การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ

The Analysis of Relationship Student Factors Affecting The Level of Academic Performance Using Decision Tree Technique : The Case Study Student of Rajamangala University of Technology Suvarnabhumi

พฤฒิพงศ์ เพ็งศิริ[1*] พันธนา ก้อนเชื้อรัตน์[1] ชัชฎา ชวรางกูร[1] และอัจฉราพรรณ คชเดช[1]

บทคัดย่อ

บทความนี้เป็นนำเสนอการวิเคราะห์กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูล ซึ่งอาศัยความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งผลลัพธ์ที่ได้นั้นจะส่งผลทำให้เอื้อประโยชน์กับแนวทางในการปรับปรุงบริหารจัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง โดยนำปัจจัยข้อมูลของนักศึกษา ได้แก่ข้อมูลทั่วไป ข้อมูลส่วนตัว ข้อมูลการศึกษา ข้อมูลระดับผลการศึกษาของนักศึกษากลุ่มที่ศึกษาในชั้นปีสุดท้ายที่สำเร็จการศึกษา มาหาความสัมพันธ์กับผลการศึกษาที่ได้ โดยวิเคราะห์จากระดับผลการเรียน (เกรดเฉลี่ย) ของการเรียนในชั้นปีที่ 1 ผลการทดลองพบว่าปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน มีทั้งหมด 7 ตัวแปร จากข้อมูลนำเข้าทั้งหมด 12 ตัวแปร ทั้งนี้ 7 ตัวแปรมาจากต้นไม้ตัดสินใจที่ได้มาสามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด โดยวัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง ดังนั้นตัวแปรที่น่าจะเป็นปัจจัยที่ส่งผลกับ ระดับผลการเรียนที่จบการศึกษาสูงสุดคือ ความสม่ำเสมอการเข้าเรียนในการเรียน

คำสำคัญ : ความสัมพันธ์ ต้นไม้ตัดสินใจ ปัจจัยข้อมูลนักศึกษา

Abstract

This paper proposes analysis of relationship by applying decision-making process with decision tree technique which was data mining. Based on association of student indicating learning level of students of Rajamangala University of Technology Suvarnabhumi which was beneficial ways and means to improved management and planning of further education courses related. These factors of student such as general information, personal Information, education level of students study in final graduation group. The correlation with the results obtained and analysis of GPA of freshly students. The results showed that the factors of the students in the class has 7 input variables from 12 variables which were derived from a decision tree can be selected only branch that has the most effect and measurement accuracy was as high as 84.78%, which is considered high. The variables are probably factors that affect it. Grade graduation is a maximum. Regular attendance in class.

Keywords: Prediction of Relationship; Decision Tree; Factor of Students

[1] คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ พระนครศรีอยุธยา หันตรา 13000

[1] Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Huntra District, Ayudhya 13000, Thailand

