

การเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์

A Forecasting Performance comparison between Back-propagation Neural Network and Naïve Bayes

นางสาวทิพย์หทัย ทองธรรมชาติ*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์ ขั้นตอนดำเนินงานวิจัย ประกอบด้วย การเตรียมข้อมูลและการวิเคราะห์ข้อมูลเบื้องต้น โดยใช้เทคนิคเคมีนในการจัดกลุ่มผลสัมฤทธิ์ทางการเรียน ใช้เทคนิค Information Gain และเทคนิค Gain Ratio ในการคัดเลือกตัวแปรสำหรับใช้ในการพยากรณ์ เทคนิคที่ใช้ในการพยากรณ์ คือ เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์ ข้อมูลที่ใช้ในงานวิจัย ได้แก่ ข้อมูลภูมิหลัง ข้อมูลผลการเรียนในชั้นปีที่ 1 และ 2 ของนักศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏกำแพงเพชร จำนวนทั้งสิ้น 358 คน จำแนกเป็นข้อมูลชุดฝึกสอน จำนวน 231 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2554–2557 ข้อมูลชุดทดสอบ จำนวน 70 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2558 และข้อมูลชุดตรวจสอบ จำนวน 57 คน ซึ่งเป็นนักศึกษาที่กำลังจะสำเร็จการศึกษาประจำปีการศึกษา 2559 ผลการวิจัยพบว่า มีเพียงข้อมูลผลการเรียนในชั้นปีที่ 1 และ 2 เท่านั้น ที่เป็นข้อมูลสำคัญในการพยากรณ์ผลสัมฤทธิ์ทางการเรียน และการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ค่าความถูกต้องดีกว่าเมื่อเทียบกับการพยากรณ์ด้วยเทคนิคเบย์ ด้วยค่าความถูกต้อง 85.71%, 85.71% และ 82.46% ในข้อมูลชุดฝึกสอน ข้อมูลชุดทดสอบ และข้อมูลชุดตรวจสอบ ตามลำดับ

คำสำคัญ: การทำเหมืองข้อมูล, การพยากรณ์

Abstract

The objective of this research was a forecasting performance comparison between Back-propagation Neural Network and Naïve Bayes. The research process includes of data preparation and preliminary analysis. K-means technique was used to transform learning achievement into categorical data. Information Gain Technique and Gain Ratio Technique were used to select the variables for forecasting. Then the Naïve Bayes Technique and Back-propagation Neural Network Technique were applied to forecasting the students' achievement.

The data used in this research include the students' background information and studied grades in the first and second years of Bachelor's degree students who studied in Business Computer Department, Faculty of Management Science, Kamphaeng Phet Rajabhat University and Kamphaeng Phet Rajabhat University Maesot. There are 358 of them which are training data (231 records) who

*วิทยาศาสตร์มหาบัณฑิต (วท.ม.) สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยนครสวรรค์ อาจารย์ประจำสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏกำแพงเพชร แม่สอด

graduated in 2011-2014, testing data (70 records) who graduated in 2015 and validation data (57 records) that will be graduated in 2016.

The research results show that only students' grades in the first and second years are important in forecasting the academic achievement of students. The prediction of Back-propagation Neural Network Technique has accuracies better than the results from Naïve Bayes Technique with the accuracies of 85.71%, 85.71% and 82.46% in training data, testing data and validation data, respectively.

Keywords: Data Mining, forecasting

บทนำ

หลักสูตรบริหารธุรกิจบัณฑิต สาขาคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยราชภัฏกำแพงเพชร มีการผลิตบัณฑิตมาแล้วหลายรุ่น ทางสาขาได้มีการปรับปรุงหลักสูตรให้มีความทันสมัยอยู่เสมอ โดยปกติแล้วจะมีนักศึกษาที่ไม่สามารถเลือกสาขาวิชาเรียนที่เหมาะสมกับความสามารถของตนเอง จนทำให้นักศึกษาบางส่วนขอย้ายสาขาวิชา ขอพักการเรียน ลาออก และพ้นสภาพจากการเป็นนักศึกษา ซึ่งสาเหตุอาจเกิดจากที่นักศึกษาสอบไม่ผ่าน ส่งผลให้มีเกรดเฉลี่ยสะสมอยู่ในระดับต่ำ จึงทำให้นักศึกษาไม่สามารถเรียนจนสำเร็จการศึกษาได้ และในการเลือกสาขาวิชาเรียนนั้น นักศึกษาไม่ทราบถึงสาขาวิชาที่เหมาะสมกับความสามารถของตนเอง ทำให้เมื่อเข้าไปเรียนแล้วพบว่าตนเองไม่ชอบในสาขาวิชาที่เลือกเรียน หรือความสามารถไม่เหมาะสมกับสาขาวิชาเหล่านั้น ทางมหาวิทยาลัยมีการจัดเก็บข้อมูลนักศึกษาไว้เป็นจำนวนมาก โดยข้อมูลเหล่านี้สามารถนำมาใช้ประโยชน์ได้ แต่ไม่ถูกนำมาใช้ให้เกิดประโยชน์อย่างจริงจัง

