Technical Education Journal : King Mongkut's University of Technology North Bangkok
Vol. 12 No. 2 May – August 2021

การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

บีสุดา ดาวเรือง[1*] จรัญ แสนราช[2] และ อนิราช มิ่งขวัญ[3]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกตัวแบบทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี โดยใช้เทคนิคเหมืองข้อมูล 3 วิธี คือวิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ และวิธีฐานกฎ นำมาเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกหาตัวแบบที่เหมาะสมเพื่อใช้ทำนายแขนงวิชาเรียนที่เหมาะสมกับนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ โดยรวบรวมข้อมูลของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี ตั้งแต่ปีการศึกษา 2548-2559 จำนวน 759 ระเบียน วิเคราะห์ข้อมูลบนพื้นฐานของวิธี 10-Fold Cross Validation โดยใช้โปรแกรม RapidMiner Studio 7 ในการสร้างแบบจำลอง ผลการศึกษาพบว่าวิธีฐานกฎมีประสิทธิภาพสูงสุด ให้ค่าความถูกต้องเท่ากับ 78.66% รองลงมาคือวิธีต้นไม้ตัดสินใจที่ให้ค่าความถูกต้องเท่ากับ 78.26% และวิธีแบบเบย์ให้ค่าความถูกต้องเท่ากับ 78.12% จากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้ สามารถนำวิธีฐานกฎไปใช้ในการทำนายแขนงวิชาเพื่อให้นักศึกษาเลือกเรียนให้เหมาะสม ซึ่งเป็นแนวทางในการเสนอแนะแขนงวิชาเรียนแก่นักศึกษาให้เรียนได้อย่างมีประสิทธิภาพ และสามารถจบการศึกษาได้ในระยะเวลาตามหลักสูตร

คำสำคัญ: เหมืองข้อมูล; วิธีต้นไม้ตัดสินใจ; วิธีแบบเบย์; วิธีฐานกฎ; การจำแนกข้อมูล; ประสิทธิภาพของโมเดล

รับพิจารณา: 20 เมษายน 2561
แก้ไข: 21 กรกฎคม 2564
ตอบรับ: 6 สิงหาคม 2564

[1] อาจารย์ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

[2] ผู้ช่วยศาสตราจารย์ ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

[3] ผู้ช่วยศาสตราจารย์ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

[*] ผู้นิพนธ์ประสานงาน โทร. +668 9920 5802 อีเมล: beesuda.p@fitm.kmutnb.ac.th

The Comparison of Data Classification Efficiency for Decision in the Selection of Information Technology Students' Academic Subjects, Faculty of Technology and Industrial Management, King Mongkut's University of Technology North Bangkok, Prachinburi Campus

Beesuda Daoruang[1*] Charun Sanrach[2] and Anirach Mingkhwan[3]

Abstract

This research proposes to compare the efficiency data mining techniques for decision in the selection of Information Technology students' academic subjects, Faculty of Technology and Industrial Management, King Mongkut's University of Technology North Bangkok, Prachinburi Campus. There are three data mining classification techniques used, Decision Tree, Naïve Bayes and Rule Base, to compare the efficiency of the model and to identify the appropriate model for predicting the appropriate learning outcomes. The data used in the research were 759 records of the students from Department of Information Technology, Faculty of Technology and Industrial Management, King Mongkut's University of Technology North Bangkok, Prachinburi Campus during the 2005-2017 academic year. The data was analyzed by using RapidMiner Studio 7 with datamining technique on 10-Fold Cross Validation. The comparative results showed that the accuracy of Rule Base algorithm is 78.26% which was higher than Decision Tree and Naïve Bayes with 78.26% and 78.12% respectively. The finding also indicated that the findings could help student's decision in the selection of Information Technology academic subjects effectively and could predict their graduation duration according to the program criteria.

