบกับชุดข้อมูลที่สมดุลกันจากผลของ

เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (Synthetic Minority Oversampling Technique: SMOTE) เพื่อให้ได้ผลการวิจัยที่หลากหลาย

ตาราง 4-2

จำนวนนิสิตในแต่ละคลาส

Class	Amount
Excellent	35
Very Good	102
Good	100
Fair	26
Total Students	**263**

ตาราง 4-3 คือ จำนวนนิสิตในแต่ละคลาสหลังจากผ่านกระบวนการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) เป็นผลให้สมาชิกในแต่ละคลาสมีจำนวนเท่ากันคือ 100 ระเบียน สำหรับคลาส Very Good ผู้วิจัยได้ตัด 2 ระเบียนท้ายสุดออก เพื่อให้ได้จำนวนสมาชิกที่เท่ากันทุกคลาส ดังนั้น จำนวนระเบียนทั้งหมดมีค่าเท่ากับ 400 ระเบียน

ตาราง 4-3

จำนวนนิสิตในแต่ละคลาสหลังทำ SMOTE

Class	Amount
Excellent	100
Very Good	100
Good	100
Fair	100
Total Students	**400**

หลังจากนำชุดข้อมูลทั้งหมดมาผ่านการทดลองด้วยอัลกอริทึมที่แตกต่างกัน ทั้งค่าพารามิเตอร์และสภาพแวดล้อมประกอบ จึงเกิดเป็นโมเดลการทำนาย 5 โมเดล ดังที่กล่าวไว้ในบทที่ 3 ผลลัพธ์ที่ได้ถูกแสดงในรูปแบบของตารางและกราฟในหัวข้อถัดไป ทั้งนี้ ผลลัพธ์ที่ได้จะประกอบไปด้วยบทวิเคราะห์และการอภิปรายผลอย่างละเอียด ซึ่งก็คือ ผลการวิจัยนั้นเอง

การจัดลำดับความสำคัญของตัวแปร (Variables Importance)

การจัดลำดับความสำคัญของตัวแปรหรือแอตทริบิวต์ในชุดข้อมูลเป็นการคัดเลือกแอตทริบิวต์ที่ส่งผลต่อความแม่นยำในการทำนายจากมากที่สุดไปยังน้อยที่สุด การเลือกแอตทริบิวต์เป็นกระบวนการลดจำนวนตัวแปรนำเข้าสำหรับการพัฒนาโมเดลจำแนกประเภท ซึ่งก็คือ การคัดเลือกคุณลักษณะ (Feature Selection) ที่เหมาะสมมาจำนวนหนึ่งจากคุณลักษณะทั้งหมดในชุดข้อมูล วิธีการคัดเลือกคุณลักษณะนี้จะช่วยปรับปรุงประสิทธิภาพของโมเดล โดยปกติแล้ว การคัดเลือกคุณลักษณะมีความเกี่ยวข้องกับการประเมินความสัมพันธ์ระหว่างตัวแปรนำเข้าแต่ละตัวกับตัวแปรเป้าหมายเพื่อค้นหาคุณลักษณะที่มีผลกระทบต่อการสร้างโมเดล มีเทคนิคการคัดเลือกคุณลักษณะอยู่หลายวิธี ในงานวิจัยนี้ได้เลือกเทคนิคที่เรียกว่า ค่าเกนความรู้ (Information Gain: IG) ซึ่งเป็นอีกหนึ่งเทคนิคที่ได้รับความนิยมเป็นอย่างมาก

ค่าเกนความรู้เป็นการคำนวณหาค่าเอนโทรปี (Entropy) สำหรับแต่ละแอตทริบิวต์ โดยที่แต่ละแอตทริบิวต์จะมีข้อมูลที่แตกต่างกัน แอตทริบิวต์ที่มีค่าเอนโทรปีมากกว่า จัดว่าเป็นแอตทริบิวต์ที่ส่งผลต่อโมเดลมากกว่า และจะเป็นตัวเลือกแรกที่ถูกนำมาใช้สร้างโมเดลในการจำแนกประเภท กรณีที่มี 2 คลาส ค่าเอนโทรปีจะอยู่ในช่วง 0 ถึง 1 แต่งานวิจัยนี้มีทั้งหมด 4 คลาส ดังนั้น ค่าเอนโทรปีอาจมากกว่า 1 ก็เป็นได้ สำหรับโปรแกรม Weka ฟังก์ชันที่รองรับเทคนิคนี้คือ InfoGainAttributeEval ภายใต้คำสั่ง Attribute Evaluator โดยค้นหาด้วยวิธีที่เรียกว่า Ranker Search Method

เทคนิคในการคัดเลือกคุณลักษณะที่กล่าวมาข้างต้นใช้สำหรับลดจำนวนตัวแปรนำเข้าเพื่อลดต้นทุนในการคำนวณสำหรับสร้างโมเดลการทำนาย และยังช่วยปรับปรุงประสิทธิให้กับตัวโมเดลเองอีกด้วย ค่าที่ได้จะนำมาใช้จัดลำดับความสำคัญของตัวแปรแต่ละตัว โดยพิจารณาจากค่าความแม่นยำ (Accuracy) แล้วนำไปคำนวณหาค่าน้ำหนักที่เป็นบรรทัดฐาน (Normalized Weight) ตัวเลขที่ได้จะถูกบันทึกลงในรูปแบบของตารางเพื่อแสดงให้เห็นลำดับความสำคัญที่จัดเรียงไว้แล้วของแต่ละตัวแปรในแต่ละชุดข้อมูล

ผลลัพธ์จากการสร้างโมเดลด้วยข้อมูลด้านประชากรของนิสิต (Student Background)

นำข้อมูลด้านประชากรที่ประกอบไปด้วยแอตทริบิวต์ เช่น เกรดเฉลี่ยจากโรงเรียนมัธยม จังหวัดที่อยู่ของนิสิตแบ่งตามภาค เพศของนิสิต อาชีพและรายได้ของบิดามารดา เป็นต้น มาสร้างเป็นโมเดลด้วยอัลกอริทึมการจำแนกประเภทชนิดต่าง ๆ ทั้งหมด 5 อัลกอริทึม ผลลัพธ์ที่ได้นำมาเปรียบเทียบกันด้วยค่าความแม่นยำ ค่าความเที่ยงตรง ค่าการระลึก และค่าเอฟ โดยใช้ 10-Fold Cross Validation เป็นหลัก สำหรับค่าผลลัพธ์ของมาตรวัดจะพิจารณาที่ทศนิยม 3 ตำแหน่ง ที่มีค่าสูงสุดคือ 1.000 และค่าต่ำสุดคือ 0.000

1. ข้อมูลด้านประชากรของนิสิต

จากผลลัพธ์ในตาราง 4-4 จะเห็นว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลด้านประชากรมีผลกระทบที่ค่อนข้างน้อยมาก ไม่ว่าจะใช้โมเดลอะไรก็ตาม โมเดลที่ให้ค่าความแม่นยำ ค่าความเที่ยงตรง ค่าการระลึก และค่าเอฟ ดีที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) นั่นคือ 0.392, 0.387, 0.392 และ 0.388

ตามลำดับ โดยเน้นตัวเลขด้วยตัวหนา (Bold) ส่วนโมเดลอื่นก็มีผลการประเมินที่น้อยกว่าโมเดลต้นไม้ตัดสินใจเล็กน้อยลดหลั่นกันลงมา และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือโมเดลแบบเบส์อย่างง่าย (NB) ที่มีค่าการประเมินเพียง 0.335, 0.338, 0.335 และ 0.333 ตามลำดับ ซึ่งหมายความว่า ข้อมูลด้านประชากรส่งผลต่อการทำนายน้อยมาก ดังนั้น การนำข้อมูลชุดนี้มาใช้จำแนกประเภทของนิสิตเพื่อคาดการณ์ผลการเรียนเมื่อจบการศึกษาจึงไม่เหมาะสมที่จะนำมาประกอบการพิจารณาเพื่อสนับสนุนด้านการเรียนให้กับนิสิต รวมถึงนำมาใช้เป็นบรรทัดฐานในการปรับปรุงหลักสูตรอีกด้วย

ตาราง 4-4

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านประชากร

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.369	0.321	0.369	0.339
NB	0.335	0.338	0.335	0.333
DT	**0.392**	**0.387**	**0.392**	**0.388**
SVM	0.354	0.354	0.354	0.349
ANN	0.357	0.355	0.357	0.356

ตาราง 4-5

ลำดับความสำคัญของตัวแปรด้านประชากรด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	SGPA	1.00	0.201
2	Father Occupation	0.44	0.088
3	Region	0.30	0.061
4	Mother Occupation	0.30	0.060
5	Mother Income	0.17	0.034
6	Father Income	0.11	0.023
7	Gender	0.04	0.008
8	Sibling	0.01	0.003

การวิเคราะห์ผลการจัดลำดับความสำคัญของแอตทริบิวต์ด้วยชุดข้อมูลด้านประชากรโดยการคำนวณหาค่าเกนความรู้ (IG) จากตาราง 4-5 พบว่า เกรดเฉลี่ยจากโรงเรียนมัธยม (SGPA) ให้ผลลัพธ์สูงที่สุดเมื่อเทียบกับปัจจัยอื่น โดยมีค่าเกนความรู้เท่ากับ 0.201 ลำดับที่ 2 คือ อาชีพบิดา (Father Occupation) โดยมีค่าเกนความรู้เป็น 0.088 จะเห็นว่าค่าเกนความรู้ในลำดับที่ 2 ห่างจากลำดับแรกอยู่มากแบบก้าวกระโดด

เพราะฉะนั้น ความสำคัญของตัวแปรตั้งแต่ลำดับที่ 2 ลงไปมีผลต่อความแม่นยำน้อยของโมเดลน้อยมาก โดยเฉพาะสองลำดับสุดท้าย คือ เพศ (Gender) และการมีพี่น้อง (Sibling) มีค่าเกนความรู้เพียง 0.008 และ 0.003ตามลำดับ ซึ่งแทบไม่มีผลต่อการทำนายเลย ดังนั้นจึงสรุปได้ว่า ปัจจัยของข้อมูลด้านประชากรมีเพียงเกรดเฉลี่ยจากโรงเรียนมัธยมเท่านั้นที่พอจะสามารถนำมาพิจารณาว่า นิสิตจะมีผลการเรียนอยู่ในคลาสใดเมื่อสำเร็จการศึกษาแล้ว

2. ข้อมูลด้านประชากรของนิสิตโดยใช้ SMOTE

ในหัวข้อนี้เป็นการนำชุดข้อมูลด้านประชากรของนิสิตมาผ่านกระบวนการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) เพื่อให้สมาชิกของแต่ละคลาสมีจำนวนเท่ากัน จากนั้นจึงนำชุดข้อมูลนี้ไปพัฒนาต่อเป็นโมเดลการทำนายต่อไป จากผลลัพธ์ที่ได้ในตาราง 4-6 จะเห็นว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลด้านประชากรที่มีสมาชิกในแต่ละคลาสสมดุลกันแล้ว ก็ยังมีผลกระทบที่ค่อนข้างน้อยมากอยู่ดี ไม่ว่าจะใช้โมเดลอะไรก็ตาม โมเดลที่ให้ค่าความแม่นยำ ค่าความเที่ยงตรง ค่าการระลึก และค่าเอฟ ดีที่สุดคือโมเดลโครงข่ายประสาทเทียม (ANN) นั่นคือ 0.580, 0.874, 0.580 และ 0.577 ตามลำดับ ส่วนโมเดลอื่นก็มีผลการประเมินที่น้อยกว่าโมเดลโครงข่ายประสาทเทียมเล็กน้อยลดหลั่นกันลงมา และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดยังคงเป็นโมเดลแบบเบส์อย่างง่าย (NB) ที่มีค่าความแม่นยำอยู่ที่ 0.518 ถึงแม้ว่าค่าความเที่ยงตรง และค่าเอฟ จะสูงกว่าโมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว (KNN) อยู่เล็กน้อย แต่ผู้วิจัยจะให้ความสำคัญกับค่าความแม่นยำมากที่สุด ดังนั้น จึงสรุปได้ว่า ไม่ว่าจะเป็นชุดข้อมูลเดิมหรือชุดข้อมูลที่มีคลาสสมดุลกันแล้ว ข้อมูลด้านประชากรก็ยังส่งผลต่อการทำนายน้อยมากอยู่ดี ด้วยเหตุนี้ การนำข้อมูลชุดนี้มาใช้จำแนกประเภทของนิสิตเพื่อคาดการณ์ผลการเรียนเมื่อจบการศึกษาจึงไม่มีความเหมาะสมอย่างยิ่ง

ตาราง 4-6

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านประชากร (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.523	0.470	0.523	0.470
NB	0.518	0.471	0.518	0.489
DT	0.563	0.527	0.563	0.538
SVM	0.575	0.551	0.575	0.560
ANN	**0.580**	**0.574**	**0.580**	**0.577**

การวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านประชากรที่ปรับขนาดคลาสให้สมดุลกันแล้วด้วยเทคนิค SMOTE จากตาราง 4-7 พบว่า เกรดเฉลี่ยจากโรงเรียนมัธยม (SGPA) ให้ผลความแม่นยำสูงที่สุดเมื่อเทียบกับปัจจัยอื่น โดยมีค่าเกนความรู้เท่ากับ 0.448 ลำดับที่ 2 คือ อาชีพบิดา (Father

Occupation) มีค่าเกนความรู้เป็น 0.178 จะเห็นว่าค่าเกนความรู้ในลำดับที่ 2 ห่างจากลำดับแรกอยู่เยอะมาก ความสำคัญของแอตทริบิวต์ตั้งแต่ลำดับที่ 2 ลงไปมีผลต่อโมเดลน้อยมาก โดยเฉพาะตัวแปรลำดับสุดท้าย คือ เพศ (Gender) มีค่าเกนความรู้เพียง 0.019 ซึ่งแทบไม่มีผลต่อการทำนายเลย ดังนั้นจึงสรุปได้ว่า ปัจจัยของข้อมูลด้านประชากรทั้งชุดข้อมูลที่มีสมาชิกในแต่ละคลาสสมดุลกันและไม่สมดุลกัน มีเพียงเกรดเฉลี่ยจากโรงเรียนมัธยมเท่านั้นที่มีผลต่อโมเดลการทำนายอยู่บ้าง นอกเหนือจากนี้ ตัวแปรอื่น ๆ ด้านประชากรไม่มีผลกระทบทางบวกต่อโมเดลอย่างมีนัยยะสำคัญ

ตาราง 4-7

ลำดับความสำคัญของตัวแปรด้านประชากรด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	SGPA	1.00	0.448
2	Father Occupation	0.40	0.178
3	Mother Income	0.31	0.140
4	Mother Occupation	0.29	0.128
5	Region	0.19	0.083
6	Father Income	0.16	0.071
7	Sibling	0.10	0.047
8	Gender	0.04	0.019

ผลลัพธ์จากการสร้างโมเดลด้วยข้อมูลด้านเกรดของแต่ละรายวิชา (Subjects Grade)

สำหรับการสร้างโมเดลการทำนายด้วยชุดข้อมูลด้านเกรดของแต่ละรายวิชายังคงใช้แนวทางเดียวกันกับชุดข้อมูลด้านประชากร โดยนำข้อมูลของเกรดที่นิสิตได้รับในแต่ละรายวิชามาประมวลผล แอตทริบิวต์หรือตัวแปรที่ใช้ คือ รายชื่อวิชาต่าง ๆ ในหลักสูตรสารสนเทศศึกษา โดยแต่ละรายวิชาได้ระบุเกรดที่เป็นไปได้ตั้งแต่ A ถึง D สำหรับเกรด F หรือ W จะไม่นำมาใช้ในการคำนวณ ชุดข้อมูลด้านเกรดของแต่ละรายวิชาจะถูกแบ่งออกเป็นกลุ่ม ๆ ได้แก่ กลุ่มรายวิชาศึกษาทั่วไป (GE Subjects) กลุ่มรายวิชาเอกบังคับ (Core Subjects) และกลุ่มรายวิชาเอกเลือก (Elective Subjects) นอกจากนี้ กลุ่มรายวิชาเอกบังคับและเอกเลือกยังสามารถแบ่งออกเป็นรายวิชาด้านสารสนเทศศาสตร์ (IS Core และ IS Elective Subjects) และด้านเทคโนโลยีสารสนเทศ (IT Core และ IT Elective Subjects) ทั้งนี้ก็เพื่อใช้สำหรับการวิเคราะห์ชุดข้อมูลที่มีลักษณะแตกต่างกัน ในส่วนของรายวิชาศึกษาทั่วไปและรายวิชาเอกเลือก นิสิตอาจลงทะเบียนเรียนไม่เหมือนกันทั้งหมด ดังนั้น ผู้วิจัยจะอาศัยข้อมูลเกรดที่ได้รับของบางรายวิชาที่มีข้อมูลอยู่ทดแทนกับรายวิชาที่ไม่มีข้อมูลโดยที่ รายวิชาทดแทนต้องมีลักษณะใกล้เคียงกับรายวิชาที่เรียนจริง ชุดข้อมูลด้านเกรดของแต่ละรายวิชาจะใช้ข้อมูลทั้งหมดของนิสิตแต่ละคน ตั้งแต่เริ่มเรียนภาคการศึกษาแรกไปจนถึงภาคการศึกษาสุดท้ายเมื่อสำเร็จการศึกษาแล้ว งานวิจัยนี้

จะไม่นำจำนวนหน่วยกิตมาคำนวณร่วมด้วยโดยอนุมานว่าทุกรายวิชามีจำนวนหน่วยกิตเท่ากัน คือ 3 หน่วยกิต การทดลองเริ่มจากสร้างโมเดลด้วยอัลกอริทึมการจำแนกประเภททั้ง 5 อัลกอริทึม ผลลัพธ์ที่ได้นำมาเปรียบเทียบกันด้วยค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) หรือค่าเอฟ (F-Measure) โดยใช้ 10-Fold Cross Validation เป็นหลัก การพิจารณาค่าของผลลัพธ์ด้วยมาตรวัดชนิดต่าง ๆ จะเป็นไปในแนวทางเดียวกัน คือ พิจารณาที่ทศนิยม 3 ตำแหน่ง ที่มีค่าสูงสุดคือ 1.000 และค่าต่ำสุดคือ 0.000

1. รายวิชาศึกษาทั่วไป (GE Subjects Grade)

ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชาศึกษาทั่วไปแสดงไว้ในตาราง 4-8 จากการศึกษาพบว่าการทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบที่ค่อนข้างน้อย อย่างไรก็ตาม ผลลัพธ์ที่ได้ให้ค่ามาตรวัดที่สูงกว่าชุดข้อมูลด้านประชากร โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.673, 0.673, 0.673 และ 0.672 ตามลำดับ ลำดับที่ 2 คือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ที่มีค่าการประเมิน 0.631, 0.629, 0.631 และ 0.629 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดหลั่นกันลงมา และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือโมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.536, 0.533, 0.536 และ 0.531 ตามลำดับ จากผลการประเมินที่ได้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาศึกษาทั่วไปส่งผลต่อการทำนายในระดับปานกลางเท่านั้น

ตาราง 4-8

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาศึกษาทั่วไป

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.597	0.607	0.597	0.580
NB	**0.673**	**0.673**	**0.673**	**0.672**
DT	0.536	0.533	0.536	0.531
SVM	0.631	0.629	0.631	0.629
ANN	0.616	0.615	0.616	0.614

การวิเคราะห์ผลการจัดลำดับความสำคัญของรายวิชาศึกษาทั่วไปด้วยเทคนิคค่าเกนความรู้ (Information Gain: IG) ซึ่งมีแอตทริบิวต์ทั้งหมดจำนวน 12 รายวิชา จากตาราง 4-9 พบว่า รายวิชาก้าวทันสังคมดิจิทัลด้วยไอซีที (Moving Forward in a Digital Society with ICT) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น ๆ โดยมีค่าความแม่นยำที่ 0.388 ลำดับที่ 2 คือ รายวิชาทักษะการใช้ภาษาไทยเพื่อการสื่อสาร (Thai Language Skills for Communication) โดยมีค่าเกนความรู้เท่ากับ 0.367 ถัดมาคือ รายวิชาภาษาอังกฤษเพื่อการสื่อสาร (English for Communication) ที่มีค่าการประเมินเท่ากับ 0.304 รายวิชาที่

เหลือให้มีค่าเกนความรู้ต่ำกว่า 0.3 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญที่ต่ำมากและไม่ส่งผลต่อการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตมากนัก ด้วยเหตุนี้ จึงสรุปได้ว่า เกรดที่นิสิตได้รับในกลุ่มรายวิชาศึกษาทั่วไปไม่สามารถนำมาใช้เป็นปัจจัยในการวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตได้ เนื่องจากผลลัพธ์จากโมเดลการจำแนกประเภทข้อมูลและค่าเกนความรู้อยู่ในเกณฑ์ปานกลางไปจนถึงต่ำมากนั่นเอง

