การเปรียบเทียบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูล เพื่อจำแนกระดับผลการเรียนของนักศึกษากลุ่มสาขาวิชาคอมพิวเตอร์

Performance comparison of Data Mining Techniques for Classification Academic Performance of Computer Student program

วรรัตน์ จงไกรจักร[1] และ ภูริทรัพย์ เดชพิพัฒน์ประชา[2]
Worrarat Jongkraijak and Bhurisub Dejpipatpracha

บทคัดย่อ

งานวิจัยนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลของเทคนิคการทำเหมืองข้อมูล 3 แบบ ได้แก่ 1) Neural Network 2) Naïve Bayes และ 3) k-Nearest Neighbor เทคนิคใดมีประสิทธิภาพในการจำแนกข้อมูลมากที่สุด โดยใช้ข้อมูลผลการเรียนระดับมัธยมศึกษาตอนปลายของนักศึกษากลุ่มสาขาวิชาคอมพิวเตอร์ มหาวิทยาลัยราชภัฏภูเก็ต ตั้งแต่ปีการศึกษา พ.ศ. 2554 – 2558 จำนวน 930 ชุด แต่ละชุดประกอบด้วยผลการเรียนเฉลี่ยภาคการศึกษาแรกเข้า ผลการเรียนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย และผลการเรียนเฉลี่ยสะสม 8 กลุ่มสาระการเรียนรู้ โดยผลการเปรียบเทียบประสิทธิภาพพบว่า เทคนิคแบบ k-Nearest Neighbor มีประสิทธิภาพในการจำแนกข้อมูลมากที่สุด ซึ่งมีค่าเฉลี่ยความถูกต้องเป็น 71.18% มากกว่าเทคนิคแบบ Neural Network และ Naïve Bayes ที่มีค่าเฉลี่ยความถูกต้องเป็น 57.74% และ 55.87% ตามลำดับ ดังนั้นจึงได้นำเทคนิคแบบ k-Nearest Neighbor มาทดสอบโดยการลดข้อมูลนำเข้าเพื่อหาปัจจัยที่ส่งผลต่อผลการเรียนของนักศึกษาภาคการศึกษาแรกเข้าพบว่า ปัจจัยที่ส่งผลกระทบมากที่สุดเรียงลำดับจากมากไปน้อย คือ 1) การงานอาชีพและเทคโนโลยี 2) ศิลปะ 3) สังคมศึกษา ศาสนาและวัฒนธรรม 4) ภาษาต่างประเทศ 5) คณิตศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ผลการเรียนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย 8) วิทยาศาสตร์ และ 9) ภาษาไทย

คำสำคัญ: การเปรียบเทียบประสิทธิภาพแบบจำลอง เทคนิคการทำเหมืองข้อมูล การจำแนกข้อมูล

ABSTRACT

This research aims to compare performance of 3 data mining techniques: Neural Network, Naïve Bayes, and k-Nearest Neighbor, that which one is the most effective technique to classify data. The import data is high school academic performance of students who were studied in Phuket Rajabhat University Computer program since year 2011 – 2015 totally 930 datasets. The Data set includes: the first semester in university average academic performance, high school GPAX, and each 8 course categories GPAX. Finally, the data mining technique comparison result found that k-Nearest Neighbor is the highest efficiency was 71.18% that more than Neural Network and Naïve Bayes techniques, which had an average accuracy of 57.74% and 55.87% respectively. Therefore, to use k-Nearest Neighbor to reduce the inputs data in order to determine the factor that affect the academic performance of students in the first semester. The descending order most affected course is Occupations and Technology, Arts, Social Studies,

[1] สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏภูเก็ต
E-mail: worrarat.j@pkru.ac.th
[2] สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏ- ภูเก็ต E-mail: bhurisub@pkru.ac.th

Foreign languages, Mathematics, Health and Physical Education, High school GPAX, Science, and Thai Language.

Keywords: Data Mining, Classification, Algorithm

บทนำ

ปัจจุบันเทคโนโลยีคอมพิวเตอร์ทางด้านระบบปัญญาประดิษฐ์ได้รับการพัฒนาให้ก้าวหน้าขึ้นเป็นอย่างมาก และมีการใช้งานอย่างแพร่หลาย โดยนำมาประยุกต์ช่วยอำนวยความสะดวกสบายให้กับมนุษย์ ทั้งทางด้านการแพทย์ที่มีการนำหุ่นยนต์ที่ใช้ระบบปัญญาประดิษฐ์ช่วยในการผ่าตัด ทางด้านอุตสาหกรรมที่มีระบบปัญญาประดิษฐ์มาใช้เพื่อลดต้นทุนการผลิต ทางด้านการทหารที่นำระบบปัญญาประดิษฐ์ประยุกต์ใช้ในเครื่องบินไร้คนขับหรือรถถังไร้คนขับ ซึ่งเห็นได้ว่าระบบปัญญาประดิษฐ์ถูกนำมาใช้กับงานได้หลากหลายด้าน แต่ในด้านการศึกษายังไม่มีการนำระบบปัญญาประดิษฐ์เข้ามาประยุกต์ใช้อย่างจริงจังโดยเฉพาะอย่างยิ่งในระดับมหาวิทยาลัย