Keywords: Data Mining; Decision Tree; Naïve Bayes; Rule-Based; Classification; Efficiency

Received: April 20, 2018
Revised: July 21, 2021
Accepted: August 6, 2021

---

[1] Lecturer, Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok

[2] Associate Professor, Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok

[3] Associate Professor, Department of Information Technology, Faculty of Technology and Industrial Management, King Mongkut's University of Technology North Bangkok

[*] Corresponding Author Tel. +668 9920 5802 e-mail: beesuda.p@fitm.kmutnb.ac.th

1. บทนำ

ปัญหานักศึกษาออกกลางคันในระหว่างเรียน สาเหตุอย่างหนึ่งเกิดจากการเลือกเรียนในแขนงวิชาที่เมื่อเรียนไประยะหนึ่งเกิดความไม่ชอบและคิดว่าไม่ตรงกับความถนัดของตนเองหรือเรียนแล้วไม่ประสบผลสำเร็จตามที่ตนเองคาดหวัง โดยภาควิชาเทคโนโลยีสารสนเทศได้จัดให้นักศึกษาหลักสูตรเทคโนโลยีบัณฑิต 4 ปี จะต้องทำการเลือกแขนงวิชาเรียนเมื่อขึ้นชั้นปีที่ 2 โดยมีทั้งหมด 4 แขนงได้แก่ แขนงวิชาวิทยาการสารสนเทศ, แขนงวิชาเทคโนโลยีมัลติมีเดีย, แขนงวิชาเทคโนโลยีเครือข่ายและแขนงวิชาเทคโนโลยีระบบธุรกิจ ซึ่งจากระยะที่ผ่านมาพบว่าจำนวนนักศึกษาออกกลางคันยังคงมีจำนวนที่มากอยู่ตลอด สาเหตุส่วนหนึ่งเกิดจากการเลือกแขนงวิชาเรียนที่ทำให้ไม่สามารถเรียนจบหลักสูตรในระยะเวลาที่กำหนด ศักรินทร์ และคณะ [1] ได้กล่าวถึงสาเหตุของการลาออกกลางคันของนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นสูง โดยใช้เทคนิคเหมืองข้อมูล ว่ามีปัจจัยที่เกี่ยวข้อง 5 อันดับแรกคือ ประเภทโรงเรียนเดิม, ประเภทหลักสูตร, เกรดเฉลี่ยก่อนเข้าศึกษา, การกู้ยืมกองทุนเพื่อการศึกษา และสาขา เช่นเดียวกับกาญจนา [2] ได้กล่าวว่าตัวแปรที่มีอิทธิพลต่อแรงจูงใจในการเลือกเรียนวิชาชีพครูของนักศึกษาภาคปกติ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเชียงรายซึ่งพบว่าตัวแปรปัจจัยส่วนบุคคลมีอิทธิพลต่อตัวแปรแรงจูงใจในการเลือกเรียนวิชาชีพครู ซึ่งตัวแปรส่วนบุคคลสามารถวัดได้จากผลสัมฤทธิ์ทางการเรียน ทัศนคติ ลักษณะมุ่งอนาคตและบุคลิกภาพ

การนำเทคโนโลยีเหมืองข้อมูล (Data Mining) มาช่วยในทำนายเพื่อการวางแผนให้ตรงกับความต้องการใช้งาน เป็นการนำความรู้จากข้อมูลที่มีอยู่เพื่อประโยชน์ในการทำนายข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต ซึ่งการทำเหมืองข้อมูลคือการกระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำมาไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม [3]

ปรียารัตน์ และกิดาการ [4] ได้นำเทคนิคการทำเหมืองข้อมูลมาช่วยในการทำนายผลสัมฤทธิ์ในการเรียนวิชาสถิติเบื้องต้นของนิสิตปริญญาตรี มหาวิทยาลัยบูรพา ด้วยการวิเคราะห์การจำแนกและข่ายงานระบบประสาทเช่นเดียวกับงานวิจัยของขนิษฐา และคณะ [5] ได้ทำการสังเคราะห์รูปแบบบทเรียนคอมพิวเตอร์ช่วยสอนบนเครือข่าย โดยการวิเคราะห์ด้วยวิธีต้นไม้ ซึ่งสามารถจำแนกผู้เรียนตามความถนัดทางการเรียนตามทฤษฎี 4MAT