ตาราง 4-9

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาศึกษาทั่วไปด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Moving Forward in a Digital Society with ICT	1.00	0.388
2	Thai Language Skills for Communication	0.95	0.367
3	English for Communication	0.78	0.304
4	Natural Disasters	0.76	0.294
5	Psychology for Living and Adjustment	0.62	0.242
6	Life and Health	0.52	0.202
7	Sufficiency Economy and Social Development	0.51	0.197
8	Collegiate English	0.48	0.186
9	Contemporary Scientific Innovation	0.43	0.167
10	English Writing for Communication	0.31	0.119
11	Exercise for Quality of Life	0.21	0.080
12	Arts and Creativity	0.18	0.070

2. รายวิชาเอกบังคับ (Core Subjects Grade)

ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชาเอกบังคับแสดงไว้ในตาราง 4-10 รายวิชาในกลุ่มนี้มีจำนวนทั้งหมด 18 รายวิชา ที่ซึ่งนิสิตทุกคนต้องเรียนเหมือนกัน จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลค่อนข้างมาก ผลลัพธ์ที่ได้ให้ค่าการประเมินที่สูงกว่าชุดข้อมูลด้านประชากรและชุดข้อมูลด้านเกรดในกลุ่มรายวิชาศึกษาทั่วไป โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.814, 0.826, 0.814 และ 0.814 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.745, 0.747, 0.745 และ 0.745 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดหลั่นกันลงมา และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.593, 0.586, 0.593 และ 0.585 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรด

ของแต่ละรายวิชาในกลุ่มรายวิชาเอกบังคับส่งผลต่อการทำนายในระดับสูง ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนใด สามารถนำตัวแปรด้านเกรดในกลุ่มรายวิชาเอกบังคับมาร่วมทำการวิเคราะห์หาผลลัพธ์เป้าหมายได้อย่างถูกต้อง โดยมีค่าความแม่นยำสูงกว่า 80% ด้วยอัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลสำหรับการทำนายผล

ตาราง 4-10

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับ

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.677	0.685	0.677	0.669
NB	**0.814**	**0.826**	**0.814**	**0.814**
DT	0.593	0.586	0.593	0.585
SVM	0.688	0.691	0.688	0.687
ANN	0.745	0.747	0.745	0.745

จากตาราง 4-11 พบว่า รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) ให้ค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีผลการคำนวณเท่ากับ 0.59 ลำดับที่ 2 คือ รายวิชาการวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification) โดยมีค่าเกนความรู้เป็น 0.482 ความสำคัญของรายวิชาในลำดับที่ 3 คือ รายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) ที่มีค่าเกนความรู้เท่ากับ 0.466 ส่วนรายวิชาระบบห้องสมุดอัตโนมัติ (Library Automation Systems) และรายวิชาการวิเคราะห์และออกแบบระบบสารสนเทศ (Information Systems Analysis and Design) มีค่าเกนความรู้ที่ 0.446 และ 0.433 ตามลำดับ สำหรับรายวิชาที่เหลือให้ค่าการประเมินต่ำกว่า 0.4 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญค่อนข้างน้อยและไม่มีน้ำหนักในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตมากนัก

ตาราง 4-11

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	0.590
2	Library of Congress Classification	0.82	0.482
3	Organization of Information Resources	0.79	0.466
4	Library Automation Systems	0.76	0.446
5	Information Systems Analysis and Design	0.73	0.433
6	Electronic Information and Record Management	0.68	0.399
7	Information and Reference Services	0.61	0.359
8	Cataloging of Information Resources	0.57	0.334
9	Reading for Information Professional	0.56	0.332
10	Information Science	0.53	0.315
11	Programming in Information Work	0.52	0.309
12	Database Management for Information Work	0.49	0.287
13	Presentation and Training in Information Work	0.47	0.28
14	Information Technology	0.37	0.216
15	Web Design for Information Work	0.35	0.209
16	Collection Development	0.33	0.197
17	Management of Information Institutes	0.33	0.194
18	Seminar on Current Issues and Trend in Information Science	0.27	0.157

3. รายวิชาเอกบังคับด้านสารสนเทศศาสตร์ (IS Core Subjects Grade)

รายวิชาในกลุ่มเอกบังคับด้านสารสนเทศศาสตร์มีจำนวนทั้งหมด 9 รายวิชา โดยพิจารณาจากคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านการจัดการสารสนเทศเป็นหลัก ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชานี้แสดงไว้ในตาราง 4-12 จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลค่อนข้างมาก ซึ่งสอดคล้องกับผลการวิเคราะห์ในหัวข้อก่อนหน้านี้ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุด คือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.753, 0.755, 0.753 และ 0.752 ตามลำดับ ลำดับที่ 2 คือ โมเดล

โครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.726, 0.723, 0.726 และ 0.724 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดลงมาเล็กน้อย และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.643, 0.636, 0.643 และ 0.630 ตามลำดับ

ตาราง 4-12

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับด้านสารสนเทศศาสตร์

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.696	0.713	0.696	0.691
NB	**0.753**	**0.755**	**0.753**	**0.752**
DT	0.643	0.636	0.643	0.630
SVM	0.703	0.705	0.703	0.704
ANN	0.726	0.723	0.726	0.724

จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ยังคงส่งผลต่อการทำนายในระดับสูง ด้วยค่าความแม่นยำที่ 75.3% ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนใด ยังสามารถนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ต่อได้ โดยใช้อัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลสำหรับการทำนายผล มีข้อสังเกตว่า ผลการประเมินที่น้อยที่สุดในกลุ่มรายวิชานี้ ให้ผลลัพธ์ในภาพรวมที่สูงกว่าในกลุ่มรายวิชาเอกบังคับทั้งหมด เนื่องด้วยผลการเรียนในบางรายวิชาที่มีความสำคัญน้อยส่งผลให้ผลการประเมินในภาพรวมน้อยลงตามไปด้วยเช่นกัน

จากตาราง 4-13 พบว่า รายวิชาการวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification) มีค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่นในกลุ่มเดียวกัน โดยมีผลการวิเคราะห์เท่ากับ 0.482 ลำดับที่ 2 คือ รายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) ที่มีค่าเกนความรู้เท่ากับ 0.466 ส่วนรายวิชาระบบห้องสมุดอัตโนมัติ (Library Automation Systems) มีค่าเกนความรู้เป็นลำดับที่ 3 ซึ่งมีค่าเท่ากับ 0.446 สำหรับรายวิชาที่เหลือให้ค่าเกนความรู้ต่ำกว่า 0.4 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญน้อยเพราะเป็นปัจจัยที่ไม่มีน้ำหนักในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตเท่าใดนัก

ตาราง 4-13

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Library of Congress Classification	1.00	0.482
2	Organization of Information Resources	0.97	0.466
3	Library Automation Systems	0.93	0.446
4	Information and Reference Services	0.74	0.359
5	Cataloging of Information Resources	0.69	0.334
6	Reading for Information Professional	0.69	0.332
7	Information Science	0.65	0.315
8	Collection Development	0.41	0.197
9	Management of Information Institutes	0.40	0.194

4. รายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ (IT Core Subjects Grade)

รายวิชาในกลุ่มเอกบังคับด้านเทคโนโลยีสารสนเทศมีจำนวน 9 รายวิชา ซึ่งผู้วิจัยได้ออกแบบให้มีจำนวนรายวิชาเท่ากันกับกลุ่มเอกบังคับต้านสารสนเทศศาสตร์ เพื่อให้ผลการวิเคราะห์เกิดความสมดุลกันในแง่ของจำนวนตัวแปรของแต่ละชุดข้อมูล รายวิชาด้านเทคโนโลยีสารสนเทศมีคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านนี้โดยเฉพาะ รวมถึงรายวิชาที่มีการนำเทคโนโลยีสมัยใหม่เข้ามาประยุกต์ใช้กับงานสารสนเทศประเภทต่าง ๆ เทคโนโลยีดังกล่าวเป็นได้ทั้งหลักการ ทฤษฎี รวมไปถึงเครื่องมือทางซอฟต์แวร์ด้วย ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-14 จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลค่อนข้างมากใกล้เคียงกับกลุ่มรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดยังคงเป็น โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.745, 0.745, 0.745 และ 0.743 ตามลำดับ ลำดับที่ 2 คือ โมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว (KNN) โดยมีค่าการประเมินเท่ากับ 0.658, 0.638, 0.658 และ 0.634 ในกรณีนี้ ประสิทธิภาพของโมเดลจะพิจารณาจากค่าความแม่นยำเป็นหลัก ในขณะที่โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) เช่นเดิม ที่มีค่าการประเมินอยู่ที่ 0.593, 0.593, 0.593 และ 0.593 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศมีค่าความแม่นยำน้อยกว่าด้านสารสนเทศศาสตร์อยู่เล็กน้อย แต่ก็ยังคงส่งผลต่อการทำนายในระดับที่ค่อนข้างสูง ด้วยค่าความแม่นยำที่ 74.5% ดังนั้น การจำแนกประเภทของนิสิตยังคงสามารถนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ได้ โดยใช้อัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลสำหรับการทำนายผล

ตาราง 4-14

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.658	0.638	0.658	0.634
NB	**0.745**	**0.745**	**0.745**	**0.743**
DT	0.593	0.593	0.593	0.593
SVM	0.650	0.650	0.650	0.650
ANN	0.639	0.639	0.639	0.638

ตาราง 4-15

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	0.590
2	Information Systems Analysis and Design	0.73	0.433
3	Electronic Information and Record Management	0.68	0.399
4	Programming in Information Work	0.52	0.309
5	Database Management for Information Work	0.49	0.287
6	Presentation and Training in Information Work	0.47	0.280
7	Information Technology	0.37	0.216
8	Web Design for Information Work	0.35	0.209
9	Seminar on Current Issues and Trend in Information Science	0.27	0.157

การจัดลำดับความสำคัญของตัวแปรในตาราง 4-15 แสดงให้เห็นว่า รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) มีค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่นในกลุ่มเดียวกัน โดยมีค่าเกนความรู้เท่ากับ 0.590 ลำดับที่ 2 คือ รายวิชาการวิเคราะห์และออกแบบระบบสารสนเทศ (Information Systems Analysis and Design) ที่มีค่าเกนความรู้เท่ากับ 0.433 ซึ่งค่อนข้างห่างจากลำดับที่ 1 ส่วนรายวิชาอื่น ๆ ในกลุ่มนี้ ให้ค่าเกนความรู้ต่ำกว่า 0.4 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญที่น้อยและไม่อาจนำมาเป็นปัจจัยในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตหลังจากสำเร็จการศึกษาแล้วได้ มีอีกหนึ่งข้อสังเกตคือ ผลการประเมินหาค่าเกนความรู้ในรายวิชาทั้ง กลุ่มด้าน

สารสนเทศศาสตร์และกลุ่มด้านเทคโนโลยีสารสนเทศ ล้วนให้ผลลัพธ์เท่ากันกับในตาราง 4-11 ที่ซึ่งใช้ชุดข้อมูลในกลุ่มรายวิชาเอกบังคับทั้งหมด แต่หากเปรียบเทียบกันระหว่างสองกลุ่มนี้ กลุ่มรายวิชาด้านสารสนเทศให้ผลลัพธ์ที่ดีกว่าอยู่เพียงเล็กน้อย ซึ่งไม่มีนัยยะสำคัญมากนัก

5. รายวิชาเอกเลือก (Elective Subjects Grade)

ผลลัพธ์ที่ได้จากการประเมินในกลุ่มรายวิชาเอกเลือก โดยมีจำนวนรายวิชาอยู่ทั้งหมด 10 รายวิชา เนื่องจากเป็นรายวิชาเอกเลือก ดังนั้น นิสิตแต่ละคนจึงเรียนไม่เหมือนกันทั้งหมด ข้อมูลชุดนี้จึงเป็นการประมาณการโดยใช้เกรดจากรายวิชาที่ใกล้เคียงกัน เช่น รายวิชาสารสนเทศทางมนุษยศาสตร์และสังคมศาสตร์ (Information in Humanities and Social Sciences) แทนด้วยรายวิชาสารสนเทศทางวิทยาศาสตร์และเทคโนโลยี (Information in Science and Technology) หรือ รายวิชาธุรกิจสารสนเทศ (Information Business) แทนด้วย รายวิชาการพิมพ์และธุรกิจการพิมพ์ (Printing and Publishing Business) เป็นต้น จากตาราง 4-16 พบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้ให้ค่าการประเมินอยู่ในเกณฑ์ดี โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) สูงที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.745, 0.745, 0.745 และ 0.744 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.681, 0.680, 0.681 และ 0.679 ส่วนโมเดลอื่นยังคงให้ผลการประเมินที่ใกล้เคียงกัน และโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) โดยมีผลการประเมินเท่ากับ 0.567, 0.560, 0.567 และ 0.562 ตามลำดับ จากผลการประเมินสามารถสรุปได้ว่า ชุดข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกส่งผลต่อการทำนายผลสัมฤทธิ์ทางการศึกษาในระดับที่ดี แต่ยังน้อยกว่าชุดข้อมูลกลุ่มรายวิชาเอกบังคับ อย่างไรก็ตาม ข้อมูลชุดนี้ยังสามารถนำมาใช้ประโยชน์ได้โดยมีโมเดลแบบเบส์อย่างง่ายที่ให้ค่าความแม่นยำ 74.5% ซึ่งเทียบเท่ากับผลลัพธ์ของกลุ่มรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ

ตาราง 4-16

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกเลือก

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.620	0.641	0.620	0.603
NB	**0.745**	**0.745**	**0.745**	**0.744**
DT	0.567	0.560	0.567	0.562
SVM	0.658	0.655	0.655	0.656
ANN	0.681	0.680	0.681	0.679

การวิเคราะห์ผลการจัดลำดับความสำคัญของรายวิชาเอกเลือกด้วยค่าเกนความรู้ (IG) แสงดไว้ในตาราง 4-17 ซึ่งสามารถสรุปได้ว่า รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าเกนความรู้เท่ากับ 0.474 ลำดับที่ 2 คือ รายวิชาการประยุกต์โปรแกรมกลุ่มโอเพนซอร์ส (Open Source Program Application) โดยมีค่าเกนความรู้เท่ากับ 0.396 ลำดับที่ 3 คือ รายวิชาการพิมพ์และธุรกิจการพิมพ์ (Printing and Publishing Business) ที่มีค่าการประเมินเป็น 0.383 ส่วนรายวิชาสารสนเทศทางวิทยาศาสตร์และเทคโนโลยี (Information in Science and Technology) และรายวิชาการค้นคืนสารสนเทศอิเล็กทรอนิกส์ (Electronic Information Retrieval) มีค่าการประเมินเท่ากับ 0.356 และ 0.346 ตามลำดับ รายวิชาที่เหลือให้ค่าการประเมินที่น้อยลงไปอีก ดังนั้นลำดับความสำคัญจึงค่อนข้างน้อยและไม่ส่งผลกระทบต่อโมเดลการทำนายผลสัมฤทธิ์ทางการศึกษามากนัก

ตาราง 4-17

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Data Communication and Computer Networks	1.00	0.474
2	Open Source Program Application	0.84	0.396
3	Printing and Publishing Business	0.81	0.383
4	Information in Science and Technology	0.75	0.356
5	Electronic Information Retrieval	0.73	0.346
6	User Behavior and Information Needs	0.72	0.339
7	English for Information Professional	0.61	0.287
8	Multimedia Technology	0.55	0.262
9	Information of Local Wisdom	0.47	0.225
10	Knowledge Management	0.38	0.179

6. รายวิชาเอกเลือกด้านสารสนเทศศาสตร์ (IS Elective Subjects Grade)

รายวิชาในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์แบ่งออกมาจากกลุ่มรายวิชาเอกเลือกทั้งหมดได้ 5 รายวิชา โดยพิจารณาจากคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านการจัดการสารสนเทศเป็นหลัก ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชานี้แสดงไว้ในตาราง 4-18 จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลอยู่ในเกณฑ์ปานกลาง โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดล

ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ซึ่งมีค่าการประเมินตามลำดับ ดังนี้ 0.643, 0.641, 0.643 และ 0.632 ผลการวิเคราะห์ในลำดับที่ 2 คือ โมเดลแบบเบส์อย่างง่าย (NB) ที่มีค่าการประเมินเท่ากับ 0.612, 0.608, 0.612 และ 0.605 ในทางกลับกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.498, 0.480, 0.498 และ 0.486 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ส่งผลต่อการทำนายในระดับปานกลาง ด้วยค่าความแม่นยำสูงสุดที่ 64.3% ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนใด ไม่ควรนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ด้วย อีกทั้งเกรดในกลุ่มรายวิชาเอกเลือกมักเกิดขึ้นในภาคการศึกษาท้าย ๆ นั่นคือ ก่อนนิสิตจะสำเร็จการศึกษาไม่นานนัก ดังนั้น การนำตัวแปรในกลุ่มนี้มาทำนายผลสัมฤทธิ์ทางการเรียน จึงไม่เหมาะสมอย่างยิ่ง

ตาราง 4-18

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกเลือกด้านสารสนเทศศาสตร์

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.551	0.598	0.551	0.526
NB	0.612	0.608	0.612	0.605
DT	0.498	0.480	0.498	0.486
SVM	**0.643**	**0.641**	**0.643**	**0.632**
ANN	0.559	0.554	0.559	0.555

ตาราง 4-19

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Printing and Publishing Business	1.00	0.383
2	User Behavior and Information Needs	0.72	0.339
3	English for Information Professional	0.61	0.287
4	Information of Local Wisdom	0.47	0.225
5	Knowledge Management	0.38	0.179

จากตาราง 4-19 พบว่า รายวิชาการพิมพ์และธุรกิจการพิมพ์ (Printing and Publishing Business) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่นในกลุ่มเดียวกัน คือ 0.383 ลำดับที่ 2 คือ รายวิชาพฤติกรรมผู้ใช้และความต้องการสารสนเทศ (User Behavior and Information Needs) มีค่าเกนความรู้เท่ากับ 0.339

สำหรับรายวิชาที่เหลือให้ค่าเกนความรู้ต่ำกว่า 0.3 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญที่น้อยมากและเป็นปัจจัยที่ไม่มีน้ำหนักในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิต

7. รายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ (IT Elective Subjects Grade)

รายวิชาในกลุ่มเอกเลือกด้านเทคโนโลยีสารสนเทศมีจำนวน 5 รายวิชา เท่ากับจำนวนรายวิชาในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์ ซึ่งผู้วิจัยได้ออกแบบให้มีจำนวนรายวิชาเท่ากันเพื่อให้ผลการวิเคราะห์เกิดความสมดุลกันในแง่ของจำนวนตัวแปรในแต่ละชุดข้อมูล รายวิชาด้านเทคโนโลยีสารสนเทศมีคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านนี้โดยเฉพาะ รวมถึงรายวิชาที่มีการนำเทคโนโลยีสมัยใหม่เข้ามาประยุกต์ใช้กับงานสารสนเทศประเภทต่าง ๆ เทคโนโลยีดังกล่าวเป็นได้ทั้งหลักการ ทฤษฎี รวมไปถึงเครื่องมือทางซอฟต์แวร์ด้วย ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-20 จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลสูงกว่าปานกลางเล็กน้อย แต่ดีกว่าผลลัพธ์จากกลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุด ได้แก่ โมเดลแบบเบส์อย่างง่าย (NB) ด้วยค่าการประเมินเท่ากับ 0.684, 0.686, 0.684 และ 0.681 ตามลำดับ ลำดับที่ 2 คือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีค่าการประเมินเท่ากับ 0.658, 0.638, 0.658 และ 0.634 ในกรณีนี้ ประสิทธิภาพของโมเดลจะพิจารณาจากค่าความแม่นยำเป็นหลัก ในขณะที่โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) เช่นเดิม ที่มีค่าการประเมินอยู่ที่ 0.650, 0.653, 0.650 และ 0.649 ตามลำดับ

ตาราง 4-20

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.601	0.644	0.601	0.579
NB	**0.684**	**0.686**	**0.684**	**0.681**
DT	0.593	0.589	0.593	0.589
SVM	0.650	0.653	0.650	0.649
ANN	0.639	0.642	0.639	0.639

จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศมีค่าความแม่นยำมากกว่าด้านสารสนเทศศาสตร์อยู่เล็กน้อย แต่ก็ยังคงส่งผลต่อการทำนายในระดับปานกลาง ด้วยค่าความแม่นยำเพียง 68.4% ดังนั้น การจำแนกประเภทของนิสิตจึงไม่ควรนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ด้วย อีกทั้งรายวิชาเอกเลือกทั้งหมดมักอยู่ในช่วงการเรียนของนิสิตปี 3 และปี 4 ซึ่งใกล้จะสำเร็จการศึกษาแล้ว เพราะฉะนั้น จึงไม่อาจนำผลลัพธ์จากรายวิชากลุ่ม

เอกเลือกทั้งด้านสารสนเทศศาสตร์และด้านเทคโนโลยีสารสนเทศมาเป็นปัจจัยการทำนายผลสัมฤทธิ์ทางการศึกษาได้

การวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศแสดงไว้ในตาราง 4-21 จากผลลัพธ์ที่ได้แสดงให้เห็นว่า รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าเกนความรู้เท่ากับ 0.474 ลำดับที่ 2 คือ รายวิชาการประยุกต์โปรแกรมกลุ่มโอเพนซอร์ส (Open Source Program Application) โดยมีค่าเกนความรู้เท่ากับ 0.396 สำหรับรายวิชาที่เหลือให้ค่าเกนความรู้ที่น้อยลงตามลำดับ ถึงแม้ว่า จะให้ผลลัพธ์ที่ดีกว่ากลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ แต่ก็ถือได้ว่ามีลำดับความสำคัญที่น้อยโดยมีน้ำหนักในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตค่อนข้างน้อยทีเดียว