มหาวิทยาลัยราชภัฏภูเก็ตเป็นสถาบันการศึกษาของรัฐที่มีหลักสูตรที่เปิดการเรียนการสอนเป็นจำนวนมาก แต่ละปีการศึกษาจะมีนักศึกษาที่มีผลการเรียนเฉลี่ยไม่ถึงตามเกณฑ์ที่มหาวิทยาลัยกำหนดส่งผลให้พ้นสภาพการเป็นนักศึกษา สาเหตุอาจมาจากสถานภาพทางครอบครัว ปัญหาส่วนตัว หรือปรับตัวไม่ทันเมื่อเริ่มเข้าเรียนในระดับมหาวิทยาลัย โดยเฉพาะกลุ่มสาขาวิชาคอมพิวเตอร์ มหาวิทยาลัยฯ มีการเปิดการเรียนการสอนตาม มคอ.1 ทั้งหมด 4 สาขาวิชา ได้แก่ สาขาวิชาวิทยาการคอมพิวเตอร์ สาขาวิชาเทคโนโลยีสารสนเทศ สาขาวิชาคอมพิวเตอร์ธุรกิจ และสาขาวิชาคอมพิวเตอร์ศึกษา ซึ่งก่อนที่นักศึกษาจะเข้ามหาวิทยาลัยฯ จะมีการจัดเก็บข้อมูลของนักศึกษาทั้งที่อยู่ในรูปแบบกระดาษและจัดเก็บในฐานข้อมูลเพื่อใช้เป็นหลักฐานในการตรวจสอบข้อมูลนักศึกษาแรกเข้าและใช้ประโยชน์ในการดำเนินงานกิจกรมต่าง ๆ ของทางมหาวิทยาลัยฯ แต่ข้อมูลนักศึกษาดังกล่าวยังไม่ได้นำไปใช้ประโยชน์อย่างเต็มที่ ดังนั้นทางคณะผู้วิจัยจึงมีแนวคิดในการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล 3 แบบ ได้แก่ 1) Neural Network 2) Naïve Bayes และ 3) k-Nearest Neighbor มาจำแนกข้อมูลผลการเรียนของนักศึกษา เพื่อเปรียบเทียบประสิทธิภาพของการจำแนกโดยเทคนิคดังกล่าว โดยมีจุดมุ่งหมายเพื่อหาเทคนิคที่เหมาะสมกับข้อมูลนักศึกษาของมหาวิทยาลัยฯ และหาปัจจัยนำเข้าที่ส่งผลกระทบต่อผลระดับผลการเรียนในภาคการศึกษาแรกเข้า ทำให้สามารถวางแผนดูแลนักศึกษาที่มีความเสี่ยงต่อการพ้นสภาพการในภาคการศึกษาแรกได้

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลของเทคนิคการทำเหมืองข้อมูล 3 แบบ คือ Neural Network Naïve Bayes และ k-Nearest Neighbor

2. เพื่อหาปัจจัยนำเข้าที่ส่งผลต่อระดับผลการเรียนในภาคการศึกษาแรกเข้าของนักศึกษากลุ่มสาขาวิชาคอมพิวเตอร์ มหาวิทยาลัยราชภัฏภูเก็ต

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

กลุ่มสาระการเรียนรู้

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 มุ่งพัฒนาผู้เรียนให้มีคุณภาพตามมาตรฐานการเรียนรู้ ซึ่งการพัฒนาผู้เรียนให้บรรลุมาตรฐานการเรียนรู้ ผู้เรียนในระดับการศึกษาขั้นพื้นฐานจำเป็นต้องเรียนรู้ตามกลุ่ม

สาระการเรียนรู้ โดยแบ่งออกเป็น 8 กลุ่ม ได้แก่ 1) ภาษาไทย 2) คณิตศาสตร์ 3) วิทยาศาสตร์ 4) สังคมศึกษา ศาสนาและวัฒนธรรม 5) สุขศึกษาและพลศึกษา 6) ศิลปะ 7) การงานอาชีพและเทคโนโลยี 8) ภาษาต่างประเทศ (กระทรวงศึกษาธิการ, 2551)

ผลการศึกษาวิจัย เรื่องการพัฒนาแบบจำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษาสาขาคอมพิวเตอร์ โดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Perceptron Neural Network) ซึ่งกำหนดชั้นซ่อน (Hidden layer) เท่ากับ 1 และใช้ฟังก์ชันการกระตุ้น คือ sigmoid function พบว่าตัวแปรที่มีอิทธิพลต่อการสมัครงานให้ตรงกับวุฒิการศึกษา คือ บุคลิกภาพที่สอดคล้องคุณลักษณะบัณฑิตของสาขาคอมพิวเตอร์ และผลการเรียนเฉลี่ยสะสมของแต่ละกลุ่มสาระการเรียนรู้ ซึ่งผลลัพธ์การพยากรณ์ของแบบจำลองมีค่าเฉลี่ยความถูกต้องอยู่ที่ 75.63% (ณัฏฐา ผิวมา, 2558)

สอดคล้องกับผลการศึกษาวิจัยเรื่องการใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี พบว่ากลุ่มสาระการเรียนรู้ในระดับชั้นมัธยมศึกษาตอนปลายมีความสำคัญต่อการเลือกสาขาวิชาในระดับปริญญาตรี (อนันต์ ปินะเตและคณะ, 2557)

จากงานวิจัยข้างต้นพบว่า ผลการเรียนเฉลี่ยสะสมของกลุ่มสาระการเรียนรู้ระดับมัธยมศึกษาตอนปลายมีความสัมพันธ์กับผลการเรียนระดับปริญญาตรี ดังนั้นงานวิจัยนี้จึงใช้ระดับผลการเรียนเฉลี่ยสะสมในแต่ละกลุ่มสาระการเรียนรู้และระดับผลการเรียนเฉลี่ยสะสมในระดับมัธยมศึกษาตอนปลาย มาใช้เป็นข้อมูลเพื่อการจำแนกกลุ่มนักศึกษาสาขาวิชาคอมพิวเตอร์ ของมหาวิทยาลัยราชภัฏภูเก็ต

เหมืองข้อมูล

เหมืองข้อมูล (Data Mining) คือ การสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Databases : KDD) หรือที่เรียกกันว่าการทำเหมืองข้อมูล เป็นเทคนิคที่ใช้ในการจัดการข้อมูลขนาดใหญ่ที่ถูกจัดเก็บอยู่ในคลังข้อมูล แล้วจะนำข้อมูลมาวิเคราะห์เพื่อดึงความรู้หรือสิ่งสำคัญออกมา ซึ่งใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) วิธีการทางคณิตศาสตร์และสถิติ วิธีการทางปัญญาประดิษฐ์ (Artificial Intelligence) หรือวิธีอื่น ๆ มาทำการวิเคราะห์เพื่อสกัดความรู้จากคลังข้อมูล เพื่อการวิเคราะห์แนวโน้ม ความสัมพันธ์ กฎ หรือรูปแบบของข้อมูลประกอบด้วย 3 กระบวนการหลัก ๆ ดังนี้

1. กระบวนการเตรียมข้อมูล (Pre-processing) เป็นกระบวนการเตรียมข้อมูลให้เหมาะสม และอยู่ในรูปแบบที่สามารถนำไปใช้งานต่อได้

2. กระบวนการทำเหมืองข้อมูล (Data Mining) เป็นการเลือกเทคนิคที่เหมาะสมสำหรับงานที่ต้องการ โดยสามารถรวมเทคนิคได้มากกว่าหนึ่งเทคนิคมาประมวลผลเพื่อดึงความรู้หรือสิ่งที่น่าสนใจจากข้อมูลที่ผ่านขั้นตอน Pre-processing แล้ว โดยผลลัพธ์ที่ได้จากขั้นตอนนี้คือ ฐานความรู้