* Corresponding author. E-mail : prudtipong.p@rmutsb.ac.th

บทนำ

จากการศึกษาในระดับปริญญาตรีของนักศึกษาในมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จะอยู่ภายใต้คณะ และสาขาวิชาต่าง ๆ จะเห็นได้ว่า บางคนมีระดับผลเรียนดี บางคนมีระดับผลการเรียนไม่ดี บางคนก็ดูตั้งใจเรียน แต่ได้ผลการเรียนที่ไม่ดี ซึ่งสาเหตุของการเรียนที่ไม่ดีในบางคน อาจเป็นเพราะปัจจัยข้อมูลของนักศึกษาเองที่ไม่เอื้ออำนวยในการศึกษาได้ เช่น ภูมิลำเนาในจังหวัดต่าง ๆ ของนักศึกษา สถานภาพบิดา มารดา รายได้ของบิดามารดา อาชีพของบิดามารดาแตกต่างกัน เป็นต้น ทางผู้บริหารและทางมหาวิทยาลัย รวมถึงผู้ปกครอง ก็ไม่ทราบถึงสาเหตุที่แท้จริงของนักศึกษาแต่ละคน เพื่อที่จะได้แนวทางในการปรับปรุง พัฒนาส่งเสริม วางแผนด้านนโยบายการศึกษา ให้นักศึกษาเรียนได้ผลการเรียนที่ดีขึ้น ให้มีปัจจัยเอื้ออำนวยในการเรียนให้ได้ผลการเรียนดีขึ้น อีกทั้งหากต้องการหาความสัมพันธ์ว่าปัจจัยนักศึกษาแบบใดที่สามารถเรียนในสาขาวิชาที่สนใจได้ดี ต้องมีลักษณะข้อมูลนักศึกษาอย่างไร เช่น สาขาวิศวกรรมเครื่องกล เป็นสาขาที่เรียนยาก เป็นที่ต้องการตลาดแรงงานสูง เป็นต้น เพื่อจะได้ส่งเสริมให้นักศึกษาสาขานี้มีผลการเรียนดี เป็นที่ยอมรับในตลาดแรงงาน ก็ยังไม่ทราบแนวทางที่จะหาความสัมพันธ์นี้ ดังนั้น สถาบันการศึกษาระดับมหาวิทยาลัยถือเป็นองค์กรหนึ่งที่เก็บข้อมูลนักศึกษาจำนวนมาก โดยข้อมูลเหล่านี้สามารถนำมาใช้ประโยชน์ได้มาก แต่ไม่ได้ถูกนำมาใช้อย่างจริงจัง การทำเหมืองข้อมูล (Data Mining) จึงเป็นวิธีการหนึ่งที่สามารถนำมาใช้สร้างแบบจำลอง (Model) โดยใช้วิธีจำแนกประเภท (Classification) ของนักศึกษาที่มีข้อมูลการเรียนมีผลการศึกษาเฉลี่ยอยู่ในเกณฑ์ ดี หรือ ไม่ดี พร้อมทั้งหาคุณสมบัติ (Attribute) ของข้อมูลที่มีผลกับผลการเรียน เมื่อสามารถหาแบบจำลองที่มีความน่าเชื่อถือได้สูง ได้ทราบความสัมพันธ์ว่าปัจจัยข้อมูลนักศึกษาแบบใด ส่งผลให้ระดับผลการเรียนที่ดี ก็สามารถนำมาเป็นข้อมูลที่เป็นประโยชน์ ในอันที่จะส่งเสริมการเรียนรู้ของนักศึกษา จะเกิดประโยชน์กับทางผู้บริหาร และผู้ปกครองเป็นอย่างมากในการที่จะหาแนวทาง ปรับปรุง ส่งเสริม พัฒนาผลการเรียนของนักศึกษาให้ดีขึ้นได้ ส่งผลให้นักศึกษามีความรู้ ความเข้าใจในการเรียนและมีความสุขในการเรียนเพิ่มขึ้น ซึ่งปัญหาที่พบในปัจจุบัน มีข้อมูลดังต่อไปนี้

1) นักศึกษาและผู้ปกครอง ไม่ทราบถึงสาเหตุบางส่วนของการที่ผลการเรียนไม่ดี อาจมาจากปัจจัยข้อมูลของนักศึกษา ของนักศึกษาในกลุ่มตัวอย่าง จึงไม่ทราบแนวทางปรับปรุง ส่งเสริม พัฒนาผลการเรียนของนักศึกษาให้ดีขึ้น

2) ผู้บริหารต้องการทราบปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา ที่ได้ผลการเรียนดี เพื่อเป็นแนวทางปรับปรุงการศึกษา ผลักดันนโยบายการศึกษา สนับสนุนปัจจัยที่จะทำให้นักศึกษามีผลการเรียนที่ดีขึ้นได้ เช่นปัจจัยข้อมูล “ประเภทแอดมิชชั่น” (ปัจจุบันมี 3 ประเภท คือ มีรับตรง โควตา และ สกอ.) หากทราบว่าปัจจัยนี้ มีผลต่อระดับผลการเรียน อาจปรับนโยบายการรับนักศึกษา เป็นประเภทที่มีความสัมพันธ์กับผลการเรียนนักศึกษาที่ได้ผลการเรียนที่ดีให้มีสัดส่วนมากขึ้น เป็นต้น

3) ผู้บริหาร ต้องการทราบคุณลักษณะปัจจัยข้อมูลของนักศึกษาที่จะเรียนสาขาวิชาที่สนใจให้ได้ผลการเรียนดี คือ คณะวิศวกรรมศาสตร์ สาขาวิศวกรรมเครื่องกล แต่ไม่ทราบว่าจะได้ข้อมูลนั้นมาอย่างไร

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมินั้น ผู้วิจัยจึงได้นำเสนอกรอบความคิดและทำการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องดังนี้

1. การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning)

เป็นวิธีหนึ่งที่จะประมาณฟังก์ชันที่มีค่าไม่ต่อเนื่อง (discrete-value function) ด้วย แผนผังต้นไม้ อาจประกอบด้วยเซตของกฎต่างๆ (ชลนิศา สาระ และเยาวดี เต็มธนาภัทร์, 2007) แบบ ถ้า-แล้ว (if-then) เพื่อให้มนุษย์สามารถอ่านแล้วเข้าใจการตัดสินใจของต้นไม้ได้ ในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบจะแสดงประเภทของวัตถุ ต้นไม้การจัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และทฤษฎีกราฟต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

