การพัฒนาตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ

The Development of Models to Forecast Numbers of New Students Using the Classification Rules with Decision Tree Technique

ลาภ พุ่มหิรัญ[1*] และ มาลีรัตน์ โสดานิล[2]

[1]นักศึกษา [2]อาจารย์ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพฯ 10800

บทคัดย่อ

การประชาสัมพันธ์และรับสมัครนักศึกษาใหม่ในระดับปริญญาตรีที่ตรงกลุ่มเป้าหมายจะช่วยลดต้นทุนด้านการประชาสัมพันธ์หรือรับสมัครนอกสถาบันได้เป็นอย่างดี เทคนิคต้นไม้ตัดสินใจสามารถนำมาใช้ในการพัฒนาตัวแบบหรือเป็นเครื่องมือช่วยในการพยากรณ์จำนวนนักศึกษาใหม่ ที่อยู่ในรูปของกฎการจำแนกประเภทที่ให้ความถูกต้องแม่นยำสำหรับช่วยสถาบันการศึกษานำไปตัดสินใจเลือกรูปแบบและสถานที่ในการประชาสัมพันธ์หรือรับสมัครนักศึกษาใหม่ได้ตรงกลุ่มเป้าหมายมากยิ่งขึ้น ตัวแบบที่พัฒนาขึ้นโดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจถูกสร้างและทดสอบตัวแบบด้วยวิธีการสุ่มตัวอย่างที่แตกต่างกัน 7 ตัวแบบ ได้แก่ วิธีการตรวจสอบไขว้ 3 ตัวแบบ วิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 3 ตัวแบบ และวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน 1 ตัวแบบ ผลการวิจัยพบว่า ตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ที่ถูกพัฒนาด้วยวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน มีค่าประสิทธิภาพสูงกว่าตัวแบบที่พัฒนาด้วยวิธีอื่น โดยมีค่าความถูกต้องเท่ากับ 94% ค่าความแม่นยำเท่ากับ 94.3% ค่าความระลึกเท่ากับ 94% และ ค่าความถ่วงดุลเท่ากับ 93.7% แสดงว่าวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน สามารถนำไปใช้พัฒนาตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องและแม่นยำในการทำนายจำนวนนักศึกษาใหม่ได้เป็นอย่างดี

Abstract

Advertisement and admission of target applicants, who will be new undergraduate students, will help to reduce costs, promote or recruit off- campus as well. The decision tree Technique can be used to develop models to forecast numbers of new students in the classification rules for accuracy. It helps to decide on the form and place in admission of new students to meet the target even more. The seven models developed by using the rules of classification techniques, decision trees were built and tested with different sampling methods: 3 models of k-fold cross-validation, 3 models of percentage split, and a model of training set and test set. The study result for forecasting new students via model developed by training set and test set method is higher performance than model developed by other methods. It was shown that the efficiency was 94%, the precision was 94.3%, the recall was 94% and the F-measure was 93.7%. Thus, this model is accurate in forecasting numbers of new students using the classification rules with decision tree technique.

คำสำคัญ : ต้นไม้ตัดสินใจ การพยากรณ์จำนวนนักศึกษาใหม่ กฎการจำแนก

Keywords : decision tree, forecasting numbers of new students, classification rules