3. กระบวนการประเมินผล (Post-processing) เป็นการนำฐานความรู้ที่ได้จากขั้นตอนการทำเหมืองข้อมูลมาทดสอบและพิจารณาว่าถูกต้องตามความต้องการหรือไม่ ซึ่งบางครั้งอาจต้องปรับแก้ค่าและนำเข้าสู่ขั้นตอนทำเหมืองข้อมูลใหม่อีกครั้งจนกว่าจะได้ความรู้หรือสิ่งที่น่าสนใจตามที่ต้องการออกมา (สายชล สินสมบูรณ์ทอง, 2558)

การจำแนกกลุ่ม

การจำแนกกลุ่ม (Classification) เป็นกระบวนการสร้างต้นแบบจัดการข้อมูลให้อยู่ในรูปแบบของกลุ่มที่กำหนดให้เพื่อแสดงให้เห็นความแตกต่างระหว่างกลุ่มของข้อมูล และเพื่อจำแนกข้อมูลว่าควรจัดอยู่กลุ่มใด ซึ่งต้นแบบที่ใช้ในการจำแนกข้อมูลออกเป็นกลุ่มตามที่กำหนด จะขึ้นอยู่กับการวิเคราะห์เซตของชุดข้อมูล (Data Set) โดยนำข้อมูลมาสอนให้เป็นระบบ

เรียนรู้ว่าข้อมูลใดอยู่ในกลุ่มเดียวกันบ้าง และต้นแบบนี้สามารถจัดกลุ่มของข้อมูลที่ยังไม่ได้จำแนกกลุ่มได้ด้วย ซึ่งจะมีขั้นตอนการจำแนกกลุ่ม 2 ขั้นตอน คือ

ขั้นตอนที่ 1 การสร้างต้นแบบ (Classifier Model) เป็นการนำชุดข้อมูล ผ่านกระบวนการของอัลกอริธึมการจำแนก (Classification Algorithm) ซึ่งผลลัพธ์ที่ได้จะอยู่ในรูปแบบของต้นแบบการจำแนก จากนั้นสามารถสร้างเป็นกฎได้

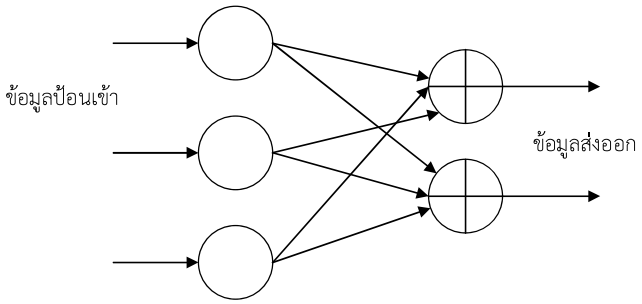
ขั้นตอนที่ 2 การใช้โมเดลเพื่อการทำนาย (Prediction) ซึ่งจุดมุ่งหมายสูงสุดในการแก้ไขปัญหา คือ การสร้างต้นแบบแล้วเมื่อมีข้อมูลใหม่ก็จะสามารถทำนายได้ โดยนำข้อมูลที่ได้รับการเปรียบเทียบและวิเคราะห์เพื่อตัดสินใจความเป็นไปได้ของข้อมูลนั้น ๆ (ภัทร์พงศ์ พงศ์ภัทรกานต์, 2553)

เทคนิคแบบ Neural Network

เทคนิคแบบ Neural Network หรือโครงข่ายประสาทเทียม เป็นศาสตร์แขนงหนึ่งทางด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) ที่สามารถนำไปประยุกต์ใช้กับงานหลายด้านได้อย่างมีประสิทธิภาพ เช่น การจำแนกรูปแบบ การทำนาย การควบคุม การหาความเหมาะสม และการจัดกลุ่ม เป็นต้น หลักการสำคัญของโครงข่ายประสาทเทียม คือ ความพยายามที่จะลอกเลียนแบบการทำงานของเซลล์ประสาทในสมองมนุษย์เพื่อทำงานได้อย่างมีประสิทธิภาพ ลักษณะทั่วไปของโครงข่ายประสาทเทียม คือ การที่โหนด (Node) ต่าง ๆ จำลองมาจากไซแนป (Synapse) ของเซลล์ประสาทระหว่างเดนไดร์ท (Dendrite) และแอกซอน (Axon) โดยมีฟังก์ชันเป็นตัวกำหนดสัญญาณส่งออก (Activation Function or Transfer Function)

ลักษณะของโครงข่ายประสาทเทียมสามารถแบ่งได้ 2 แบบ ดังนี้

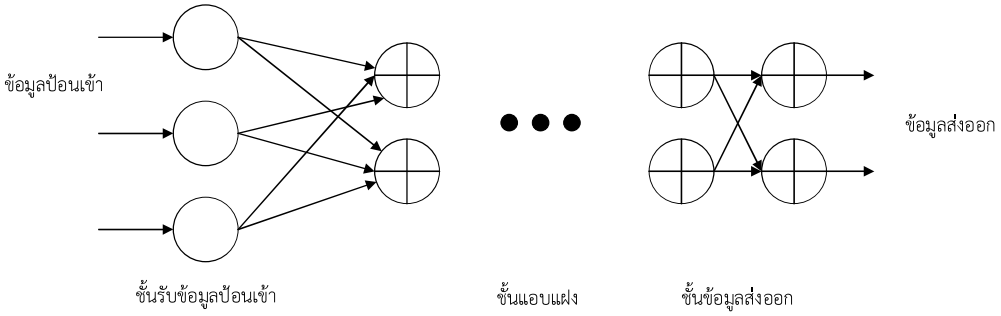
1. โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer) จะมีเพียงชั้นสัญญาณประสาทขาเข้า และชั้นสัญญาณประสาทขาออกเท่านั้น เช่น โครงข่ายเพอเซบตรอนอย่างง่าย (Simple Perceptron) และโครงข่ายโฮปฟิลด์ (Hopfield networks) เป็นต้น ตัวอย่างโครงข่ายประสาทเทียมแบบชั้นเดียว แสดงดังภาพที่ 3.1



ภาพที่ 3.1 แสดงโครงข่ายประสาทเทียมแบบชั้นเดียว. จาก ธนาวุฒิ ประกอบผล. (2552). *โครงข่ายประสาทเทียม Artificial Neural Network.* วารสาร มฉก.วิชาการ.

