การพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง*

PREDICTING ACADEMIC PERFORMANCE FOR PLANNING REGISTER OF SPORTS SCIENCE UNDER GRADUATES USING MACHINE LEARNING ALGORITHMS

เอกวิจัย เมยไธสง[1], ฉวีวรรณ สีสม[2], สุเทพ เมยไธสง[3]
Eckwijai Maythaisong[1], Chaweewan Seesom[2], Suthep Maythaisong[3]
มหาวิทยาลัยการกีฬาแห่งชาติ วิทยาเขตมหาสารคาม[1,2]
Thailand National Sports University, Mahasarakham Campus[1,2]
มหาวิทยาลัยมหามกุฏราชวิทยาลัย วิทยาเขตร้อยเอ็ด[3]
Mahamakut Buddhist University, Roi Et Campus[3]
Email : eckwijai@gmail.com

บทคัดย่อ

บทความวิจัยนี้มีวัตถุประสงค์ 1)เพื่อสร้างแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง 2)เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นข้อมูลของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬา คณะวิทยาศาสตร์การกีฬาและสุขภาพ มหาวิทยาลัยการกีฬาแห่งชาติ วิทยาเขตมหาสารคาม ที่เข้าศึกษาระหว่างปีการศึกษา 2556-2557 จำนวน 132 ระเบียน เครื่องมือที่ใช้ในการวิจัยเป็นการสร้างแบบจำลองด้วยวิธีการเรียนรู้ของเครื่อง ได้แก่ โครงข่ายประสาทเทียม การเรียนรู้แบบต้นไม้ตัดสินใจ การค้นหาเพื่อนบ้านใกล้ที่สุด การเรียนรู้เบย์อย่างง่าย และการสุ่มป่าไม้ สถิติที่ใช้ในการวิเคราะห์ข้อมูลเป็นค่าร้อยละโดยใช้เกณฑ์การประเมินประสิทธิภาพแบบจำลอง ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล

ผลการวิจัยพบว่า วิธีโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นให้ค่าความถูกต้องสูงที่สุดจากชุดข้อมูลรายวิชาทั้งหมดให้ค่าความถูกต้องคิดเป็นร้อยละ 87.9 และชุดข้อมูลเฉพาะรายวิชาเลือกให้ค่าความถูกต้องคิดเป็นร้อยละ 75 ตามลำดับ

คำสำคัญ : 1. การพยากรณ์ 2. การวางแผนการลงทะเบียน 3. การเรียนรู้ของเครื่อง

*Received: April 27, 2022; Revised: June 15, 2022; Accepted: June 21, 2022

ABSTRACT

The objectives of this research article were 1) to create a model pattern discovery for course enrollment planning of undergraduate students in sports science using machine learning algorithm 2) to compare the efficiency of the model for pattern discovery for course enrollment planning of undergraduate students in sports science using machine learning. The samples used in the research were 132 records of students from the Faculty of Sports and Health Science, Thailand National Sports University, Mahasarakham Campus, during the 2013-2014 academic year. The research instrument was conducted to build the model with machine learning. The algorithm sets comprise Artificial Neural Network, Decision Tree Learning, K-Nearest Neighbors, Naive Bayes and Random Forest. The statistics used to analyze the data are percentage values by which performance is measured by Accuracy, Precision, Recall and F-Measure.

The result showed that Artificial Neural Network (Multi-Layer Perceptron) obtained the highest values for both levels of analysiswith accuracy performances for all subjects 87.9% and elective subjects 75% respectively.

Keywords : 1. Prediction 2. Registration Planning 3. Machine Learning

1. ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

การศึกษาระดับอุดมศึกษาปัจจุบันได้มีการมุ่งเน้นในการพัฒนาหลักสูตรเพื่อให้นักศึกษาที่เข้าศึกษาในมหาวิทยาลัยมีความรู้ความสามารถพัฒนาตนเองให้มีคุณภาพตรงตามความต้องการของตลาดงานสามารถทำงานเพื่อดำรงชีพตนเองและช่วยเหลือสังคม อย่างไรก็ตามเป้าหมายของมหาวิทยาลัยต้องการให้นักศึกษาสำเร็จการศึกษาตามระยะเวลาที่หลักสูตรกำหนด ซึ่งแต่ละหลักสูตรมีรายวิชาที่มีความหลากหลายแตกต่างกันไปสำหรับรองรับกับความต้องการของนักศึกษาที่มีความสนใจในแต่ละด้าน ในบางรายวิชาก็มีจำนวนจำกัดทำให้การลงทะเบียนเรียนของนักศึกษาในแต่ละภาคการศึกษานั้นจะต้องมีการวางแผนการลงทะเบียนก่อนเพื่อให้ผ่านหลักสูตรตามที่ทางคณะกำหนด แต่ละสาขาวิชามีแผนการลงทะเบียนเรียนไว้ให้นักศึกษาดูเป็นแบบอย่างในการลงทะเบียนเลือกลงวิชาต่างๆ ซึ่งตามเกณฑ์มาตรฐานหลักสูตรระดับปริญญาตรีกำหนดให้แต่ละหลักสูตรต้องศึกษาไม่น้อยกว่า 120 หน่วยกิต ใช้เวลาศึกษาไม่เกิน 8 ปีการศึกษา และต้องได้รับเกรดเฉลี่ยไม่น้อยกว่า 2.00 เมื่อจบการศึกษา ทำให้นักศึกษาจะต้องมีการวางแผนก่อนการลงทะเบียนเลือกรายวิชาที่เหมาะสมให้