2. อัลกอริทึมที่ใช้ในแบบจำลองต้นไม้ตัดสินใจ

2.1 ID3 ในปัจจุบันนั้นมีการพัฒนาขั้นตอนวิธี (ชุติมา อุตมะมุณีย์ และประสงค์ ประณีตพลกรัง, 2010) ในการสอน (training) ต้นไม้การตัดสินใจมากมาย ซึ่งส่วนมากมาจากวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบละโมภ (Greedy search) จากบนลงล่าง (top-down) ชื่อว่า ID3 ซึ่งถูกพัฒนาโดยJohn Ross Quinlan ในปี 1986

2.1.1 เอนโทรปี (Entropy) ID3 นั้นสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด (ขอใช้คำว่าลักษณะแทนตัวแปรต้น) ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำๆไปเรื่อยๆเพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (Rrecursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ E (S) ดังนี้

$$E(S) = -\sum_{j=1}^{n} p_s(j) \log_2 p_s(j)$$

เมื่อ

- *S* คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี
- *Ps(j)* คืออัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (boolean) ใช่กับไม่ใช่ เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปีคือ

$$E(S) = -p_{yes} \log_2(p_{yes}) - p_{no} \log_2(p_{no})$$

2.1.2 เกนความรู้ (Information Gain) จากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆมีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามได้โดย

$$Gain(S,A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

เมื่อ

- *S* คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี
- *E* คือเอนโทรปีของตัวอย่าง
- *A* คือตัวแปรต้นที่พิจารณา
- *value (A)* คือเซตของค่าของ *A* ที่เป็นไปได้
- S_v คือตัวอย่างที่ *A* มีค่า *v* ทั้งหมด

2.2 C4.5 เป็นอัลกอริทึมในการสร้างต้นไม้ พัฒนาโดย J. Ross Quinlan (1993) โดยนำเอา ID3 มาปรับปรุงให้มีความสามารถมากขึ้นใช้วิธีการ Information Gain เพิ่มเติมการจัดการกับข้อมูล ตัวเลข ข้อมูลที่ขาดไป ไม่สมบูรณ์ (Missing Value, Noisy Data) และการ Prune ด้วยการแทน Branch (กิ่ง) ที่ไม่ช่วยการตัดสินใจด้วย Leaf Node ที่ตัดสินใจได้ดีกว่า

2.3 CART เป็นอัลกอริทึมในการสร้างต้นไม้ (Classification and Regression Trees) คือต้นไม้ตัดสินใจด้านสถิติเพื่อการทำนาย และจำแนกกลุ่มที่พัฒนามาเพื่อสร้างต้นไม้ ที่ใช้จัดการกับการจำแนกกลุ่มแบบตัวเลขให้ดียิ่งขึ้น พัฒนาโดย Breiman et al. (1984) โดยอาศัยกลักการเดียวกันกับ C4.5 ในด้านการจัดการ Missing Values และมีการจัดการได้ดีกับข้อมูลหลายชนิด เช่น Floats, Integer และค่าที่ไม่ใช่ตัวเลข

3. งานวิจัยที่เกี่ยวข้อง

นางสาวรติกร คงประเสริฐ (2554) แม่แบบระบบคลังข้อมูลและเหมืองข้อมูลสำหรับหน่วยงานด้านการศึกษา มีวัตถุประสงค์ในการวิจัยเพื่อต่อยอดการออกแบบและพัฒนาแม่แบบคลังข้อมูล และเหมืองข้อมูลสำหรับงานด้านการศึกษา สามารถแสดงผลรายงานในรูปแบบ Dynamic Report ได้ และเป็นการประยุกต์ใช้เทคนิค Data mining ในการออกรายงานเชิงวิเคราะห์หรือเชิงพยากรณ์ จากข้อมูลสำหรับงานด้านการศึกษา วิธีการของผู้ทำการวิจัยคือ พัฒนาแม่แบบสำหรับหน่วยงานทางด้านการศึกษาเพิ่มจากโครงงานที่มีอยู่เดิม เพิ่มส่วนการออกรายงานแบบ Dynamic แสดงผลในรูปแบบของรายงานและกราฟ โดยนำเทคนิค Decision Tree มาใช้ในการทำนายเพื่อช่วยในการตัดสินใจเลือกภาควิชาในกลุ่มนักศึกษากลุ่มตัวอย่าง ผลที่ได้จากการวิจัยทำให้แม่แบบการพัฒนาระบบ BI สำหรับหน่วยงานด้านการศึกษาที่มีอยู่เดิม สามารถออกรายงานในรูปแบบของ Dynamic รวมถึงการออกรายงานเชิงวิเคราะห์พยากรณ์ และนำแม่แบบดังกล่าวไปใช้พัฒนาระบบ BI ในหน่วยงานด้านการศึกษาอื่น ๆ ต่อไปได้