ในการวิจัยครั้งนี้ ได้นำข้อมูลของนักศึกษาตั้งแต่ปีการศึกษา 2548-2559 จำนวน 759 คน มาวิเคราะห์เพื่อพัฒนาโมเดลที่สามารถทำนายแขนงวิชาที่เหมาะสมกับผู้เรียน เพื่อเป็นแนวทางประกอบการตัดสินใจการเลือกแขนงวิชาเรียนของนักศึกษา ซึ่งมีตัวแปรต้นได้แก่ ข้อมูลหลักสูตรที่จบการศึกษาก่อนหน้า (สายสามัญหรือสายอาชีพ), เกรดเฉลี่ยก่อนเข้าศึกษา, เกรดเฉลี่ยเมื่อจบการศึกษาในระดับปริญญาตรี และผลการเรียนรายวิชาหลักของสาขาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

2. วัตถุประสงค์

เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกตัวแบบทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

3. ทฤษฎีที่เกี่ยวข้อง

3.1 เหมืองข้อมูล

เหมืองข้อมูล (Data Mining) คือการค้นหาหรือการสกัดความรู้จากฐานข้อมูลขนาดใหญ่ กระบวนการที่

กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบแนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักการทางคณิตศาสตร์ สถิติ การรู้จำ การเรียนรู้ทางเครื่องจักร เพื่อนำความรู้ที่ได้นั้นมาใช้ในการแก้ปัญหา วางแผน หรือดำเนินกลยุทธ์ขององค์กรให้ประสบความสำเร็จสูงสุด ซึ่งการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลสามารถแบ่งได้เป็น 2 ประเภทหลัก ๆ คือ [6]

3.1.1 เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) จะเน้นที่การพิจารณาข้อมูลเป็นหลัก เช่น พิจารณาว่าข้อมูลมีความสัมพันธ์กันในลักษณะใดบ้าง เทคนิคในประเภทนี้จะแบ่งย่อยได้อีกคือ เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule) และการแบ่งกลุ่มข้อมูล (Clustering)

3.1.2 เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตเพื่อนำมาสร้างโมเดลสำหรับทำนายหรือคาดการณ์สิ่งที่เกิดขึ้นในอนาคต สามารถแบ่งย่อยได้อีกคือ การจำแนกประเภทข้อมูล (Classification) และการประมาณค่าข้อมูล (Regression) ซึ่งทั้งสองเทคนิคจะมีความแตกต่างกันที่คำตอบที่ต้องการทำนาย ซึ่งการจำแนกประเภทข้อมูล (Classification) จะทำนายข้อมูลที่มีค่าเป็น Nominal เช่น เพศชาย หญิง หรือค่าที่ไม่ใช่ตัวเลขนั่นเอง ส่วนการประมาณค่าข้อมูล (Regression) จะใช้กับข้อมูลคำตอบที่เป็นตัวเลขเท่านั้น

3.2 เทคนิคเหมืองข้อมูล

ในการวิจัยครั้งนี้ ได้ใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) แบบการจำแนกประเภทข้อมูล (Classification) โดยใช้อัลกอริธึมเพื่อใช้ทำนายแนวโน้มให้แก่ข้อมูลใหม่ 3 วิธีดังนี้ [7]

3.2.1 วิธีต้นไม้ตัดสินใจ (Decision Tree)

เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้ สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ล่วงหน้าและพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วยรูปแบบของ Tree โครงสร้างประกอบด้วย Root Node, Child และ Leaf Node

3.2.2 วิธีแบบเบย์ (Naïve Bayes)

คือการทำเหมืองข้อมูลที่ถูกสร้างขึ้นโดยหลักความน่าจะเป็น ซึ่งจะใช้วิเคราะห์หาความน่าจะเป็นของสิ่งที่ยังไม่เคยเกิดขึ้นด้วยการคาดเดาจากสิ่งที่เคยเกิดขึ้นมาก่อน โดนใช้ทฤษฎีของ Bayes ในการแก้ปัญหา