ตาราง 4-21

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Data Communication and Computer Networks	1.00	0.474
2	Open Source Program Application	0.84	0.396
3	Information in Science and Technology	0.75	0.356
4	Electronic Information Retrieval	0.73	0.346
5	Multimedia Technology	0.55	0.262

8. รายวิชาทั้งหมด (All Subjects Grade)

หัวข้อนี้ได้นำรายวิชาทั้งหมดซึ่งประกอบไปด้วย 40 รายวิชา มาสร้างโมเดลการทำนายเพื่อประเมินสิทธิภาพ ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-22 การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลมากที่สุด เนื่องจากเป็นข้อมูลเกรดจากรายวิชาทั้งหมด ยกเว้นเพียงรายวิชาฝึกงานเท่านั้น ดังนั้น ผลลัพธ์ที่ได้จึงสูงกว่าทุกชุดข้อมูลที่ผ่านมาทั้งหมด สำหรับโมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) สูงที่สุดได้แก่ โมเดลแบบเบส์อย่างง่าย (NB) โดยมีค่าการประเมินเท่ากับ 0.844, 0.854, 0.844 และ 0.845 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.806, 0.807, 0.806 และ 0.804 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดหลั่นกันลงมาแต่ก็ยังอยู่ในเกณฑ์ดี ยกเว้นโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.616, 0.618, 0.616 และ 0.615 ตามลำดับ ซึ่งจัดว่าอยู่ในเกณฑ์ปานกลางเท่านั้น จากผลการประเมินนี้สรุปได้ว่า หากใช้ชุดข้อมูลด้านเกรดของ

รายวิชาทั้งหมดจะมีความแม่นยำของการทำนายในระดับสูง ที่ซึ่งมีถึง 2 โมเดลที่มีค่าความแม่นยำมากกว่า 80% โดยที่โมเดลแบบเบส์อย่างง่ายให้ค่าความแม่นยำสูงถึง 84.4% ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนกลุ่มใด สามารถนำตัวแปรด้านเกรดของรายวิชาทั้งหมดมาร่วมวิเคราะห์หาผลลัพธ์เป้าหมายได้อย่างถูกต้อง อย่างไรก็ตาม การที่จะทราบเกรดของรายวิชาทั้งหมด นิสิตต้องผ่านเส้นทางของการศึกษาในหลักสูตรไปแล้วถึงสามปีครึ่ง ด้วยเหตุนี้ ในทางปฏิบัติจึงไม่อาจใช้ประโยชน์ชุดข้อมูลนี้ได้ แต่อาจใช้เพื่อเปรียบเทียบผลการประเมินสูงสุดที่เป็นไปได้เท่านั้น

ตาราง 4-22

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาทั้งหมด

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.738	0.770	0.738	0.740
NB	**0.844**	**0.854**	**0.844**	**0.845**
DT	0.616	0.618	0.616	0.615
SVM	0.768	0.769	0.768	0.768
ANN	0.806	0.807	0.806	0.804

ในส่วนการวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดของรายวิชาทั้งหมด ผลลัพธ์ที่ได้สอดคล้องตรงกันกับชุดข้อมูลกลุ่มต่าง ๆ ก่อนหน้านี้ นั่นคือ รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) ให้ค่าความเกนความรู้สูงที่สุด โดยมีผลการคำนวณเท่ากับ 0.59 ลำดับที่ 2 คือ รายวิชาการวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification) โดยมีค่าเกนความรู้เป็น 0.482 ซึ่งทั้ง 2 รายวิชานี้อยู่กลุ่มรายวิชาเอกบังคับ ส่วนรายวิชาในกลุ่มเอกเลือกที่มีความสำคัญสูงสุดจัดอยู่ในลำดับที่ 3 นั่นคือ รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ที่มีค่าเกนความรู้เท่ากับ 0.474 ดังแสดงไว้ในตาราง 4-23 จะเห็นได้ว่า รายวิชาข้างต้นเป็นรายวิชาที่อยู่ตารางเรียนของนิสิตปี 2 และ 3 ดังนั้น จึงสามารถนำมาใช้เป็นปัจจัยในการวิเคราะห์หาความสัมพันธ์ระหว่างเกรดของรายวิชากับผลสัมฤทธิ์ทางการเรียนได้ เพื่อให้อาจารย์หรือกรรมการบริหารหลักสูตรสามารถใช้เป็นข้อมูลสำหรับการพัฒนานิสิตในกลุ่มเป้าหมายต่าง ๆ ให้มีผลการเรียนที่ดีขึ้นเมื่อนิสิตสำเร็จการศึกษาแล้ว

ตาราง 4-23

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาทั้งหมดด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	0.590
2	Library of Congress Classification	0.82	0.482
3	Data Communication and Computer Networks	0.80	0.474
4	Organization of Information Resources	0.79	0.466
5	Library Automation Systems	0.76	0.446
6	Information Systems Analysis and Design	0.73	0.433
7	Electronic Information and Record Management	0.68	0.399
8	Open Source Program Application	0.67	0.396
9	Moving Forward in a Digital Society with ICT	0.66	0.388
10	Printing and Publishing Business	0.65	0.384
11	Thai Language Skills for Communication	0.62	0.367
12	Information and Reference Services	0.61	0.359
13	Information in Science and Technology	0.60	0.356
14	Electronic Information Retrieval	0.59	0.346
15	User Behavior and Information Needs	0.57	0.339
16	Cataloging of Information Resources	0.57	0.334
17	Reading for Information Professional	0.56	0.332
18	Information Science	0.53	0.315
19	Programming in Information Work	0.52	0.309
20	English for Communication	0.52	0.304
21	Natural Disasters	0.50	0.294
22	Database Management for Information Work	0.49	0.287
23	English for Information Professional	0.49	0.287
24	Presentation and Training in Information Work	0.47	0.280
25	Multimedia Technology	0.44	0.262
26	Psychology for Living and Adjustment	0.41	0.242

Order	Variables	Normalized Weight	Information Gain
27	Information of Local Wisdom	0.38	0.225
28	Information Technology	0.37	0.216
29	Web Design for Information Work	0.35	0.209
30	Life and Health	0.34	0.202
31	Collection Development	0.34	0.198
32	Sufficiency Economy and Social Development	0.33	0.197
33	Management of Information Institutes	0.33	0.194
34	Collegiate English	0.32	0.186
35	Knowledge Management	0.3	0.179
36	Contemporary Scientific Innovation	0.28	0.167
37	Seminar on Current Issues and Trend in Information Science	0.27	0.157
38	English Writing for Communication	0.20	0.119
39	Exercise for Quality of Life	0.14	0.800
40	Arts and Creativity	0.12	0.700

ตาราง 4-24 เป็นการเปรียบเทียบชุดข้อมูลจากกลุ่มรายวิชาที่แตกต่างกันโดยอาศัยมาตรวัดที่เป็นค่าความแม่นยำ (Accuracy) เท่านั้น กลุ่มรายวิชาเอกบังคับมีผลการประเมินสูงที่สุดเมื่อเทียบกับกลุ่มรายวิชาศึกษาทั่วไปและกลุ่มรายวิชาเอกเลือก โดยมีค่าความแม่นยำคิดเป็น 81.4% รองลงมาคือกลุ่มรายวิชาเอกเลือกที่มีค่าความแม่นยำคิดเป็น 74.5% ส่วนกลุ่มรายวิชาศึกษาทั่วไปให้การประเมินที่น้อยที่สุดคือ 67.3% ซึ่งทั้งหมดเป็นการประเมินจากโมเดลแบบเบส์อย่างง่าย (NB) เมื่อแบ่งกลุ่มรายวิชาเอกบังคับและเอกเลือกเป็นกลุ่มรายวิชาด้านสารสนเทศศาสตร์และกลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศ พบว่า ชุดข้อมูลจากทั้งสองกลุ่มรายวิชายังคงให้ผลลัพธ์ที่ดีสำหรับกลุ่มรายวิชาเอกบังคับ โดยมีค่าความแม่นยำคิดเป็น 75.3% และ 74.5% ตามลำดับ ส่วนกลุ่มรายวิชาเอกเลือก ทั้งสองชุดข้อมูลให้ผลการประเมินที่ใกล้เคียงกับชุดข้อมูลจากกลุ่มรายวิชาศึกษาทั่วไป ซึ่งอยู่ในเกณฑ์ที่ต่ำกว่า 70%

เมื่อเปรียบเทียบผลลัพธ์ระหว่างกลุ่มรายวิชาด้านสารสนเทศศาสตร์และกลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศ พบว่า กลุ่มรายวิชาด้านสารสนเทศศาสตร์มีค่าความแม่นยำที่ดีกว่าในกลุ่มรายวิชาเอกบังคับ ในทางกลับกัน กลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศมีผลการประเมินที่สูงกว่าในกลุ่มรายวิชาเอกเลือก อย่างไรก็ตาม งานวิจัยนี้ต้องการค้นหาความสัมพันธ์ระหว่างข้อมูลก่อนที่นิสิตจะสำเร็จการศึกษากับผลสัมฤทธิ์

ทางการศึกษาของนิสิตคนนั้นในระยะต้นและระยะกลางของหลักสูตร ดังนั้น ชุดข้อมูลด้านเกรดของรายวิชาเอกบังคับจึงเป็นปัจจัยสำคัญที่จะนำมาวิเคราะห์และสร้างเป็นโมเดลการทำนายที่เหมาะสมที่สุด

ตาราง 4-24

การเปรียบเทียบผลลัพธ์จากกลุ่มรายวิชาทั้งหมดด้วยค่าความแม่นยำ

Model	Accuracy (%)							
	GE	Core	IS-Core	IT-Core	Elec.	IS-Elec.	IT-Elec.	All
KNN	0.597	0.677	0.696	0.658	0.620	0.551	0.601	0.738
NB	0.673	0.814	0.753	0.745	0.745	0.612	0.684	0.844
DT	0.536	0.593	0.643	0.593	0.567	0.498	0.593	0.616
SVM	0.631	0.688	0.703	0.650	0.658	0.643	0.650	0.768
ANN	0.616	0.745	0.726	0.639	0.681	0.559	0.639	0.806

ผลลัพธ์จากการสร้างโมเดลด้วยข้อมูลด้านเกรดของแต่ละรายวิชาโดยใช้ SMOTE

สำหรับการสร้างโมเดลการทำนายด้วยชุดข้อมูลด้านเกรดของแต่ละรายวิชาโดยใช้เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) ซึ่งยังคงมีกระบวนการเป็นไปในแนวทางเดียวกันกับชุดข้อมูลก่อนหน้านี้ โดยนำข้อมูลของเกรดที่นิสิตได้รับในแต่ละรายวิชามาประมวลผล ในส่วนของแอตทริบิวต์ก็คือ รายชื่อวิชาต่าง ๆ ในหลักสูตรสารสนเทศศึกษา โดยแต่ละรายวิชาได้ระบุเกรดที่เป็นไปได้ตั้งแต่ A ถึง D สำหรับเกรด F หรือ W จะไม่นำมาใช้ในการคำนวณ ชุดข้อมูลด้านเกรดของแต่ละรายวิชาจะถูกแบ่งออกเป็นกลุ่ม ๆ ได้แก่ กลุ่มรายวิชาศึกษาทั่วไป กลุ่มรายวิชาเอกบังคับ และกลุ่มรายวิชาเอกเลือก นอกจากนี้ กลุ่มรายวิชาเอกบังคับและเอกเลือกยังสามารถแบ่งออกเป็นรายวิชาด้านสารสนเทศศาสตร์และด้านเทคโนโลยีสารสนเทศ ทั้งนี้ก็เพื่อใช้สำหรับการวิเคราะห์ชุดข้อมูลที่มีลักษณะแตกต่างกัน ชุดข้อมูลด้านเกรดของแต่ละรายวิชาจะใช้ข้อมูลทั้งหมดของนิสิตแต่ละคน ตั้งแต่เริ่มเรียนภาคการศึกษาแรกไปจนถึงภาคการศึกษาสุดท้ายเมื่อสำเร็จการศึกษาแล้ว งานวิจัยนี้ทดลองด้วยการสร้างโมเดลจากอัลกอริทึมการจำแนกประเภท 5 อัลกอริทึมด้วยกัน ผลลัพธ์ที่ได้นำมาเปรียบเทียบกันด้วยค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) หรือค่าเอฟ (F-Measure) โดยใช้ 10-Fold Cross Validation การพิจารณาค่าของผลลัพธ์ด้วยมาตรวัดชนิดต่าง ๆ จะเป็นไปในแนวทางเดียวกัน คือ พิจารณาที่ทศนิยม 3 ตำแหน่ง ที่มีค่าสูงสุดคือ 1.000 และค่าต่ำสุดคือ 0.000

1. รายวิชาศึกษาทั่วไป (GE Subjects Grade) โดยใช้ SMOTE

ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชาศึกษาทั่วไปอยู่ในตาราง 4-25 จากการศึกษาพบว่าการทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบอยู่ในระดับดี โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีค่าการประเมินเท่ากับ 0.773, 0.767, 0.773 และ 0.768 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเป็น 0.765, 0.763, 0.765 และ 0.764 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดหลั่นกันลงมา และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด *K* ตัว (KNN) ที่มีค่าการประเมินอยู่ที่ 0.68, 0.677, 0.68 และ 0.648 ตามลำดับ ทั้งนี้การจัดอันดับจะพิจารณาจากค่าความแม่นยำ (Accuracy) เป็นตัวตั้ง จากผลการประเมินที่ได้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาศึกษาทั่วไปส่งผลต่อการทำนายในระดับที่ค่อนข้างดีทีเดียว

ตาราง 4-25

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาศึกษาทั่วไป (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.680	0.677	0.680	0.648
NB	0.758	0.751	0.758	0.753
DT	0.690	0.672	0.690	0.678
SVM	**0.773**	**0.767**	**0.773**	**0.768**
ANN	0.765	0.763	0.765	0.764

การวิเคราะห์ผลการจัดลำดับความสำคัญของรายวิชาศึกษาทั่วไปด้วยเทคนิค Information Gain (IG) โดยใช้ชุดข้อมูลที่ผ่านการปรับสมดุลของคลาสเป้าหมายแล้ว ซึ่งมีแอตทริบิวต์ทั้งหมดจำนวน 12 รายวิชา จากตาราง 4-26 รายวิชาก้าวทันสังคมดิจิทัลด้วยไอซีที (Moving Forward in a Digital Society with ICT) ให้ค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าความเกนความรู้เท่ากับ 0.748 ซึ่งสูงกว่าลำดับที่ 2 อยู่มากทีเดียว ลำดับที่ 2 คือ รายวิชาทักษะการใช้ภาษาไทยเพื่อการสื่อสาร (Thai Language Skills for Communication) โดยมีค่าเกนความรู้เป็น 0.554 ถัดมาคือ รายวิชาภาษาอังกฤษเพื่อการสื่อสาร (English for Communication) ที่มีค่าการประเมินเท่ากับ 0.501 ตามด้วยรายวิชาภัยพิบัติทางธรรมชาติ (Natural Disasters) ด้วยค่าการประเมินเท่ากับ 0.5 รายวิชาที่เหลือให้ค่าเกนความรู้ต่ำกว่า 0.5 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญน้อยและส่งผลต่อการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตอยู่ในเกณฑ์ปานกลางถึงต่ำ

เมื่อนำผลที่ได้จากทั้ง 5 โมเดล ของชุดข้อมูลรายวิชาศึกษาทั่วไปทั้งก่อนและหลังการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) พบว่า เมื่อชุดข้อมูลมีความสมดุลมากขึ้นก็ส่งผลให้ผลลัพธ์ของการ

ทำนายดีขึ้น โดยมีค่าความแม่นยำเท่ากับ 77.3% จากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (SVM) นอกจากนี้ สำหรับโมเดลการทำนายอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมด ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น ซึ่งแสดงไว้ในตารางที่ 4-27

ตาราง 4-26

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาศึกษาทั่วไปด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Moving Forward in a Digital Society with ICT	1.00	0.748
2	Thai Language Skills for Communication	0.74	0.554
3	English for Communication	0.67	0.501
4	Natural Disasters	0.67	0.500
5	Psychology for Living and Adjustment	0.60	0.452
6	Life and Health	0.59	0.439
7	Sufficiency Economy and Social Development	0.56	0.419
8	Collegiate English	0.54	0.405
9	Contemporary Scientific Innovation	0.45	0.337
10	English Writing for Communication	0.34	0.258
11	Exercise for Quality of Life	0.32	0.236
12	Arts and Creativity	0.15	0.114

ตาราง 4-27

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาศึกษาทั่วไประหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.597	**0.680**
NB	0.673	**0.758**
DT	0.536	**0.690**
SVM	0.631	**0.773**
ANN	0.616	**0.765**

2. รายวิชาเอกบังคับ (Core Subjects Grade) โดยใช้ SMOTE

ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชาเอกบังคับโดยใช้เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) แสดงไว้ในตาราง 4-28 รายวิชาในกลุ่มนี้มีจำนวนทั้งหมด 18 รายวิชา ที่ซึ่งนิสิตทุกคนต้องเรียนเหมือนกัน จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลเป็นอย่างมาก ผลลัพธ์ที่ได้ให้ค่าการประเมินที่สูงกว่าชุดข้อมูลด้านประชากรและชุดข้อมูลด้านเกรดในกลุ่มรายวิชาศึกษาทั่วไป โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.878, 0.879, 0.878 และ 0.878 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.843, 0.843, 0.843 และ 0.843 ส่วนโมเดลอื่นก็มีผลการประเมินที่น้อยลงมาตามลำดับ และในทำนองเดียวกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.745, 0.744, 0.745 และ 0.744 จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกบังคับหลังจากปรับสมดุลให้กับชุดข้อมแล้วแล้ว ส่งผลต่อการทำนายในระดับสูงมาก ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนใด สามารถนำตัวแปรด้านเกรดในกลุ่มรายวิชาเอกบังคับมาร่วมทำการวิเคราะห์หาผลลัพธ์เป้าหมายได้อย่างถูกต้อง ที่ซึ่งมีค่าความแม่นยำสูงถึง 87.8% โดยอาศัยอัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลสำหรับการทำนายผล

ตาราง 4-28

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับ (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.763	0.774	0.763	0.745
NB	**0.878**	**0.879**	**0.878**	**0.878**
DT	0.745	0.744	0.745	0.744
SVM	0.790	0.788	0.790	0.789
ANN	0.843	0.843	0.843	0.843

จากตาราง 4-29 พบว่า รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) ให้ค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีผลการคำนวณเท่ากับ 0.989 ลำดับที่ 2 มีค่าเกนความรู้เท่ากัน คือ รายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) และ รายวิชาบริการสารสนเทศและช่วยการค้นคว้า (Information and Reference Services) โดยมีค่าเกนความรู้เท่ากับ 0.82 ส่วนรายวิชาการวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification) และรายวิชาการวิเคราะห์และออกแบบระบบสารสนเทศ (Information Systems Analysis and Design) มีค่าเกนความรู้ที่ 0.779 และ 0.778 ตามลำดับ

สำหรับรายวิชาที่เหลือให้ค่าการประเมินที่น้อยลงมาแต่ก็ยังอยู่ในเกณฑ์ที่ค่อนข้างสูง ดังนั้น รายวิชาที่อยู่ในอันดับต้น ๆ สามารถนำมาใช้เพื่อทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตได้เป็นอย่างดี

ตาราง 4-29

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	0.989
2	Organization of Information Resources	0.83	0.820
3	Information and Reference Services	0.83	0.820
4	Library of Congress Classification	0.79	0.779
5	Information Systems Analysis and Design	0.79	0.778
6	Electronic Information and Record Management	0.78	0.768
7	Cataloging of Information Resources	0.72	0.713
8	Library Automation Systems	0.70	0.697
9	Programming in Information Work	0.65	0.647
10	Reading for Information Professional	0.62	0.616
11	Information Science	0.61	0.607
12	Database Management for Information Work	0.56	0.554
13	Collection Development	0.55	0.548
14	Presentation and Training in Information Work	0.51	0.505
15	Management of Information Institutes	0.48	0.472
16	Web Design for Information Work	0.44	0.433
17	Information Technology	0.43	0.429
18	Seminar on Current Issues and Trend in Information Science	0.27	0.267

ผลการเปรียบเทียบด้วยชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วแสดงไว้ในตารางที่ 4-30 พบว่าชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าอย่างเห็นได้ชัดสำหรับทุกอัลกอริทึมจำแนกประเภท ด้วยค่าความแม่นยำเท่ากับ 87.8% จากอัลกอริทึมแบบเบส์อย่างง่าย (NB) นอกจากนี้ สำหรับโมเดลการทำนายอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมด ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น

ตาราง 4-30

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกบังคับระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.677	**0.763**
NB	0.814	**0.878**
DT	0.593	**0.745**
SVM	0.688	**0.790**
ANN	0.745	**0.843**