การพยากรณ์ หมายถึง การคาดการณ์เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศประกอบการตัดสินใจซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดที่จะต้องทำก่อนการวางแผนหรือการเตรียมการที่จะเริ่มทำอะไรเพื่อความถูกต้องและแม่นยำในการตัดสินใจ (มนต์ชัย เทียนทอง, 2548) ปกติแล้วการพยากรณ์จะใช้หลักการทางสถิติเป็นส่วนใหญ่ นอกจากวิธีการทางสถิติทั่วไปแล้ว เทคนิคเหมืองข้อมูล (Data Mining Techniques) ก็เป็นอีกวิธีหนึ่งที่นักวิจัยนิยมนำมาใช้ในการพยากรณ์สิ่งที่ต้องการรู้ที่จะเกิดขึ้นในอนาคต เป็นการพยากรณ์หาค่าที่ต้องการรู้จากข้อมูลที่มีอยู่ เนื่องจากการทำเหมืองข้อมูลเป็นกระบวนการสกัดความรู้ที่น่าสนใจจากข้อมูลที่มีอยู่จำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ถูกซ่อนอยู่ในข้อมูล (ชนวัฒน์ ศรีสอาน, 2551) จะพบว่านักวิจัยหลายท่านได้นำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้กับปัญหาต่างๆ และได้ผลการทดลองที่มีประสิทธิภาพ เช่น การพัฒนาแบบจำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษา สาขาคอมพิวเตอร์ โดยใช้โครงข่ายประสาทเทียม (ฉันทฐา ฝิวมา, 2558, หน้า 1-16) การศึกษาเชิงเปรียบเทียบในการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำเหมืองข้อมูล เพื่อพยากรณ์โอกาสการสำเร็จการศึกษาของนักศึกษา (พรเทพ คงไชย, 2555) การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา (ชุติมา อุตมะมุณี และ ประสงค์ ปรานีตพลกรัง, 2553, หน้า 39-48) การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา (เสกสรรค์ วิสัยลักษณ์, วิภา เจริญภัณฑารักษ์ และ ดวงดาว วิชาดากุล, 2558, หน้า 1-17) การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา (สุธีรา วงศ์

อนันทรัพย์, ต่องใจ แยมผกา และ อรวรรณ มุสิกะ, 2559, หน้า 12-19) การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซอฟต์แวร์แมชชีน (พรณิภา บุตรเอก และสุรเดช บุญลือ, 2557, หน้า 40-49) การเปรียบเทียบอัลกอริทึมเหมือนข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อผลการเรียนของนักศึกษา (เยาวภา ภารสำเร็จ, จิรัฐธา ภูบุญยอบ และ วิรัตน์ พงษ์ศิริ, 2556, หน้า 281-289) การหาปัจจัยที่ส่งต่อความเสี่ยงของนักเรียนอ่อนด้วยเทคนิคกฎความสัมพันธ์ กรณีศึกษา: มหาวิทยาลัยสงขลานครินทร์ (ณัฐธิดา สุวรรณโณ และอันธิกา สิงห์เอี่ยม, 2554, หน้า 65-79) การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์ (นิภาพร ชนะมาร และพรณิ สิทธิเดช, 2557, หน้า 31-45) เป็นต้น เมื่อสามารถหาแบบจำลองที่มีความน่าเชื่อถือได้ ก็สามารถที่จะนำข้อมูลนักศึกษาที่เรียนในระดับชั้นปีที่ 1 มาทำการทดสอบโดยผ่านแบบจำลองนี้ จะสามารถพยากรณ์ได้ว่านักศึกษามีโอกาสสำเร็จการศึกษาในสาขาวิชาที่ตนเองได้เลือกเรียนหรือไม่ และถ้านักศึกษาคนดังกล่าวอยู่ในกลุ่มที่มีผลการเรียนต่ำ ทางสาขาวิชาจะได้ให้คำแนะนำหรือเข้าไปแก้ปัญหาให้ได้อย่างถูกต้องและเหมาะสม