ผู้นิพนธ์ประสานงานไปรษณีย์อิเล็กทรอนิกส์ lap_p@windowslive.com โทร. 089-172-2190

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การได้มาซึ่งนักศึกษาใหม่ในระดับปริญญาตรี ของแต่ละสถาบันอุดมศึกษาเกิดจากกระบวนการประชาสัมพันธ์ การรับสมัครและการสอบคัดเลือกเป็นหลัก โดยเฉพาะการประชาสัมพันธ์และการรับสมัครเป็นกระบวนการแรกเริ่มที่ได้มาซึ่งผู้สมัครเข้าศึกษาต่อ สถาบันการศึกษาจำเป็นต้องมีรูปแบบวิธีการประชาสัมพันธ์ต่างๆ ผ่านสื่อวิทยุโทรทัศน์ สื่อสิ่งพิมพ์ รวมถึงการไปรับสมัครผู้สมัครนอกสถาบัน เช่น โอเพ่นเฮ้าส์ (Open house) ด้วยการเดินทางไปรับสมัครตามโรงเรียนหรือจังหวัดต่างๆ ทั่วประเทศ เพื่อให้ได้ผู้สมัครเข้าศึกษาต่อปริมาณมากขึ้น และเพิ่มโอกาสในการคัดเลือกผู้สมัครที่มีความสามารถเป็นนักศึกษาใหม่ได้ดียิ่งขึ้น แต่กระบวนการประชาสัมพันธ์และรับสมัครนอกสถาบันดังกล่าวจำเป็นต้องใช้งบประมาณสูงและอาจได้รับผู้สมัครที่ไม่ตรงกลุ่มเป้าหมาย โดยทั่วไปแต่ละสถาบันการศึกษาจะมีระบบฐานข้อมูลสำหรับจัดเก็บข้อมูลผู้สมัครเข้าศึกษาต่อแต่ละปีการศึกษาไว้ แต่มิได้นำข้อมูลเหล่านั้นมาใช้ประโยชน์ ซึ่งแท้จริงแล้วข้อมูลดังกล่าวหากถูกนำมาประมวลผลด้วยเทคนิคเหมืองข้อมูล (Data mining) จะช่วยให้ได้รับสารสนเทศที่เป็นประโยชน์ต่อสถาบันการศึกษาสำหรับนำไปประยุกต์ใช้ในกระบวนการประชาสัมพันธ์หรือรับสมัครนอกสถาบันได้

1.2 วัตถุประสงค์การวิจัย

งานวิจัยนี้ มีวัตถุประสงค์เพื่อพยากรณ์จำนวนนักศึกษาใหม่ที่จะเข้าศึกษาต่อในระดับปริญญาตรี โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ และเพื่อวัดประสิทธิภาพตัวแบบที่พัฒนาขึ้น ตัวแบบที่มีประสิทธิภาพและความถูกต้องแม่นยำมากที่สุดสามารถนำไปประยุกต์ใช้สำหรับให้คำแนะนำสถาบันการศึกษาในการประชาสัมพันธ์และการรับสมัครนักศึกษาใหม่ได้อย่างเหมาะสมกับกลุ่มเป้าหมายที่จะเป็นนักศึกษาในอนาคต

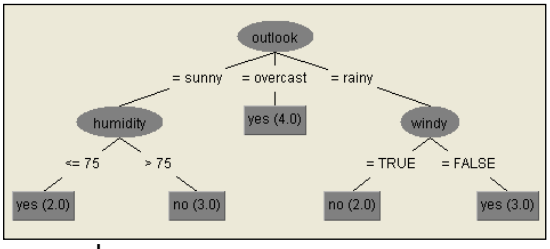
1.3 ทฤษฎีที่ใช้ในการวิจัย

การพัฒนาตัวแบบพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ประกอบด้วยทฤษฎีต่าง ๆ ดังนี้

1. การทำเหมืองข้อมูล เป็นขบวนการในการกรองข้อมูลและสืบค้นความรู้ที่เป็นประโยชน์จากฐานข้อมูลขนาดใหญ่ (Large information) [7] โดยการทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญในการค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge discovery in database) หรือที่เรียกว่า KDD จะนำข้อมูลที่มีอยู่มาวิเคราะห์และสืบค้นความรู้สารสนเทศหรือสิ่งที่สำคัญที่ยังไม่รู้ออกมา เพื่อให้ได้สารสนเทศที่สามารถนำไปใช้งานได้ (Actionable) ประกอบในการช่วยตัดสินใจใน เช่น การวิเคราะห์พฤติกรรมผู้บริโภคในการซื้อสินค้า เพื่อใช้ในการส่งเสริมการขายสินค้าในธุรกิจการพยากรณ์หุ้น การพยากรณ์อากาศ เป็นต้น [5]

2. กฎการจำแนกประเภท (Classification rules) เป็นกระบวนการจัดแบ่งข้อมูลตามลักษณะของวัตถุประสงค์นั้นๆ ด้วยการวิเคราะห์เซตของกลุ่มข้อมูล (Data object) ที่ยังไม่จัดแบ่งประเภท เพื่อสร้างโมเดลจัดการข้อมูลให้อยู่ในรูปชุดข้อมูล (Class) ที่กำหนด โดยจะนำข้อมูลส่วนหนึ่งจากข้อมูลทั้งหมดเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ (Training data) เพื่อจำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้ ผลลัพธ์ที่ได้จากการเรียนรู้คือ โมเดลจัดประเภทข้อมูล (Classifier model) แล้วจึงนำข้อมูลส่วนที่เหลือมาใช้สำหรับทดสอบ (Testing data) และนำผลที่ได้มาใช้เปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลเพื่อทดสอบความถูกต้องโดยกฎการจำแนกที่ได้มีรูปแบบของ IF <Conditions> THEN <Class> หรือถ้า <เงื่อนไข> แล้ว <คลาส> นั่งเอง