3.2.3 วิธีฐานกฎ (Rule-Based Classification) [8]

การจำแนกข้อมูลด้วยกฎจะเป็นโมเดลที่จะแสดงผลด้วยเซตของกฎที่มีลักษณะแบบ IF-THEN โดยกฎหนึ่งๆ จะถูกแสดงอยู่ในรูปฟอร์ม IF condition THEN conclusion

3.3 ตัววัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล [9]

การนำโมเดลไปใช้งานจริงได้นั้น จำเป็นจะต้องทราบประสิทธิภาพของโมเดลก่อน โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงานต่าง ๆ อยู่ 5 ค่า คือ

- ค่าความแม่นยำ (Precision) คือค่าที่ดูสิ่งที่ทำนายออกมาแล้วทายถูกได้กี่เปอร์เซ็นต์
- ค่าความระลึก (Recall) คือจำนวนที่ทำนายถูกกี่ตัว เป็นการวัดความถูกต้องของโมเดล
- ค่าความถ่วงดุล (F-measure) คือค่าเฉลี่ยของค่าความแม่นยำและค่าความระลึก
- ค่าความถูกต้อง (Accuracy) คือจำนวนข้อมูลที่ทำนายถูกทุกคลาส เป็นการวัดความถูกต้องของโมเดลโดยพิจารณารวมทุกคลาส

3.4 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูลด้วยวิธี Cross-validation Test

เป็นวิธีที่นิยมใช้ในการทดสอบประสิทธิภาพของโมเดล เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี Cross-validation นี้ จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross-

validation คือทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือการแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้

4. งานวิจัยที่เกี่ยวข้อง

พุทธิดา และคณะ [10] ได้นำเสนอผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลโดยใช้เทคนิคการรวมกลุ่มจำแนกข้อมูล (ENSEMBLE) ทั้ง 3 เทคนิคคือ Vote Ensemble, Bootstrap Aggregating และ Random Forest และนำอัลกอริธึมที่มีประสิทธิภาพสูงสุดมาใช้วิเคราะห์ปัจจัยที่ส่งผลต่อการสอบวัดผลการทดสอบทางการศึกษาระดับชาติทางด้านอาชีวศึกษาโดยการลดการนำเข้าทีละตัวแปร ผลการเปรียบเทียบพบว่า Random Forest ให้ค่าประสิทธิภาพความถูกต้องสูงที่สุด 100%

ชัชชฎา และคณะ [11] ได้ทำการวิจัยเพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษาโดยใช้เทคนิคแบบจำลองต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และการเรียนรู้แบบเบย์ จากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบต้นไม้ตัดสินใจมีประสิทธิภาพในการจำแนกสูงสุด

ภรัณยา และคณะ [12] ได้เสนอการเปรียบเทียบประสิทธิภาพของแบบจำลองในการจำแนกข้อมูลของโรคลมร้อน โดยใช้วิธีการจำแนกข้อมูลด้วยโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอน แบบหลายชั้นและซัพพอร์ตเวกเตอร์แมชชีน การวัดประสิทธิภาพจากความถูกต้องของการจำแนกประเภทของข้อมูล ซึ่งพบว่าการใช้แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็บตรอนหลายชั้นในการจำแนกข้อมูลมีประสิทธิภาพดีที่สุด

5. วิธีการดำเนินการวิจัย

5.1 ศึกษาปัญหา

งานวิจัยชิ้นนี้ใช้ข้อมูลประวัติการเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี ตั้งแต่ปีการศึกษา 2548-2559 ซึ่งมีจำนวนข้อมูลทั้งหมด 759 ระเบียน ซึ่งมีจำนวนแอทริบิวต์ ทั้งหมด 12 แอทริบิวต์ ได้แก่ เกรดเฉลี่ยจบการศึกษา, ระดับผลการเรียนของโรงเรียนเดิม, ผลการเรียนในรายวิชาหลัก 8 วิชา, หลักสูตรที่จบการศึกษาก่อนเข้าศึกษาปริญญาตรี (สายสามัญหรือสายอาชีพ) และแขนงวิชาเรียน โดยมี 4 คลาส ได้แก่ คลาส 1 คือแขนงวิชาวิทยาการสารสนเทศ, คลาส 2 คือแขนงวิชาเทคโนโลยีมัลติมีเดีย, คลาส 3 คือแขนงวิชาเทคโนโลยีเครือข่าย และคลาส 4 คือแขนงวิชาเทคโนโลยีระบบธุรกิจ