มีผลการเรียนของตนเองอยู่ในเกณฑ์ดี ซึ่งในยุคปัจจุบันด้วยการพัฒนาอย่างรวดเร็วของคอมพิวเตอร์และเทคโนโลยี การวิเคราะห์ข้อมูลทางการศึกษาเพื่อค้นหาข้อมูลที่ซ่อนอยู่เพื่อให้ผู้เรียน ผู้สอน หรือผู้บริหารได้รับประโยชน์จากข้อมูลนั้น การทำเหมืองข้อมูลทางการศึกษา(Educational Data Mining) เป็นหนึ่งในวิธีที่สำคัญในการวิเคราะห์ข้อมูลด้านการศึกษาอย่างมีประสิทธิภาพโดยประยุกต์ใช้การเรียนรู้ของเครื่อง(Machine Learning) และการทำเหมืองข้อมูล (Data Mining) Bo Guo et al. (2015)มีหลายงานวิจัยที่ได้นำเทคนิคการทำเหมืองข้อมูลทางการศึกษามาแก้ปัญหาในการวิเคราะห์ประสิทธิภาพการเรียนของนักศึกษา ดังตัวอย่างในงานวิจัยของ ปรียารัตน์ นาคสุวรรณ์ และกิดาการ สายธนู (2555) ได้วิเคราะห์การจำแนกและข่ายงานระบบประสาทเพื่อทำนายผลสัมฤทธิ์ในการเรียนวิชาสถิติเบื้องต้นของนิสิตปริญญาตรี งานวิจัยของ บีสุดา ดาวเรือง และคณะ (2564) ได้เปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศเพื่อใช้ทำนายแขนงวิชาเรียนที่เหมาะสมกับนักศึกษาในภาควิชา งานวิจัยของพงษ์เทพ รักผกาวงศ์และอุไรวรรณ์ รักผกาวงศ์ (2558) ได้ใช้การใช้กฎความสัมพันธ์ร่วมกับฟัซซี่กฎความสัมพันธ์เพื่อคาดการณ์ผลการเรียนของนักศึกษาเพื่อให้นักศึกษามีผลการเรียนเฉลี่ยรวมที่สูงขึ้นและสามารถผ่านเกณฑ์ที่กำหนดได้เมื่อเรียนจนครบหลักสูตรงานวิจัยของ อนันต์ ปินะเต และคณะ (2557) ได้นำเทคนิคการทำเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี ซึ่งจะเห็นได้ว่าการทำเหมืองข้อมูลทางการศึกษามีความสำคัญอย่างยิ่งสำหรับการแก้ปัญหาประสิทธิภาพการเรียนของนักศึกษาในมหาวิทยาลัย จากข้อมูลนักศึกษาปีการศึกษา 2553-2557 หลักสูตรสาขาวิชาวิทยาศาสตร์การกีฬา คณะวิทยาศาสตร์การกีฬาและสุขภาพ มหาวิทยาลัยการกีฬาแห่งชาติ วิทยาเขตมหาสารคาม มีนักศึกษาที่พ้นสภาพการเป็นนักศึกษาเนื่องจากผลการเรียนรวมเฉลี่ยไม่ถึง 2.00 คิดเป็นร้อยละ 40 จากนักศึกษาในสาขาวิชาทั้งหมด ซึ่งเห็นได้ว่านักศึกษาที่พ้นสภาพการเป็นนักศึกษาเนื่องจากมีผลการเรียนต่ำกว่าเกณฑ์การประเมินมีเป็นจำนวนมาก

จากข้อมูลดังกล่าวการหาสาเหตุที่นักศึกษามีผลการเรียนลดลง เนื่องจากผลการเรียนเป็นสิ่งสำคัญที่ใช้ในการสำเร็จการศึกษาอีกทั้งการวางแผนในการลงทะเบียนรายวิชาใดถึงจะได้ผลการเรียนที่ดีขึ้นหรือผลการเรียนใดควรจะพัฒนาในด้านการเรียนการสอน จึงเป็นที่มาของการสร้างแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยประยุกต์ใช้วิธีการเรียนรู้ของเครื่องมาช่วยวิเคราะห์ในการพยากรณ์ผลการเรียนแล้วทำการเปรียบเทียบผลที่ได้จากการวิเคราะห์ทั้ง 5 วิธี ได้แก่ โครงข่ายประสาทเทียม(Artificial Neural Network) การเรียนรู้แบบต้นไม้ตัดสินใจ(Decision Tree Learning) การค้นหาเพื่อนบ้านใกล้ที่สุด(K-Nearest Neighbors) การเรียนรู้เบย์อย่างง่าย(Naive Bayes) และการสุ่มป่าไม้(Random Forest) ว่าวิธีการไหนจะสามารถทำนายผลการเรียนได้มีประสิทธิภาพดีที่สุด ในการคำนวณผลการเรียนจะขึ้นอยู่กับหมวดวิชาบังคับและหมวดวิชาเลือกตามเกณฑ์มาตรฐานหลักสูตรสาขาวิชาวิทยาศาสตร์การกีฬา