นางสาวพัชราภรณ์ ราชประดิษฐ์(2553) ระบบผู้เชี่ยวชาญเพื่อวินิจฉัยโรคข้าว มีวัตถุประสงค์ในการวิจัยเพื่อพัฒนาระบบผู้เชี่ยวชาญเพื่อวินิจฉัยโรคข้าวที่ให้คำตอบการวินิจฉัยโรคข้าวได้อย่างถูกต้อง และเพื่อรวบรวมความจริงเกี่ยวกับข้าวและพัฒนาฐานความรู้โรคข้าว สุดท้าย เพื่อให้ผู้ที่เกี่ยวข้องด้านเกษตรกรรมหรือผู้ที่สนใจนำไปใช้ในการวินิจฉัยโรคข้าว วิธีการของผู้ทำการวิจัย คือใช้เครื่องมือการสร้าง Export System ชื่อว่า CLIPS มาใช้สร้างกฎและอนุมาน โดยจำแนกลักษณะอาการของโรค แล้วนำมาวิเคราะห์สร้างเป็น Decision Tree ตามอัลกอริทึม C4.5 โดยใช้โปรแกรม WEKA เมื่อสร้างกฎเสร็จเรียบร้อยแล้วจะทำการตรวจสอบความถูกต้องของกฎโดยมีผู้เชี่ยวชาญโรคข้าวเป็นผู้ตรวจสอบ เปรียบเทียบผลที่ได้จากการวินิจฉัยของผู้เชี่ยวชาญ กับระบบผู้เชี่ยวชาญ ผลการประเมินประสิทธิภาพของระบบพบว่าสามารถวินิจฉัยโรคได้ตรงกับผู้เชี่ยวชาญได้ถึง 94.5%

วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการหาความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีต่อระดับผลการเรียน ของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ โดยได้นำข้อมูลนักศึกษา และผลการเรียนของนักศึกษาจากฐานข้อมูล ระบบทะเบียนและประมวลผล ของมหาวิทยาลัยแห่งนี้ มาเลือกใช้ คัดกรองเฉพาะข้อมูลที่สนใจ แล้วใช้โปรแกรม KNIME (KNIME.com AG, 2013) เป็นเครื่องมือในการวิเคราะห์ข้อมูล เพื่อได้ปัจจัยข้อมูลนักศึกษา ที่มีผลต่อระดับผลการเรียน

งานวิจัยนี้ใช้ข้อมูลนักศึกษาจากฐานข้อมูลระบบทะเบียนและประมวลผล ของมหาวิทยาลัยสังกัดรัฐบาลแห่งหนึ่ง ซึ่งตั้งอยู่ในเขตกรุงเทพฯและปริมณฑล เป็นระบบสารสนเทศที่จัดเก็บข้อมูลด้วยโปรแกรมจัดการฐานข้อมูล Microsoft SQL Server 2008 โดยเงื่อนไขการนำข้อมูลมาใช้ในงานวิจัยนี้ เงื่อนไขใน SQL Query ที่เลือก คือ

- เป็นนักศึกษาที่ในระดับปริญญาตรี
- เป็นนักศึกษา หลักสูตร 4 ปี
- เป็นนักศึกษาที่มีข้อมูลผลการเรียนใน ชั้นปีที่ 1 (ใช้ข้อมูลผลการเรียนเฉพาะ วิชาที่เรียนปีที่ 1)

โดยจากตัวแปรตาม (Dependent Variables) ที่ตั้งไว้คือระดับผลการเรียนของนักศึกษา (ดี/ไม่ดี) โดยนับข้อมูลเฉพาะผลการเรียนในชั้นปีที่ 1 เท่านั้น โดยที่

- “ดี” ในงานวิจัยนี้หมายถึงเกรดเฉลี่ยทุกวิชามีค่าตั้งแต่ 2.5 ขึ้นไป
- -“ไม่ดี” ในงานวิจัยนี้หมายถึงเกรดเฉลี่ยทุกวิชามีค่าน้อยกว่า 2.5