5.2 การเตรียมข้อมูลสำหรับทำเหมืองข้อมูล

ทำการเตรียมข้อมูลทั้ง 759 ระเบียนดังตารางที่ 1

5.3 การออกแบบโมเดล

ทำการออกแบบโมเดลตามรูปแบบการจำแนกประเภทข้อมูล (Classification) และใช้โปรแกรม RapidMiner Studio 7 วิเคราะห์โดยใช้อัลกอริธึม วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ และวิธีฐานกฎ เพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูล

ตารางที่ 1 รายละเอียดข้อมูล

Variable	Detail	Key code
IT_GPA	เกรดเฉลี่ยจบการศึกษา	ตัวเลขเกรดเฉลี่ยจบการศึกษา
ComPro	ผลการเรียนวิชา Computer Programming	High = A และ B+ Normal = B และ C+ Medium = C และ D+ Low = D และ F
ComProLab	ผลการเรียนวิชา Computer Programming Lab	
Math	ผลการเรียนวิชา Math	
DataStruc	ผลการเรียนวิชา Data Structure	

ตารางที่ 1 รายละเอียดข้อมูล (ต่อ)

Variable	Detail	Key code
Eng1	ผลการเรียนวิชา English I	
Eng2	ผลการเรียนวิชา English II	
IntroIT	ผลการเรียนวิชา Introduction to IT	
OOP	ผลการเรียนวิชา Object Orientation Programming	
School_GPA	ระดับผลการเรียนจากโรงเรียนเดิม	Good=3.00 ขึ้นไป Normal=2.50-2.99 Bad=น้อยกว่า 2.49
Level	วุฒิการศึกษาเดิมที่จบ	ปวช, สายสามัญ
Field	แขนงวิชา	A=วิทยาการสารสนเทศ B=เทคโนโลยีมัลติมีเดีย C=เทคโนโลยีเครือข่าย D=เทคโนโลยีระบบธุรกิจ

ทำการทดสอบประสิทธิภาพในการจำแนกข้อมูลโดยการวัดความถูกต้องของทั้ง 3 อัลกอริธึม ด้วยวิธีการ 10-fold cross-validation ซึ่งจะแบ่งข้อมูลออกเป็น 10 ชุดเท่า ๆ กันแล้วให้ใช้ 1 กลุ่มมาเป็นกลุ่มทดสอบ (Test set) ส่วนที่เหลือ 9 ชุด นำมาใช้เป็นกลุ่มเรียนรู้ (Training Set) แล้วทำการวนทำเป็นจำนวน 10 รอบ โดยเปลี่ยนกลุ่มทดสอบไปเรื่อย ๆ จนครบ ในการวัดประสิทธิภาพของการจำแนกประเภทข้อมูลได้แก่ การหาค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F-measure) และค่าความถูกต้อง (Accuracy)

6. ผลการวิจัย

6.1 ผลการเปรียบเทียบประสิทธิภาพ

จากการทดสอบประสิทธิภาพด้วยวิธีการ 10-fold cross-validation เพื่อทำการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล 3 อัลกอริธึม คือ แบบวิธีต้นไม้ตัดสินใจ วิธีแบบเบย์และวิธีฐานกฎค่าความถูกต้องของแต่ละอัลกอริธึม พบว่าการจำแนกประเภทแบบวิธีฐานกฎ ให้ค่าความถูกต้องสูงที่สุด (78.66) รองลงมาคือวิธีต้นไม้ตัดสินใจ (78.26) และวิธีแบบเบย์ (78.12) ตามลำดับ ผลการทดลองแสดงดังตารางที่ 2