กำหนดให้นักศึกษาจะต้องลงทะเบียนเรียนไม่น้อยกว่า 40 รายวิชา เพื่อเป็นแนวทางให้กับนักศึกษาที่กำลังวางแผนจะลงทะเบียนเรียนเลือกเรียนรายวิชาที่ทำให้ผลการเรียนของตนเองอยู่ในเกณฑ์ที่ผ่านตามหลักสูตรมีแนวโน้มที่จะสามารถสำเร็จการศึกษาได้

2. วัตถุประสงค์ของการวิจัย

2.1 เพื่อสร้างแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง

2.2 เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง

3. ประโยชน์ที่ได้รับจากการวิจัย

3.1 ได้แบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง

3.2 ได้ผลการเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง

4. วิธีดำเนินการวิจัย

ผู้วิจัยได้ทำการแบ่งการวิจัยออกเป็น 3 ขั้นตอน ประกอบด้วย การเตรียมข้อมูล การสร้างแบบจำลอง และการประเมินผล มีรายละเอียดดังนี้

1. การเตรียมข้อมูล (Data Preparation) การวิจัยครั้งนี้ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลรายวิชาที่ลงทะเบียนเรียน ผลการเรียน และเกรดเฉลี่ยสะสม ของนักศึกษาระดับปริญญาตรี ภาคปกติ สาขาวิชาวิทยาศาสตร์การกีฬา คณะวิทยาศาสตร์การกีฬาและสุขภาพ มหาวิทยาลัยการกีฬาแห่งชาติ วิทยาเขตมหาสารคาม ที่เริ่มเข้าศึกษาตั้งแต่ปีการศึกษา 2556-2557 (หลักสูตรปรับปรุง พ.ศ. 2556) และจบการศึกษาตามเกณฑ์มาตรฐานหลักสูตร (เกรดเฉลี่ยสะสมตั้งแต่ 2.00 ขึ้นไป) และทำการแปลงข้อมูล (Data Transformation) โดยเริ่มจากการทำความสะอาดข้อมูล (Data Cleaning) นำข้อมูลที่ไม่สามารถประมวลผลได้ออกไป ด้วยการลบนักศึกษาที่ไม่ได้ลงทะเบียนเรียนตามหลักสูตรรายวิชาที่ไม่มีนักศึกษาลงทะเบียนเรียน นักศึกษาที่มีรายวิชาได้ผลการเรียน F จะถูกแทนที่ด้วยผลการเรียนที่ลงทะเบียนแก้แล้วผ่านล่าสุด จะได้จำนวนรายวิชาในการลงทะเบียนทั้งหมด 59 รายวิชา แบ่งออกเป็น 2 หมวด คือ หมวดวิชาบังคับ 38 วิชาและ หมวดวิชาเลือก 21 วิชา ซึ่งแต่ละวิชาจะมีจำนวนรหัส 8 หลัก จากนั้นแบ่งกลุ่มเกรดเฉลี่ยสะสมออกเป็น 4 กลุ่ม ได้แก่ 1)กลุ่มที่มีเกรดเฉลี่ยสะสมตั้งแต่ 2.00 ถึง 2.74 (เกณฑ์สำเร็จการศึกษา) 2)กลุ่มที่มีเกรดเฉลี่ยสะสมตั้งแต่ 2.75 ถึง 3.24

3)กลุ่มที่มีเกรดเฉลี่ยสะสมตั้งแต่ 3.25 ถึง 3.59 (เกียรตินิยมอันดับ 2) 4)กลุ่มที่มีเกรดเฉลี่ยสะสมตั้งแต่ 3.60 ถึง 4.00 (เกียรตินิยมอันดับ 1) ดังนั้นจะทำให้มีข้อมูลที่ใช้ในการทดลองรวมทั้งสิ้น 132 ระเบียน