จากภาพที่ 3.1 โหนดในชั้นรับข้อมูลป้อนเข้าทำหน้าที่รับข้อมูลเข้า (input value) แล้วส่งข้อมูลผ่านเส้นเชื่อมโยงต่าง ๆ ไปให้โหนดในชั้นส่งข้อมูลออก ความเข้มของสัญญาณ หรือปริมาณข้อมูลที่นำเข้าสู่โหนดในชั้นส่งข้อมูลออก จะขึ้นอยู่กับค่าน้ำหนักที่อยู่บนเส้นเชื่อมโยงโหนดในชั้นส่งข้อมูลออกจะนำข้อมูลที่ได้รับมาคำนวณโดยใช้ฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันการแปลง (transfer function) ที่เหมาะสมกับปัญหา แล้วส่งผลลัพธ์ที่ได้ออกมาเป็นข้อมูลส่งออก

2. โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer) มีลักษณะเช่นเดียวกับโครงข่ายประสาทเทียมแบบชั้นเดียว แต่จะมีชั้นแอบแฝง (Hidden Node) เพิ่มขึ้น โดยอยู่ส่วนกลางระหว่างชั้นนำข้อมูลป้อนเข้าและชั้นส่งข้อมูลออก ดังแสดงในภาพที่ 3.2



ภาพที่ 3.2 แสดงโครงข่ายประสาทเทียมแบบชั้นเดียว. จาก ธนาวุฒิ ประกอบผล. (2552). *โครงข่ายประสาทเทียม Artificial Neural Network.* วารสาร มฉก.วิชาการ.

การแบ่งโครงข่ายประสาทเทียมตามประเภทการเรียนรู้ของโครงข่าย สามารถแบ่งได้ 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) โดยในปัจจุบันการพัฒนาโครงข่ายประสาทเทียมยังคงมีการดำเนินการอย่างต่อเนื่องและคาดว่าจะมีบทบาทอย่างมากในด้านการจำแนกรูปแบบ การพยากรณ์ การควบคุม การหาความเหมาะสมและการจัดกลุ่ม (ธนาวุฒิ ประกอบผล, 2552)

เทคนิคแบบ Naïve Bayes

เทคนิคแบบ Naïve Bayes เป็นตัวจำแนกประเภทแบบหนึ่งที่ใช้งานได้ดี เหมาะกับกรณีของเซตตัวอย่างมีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน มีการนำอัลกอริทึม Naïve Bayes ไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ (Text Classification) การวินิจฉัย (Diagnosis) และพบว่าใช้งานได้ดีไม่ต่างจากการจำแนกประเภทวิธีการอื่น เช่น การเรียนรู้ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม เป็นต้น ขั้นตอนการจำแนกข้อมูลโดยใช้อัลกอริทึม Naïve Bayes ทำได้ดังนี้

กำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นกลุ่มสำหรับข้อมูลที่มีแอทริบิวต์ทั้งหมด n ตัว จะได้

$$P = \{a_1, a_2, a_3, \ldots, a_n\} \text{ หรือ } P = \{a_1, a_2, a_3, \ldots, a_n \mid V_j\}$$

จากนั้นคำนวณค่าผลคูณของความน่าจะเป็นโดยใช้สมการดังนี้

$$P = \{a_1, a_2, a_3, \ldots, a_n \mid V_j\} = \prod_{i=1}^{n} P(a_i|V_j)$$

โดยที่ $\prod$ หมายถึงการนำค่า $P = \{a_i, \mid V_j\}$ ทั้งหมดมาคูณกัน สมการนี้ถ้าใช้กฎลูกโซ่มาคำนวณค่าความน่าจะเป็นที่ด้านซ้ายของสูตรจะได้เท่ากับ $P = \{a_i, \mid V_j\} \times P = \{a_2, \mid a_1, V_j\} \times P = \{a_2, \mid a_2, a_1, V_j\} \times \ldots \times P = \{a_n, \mid a_{a-1}, a_{n-2}, \ldots, a_1, V_j\}$ ดังนั้นค่าความน่าจะเป็นทางด้านซ้ายของสมการจะเท่ากับผลคูณค่าความน่าจะเป็นทางด้านขวาก็ต่อเมื่อคุณสมบัติ $a_1, a_2, \ldots, a_n$ ไม่ขึ้นต่อกัน นำค่าที่ได้มาเปรียบเทียบกับกลุ่มที่มีค่าความน่าจะเป็นสูงสุดโดยใช้สมการดังนี้

$$V_{NB} = \arg\max_{v_j \in V} P(v_j) \times \prod_{i=1}^{n} (a_i|v_j)$$

ข้อดีของเทคนิคแบบ Naïve Bayes คือ สามารถใช้ข้อมูลและความรู้ก่อนหน้า (Prior Knowledge) เข้ามาช่วยในการเรียนรู้ได้ด้วย ความรู้ก่อนหน้าหมายถึงความรู้ที่มีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่จะเก็บข้อมูล เมื่อใช้งานจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง พบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น (บุญเสริม กิจศิริกุล, 2546)

เทคนิคแบบ k-Nearest Neighbor

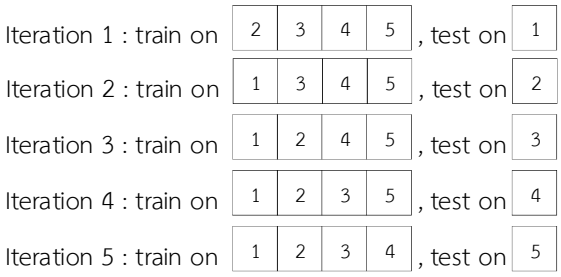
การจำแนกข้อมูลด้วยเทคนิคแบบ k-Nearest Neighbor จะเป็นการเรียนรู้โดยการเปรียบเทียบกันระหว่างเรคคอร์ดของข้อมูลที่ต้องการจำแนก/ทำนายหมวดหมู่กับเรคคอร์ดทั้งหมดในชุดข้อมูลสอนที่มีลักษณะเหมือนกันหรือใกล้เคียงกันด้วยการพิจารณาข้อมูลแอทริบิวต่าง ๆ ในการที่จะหาความเหมือนกันหรือต่างกัน หรือค่าความใกล้เคียงกันระหว่าเรคคอร์ดใด ๆ สามารถประยุกต์ใช้มาตรวัดความแตกต่างต่าง ๆ เช่น การหาความใกล้เคียงโดยใช้ Euclidean distance ที่ซึ่งจะทำการพิจารณา 2 เรคคอร์ด $X_1=(x_{11}, x_{12}, ..., x_{1n})$ และ $X_2=(x_{21}, x_{22}, ..., x_{2n})$ ใด ๆ จากนั้นทำการพิจารณาความแตกต่างระหว่างค่าที่ปรากฏขึ้นในแต่ละแอทริบิวของทั้งสองเรคคอร์ดแล้วนามารวมเป็นค่าความแตกต่างรวมที่ซึ่งจะสามารถคำนวณได้ดังนี้