โดยการสร้างแบบจำลองต้นไม้ตัดสินใจ เพื่อหาความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีต่อระดับผลการเรียน ตัวแปรทั้งหมด 12ตัวแปร ได้แก่ ปีที่เข้าศึกษา, เพศ, จังหวัดภูมิลำเนา, สถานะทางร่างกาย, ศาสนา, ประเภทแอดมิชชั่น ,วุฒิการศึกษาสูงสุด, จังหวัดที่จบการศึกษาสูงสุด, สถานภาพบิดา, รายได้บิดา, สถานภาพมารดา, รายได้มารดา มีการวิเคราะห์จากสมมติฐานที่ตั้งเอาไว้ 5 สมมติฐาน เป็นการหาปัจจัยข้อมูลนักศึกษาใน 2 ด้านดังนี้

Table 1 The Concept Information for Hypothesis Testing

	ด้านที่ 1	ด้านที่ 2
สมมติฐาน	สมมติฐานที่ 1 ภาคของจังหวัดภูมิลำเนาของนักศึกษา ที่เป็นจังหวัดใน 3 จังหวัดชายแดนภาคใต้มีผลต่อระดับผลการเรียนของนักศึกษา สมมติฐานที่ 2 สถานภาพของบิดา มีผลต่อระดับผลการเรียนของนักศึกษา สมมติฐานที่ 3 สถานภาพของมารดา มีผลต่อระดับผลการเรียนของนักศึกษา สมมติฐานที่ 4 วุฒิการศึกษาสูงสุด มีผลต่อระดับผลการเรียนของนักศึกษา	สมมติฐานที่ 5 ปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน คณะวิศวกรรมศาสตร์ สาขาวิศวกรรมเครื่องกล คือ จังหวัดภูมิลำเนาอยู่ในกรุงเทพและปริมณฑล สถานภาพของบิดาคือมีชีวิตอยู่ สถานภาพของมารดาคือมีชีวิตอยู่
จุดประสงค์	หาปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียน เลือกวิจัยทุกสาขาวิชา เพื่อเป็นภาพรวมทั้งมหาวิทยาลัย	หาปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียน เลือกวิจัยเฉพาะสาขาวิชาวิศวกรรมเครื่องกล
ทำให้แบ่งชุดข้อมูลในงานวิจัยนี้	**ข้อมูลชุดที่ 1** : ข้อมูลนักศึกษา ทุกสาขาวิชา	**ข้อมูลชุดที่ 2** : ข้อมูลนักศึกษา วิศวกรรมเครื่องกล

กรอบแนวความคิดในการวิจัย

ตัวแปรอิสระ (Independent Variables) ตัวแปรตาม (Dependent Variables

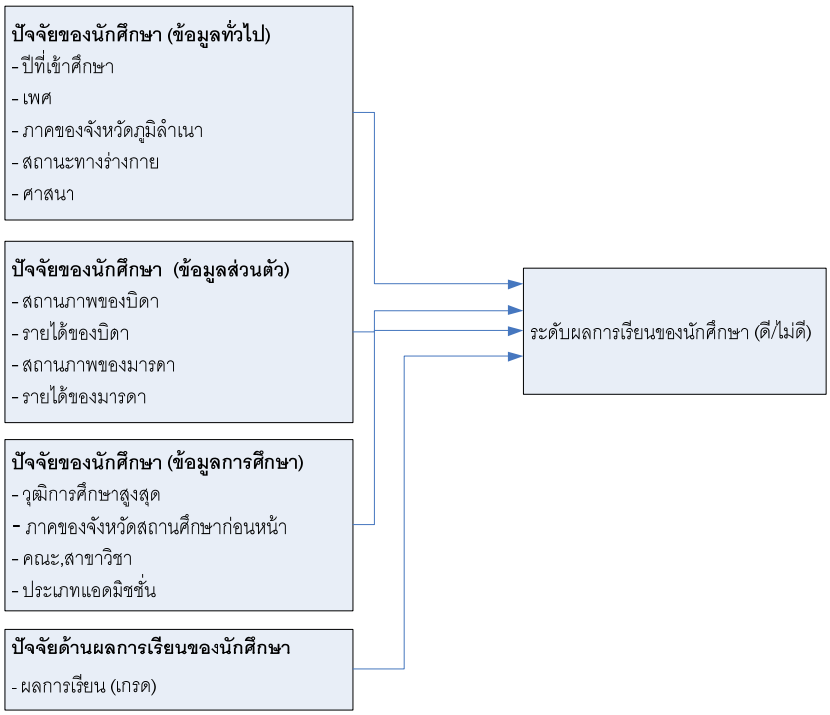

Figure 1 The Concept of Frame Work

ผลการวิจัย

ได้นำข้อมูลผ่านขั้นตอนก่อนการประมวลผล (Preprocessing) โดยการทำการกลับกรองข้อมูล (Data Cleaning) ทำให้ได้12 ตัวแปร แสดงให้เห็นการตัดสินใจใน Table 2