2. การสร้างแบบจำลอง(Modeling) เป็นขั้นตอนการออกแบบและสร้างแบบจำลองการพยากรณ์ผลการเรียนโดยประยุกต์ใช้การเรียนรู้ของเครื่อง การวิเคราะห์ด้วยการเรียนรู้ของเครื่องเพื่อสร้างแบบจำลองสำหรับการพยากรณ์ การจำแนกประเภทข้อมูล หรือเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ของข้อมูล มีการแบ่งการวิเคราะห์ออกเป็น 2 ส่วน ประกอบด้วย วิเคราะห์จากรายวิชาทั้งหมดและวิเคราะห์จากรายวิชาเลือก โดยเครื่องมือที่ใช้ในการวิจัยได้ใช้โปรแกรม Orange Data Mining เป็นซอฟต์แวร์โอเพนซอร์ซสำหรับการเรียนรู้ของเครื่อง สามารถแสดงกระบวนการวิเคราะห์ข้อมูล(Data Analysis Workflow) ด้วยภาพ ในการวิจัยนี้ใช้การสร้างแบบจำลองจาก 5 วิธี ได้แก่ 1)โครงข่ายประสาทเทียม เป็นโครงข่ายคอมพิวเตอร์ที่ได้รับแรงบันดาลใจจากการทำงานของสมองมนุษย์ทำให้คอมพิวเตอร์รู้จักและจดจำข้อมูลต่างๆ ได้จากการเรียนรู้ผ่านตัวอย่าง ในงานวิจัยนี้เลือกใช้แบบเพอร์เซ็ปตรอนหลายชั้น(Multi-Layer Perceptron) ที่มีขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) พื้นฐานโครงข่ายประสาทเทียมประกอบด้วย 3 ส่วน ประกอบด้วยข้อมูลชั้นขาเข้า (Input Layer) ชั้นซ่อนตัว(Hidden Layer) และชั้นขาออก(Output Layer) โดยกำหนดชั้นซ่อนตัวสูงสุด 100 ชั้น และรอบการเรียนรู้จำนวน 200 รอบ 2)การเรียนรู้แบบต้นไม้ตัดสินใจ เป็นการเรียนรู้โดยการจำแนกประเภท(Classification) ออกเป็นกลุ่มต่างๆ โดยใช้คุณลักษณะข้อมูลในการจำแนกประเภทที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณลักษณะใดที่เหมาะสมเป็นตัวกำหนดการจำแนกประเภท ซึ่งมีประโยชน์ช่วยให้สามารถวิเคราะห์ข้อมูลและตัดสินใจได้ดียิ่งขึ้น อีกทั้งผลลัพธ์ที่เป็นโครงสร้างแบบต้นไม้ทำให้มองเห็นความสัมพันธ์และสามารถอธิบายได้ง่าย ต้นไม้ตัดสินใจประกอบด้วยโหนด (Node) กิ่ง (Branch) และใบ (Leaves) JankaKabathova and Martin Drlik (2021) ในงานวิจัยนี้กำหนดค่าไม่เกิน 100 โหนดและหยุดแบ่งโหนดเมื่อโหนดส่วนใหญ่ถึงกำหนดร้อยละ 95 3)การค้นหาเพื่อนบ้านใกล้ที่สุด เป็นการสร้างแบบจำลองจำแนกประเภทข้อมูลโดยการเปรียบเทียบข้อมูลโดยจะค้นหาข้อมูลเดิมที่มีลักษณะใกล้เคียงข้อมูลใหม่มากที่สุดจำนวน k ตัวมาร่วมกันตัดสินว่าข้อมูลใหม่นี้ควรเป็นประเภทใด ในงานวิจัยนี้กำหนดค่า k เท่ากับ 5 และใช้วิธีการวัดระยะทางแบบ Euclidean 4)การเรียนรู้เบย์อย่างง่าย เป็นการวิเคราะห์ที่อาศัยหลักการของความน่าจะเป็นเข้ามาช่วยในการหาคำตอบของประเภทตัวอย่างใหม่เพื่อสร้างเงื่อนไขความน่าจะเป็นของแต่ละความสัมพันธ์ให้อยู่ในรูปอย่างง่าย 5)การสุ่มป่าไม้ เป็นวิธีที่ใช้การสุ่มเลือกคุณลักษณะต่าง ๆ ของชุดข้อมูลจากนั้นนำเอาชุดข้อมูลและคุณลักษณะนี้มาทำการสร้างแบบจำลองการพยากรณ์ด้วยต้นไม้ตัดสินใจหลาย ๆ ต้นและเลือกใช้แบบจำลองที่ได้ประสิทธิภาพดีที่สุด ในงานวิจัยนี้กำหนดค่าจำนวนต้นไม้เท่ากับ 10 และหยุดแบ่งโหนดเมื่อโหนดน้อยกว่า 5 Mustafa Yagci (2022)

3. การประเมินผล(Evaluation) การประเมินประสิทธิภาพของแบบจำลองนั้นจะใช้วิธีการตรวจสอบไขว้(K-Fold Cross Validation) กำหนดค่า K เท่ากับ 10 โดยแบ่งข้อมูลเป็น 10 ชุด ไม่ซ้ำกันโดยใช้ชุดข้อมูลสอน(Training Set) 9 ชุดและชุดข้อมูลทดสอบ(Test Set) 1 ชุด แล้ววนทำซ้ำจนข้อมูลทุกส่วนที่ถูกนำมาทดสอบและใช้ผลจากตาราง Confusion Matrix โดยประเมินจากตัวชี้วัด 4 ตัว ได้แก่ ค่าความถูกต้อง(Accuracy) เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณารวมทุกคลาส ค่าความแม่นยำ(Precision) เป็นการวัดความแม่นยำของข้อมูล โดยพิจารณาแยกทีละคลาส ค่าความระลึก(Recall) เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณาแยกทีละคลาส และค่าความถ่วงดุล(F-Measure) เป็นค่าเฉลี่ยของค่าความแม่นยำและค่าความระลึกสำหรับทุกคลาส Aderibigbe Israel Adekitan and Odunayo Salau (2019)