$$dist(X_1,X_2)=\sqrt{\sum_{i=1}^{n}(x_{1i}-x_{2i})^2}$$

หลังจากทำการหาค่าความแตกต่างระหว่างเรคคอร์ดที่ต้องการจำแนกหมวดหมู่กับเรคคอร์ดทั้งหมดในชุดข้อมูลสอนแล้ว จะทำการจำแนกหมวดหมู่ข้อมูลด้วยการพิจารณาจากหมวดหมู่ของข้อมูล k เรคคอร์ดที่มีค่าความแตกต่างน้อยที่สุดอันดับแรก เมื่อเทียบกับเรคคอร์ดที่ต้องการจำแนกหมวดหมู่ด้วยการค้นหาว่าใน k k เรคคอร์ดพบเจอหมวดหมู่ข้อมูลใดมากที่สุด จะสรุปว่าเรคคอร์ดที่ต้องการจำแนกหมวดหมู่จะมีหมวดหมู่ตามหมวดหมู่ที่มากที่สุดที่พิจารณาก่อนหน้า (โกเมศ อัมพวัน , 2558)

การวัดประสิทธิภาพ

การวัดค่าประสิทธิภาพของเทคนิควิธีต่าง ๆ จะต้องทำการเลือกข้อมูลสำหรับเรียนรู้ (Training Set) และข้อมูลสำหรับทดสอบ (Testing Set) ในงานวิจัยนี้เลือกใช้วิธีสุ่มเลือกแบ่งข้อมูลแบบความเที่ยงตรง k กลุ่ม (k-Fold Cross Validation) โดยเริ่มจากการแบ่งชุดข้อมูลออกเป็นส่วน ๆ ที่เท่ากัน นำข้อมูลบางส่วนมาทำการเรียนรู้และนำข้อมูลบางส่วนมาทำการทดสอบแบบจำลองที่ได้จากการเรียนรู้ โดยในการทำงานจะทำการเลือกสุ่มข้อมูลออกเป็น k ชุดเท่ากัน ดังตัวอย่างในภาพที่ 3.3



ภาพที่ 3.3 แสดงการวัดประสิทธิภาพโดยใช้ 5 - fold Cross Validation. จาก เยาวภา ภารสาเร็จ. จิรัฏฐา ภูบุญอบ. วิรัตน์ พงษ์ศิริ. (2556). การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, ครั้งที่ 9

การทดลองครั้งแรกข้อมูลชุดที่ 1 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดเรียนรู้ ในการทดลองครั้งที่ 2 ข้อมูลชุดที่ 2 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดเรียนรู้ ทำจนกระทั่งข้อมูลทุกชุดได้ถูกนำมาเป็นข้อมูลชุดทดสอบและชุดเรียนรู้ ซึ่งจะมีการทดลองทั้งหมด k ครั้ง (เยาวภา ภารสาเร็จ, จิรัฏฐา ภูบุญอบ, วิรัตน์ พงษ์ศิริ, 2556)

วิธีการดำเนินการวิจัย

ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการจัดเตรียมข้อมูลจะใช้ข้อมูลของกลุ่มนักศึกษาสาขาวิชาคอมพิวเตอร์ ประกอบด้วย 4 สาขาวิชา คือ คอมพิวเตอร์ธุรกิจ วิทยาการคอมพิวเตอร์ เทคโนโลยีสารสนเทศ และคอมพิวเตอร์ศึกษา ตั้งแต่ปีการศึกษา พ.ศ. 2554 - 2558 จำนวน 930 ชุดข้อมูล ประกอบด้วยข้อมูลระดับผลการเรียนเฉลี่ย 8 กลุ่มสาระการเรียนรู้ ผลการเรียนเฉลี่ยสะสมระดับขั้นมัธยมศึกษาตอนปลาย และผลการเรียนเฉลี่ยสะสมภาคการศึกษาแรกเข้า ดังแสดงตารางที่ 4.1

ตารางที่ 4.1 แสดงตัวอย่างข้อมูลที่ใช้ในงานวิจัย

GPA_TH	GPA_Math	GPA_Sci	GPA_So	GPA_Gym	GPA_Art	GPA_Tech	GPA_EN	GPA_Entry	GPA_T1
2.43	1.62	2.77	4.00	3.00	2.55	3.43	2.42	2.83	2.78
1.80	2.16	2.08	2.29	2.38	3.46	3.27	1.82	2.42	1.85
2.35	1.82	2.33	2.88	3.25	3.75	3.47	1.41	2.80	3.14
...	...	...	...	...	...	...	...	...	...

จากตารางที่ 4.1 แสดงตัวอย่างข้อมูลที่ใช้ในงานวิจัย ซึ่งประกอบด้วยข้อมูลดังนี้

1. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ภาษาไทย (GPA_TH)
2. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้คณิตศาสตร์ (GPA_Math)
3. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้วิทยาศาสตร์ (GPA_Sci)
4. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม (GPA_So)
5. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้สุขศึกษาและพลศึกษา (GPA_Gym)
6. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ศิลปะ (GPA_Art)
7. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี (GPA_Tech)
8. ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (GPA_EN)
9. ผลการเรียนเฉลี่ยสะสมชั้นมัธยมศึกษาตอนปลาย (GPA_Entry)
10. ผลการเรียนเฉลี่ยสะสมภาคการศึกษาแรกเข้า (GPA_T1)

จากชุดข้อมูลที่ได้กล่าวมา ผลการเรียนเฉลี่ยสะสมภาคการศึกษาแรกเข้า (GPA_T1) จะเป็นค่าเป้าหมายสำหรับการจัดกลุ่ม โดยผลการเรียนจะถูกจัดเป็น 3 กลุ่ม คือ 1) กลุ่มผลการเรียนดีมาก (Excellent) มีระดับคะแนนเฉลี่ยอยู่ระหว่าง 3.00 – 4.00 2) กลุ่มผลการเรียนดี (Good) มีระดับคะแนนเฉลี่ยอยู่ระหว่าง 2.00 – 2.99 และ 3) กลุ่มผลการเรียนที่ควรปรับปรุง (Poor) มีระดับคะแนนเฉลี่ยอยู่ระหว่าง 1.00 – 1.99 ซึ่งเมื่อจัดกลุ่มข้อมูลแล้ว จะได้ข้อมูลดังตารางที่ 4.2