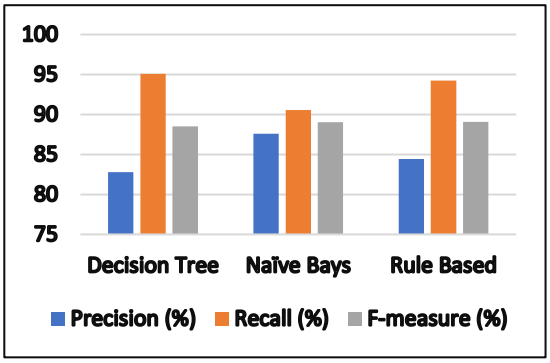
ตารางที่ 2 ผลลัพธ์ความถูกต้องในการจำแนกข้อมูล

Algorithms	Accuracy Rate (%)
Decision Tree	78.26
Naïve Bays	78.12
Rule Based	78.66

การประเมินประสิทธิภาพโดยใช้การวัดค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) ผลลัพธ์แสดงในตารางที่ 3 และรูปที่ 1

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของอัลกอริธึมสำหรับจำแนกข้อมูล

Algorithms	Precision (%)	Recall (%)	F-measure (%)
Decision Tree	82.79	95.10	88.51
Naïve Bays	87.58	90.54	89.03
Rule Based	84.42	94.26	89.06



รูปที่ 1 กราฟเปรียบเทียบค่าความแม่นยำ, ค่าความระลึก และค่าความถ่วงดุลของการจำแนกข้อมูล

6.2 ตัวแบบที่ได้จากการเปรียบเทียบประสิทธิภาพ

จากผลการเปรียบเทียบประสิทธิภาพพบว่าอัลกอริธึมวิธีฐานกฎให้ค่าความถูกต้องสูงที่สุด ซึ่งได้ผลการวิเคราะห์จากโปรแกรม RapidMiner Studio 7 แสดงผลการวิเคราะห์ข้อมูลจนได้ตัวแบบตามวิธีฐานกฎตามรูปที่ 2

RuleModel

```
if OOP = Low and Eng1 = Low then 2.5  (248 / 11 / 4 / 6 / 0)
if ComPro = Low and Math = Low then 2.5  (143 / 11 / 1 / 7 / 0)
if ComProLab = Low and School_GPA = Normal then 2.5  (52 / 0 / 4 / 3 / 1)
if OOP = Low and Eng2 = Low then 2.5  (13 / 0 / 0 / 0 / 0)
if OOP = Low and DataStruc = Low then 2.5  (23 / 0 / 4 / 3 / 0)
if IntroIT = Low then 2.5  (23 / 2 / 1 / 0 / 0)
if Eng2 = Low and OOP = Medium then 2.5  (21 / 0 / 2 / 5 / 0)
if Math = Low and Eng2 = Medium then 2.5  (15 / 0 / 1 / 2 / 0)
if ComPro = High and Eng1 = Medium then 3.5  (5 / 0 / 14 / 1 / 1)
if ComProLab = Low and DataStruc = Medium then 2.5  (10 / 0 / 0 / 3 / 0)
if Fields = A and Eng2 = High then 3.5  (2 / 0 / 17 / 4 / 5)
if Math = Low and ComProLab = High then 2.5  (6 / 0 / 1 / 0 / 0)
if ComPro = Low and Eng1 = Medium then 2.5  (6 / 0 / 3 / 1 / 0)
if Math = High and Eng2 = Medium then 3.5  (3 / 0 / 9 / 4 / 1)
if ComProLab = Low and Eng2 = Low then 2.5  (7 / 1 / 1 / 1 / 0)
if Eng2 = Low and IntroIT = High then 3.5  (5 / 0 / 7 / 2 / 0)
if Math = Low and Fields = B then 2.5  (3 / 0 / 0 / 0 / 0)
if DataStruc = Low and ComProLab = High then 3.0  (0 / 0 / 0 / 2 / 0)
if Math = Low and Eng1 = High then 3.5  (1 / 0 / 3 / 0 / 0)
if ComProLab = Medium and School_GPA = Good then 3.0  (0 / 0 / 0 / 3 / 0)
if OOP = High and ComPro = High then 4.0  (1 / 0 / 1 / 0 / 5)
if ComPro = Low and Math = Medium then 2.5  (4 / 0 / 1 / 0 / 0)
if OOP = Low then 3.0  (0 / 0 / 0 / 2 / 0)
if Math = High and Eng1 = High then 3.5  (0 / 0 / 3 / 0 / 0)
if Math = Medium and Fields = B then 3.0  (0 / 0 / 0 / 2 / 0)
if Math = High then 2.5  (1 / 0 / 0 / 0 / 0)
else 4.0  (0 / 0 / 0 / 0 / 0)

correct: 642 out of 758 training examples.
```