5. ผลการวิจัย

ผลการเตรียมข้อมูล การสร้างแบบจำลอง และการประเมินประสิทธิภาพของแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬา โดยใช้วิธีการเรียนรู้ของเครื่อง ใช้การวิเคราะห์ทั้งหมด 5 วิธี ได้แก่ โครงข่ายประสาทเทียม การเรียนรู้แบบต้นไม้ตัดสินใจ การค้นหาเพื่อนบ้านใกล้ที่สุด การเรียนรู้เบย์อย่างง่าย และการสุ่มป่าไม้ ผ่านการใช้งานโปรแกรม Orange Data Mining ในการทำการทดลอง โดยข้อมูลที่ใช้ในการวิจัยนี้เป็นข้อมูลรายวิชาที่ลงทะเบียนเรียน ผลการเรียน และเกรดเฉลี่ยสะสมของนักศึกษาระดับปริญญาตรี ภาคปกติ สาขาวิชาวิทยาศาสตร์การกีฬา คณะวิทยาศาสตร์การกีฬาและสุขภาพ มหาวิทยาลัยการกีฬาแห่งชาติ วิทยาเขตมหาสารคาม ที่เริ่มเข้าศึกษาตั้งแต่ปีการศึกษา 2556-2557 และจบการศึกษาที่ผ่านการแปลงข้อมูลจำนวนทั้งสิ้น 132 ระเบียน ผลการเตรียมข้อมูลในส่วนการแปลงข้อมูล แสดงรายละเอียดดังตารางที่ 1

ตารางที่ 1 ผลการเตรียมข้อมูลในส่วนการแปลงข้อมูล

รายละเอียด	ข้อมูลเดิม	การแปลงข้อมูล
รหัสวิชาบังคับ	ศทxxxxxx, วทxxxxxx	C_GExxxxxx, C_SCxxxxxx
รหัสวิชาเลือก	ศทxxxxxx, วทxxxxxx	S_GExxxxxx, S_SCxxxxxx
ผลการเรียนรายวิชา	A, B+, B, C+, C, D+, D	ทำการแปลงข้อมูลออกเป็น 7 กลุ่ม ดังนี้ A แทนด้วย 4 B+ แทนด้วย 3.5 B แทนด้วย 3

รายละเอียด	ข้อมูลเดิม	การแปลงข้อมูล
		C+ แทนด้วย 2.5 C แทนด้วย 2 D+ แทนด้วย 1.5 D แทนด้วย 1
เกรดเฉลี่ยสะสม	ค่าตัวเลขตั้งแต่ 2.00-4.00	ทำการแปลงข้อมูลออกเป็น 4 กลุ่ม ดังนี้ 3.60=<GPA<=4.00 3.25=<GPA<=3.59 2.75=<GPA<=3.24 2.00=<GPA<=2.74

การสร้างแบบจำลองจะแสดงผลตัวอย่างที่ได้จากการเรียนรู้แบบต้นไม้ตัดสินใจเนื่องจากต้นไม้ตัดสินใจง่ายต่อการเข้าใจและทำให้เห็นความสัมพันธ์ของรายวิชาเพื่อวางแผนการลงทะเบียนได้อย่างชัดเจน จากผลการทดลองสามารถนำมาแปลผลเป็นกฎการตัดสินใจที่เกิดขึ้นจากรายวิชาบังคับและวิชาเลือกทั้งหมด 59 รายวิชา สามารถแยกออกมาเป็นตัวอย่างกฎที่รายวิชาเกิดความสัมพันธ์จำนวน 10 กฎ ดังตารางที่ 2

ตารางที่ 2 กฎการตัดสินใจที่เกิดกับรายวิชาทั้งหมด

ลำดับ	รายละเอียด	ผลลัพธ์
กฎข้อที่ 1	If C_SC021004 > 2.5 and C_SC102002> 2.5 and S_SC081249 > 3.5 and C_SC081041 > 3	3.60=<GPA<=4.00
กฎข้อที่ 2	If C_SC021004 > 2.5 and C_SC102002> 2.5 and S_SC081249 > 3.5 and C_SC081041 <= 3	3.25=<GPA<=3.59
กฎข้อที่ 3	If C_SC021004 > 2.5 and C_SC102002> 2.5 and S_SC081249 <= 3.5 and C_GE071004 > 2 and S_SC081256 > 3	3.25=<GPA<=3.59
กฎข้อที่ 4	If C_SC021004 > 2.5 and C_SC102002> 2.5 and S_SC081249 <= 3.5 and C_GE071004 > 2 and S_SC081256 <= 3	3.25=<GPA<=3.59