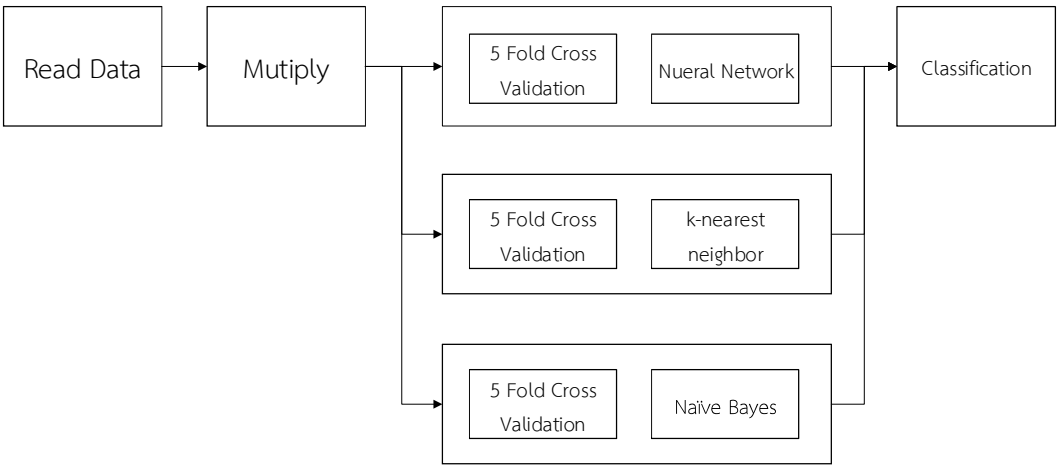
ตารางที่ 4.2 แสดงตัวอย่างข้อมูลที่ได้จัดกลุ่มข้อมูลแล้ว

GPA_TH	GPA_Math	GPA_Sci	GPA_So	GPA_Gym	GPA_Art	GPA_Tech	GPA_EN	GPA_Entry	GPA_T1
2.43	1.62	2.77	4.00	3.00	2.55	3.43	2.42	2.83	Good
1.80	2.16	2.08	2.29	2.38	3.46	3.27	1.82	2.42	Poor
2.35	1.82	2.33	2.88	3.25	3.75	3.47	1.41	2.80	Excellent
...	...	...	...	...	...	...	...	...	...

เมื่อได้จัดกลุ่มข้อมูลเรียบร้อยแล้วจะได้ข้อมูลตามตารางที่ 4.2 จากนั้นจะส่งข้อมูลเหล่านี้ไปยังกระบวนการสร้างและทดสอบแบบจำลองเพื่อใช้ในการจำแนกประเภทข้อมูลต่อไป

การสร้างและทดสอบแบบจำลอง

ขั้นตอนการสร้างแบบจำลองจะนำข้อมูลที่ได้จากขั้นตอนการเตรียมข้อมูลมาจำแนกโดยใช้เทคนิคการทำเหมืองข้อมูล 3 แบบ คือ 1) Neural Network แบบ feed-forward neural network 2) Naïve Bayes และ 3) k-Nearest Neighbor กำหนดค่า k = 1 ใช้โปรแกรม RapidMiner Studio ในการจำแนกข้อมูล ระบบจำแนกข้อมูลแสดงดังภาพที่ 4.1



ภาพที่ 4.1 แสดงระบบการจำแนกข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูล 3 แบบ

จากภาพที่ 4.1 ซึ่งแสดงระบบการจำแนกข้อมูลโดยใช้เทคนิค 3 แบบ ผลลัพธ์ที่ได้จะถูกทดสอบประสิทธิภาพโดยใช้วิธีการ k-fold cross validation โดยได้กำหนดให้ K = 5 ซึ่งผลลัพธ์จากขั้นตอนนี้คือค่าเฉลี่ยความถูกต้อง (Accuracy) และค่าเฉลี่ยความเที่ยงตรง (Precision) ในการจำแนกข้อมูลของเทคนิคแต่ละแบบ โดยจะนำมาเปรียบเทียบกันเพื่อหาเทคนิคที่มีค่าเฉลี่ยความถูกต้องและเที่ยงตรงมากที่สุดในการจำแนกข้อมูลมากที่สุด

การหาปัจจัยที่ส่งผลต่อระดับผลการเรียน

ขั้นตอนนี้จะค้นหาปัจจัยที่มีความสำคัญมากที่สุด โดยนำเทคนิคที่มีประสิทธิภาพมากที่สุดจากการเปรียบเทียบมาค้นหาปัจจัยที่ส่งผลต่อระดับผลการเรียน ด้วยการตรวจสอบปัจจัยย้อนกลับ โดยลดการนำเข้าทีละแอทริบิวต์ ตรวจสอบค่าความถูกต้อง (Accuracy) ที่ลดลงมากที่สุด คือ แอทริบิวต์ที่มีความสำคัญหรือเป็นปัจจัยที่มีความสำคัญที่สุด

ผลการวิจัย

ผลการจำแนกข้อมูลนักศึกษาโดยใช้เทคนิคทั้ง 3 แบบ คือ 1) Neural Network, 2) Naïve Bayes และ 3) k-Nearest Neighbor แสดงดังตารางที่ 5.1 ตารางที่ 5.2 และตารางที่ 5.3 ตามลำดับ

ตารางที่ 5.1 แสดงผลการจำแนกข้อมูลโดยใช้เทคนิคแบบ Neural Network

การจำแนกประเภท (Classification)	**ค่าจริง (Actual)**			**ค่าเฉลี่ยความเที่ยงตรง**
	กลุ่มผลการเรียนดีเยี่ยม	**กลุ่มผลการเรียนดี**	**กลุ่มผลการเรียนที่ควรปรับปรุง**	
กลุ่มผลการเรียนดีเยี่ยม	169	75	20	64.02%
กลุ่มผลการเรียนดี	84	155	88	47.40%
กลุ่มผลการเรียนที่ควรปรับปรุง	33	93	213	62.83%
ค่าเฉลี่ยความถูกต้อง	59.09%	47.99%	66.36%	

จากข้อมูลในตารางที่ 5.1 ค่าเฉลี่ยความถูกต้องของการจำแนกประเภทของทุกกลุ่ม คิดเป็น 57.74% เมื่อสังเกตค่าเฉลี่ยความถูกต้องของแต่ละกลุ่ม พบว่ากลุ่มผลการเรียนที่ควรปรับปรุงมีค่าเฉลี่ยความถูกต้องสูงที่สุด คิดเป็น 66.36% รองลงมาเป็นกลุ่มผลการเรียนดีเยี่ยมมีค่าเฉลี่ยความถูกต้องคิดเป็น 59.09% และสุดท้ายกลุ่มผลการเรียนดีจำแนกได้ถูกต้องน้อยที่สุด คิดเป็น 47.99% เนื่องจากกลุ่มนี้จะถูกจำแนกไปเป็นกลุ่มผลการเรียนดีเยี่ยม 75 ชุดข้อมูล และถูกจำแนกไปเป็นกลุ่มผลการเรียนที่ควรปรับปรุงจำนวน 93 ข้อมูล