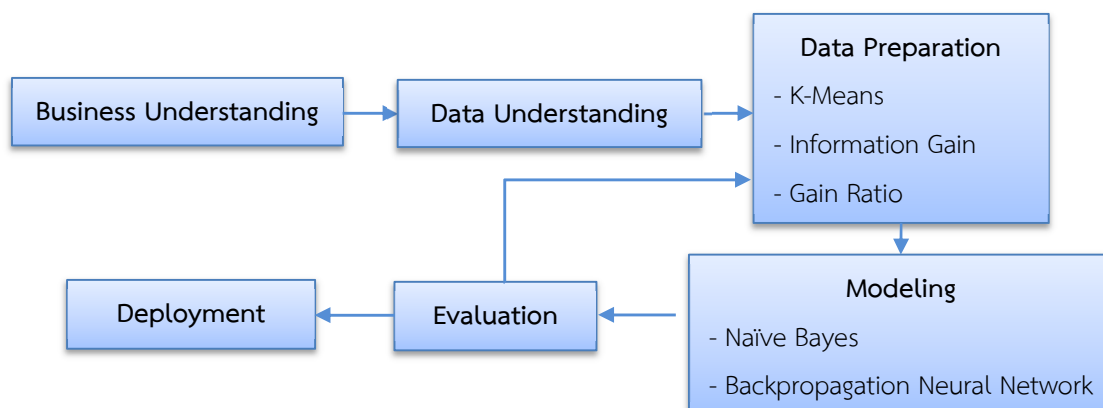
ดังนั้นในงานวิจัยนี้ จึงมีวัตถุประสงค์เพื่อการเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอ็พเบย์ เพื่อทำให้นักศึกษาสามารถปรับปรุงผลการเรียนของตนเองให้ดีขึ้นและทันเวลา โดยใช้ข้อมูลภูมิหลัง และผลการเรียนในชั้นปีที่ 1 และ 2 เป็นหลัก จากการศึกษางานวิจัยที่ผ่านมา ผู้วิจัยจึงเลือกใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ มาเปรียบเทียบกับเทคนิคเบย์ พบว่าสามารถใช้แก้ปัญหาด้วยการลดข้อผิดพลาดในการพยากรณ์

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอ็พเบย์

ระเบียบวิธีการวิจัย

การวิจัยนี้ได้ดำเนินงานโดยประยุกต์ตามแนวทางในการทำเหมืองข้อมูล ที่เรียกว่า กระบวนการมาตรฐานอุตสาหกรรม หรือ CRISP-DM (Chapman et al. 2000) ที่ได้รับความนิยมมากในปัจจุบัน ซึ่งมีขั้นตอนการดำเนินการวิจัย ดังภาพที่ 1 รายละเอียดการทำงานแต่ละขั้นตอน มีดังนี้



ภาพที่ 1 ขั้นตอนการดำเนินการวิจัย

1. ศึกษาโจทย์ที่ต้องการทำ (Business Understanding)

ผู้วิจัยได้ทำการศึกษาโครงสร้างหลักสูตรบริหารธุรกิจบัณฑิต สาขาวิชาคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยราชภัฏกำแพงเพชร จากโครงสร้างของหลักสูตรได้มีการกำหนดเป็นหมวดวิชาศึกษาทั่วไป (กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชาคณิตศาสตร์ วิทยาศาสตร์ และเทคโนโลยี) หมวดวิชาเฉพาะ (จำแนกเป็นกลุ่มวิชาแกน กลุ่มวิชาบังคับ และกลุ่มวิชาเลือก) และหมวดวิชาเลือกเสรี

2. ศึกษาข้อมูลที่ใช้ในงานวิจัย (Data Understanding)

ผู้วิจัยได้เลือกใช้ข้อมูลภูมิหลังและข้อมูลผลการเรียนรายวิชาในชั้นปีที่ 1 และ 2 ของนักศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจ ที่สำเร็จการศึกษาประจำปีการศึกษา 2554 – 2557 ได้ตัวแปรทั้งหมดจำนวน 35 ตัวแปร สามารถจำแนกเป็นตัวแปรต้น 34 ตัวแปร และตัวแปรตาม 1 ตัวแปร

3. การเตรียมข้อมูล (Data Preparation)

3.1 ทำความสะอาดข้อมูล หลังจากสำรวจข้อมูลแล้ว พบว่าข้อมูลยังไม่สมบูรณ์ เช่น ค่าว่าง (Missing Value) และมีสิ่งรบกวน (Noisy Data) แก้ไขโดยการแทนค่าข้อมูลที่ถูกต้องไปแทนที่ข้อมูลเดิม