Table 2 Branches of the decision tree of the only Grade = "good"

กิ่งที่	ตัวแปรลำดับชั้นที่ 1	ตัวแปร ลำดับชั้นที่ 2	ตัวแปร ลำดับชั้นที่ 3	ตัวแปรลำดับ ชั้นที่ 4
1	ภาคจังหวัดที่จบการศึกษาสูงสุด= ภาคตะวันออก	-	-	-
2	ภาคจังหวัดที่จบการศึกษาสูงสุด= ภาคเหนือ	รายได้มารดา = ไม่มีรายได้	-	-
3	ภาคจังหวัดที่จบการศึกษาสูงสุด= ภาคกลาง	วุฒิการศึกษาสูงสุด= ม.ปลาย	-	-
4	ภาคจังหวัดที่จบการศึกษาสูงสุด= กรุงเทพฯและปริมณฑล	ที่เข้าศึกษา= 2554	สถานภาพมารดา= ไม่ระบุ	-
5	ภาคจังหวัดที่จบการศึกษาสูงสุด= กรุงเทพฯและปริมณฑล	ปีที่เข้าศึกษา= 2553	รายได้มารดา= 150,000-300,000 ต่อปี	-

กิ่งที่	ตัวแปรลำดับชั้นที่ 1	ตัวแปร ลำดับชั้นที่ 2	ตัวแปร ลำดับชั้นที่ 3	ตัวแปรลำดับชั้นที่ 4
6	ภาคจังหวัดที่จบการศึกษาสูงสุด= กรุงเทพฯและปริมณฑล	ปีที่เข้าศึกษา= 2553	รายได้มารดา= ไม่มีรายได้	-
7	ภาคจังหวัดที่จบการศึกษาสูงสุด= กรุงเทพฯและปริมณฑล	ที่เข้าศึกษา= 2550	ภาคจังหวัดภูมิลำเนา= ภาคกลาง	-
8	ภาคจังหวัดที่จบการศึกษาสูงสุด= กรุงเทพฯและปริมณฑล	ปีที่เข้าศึกษา= 2550	ภาคจังหวัดภูมิลำเนา= กรุงเทพฯและปริมณฑล	สถานภาพบิดา=ถึงแก่กรรม

ทำการหาค่าความแม่นยำ (Accuracy) ที่ได้ออกมา (Output) แสดงดังตารางด้านล่างจะได้ข้อมูลที่ได้ค่าความแม่นยำสูงสุดในทั้งหมด 12 ครั้งคือ ข้อมูลชุดที่ 1 (ข้อมูลนักศึกษาทุกสาขา) ครั้งที่ 7 มีความถูกต้องที่ 68.17% และข้อมูลชุดที่ 2 (ข้อมูลนักศึกษาวิศวกรรมเครื่องกล) ครั้งที่ 1 มีความถูกต้องที่ 84.78%

จากผลการทดสอบ ชุดข้อมูลที่ 1(ข้อมูลนักศึกษาทุกสาขาวิชา) เพื่อพิสูจน์สมมติฐานใน 4 ข้อแรก คือ

สมมติฐานที่ 1 ภาคของจังหวัดภูมิลำเนาของนักศึกษา ที่เป็นจังหวัดใน 3 ตัวหวัดชายแดนภาคใต้ มีผลต่อระดับผลการเรียนของนักศึกษา คือผลการเรียนไม่ดี

สมมติฐานที่ 2 สถานภาพของบิดา มีผลต่อระดับผลการเรียนของนักศึกษาแตกต่างกัน

สมมติฐานที่ 3 สถานภาพของมารดา มีผลต่อระดับผลการเรียนของนักศึกษาแตกต่างกัน

สมมติฐานที่ 4 วุฒิการศึกษาสูงสุด มีผลต่อระดับผลการเรียนของนักศึกษาแตกต่างกัน

Table 3 The Result form Experimentation

ครั้งที่ทดสอบ	Input			Output	
	Quality measure	Pruning method	Min number records per node	Accuracy ข้อมูลชุดที่ 1 (%)	Accuracy ข้อมูลชุดที่ 2 (%)
1	Gini index	No Pruning	2	67.29	84.78
2	Gini index	No Pruning	3	66.67	82.07
3	Gini index	No Pruning	4	65.69	79.35
4	Gini index	MDL	2	60.47	71.74
5	Gini index	MDL	3	60.47	71.74
6	Gini index	MDL	4	60.47	71.74
7	Gain Ratio	No Pruning	2	68.17	84.24
8	Gain Ratio	No Pruning	3	66.83	81.52
9	Gain Ratio	No Pruning	4	66.21	79.35
10	Gain Ratio	MDL	2	60.70	72.83