ตารางที่ 5.2 แสดงผลการจำแนกข้อมูลโดยใช้เทคนิคแบบ Naïve Bayes

การจำแนกประเภท (Classification)	**ค่าความจริง (Actual)**			**ค่าเฉลี่ยความเที่ยงตรง**
	กลุ่มผลการเรียนดีเยี่ยม	**กลุ่มผลการเรียนดี**	**กลุ่มผลการเรียนที่ควรปรับปรุง**	
กลุ่มผลการเรียนดีเยี่ยม	181	107	36	55.86%
กลุ่มผลการเรียนดี	67	126	72	47.55%
กลุ่มผลการเรียนที่ควรปรับปรุง	38	90	212	62.35%
ค่าเฉลี่ยความถูกต้อง	63.29%	39.01%	66.25%	

จากข้อมูลในตารางที่ 5.2 ผลการจำแนกประเภทมีค่าเฉลี่ยความถูกต้องของทุกกลุ่ม คิดเป็น 55.87% โดยผลการจำแนกข้อมูลกลุ่มผลการเรียนที่ควรปรับปรุงมีค่าเฉลี่ยความถูกต้องมากที่สุด คิดเป็น 66.25% รองลงมาเป็นกลุ่มผลการเรียนดีเยี่ยม มีค่าเฉลี่ยความถูกต้องเป็น 63.29% และลำดับสุดท้ายเป็นกลุ่มผลการเรียนดี มีค่าเฉลี่ยเป็น 39.01 เนื่องจากถูกจำแนกไปเป็นกลุ่มผลการเรียนดีเยี่ยม จำนวน 107 ชุดข้อมูล และไปเป็นกลุ่มผลการเรียนที่ควรปรับปรุงจำนวน 90 ชุดข้อมูล

ตารางที่ 5.3 แสดงผลการจำแนกข้อมูลโดยใช้เทคนิคแบบ k-Nearest Neighbor

การจำแนกประเภท (Classification)	**ค่าความจริง (Actual)**			**ค่าเฉลี่ยความเที่ยงตรง**
	กลุ่มผลการเรียนดีเยี่ยม	**กลุ่มผลการเรียนดี**	**กลุ่มผลการเรียนที่ควรปรับปรุง**	
กลุ่มผลการเรียนดีเยี่ยม	272	57	26	76.62%
กลุ่มผลการเรียนดี	14	179	84	64.62%
กลุ่มผลการเรียนที่ควรปรับปรุง	0	87	211	70.81%
ค่าเฉลี่ยความถูกต้อง	95.10%	55.42%	65.73%	

จากข้อมูลในตารางที่ 5.3 ผลการจำแนกประเภทข้อมูลมีค่าเฉลี่ยความถูกต้องของทุกกลุ่มคิดเป็น 71.18% ซึ่งสามารถจำแนกกลุ่มผลการเรียนดีเยี่ยมได้ถูกต้องที่สุด คิดเป็น 95.10% รองลงมาเป็นกลุ่มผลการเรียนที่ควรปรับปรุง คิดเป็น 65.73% และลำดับสุดท้ายเป็นกลุ่มผลการเรียนดี เป็น 55.42% เนื่องจากถูกจำแนกไปเป็นกลุ่มผลการเรียนดีเยี่ยม จำนวน 57 ชุดข้อมูล และผลการเรียนที่ควรปรับปรุงจำนวน 87 ชุดข้อมูล

เมื่อนำผลการจำแนกข้อมูลของอัลกอริทึมแต่ละแบบมาเปรียบเทียบค่าเฉลี่ยความถูกต้องและค่าเฉลี่ยความเที่ยงตรงจะได้ดังตารางที่ 5.4

ตารางที่ 5.4 แสดงการเปรียบเทียบค่าเฉลี่ยความถูกต้องและเที่ยงตรงของเทคนิคแต่ละแบบ

เทคนิคแบบ	ค่าเฉลี่ยความถูกต้อง (Accuracy)	ค่าเฉลี่ยความเที่ยงตรง (Precision)
Neural Network	57.74%	58.08%
Naïve Bayes	55.87%	55.25%
k-Nearest Neighbor	71.18%	70.68%

เมื่อทำการเปรียบเทียบค่าเฉลี่ยความถูกต้องของผลการจำแนกข้อมูลนักศึกษาโดยใช้เทคนิคที่แตกต่างกัน จะเห็นได้ว่าค่าเฉลี่ยความถูกต้องและค่าเฉลี่ยความเที่ยงตรงของเทคนิคแต่ละแบบจะมีค่าใกล้เคียงกัน ซึ่งหมายความว่าทุกเทคนิคสามารถจำแนกข้อมูลได้อย่างถูกต้องและเที่ยงตรง แต่เทคนิคแบบ k-Nearest Neighbor จะมีค่าสูงกว่าเมื่อเทียบกับเทคนิคแบบ Neural Network และ Naïve Bayes จึงได้เลือกวิธีนี้เพื่อไปทดสอบหาปัจจัยที่ส่งผลต่อระดับผลการเรียนต่อไป

การหาปัจจัยที่ส่งผลต่อระดับผลการเรียน

เมื่อได้เทคนิคที่สามารถจำแนกข้อมูลนักศึกษาได้ถูกต้องที่สุดแล้ว จะนำมาหาปัจจัยนำเข้าที่มีผลต่อค่าเป้าหมาย (ระดับผลการเรียนเฉลี่ยในภาคการศึกษาแรกเข้า) มากที่สุด โดยการลดข้อมูลนำเข้าทีละแอทริบิวต์ซึ่งสามารถเรียงลำดับปัจจัยที่มีผลกระทบต่อค่าเป้าหมายมากที่สุดไปหาน้อยที่สุดได้ ดังตารางที่ 5.5