3.2 แปลงข้อมูล เนื่องจากข้อมูลมีทั้งที่เป็นตัวเลข และข้อมูลที่เป็นตัวอักษร ไม่อยู่ในรูปแบบที่สามารถวิเคราะห์ได้ จึงต้องทำการแทนค่าข้อมูลให้อยู่ในรูปแบบที่สามารถวิเคราะห์ได้

3.3 ผู้วิจัยได้ทำการคัดเลือกข้อมูล โดยทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้น กับตัวแปรตาม

3.4 ศึกษาข้อมูลผลสัมฤทธิ์ทางการเรียนของนักศึกษา พบว่าอยู่ระหว่าง 2.00-3.64 ดังนั้นผู้วิจัยจึงใช้เทคนิคเคมีน ($k=3$) แบ่งกลุ่มผลสัมฤทธิ์ทางการเรียนออกเป็น 3 กลุ่ม คือ กลุ่มเรียนอ่อน (L) มีผลสัมฤทธิ์ทางการเรียนระหว่าง 2.00-2.59 กลุ่มเรียนปานกลาง (M) มีผลสัมฤทธิ์ทางการเรียนระหว่าง 2.60-2.99 และกลุ่มเรียนเก่ง (H) คือ มีผลสัมฤทธิ์ทางการเรียนเท่ากับ 3.00 ขึ้นไป

3.5 ใช้เทคนิคการคัดเลือกชุดตัวแปร Information Gain และ Gain Ratio ร่วมกับวิธีการค้นหาแบบจัดลำดับช่วยลดจำนวนตัวแปร ผลของการคัดเลือกตัวแปรจากทั้งหมด 34 ตัวแปร ลดเหลือเพียง 25 ตัวแปร ข้อมูลที่ใช้ในงานวิจัยมีจำนวนทั้งสิ้น 358 คน จำแนกเป็นข้อมูลชุดฝึกสอน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2554 – 2557 จำนวน 231 คน ข้อมูลชุดทดสอบ จำนวน 70 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2558 และข้อมูลชุดตรวจสอบ จำนวน 57 คน ซึ่งเป็นนักศึกษาที่กำลังจะสำเร็จการศึกษาประจำปีการศึกษา 2559

4. การสร้างตัวแบบ (Modeling)

ในขั้นตอนนี้ เป็นการนำข้อมูลหลังการคัดเลือกตัวแปรที่สำคัญด้วยเทคนิค Information Gain และ Gain Ratio ที่เหลือจำนวนตัวแปรต้น 25 ตัวแปร เพื่อทดลองการจำแนกประเภทข้อมูลด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคเบย์ และเปรียบเทียบการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักศึกษา พบว่าค่าความถูกต้องต่ำจึงได้ทำการรวมกลุ่มวิชาเข้าด้วยกัน ปรากฏว่าเหลือตัวแปรต้น 5 ตัวแปร คือ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชาคณิตศาสตร์ วิทยาศาสตร์ และเทคโนโลยี กลุ่มวิชาบังคับ และตัวแปรตาม

1 ตัวแปร คือ ผลสัมฤทธิ์ทางการเรียน

จากผลการทดลองจะเห็นว่า ข้อมูลภูมิหลัง ที่ถูกตัดออกไปทำให้ค่าความถูกต้องสูงขึ้น ดังนั้นหากต้องการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักศึกษา ก็สามารถเลือกใช้ข้อมูลที่จำเป็น นั่นคือตัวแปรต้น 5 ตัวแปร และแสดงค่าความถูกต้อง ดังตารางที่ 1

ตารางที่ 1 ค่าความถูกต้องของข้อมูลชุดฝึกสอน

	Naïve Bayes	Backpropagation Neural Network
ชุดฝึกสอน	85.71%	85.71%

5. การประเมินผล (Evaluation)

ในงานวิจัยนี้ใช้การทดสอบแบบไขว้ทับ (k - Fold Cross Validation) แบบ 10 ส่วน ดังภาพที่ 2 แล้วทำการทดสอบเพื่อประเมินประสิทธิภาพ