3. รายวิชาเอกบังคับด้านสารสนเทศศาสตร์ (IS Core Subjects Grade) โดยใช้ SMOTE

รายวิชาในกลุ่มเอกบังคับด้านสารสนเทศศาสตร์มีจำนวนทั้งหมด 9 รายวิชา โดยพิจารณาจากคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านการจัดการสารสนเทศเป็นหลัก ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชานี้แสดงไว้ในตาราง 4-31 จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้หลังการจากที่ปรับสมดุลของคลาสเป้าหมายแล้วให้ผลกระทบต่อโมเดลค่อนข้างมาก ซึ่งสอดคล้องกับผลการวิเคราะห์ในหัวข้อก่อนหน้านี้ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.858, 0.858, 0.858 และ 0.857 ตามลำดับ ลำดับที่ 2 คือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ที่มีค่าการประเมินเท่ากับ 0.83, 0.832, 0.83 และ 0.831 ส่วนโมเดลอื่นก็มีผลการประเมินที่ไม่ห่างกันมาก ส่วนโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.775, 0.777, 0.775 และ 0.776 ทั้งนี้การพิจารณาลำดับจะพิจารณาจากค่าความแม่นยำเป็นตัวตั้งแล้วตามด้วยค่าความเที่ยงตรงตามลำดับ

ตาราง 4-31

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.775	0.786	0.775	0.761
NB	**0.858**	**0.858**	**0.858**	**0.857**
DT	0.775	0.777	0.775	0.776
SVM	0.830	0.832	0.830	0.831
ANN	0.805	0.805	0.805	0.805

จากตาราง 4-32 พบว่า รายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) และ รายวิชาบริการสารสนเทศและช่วยการค้นคว้า (Information and Reference Services) จัดอยู่ในลำดับที่ 1 เหมือนกัน โดยมีค่าเกนความรู้เท่ากับ 0.82 ลำดับที่ 2 คือ รายวิชาการวิเคราะห์หมวดหมู่ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification) มีค่าเกนความรู้เท่ากับ 0.779 และรายวิชาในลำดับที่ 3 คือ โดยมีค่าเกนความรู้เป็น 0.713 สำหรับรายวิชาที่เหลือให้ค่าการประเมินที่น้อยกว่า 0.7 ดังนั้น เราควรพิจารณาเฉพาะรายวิชาที่อยู่ในอันดับต้น ๆ สำหรับนำไปสร้างโมเดลการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิต

ตาราง 4-32

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Organization of Information Resources	1.00	0.820
2	Information and Reference Services	1.00	0.820
3	Library of Congress Classification	0.95	0.779
4	Cataloging of Information Resources	0.87	0.713
5	Library Automation Systems	0.85	0.697
6	Reading for Information Professional	0.75	0.616
7	Information Science	0.74	0.607
8	Collection Development	0.67	0.548
9	Management of Information Institutes	0.58	0.472

ตาราง 4-33

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.696	**0.775**
NB	0.753	**0.858**
DT	0.643	**0.775**
SVM	0.703	**0.830**
ANN	0.726	**0.805**

สำหรับการเปรียบเทียบด้วยชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วแสดงไว้ในตาราง 4-33 จะเห็นได้ว่า ชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าอย่างเห็นได้ชัดอีกเช่นเคย ในทุกอัลกอริทึมของการจำแนกประเภท ด้วยค่าความแม่นยำเท่ากับ 85.8% จากอัลกอริทึมแบบเบส์อย่างง่าย (NB) นอกจากนี้สำหรับโมเดลการทำนายอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมดเช่นกัน ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น และอยู่ในเกณฑ์ที่ให้ค่าความแม่นยำในระดับสูง

4. รายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ (IT Core Subjects Grade) โดยใช้ SMOTE

รายวิชาในกลุ่มเอกบังคับด้านเทคโนโลยีสารสนเทศมีจำนวน 9 รายวิชา ซึ่งผู้วิจัยได้ออกแบบให้มีจำนวนรายวิชาเท่ากันกับกลุ่มเอกบังคับด้านสารสนเทศศาสตร์ เพื่อให้ผลการวิเคราะห์เกิดความสมดุลกันในแง่ของจำนวนตัวแปรของแต่ละชุดข้อมูล หัวข้อนี้ได้ใช้ชุดข้อมูลที่ผ่านกระบวนการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) มาแล้ว ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-34 จากการศึกษาพบว่าการทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลค่อนข้างมากใกล้เคียงกับกลุ่มรายวิชาเอกบังคับด้านสารสนเทศศาสตร์ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดยังคงเป็น โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.83, 0.831, 0.83 และ 0.83 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) โดยมีค่าการประเมินเท่ากับ 0.79, 0.79, 0.79 และ 0.79 ในขณะที่โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (KNN) ที่มีค่าการประเมินอยู่ที่ 0.735, 0.742, 0.735 และ 0.713 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศมีค่าความแม่นยำน้อยกว่าด้านสารสนเทศศาสตร์อยู่เล็กน้อย แต่ก็ยังคงส่งผลต่อการทำนายในระดับที่ค่อนข้างสูงด้วยค่าความแม่นยำสูงถึง 83% ดังนั้น การจำแนกประเภทของนิสิตยังคงสามารถนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ได้ โดยใช้อัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลสำหรับการทำนายผล

ตาราง 4-34

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.735	0.742	0.735	0.713
NB	**0.830**	**0.831**	**0.830**	**0.830**
DT	0.740	0.746	0.740	0.743
SVM	0.785	0.787	0.785	0.786
ANN	0.790	0.790	0.790	0.790

การจัดลำดับความสำคัญของตัวแปรในตาราง 4-35 แสดงให้เห็นว่า รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) มีค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่นในกลุ่มเดียวกัน โดยมีค่าเกนความรู้เท่ากับ 0.989 ลำดับที่ 2 คือ รายวิชาการวิเคราะห์และออกแบบระบบสารสนเทศ (Information Systems Analysis and Design) ที่มีค่าเกนความรู้เท่ากับ 0.778 ซึ่งค่อนข้างห่างจากลำดับที่ 1 ลำดับที่ 3 คือ รายวิชาการจัดการเอกสารและสารสนเทศอิเล็กทรอนิกส์ (Electronic Information and Record Management) ที่มีค่าการประเมินเท่ากับ 0.768 ส่วนรายวิชาอื่น ๆ ในกลุ่มนี้ ให้ค่าเกนความรู้ต่ำกว่า 0.7 ดังนั้น จึงถือได้ว่ามีลำดับความสำคัญไม่มากนักและไม่ควรนำมาเป็นปัจจัยในการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิต มีอีกหนึ่งข้อสังเกตคือ ผลการประเมินหาค่าเกนความรู้ในรายวิชาทั้งกลุ่มด้านสารสนเทศศาสตร์และกลุ่มด้านเทคโนโลยีสารสนเทศ ล้วนให้ผลลัพธ์เท่ากันกับในตาราง 4-29 ที่ซึ่งใช้ชุดข้อมูลในกลุ่มรายวิชาเอกบังคับทั้งหมด แต่หากเปรียบเทียบกันระหว่างสองกลุ่มนี้ กลุ่มรายวิชาด้านสารสนเทศให้ผลลัพธ์ที่ดีกว่าอยู่เล็กน้อย ซึ่งไม่มีนัยยะสำคัญมากนัก

ตาราง 4-35

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	0.989
2	Information Systems Analysis and Design	0.79	0.778
3	Electronic Information and Record Management	0.78	0.768
4	Programming in Information Work	0.65	0.647
5	Database Management for Information Work	0.56	0.554
6	Presentation and Training in Information Work	0.51	0.505
7	Web Design for Information Work	0.44	0.433
8	Information Technology	0.43	0.429
9	Seminar on Current Issues and Trend in Information Science	0.27	0.267

สำหรับการเปรียบเทียบด้วยชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วที่แสดงไว้ในตารงที่ 4-36 ยังคงให้ผลลัพธ์ไปในทิศทางเดียวกัน นั่นคือ ชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าในทุกอัลกอริทึมของการจำแนกประเภท ด้วยค่าความแม่นยำสูงสุดเท่ากับ 83% จากอัลกอริทึมแบบเบส์อย่างง่าย

(NB) นอกจากนี้ สำหรับโมเดลการทำนายอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมดเช่นกัน ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น และอยู่ในเกณฑ์ที่ให้ค่าความแม่นยำในระดับที่ค่อนข้างสูงทีเดียว

ตาราง 4-36

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.658	**0.735**
NB	0.745	**0.830**
DT	0.593	**0.740**
SVM	0.650	**0.785**
ANN	0.639	**0.790**

5. รายวิชาเอกเลือก (Elective Subjects Grade) โดยใช้ SMOTE

ในกลุ่มของรายวิชาเอกเลือกมีจำนวนรายวิชาอยู่ทั้งหมด 10 รายวิชา เนื่องจากเป็นรายวิชาเอกเลือก ดังนั้น นิสิตแต่ละคนจึงเรียนไม่เหมือนกันทั้งหมด ข้อมูลชุดนี้จึงเป็นการประมาณการโดยใช้เกรดจากรายวิชาที่ใกล้เคียงกันเพื่อทดแทนรายวิชาที่ขาดหายไป ชุดข้อมูลรายวิชาเอกเลือกนี้จะใช้ชุดข้อมูลที่ผ่านการปรับสมดุลแล้วด้วยเทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) มาสร้างเป็นโมเดลการทำนายจาก 5 อัลกอริทึมที่แตกต่างกัน ผลลัพธ์ที่ได้แสดงไว้ในตาราง 4-37 ซึ่งจะเห็นได้ว่า ผลการทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้อยู่ในเกณฑ์ดีมาก โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) สูงที่สุดคือ โมเดลแบบเบส์อย่างง่าย (NB) นั่นคือ 0.848, 0.848, 0.848 และ 0.847 ตามลำดับ ลำดับที่ 2 คือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ที่มีค่าการประเมิน 0.818, 0.815, 0.818 และ 0.816 ส่วนโมเดลอื่นยังคงให้ผลการประเมินที่น้อยลงมาไม่มากนัก และโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) โดยมีผลการประเมินเท่ากับ 0.725, 0.713, 0.725 และ 0.717 ตามลำดับ จากผลการประเมินสามารถสรุปได้ว่า ชุดข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกที่ปรับสมดุลแล้วส่งผลต่อการทำนายผลสัมฤทธิ์ทางการศึกษาในระดับที่ดี แต่ยังน้อยกว่าชุดข้อมูลกลุ่มรายวิชาเอกบังคับอยู่บ้าง อย่างไรก็ตาม ข้อมูลชุดนี้ยังสามารถนำมาใช้ประโยชน์ได้โดยมีโมเดลแบบเบส์อย่างง่าย (NB) ที่ให้ค่าความแม่นยำสูงสุดคือ 84.8%

ตาราง 4-37

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกเลือก (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.743	0.755	0.743	0.720
NB	**0.848**	**0.848**	**0.848**	**0.847**
DT	0.725	0.713	0.725	0.717
SVM	0.818	0.815	0.818	0.816
ANN	0.813	0.809	0.813	0.810

การวิเคราะห์ผลการจัดลำดับความสำคัญของรายวิชาเอกเลือกด้วยเทคนิค Information Gain (IG) จากตาราง 4-38 สามารถสรุปได้ว่า รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ให้ค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าการประเมินเท่ากับ 0.812 ลำดับที่ 2 คือ รายวิชาการค้นคืนสารสนเทศอิเล็กทรอนิกส์ (Electronic Information Retrieval) โดยมีค่าเกนความรู้เท่ากับ 0.713 สำหรับรายวิชาอื่นในกลุ่มนี้มีค่าเกนความรู้ต่ำกว่า 0.7 ดังนั้น จึงอาจยกเว้นที่จะนำมาเป็นปัจจัยในการสร้างโมเดลการทำนายสำหรับพัฒนาผู้เรียนในลำดับต่อไป

ตาราง 4-38

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Data Communication and Computer Networks	1.00	0.812
2	Electronic Information Retrieval	0.88	0.713
3	User Behavior and Information Needs	0.85	0.689
4	Open Source Program Application	0.81	0.659
5	Printing and Publishing Business	0.78	0.631
6	Information in Science and Technology	0.77	0.626
7	English for Information Professional	0.75	0.610
8	Multimedia Technology	0.67	0.548
9	Information of Local Wisdom	0.62	0.506
10	Knowledge Management	0.45	0.367

ตาราง 4-39

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกเลือกระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.620	**0.743**
NB	0.745	**0.848**
DT	0.567	**0.725**
SVM	0.658	**0.818**
ANN	0.681	**0.813**

การเปรียบเทียบของชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วในกลุ่มรายวิชาเอกเลือกได้แสดงไว้ในตารงที่ 4-39 ซึ่งยังคงให้ผลลัพธ์ไปในทิศทางเดียวกัน นั่นคือ ชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าในทุกอัลกอริทึมของการจำแนกประเภท ด้วยค่าความแม่นยำสูงสุดคือ 84.8% ซึ่งอยู่ในระดับดีมาก อัลกอริทึมที่ดีที่สุดสำหรับสร้างโมเดลคือ อัลกอริทึมแบบเบส์อย่างง่าย (NB) สำหรับโมเดลอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมดเช่นกันกรณีที่ใช้ชุดข้อมูลสมดุล ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจึงมีส่วนสำคัญในการช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น

6. รายวิชาเอกเลือกด้านสารสนเทศศาสตร์ (IS Elective Subjects Grade) โดยใช้ SMOTE

รายวิชาในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์แบ่งออกมาจากกลุ่มรายวิชาเอกเลือกทั้งหมดได้ 5 รายวิชา โดยพิจารณาจากคำอธิบายรายวิชาที่เกี่ยวข้องกับศาสตร์ด้านการจัดการสารสนเทศเป็นหลัก ผลลัพธ์ที่ได้จากการประเมินเฉพาะกลุ่มรายวิชานี้แสดงไว้ในตาราง 4-40

ตาราง 4-40

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.665	0.650	0.665	0.637
NB	0.760	0.752	0.760	0.752
DT	0.688	0.678	0.688	0.680
SVM	**0.763**	**0.762**	**0.763**	**0.760**
ANN	0.715	0.708	0.715	0.711

จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลอยู่ในเกณฑ์ดี โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุดคือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ซึ่งมีค่าการประเมินตามลำดับ ดังนี้ 0.763, 0.762, 0.763 และ 0.76 ผลการประเมินในลำดับที่ 2 คือ โมเดลแบบเบส์อย่างง่าย (NB) ที่มีค่าการประเมินเท่ากับ 0.76, 0.752, 0.76 และ 0.752 ในทางกลับกัน โมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (KNN) ที่มีค่าการประเมินอยู่ที่ 0.665, 0.65, 0.665 และ 0.637 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ส่งผลต่อการทำนายในระดับดี ด้วยค่าความแม่นยำสูงสุดที่ 76.3% ดังนั้น การจำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนกลุ่มใดสามารถนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ด้วยได้ อย่างไรก็ตาม เกรดในกลุ่มรายวิชาเอกเลือกนี้มักเกิดขึ้นในช่วงครึ่งหลังของการศึกษา นั่นคือ ก่อนนิสิตจะสำเร็จการศึกษาไม่นานนัก ดังนั้น การนำตัวแปรในกลุ่มนี้มาทำนายผลสัมฤทธิ์ทางการเรียนจึงอาจไม่เหมาะสมนัก

ตาราง 4-41

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	User Behavior and Information Needs	1.00	0.689
2	Printing and Publishing Business	0.92	0.631
3	English for Information Professional	0.89	0.610
4	Information of Local Wisdom	0.73	0.506
5	Knowledge Management	0.53	0.367

สำหรับการจัดลำดับความสำคัญของรายวิชาในกลุ่มนี้ด้วยค่าเกนความรู้แสดงไว้ในตาราง 4-41 ผลที่ได้แสดงให้เห็นว่า รายวิชาพฤติกรรมผู้ใช้และความต้องการสารสนเทศ (User Behavior and Information Needs) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่นในกลุ่มเดียวกัน คือ 0.689 ลำดับที่ 2 คือ รายวิชาการพิมพ์และธุรกิจการพิมพ์ (Printing and Publishing Business) มีค่าเกนความรู้เท่ากับ 0.631 อย่างไรก็ตาม ค่าเกนความรู้ในกลุ่มนี้ต่ำกว่า 0.7 ในทุกรายวิชา ดังนั้น จึงถือได้ว่ารายวิชาในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์มีลำดับความสำคัญที่ค่อนข้างน้อย จึงไม่มีน้ำหนักมากนักที่จะนำมาใช้ประกอบการสร้างโมเดลการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิต

ตาราง 4-42 แสดงการเปรียบเทียบชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วของชุดข้อมูลในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์ ผลที่ได้คือ ชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำมากกว่าพอสมควรสำหรับทุกอัลกอริทึมของการจำแนกประเภท โดยมีค่าความแม่นยำสูงสุดเท่ากับ 76.3% ซึ่งสร้างจากโมเดลแบบเบส์อย่างง่าย (NB) การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น และอยู่ในเกณฑ์ที่ให้ค่าความแม่นยำในระดับปานกลางถึงค่อนข้างดี

ตาราง 4-42

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.551	**0.665**
NB	0.612	**0.760**
DT	0.498	**0.688**
SVM	0.643	**0.763**
ANN	0.559	**0.715**

7. รายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศ (IT Elective Subjects Grade) โดยใช้ SMOTE

รายวิชาในกลุ่มเอกเลือกด้านเทคโนโลยีสารสนเทศมีจำนวน 5 รายวิชา เท่ากับจำนวนรายวิชาในกลุ่มเอกเลือกด้านสารสนเทศศาสตร์ โดยใช้ชุดข้อมูลที่ผ่านการปรับสมดุลแล้วด้วยเทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-43

ตาราง 4-43

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาเอกบังคับด้านเทคโนโลยีสารสนเทศ (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.723	0.714	0.723	0.708
NB	**0.800**	**0.795**	**0.800**	**0.796**
DT	0.740	0.728	0.740	0.730
SVM	0.793	0.790	0.793	0.791
ANN	0.763	0.758	0.763	0.759

จากการศึกษาพบว่า การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลอยู่ในเกณฑ์ดี และดีกว่าผลลัพธ์จากกลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ โมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) ดีที่สุด ได้แก่ โมเดลแบบเบส์อย่างง่าย (NB) ด้วยค่าการประเมินเท่ากับ 0.8, 0.795, 0.8 และ 0.796 ตามลำดับ ลำดับที่ 2 คือ โมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีค่าการประเมินเท่ากับ 0.793, 0.79, 0.793 และ 0.791 สำหรับโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (KNN) ที่มีค่าการประเมินอยู่ที่ 0.723, 0.714, 0.723 และ 0.708 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า ข้อมูลด้านเกรดของแต่ละรายวิชาในกลุ่มรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศมีค่าความแม่นยำมากกว่าด้านสารสนเทศศาสตร์อยู่เล็กน้อย จำแนกประเภทของนิสิตสามารถนำตัวแปรด้านเกรดในรายวิชากลุ่มนี้มาร่วมทำการวิเคราะห์ร่วมด้วยได้ อย่างไรก็ตาม รายวิชาเอกเลือกทั้งหมดมักอยู่ในช่วงการเรียนของนิสิตปี 3 และปี 4 ซึ่งใกล้จะสำเร็จการศึกษาแล้ว เพราะฉะนั้น ในทางปฏิบัติจึงไม่อาจนำผลลัพธ์จากรายวิชากลุ่มเอกเลือกทั้งด้านสารสนเทศศาสตร์และด้านเทคโนโลยีสารสนเทศมาเป็นปัจจัยการทำนายผลสัมฤทธิ์ทางการศึกษาได้

ตาราง 4-44

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Data Communication and Computer Networks	1.00	0.812
2	Electronic Information Retrieval	0.88	0.713
3	Open Source Program Application	0.81	0.659
4	Information in Science and Technology	0.77	0.626
5	Multimedia Technology	0.67	0.548

การวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศที่ผ่านการปรับสมดุลข้อมูลแล้ว แสดงไว้ในตาราง 4-44 จากผลลัพธ์ที่ได้แสดงให้เห็นว่ารายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ให้ค่าเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าเกนความรู้เท่ากับ 0.812 ลำดับที่ 2 คือ รายวิชาการค้นคืนสารสนเทศอิเล็กทรอนิกส์ (Electronic Information Retrieval) โดยมีค่าเกนความรู้เท่ากับ 0.713 สำหรับรายวิชาที่เหลือให้ค่าเกนความรู้ที่น้อยลงตามลำดับและไม่เกิน 0.7 ถึงแม้ว่าผลลัพธ์นี้จะดีกว่ากลุ่มรายวิชาเอกเลือกด้านสารสนเทศศาสตร์ แต่ก็ถือได้ว่ามีลำดับความสำคัญที่ค่อนข้างน้อย จึงต้องพิจารณาให้ดีก่อนนำมาเป็นปัจจัยสำหรับการทำนายผลการเรียน

ตาราง 4-45

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.601	**0.723**
NB	0.684	**0.800**
DT	0.593	**0.740**
SVM	0.650	**0.793**
ANN	0.639	**0.763**

การเปรียบเทียบของชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วในกลุ่มรายวิชาเอกเลือกด้านเทคโนโลยีสารสนเทศได้แสดงไว้ในตารางที่ 4-45 ชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าชุดข้อมูลดั้งเดิมในทุกโมเดล โดยที่โมเดลแบบเบส์อย่างง่าย (NB) มีค่าความแม่นยำสูงสุดเท่ากับ 80% ซึ่งอยู่ในเกณฑ์ดี โมเดลอื่นที่ใช้ชุดข้อมูลที่ปรับสมดุลแล้วก็ให้ผลลัพธ์ที่ดีกว่าโมเดลที่ใช้ชุดข้อมูลดั้งเดิมทั้งหมด ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจึงมีส่วนสำคัญในการช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้น

8. รายวิชาทั้งหมด (All Subjects Grade) โดยใช้ SMOTE

หัวข้อนี้ได้นำรายวิชาทั้งหมดซึ่งประกอบไปด้วย 40 รายวิชา มาสร้างโมเดลการทำนายเพื่อประเมินสิทธิภาพ ผลลัพธ์ที่ได้จากการประเมินแสดงไว้ในตาราง 4-46 การทำนายผลสัมฤทธิ์ทางการเรียนด้วยข้อมูลชุดนี้มีผลกระทบต่อโมเดลมากที่สุด เนื่องจากเป็นข้อมูลเกรดจากรายวิชาทั้งหมด ยกเว้นเพียงรายวิชาฝึกงานเท่านั้น ดังนั้น ผลลัพธ์ที่ได้จึงสูงกว่าทุกชุดข้อมูลที่ผ่านมาทั้งหมด อีกทั้งกรณีนี้ยังใช้ชุดข้อมูลที่ผ่านการปรับสมดุลแล้ว จึงส่งผลให้โมเดลมีความแม่นยำสูง สำหรับโมเดลที่ให้ค่าความแม่นยำ (Accuracy) ค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) และค่าเอฟ (F-Measure) สูงที่สุดได้แก่ โมเดลแบบเบส์อย่างง่าย (NB) โดยมีค่าการประเมินเท่ากับ 0.913, 0.918, 0.913 และ 0.914 ตามลำดับ ลำดับที่ 2 คือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าการประเมินเท่ากับ 0.878, 0.88, 0.878 และ 0.878 ส่วนโมเดลอื่นก็มีผลการประเมินที่ลดหลั่นกันลงมาแต่ก็ยังอยู่ในเกณฑ์ดีมาก สำหรับโมเดลที่ให้ค่าการประเมินน้อยที่สุดคือ โมเดลต้นไม้ตัดสินใจ (DT) ที่มีค่าการประเมินอยู่ที่ 0.75, 0.75, 0.75 และ 0.748 ตามลำดับ จากผลการประเมินนี้สรุปได้ว่า หากใช้ชุดข้อมูลด้านเกรดของรายวิชาทั้งหมดจะมีความแม่นยำของการทำนายในระดับที่สูงมาก โดยที่โมเดลลำดับที่ 1 มีค่าความแม่นยำสูงถึง 91.3% โดยอาศัยอัลกอริทึมแบบเบส์อย่างง่าย (NB) ดังนั้น การ

จำแนกประเภทนิสิตว่าอยู่ในกลุ่มผลสัมฤทธิ์ทางการเรียนกลุ่มใด สามารถนำตัวแปรด้านเกรดของรายวิชาทั้งหมดมาร่วมวิเคราะห์หาผลลัพธ์เป้าหมายได้อย่างถูกต้อง อย่างไรก็ตาม การที่จะทราบเกรดของรายวิชาทั้งหมดต้องใช้ข้อมูลของนิสิตปี 4 ที่ผ่านการเรียนในภาคการศึกษาแรกมาแล้ว ด้วยเหตุนี้ ในทางปฏิบัติจึงไม่อาจใช้ประโยชน์ชุดข้อมูลนี้ได้ แต่อาจใช้เพื่อเปรียบเทียบผลการประเมินสูงสุดที่เป็นไปได้เท่านั้น

ตาราง 4-46

การเปรียบเทียบผลลัพธ์จากข้อมูลรายวิชาทั้งหมด (SMOTE)

Model	Accuracy	Precision	Recall	F-Measure
KNN	0.780	0.800	0.780	0.765
NB	**0.913**	**0.918**	**0.913**	**0.914**
DT	0.750	0.750	0.750	0.748
SVM	0.850	0.853	0.850	0.850
ANN	0.878	0.880	0.878	0.878

ในส่วนการวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดของรายวิชาทั้งหมดที่ปรับสมดุลของคลาสเป้าหมายแล้ว ผลลัพธ์ที่ได้สอดคล้องตรงกันกับชุดข้อมูลกลุ่มต่าง ๆ ก่อนหน้านี้ นั่นคือ รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) ให้ค่าความเกนความรู้สูงที่สุด โดยมีผลการคำนวณเท่ากับ 1.017 ลำดับที่ 2 คือ รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ที่มีค่าเกนความรู้เท่ากับ 0.863 ลำดับที่ 3 คือ รายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) ที่มีค่าเกนความรู้เท่ากับ 0.862 ส่วนลำดับที่ 4 คือ รายวิชาบริการสารสนเทศและช่วยการค้นคว้า (Information and Reference Services) โดยมีค่าเกนความรู้เท่ากับ 0.822 ซึ่งแสดงไว้ในตาราง 4-47 จะเห็นได้ว่า รายวิชาข้างต้นเป็นรายวิชาที่อยู่ตารางเรียนของนิสิตปี 2 และ 3 ดังนั้น จึงสามารถนำมาใช้เป็นปัจจัยในการวิเคราะห์หาความสัมพันธ์ระหว่างเกรดของรายวิชากับผลสัมฤทธิ์ทางการเรียนได้เป็นอย่างดี

ตาราง 4-47

ลำดับความสำคัญของตัวแปรด้านเกรดรายวิชาทั้งหมดด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	Research and Statistics in Information Studies	1.00	1.017
2	Data Communication and Computer Networks	0.85	0.863
3	Organization of Information Resources	0.85	0.862
4	Information and Reference Services	0.81	0.822
5	Information Systems Analysis and Design	0.76	0.778
6	Electronic Information and Record Management	0.76	0.777
7	Library of Congress Classification	0.75	0.763
8	Moving Forward in a Digital Society with ICT	0.74	0.757
9	Library Automation Systems	0.72	0.732
10	Open Source Program Application	0.70	0.709
11	Electronic Information Retrieval	0.69	0.703
12	Printing and Publishing Business	0.68	0.691
13	Information in Science and Technology	0.68	0.689
14	User Behavior and Information Needs	0.67	0.679
15	Cataloging of Information Resources	0.66	0.667
16	Programming in Information Work	0.65	0.663
17	Information Science	0.65	0.661
18	Thai Language Skills for Communication	0.64	0.654
19	Reading for Information Professional	0.62	0.634
20	English for Information Professional	0.6	0.614
21	Natural Disasters	0.57	0.584
22	Database Management for Information Work	0.57	0.577
23	English for Communication	0.56	0.567
24	Collection Development	0.53	0.540
25	Information of Local Wisdom	0.52	0.530
26	Psychology for Living and Adjustment	0.51	0.523

Order	Variables	Normalized Weight	Information Gain
27	Multimedia Technology	0.51	0.517
28	Presentation and Training in Information Work	0.5	0.508
29	Information Technology	0.48	0.490
30	Collegiate English	0.47	0.482
31	Management of Information Institutes	0.47	0.478
32	Web Design for Information Work	0.46	0.466
33	Knowledge Management	0.43	0.439
34	Life and Health	0.41	0.416
35	Sufficiency Economy and Social Development	0.37	0.376
36	Contemporary Scientific Innovation	0.32	0.328
37	English Writing for Communication	0.28	0.284
38	Seminar on Current Issues and Trend in Information Science	0.28	0.283
39	Arts and Creativity	0.18	0.181
40	Exercise for Quality of Life	0.07	0.670

ตาราง 4-48

การเปรียบเทียบค่าความแม่นยำจากข้อมูลรายวิชาทั้งหมดระหว่างข้อมูลดั้งเดิมและข้อมูลที่ปรับความสมดุลแล้ว

Model	Accuracy (Original Dataset)	Accuracy (Balanced Dataset)
KNN	0.738	**0.780**
NB	0.844	**0.913**
DT	0.616	**0.750**
SVM	0.768	**0.850**
ANN	0.806	**0.878**

ผลการเปรียบเทียบด้วยชุดข้อมูลดั้งเดิมกับชุดข้อมูลที่ปรับสมดุลแล้วแสดงไว้ในตารางที่ 4-48 พบว่าชุดข้อมูลที่มีความสมดุลให้ผลการทำนายที่แม่นยำกว่าอย่างเห็นได้ชัดสำหรับทุกอัลกอริทึมจำแนกประเภทด้วยค่าความแม่นยำสูงสุดเท่ากับ 91.3% ในขณะที่ชุดข้อมูลดั้งเดิมมีค่าความแม่นยำสูงสุดอยู่ที่ 84.4% ซึ่งห่าง

กันพอสมควร ผลลัพธ์ทั้งสองเป็นผลจากการใช้อัลกอริทึมแบบเบส์อย่างง่าย (NB) เพื่อสร้างเป็นโมเดลการจำแนกประเภท นอกจากนี้ สำหรับโมเดลการทำนายอื่นก็ให้ผลลัพธ์ที่ดีกว่าทั้งหมด ดังนั้น การปรับสมดุลให้กับชุดข้อมูลก่อนสร้างโดเมลจะช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนมีประสิทธิภาพดียิ่งขึ้นสำหรับทุกกลุ่มชุดข้อมูลด้านเกรดของแต่ละรายวิชาทั้งหมด

ตาราง 4-49 เป็นการเปรียบเทียบชุดข้อมูลจากกลุ่มรายวิชาที่แตกต่างกันโดยอาศัยมาตรวัดที่เป็นค่าความแม่นยำ (Accuracy) เท่านั้น กลุ่มรายวิชาเอกบังคับมีผลการประเมินสูงที่สุดเมื่อเทียบกับกลุ่มรายวิชาศึกษาทั่วไปและกลุ่มรายวิชาเอกเลือก โดยมีค่าความแม่นยำคิดเป็น 87.8% จากโมเดลแบบเบส์อย่างง่าย (NB) รองลงมาคือกลุ่มรายวิชาเอกเลือกที่มีค่าความแม่นยำคิดเป็น 84.8% จากโมเดลแบบเบส์อย่างง่าย (NB) เช่นกัน ส่วนกลุ่มรายวิชาศึกษาทั่วไปให้การประเมินที่น้อยที่สุดคือ 77.3% ซึ่งเป็นการประเมินจากโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) เมื่อแบ่งกลุ่มรายวิชาเอกบังคับและเอกเลือกเป็นกลุ่มรายวิชาด้านสารสนเทศศาสตร์และกลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศ พบว่า ชุดข้อมูลจากทั้งสองกลุ่มรายวิชายังคงให้ผลลัพธ์ที่ดีสำหรับกลุ่มรายวิชาเอกบังคับ โดยมีค่าความแม่นยำคิดเป็น 85.8% และ 83% ตามลำดับ ส่วนกลุ่มรายวิชาเอกเลือกทั้งสองชุดข้อมูลให้ผลการประเมินที่ใกล้เคียงกับชุดข้อมูลจากกลุ่มรายวิชาศึกษาทั่วไป ซึ่งให้ค่าความแม่นยำตั้งแต่ 80% ลงไป

ตาราง 4-49

การเปรียบเทียบผลลัพธ์จากกลุ่มรายวิชาทั้งหมดด้วยค่าความแม่นยำ (SMOTE)

Model	Accuracy (%)							
	GE	Core	IS-Core	IT-Core	Elec.	IS-Elec.	IT-Elec.	All
KNN	0.680	0.763	0.775	0.735	0.743	0.665	0.723	0.780
NB	0.758	0.878	0.858	0.830	0.848	0.760	0.800	0.913
DT	0.690	0.745	0.775	0.740	0.725	0.688	0.740	0.750
SVM	0.773	0.790	0.830	0.785	0.818	0.763	0.793	0.850
ANN	0.765	0.843	0.805	0.790	0.813	0.715	0.763	0.878

เมื่อเปรียบเทียบผลลัพธ์ระหว่างกลุ่มรายวิชาด้านสารสนเทศศาสตร์และกลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศด้วยชุดข้อมูลที่ปรับสมดุลแล้ว พบว่า กลุ่มรายวิชาด้านสารสนเทศศาสตร์มีค่าความแม่นยำที่ดีกว่าในกลุ่มรายวิชาเอกบังคับ ในทางกลับกัน กลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศมีผลการประเมินที่สูงกว่าในกลุ่มรายวิชาเอกเลือก อย่างไรก็ตาม งานวิจัยนี้ต้องการค้นหาความสัมพันธ์ระหว่างข้อมูลก่อนที่นิสิตจะสำเร็จการศึกษากับผลสัมฤทธิ์ทางการศึกษาของนิสิตคนนั้นในระยะต้นและระยะกลางของหลักสูตร ดังนั้น ชุดข้อมูลด้านเกรดของรายวิชาเอกบังคับจึงเป็นปัจจัยสำคัญที่จะนำมาวิเคราะห์และสร้างเป็นโมเดลการทำนายที่เหมาะสมที่สุด

ผลลัพธ์จากการสร้างโมเดลด้วยข้อมูลด้านเกรดเฉลี่ยในแต่ละภาคการศึกษาและเกรดเฉลี่ยรวม (Early Years GPA)

สำหรับการสร้างโมเดลการทำนายด้วยชุดข้อมูลด้านเกรดเฉลี่ยได้ใช้แนวทางเดียวกันกับชุดข้อมูลด้านประชากรและด้านเกรดของแต่ละรายวิชา โดยนำข้อมูลด้านเกรดเฉลี่ยที่ประกอบไปด้วยแอตทริบิวต์ เช่น เกรดเฉลี่ยจากภาคเรียนที่ 1 (GPA1) เกรดเฉลี่ยจากภาคเรียนที่ 2 (GPA2) เกรดเฉลี่ยจากภาคเรียนที่ 3 (GPA3) รวมถึงเกรดเฉลี่ยรวม (AGPA) เมื่อสิ้นสุดการเรียนในแต่ละภาคการศึกษาเริ่มจากภาคเรียนที่ 2 ถึง 6 เป็นต้น ข้อมูลเกรดเฉลี่ยนี้เป็นของภาคการศึกษาที่ 1 ถึง 6 นั่นคือ จะใช้เฉพาะข้อมูลของนิสิตใน 3 ปีแรกเท่านั้น และเนื่องจากตัวแปรด้านเกรดเฉลี่ยบางตัวมีความเกี่ยวข้องกัน งานวิจัยนี้จึงแยกพิจารณาตัวแปรแต่ละตัวออกจากกัน การทดลองเริ่มจากสร้างโมเดลด้วยอัลกอริทึมการจำแนกประเภทชนิดต่าง ๆ ทั้งหมด 5 อัลกอริทึม ผลลัพธ์ที่ได้จะนำมาเปรียบเทียบกันโดยพิจารณาจากค่าความแม่นยำ (Accuracy) เท่านั้น และใช้เทคนิคการสร้างโมเดลแบบ 10-Fold Cross Validation สำหรับการพิจารณาค่าของผลลัพธ์จะเป็นไปในแนวทางเดียวกัน คือ พิจารณาที่ทศนิยม 3 ตำแหน่ง เช่นเดียวกันกับหัวข้อที่ผ่านมา

1. ข้อมูลด้านเกรดเฉลี่ยในแต่ละภาคการศึกษาและเกรดเฉลี่ยรวม

จากตาราง 4-50 ค่าความแม่นยำที่ดีที่สุดในแต่ละโมเดลระบุได้ด้วยตัวหนา เกรดเฉลี่ยในภาคการศึกษาที่ 4 (GPA4) สำหรับโมเดลแบบเบส์อย่างง่าย (NB) มีผลลัพธ์ของการประเมินสูงที่สุด ด้วยค่าความแม่นยำคิดเป็น 76.8% สำหรับ 4 โมเดลที่เหลือ มีค่าความแม่นยำสูงสุดเท่ากันคือ 72.2% ซึ่งเป็นผลการประเมินจากตัวแปรเกรดเฉลี่ยรวมของ 6 ภาคการศึกษา (AGPA6) สังเกตได้ว่า เกรดเฉลี่ยในแต่ละภาคการศึกษาให้ผลลัพธ์การประเมินที่แตกต่างกัน แต่ในภาพรวมผลการประเมินค่อย ๆ เพิ่มขึ้นเมื่อเข้าใกล้ปลายทางของการศึกษา นั่นเพราะยิ่งมีจำนวนภาคการศึกษาเพิ่มขึ้นมากเท่าไหร่ เกรดเฉลี่ยรวมก็ยิ่งเข้าใกล้ผลลัพธ์เป้าหมายมากขึ้นเท่านั้นนั่นเอง อาจมีกรณีที่บางภาคการศึกษามีค่าความแม่นยำน้อย นั่นเป็นเพราะมีรายวิชาที่ทำให้นิสิตได้เกรดน้อยลงเมื่อเทียบกับภาคการศึกษาก่อนหน้านี้ เช่น นิสิตได้เกรด D หรือ F ในบางรายวิชา เป็นต้น สำหรับภาคเรียนที่ประเมินได้น้อยที่สุดยังคงเป็นภาคเรียนแรก เนื่องจากยังเหลือระยะเวลายาวนานกว่าจะสิ้นสุดการศึกษา ส่งผลให้โมเดลมีความแม่นยำค่อนข้างต่ำ

ตาราง 4-50

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านเกรดเฉลี่ยด้วยโมเดลที่แตกต่างกัน

GPA	Accuracy (%)				
	KNN	NB	DT	SVM	ANN
GPA1	59.7	59.7	59.7	59.7	57.4
GPA2	64.3	64.3	64.3	64.0	64.3
GPA3	62.4	62.4	62.4	62.4	62.4
GPA4	71.1	**76.8**	71.1	71.1	71.1
GPA5	68.4	68.4	68.4	68.4	68.4
GPA6	64.3	64.0	64.3	64.3	65.8
AGPA2	60.1	60.1	60.1	60.1	59.3
AGPA3	62.4	63.5	63.5	63.5	63.5
AGPA4	65.0	64.3	65.0	65.0	65.8
AGPA5	66.5	65.0	66.5	66.5	66.9
AGPA6	**72.2**	72.2	**72.2**	**72.2**	**72.2**

ตาราง 4-51

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านเกรดเฉลี่ยด้วยโมเดลที่แตกต่างกันแบบเพิ่มขึ้น

GPA	Accuracy (%)				
	KNN	NB	DT	SVM	ANN
GPA1	59.7	59.7	59.7	59.7	57.4
GPA2	63.1	63.5	63.9	65.8	64.6
GPA3	68.4	72.2	73.8	71.5	70.3
GPA4	77.9	79.1	76.4	75.7	76.0
GPA5	81.4	83.7	73.8	81.4	79.5
GPA6	83.3	**85.6**	75.7	84.4	82.1

การวิเคราะห์ด้วยการเพิ่มข้อมูลด้านเกรดเฉลี่ย (GPA) จากเกรดเฉลี่ยในภาคการศึกษาแรก (GPA1) ไปจนถึงภาคการศึกษาที่ 6 (GPA6) แสดงไว้ในตาราง 4-51 หมายความว่า ในตัวแปรแรก (GPA1) จะมีเพียงเกรดเฉลี่ยของภาคการศึกษาแรกเท่านั้น ในขณะที่ตัวแปรที่สอง (GPA2) จะประกอบไปด้วยเกรดเฉลี่ยของภาคการศึกษาแรกและภาคการศึกษาที่ 2 ในขณะทำนองเดียวกัน ตัวแปรที่สาม (GPA3) ประกอบไปด้วยเกรด

เฉลี่ยของภาคการศึกษาแรก ภาคการศึกษาที่ 2 และภาคการศึกษาที่ 3 เป็นลำดับไปเช่นนี้จนถึงตัวแปรตัวที่ 6 (GPA6)

ผลลัพธ์จากการเพิ่มข้อมูลเกรดเฉลี่ยของภาคการศึกษาที่ติดต่อกันจะช่วยเพิ่มความแม่นยำให้กับโมเดล ดังนั้นผลลัพธ์ที่ดีที่สุดมีความแม่นยำเท่ากับ 85.6% ซึ่งเป็นผลมาจากเกรดเฉลี่ยของทั้ง 6 ภาคการศึกษา ด้วยโมเดลแบบเบส์อย่างง่าย (NB) โดยที่ผลลัพธ์เหล่านี้แตกต่างจากตัวแปรเกรดเฉลี่ยรวมของ 6 ภาคการศึกษา (AGPA6) ที่เคยประเมินไว้ก่อนหน้านี้ อย่างไรก็ตาม ผลลัพธ์เหล่านี้มีรูปแบบที่คล้ายกัน นั่นคือ ความแม่นยำของโมเดลการทำนายสูงขึ้นเมื่อมีข้อมูลด้านเกรดเฉลี่ยที่เข้าใกล้จุดสิ้นสุดของการศึกษามากขึ้น