รูปที่ 2 ตัวบบที่ได้จากวิธีฐานกฎด้วยการวิเคราะห์จากโปรแกรม RapidMiner Studio 7

การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมนิ่งโดยโปรแกรม RapidMiner Studio 7 ได้วิธีฐานกฎซึ่งเป็นอัลกอริธึมที่ให้ค่าความถูกต้องสูงสุด พบว่าตัวแบบที่ได้นั้นสามารถนำไปทำนายแขนงวิชาเรียนให้แก่นักศึกษา ซึ่งจะต้องเลือกเรียนในชั้นปีที่ 2 ซึ่งมีอยู่ 4 แขนง ได้แก่ แขนงวิชาวิทยาการสารสนเทศ, แขนงวิชาเทคโนโลยีมัลติมีเดีย, แขนงวิชาเทคโนโลยีเครือข่าย และแขนงวิชาเทคโนโลยีระบบธุรกิจของภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี

ผลจากการทำนายทำให้ทราบว่านักศึกษาที่มีเกรดเฉลี่ยก่อนเข้าศึกษาต่อในระดับใด และมีผลการเรียนในรายวิชาบังคับเป็นเท่าไหร่ ควรจะเลือกเรียนในแขนงวิชาใดและจะมีเกรดเฉลี่ยตอนจบการศึกษาเท่าไหร่ ซึ่งสามารถนำไปใช้เป็นแนวทางในการช่วยตัดสินใจเลือกแขนงวิชาเรียนของนักศึกษาชั้นปีที่ 2 ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือต่อไปได้

7. สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้ทำการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี เพื่อค้นคว้าหาตัวแบบที่ใช้ในการจำแนกประเภทข้อมูลที่ดีที่สุด โดยตัวแปรที่ใช้มีทั้งหมด 12 ตัวแปร ได้แก่ เกรดเฉลี่ยจบการศึกษา, ระดับผลการเรียนของโรงเรียนเดิม, ผลการเรียนในรายวิชาหลัก 8 วิชา, หลักสูตรที่จบการศึกษาก่อนเข้าศึกษาปริญญาตรี (สายสามัญหรือสายอาชีพ) และแขนงวิชาเรียน โดยได้ทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูล 3 วิธีคือ แบบวิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ และวิธีฐานกฎ ผลการเปรียบเทียบอัลกอริธึมทั้ง 3 วิธี พบว่าการจำแนกประเภทแบบวิธีฐานกฎ (Rule Base) ให้ค่าความถูกต้องสูงที่สุด คือ 78.66% รองลงมาคือวิธีต้นไม้ตัดสินใจ (Decision Tree) ให้ความถูกต้อง 78.26% และวิธีแบบเบย์ 78.12% ตามลำดับ ซึ่งสอดคล้องกับงานวิจัยของนงลักษณ์และคณะ [13] ได้เปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคการทำเหมือง ระหว่างเคเนียเรสเนเบอร์ กฎพื้นฐาน และต้นไม้ตัดสินใจ พบว่าอัลกอริธึม Rule Base เหมาะสมกับข้อมูลที่มีปริมาณแอทริบิวต์สูงและข้อมูลที่มีจำนวนกลุ่มต่ำ