ตารางที่ 5.5 แสดงปัจจัยที่มีผลกระทบต่อค่าเป้าหมายมากที่สุดจากมากไปน้อย

ปัจจัยนำเข้า	ค่าเฉลี่ยความถูกต้อง	ลำดับความสำคัญที่มีผลการกระทบต่อค่าเป้าหมาย
ผลการเรียนเฉลี่ยสะสมการงานอาชีพและเทคโนโลยี	54.25%	1
ผลการเรียนเฉลี่ยสะสมศิลปะ	54.62%	2
ผลการเรียนเฉลี่ยสะสมสังคมศึกษา ศาสนาและวัฒนธรรม	55.11%	3
ผลการเรียนเฉลี่ยสะสมภาษาต่างประเทศ	55.22%	4
ผลการเรียนเฉลี่ยสะสมคณิตศาสตร์	55.33%	5
ผลการเรียนเฉลี่ยสะสมสุขศึกษาและพลศึกษา	55.54%	6
ผลการเรียนเฉลี่ยสะสมในชั้นมัธยมศึกษาตอนปลาย	55.87%	7
ผลการเรียนเฉลี่ยสะสมวิทยาศาสตร์	56.08%	8
ผลการเรียนเฉลี่ยสะสมภาษาไทย	69.46%	9

อภิปรายผล

ผลการวิจัยการเปรียบเทียบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลเพื่อจำแนกระดับผลการเรียนของนักศึกษากลุ่มสาขาวิชาคอมพิวเตอร์ ทั้ง 3 เทคนิค คือ 1) Neural Network 2) Naïve Bayes และ 3) k-Nearest Neighbor สรุปได้ว่า เทคนิคแบบ k-Nearest Neighbor มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลได้ถูกต้องที่สุด เมื่อเทียบกับเทคนิคแบบ Naïve Bayes และแบบ Neural Network โดยสามารถจำแนกกลุ่มข้อมูลได้ถูกต้องคิดเป็นค่าเฉลี่ย 71.18% และมีค่าความเที่ยงตรงเฉลี่ยเป็น 70.68% ส่วนเทคนิคแบบ Naïve Bayes และแบบ Neural Network จำแนกประเภทได้ถูกต้องน้อยกว่าเนื่องจากทั้ง 2 เทคนิค ยังจำแนกกลุ่มข้อมูลประเภทกลุ่มผลการเรียนดีได้ไม่ดีเท่าที่ควร สังเกตได้จากความเฉลี่ยความถูกต้อง

ของกลุ่มนี้ที่มีค่าเป็น 39.01% ในเทคนิคแบบ Naïve Bayes และ 47.99% ในเทคนิคแบบ Neural Network เนื่องจากถูกจำแนกไปเป็นกลุ่มผลการเรียนดีเยี่ยมและกลุ่มผลการเรียนที่ควรปรับปรุงค่อนข้างมาก

เมื่อได้เทคนิคที่มีความถูกต้องมากที่สุดแล้ว ได้ทดสอบลดปัจจัยนำเข้าเพื่อหาปัจจัยที่มีผลกระทบต่อระดับผลการเรียนเฉลี่ยในภาคการศึกษาแรกเข้าของนักศึกษา ซึ่งผลการวิจัยพบว่าปัจจัยนำเข้าที่ส่งผลกระทบมากที่สุดเรียงลำดับจากมากไปน้อย ได้แก่ 1) ผลการเรียนเฉลี่ยสะสมการงานอาชีพและเทคโนโลยี 2) ศิลปะ 3) สังคมศึกษา ศาสนาและวัฒนธรรม 4) ภาษาต่างประเทศ 5) คณิตศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ผลการเรียนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย 8) วิทยาศาสตร์ และ 9) ภาษาไทย ตามลำดับ

ข้อเสนอแนะ

จากผลการวิจัยที่ได้หากต้องการให้ประสิทธิภาพของการจำแนกข้อมูลสูงขึ้น ควรมีการวิเคราะห์ความสัมพันธ์ของข้อมูลในเชิงสถิติ เพื่อหากลุ่มของข้อมูลที่มีความสัมพันธ์กันและกลุ่มของข้อมูลที่ไม่มีความสัมพันธ์กัน ซึ่งจะทำให้ทราบสาเหตุที่ทำให้ความถูกต้องของการจำแนกประเภทข้อมูลในบางกลุ่มยังมีค่าน้อย และควรเก็บข้อมูลของนักศึกษาเพิ่มเติม เช่น ข้อมูลส่วนตัว ข้อมูลพื้นฐานเกี่ยวกับครอบครัว เป็นต้น โดยที่อาจจะส่งผลกระทบต่อผลระดับผลการเรียนในแต่ละรายวิชาของนักศึกษา และผลการวิจัยในครั้งนี้สามารถนำปัจจัยนำเข้าที่ส่งผลกระทบต่อระดับผลการเรียน ไปใช้ประโยชน์ในการเฝ้าระวังกลุ่มนักศึกษาแรกเข้าที่จัดอยู่ในกลุ่มผลการเรียนไม่ดี โดยการดูแลให้คำแนะนำหรือปรับพื้นฐานในด้านวิชาการให้กับนักศึกษากลุ่มนี้ก่อนเข้าเรียนในภาคการศึกษาแรกเข้าซึ่งจะสามารถลดความเสี่ยงการพ้นสภาพของนักศึกษาให้มีจำนวนลดลงได้

เอกสารอ้างอิง

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* โรงพิมพคุรุสภาลาดพราว.

ณัฏฐา ผิวมา. (2558). *การพัฒนาแบบจำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษา สาขาคอมพิวเตอร์โดยใช้ โครงข่ายประสาทเทียม.* วารสารปัญญาภิวัฒน์ ปีที่ 7 ฉบับที่ 2.

ธนาวุฒิ ประกอบผล. (2552). *โครงข่ายประสาทเทียม Artificial Neural Network.* วารสาร มฉก.วิชาการ.

บุญเสริม กิจศิริกุล. (2546). *ปัญญาประดิษฐ์ Artificial Intelligent.* สำนักพิมพ์จุฬาฯ.

ภัทร์พงศ์ พงศ์ภัทรกานต์. (2553). *การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ระดับปริญญาตรี โดยใช้คอมมิตตีแมชชีน.* NCCIT, 2010.

เยาวภา ภารสาเร็จ. จิรัฏฐา ภูบุญอบ. วิรัตน์ พงษ์ศิริ. (2556). *การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา.* วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, ครั้งที่ 9.

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

อนันต์ ปินะเต ฉัตรเกล้า เจริญผล และแกวกาญจน์ สมประเสริฐศรี. (2557). *การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสาหรับการศึกษาต่อระดับปริญญาตรี.* วารสารมหาวิทยาลัยมหาสารคาม, ปีที่ 33, หน้า 648 – 656.