ลำดับ	รายละเอียด	ผลลัพธ์
กฎข้อที่ 5	If C_SC021004 > 2.5 and C_SC102002> 2.5 and S_SC081249 <= 3.5 and C_GE071004 <= 2	2.75=<GPA<=3.24
กฎข้อที่ 6	If C_SC021004 > 2.5 and C_SC102002<= 2.5 and S_SC084004 > 1.5	2.75=<GPA<=3.24
กฎข้อที่ 7	If C_SC021004 > 2.5 and C_SC102002<= 2.5 and S_SC084004 <= 1.5	2.00=<GPA<=2.74
กฎข้อที่ 8	If C_SC021004 <= 2.5 and C_GE051002 > 2.5 and C_SC081041 > 2.5 and C_SC082141 > 2.5	2.75=<GPA<=3.24
กฎข้อที่ 9	If C_SC021004 <= 2.5 and C_GE051002 > 2.5 and C_SC081041 > 2.5 and C_SC082141 <= 2.5	2.75=<GPA<=3.24
กฎข้อที่ 10	If C_SC021004 <= 2.5 and C_GE051002 > 2.5 and C_SC081041 > 2.5 and C_SC102002> 2.5 and S_SC081249 > 2	2.75=<GPA<=3.24

ผลการทดลองเฉพาะรายวิชาเลือกทั้งหมด 21 รายวิชา สามารถแยกออกมาเป็นตัวอย่างกฎที่รายวิชาเกิดความสัมพันธ์จำนวน 10 กฎ ดังตารางที่ 3

ตารางที่ 3 กฎการตัดสินใจที่เกิดกับรายวิชาเลือก

ลำดับ	รายละเอียด	ผลลัพธ์
กฎข้อที่ 1	If S_GE072006 > 3 and S_SC081249 > 3 and S_SC081203 > 2.5 and S_GE023013 > 2.5	3.60=<GPA<=4.00
กฎข้อที่ 2	If S_GE072006 > 3 and S_SC081249 > 3 and S_SC081203 > 2.5 and S_GE023013 <= 2.5	3.25=<GPA<=3.59
กฎข้อที่ 3	If S_GE072006 > 3 and S_SC081249 > 3 and S_SC081203 <= 2.5	2.75=<GPA<=3.24
กฎข้อที่ 4	If S_GE072006 > 3 and S_SC081249 <= 3 and S_GE023012 > 2.5 and S_GE062012 > 3	3.25=<GPA<=3.59
กฎข้อที่ 5	If S_GE072006 > 3 and S_SC081249 <= 3 and S_GE023012 > 2.5 and S_GE062012 <= 3	2.75=<GPA<=3.24

ลำดับ	รายละเอียด	ผลลัพธ์
กฎข้อที่ 6	If S_GE072006 > 3 and S_SC081249 <= 3 and S_GE023012 > 2.5 and S_GE032015 > 2	2.75=<GPA<=3.24
กฎข้อที่ 7	If S_GE072006 > 3 and S_SC081249 <= 3 and S_GE023012 <= 2.5 and S_GE032015 <= 2	2.00=<GPA<=2.74
กฎข้อที่ 8	If S_GE072006 > 3 and S_GE023012 <= 2.5 and S_SC081203 > 1.5 and S_GE032015 > 2	2.75=<GPA<=3.24
กฎข้อที่ 9	If S_GE072006 <= 3 and S_GE023012 <= 2.5 and S_SC081203 > 1.5 and S_GE032015 <= 2	2.00=<GPA<=2.74
กฎข้อที่ 10	If S_GE072006 <= 3 and S_GE023012 <= 2.5 and S_SC081203 <= 1.5	2.00=<GPA<=2.74

การประเมินผลเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองโดยวัดประสิทธิภาพของแบบจำลองด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล แสดงผลการทดลองออกเป็น 2 ส่วน คือ การเปรียบเทียบประสิทธิภาพของแบบจำลองกับชุดข้อมูลรายวิชาทั้งหมด และการเปรียบเทียบประสิทธิภาพของแบบจำลองกับชุดข้อมูลเฉพาะรายวิชาเลือก แสดงดังตารางที่ 4 และตารางที่ 5

ตารางที่ 4 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองกับชุดข้อมูลรายวิชาทั้งหมด

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
Neural Network	87.9%	0.885	0.879	0.879
Decision Tree	72.0%	0.714	0.720	0.715
K-NN	83.3%	0.837	0.833	0.834
Naive Bayes	69.7%	0.774	0.697	0.716
Random Forest	81.8%	0.809	0.818	0.806