3. ต้นไม้ตัดสินใจ เป็นแบบจำลองวิธีหนึ่งที่นิยมใช้ในการพยากรณ์ (Prediction) หรือการจำแนกข้อมูล (Classification) โดยโครงสร้างจะมีลักษณะ



รูปที่ 1 ต้นไม้ตัดสินใจการเล่นกีฬากอล์ฟ

เหมือนโครงสร้างต้นไม้ ที่แต่ละโหนด (Node) แสดงคุณลักษณะประจำ (Attribute) ที่ใช้ทดสอบข้อมูล แต่ละกิ่งแสดงผลลัพธ์ในการทดสอบตามเงื่อนไข และลีฟโหนด (Leaf node) แสดงกลุ่มหรือคลาส (Class) ที่กำหนดไว้ตัวอย่างดังรูปที่ 1

1.4 งานวิจัยที่เกี่ยวข้อง

มีหลากหลายงานวิจัยที่นำเทคนิคเหมืองข้อมูลมาใช้สำหรับวิเคราะห์และสร้างกฎการจำแนกข้อมูลแล้วให้ผลลัพธ์น่าสนใจ ชลนิศา สาระ [2] นำแบบจำลองต้นไม้ตัดสินใจ (Decision tree) มาใช้สำหรับพยากรณ์โอกาสการสำเร็จการศึกษา และทราบถึงความสัมพันธ์ของปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษา และสามารถคาดเดาระยะเวลาในการสำเร็จการศึกษาได้อย่างแม่นยำ

บุญมา เพ่งช่วน [3] ใช้เทคนิคต้นไม้ตัดสินใจ พัฒนาตัวแบบสำหรับทำนายแนวโน้มการเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี เพื่อเป็นแนวทางในการผลิตบัณฑิตตามสายอาชีพที่เหมาะสมต่อไปได้

กฤษณะ ไวยมัย และคณะ [1] ได้นำเสนอเทคนิคเหมืองข้อมูล ได้แก่ การค้นหากฎการจำแนกข้อมูล (Data classification) การจำแนกเชิงความสัมพันธ์ (Association rule discovery) การพยากรณ์ข้อมูล (Data prediction) มาประยุกต์ใช้สำหรับช่วยแนะนำนิสิตเลือกสาขาที่เหมาะสมที่สุด พร้อมทำนายผลการเรียนแต่ละรายวิชาในภาคการศึกษาถัดไป โดยนำเสนอแบบจำลองการจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ จากงานวิจัยสามารถทำนายสาขาวิชาที่เหมาะสมที่สุดให้กับนิสิตได้และมีความถูกต้องค่อนข้างสูง แต่แบบจำลองที่ได้มีความโน้มเอียงไปหาสาขาวิชาที่มีจำนวนนิสิตมากส่งผลให้ความถูกต้องแม่นยำในการพยากรณ์ลดลง

งานวิจัยข้างต้น อาศัยเทคนิคเหมืองข้อมูลมาช่วยในการสนับสนุนการตัดสินใจเพื่อค้นหาทางเลือกที่เหมาะสม รวมถึงการประยุกต์ใช้ในการพยากรณ์ด้านต่าง ๆ โดยเฉพาะการใช้เทคนิคต้นไม้ตัดสินใจอันเป็นเทคนิคเดียวกับที่ใช้ในงานวิจัยนี้

2. วิธีการศึกษา

วิธีการวิจัยแบ่งออกเป็น 3 ขั้นตอนคือ การเตรียมข้อมูล การสร้างและทดสอบตัวแบบการพยากรณ์ และการวัดค่าประสิทธิภาพของตัวแบบการพยากรณ์

2.1 การเตรียมข้อมูล

ข้อมูลที่นำมาใช้ในการวิจัยได้จากข้อมูลผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี มหาวิทยาลัยราชภัฏสวนสุนันทา ปีการศึกษา 2552 และ 2553 จำนวน 25,241 ชุดข้อมูล โดยคัดเลือกปัจจัยหรือคุณลักษณะประจำ (Attribute) ที่สนใจนำมาสร้างตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ ดังนี้