การวัดค่าประสิทธิภาพของเทคนิควิธีต่างๆ จะต้องทำการเลือกข้อมูลสำหรับเรียนรู้ (Training Set) และข้อมูลสำหรับทดสอบ (Testing Set) ในงานวิจัยนี้เลือกใช้วิธีสุ่มเลือกแบ่งข้อมูลแบบความเที่ยงตรง k กลุ่ม (k-Fold Cross Validation) โดยเริ่มจากการแบ่งชุดข้อมูลออกเป็น ส่วนๆ ให้เท่าๆ กัน ต่อจากนั้นให้นำข้อมูลบางส่วนมาทำการเรียนรู้ และนำข้อมูลบางส่วนมาทำการทดสอบแบบจำลองที่ได้จากการเรียนรู้ โดยในการทำงานจะทำการเลือกสุ่มข้อมูลออกเป็น k ชุดที่เท่าๆ กัน ในการทดลองครั้งแรก ข้อมูลชุดที่ 1 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดเรียนรู้ ในการทดลองครั้งที่ 2 ข้อมูลชุดที่ 2 เป็นข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดเรียนรู้ ทำจนกระทั่งข้อมูลทุกชุดได้ถูกนำมาเป็นข้อมูลชุดทดสอบและข้อมูลชุดเรียนรู้ ซึ่งจะมีการทดลองทั้งหมด k ครั้ง ในงานวิจัยนี้ได้เลือกใช้ค่า k = 10 ดังอธิบายในตารางที่ 2

ตารางที่ 2 10-Fold Cross Validation

รอบที่	ข้อมูลชุดเรียนรู้									ข้อมูลชุดทดสอบ
	2	3	4	5	6	7	8	9	10	
รอบที่ 1										1
รอบที่ 2	1									2
รอบที่ 3	1	2								3
รอบที่ 4	1	2	3							4
รอบที่ 5	1	2	3	4						5
รอบที่ 6	1	2	3	4	5					6
รอบที่ 7	1	2	3	4	5	6				7
รอบที่ 8	1	2	3	4	5	6	7			8
รอบที่ 9	1	2	3	4	5	6	7	8		9
รอบที่ 10	1	2	3	4	5	6	7	8	9	10

การคำนวณประสิทธิภาพของแบบจำลอง สามารถคำนวณได้จากตาราง เมทริกซ์ความสับสน ซึ่งเป็นตารางสรุปจำนวนข้อมูลที่ตัวแบบมีการจำแนกได้อย่างถูกต้องและไม่ถูกต้อง ผลการทดลองแสดงดังตารางที่ 3

ตารางที่ 3 เมทริกซ์ความสับสน (confusion matrix)

		ผลการจำแนก	
		คำตอบเป็นบวก	คำตอบเป็นลบ
ค่าที่แท้จริง	คำตอบเป็นบวก	TP (True Positive)	FN (False Negative)
	คำตอบเป็นลบ	FP (False Positive)	TN (True Negative)

โดยที่

TP คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นบวก

TN คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นลบ

FP คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นบวก ซึ่งค่าที่แท้จริงเป็นลบ

FN คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นลบ ซึ่งค่าที่แท้จริงเป็นบวก

แล้วทำการวัดประสิทธิภาพตัวแบบการพยากรณ์โดยใช้เกณฑ์การวัดประสิทธิภาพของตัวแบบรู้จำด้วยวิธี Predictive Modeling (Bramer, 2007) ซึ่งประกอบด้วยค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าประสิทธิภาพโดยรวม (F-Measure) และค่าความถูกต้อง (Accuracy) ซึ่งมีค่าอยู่ระหว่าง 0 - 1 ซึ่ง 1 หมายถึง ประสิทธิภาพดี ดังสมการ (1) (2) (3) และ (4) ตามลำดับ

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

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

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

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

ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ย (Root Mean Squared Error: RMSE) ดังสมการ (5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (5)$$

โดยที่

y_j คือ ค่าข้อมูลจริง

\hat{y}_j คือ ค่าที่ได้จากการพยากรณ์

ในงานวิจัยนี้ได้เลือกใช้วิธีประเมินประสิทธิภาพของการพยากรณ์ด้วยค่าความถูกต้อง โดยใช้ข้อมูลฝึกสอน ทดลองปรับค่าพารามิเตอร์ที่เหมาะสมของทั้งเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอ์ฟเบย์และได้นำมาทดลองกับข้อมูลชุดทดสอบเปรียบเทียบประสิทธิภาพการทำงานของข้อมูลทั้งสองชุดเพื่อป้องกันการเกิด Over-fitting นอกจากนั้นยังได้ทำการทวนสอบผลการทดลองกับข้อมูลอีกชุดหนึ่ง คือ ชุดตรวจสอบ เพื่อเพิ่มความเชื่อมั่นของตัวจำแนกประเภท ดังแสดงในตารางที่ 4

ตารางที่ 4 ค่าความถูกต้องของข้อมูลชุดทดสอบ และชุดตรวจสอบ

	Naïve Bayes	Backpropagation Neural Network
ชุดทดสอบ	81.43%	85.71%
ชุดตรวจสอบ	80.70%	82.46%