จากตารางที่ 4 พบว่า วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องสูงที่สุดร้อยละ 87.9 วิธีการค้นหาเพื่อนบ้านใกล้ที่สุดให้ค่าความถูกต้องร้อยละ 83.3 วิธีการสุ่มป่าไม้ให้ค่าความถูกต้องร้อยละ 81.8วิธีการเรียนรู้แบบต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 72 และวิธีการเรียนรู้เบย์อย่างง่ายให้ค่าความถูกต้องร้อยละ 69.7 ตามลำดับ

ตารางที่ 5 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองกับชุดข้อมูลเฉพาะรายวิชาเลือก

Algorithm	Accuracy (%)	Precision	Recall	F-Measure
Neural Network	75.0%	0.762	0.750	0.747
Decision Tree	65.9%	0.657	0.659	0.658
K-NN	70.2%	0.716	0.720	0.717
Naive Bayes	61.7%	0.645	0.617	0.633
Random Forest	70.5%	0.702	0.705	0.696

จากตารางที่ 5 พบว่า วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องสูงที่สุดร้อยละ 75 วิธีการสุ่มป่าไม้ให้ค่าความถูกต้องร้อยละ 70.5 วิธีการค้นหาเพื่อนบ้านใกล้ที่สุดให้ค่าความถูกต้องร้อยละ 70.2 วิธีการเรียนรู้แบบต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 65.9 และวิธีการเรียนรู้เบย์อย่างง่ายให้ค่าความถูกต้องร้อยละ 61.7 ตามลำดับ

6. อภิปรายผลการวิจัย

งานวิจัยนี้เป็นการสร้างแบบจำลองการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่องและเปรียบเทียบประสิทธิภาพการพยากรณ์ผลซึ่งวิธีการเรียนรู้ของเครื่องเป็นส่วนหนึ่งของการประยุกต์ใช้ในการทำเหมืองข้อมูลทางการศึกษา โดยพิจารณาจากข้อมูลรายวิชาที่ลงทะเบียนเรียน ผลการเรียน และเกรดเฉลี่ยสะสม แบบจำลองที่ใช้ในการเปรียบเทียบ 5 วิธี ได้แก่ โครงข่ายประสาทเทียม การเรียนรู้แบบต้นไม้ตัดสินใจ การค้นหาเพื่อนบ้านใกล้ที่สุด การเรียนรู้เบย์อย่างง่าย และการสุ่มป่าไม้ ผลการเปรียบเทียบทั้ง 5 วิธี พบว่าแบบจำลองการพยากรณ์ผลกับชุดข้อมูลรายวิชาทั้งหมด วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องสูงที่สุดคิดเป็นร้อยละ 87.9 สำหรับแบบจำลองกับชุดข้อมูลเฉพาะรายวิชาเลือก วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องสูงที่สุดคิดเป็นร้อยละ 75 ซึ่งสอดคล้องกับงานวิจัยของ สุวิมล สิทธิชาติ (2560) ที่ได้ศึกษาเกี่ยวกับการวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูลระหว่างวิธีโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นและวิธีการเรียนรู้แบบต้นไม้ตัดสินใจ พบว่าวิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องสูงที่สุด นอกจากนี้ผู้วิจัยได้อธิบายตัวอย่างความสัมพันธ์ของรายวิชาในรูปแบบของกฎด้วยวิธีการเรียนรู้แบบต้นไม้ตัดสินใจซึ่งพบว่ามีกฎที่แสดงความสัมพันธ์ของรายวิชาจากรายวิชาทั้งหมดที่มีแนวโน้มที่จะสามารถสำเร็จการศึกษาแล้วได้เกียรตินิยมอันดับ 1 คือ ถ้าเรียนวิชา วท021004 เคมีทั่วไป (วิชาบังคับ) ได้ผลการเรียนมากกว่า C+ และ วท102002 สถิติทางวิทยาศาสตร์การกีฬา (วิชาบังคับ) ได้ผลการเรียน

มากกว่า C+ และ วท081249 โยคะ (วิชาเลือก) ได้ผลการเรียนมากกว่า B+ และ วท081041 พื้นฐานวิทยาศาสตร์การกีฬา (วิชาบังคับ) ได้ผลการเรียนมากกว่า Bส่วนกฎที่แสดงความสัมพันธ์ของรายวิชาจากรายวิชาเลือกที่มีแนวโน้มที่จะสามารถสำเร็จการศึกษาแล้วได้เกียรตินิยมอันดับ 1 คือ ถ้าเรียนวิชา ศท072006 การใช้โปรแกรมสำเร็จรูปในชีวิตประจำวัน ได้ผลการเรียนมากกว่า B และ วท 081249 โยคะ ได้ผลการเรียนมากกว่า B และ วท081203 กระบี่กระบอง ได้ผลการเรียนมากกว่า C+ และ ศท023013 การพัฒนาทักษะการเขียนภาษาอังกฤษ ได้ผลการเรียนมากกว่า B