1. รูปแบบการสมัคร/สถานที่รับสมัคร (Location) มีขอบเขตข้อมูล 3 ค่า คือ สมัครที่สถาบัน/มหาวิทยาลัย (University) สมัครผ่านอินเตอร์เน็ต (Internet) และเดินทางไปรับสมัครนอกสถาบัน (Open house)

2. จังหวัดผู้สมัคร (Province) ขอบเขตข้อมูลประกอบด้วยจังหวัดต่าง ๆ จำนวน 76 จังหวัด

3. โปรแกรมวิชา (Program) ขอบเขตข้อมูลประกอบด้วยรหัสโปรแกรมที่มีผู้มาสมัครในปีการศึกษา 2552 และ 2553 จำนวน 209 โปรแกรม

4. คะแนนเฉลี่ยระดับมัธยมศึกษาตอนปลายหรือเทียบเท่า (GPA)

5. ผลการสอบคัดเลือกเป็นนักศึกษาใหม่ (Isstudent) เป็นผลลัพธ์หรือคลาส (Class) จำนวน 2 คลาส คือ คลาสผู้สมัครที่สอบผ่านการคัดเลือกเป็นนักศึกษาใหม่ และคลาสของผู้สมัครที่สอบไม่ผ่านการคัดเลือกเป็นนักศึกษาใหม่

แล้วนำข้อมูลเหล่านี้มาแปลงให้อยู่ในรูปแบบของไฟล์ ARFF สำหรับเตรียมที่จะนำไปสร้างและทดสอบตัวแบบด้วยโปรแกรมเวกา (WEKA) โดยภายในแฟ้มข้อมูลจะมีส่วนประกอบต่าง ๆ ได้แก่ ส่วนหัวของไฟล์ (Relation) คุณลักษณะประจำ (Attribute) และข้อมูล (Data) ตัวอย่างดังรูปที่ 2

```
@relation student_location
@attribute location {University, Internet, Openhouse}
@attribute program {1101,1102,1103,1104,1105,1106,1201,1401,
,7322,7323,7324,7401,8301}
@attribute gpa real
@attribute province {Krabi,Bangkok,Kanchanaburi,Kalasin,Kampha
@attribute isstudent {Yes, No
@data
University,2507,2.76,Bangkok,Yes
University,1402,2.14,Ratchaburi,No
University,2507,2.32,Nonthaburi,Yes
```

รูปที่ 2 ข้อมูลที่ถูกจัดในรูปแบบ ARFF ไฟล์

2.2 การสร้างและทดสอบตัวแบบการพยากรณ์

วิธีการสร้างตัวแบบการพยากรณ์ดังกล่าว เลือกใช้โปรแกรมเวกา มาช่วยในการสร้างและทดสอบตัวแบบด้วยเทคนคดต้นไม้การตัดสินใจภายใต้อัลกอริทึม C4.5 (หรือ J48) โดยตัวแบบที่ได้จะอยู่ในรูปของกฎการจำแนกประเภทข้อมูลจากการเรียนรู้ด้วยข้อมูลชุดเรียนรู้ (Training set) หรือข้อมูลชุดสร้างตัวแบบ แล้วนำไปทดสอบด้วยข้อมูลชุดทดสอบ (Test set) ตามลำดับดังรูปที่ 3 ตัวแบบที่พัฒนาขึ้นแต่ละตัวแบบมีวิธีการแบ่งข้อมูลที่แตกต่างกันด้วยวิธีการสุ่มข้อมูล 3 วิธีการ คือ วิธีการตรวจสอบไขว้ (K-fold cross-validation) วิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (Percentage split) และวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน (Training set and test set) แต่ละวิธีมีการสร้างตัวแบบดังนี้