ผู้วิจัยจะนำตัวแบบซึ่งได้จากการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลจากงานวิจัยนี้ไปใช้ทำนายเพื่อวิเคราะห์แนวทางการเสนอแนะแขนงวิชาเรียนแก่นักศึกษาชั้นปีที่ 2 ของภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม เพื่อให้นักศึกษาสามารถใช้ความสามารถได้เต็มศักยภาพ ลดภาวการณ์ออกกลางคันในระหว่างเรียนของนักศึกษาของภาควิชา และสามารถจบการศึกษาได้ในระยะเวลาตามหลักสูตร

8. เอกสารอ้างอิง

[1] S. Phupanna, P. Wanchai, V. Songpan and N. Benjamas, " Predictive Analysis for student Dropout in High Vocational Certificate Using Data Mining Technique," in *The 13th National Conference on Computing and Information Technology: NCCIT2017*, Thailand, 2017. (in Thai)

[2] K. Yajaiman, *The Development of a Model of the Causal Factors that Influence the Motivation to Choose the Teaching Profession for Students of the Faculty of Education, Chiang Rai Rajabhat University,* Chaing Rai Rajabhat: Master of Education, 2013. (in Thai)

[3] A. Ngamyim, "Data Mining," 2016. [Online]. Available: http: / / compcenter. bu. ac. th/ news-information/ data-mining. [Accessed 19 December 2017]. (in Thai)

[4] P. Naksuwan and K. Saithanu, "Prediction of Learning Achievement on Elementary Statistics Course of Burapha University Undergraduate Students by Discriminant Analysis and Neural Networks," *Burapha Science Journal,* vol. 17, no. 1, pp. 59-68, 2012. (in Thai)

[5] K. Deesubin, D. Phonak and M. Tiantong, "A Synthesis forms the CAI Network Students Classified by Aptitude Theoretical 4MAT by Decision Tree Methodology," *Technical Education Journal King Mongkut' s University of Technology North Bangkok,* vol. 7, no. 1, pp. 130-137, 2016. (in Thai)

[6] O. Pantho and M. Tiantong, "A Comparative Efficiency of Classification of VARK Learning Style Using Data Mining Techniques," *Journal of Industrial Technology Ubon Ratchathani Rajabhat University,* vol. 4, no. 1, pp. 1-11, 2014. (in Thai)

[7] W. Buathong, "Data Mining Classification," 2015. [Online]. Available: https://wipawan blog. [Accessed 19 December 2017]. (in Thai)

[8] K. Amphawan, "Data mining," 2016. [Online]. Available: https://staff.informatics.buu.ac.th / ~komate/ 886464/ % 5B6% 5D- Classification. pdf. [Accessed 19 December 2017]. (in Thai)

[9] E. Pacharawongsak, An Introduction to Data Mining Techniques, Thailand: Asia Digital Printing Company, 2014. (in Thai)

[10] P. Sakulviriyakitkul, C. Sanrach and T. Sintanakul, " The comparison of Data Classification Efficiency by using ENSEMBLE Technique for Analysis of Factors Affecting on the National Educational Test Results of Vocational Education (V-NET)," in *The 13th National Conference on Computing and Information Technology: NCCIT2017*, Thailand, 2017. (in Thai)

[11] C. Vandee, J. Phuboon- ob and C. Jareanpon, " Comparative Efficiency of Classification Choosing a Career with Data Mining Technique," in *The 9th National Conference on Computing and Information Technology: NCCIT2013*, Thailand, 2013. (in Thai)

[12] P. Ammaruekarat and P. Meesad, " A Comparative Efficiency of Neural Network Classification for the Diagnosis of Heatstork," in *The 6th National Conference on Computing and Information Technology: NCCIT2010*, Thailand, 2010. (in Thai)

[13] N. Promthong, S. Penchamrush and S. Tangprasert, " Comparative Efficiency of Classification with Data Mining Technique between K-nearest neighbor Rule Base and Decision Tree," in *The 8th National Conference on Computing and Information Technology: NCCIT2012*, Thailand, 2012. (in Thai)