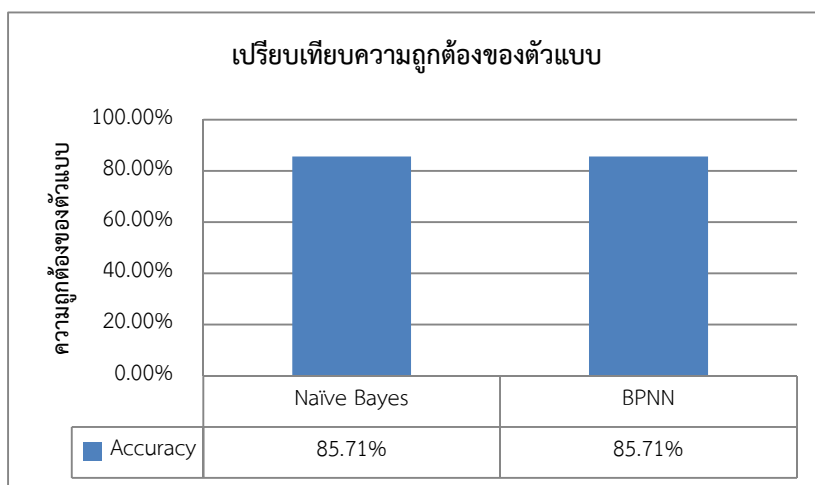
6. นำตัวแบบมาใช้งาน (Deployment)

หลังจากทำการประเมินผลตัวจำแนกของข้อมูลชุดฝึกสอนข้อมูลชุดทดสอบ และข้อมูลชุดตรวจสอบเรียบร้อยแล้ว ได้ผล สามารถนำตัวแบบที่ได้สร้างขึ้นมาใช้ประโยชน์จริงในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักศึกษาที่จะสำเร็จการศึกษาในปี พ.ศ. 2559 และในรุ่นต่อไปได้

ผลการศึกษาหรือผลการวิจัย

ผลการวิเคราะห์เปรียบเทียบค่าความถูกต้องของตัวแบบ

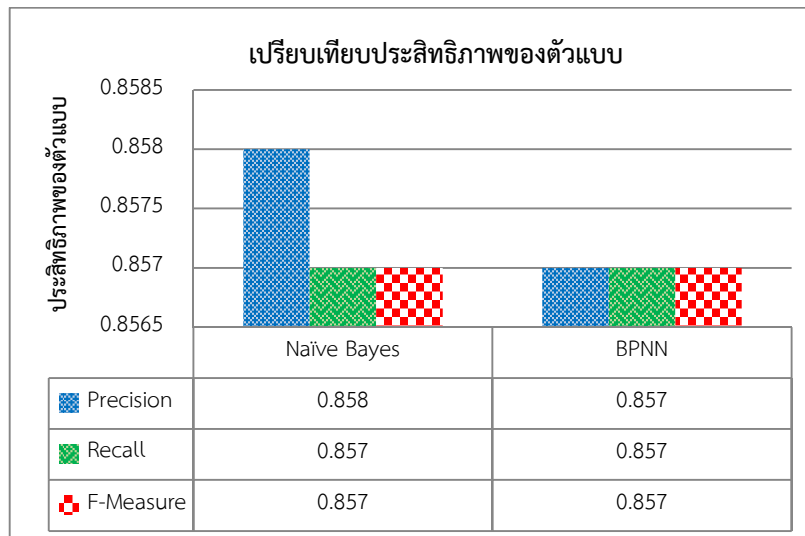
จากการนำข้อมูลภูมิหลังและข้อมูลผลการเรียนรายวิชาในชั้นปีที่ 1 และ 2 ของนักศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏกำแพงเพชร ที่สำเร็จการศึกษาประจำปีการศึกษา 2554 – 2557 จำนวนทั้งสิ้น 358 คน จำแนกเป็นข้อมูลชุดฝึกสอน จำนวน 231 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2554–2557 ข้อมูลชุดทดสอบ จำนวน 70 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2558 และข้อมูลชุดตรวจสอบ จำนวน 57 คน ซึ่งเป็นนักศึกษาที่กำลังจะสำเร็จการศึกษาประจำปีการศึกษา 2559 ไปพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์ ด้วยวิธีไขว้ทบ 10 ส่วน (10-Folds Cross Validation) พบว่าตัวแบบที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ค่าความถูกต้องในการพยากรณ์สูงกว่าตัวแบบที่ใช้เทคนิคนาอิวเบย์ แสดงดังภาพที่ 2