วิธีแรกคือ วิธีการตรวจสอบไขว้ จะนำข้อมูลทั้งหมดจำนวน N ชุดข้อมูล มาแบ่งออกเป็นส่วนย่อยๆ จำนวน k ส่วน (Fold) แต่ละส่วนมีขนาดเท่ากับ N หารด้วย k จากนั้นนำข้อมูลส่วนแรกไปทดสอบกับข้อมูลส่วนที่เหลือทีละส่วนจนครบ แล้วจึงเปลี่ยนมาใช้ข้อมูลส่วนถัดไปเป็นข้อมูลชุดทดสอบแทน ซึ่งวิธีการนี้เป็นการเปลี่ยนข้อมูลชุดทดสอบไปจนครบ k ส่วนนั่นเอง สำหรับวิธีนี้ผู้วิจัยกำหนดค่า k ไว้จำนวน 3 ค่า คือ 5 10 และ 100 ตามลำดับ ทำให้ได้ตัวแบบการพยากรณ์จำนวน 3 ตัวแบบออกมา

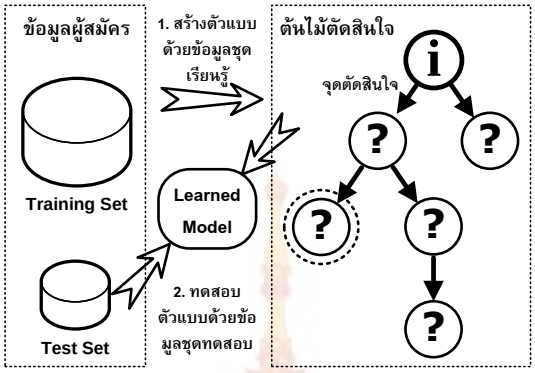
วิธีที่สองคือ วิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ เป็นการแบ่งข้อมูลชุดเรียนรู้และชุดทดสอบด้วยวิธีการสุ่ม โดยกำหนดขนาดของข้อมูลชุดทดสอบเป็นร้อยละ 10 20 และ 66 ทำให้ได้ตัวแบบการพยากรณ์ได้ 3 ตัวแบบตามลำดับ

วิธีสุดท้ายคือ วิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster sampling) และการสุ่มตัวอย่างแบบแบ่งชั้น (Stratified random sampling) ตามลำดับ โดยการแบ่งกลุ่มข้อมูลตามจังหวัดและตามคลาสผลลัพธ์ทำให้ได้ข้อมูลสองชุดที่มีการกระจายตัวทั่วถึง คือ ข้อมูลชุดเรียนรู้สำหรับสร้างตัวแบบและข้อมูลชุดทดสอบ ซึ่งผู้วิจัยกำหนดอัตราส่วนระหว่างข้อมูลทั้งสองชุดเป็นร้อยละ 80 และ 20 จากชุดข้อมูลจำนวนทั้งสิ้น 25,241 ชุดข้อมูล ทำให้ได้ข้อมูลชุดเรียนรู้และทดสอบ คือ 20,193 และ 5,048 ชุดข้อมูลตามลำดับ

การพัฒนาตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ด้วยโปรแกรมเวกาภายใต้อัลกอริทึม C4.5 มีขั้นตอนดังนี้

1. นำข้อมูลทั้งหมดที่เตรียมไว้ในรูปแบบไฟล์ ARFF สำหรับสร้างตัวแบบเข้าสู่โปรแกรมเวกา

2. สร้างและทดสอบตัวแบบตามโมเดลของกฎการจำแนกเทคนิคต้นไม้ตัดสินใจดังรูปที่ 3 ด้วยวิธีการสุ่มข้อมูลต่าง ๆ โดยกำหนดค่าความเชื่อมั่น (confidence factor) เท่ากับ 0.25 ซึ่งเป็นค่าโดยปริยายของโปรแกรมเวกา



รูปที่ 3 โมเดลการพัฒนาตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ

ตัวแบบที่ถูกพัฒนาขึ้นจะอยู่ในรูปของกฎการจำแนกอาจมีเพียงไม่กี่กฎหรือหลายร้อยกฎซึ่งปริมาณกฎที่ได้ขึ้นอยู่กับความน่าจะเป็นของคุณลักษณะประจำและการเรียนรู้ที่ได้จากข้อมูลชุดเรียนรู้ ตัวอย่างกฎที่ได้ เช่น If location = University and province = Krabi then “Yes” หมายความว่า ถ้าผู้สมัครมาสมัครด้วยตนเองที่มหาวิทยาลัยและมาจากจังหวัดกระบี่แล้วจะเป็นผู้สอบผ่านการคัดเลือกเป็นนักศึกษาใหม่