ตาราง 4-52

ลำดับความสำคัญของตัวแปรด้านเกรดเฉลี่ยด้วยค่าเกนความรู้

Order	Variables	Normalized Weight	Information Gain
1	AGPA6	1.00	0.918
2	AGPA5	0.91	0.831
3	AGPA4	0.86	0.787
4	GPA6	0.85	0.779
5	GPA5	0.83	0.762
6	GPA4	0.79	0.725
7	GPA3	0.76	0.697
8	AGPA3	0.74	0.682
9	AGPA2	0.63	0.579
10	GPA2	0.62	0.567
11	GPA1	0.58	0.534

การวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดเฉลี่ย โดยใช้เทคนิคการคำนวณหาค่าเกนความรู้ (Information Gain: IG) จากตาราง 4-52 พบว่า เกรดเฉลี่ยรวมของ 6 ภาคการศึกษา (AGPA6) ให้ผลลัพธ์สูงที่สุดเมื่อเทียบกับตัวแปรอื่น โดยมีค่าเกนความรู้เท่ากับ 0.918 ลำดับที่ 2 คือ เกรดเฉลี่ยรวมของ 5 ภาคการศึกษา (AGPA5) โดยมีค่าเกนความรู้เป็น 0.813 ถัดมาเป็น เกรดเฉลี่ยรวมของ 4 ภาคการศึกษา (AGPA4) และเกรดเฉลี่ยของภาคการศึกษาที่ 6 (GPA6) ด้วยค่าเกนความรู้เป็น 0.787 และ 0.779 ตามลำดับ จะเห็นว่า ตัวแปรยิ่งเข้าใกล้จุดสิ้นสุดของการศึกษามากเท่าไหร่ความสำคัญก็ยิ่งเพิ่มขึ้นมากเท่านั้น หากแต่วัตถุประสงค์ของงานวิจัยนี้คือ ต้องการทราบล่วงหน้าว่าผลสัมฤทธิ์ทางการศึกษาของนิสิตจะเป็นอย่างไรเมื่อสำเร็จการศึกษาแล้ว ดังนั้น จึงต้องอาศัยข้อมูลในระยะแรกถึงระยะกลางของการศึกษา ซึ่งผลการวิจัยชี้ให้เห็นว่า เราสามารถประเมินได้ว่านิสิตจะมีผลสัมฤทธิ์ทางการศึกษาเช่นใดด้วยข้อมูลใน 2 ปีแรก

โดยที่มีค่าเกนความรู้สูงถึง 0.787 ซึ่งเป็นตัวเลขที่สูงทีเดียว จึงสามารถนำข้อมูลนี้ไปใช้เพื่อกำหนดรูปแบบการสนับสนุนการศึกษาให้กับนิสิตต่อไปได้ในอีก 2 ปีที่เหลือของหลักสูตร

2. ข้อมูลด้านเกรดเฉลี่ยในแต่ละภาคการศึกษาและเกรดเฉลี่ยรวมโดยใช้ SMOTE

ในหัวข้อนี้จะเป็นการนำชุดข้อมูลด้านเกรดเฉลี่ยในแต่ละภาคการศึกษาและเกรดเฉลี่ยรวมมาปรับสมดุลให้กับคลาสเป้าหมายด้วยเทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) หลังจากนั้นจึงนำชุดข้อมูลไปสร้างโมเดลการทำนายทั้ง 5 แบบ ผลการประเมินแสดงไว้ในตาราง 4-53 เกรดเฉลี่ยในภาคการศึกษาที่ 5 (GPA5) ให้ค่าความแม่นยำสูงที่สุด คือ 77.8% โดยที่ทุกโมเดลให้ผลลัพธ์เท่ากัน รองลงมาคือเกรดเฉลี่ยในภาคการศึกษาที่ 6 (GPA6) และเกรดเฉลี่ยในภาคการศึกษาที่ 4 (GPA4) ที่มีค่าความแม่นยำเท่ากับ 76% และ 74% ตามลำดับ สังเกตได้ว่า ครั้งนี้ผลการประเมินที่ดีกว่าเป็นผลมาจากตัวแปรเกรดเฉลี่ยในแต่ละภาคการศึกษา ไม่ใช่เกรดเฉลี่ยรวมเหมือนกับการใช้ชุดข้อมูลดั้งเดิม แสดงว่า เทคนิค SMOTE ช่วยให้เกิดผลลัพธ์ที่แตกต่างออกไป แต่ในภาพรวมยังคงเหมือนเดิมคือ ผลการประเมินค่อย ๆ เพิ่มขึ้นเมื่อเข้าใกล้จุดสิ้นสุดของการศึกษา นั่นเพราะยิ่งมีจำนวนภาคการศึกษาเพิ่มขึ้นมากเท่าไหร่ เกรดเฉลี่ยรวมก็ยิ่งเข้าใกล้ผลลัพธ์เป้าหมายมากขึ้นเท่านั้น สำหรับภาคเรียนที่ประเมินได้น้อยที่สุดยังคงเป็นเกรดเฉลี่ยจากภาคเรียนแรกเนื่องด้วยยังเหลือระยะเวลายาวนานกว่าจะสิ้นสุดการศึกษา ส่งผลให้โมเดลมีความแม่นยำค่อนข้างต่ำ

ตาราง 4-53

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านเกรดเฉลี่ยด้วยโมเดลที่แตกต่างกัน (SMOTE)

GPA	Accuracy (%)				
	KNN	NB	DT	SVM	ANN
GPA1	57.5	57.5	57.5	56.0	59.5
GPA2	58.0	58.0	58.0	58.0	58.0
GPA3	63.5	63.5	63.5	63.5	62.5
GPA4	74.0	74.0	74.0	74.0	74.0
GPA5	**77.8**	**77.8**	**77.8**	**77.8**	**77.8**
GPA6	76.0	76.0	76.0	76.0	76.0
AGPA2	62.0	62.0	62.0	62.0	62.0
AGPA3	62.8	63.5	63.5	63.5	63.5
AGPA4	68.5	68.5	68.5	68.5	68.5
AGPA5	68.8	68.8	68.8	68.8	68.8
AGPA6	71.0	71.0	71.0	71.0	71.0

ตาราง 4-54

การเปรียบเทียบผลลัพธ์จากข้อมูลด้านเกรดเฉลี่ยด้วยโมเดลที่แตกต่างกันแบบเพิ่มขึ้น (SMOTE)

GPA	Accuracy (%)				
	KNN	NB	DT	SVM	ANN
GPA1	57.5	57.5	57.5	56.0	59.5
GPA2	63.8	65.5	64.5	64.3	65.0
GPA3	74.0	75.3	76.5	77.5	76.3
GPA4	81.0	83.5	84.3	80.0	83.5
GPA5	83.5	86.5	82.3	84.5	81.5
GPA6	87.3	**90.5**	86.0	89.8	87.3

การวิเคราะห์ด้วยการเพิ่มข้อมูลด้านเกรดเฉลี่ย (GPA) จากเกรดเฉลี่ยในภาคการศึกษาแรก (GPA1) ไปจนถึงภาคการศึกษาที่ 6 (GPA6) แสดงไว้ในตาราง 4-54 ผลลัพธ์จากการเพิ่มข้อมูลเกรดเฉลี่ยของภาคการศึกษาที่ติดต่อกันจะช่วยเพิ่มความแม่นยำให้กับโมเดลสูงขึ้นเรื่อย ๆ และผลลัพธ์ที่ดีที่สุดคือ เกรดเฉลี่ยของทั้ง 6 ภาคการศึกษา ที่มีความแม่นยำสูงถึง 90.5% ผลลัพธ์เหล่านี้มีรูปแบบที่คล้ายกัน นั่นคือ ความแม่นยำของโมเดลการทำนายสูงขึ้นเมื่อมีข้อมูลด้านเกรดเฉลี่ยที่เข้าใกล้จุดสิ้นสุดของการศึกษามากขึ้น การประยุกต์ใช้เทคนิค SMOTE เพิ่มประสิทธิการทำนายให้กับโมเดล ซึ่งจะเห็นจากค่าความแม่นยำที่สูงขึ้นในทุกกรณี เมื่อพิจารณาจากข้อมูลในช่วงสองปีแรกของการศึกษา นั่นคือ เกรดเฉลี่ยในภาคการศึกษาแรก (GPA1) ไปจนถึงภาคการศึกษาที่ 4 (GPA4) ให้ผลการทำนายที่ค่อนข้างแม่นยำด้วยโมเดลต้นไม้ตัดสินใจ (DT) ที่ให้มีค่าความแม่นยำเท่ากับ 84.3% ซึ่งเพียงพอต่อการนำมาจำแนกประเภทของนิสิต อย่างไรก็ตาม หากมีข้อมูลปีที่สามประกอบด้วยก็จะส่งผลให้การทำนายแม่นยำยิ่งขึ้น

การวิเคราะห์ผลการจัดลำดับความสำคัญของตัวแปรด้วยชุดข้อมูลด้านเกรดเฉลี่ย โดยใช้เทคนิคการคำนวณหาค่าเกนความรู้ (Information Gain: IG) จากตาราง 4-55 พบว่า เกรดเฉลี่ยรวมของ 6 ภาคการศึกษา (AGPA 6) ให้ผลลัพธ์สูงที่สุดเมื่อเทียบกับตัวแปรอื่น โดยมีค่าเกนความรู้เท่ากับ 1.188 ลำดับที่ 2 คือ เกรดเฉลี่ยของภาคการศึกษาที่ 5 (GPA5) ซึ่งมีค่าเกนความรู้เป็น 1.125 ถัดมาเป็น เกรดเฉลี่ยรวมของ 5 ภาคการศึกษา (AGPA5) และเกรดเฉลี่ยภาคการศึกษาที่ 6 (GPA6) ด้วยค่าเกนความรู้เป็น 1.108 และ 1.097 ตามลำดับ จะเห็นว่า ตัวแปรยิ่งเข้าใกล้จุดสิ้นสุดของการศึกษามากเท่าไหร่ความสำคัญก็ยิ่งเพิ่มขึ้น หากแต่วัตถุประสงค์ของงานวิจัยนี้คือ ต้องการทราบล่วงหน้าว่าผลสัมฤทธิ์ทางการศึกษาของนิสิตจะเป็นอย่างไรเมื่อสำเร็จการศึกษาแล้ว ดังนั้น จึงต้องอาศัยข้อมูลในระยะแรกถึงระยะกลางของการศึกษา ซึ่งผลการทำนายที่ดีที่สุดใน 2 ปีแรก ต้องใช้ข้อมูลเกรดเฉลี่ยรวมของ 4 ภาคการศึกษา (AGPA4) ที่มีค่าเกนความรู้เท่ากับ 1.079 ซึ่งจัดว่าเป็นผลลัพธ์ที่สูงพอสมควร ด้วยเหตุนี้จึงสามารถนำข้อมูลใน 2 ปีแรก ไปใช้เพื่อกำหนดรูปแบบการสนับสนุนการศึกษาให้กับนิสิตต่อไปได้ในอีก 2 ปีที่เหลือของหลักสูตร

ตาราง 4-55

ลำดับความสำคัญของตัวแปรด้านเกรดเฉลี่ยด้วยค่าเกนความรู้ (SMOTE)

Order	Variables	Normalized Weight	Information Gain
1	AGPA6	1.00	1.188
2	GPA5	0.95	1.125
3	AGPA5	0.93	1.108
4	GPA6	0.92	1.097
5	AGPA4	0.91	1.079
6	GPA4	0.85	1.009
7	GPA3	0.84	0.992
8	AGPA3	0.76	0.906
9	AGPA2	0.7	0.837
10	GPA1	0.69	0.825
11	GPA2	0.65	0.768

ผลการวิจัยทั้งหมดแบ่งชุดข้อมูลออกเป็นสองรูปแบบ คือ แบบที่เป็นชุดข้อมูลดั้งเดิมและแบบที่เป็นชุดข้อมูลที่ผ่านการปรับสมดุลแล้ว สำหรับชุดข้อมูลดั้งเดิม ตัวแปรที่มีผลต่อการทำนายดีที่สุด ได้แก่ เกรดเฉลี่ยของภาคการศึกษาที่ 4 (GPA4) โดยมีค่าความแม่นยำเท่ากับ 76.8% และถ้านำข้อมูลเกรดเฉลี่ยของภาคการศึกษาที่ 1 ถึง 3 มาร่วมวิเคราะห์ด้วย จะยิ่งส่งผลให้มีค่าความแม่นยำที่สูงขึ้นไปเป็น 79.1% โดยใช้โมเดลแบบเบส์อย่างง่าย (NB) แต่ถ้าใช้ชุดข้อมูลที่ปรับสมดุลแล้ว ตัวแปรที่มีผลต่อการทำนายดีที่สุด คือ เกรดเฉลี่ยของภาคการศึกษาที่ 5 (GPA5) ด้วยค่าความแม่นยำเท่ากับ 77.8% ซึ่งจะเข้าสู่ครึ่งหลังของหลักสูตร ดังนั้น ถ้าต้องการจำแนกประเภทนิสิตในช่วง 2 ปีแรก ต้องใช้ เกรดเฉลี่ยของภาคการศึกษาที่ 4 (GPA4) เป็นตัวแปรหลัก ซึ่งมีค่าความแม่นยำเท่ากับ 74% และถ้าได้ข้อมูลเกรดเฉลี่ยของภาคการศึกษาที่ 1 ถึง 3 มาร่วมวิเคราะห์ด้วย ค่าความแม่นยำจะเพิ่มขึ้นไปเป็น 84.3% โดยใช้โมเดลต้นไม้ตัดสินใจ (DT) จะเห็นว่า ชุดข้อมูลที่ปรับสมดุลแล้วให้ผลลัพธ์ดีกว่าชุดข้อมูลดั้งเดิม สรุปได้ว่า เกรดเฉลี่ยสะสมจากสองปีแรกจะให้ข้อมูลที่เพียงพอในการจำแนกประเภทนิสิตว่าจะมีผลสัมฤทธิ์ทางการเรียนอยู่ในกลุ่มใด และการใช้เทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) ช่วยให้ผลการทำนายดีขึ้น ผลลัพธ์จากกระบวนการค้นพบความรู้นี้จะเป็นประโยชน์ต่อผู้รับผิดชอบหลักสูตรหรือคณาจารย์ในภาคฯ สำหรับการออกแบบการให้ความช่วยเหลือ การตัดสินใจ หรือการวางแผนสนับสนุนการศึกษาให้กับนิสิตได้อย่างเหมาะสม รวมถึงเป็นปัจจัยสำคัญที่จะช่วยให้การปรับปรุงหลักสูตรมีประสิทธิภาพดียิ่งขึ้น

บทที่ 5
อภิปรายและสรุปผลการวิจัย

อภิปรายผล

การทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตในระยะแรกของการศึกษาสามารถช่วยแก้ปัญหาผลการเรียนตกต่ำให้กับนิสิตที่มีความเสี่ยงสูงได้อย่างทันท่วงที การช่วยเหลือนิสิต อย่างเช่น การแนะนำเกี่ยวกับหลักสูตร การปฏิบัติตัวในมหาวิทยาลัย หรือการฝึกอบรมพิเศษ จะช่วยเพิ่มอัตราความสำเร็จทางการศึกษาของนิสิตได้เป็นอย่างดี ดังนั้นการนำเทคนิคการทำเหมืองข้อมูลทางการศึกษาจึงมีความสำคัญในการพัฒนาโมเดลการทำนายผลสัมฤทธิ์ทางการเรียนเพื่อปรับปรุงความสำเร็จของนิสิตเมื่อสิ้นสุดการศึกษาในมหาวิทยาลัยแล้ว

งานวิจัยนี้ได้มีการศึกษาปัจจัยที่ใช้เป็นตัวแปรสำหรับการสร้างโมเดลด้วยอัลกอริทึมการเรียนรู้ของเครื่องที่หลากหลายและได้รับความนิยม ที่ซึ่งระบุไว้ในวรรณกรรมที่เกี่ยวข้องตั้งแต่อดีตถึงปัจจุบัน ทั้งนี้ก็เพื่อทำนายความสำเร็จทางวิชาการของผู้เรียนซึ่งวัดได้จากผลสัมฤทธิ์ทางการเรียน โดยทั่วไปมักนิยมนำปัจจัยด้านคะแนนหรือเกรดมาใช้ในการวิเคราะห์ผล อย่างไรก็ตามคุณลักษณะด้านประชากรของผู้เรียน ไม่ว่าจะเป็น เพศ อายุ เชื้อชาติ ข้อมูลผู้ปกครอง หรือจำนวนพี่น้อง ก็อาจมีส่วนสำคัญเช่นกัน

แม้ว่าในปัจจุบันจะมีเทคนิคการทำเหมืองข้อมูลและซอฟต์แวร์เฉพาะให้เลือกใช้อย่างมากมาย แต่ก็ยังเป็นเรื่องท้าทายสำหรับนักวิจัยมือใหม่ในการนำเครื่องมือเหล่านั้นไปประยุกต์ใช้ได้อย่างเหมาะสม การวิเคราะห์ข้อมูลทางการศึกษาทำได้หลากหลายแนวทาง ทั้งการจำแนกประเภทของชุดข้อมูล หรือการจัดกลุ่มไปจนถึงการค้นหาความสัมพันธ์ระหว่างชุดข้อมูล อีกทั้งยังมีการนำอัลกอริทึมการทำเหมืองข้อมูลที่มีอยู่เป็นจำนวนมากมาใช้ในสร้างแบบจำลองเพื่อทำการทดลองด้วยเงื่อนไขที่หลากหลาย แต่การวิเคราะห์ที่ได้รับความนิยมสูงที่สุดคือ การทำนายผลสำเร็จทางวิชาการเมื่อสิ้นสุดการศึกษาแล้ว

งานวิจัยนี้ได้นำโมเดลการจำแนกประเภท 5 โมเดล ได้แก่ การค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (K-Nearest Neighbor: KNN) วิธีแบบเบส์อย่างง่าย (Naïve Bayes: NB) ต้นไม้ตัดสินใจ (Decision Tree: DT) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และ โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เพื่อวิเคราะห์ผลสัมฤทธิ์ทางการศึกษาของนิสิตภาควิชาสารสนเทศศึกษาที่สำเร็จการศึกษาไปแล้วระหว่างปีการศึกษา 2562 ถึง 2565 จำนวน 275 คน โดยใช้กระบวนการการทำเหมืองข้อมูลทางการศึกษา การวิเคราะห์ผลงานวิจัยได้ใช้ชุดข้อมูล 3 ชุด ประกอบไปด้วย ชุดข้อมูลด้านประชากรของนิสิต ชุดข้อมูลเกรดที่นิสิตได้รับในแต่ละรายวิชา และชุดข้อมูลเกรดเฉลี่ยของนิสิตในแต่ละภาคการศึกษา ทั้งนี้เพื่อค้นหาสัมพันธ์ระหว่างชุดข้อมูลในแต่ละกลุ่มกับผลสัมฤทธิ์ทางการเรียนเมื่อนิสิตสำเร็จการศึกษาแล้ว นอกจากนี้ ผู้วิจัยยังค้นหาความสำคัญของแต่ละตัวแปรในชุดข้อมูลที่มีความเชื่อมโยงกับข้อมูลเป้าหมายโดยใช้วิธีเปรียบเทียบค่าของแต่ละตัวแปรด้วยค่าเกนความรู้ (Information Gain: IG) อีกทั้งยังได้ปรับจำนวนข้อมูลเป้าหมายที่ไม่สมดุลกันให้มีความสมดุลกันด้วยเทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์

(SMOTE) และได้นำชุดข้อมูลที่ปรับสมดุลแล้วมาผ่านกระบวนการการทำเหมืองข้อมูลเพื่อเปรียบเทียบกับชุดข้อมูลที่ยังไม่ได้ปรับสมดุล

งานวิจัยนี้ช่วยยืนยันว่าการสร้างโมเดลการทำนายผลสัมฤทธิ์ทางการศึกษาของนิสิตในช่วงครึ่งแรกของการศึกษาสามารถทำได้อย่างมีประสิทธิภาพ โดยมีค่าความแม่นยำสูงถึงร้อยละ 84.3 ด้วยโมเดลต้นไม้ตัดสินใจ (DT) ที่อาศัยชุดข้อมูลด้านเกรดเฉลี่ยที่ผ่านการปรับสมดุลแล้วมาทำการประเมินผล เมื่ออาจารย์สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้เช่นนี้ อาจารย์ก็สามารถวางแผนการสนับสนุนการศึกษาให้กับนิสิตที่มีผลการเรียนในระดับต่าง ๆ ได้อย่างเหมาะสม ผู้วิจัยเห็นว่า โมเดลการทำนายนี้จะช่วยให้อาจารย์สามารถตัดสินใจได้อย่างทันท่วงทีในการให้ความช่วยเหลือกลุ่มนิสิตที่มีแนวโน้มว่าจะสำเร็จการศึกษาด้วยเกรดเฉลี่ยที่ค่อนข้างต่ำ และช่วยส่งเสริมให้นิสิตที่มีแนวโน้มจะสำเร็จการศึกษาด้วยเกรดเฉลี่ยที่ดีให้ได้รับผลการเรียนสูงขึ้นกว่าที่คาดการณ์ไว้ การวางแผนสนับสนุนการศึกษาลักษณะนี้จะเป็นการจัดการแบบเชิงรุกและกระทำได้ทันต่อเหตุการณ์ ทั้งนี้ก็เพื่อเพิ่มคุณภาพของนิสิตก่อนเรียนจบหลักสูตรที่ซึ่งเป็นเป้าหมายหลักของมหาวิทยาลัยในการผลิตบัณฑิตที่มีคุณภาพออกสู่สังคมอย่างยั่งยืน