ภาพที่ 2 เปรียบเทียบความถูกต้องของตัวแบบ

ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพของตัวแบบ

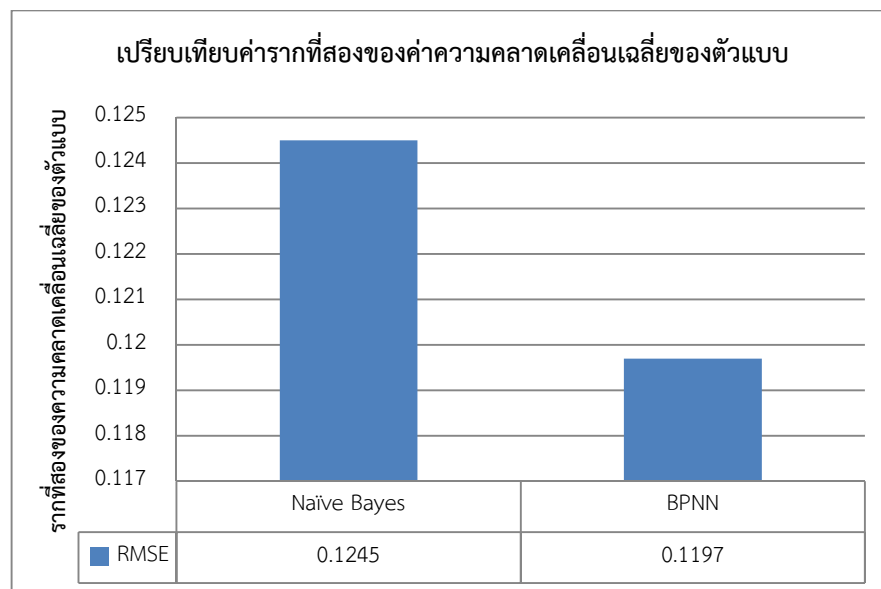
ผลการทดสอบประสิทธิภาพของตัวแบบโดยใช้ค่าความแม่นยำ ค่าความระลึกลับ และค่าประสิทธิภาพโดยรวม พบว่าตัวแบบที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ประสิทธิภาพในการพยากรณ์สูงกว่าตัวแบบที่ใช้เทคนิคนาอิวเบย์ แสดงดังภาพที่ 3



ภาพที่ 3 เปรียบเทียบประสิทธิภาพของตัวแบบ

ผลการวิเคราะห์เปรียบเทียบค่ารากที่สองของความคลาดเคลื่อนของตัวแบบ

นอกจากนี้ เมื่อทำการเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อน พบว่าตัวแบบที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ ให้ค่าความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.1245 ตัวแบบที่ใช้เทคนิคนาอิวเบย์ ให้ค่าความคลาดเคลื่อนรองลงมาเท่ากับ 0.1197 แสดงดังภาพที่ 4



ภาพที่ 4 เปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยของตัวแบบ

ดังนั้นตัวแบบที่เหมาะสมที่สุดในการแนะแนวทางการศึกษาที่เหมาะสมแก่นักศึกษา คือ ตัวแบบที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ เนื่องจากเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับนั้นให้ประสิทธิภาพค่าความถูกต้องในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักศึกษาดีที่สุด และให้ค่าความคลาดเคลื่อนน้อยที่สุด

อภิปรายผล

การเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์ ให้ค่าความถูกต้อง 85.71%, 85.71% และ 82.46% ในข้อมูลชุดฝึกสอน ข้อมูลชุดทดสอบ และข้อมูลชุดตรวจสอบ ตามลำดับ ซึ่งผลการวิจัยนี้สอดคล้องกับงานวิจัยของ ปวีณา ชัยวนารมย์ (2559, หน้า 406-416) ที่ได้วิจัยเรื่อง การพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดความเครียดในหลายระดับด้วยเทคนิคการทำเหมืองข้อมูล ผลของงานวิจัยนี้พบว่า แบบจำลองที่เหมาะสมที่สุดในการนำมาใช้เพื่อพยากรณ์ความเครียด คือ แบบจำลองที่สร้างจากอัลกอริทึม Multilayer Perceptron (MLP) ให้ค่าความถูกต้องสูงที่สุด เท่ากับ 81%

ซึ่งจะเห็นได้ว่าการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ เหมาะสำหรับนำไปพัฒนาระบบเพื่อใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนต่อไป

ข้อเสนอแนะ

ข้อเสนอแนะที่ได้จากการวิจัย

การเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และเทคนิคนาอิวเบย์ มีความแม่นยำสูง สามารถนำไปใช้ในการแนะนำหรือประกอบการตัดสินใจเรียนของนักศึกษา และเป็นการวางแผนการเรียน สร้างแรงจูงใจในการศึกษาเพื่อให้นักศึกษามีผลการเรียนที่สูงขึ้น ทำให้สามารถวางแผนการเรียนได้อีกทั้งเป็นเครื่องมือให้อาจารย์ที่ปรึกษาให้คำแนะนำในการศึกษาต่อแก่นักศึกษาได้