ดังนั้น : จะได้ว่าสมมติฐานที่ตั้ง **ไม่ถูกต้อง**

เพราะ : ปัจจัยข้อมูลของนักศึกษาชุดข้อมูลที่ 1 : ข้อมูลทุกสาขาวิชา ซึ่งเป็น ภาพรวมทั้งมหาวิทยาลัย ที่เรียนดี มีลักษณะใดลักษณะหนึ่งใน 4 ลักษณะดังต่อไปนี้

1. สภาพร่างกาย =บกพร่องทางการพูด หรือ
2. สภาพร่างกาย = ไม่พิการ และ เพศ=หญิง และ ศาสนา = ฮินดู หรือ
3. สภาพร่างกาย = ไม่พิการ และ เพศ=หญิง และ ศาสนา = อิสลาม และ รายได้มารดา=ไม่ระบุ
4. สภาพร่างกาย = ไม่พิการ และ เพศ=หญิง และ ศาสนา = อิสลาม และ รายได้มารดา=150,000-300,000 ต่อปี

เหตุที่เลือก 4 กิ่งนี้มาเป็นคำตอบเพราะเป็นกิ่งที่มีลำดับชั้นจำนวนน้อย กว่ากิ่งอื่น ๆ ซึ่งการแยกข้อมูลของต้นไม้ หากสามารถจบการแยกข้อมูลได้ภายในระดับชั้นน้อย ถือว่ามีความสามารถในการจำแนกข้อมูลได้มากกว่า วัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงสุดเป็น 68.17% ซึ่งถือว่าอยู่ในเกณฑ์ไม่สูงมาก ยังไม่เป็นการทดสอบที่เชื่อถือได้ อาจปรับปรุงข้อมูลโดยเลือกข้อมูล ให้น้อยลง

จากการหาปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียน จากผลการทดสอบ ชุดข้อมูลที่ 2 (ข้อมูลนักศึกษาวิศวกรรมเครื่องกล) เพื่อพิสูจน์สมมติฐานข้อที่ 5 คือ

<u>สมมติฐานที่ 5</u> ปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน คณะวิศวกรรมศาสตร์ สาขาวิศวกรรมเครื่องกล คือ จังหวัดภูมิลำเนาอยู่ในกรุงเทพและปริมณฑล สถานภาพของบิดาคือมีชีวิตอยู่ สถานภาพของมารดาคือมีชีวิตอยู่

จากการทดลองพบว่า ปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียนคณะวิศวกรรมศาสตร์ สาขาวิศวกรรมเครื่องกล มีทั้งหมด 7 ตัวแปร จาก Input ทั้งหมด 12 ตัวแปร ทั้งนี้7 ตัวแปรมาจากต้นไม้ตัดสินใจที่โปรแกรมได้แสดงออกมา โดยสามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด คือนักศึกษาต้องมีลักษณะ ใดลักษณะหนึ่งใน 3 ลักษณะดังต่อไปนี้

1. ภาคของจังหวัดที่จบการศึกษาสูงสุด = “ภาคตะวันออก”

2. ภาคของจังหวัดที่จบการศึกษาสูงสุด = “ภาคเหนือ” และ รายได้มารดา = “ไม่มีรายได้”

3. ภาคของจังหวัดที่จบการศึกษาสูงสุด = “ภาคกลาง” , วุฒิการศึกษาสูงสุด= “ม.ปลาย”

ซึ่งข้อมูลเหล่านี้ได้เลือก ที่ได้ข้อมูลนักศึกษา (สาขาวิศวกรรมเครื่อง) ที่มีผลการเรียนดี จำนวน 8 กิ่ง จึงได้เลือกมา กิ่งที่สำคัญ 3 กิ่ง ได้กิ่งที่ 1 กิ่งที่ 2 และ กิ่งที่ 3 เหตุที่เลือก 3 กิ่งนี้มาเป็นคำตอบเพราะเป็นกิ่งที่มีลำดับชั้นจำนวนน้อย กว่ากิ่งอื่น ๆ ซึ่งการแยกข้อมูลของต้นไม้ หากสามารถจบการแยกข้อมูลได้ภายในระดับชั้นน้อย ถือว่ามีความสามารถในการจำแนกข้อมูลได้มากกว่า วัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง เป็นการทดสอบที่เชื่อถือได้ อธิบายได้ว่าพารามิเตอร์ที่มีความเหมาะสมดี