2.3 การวัดค่าประสิทธิภาพของตัวแบบการพยากรณ์

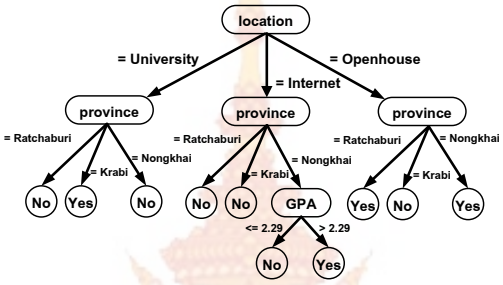
ค่าประสิทธิภาพของตัวแบบการพยากรณ์ได้จากนำตัวแบบการพยากรณ์ที่ได้จากข้อมูลชุดเรียนรู้มาทดสอบด้วยข้อมูลชุดทดสอบ ที่ให้ค่าความถูกต้อง (Correct) แสดงอยู่ในค่า Correctly classified instance ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) โดยค่าความแม่นยำคำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากจำนวนข้อมูลทั้งหมดที่ถูกจำแนกมีผลลัพธ์เดียวกัน ค่าความระลึกคำนวณจากค่าของข้อมูลที่ผลลัพธ์ถูกต้องโดยพิจารณาจากข้อมูลของผลลัพธ์เดียวกัน และค่าความถ่วงดุลคำนวณได้จากค่าเฉลี่ยระหว่างค่าความแม่นยำและความระลึก

3. ผลการศึกษาและอภิปรายผล

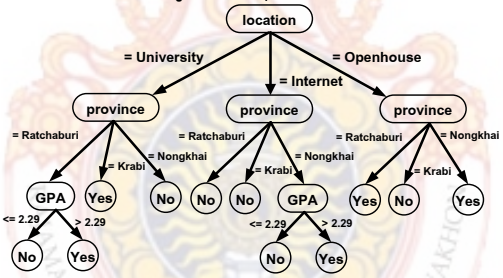
จากการพัฒนาตัวแบบในการพยากรณ์จำนวนนักศึกษาใหม่ ด้วยวิธีการตรวจสอบไขว้ วิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ และวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน ด้วยข้อมูลจำนวน 25,241 ชุดข้อมูล แสดงกฎและค่าประสิทธิภาพต่าง ๆ ได้แก่ ค่าความถูกต้องในการจำแนกประเภทของตัวแบบพยากรณ์ ความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ดังตารางที่ 1 พบว่าตัวแบบการพยากรณ์ ที่ถูกพัฒนาด้วยวิธีการตรวจสอบไขว้ และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ รวมทั้ง 6 ตัวแบบให้กฎการจำแนกประเภทการสอบผ่านและสอบไม่ผ่านการคัดเลือกเป็นนักศึกษาใหม่ จำนวน 229 กฎตัวอย่างกฎดังรูปที่ 4 โดยกฎทั้ง 6 ตัวแบบจะมีลักษณะเช่นเดียวกัน ส่วนวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกันให้กฎการจำแนกประเภทจำนวน 230 กฎ ตัวอย่างกฎดังรูปที่ 5 โดยกฎที่แตกต่างเป็นกฎที่ได้จากข้อมูล

ตารางที่ 1 ค่าประสิทธิภาพจากการทดสอบตัวแบบวิธีต่าง ๆ

วิธีสร้างตัวแบบการพยากรณ์	จำนวนกฎที่ได้ (กฎ)	ค่าความถูกต้อง (%)	ค่าความแม่นยำ (%)	ค่าความระลึก (%)	ค่าความถ่วงดุล (%)
การตรวจสอบไขว้ (5-Fold)	229	93.97	94.30	94.00	93.60
การตรวจสอบไขว้ (10-Fold)	229	93.97	94.30	94.00	93.60
การตรวจสอบไขว้ (100-Fold)	229	93.97	94.30	94.00	93.60
การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 10	229	92.08	92.70	92.10	91.40
การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 20	229	92.63	93.20	92.60	92.00
การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 66	229	93.73	94.10	93.70	93.30
การแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน (อัตราส่วน 80:20)	230	94.00	94.30	94.00	93.70



รูปที่ 4 ส่วนหนึ่งของกฎ 10 กฎที่ได้จากผู้สมัครจังหวัดราชบุรี กระบี่และหนองคายที่พัฒนาด้วยวิธีการตรวจสอบไขว้ และการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ



รูปที่ 5 ตัวอย่างส่วนหนึ่งของกฎ 11 กฎที่ได้จากผู้สมัครจังหวัดราชบุรี กระบี่และหนองคายที่พัฒนาด้วยวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน