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการเพิ่มตัวแปรอื่นๆ ที่น่าจะมีผลต่อการเรียนของนักศึกษามาพิจารณาเป็นตัวแปรในการพยากรณ์
2. ควรมีการเพิ่มจำนวนข้อมูลของนักศึกษาให้มากพอในการทำเหมืองข้อมูล
3. ควรใช้เทคนิคเหมืองข้อมูลเทคนิคอื่นๆ มาเปรียบเทียบเพื่อให้ได้ผลการพยากรณ์ที่มีประสิทธิภาพมากขึ้น
4. นำผลการวิจัยที่ได้มาประยุกต์ใช้ในการพัฒนาระบบในการแนะแนวทางการศึกษาที่เหมาะสมแก่นักศึกษา

เอกสารอ้างอิง

- ชนวัฒน์ ศรีสอาน. (2551). ฐานข้อมูล คลังข้อมูลและเหมืองข้อมูล (พิมพ์ครั้งที่ 2). ปทุมธานี: มหาวิทยาลัยรังสิต.
- ชุตินา อุตมะมุณี และ ประสงค์ ปราณีตพลกรัง. (2553, กรกฎาคม-ธันวาคม). การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ, 1(2), 39-48.
- ณัฐภา ผิวมา. (2558, พฤษภาคม - สิงหาคม). การพัฒนาแบบจำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษา คอมพิวเตอร์ โดยใช้โครงข่ายประสาทเทียม. วารสารปัญญาภิวัฒน์, 7(2), 1-16.
- ณัฐธิดา สุวรรณโณ และอันธิกา สิงห์เอี่ยม. (2554, มกราคม - มิถุนายน). การหาปัจจัยที่ส่งผลต่อความเสถียรของนักศึกษาเรียนอ่อน ด้วยเทคนิคกฎความสัมพันธ์ กรณีศึกษา : มหาวิทยาลัยสงขลานครินทร์. วารสารวิทยาการจัดการ, 28(1), 65-79.
- นิภาพร ชนะมาร และพรณี สิทธิเดช. (2557, กรกฎาคม - ธันวาคม). การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์. วารสารมหาวิทยาลัยราชภัฏสกลนคร, 6(12), 31-45.

- ปวีณา ชัยวนารมย์. (2559). การพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดความเครียดในหลายระดับด้วยเทคนิคการทำเหมืองข้อมูล ในการประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ครั้งที่ 1. วันที่ 22 มิถุนายน พ.ศ. 2559 (หน้า 406-416). พระนครศรีอยุธยา: มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ.
- พรรณนิภา บุตรเอก และสุรเดช บุญลือ. (2557, พฤศจิกายน – ธันวาคม). การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซอฟต์แวร์เวกเตอร์แมชชีน. Veridian E-Journal Science and Technology Silpakorn University, 1(6), 40-49.
- พรเทพ คงไชย. (2555). การศึกษาเชิงเปรียบเทียบในการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำเหมืองข้อมูลเพื่อพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์.
- มนต์ชัย เทียนทอง. (2548). สถิติและวิธีการวิจัยทางเทคโนโลยีสารสนเทศ. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- เยาวภา ภารสำเร็จ, จิรัฏฐา ภูบุญอบ และ วิรัตน์ พงษ์ศิริ. (2556). การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อระดับผลการเรียนของนักศึกษา. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, ฉบับพิเศษ การประชุมทางวิชาการ มหาวิทยาลัยมหาสารคาม ครั้งที่ 9, 281-289.
- สุธีรา วงศ์อนันทรัพย์, ต้องใจ แยมผกา และ อรวรรณ มุสิกะ. (2559, มกราคม - มิถุนายน). การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา. วารสารสังคมศาสตร์, 5(1), 12-19.
- เสกสรรค์ วิลัยลักษณ์, วิภา เจริญภัณฑารักษ์ และ ดวงดาว วิชาตากุล. (2558, กรกฎาคม-ธันวาคม). การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา. Veridian E-Journal Science and Technology Silpakorn University, 2(2), 1-17.
- Chapman, P. et al. (2000). CRISP-DM 1.0-Step-by-step data mining guide. Technical report. The CRISP-DM.
- Max Bramer. (2007). Principles of Data Mining. Springer-Verlag London Limited.