อย่างไรก็ตามแบบจำลองการพยากรณ์ที่ได้นี้เป็นการวิเคราะห์การวางแผนเพื่อที่จะได้เกรดเฉลี่ยสะสมสุดท้ายของการเรียนที่มีแนวโน้มที่จะสามารถสำเร็จการศึกษาได้ ความสัมพันธ์ที่ได้มาจากรายวิชาที่มีความต่อเนื่องกันนี้ไม่ได้คำนึงถึงแผนการลงทะเบียนเรียนตามชั้นปีแต่จะวิเคราะห์ว่าวางแผนการเรียนและมีผลการเรียนเท่าไรจึงจะสามารถจบตามเกณฑ์ที่นักศึกษาต้องการได้ แบบจำลองที่เหมาะสมจะนำไปพัฒนาเพื่อให้สามารถนำไปใช้ในวางแผนการลงทะเบียนของนักศึกษาสาขาวิชาวิทยาศาสตร์การกีฬาในระหว่างที่ทำการศึกษาอยู่เพื่อลดการพ้นสภาพนักศึกษาและสามารถจบการศึกษาได้ในระยะเวลาตามหลักสูตรกำหนด อีกทั้งความถูกต้องของแบบจำลองที่ใช้วิธีการเรียนรู้ของเครื่องขึ้นอยู่กับจำนวนข้อมูลที่นำมาใช้ในการสร้างแบบจำลองดังนั้นในอนาคตอาจจะต้องมีการเพิ่มเติมจำนวนข้อมูลของนักศึกษาที่จบการศึกษาตามหลักสูตรที่ปรับปรุงเพิ่มเติมเพื่อให้แบบจำลองที่สร้างขึ้นมีความครอบคลุมและมีประสิทธิภาพเพิ่มมากยิ่งขึ้น

7. ข้อเสนอแนะ

7.1 ข้อเสนอแนะเชิงนโยบาย

7.1.1การกำหนดรายวิชาเรียนในหลักสูตรควรพิจารณาความยากง่ายของแต่ละรายวิชารวมทั้งการเปิดรายวิชาให้มีลำดับก่อนหลังที่เหมาะสมเพื่อเอื้อให้นักศึกษาที่มีผลการเรียนต่ำกว่าเกณฑ์สามารถสำเร็จการศึกษาได้

7.2 ข้อเสนอแนะสำหรับผู้ปฏิบัติ

7.2.1 เนื่องจากข้อมูลที่นำมาเรียนรู้มีจำนวนข้อมูลค่อนข้างน้อยหากนำเทคนิคการเพิ่มจำนวนข้อมูลเพื่อช่วยเพิ่มประสิทธิภาพในการพยากรณ์ผลของแบบจำลองให้ดียิ่งขึ้น

7.2.2 ข้อมูลการลงทะเบียนของนักศึกษายังเป็นรายวิชาของหลักสูตรเก่าอยู่ ซึ่งในปัจจุบันได้ใช้หลักสูตรปรับปรุง พ.ศ. 2563 ดังนั้นการเก็บข้อมูลรายวิชาของหลักสูตรใหม่ที่เป็นปัจจุบันจะทำให้การพยากรณ์ผลการเรียนของนักศึกษาหลักสูตรปัจจุบันสอดคล้องกับหลักสูตรที่ศึกษา

7.3 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

7.3.1 การสร้างแบบจำลองอาจเพิ่มวิธีการอื่นๆ เข้ามาเพื่อสามารถเรียนรู้และพยากรณ์ผลออกมาได้ค่าความถูกต้องและมีประสิทธิภาพที่ดีกว่า

8. บรรณานุกรม

บีสุดา ดาวเรือง, จรัญ แสนราช และอนิราช มิ่งขวัญ. (2564). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือวิทยาเขตปราจีนบุรี. **วารสารวิชาการครุศาสตร์อุตสาหกรรม พระจอมเกล้าพระนครเหนือ.** 12(2). 136-144.

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

พงษ์เทพ รักผกาวงศ์ และอุไรวรรณ์ รักผกาวงศ์. (2558). การใช้กฎความสัมพันธ์ ร่วมกับฟัซซี่กฎความสัมพันธ์ เพื่อคาดการณ์ผลการเรียนของนักศึกษา. **วารสารวิทยาศาสตร์ มข.** 43(3). 542-551.

สุวิมล สิทธิชาติ. (2560). การวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูล. **วารสารเทคโนโลยีสารสนเทศ.**13(2). 20-28.

อนันต์ ปินะเต, ฉัตรเกล้า เจริญผล และแกมกาญจน์ สมประเสริฐศรี. (2557). การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี. **วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม.** 33(6). 648-656.

Adekitan, A. I., &Salau, O. (2019). Toward an improved learning process: the relevance of ethnicity to data mining prediction of students' performance. **SN Applied Sciences.** 2(1). 1-15.

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). **Predicting Students Performance in Educational Data Mining.** 2015 International Symposium on Educational Technology (ISET). Central China Normal University in Wuhan, China. 27-29 July 2015. 125-128.

Kabathova, J., &Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. **Applied Sciences.** 11(7). 3130.

Yagci, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. **Smart Learning Environments.** 9(11). 1-19.