ในด้านวิชาการ รายวิชาต่าง ๆ ในหลักสูตรก็มีผลสะท้อนให้เห็นถึงลำดับความสำคัญของรายวิชาที่มีต่อผลสัมฤทธิ์ทางการเรียน โดยเฉพาะอย่างยิ่งเมื่อใช้ชุดข้อมูลที่ปรับสมดุลแล้ว กลุ่มรายวิชาที่เป็นปัจจัยสำคัญต่อความสำเร็จคือ กลุ่มรายวิชาเอกบังคับ ที่มีค่าความแม่นยำร้อยละ 87.8 โดยใช้อัลกอริทึมแบบเบส์อย่างง่าย (NB) สำหรับการสร้างโมเดลการทำนาย เมื่อผู้รับผิดชอบหลักสูตรหรือคณาจารย์สามารถคาดการณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตได้ตั้งแต่เนิ่น ๆ ก็จะช่วยให้สามารถวางแผนหรือออกมาตรการจัดการต่าง ๆ ที่เพิ่มประสิทธิผลให้กับการศึกษา ตัวอย่างเช่น การบริการทางด้านการให้คำปรึกษาทางจิตวิทยาแก่นิสิต การสร้างระบบการกวดวิชาหรือสอนพิเศษให้กับนิสิตที่มีผลการเรียนต่ำ การสร้างระบบสนับสนุนระหว่างเพื่อนนิสิตด้วยกันให้ชักจูงกันไปในทางที่ดี การฝึกอบรมเพิ่มเติมในบางหัวข้อของรายวิชาสำคัญ หรือการปรับปรุงบทเรียนให้มีความร่วมสมัยและเหมาะสมกับความสนใจของนิสิตในปัจจุบัน เป็นต้น

ในอีกด้านหนึ่ง วิธีการสนับสนุนการศึกษาที่กล่าวมาข้างต้นนอกจากจะช่วยพัฒนาสติปัญญาของกลุ่มนิสิตที่มีผลการเรียนต่ำแล้ว ยังช่วยสนับสนุนหรือสร้างแรงจูงใจพิเศษเพื่อเสริมสร้างศักยภาพให้กลุ่มนิสิตที่มีผลการเรียนดีอีกด้วย ตัวอย่างเช่น การสร้างแรงบันดาลใจในการศึกษาแห่งอนาคต หรือการจัดทำโครงการด้านเกียรตินิยมเพื่อดึงดูดความสามารถของนิสิตในกลุ่มเรียนดีให้พัฒนาทักษะให้เป็นที่ประจักษ์ต่อสังคม เป็นต้น โครงการเหล่านี้จะช่วยให้มหาวิทยาลัยแข่งขันกับสถาบันอื่น ๆ ได้ อีกทั้งยังดึงดูดผู้เรียนที่มีพรสวรรค์ และผู้มีพรสวรรค์เหล่านั้นจะเป็นกำลังสำคัญในการสร้างชื่อเสียงให้กับทางมหาวิทยาลัยและเสริมสร้างความแข็งแกร่งทางวิชาการได้อีกด้วย

สรุปประเด็นสำคัญจากงานวิจัย

งานวิจัยนี้มุ่งเน้นการวิเคราะห์ข้อมูลด้วยการทำนายผลสัมฤทธิ์ทางการเรียนในระดับปริญญาตรีโดยใช้อัลกอริทึมที่ได้รับความนิยม 5 อัลกอริทึม เพื่อทำการประเมินผลลัพธ์อย่างเป็นระบบ งานวิจัยนี้วัดประสิทธิภาพของโมเดลการจำแนกประเภทด้วยตารางเมทริกซ์ความสับสนซึ่งเป็นวิธีมาตรฐานสำหรับการประเมินผลของโมเดลลักษณะนี้ โดยพิจารณาจากค่าที่อยู่ในตารางเมทริกซ์ความสับสน มาตรวัดที่ผู้วิจัยให้ค่าน้ำหนักมากที่สุดคือ ค่าความแม่นยำ (Accuracy) ซึ่งได้ถูกนำมาใช้เปรียบเทียบและวิเคราะห์ในหลากหลายชุดข้อมูล การคาดคะเนผลการเรียนของนิสิตตั้งแต่เนิ่น ๆ สามารถช่วยให้ภาควิชาสามารถวางแผนและปรับปรุงผลสัมฤทธิ์ทางการเรียนของนิสิตได้อย่างเหมาะสม ด้วยเทคนิคการทำเหมืองข้อมูลทางการศึกษา ซึ่งมีความเกี่ยวข้องกับศาสตร์ด้านเทคโนโลยีสารสนเทศอยู่หลากหลายแขนงด้วยกัน ได้แก่ สถิติศาสตร์ ศาสตร์ด้านการทำเหมืองข้อมูล ศาสตร์ด้านการเรียนรู้ของเครื่อง และศาสตร์ด้านปัญญาประดิษฐ์ ผลลัพธ์ที่ได้สามารถนำไปใช้ประกอบการออกแบบหลักสูตรสารสนเทศศึกษาในอนาคต เพื่อให้สอดคล้องกับมาตรฐานของเครือข่ายการประกันคุณภาพมหาวิทยาลัยอาเซียน (AUN-QA version 4) ได้อีกด้วย ผลลัพธ์ที่ได้จากงานวิจัยสามารถสรุปเป็นประเด็นสำคัญได้ดังต่อไปนี้

1. ข้อมูลด้านประชากรมีผลการประเมินในระดับปานกลางเท่านั้น โดยมีค่าความแม่นยำสูงสุดเพียงร้อยละ 58 จากการสร้างโมเดลการจำแนกประเภทด้วยอัลกอริทึมโครงข่ายประสาทเทียม (ANN) รองลงมาคือผลการประเมินจากโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยมีค่าความแม่นยำเท่ากับร้อยละ 57.5 เมื่อพิจารณาเฉพาะปัจจัยหรือตัวแปรที่มีผลต่อการทำนายพบว่า เกรดเฉลี่ยจากโรงเรียนมัธยม (SGPA) มีน้ำหนักมากที่สุดและห่างจากปัจจัยลำดับที่สองเป็นอย่างมาก นั่นคือ มีค่าเกนความรู้เท่ากับ 0.448 ส่วนลำดับที่สองคืออาชีพของบิดาโดยมีค่าเกนความรู้เพียง 0.178 เท่านั้น ปัจจัยอื่น ๆ ด้านประชากรให้ผลลัพธ์ของการประเมินที่น้อยลงไปอีก ดังนั้น จากผลการประเมินที่ได้สามารถสรุปได้ ชุดข้อมูลด้านประชากรของนิสิตแทบไม่มีผลต่อผลสัมฤทธิ์ทางการเรียนของนิสิตในมหาวิทยาลัยเลย ยกเว้นเพียงเกรดเฉลี่ยจากโรงเรียนมัธยมเท่านั้นที่พอจะนำมาใช้ประกอบการพิจารณาผลสัมฤทธิ์ทางการเรียนได้ ส่วนปัจจัยที่มีน้ำหนักต่อการประเมินน้อยที่สุดได้แก่ เพศของนิสิต รองลงมาคือ สถานการมีพี่น้อง ซึ่งมีค่าเกนความรู้เพียง 0.019 และ 0.047 ตามลำดับ ในขณะที่ปัจจัยอื่น ๆ เช่น อาชีพหรือรายได้ของบิดามารดา หรือ ศาสนาที่นับถือ ล้วนมีผลการประเมินที่ต่ำมาก ดังนั้น การที่นิสิตเป็นเพศอะไร มีพี่น้องหรือไม่ นับถือศาสนาใด หรือบิดามารดาทำอาชีพอะไร ล้วนส่งผลต่อผลสัมฤทธิ์ทางการศึกษาของนิสิตคนนั้นอยู่ในเกณฑ์ที่ต่ำมาก

2. ข้อมูลด้านเกรดของแต่ละรายวิชามีผลต่อผลสัมฤทธิ์ทางการศึกษาสูงกว่าข้อมูลด้านประชากรอย่างมีนัยยะสำคัญ หากนำรายวิชาของนิสิตทั้งหมดที่เรียนไปแล้วยกเว้นรายวิชาฝึกงานมาทำการวิเคราะห์จะได้ค่าความแม่นยำสูงถึงร้อยละ 91.3 ด้วยโมเดลแบบเบส์อย่างง่าย (NB) อย่างไรก็ตาม เนื่องจากวัตถุประสงค์ของการวิจัยมุ่งเน้นไปที่การทำนายผลสัมฤทธิ์ทางการเรียนล่วงหน้าเพื่อที่จะหาแนวทางปรับปรุงผลการเรียนให้กับนิสิตโดยเฉพาะในช่วงสองปีแรกของการศึกษา ดังนั้นการใช้ชุดข้อมูลด้านเกรดของรายวิชาทั้งหมดจึงไม่สอดคล้องในทางปฏิบัติ หากแต่การวิเคราะห์ข้อมูลลักษณะนี้จะช่วยในเรื่องของการปรับปรุงหลักสูตรให้มี

ความเหมาะสมยิ่งขึ้น ซึ่งจะสรุปประเด็นสำคัญที่เกี่ยวกับชุดข้อมูลด้านเกรดของแต่ละกลุ่มรายวิชาในหัวข้อถัดไป

3. ข้อมูลด้านเกรดของกลุ่มรายวิชาศึกษาทั่วไปมีผลการประเมินอยู่ในเกณฑ์ดี โดยมีค่าความแม่นยำร้อยละ 77.3 จากการทดลองด้วยโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) โมเดลที่ให้ผลการประเมินรองลงมาคือโมเดลโครงข่ายประสาทเทียม (ANN) โดยมีค่าความแม่นยำร้อยละ 76.5 หากพิจารณาเฉพาะรายวิชาทีละวิชา พบว่า รายวิชาก้าวทันสังคมดิจิทัลด้วยไอซีที (Moving Forward in a Digital Society with ICT) มีค่าน้ำหนักโดดเด่นกว่ารายวิชาอื่น โดยมีค่าเกนความรู้เท่ากับ 0.748 ในขณะที่รายวิชาลำดับที่สองมีค่าเกนความรู้เพียง 0.554 ซึ่งค่อนข้างห่างจากรายวิชาแรกพอสมควร ดังนั้น ผู้วิจัยจึงให้น้ำหนักกับรายวิชาก้าวทันสังคมดิจิทัลด้วยไอซีทีเพียงวิชาเดียวที่มีความสำคัญเพียงพอที่จะนำมาประกอบการพิจารณาหาผลสัมฤทธิ์ทางการศึกษาของนิสิต ส่วนรายวิชาที่เหลือในกลุ่มรายวิชาศึกษาทั่วไปไม่สามารถนำมาประกอบการวิเคราะห์ข้อมูลเนื่องด้วยมีความสัมพันธ์กับตัวแปรเป้าหมายที่ค่อนข้างต่ำ

4. ข้อมูลด้านเกรดของกลุ่มรายวิชาเอกบังคับมีผลต่อผลสัมฤทธิ์ทางการเรียนอยู่ในระดับดีมาก โดยมีค่าความแม่นยำสูงถึงร้อยละ 87.8 จากการทดลองด้วยโมเดลแบบเบส์อย่างง่าย (NB) โมเดลที่ให้ผลการประเมินในลำดับที่สองคือ โมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าความแม่นยำร้อยละ 84.3 ซึ่งยังไงให้ผลลัพธ์ที่อยู่ในเกณฑ์ดี เมื่อพิจารณาเฉพาะรายวิชาพบว่า รายวิชาการวิจัยและสถิติทางสารสนเทศศึกษา (Research and Statistics in Information Studies) มีค่าเกนความรู้สูงที่สุด คือ 0.989 รองลงมาคือรายวิชาการจัดระบบทรัพยากรสารสนเทศ (Organization of Information Resources) และ รายวิชาบริการสารสนเทศและช่วยการค้นคว้า (Information and Reference Services) ซึ่งมีน้ำหนักความสำคัญเท่ากัน โดยมีค่าเกนความรู้เท่ากับ 0.82 รายวิชาที่เหลือในลำดับรองลงมาก็ยังคงมีผลการประเมินอยู่ในเกณฑ์ดี ดังนั้นข้อมูลชุดนี้จึงมีค่าความแม่นยำสูงที่สุด ที่ซึ่งเป็นผลสืบเนื่องจากจำนวนรายวิชาที่มากอีกทั้งยังเป็นรายวิชาเอกบังคับ ซึ่งเป็นหัวใจสำคัญของหลักสูตรนั่นเอง

5. ข้อมูลด้านเกรดของกลุ่มรายวิชาเอกเลือกให้ผลการประเมินน้อยกว่ากลุ่มรายวิชาเอกบังคับอยู่เล็กน้อยแต่ก็ยังถือว่ามีผลต่อผลสัมฤทธิ์ทางการเรียนอยู่ในระดับดีมากเช่นกัน โดยที่โมเดลที่สร้างจากอัลกอริทึมแบบเบส์อย่างง่าย (NB) ให้ค่าความแม่นยำสูงที่สุด ซึ่งก็คือร้อยละ 84.8 รองลงมาคือผลการทำนายจากโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ด้วยค่าความแม่นยำร้อยละ 81.8 เมื่อพิจารณาจากแต่ละรายวิชาพบว่า รายวิชาการสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (Data Communication and Computer Networks) ให้ค่าความเกนความรู้สูงที่สุดเมื่อเทียบกับรายวิชาอื่น โดยมีค่าเกนความรู้เท่ากับ 0.812 ลำดับที่สองคือ รายวิชาการค้นคืนสารสนเทศอิเล็กทรอนิกส์ (Electronic Information Retrieval) โดยมีค่าเกนความรู้เท่ากับ 0.713 จะเห็นว่าลำดับที่หนึ่งและสองมีค่าเกนความรู้ห่างกันพอสมควร ส่วนรายวิชาในลำดับที่สามลงไปมีค่าน้ำหนักที่น้อยลง ดังนั้น ผลการประเมินในกลุ่มรายวิชาเอกเลือกมีผลต่อผลสัมฤทธิ์ทางการศึกษาของนิสิตน้อยกว่ากลุ่มรายวิชาเอกบังคับและมีระยะห่างของค่าน้ำหนักอยู่พอสมควร ด้วยเหตุนี้ อาจารย์จึงควรให้ความสำคัญกับรายวิชาในกลุ่มเอกบังคับมากกว่ากลุ่มรายวิชาอื่น ๆ เพื่อการปรับปรุงหลักสูตรใหม่ที่ดีขึ้น

6. เมื่อพิจารณาข้อมูลในกลุ่มรายวิชาด้านสารสนเทศศาสตร์และเทคโนโลยีสารสนเทศ พบว่า ในส่วนที่เป็นรายวิชาเอกบังคับ กลุ่มรายวิชาด้านสารสนเทศศาสตร์มีผลการทำนายที่แม่นยำมากกว่ากลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศ โดยมีค่าความแม่นยำร้อยละ 85.8 และร้อยละ 83 ตามลำดับ ซึ่งยังคงไม่ห่างกันมากนักด้วยโมเดลเดียวกันคือ โมเดลที่สร้างจากอัลกอริทึมแบบเบส์อย่างง่าย (NB) ในทางกลับกัน ส่วนที่เป็นรายวิชาเอกเลือกกลับพบว่า กลุ่มรายวิชาด้านเทคโนโลยีสารสนเทศให้ผลลัพธ์ที่ดีกว่ากลุ่มรายวิชาด้านสารสนเทศศาสตร์อยู่เล็กน้อย นั่นคือ มีค่าความแม่นยำร้อยละ 80 และ 76.3 ด้วยโมเดลแบบเบส์อย่างง่าย (NB) และโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ตามลำดับ หากพิจารณาในภาพรวมสรุปได้ว่า กลุ่มรายวิชาทั้งสองกลุ่มมีผลการประเมินที่ใกล้เคียงกัน นั่นอาจเป็นผลมากจากการออกแบบหลักสูตรที่ให้น้ำหนักในกลุ่มรายวิชาทั้งด้านสารสนเทศศาสตร์และด้านเทคโนโลยีสารสนเทศอย่างเท่าเทียมกัน

7. ข้อมูลในกลุ่มของเกรดเฉลี่ยให้ผลลัพธ์อยู่ในเกณฑ์ดีมาก อีกทั้งยังเป็นชุดข้อมูลที่นำมาใช้ประโยชน์ได้ดีสำหรับการพยากรณ์ผลการเรียนในอนาคต หากได้ข้อมูลเกรดเฉลี่ยในสามปีแรกจะมีค่าความแม่นยำของการทำนายอยู่ที่ร้อยละ 90.5 ด้วยโมเดลแบบเบส์อย่างง่าย (NB) โมเดลในลำดับที่สองก็ยังคงให้ค่าความแม่นยำสูงถึงร้อยละ 89.8 จากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (SVM) แต่ถ้าพิจารณาเฉพาะข้อมูลในสองปีแรกของการศึกษา ค่าความแม่นยำจะลดลงมาที่ร้อยละ 84.3 ด้วยโมเดลต้นไม้ตัดสินใจ (DT) ถัดมาเป็นโมเดลแบบเบส์อย่างง่าย (NB) และโมเดลโครงข่ายประสาทเทียม (ANN) ที่มีค่าความแม่นเท่ากันคือร้อยละ 83.5 ข้อมูลชุดนี้สามารถนำมาใช้ทำนายได้จริงเนื่องจากเป็นชุดข้อมูลในสองปีแรก ซึ่งต่างจากชุดข้อมูลเกรดของแต่ละรายวิชาที่ซึ่งนิสิตอาจลงทะเบียนในภาคการศึกษาที่แตกต่างกัน และบางรายวิชาลงทะเบียนเรียนในช่วงท้ายของการศึกษาก่อนจะออกไปฝึกงานจึงใช้ประโยชน์จริงได้น้อย สำหรับกรณีการพิจารณาเฉพาะเกรดเฉลี่ยในแต่ละภาคการศึกษา เกรดเฉลี่ยในภาคการศึกษาที่ 4 (GPA4) มีผลลัพธ์เป็นที่น่าพอใจโดยมีค่าเกนความรู้เท่ากับ 1.009 และถ้านำเกรดเฉลี่ยสะสมใน 4 ภาคการศึกษาแรกมาวิเคราะห์ ค่าเกนความรู้ที่ได้ยิ่งสูงขึ้น แต่ที่น่าสังเกตคือ ค่าความแม่นยำจากเกรดเฉลี่ยสะสมกลับมีค่าน้อยกว่าค่าความแม่นยำของเกรดเฉลี่ยในแต่ละภาคการศึกษา นั่นเพราะเกรดเฉลี่ยในช่วงปีแรกยังไม่อาจคาดการณ์ผลสัมฤทธิ์ทางการศึกษาได้ดีนัก และเมื่อนำให้คำนวณร่วมกับข้อมูลในปีที่สองยิ่งทำให้ความแม่นยำในการทำนายลดลง การที่ผลการเรียนในปีแรกมีผลต่อการทำนายไม่มากนักอาจเป็นเพราะการปรับตัวในระยะแรกของการเป็นนิสิตในรั่วมหาวิทยาลัยนั่นเอง

8. การศึกษาประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการศึกษาด้วยโมเดลทั้ง 5 โมเดล ให้ผลลัพธ์ที่แตกต่างกัน บางโมเดลให้ผลลัพธ์ดีในชุดข้อมูลหนึ่ง แต่ไม่ดีในชุดข้อมูลอื่น ดังนั้นจึงควรทดสอบกับโมเดลที่หลากหลาย เพื่อค้นหาแนวทางการทำนายผลสัมฤทธิ์ทางการเรียนที่ดีที่สุด เมื่อพิจารณาโมเดลที่ให้ค่าความแม่นยำสูงสุดบ่อยครั้ง พบว่า อัลกอริทึมแบบเบส์อย่างง่าย (NB) ให้ผลลัพธ์ดีที่สุดในหลายชุดข้อมูล ในขณะที่โมเดลโครงข่ายประสาทเทียม (ANN) และโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) ให้ผลลัพธ์ที่ดีรองจากโมเดลแบบเบส์อย่างง่าย (NB) อยู่เล็กน้อย ส่วนโมเดลต้นไม้ตัดสินใจ (DT) และโมเดลการค้นหาเพื่อนบ้านที่ใกล้ที่สุด K ตัว (KNN) ให้ผลการทำนายที่มีความแม่นยำน้อยกว่าโมเดลที่กล่าวไว้ข้างต้น ดังนั้น หากต้องเลือกโมเดลการ

ทำนายมาหนึ่งโมเดล โมเดลแบบเบส์อย่างง่าย (NB) จึงเป็นตัวเลือกที่ดีที่สุดสำหรับการทำนายผลสัมฤทธิ์ทางการศึกษาสำหรับชุดข้อมูลในกรณีศึกษานี้