ผู้สมัครซึ่งเดินทางมาสมัครที่สถาบันการศึกษาด้วยตนเอง และผู้สมัครมีภูมิลำเนาอยู่จังหวัดราชบุรีแล้วจะเป็นผู้สอบไม่ผ่านการคัดเลือกมีกฎที่ได้ คือ If location = University and province = Ratchaburi then “No” เมื่อเปรียบเทียบกับตัวแบบการพยากรณ์ที่ได้จากวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน จะสนใจคะแนนเฉลี่ยสะสม GPA ของผู้สมัครดังกล่าวด้วย โดยผู้ที่สอบผ่านการคัดเลือกจะต้องมี GPA มากกว่า 3.01 ขึ้นไป ทำให้กฎดังกล่าวถูกแตกออกเป็นสองกฎ ดังนี้ If location = University and province = Ratchaburi and GPA<=3.01 then “No” และ If location = University and province = Ratchaburi and GPA>3.01 then “Yes” สำหรับจำนวนกฎที่มากถึง 230 กฎ จำนวนหรือปริมาณของกฎขึ้นอยู่กับความน่าจะเป็นและการกำหนดค่าความเชื่อมั่นซึ่งผู้วิจัยกำหนดไว้ที่ 0.25 หรือ 25%

เมื่อพิจารณาในแต่ละกฎที่ถูกสร้างขึ้นมาพบว่า มีเฉพาะค่าคุณลักษณะประจำเพียง 3 ค่า คือ สถานที่รับสมัครจังหวัดผู้สมัคร และคะแนนเฉลี่ยสะสม GPA มีเพียงโปรแกรมวิชาเท่านั้นที่ไม่ได้ถูกนำมาสร้างเป็นกฎ แสดงว่าค่า

คุณลักษณะโปรแกรมวิชาไม่มีผลต่อการตัดสินใจหรือจำแนกประเภทการสอบผ่านและสอบไม่ผ่านการคัดเลือกเป็นนักศึกษาใหม่ ส่วนคะแนนเฉลี่ยสะสมจะปรากฏในกฎที่ผู้สมัครมีภูมิลำเนามาจากจังหวัดราชบุรี และหนองคายเท่านั้น ส่วนจังหวัดอื่น ๆ ไม่ได้นำคะแนนเฉลี่ยสะสมมาใช้ในการสร้างกฎ ทำให้ทราบได้ว่าค่าคะแนนเฉลี่ยสะสมมีผลต่อการตัดสินใจเฉพาะสองจังหวัดดังกล่าวเท่านั้น

จากตารางที่ 1 พบว่าค่าประสิทธิภาพต่าง ๆ ที่วัดได้จะมีค่าใกล้เคียงกันหรือมีค่าเท่ากันในบางตัวแบบ โดยตัวแบบการพยากรณ์ที่พัฒนาด้วยวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน วัดค่าความถูกต้อง ได้เท่ากับร้อยละ 94 ค่าความแม่นยำเท่ากับร้อยละ 94.3 ค่าความระลึกเท่ากับร้อยละ 94 และค่าความถ่วงดุลเท่ากับร้อยละ 93.7 และมีค่าประสิทธิภาพทุกค่าสูงกว่าตัวแบบอื่น ๆ แสดงว่าตัวแบบการพยากรณ์ที่พัฒนาด้วยวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกันมีความถูกต้องและแม่นยำในการพยากรณ์นักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจมากที่สุด

4. สรุป

การพัฒนาตัวแบบการพยากรณ์จำนวนนักศึกษาใหม่ระดับปริญญาตรี โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ จากการพัฒนาตัวแบบการพยากรณ์ 3 วิธี ได้แก่ การตรวจสอบไขว้ การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ และการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน จะได้ตัวแบบที่อยู่ในรูปของกฎการจำแนกจำนวน 229 ถึง 230 กฎ เมื่อวัดค่าประสิทธิภาพของตัวแบบ ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุลของทั้ง 3 วิธีการจะมีค่าประสิทธิภาพมากกว่าร้อยละ 90 ขึ้นไป ซึ่งการพัฒนาตัวแบบการพยากรณ์ด้วยวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกัน จะมีค่าความถูกต้องเท่ากับร้อยละ 94 ค่าความแม่นยำเท่ากับร้อยละ 94.3 ค่าความระลึกเท่ากับร้อยละ 94 และค่าความถ่วงดุลเท่ากับร้อยละ 93.7 ซึ่งมีความประสิทธิภาพทุกค่าสูงกว่าวิธีการอื่น แสดงว่าวิธีการแบ่งข้อมูลชุดเรียนรู้และทดสอบออกจากกันสามารถนำไปใช้ในการพัฒนาตัวแบบการพยากรณ์นักศึกษาใหม่โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องแม่นยำสูง และเหมาะสมกว่าวิธีอื่น หากนำกฎการจำแนกจำนวน 230 กฎที่ได้จากตัวแบบการพยากรณ์มาสร้างเป็นแอพพลิเคชั่นสำหรับพยากรณ์หรือทำนาย โดยนำข้อมูลของผู้สมัครในปีถัดไปมาทดสอบกับตัวแบบหรือกฎที่ได้ จะช่วยให้ทราบถึงจำนวนนักศึกษาใหม่ ในปีนั้นๆ ต่อไปได้และเป็นการแนะนำสถาบันการศึกษาในการกำหนดรูปแบบและวิธีการประชาสัมพันธ์หรือรับสมัครนักศึกษาใหม่ได้อย่างเหมาะสมกับกลุ่มเป้าหมายที่จะเป็นนักศึกษาในอนาคตต่อไปได้

5. กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยราชภัฏสวนสุนันทา ที่อนุเคราะห์ให้ข้อมูลเพื่อนำมาใช้ในการวิจัยครั้งนี้ และมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ให้การสนับสนุนและช่วยเหลือในการให้คำปรึกษาและการทำวิจัย

6. เอกสารอ้างอิง

กฤษณะ ไวยมัย, ชิดชนก ส่งศิริ และธนาวินท์ รักธรรมานนท์. 2544. การใช้เทคนิคดาต้าไมน์นิ่งเพื่อพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์ คณะวิศวกรรมศาสตร์.**วารสารเนคเทค,**ฉบับที่ 3. ลำดับที่ 11. หน้า 134-142.

ชลนิศา สาระ .2550. **การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ.**ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

บุญมา เพ่งช่วน. 2548. **การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบสนับสนุนการตัดสินใจด้านการผลิตบัณฑิตระดับปริญญาตรี.**ภาควิชาคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร.

ปรีชา ยามันสะบีดีน, บุญเสริม กิจศิริกุลม ปิยะวัฒน์ จิระพงษ์สุวรรณ และประสงค์ ประณีตพลกรัง**.** 2548**.การประยุกต์ใช้ดาต้าไมนิงในการบริหารลูกค้าสัมพันธ์สำหรับนักศึกษาระดับอุดมศึกษา.**บัณฑิตวิทยาลัย มหาวิทยาลัยศรีปทุม.

ไพฑูรย์ จันทร์เรือง. 2550. **ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิค**ต้นไม้ตัดสินใจ.มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

แสงจันทร์ เรืองอ่อน, ประณต บุญไชยอภิสิทธ์, ประสงค์ ปราณิตพลกรัง และปิยวัฒน์ จิรพงษ์สุวรรณ. 2545. ดาต้าไมน์นิงเชิง XML.**วารสารเนคเทค.** ฉบับที่ 48. หน้าที่ 24-29.

สุมาลี ชาญกาญจน์. 2530. **การเปรียบเทียบตัวแปรที่เกี่ยวข้องกับการสำเร็จการศึกษาและไม่สำเร็จการศึกษาของนักศึกษามหาวิทยาลัยรามคำแหง: เฉพาะกรณีที่ขอแจ้งจบ.** สำนักบริการทางวิชาการและทดสอบประเมินผล. มหาวิทยาลัยรามคำแหง.

L. Talavera and E. Gaudioso. August 22nd - 27th 2004. Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces. **The 16th European Conference on Artificial Intelligence**. Valencia. Spain .

P.L. Hsua, R. Laia, C.C. Chiub, and C.I. Hsub. 2003. The hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance**. Published in Expert Systems with Application**.. pp. 51-62.

G. John Hendricks. 2000. **An Analysis of Student Graduation Trends in Texas State Technical Colleges Utilizing Data Mining and Other Statistical Techniques**. Doctoral Thesis of Educational Administration. Baylor University. Texas. U.S.A.