สรุปการวิจัย

จากการตั้งสมมติฐานทั้ง 5 ข้อ แล้วแบ่งการทดสอบข้อมูลเป็น 2 ชุด พบว่าได้ผลที่ไม่ตรงกับสมมติฐาน เลือกตัวแปรที่ตั้งสมมติฐานที่คาดว่าจะมีผลต่อระดับผลการเรียน แต่หลังจากทำการวิจัยแล้วพบว่ามีผลน้อยมากจนถือได้ว่า ไม่มีผล หรือบางตัวแปร ก็ไม่มีผลเลย ต่อระดับผลการเรียน เพราะพบว่า ไม่ได้อยู่ในNodeของต้นไม้ตัดสินใจระดับบน หากต้องการให้มีผล อาจเลือกตัวแปรที่น่าจะเป็นปัจจัยที่ส่งผลกับการเรียน เช่น ระดับผลการเรียนที่จบการศึกษาสูงสุด ความสม่ำเสมอการเข้าเรียนในการเรียนที่จบการศึกษาสูงสุด เป็นต้น น่าจะส่งผลให้สมมติฐานที่ตั้งเอาไว้มีความถูกต้องมากกว่านี้ ส่วนในอนาคตจะไปประยุกต์ใช้กับงานเหมืองข้อมูลด้านอื่น ๆ ได้ เพื่อใช้ในกาค้นหาความจริง การทำนายข้อมูล การจำแนกข้อมูล เช่น งานวิเคราะห์ข้อมูลการศึกษา ข้อมูลวัดคุณภาพประมาณงานทั้งงานด้านเอกชนและรัฐบาล เป็นต้น

กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิที่เอื้อเฟื้อข้อมูลนักศึกษา และสถาบันบัณฑิตพัฒนบริหารศาสตร์ที่ชี้แนะเทคนิคต้นไม้ตัดสินใจ

เอกสารอ้างอิง

ชัดชัย แก้วตา และ อัจฉรา มหาวีรวัฒน์. การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ. ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี, 2553

ชลนิศา สาระ และเยาวดี เต็มธนาภัทร์, 2007. การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ. Available URL: http://www.gits.kmutnb.ac.th/ethesis/files/page34.html.

ชุติมา อุตมะมุณีย์ และประสงค์ ประณีตพลกรัง, 2010. การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาเรียนของนักศึกษาระดับอุดมศึกษา. Available URL: http://ist-journal.mut.ac.th/Journal/vol-1-2/page%2039-48.pdf.

จิโรจน์ ภาคคีรี และ กาญจนา วิริยะพันธ์, 2550. การวิเคราะห์รูปแบบการบุกรุกข้อมูลบนเครือข่าย โดยใช้เทคนิคดาต้าไมน์นิง, วารสารเทคโนโลยีสารสนเทศ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

พัชาภรณ์ ราชประดิษฐ์, 2011. ระบบผู้เชี่ยวชาญเพื่อวินิจฉัยโรคข้าว. Available URL:http://www.thapra.lib.su.ac.th/objects/thesis/fulltext/snamcn/Patcharaporn_Ratchapradit/fulltext.pdf.

อนุชิต ศรวิชัย และ วงกต ศรีอุไร, 2556. ระบบสนับสนุนการตัดสินใจเลือกซื้อโทรศัพท์มือถือโดยใช้เทคนิคต้นไม้ตัดสินใจ, ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

สำนักงานคณะกรรมการการอุดมศึกษา และมหาวิทยาลัยเทคโนโลยีพระจอมเหล้าพระนครเหนือ, 2013. โครงการเครือข่ายห้องสมุดในประเทศไทย. Available URL: http://202.28.199.11/tdc/.

KNIME.com AG. 2013. KNIME Data mining Product. Available URL: http://www.knime.org/.

KNIME.com GmbH, 2013. Knime Decision tree learner configuration. Available URL: http://tech.knime.org/forum/knime-users/knime-decision-tree-learner-configuration.

KNIME.com GmbH, 2013. Finding Percentage. Available URL: http://tech.knime.org/forum/knime-general/finding-percentage.

Rosaria, P. 2013. DMR - Data Mining and Reporting. Available URL: http://www.dataminingreporting.com/2/post/2013/05/knime-and-big-data.html.