9. ชุดข้อมูลในทุกกลุ่มข้อมูลมีจำนวนคลาสเป้าหมายที่ไม่เท่ากันผลการศึกษาในระดับพอใช้ (Fair) มีจำนวนระเบียนน้อยที่สุด ในขณะที่ผลการศึกษาในระดับดีมาก (Very Good) มีจำนวนระเบียนมากที่สุด เมื่อนำชุดข้อมูลที่ไม่สมดุลกันมาทำการทดลองจึงอาจให้ผลลัพธ์ที่คลาดเคลื่อนจากความเป็นจริงได้ ดังนั้นงานวิจัยนี้จึงได้นำชุดข้อมูลมาปรับจำนวนระเบียนให้คลาสเป้าหมายมีความสมดุลกันด้วยเทคนิคการสุ่มตัวอย่างเกินของชนกลุ่มน้อยสังเคราะห์ (SMOTE) แล้วจึงทำการทดลองกับชุดข้อมูลกลุ่มต่าง ๆ และเปรียบเทียบกับผลที่ได้จากชุดข้อมูลที่ยังไม่ได้ปรับสมดุล ผลการวิจัยพบว่า ชุดข้อมูลที่ได้รับการปรับสมดุลแล้วให้ค่าความแม่นยำสูงกว่าในทุกกลุ่มข้อมูล ดังนั้น การวิเคราะห์ผลสัมฤทธิ์ทางการศึกษาควรใช้ชุดข้อมูลที่ผ่านการปรับสมดุลแล้วจึงจะให้ผลลัพธ์ที่ดีที่สุดในทุกกรณี

ข้อจำกัดและงานวิจัยในอนาคต

งานวิจัยนี้มีข้อจำกัดบางประการ ได้แก่ ชุดข้อมูลยังมีจำนวนระเบียนข้อมูลไม่มากนักและจำนวนระเบียนในแต่ละคลาสก็ไม่สมดุลกัน รวมถึงรายวิชาที่เรียนของนิสิตแต่ละคนไม่เหมือนกันโดยเฉพาะในกลุ่มของรายวิชาศึกษาทั่วไปและรายวิชาเอกเลือก ส่งผลให้การทำนายอาจมีความคลาดเคลื่อนไปบ้าง ในอนาคตอาจมีการรวบรวมข้อมูลให้มากขึ้น และใช้ชุดข้อมูลที่มีรายวิชาตรงกันเพื่อศึกษาเชิงลึกต่อไป

โมเดลการทำนายที่พัฒนาขึ้นในงานวิจัยนี้ได้ถูกทดลองเฉพาะกับกลุ่มนิสิตของภาควิชาสารสนเทศศึกษา ซึ่งมีลักษณะเฉพาะด้านโครงสร้างหลักสูตร ความถนัดพื้นฐาน และแนวทางการเรียนรู้บางประการ การจำกัดกลุ่มตัวอย่างเช่นนี้อาจส่งผลต่อการสรุปผลในเชิงทั่วไปของโมเดล เนื่องจากไม่สามารถสะท้อนความหลากหลายของผู้เรียนในระดับสถาบันหรือระดับสาขาวิชาอื่นได้อย่างครบถ้วน

การขยายขอบเขตการศึกษาไปยังนิสิตจากหลักสูตรอื่น ๆ ที่มีความแตกต่างกันทั้งในเชิงวิชาการ (เช่น หลักสูตรด้านวิทยาศาสตร์ วิศวกรรมศาสตร์ หรือศิลปศาสตร์) และเชิงประชากรศาสตร์ (เช่น เพศ อายุ หรือประสบการณ์การเรียนรู้ก่อนหน้า) จะช่วยให้สามารถประเมินประสิทธิภาพของโมเดลในบริบทที่กว้างขวางมากขึ้น อีกทั้งยังอาจนำไปสู่ข้อค้นพบใหม่ที่สะท้อนถึงปัจจัยเฉพาะของแต่ละกลุ่ม ซึ่งผู้เรียนที่มีพื้นฐานทางด้านศิลปะศาสตร์อาจส่งผลต่อผลการเรียนรู้ในรายวิชาด้านเทคโนโลยีสารสนเทศ ตัวอย่างเช่น ผู้เรียนที่มีพื้นฐานทางด้านศิลปศาสตร์อาจพบความท้าทายในการเรียนรู้เนื้อหาด้านเทคโนโลยี ซึ่งมักต้องอาศัยความเข้าใจเชิงตรรกะ กระบวนการเชิงคำนวณ และทฤษฎีที่เป็นนามธรรมมากขึ้น ส่งผลให้ผลการเรียนอาจมีแนวโน้มลดลงเมื่อเทียบกับผู้เรียนที่มีพื้นฐานด้านวิทยาศาสตร์หรือวิศวกรรมศาสตร์ ซึ่งคุ้นเคยกับลักษณะการเรียนรู้แบบดังกล่าว การทำความเข้าใจในประเด็นเหล่านี้จะเป็นประโยชน์ต่อการออกแบบกลยุทธ์การเรียนรู้เฉพาะกลุ่มและการสนับสนุนการศึกษาที่เหมาะสมกับความต้องการของผู้เรียนแต่ละประเภทอย่างมีประสิทธิภาพ

ในการวิจัยครั้งนี้ การเก็บรวบรวมข้อมูลอาศัยข้อมูลเชิงวิชาการจากหลักสูตรสารสนเทศศึกษา ร่วมกับข้อมูลจากระบบบริการการศึกษาซึ่งส่วนใหญ่เป็นข้อมูลเชิงปริมาณ เช่น คะแนนผลการเรียนในแต่ละรายวิชาหรือโครงสร้างการลงทะเบียนเรียน อย่างไรก็ตาม ขอบเขตของข้อมูลที่ใช้ยังค่อนข้างจำกัดเฉพาะด้าน ส่งผลให้มิติของการวิเคราะห์อาจยังไม่ครอบคลุมปัจจัยสำคัญอื่นที่ส่งผลต่อผลสัมฤทธิ์ทางการศึกษา หากมีการบูรณาการข้อมูลจากแหล่งอื่นเพิ่มเติม เช่น ข้อมูลจากแบบสอบถามที่สะท้อนทัศนคติ แรงจูงใจ หรือพฤติกรรมการเรียนรู้ของนิสิต ข้อมูลจากการมีส่วนร่วมในกิจกรรมเสริมหลักสูตรของภาควิชาและคณะ หรือข้อมูลเชิงพฤติกรรมจากระบบการเรียนการสอนออนไลน์ (เช่น จำนวนครั้งในการเข้าชั้นเรียนออนไลน์ ระยะเวลาในการเรียน หรือรูปแบบการโต้ตอบกับเนื้อหา) ก็จะสามารถเพิ่มความลึกและความหลากหลายให้กับตัวแปรอธิบาย ซึ่งอาจนำไปสู่การพัฒนาโมเดลการทำนายที่มีความแม่นยำและมีประสิทธิภาพสูงยิ่งขึ้น การใช้ข้อมูลแบบสหวิถีในการวิเคราะห์ยังสามารถสะท้อนลักษณะพหุปัจจัยของกระบวนการเรียนรู้ของนิสิตได้อย่างครบถ้วนมากขึ้น ซึ่งเป็นแนวทางที่ได้รับความสนใจในงานวิจัยด้านการศึกษายุคใหม่ โดยเฉพาะอย่างยิ่งในบริบทของการวิเคราะห์เชิงพฤติกรรมและการออกแบบระบบสนับสนุนการเรียนรู้แบบเฉพาะบุคคล

ในอนาคต ผู้วิจัยมีความเชื่อมั่นว่า การบูรณาการองค์ความรู้จากงานวิจัยด้านการทำเหมืองข้อมูลทางการศึกษาเข้ากับกระบวนการพัฒนาหลักสูตร จะสามารถส่งเสริมการปรับปรุงหลักสูตรให้มีประสิทธิภาพและตอบสนองต่อความต้องการของผู้เรียนได้มากยิ่งขึ้น โดยเฉพาะเมื่อมีข้อมูลเชิงประจักษ์จากผลการเรียนของนิสิตในรุ่นต่าง ๆ ทั้งในอดีตและปัจจุบัน ซึ่งสามารถใช้เป็นฐานข้อมูลในการวิเคราะห์ลำดับความสำคัญของรายวิชาต่าง ๆ ภายในหลักสูตร ทั้งในด้านความยากง่าย ผลสัมฤทธิ์ และความสัมพันธ์กับรายวิชาอื่น การบูรณาการข้อมูลดังกล่าวร่วมกับข้อมูลเชิงคุณภาพ จะช่วยให้การออกแบบหลักสูตรมีลักษณะเป็นแบบองค์รวม ที่สามารถพัฒนาผู้เรียนได้อย่างรอบด้านมากยิ่งขึ้น

นอกจากนี้ ผลที่ได้จากงานวิจัยยังสามารถนำไปประยุกต์ใช้ในการสนับสนุนการเรียนรู้ของนิสิตในระหว่างการศึกษา เช่น การออกแบบระบบแนะนำรายวิชา การเตือนภัยล่วงหน้าสำหรับผู้ที่มีความเสี่ยงทางการเรียน หรือการจัดกลุ่มกิจกรรมส่งเสริมเฉพาะกลุ่ม ซึ่งสิ่งเหล่านี้ล้วนเป็นองค์ประกอบสำคัญที่ช่วยยกระดับผลการเรียนโดยรวมของนิสิต ซึ่งท้ายที่สุดแล้ว เป้าหมายของการประยุกต์ใช้ผลการวิจัยในลักษณะนี้คือการผลิตบัณฑิตที่มีคุณภาพ ตอบโจทย์ต่อทั้งความต้องการของตลาดแรงงานและการเป็นพลเมืองที่มีคุณภาพในสังคมยุคดิจิทัลอย่างแท้จริง

บรรณานุกรม

นนทวัฒน์ ทวีชาติ, อรยา เพ็งประจญ, วิไลรัตน์ ยาทองไชย และชูศักดิ์ ยาทองไชย. (2563). ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์มหาวิทยาลัย ราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏบุรีรัมย์*, *4*(1), 47-60.

ปรียารัตน์ นาคสุวรรณ์ และกิดาการ สายธนู. (2555). การทำนายผลสัมฤทธิ์ในการเรียนวิชาสถิติเบื้องต้นของนิสิตปริญญาตรี มหาวิทยาลัยบูรพา ด้วยการวิเคราะห์การจำแนกและข่ายงานระบบประสาท. *วารสารวิทยาศาสตร์บูรพา*, *17*(1), 59-68.

วันทนี ประจวบศุภกิจ, ธัญญาภรณ์ บุญยัง และสุกัญญา บุญศรี. (2564). การสร้างตัวแบบเพื่อทำนายผลสัมฤทธิ์ทางการศึกษาจากพฤติกรรมการใช้สมาร์ตโฟนด้วยเทคนิคการทำเหมืองข้อมูล. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, *31*(3), 550-560.

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13-49. https://doi.org/10.1016/j.tele.2019.01.007

Almarabeh, H. (2017). Analysis of students' performance by using different data mining classifiers. *International Journal of Modern Education and Computer Science*, *9*(8), 9–15. https://doi.org/10.5815/ijmecs.2017.08.02

Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, *40*(1), 133–148. https://doi.org/10.1080/01587919.2018.1553562

Alyahyan, E., & Dustegor, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(3), 1-27. https://doi.org/10.1186/s41239-020-0177-7

Anoopkumar, M., & Rahman, A. M. J. M. Z. (2016). A review on data mining techniques and factors used in educational data mining to predict student amelioration. In *Proceedings of the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (pp. 122–133). IEEE. https://doi.org/10.1109/SAPIENCE.2016.7684113

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., Hashim, W., & Tiong, S. K. (2022). Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs). *Applied Sciences*, *12*(3), 1-16. https://doi.org/10.3390/app12031289

Baashar, Y., Hamed, Y., Alkawsi, G., Capretz, L. F., Alhussian, H., Alwadain, A., & Al-amri, R. (2022). Evaluation of postgraduate academic performance using artificial intelligence models. *Alexandria Engineering Journal*, *61*, 9867-9878. https://doi.org/10.1016/j.aej.2022.03.021

Baker, R. S. J. D., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3–16. https://doi.org/10.5281/zenodo.3554657

Garg, R. (2018). Predict student performance in different regions of Punjab. *International Journal of Advanced Research in Computer Science*, *9*(1), 236–241. https://doi.org/10.26483/ijarcs.v9i1.5234

Jayaprakash, S. (2018). A Survey on Academic Progression of Students in Tertiary Education using Classification Algorithms. *International Journal of Engineering Technology Science and Research (IJETSR)*, *5*(2), 136–142.

Migueis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, *115*(4), 36-51. https://doi.org/10.1016/j.dss.2018.09.001

Parker, J. D., Hogan, M. J., Eastabrook, J. M., Oke, A., & Wood, L. M. (2006). Emotional intelligence and student retention: Predicting the successful transition from high school to university. *Personality and Individual differences*, *41*(7), 1329–1336. https://doi.org/10.1016/j.paid.2006.04.022

Pérez, B., Castellanos, C., & Correal, D. (2018). Predicting student drop-out rates using data mining techniques: A case study. In *Advances in Artificial Intelligence – IBERAMIA 2018* (pp. 111–125). Springer. https://doi.org/10.1007/978-3-030-03023-0_10

Pérez, J., Iturbide, E., Olivares, V., Hidalgo, M., Almanza, N., & Martínez, A. (2015). A data preparation methodology in data mining applied to mortality population databases. *Journal of Medical Systems, 39*(11), 1-6. https://doi.org/10.1007/s10916-015-0312-5

Raheela, A., Agathe, M., Syed, A. A., & Najmi, G. H. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*. *113*. 177-194. https://doi.org/10.1016/j.compedu.2017.05.007

Richard-Eaglin, A. (2017). Predicting student success in nurse practitioner programs. *Journal of the American Association of Nurse Practitioners*, *29*(10), 600–605. https://doi.org/10.1002/2327-6924.12502

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting Student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: A preprocessing engine. *Journal of Computer Science, 2*(9), 735–739. https://doi.org/10.3844/jcssp.2006.735.739

Xing, W. (2019). Exploring the influences of MOOC design features on student performance and persistence. *Distance Education*, *40*(1), 98–113. https://doi.org/10.1080/01587919.2018.1553560

Zimmermann, J., Brodersen, K. H., Heinimann, H. R., & Buhmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. *Journal of Educational Data Mining*, *7*(3), 151-176. https://doi.org/10.5281/zenodo.3554733

ภาคผนวก

บันทึกข้อความ

ส่วนงาน กองบริหารการวิจัยและนวัตกรรม งานมาตรฐานและจริยธรรมในการวิจัย โทร. ๒๖๒๐
ที่ อว ๘๑๐๐/๑๐๔๘๔ **วันที่** ๖ ตุลาคม พ.ศ. ๒๕๖๖
เรื่อง ขอส่งเอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา

เรียน คณบดีคณะมนุษยศาสตร์และสังคมศาสตร์

ตามที่นักวิจัยในหน่วยงานของท่าน ได้ยื่นเอกสารคำร้องเพื่อขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ชุดที่ ๒ (กลุ่มมนุษยศาสตร์และสังคมศาสตร์) รหัสโครงการวิจัย HU 075/2566 โครงการวิจัยเรื่อง การวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี หลักสูตรสารสนเทศศึกษา ด้วยเทคนิคการทำเหมืองข้อมูลทางการศึกษา โดยมี ดร.ณิชพน โชติจันทรกุล เป็นหัวหน้าโครงการวิจัย นั้น

บัดนี้ คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาโครงการฉบับนี้ ตามวิธีดำเนินการมาตรฐาน (Standard Operating Procedures, SOPs) คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ฉบับที่ ๒.๐ ประกาศ ณ วันที่ ๔ มกราคม พ.ศ. ๒๕๖๖ แล้วว่าโครงการวิจัยดังกล่าวเป็นโครงการวิจัยที่สามารถให้การรับรองโดยยกเว้นการลงมติจากที่ประชุม (Exemption Determination) ตามข้อ ๔.) เป็นการวิจัยที่ใช้ผลการทดสอบทางการศึกษา (Cognitive, Diagnostic, Attitude, Achievement) หรือใช้ข้อมูลในแบบบันทึกข้อมูลของหน่วยงานการศึกษา โดยได้รับความยินยอมจากผู้รับผิดชอบข้อมูลแล้ว ซึ่งข้อมูลดังกล่าวไม่สามารถเชื่อมโยงถึงเจ้าของข้อมูลได้ (ทั้งนี้ โครงการวิจัยต้องเป็นการเก็บข้อมูลการวิจัยในนักเรียน นิสิตหรือนักศึกษา โดยผู้วิจัยไม่ใช่ครูหรืออาจารย์ที่มีอำนาจครอบงำ หรือผู้วิจัยสามารถแสดงให้เห็นว่าวิธีการเข้าถึงกลุ่มตัวอย่างไม่ส่งผลกระทบต่อความเป็นอิสระในการตัดสินใจในการเข้าร่วมโครงการของกลุ่มตัวอย่าง) จึงเห็นสมควรให้ดำเนินการวิจัยได้ พร้อมนี้ ได้แนบเอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์ (หมายเลขใบรับรองที่ IRB2-095/2566) มายังท่าน เพื่อแจ้งนักวิจัยที่มีรายชื่อข้างต้น นำไปใช้ในการเก็บข้อมูลจริงจากผู้เข้าร่วมโครงการวิจัยต่อไป โดยห้ามนักวิจัยเบี่ยงเบนรายละเอียดต่างๆ ของโครงการวิจัยที่ยื่นมาขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา และเมื่อนักวิจัยดำเนินการวิจัยเสร็จเรียบร้อยแล้ว ขอให้แจ้งปิดโครงการวิจัย (Final Report) มายังคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ด้วย

จึงเรียนมาเพื่อโปรดแจ้งให้นักวิจัยทราบ จะขอบคุณยิ่ง

(นายเจนวิทย์ นวลแสง)
ประธานคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา
ชุดที่ ๒ (กลุ่มมนุษยศาสตร์และสังคมศาสตร์)

หมายเหตุ : ผู้วิจัยสามารถดาวน์โหลดเอกสารชี้แจงผู้เข้าร่วมโครงการวิจัย เอกสารแสดงความยินยอมของผู้เข้าร่วมโครงการวิจัย และเอกสารเครื่องมือที่ใช้ในการวิจัยต่างๆ ซึ่งผ่านการประทับตรารับรองเรียบร้อยแล้ว ได้ที่ระบบการขอรับพิจารณาจริยธรรมการวิจัยแบบออนไลน์ (BUU Ethics Submission Online) เพื่อนำไปใช้ในการเก็บข้อมูลจริงจากผู้เข้าร่วมโครงการวิจัยต่อไป

สำเนา

ที่ IRB2-095/2566

เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์

มหาวิทยาลัยบูรพา

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาโครงการวิจัย

รหัสโครงการวิจัย : HU075/2566

โครงการวิจัยเรื่อง : การวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี หลักสูตรสารสนเทศศึกษา ด้วยเทคนิคการทำเหมืองข้อมูลทางการศึกษา

หัวหน้าโครงการวิจัย : นายณิชพน โชติจันทรกุล

หน่วยงานที่สังกัด : คณะมนุษยศาสตร์และสังคมศาสตร์

วิธีพิจารณา : ☑ Exemption Determination ☐ Expedited Reviews ☐ Full Board

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาแล้วเห็นว่า โครงการวิจัยดังกล่าวเป็นไปตามหลักการของจริยธรรมการวิจัยในมนุษย์ โดยที่ผู้วิจัยเคารพสิทธิและศักดิ์ศรีในความเป็นมนุษย์ไม่มีการล่วงละเมิดสิทธิ สวัสดิภาพ และไม่ก่อให้เกิดภยันตรายแก่ตัวอย่างการวิจัยและผู้เข้าร่วมโครงการวิจัย

จึงเห็นสมควรให้ดำเนินการวิจัยในขอบข่ายของโครงการวิจัยที่เสนอได้ (ดูตามเอกสารตรวจสอบ)

1. แบบเสนอเพื่อขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ ฉบับที่ 2 วันที่ 14 เดือน กันยายน พ.ศ. 2566
2. โครงการวิจัยฉบับภาษาไทย ฉบับที่ 2 วันที่ 14 เดือน กันยายน พ.ศ. 2566
3. เอกสารชี้แจงผู้เข้าร่วมโครงการวิจัย ฉบับที่ - วันที่ - เดือน - พ.ศ. -
4. เอกสารแสดงความยินยอมของผู้เข้าร่วมโครงการวิจัย ฉบับที่ - วันที่ - เดือน - พ.ศ. -
5. แบบเก็บรวมรวบข้อมูล เช่น แบบบันทึกข้อมูล (Data Collection Form)
แบบสอบถาม หรือสัมภาษณ์ หรืออื่น ๆ ที่เกี่ยวข้อง ฉบับที่ 1 วันที่ 16 เดือน สิงหาคม พ.ศ. 2566
6. เอกสารอื่น ๆ (ถ้ามี)
6.1 รายละเอียดเครื่องมือการวิจัย ฉบับที่ 1 วันที่ 16 เดือน สิงหาคม พ.ศ. 2566
6.2 หนังสือขออนุญาตเก็บข้อมูล ฉบับที่ 1 วันที่ 16 เดือน สิงหาคม พ.ศ. 2566

วันที่รับรอง : วันที่ 27 เดือน กันยายน พ.ศ. 2566
วันที่หมดอายุ : วันที่ 27 เดือน กันยายน พ.ศ. 2567

ลงนาม *อาจารย์เจนวิทย์ นวลแสง*
(อาจารย์เจนวิทย์ นวลแสง)
ประธานคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา