

การค้นหาคำคุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่ม
โปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม ด้วยเทคนิคเหมืองข้อมูล
The Knowledge Discovery of Students' Critical Characteristics towards
Learning Achievement of Nakhon Pathom Rajabhat University
Students' Program in Computer via Data Mining Technique.

ไกรรุ่ง เสงพะพรหม¹,
สุพจน์ เสงพะพรหม² และสุวิมล มรรควิบูลย์ชัย³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อค้นหาคำคุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม ด้วยเทคนิคเหมืองข้อมูล เพื่อสร้างโมเดลหรือต้นแบบ สำหรับการพัฒนาระบบสนับสนุนการตัดสินใจการเข้าศึกษาสำหรับผู้สนใจเรียนในสาขาวิชาของกลุ่มโปรแกรมคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม โดยได้ทำการศึกษาข้อมูลจากนักศึกษาที่ศึกษาอยู่ในสาขาคอมพิวเตอร์ชั้นปีที่ 3 และปีที่ 4 ในปีการศึกษา 2556 โดยใช้ระดับผลการเรียนเฉลี่ย ณ ปัจจุบันเป็นเกณฑ์ในการจำแนกกลุ่มนักศึกษา

ในการศึกษาผู้วิจัยได้ทำการทดลองกับเทคนิคทางเหมืองข้อมูล 2 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ C4.5 และข่ายงานความเชื่อเบย์ (BayesNet) โดยแบ่งการทดลองออกเป็น 2 ลักษณะ คือ การทดลองกับชุดข้อมูลทั้งหมดทุกแอททริบิวต์ และการทดลองด้วยการเลือกทุกแอททริบิวต์ด้วยวิธีการ SNR (signal to noise ratio) เพื่อค้นหาคำคุณลักษณะสำคัญที่มีผลต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์สำหรับสร้างเป็นต้นแบบการพัฒนากระบวนการตัดสินใจเลือกเรียนในสาขาวิชาของกลุ่มโปรแกรมคอมพิวเตอร์ จากผลการทดลองทั้งสองแบบยืนยันว่า วิธีต้นไม้ตัดสินใจ C4.5 ให้ประสิทธิภาพในการทำงานที่ดีกว่าเมื่อเทียบกับวิธีการโครงข่ายเบย์ (BayesNet)

คำสำคัญ: คำค้นหาไม่มี, C4.5, ต้นไม้ตัดสินใจ, ข่ายงานความเชื่อเบย์, ระบบสนับสนุนการตัดสินใจ

Abstract

The purpose of this research is to discover the critical characteristics of students' success in computer studies in programme in computer in Nakhon Pathom Rajabhat University using data mining techniques to create the model for developing a decision making process of those who are interested in enrolling in computer related program in the faculty of science and technology at Nakhon Pathom Rajabhat University. The dataset obtained from 3rd and 4th year of computer related program students in academics year 2013 were used in this study. The current Grade Point Average was used to categorize students into group.

In this research, the researchers used two data mining techniques, which are decision tree and BayesNet. The experiment was divided into 2 sets. First, the experiment was done to all characteristic. Second, the experiment is being done to those who were chosen by the techniques called SNR to discover the critical characteristics that affect to the effectiveness of learning for the programme in

^{1,2,3} โปรแกรมวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม 85 ถ.มาลัยแมน อ.เมือง จ.นครปฐม 73000

¹ E-mail: kairung2011.heng@gmail, ² E-mail: supojn@yahoo.com and ³ E-mail: suvimol@npru.ac.th

computer and become to be a prototype of the decision support system to enroll in programme in computer. The result from both experiments has shown that C4.5 performed better compared to the BayesNet.

Keywords: data mining, C4.5, decision tree, BayesNet, decision support system

บทนำ

การสมัครเข้าศึกษาในระดับอุดมศึกษานั้นเป็นกระบวนการที่ยุ่งยากซับซ้อน เพราะนอกจากจะต้องรู้จักตัวเองแล้ว ผู้เลือกเรียนจะต้องรู้เนื้อหาของหลักสูตร คือ รู้เนื้อหาโดยละเอียดของหลักสูตรซึ่งอาจจะยากต่อทำความเข้าใจสำหรับนักเรียนที่เพิ่งสำเร็จการศึกษาในระดับชั้นมัธยมศึกษาตอนปลาย ปัญหาที่พบส่วนใหญ่เนื่องจากนักเรียนยังไม่ทราบถึงความถนัดหรือความชอบของตนเอง โดยกระบวนการตัดสินใจเข้าศึกษาต่อที่ผู้สมัครจำเป็นต้องทราบความสนใจ ความสามารถ และความถนัดของตนเองจึงจะทำให้ตัดสินใจได้อย่างถูกต้องเหมาะสม บ่อยครั้งที่พบว่านักศึกษาที่เข้าศึกษาต่อในระดับอุดมศึกษามีแรงจูงใจในการเลือกหลักสูตรมาจากเพื่อน ผู้ปกครองเป็นหลัก ดังนั้นเมื่อเข้าไปศึกษาจริง ๆ ทำให้ไม่ประสบผลสำเร็จในการเรียนเนื่องจากได้เลือกเรียนในสาขาวิชาที่ไม่เหมาะสมกับบุคลิก ความสามารถ ความสนใจ รวมถึงความถนัดของตนเอง ผลสุดท้ายนักศึกษาต้องทบทวนศึกษาต่อจนสำเร็จการศึกษา และผลที่ตามมาคือเมื่อสำเร็จการศึกษาแล้ว ทำให้ไม่สามารถทำงานตามสาขาที่จบการศึกษาได้หรือทำงานอย่างไม่เต็มประสิทธิภาพทำให้มีโอกาสก้าวหน้าในการทำงานน้อย

จากการศึกษาทางวิจัยที่เกี่ยวข้องพบว่าได้มีงานวิจัยเป็นจำนวนมากที่ให้ความสำคัญและได้ทำการศึกษาถึงปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการศึกษา รวมถึงการพัฒนาตัวแบบสำหรับการสร้างระบบสนับสนุนการตัดสินใจหลากหลายรูปแบบ ได้แก่ การวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาระดับปริญญาตรีโดยใช้คอมพิวเตอร์แมชชีน (ภัทร์พงศ์ พงศ์ภัทรกานต์, 2553) โดยงานวิจัยนี้ใช้ชุดข้อมูลนักศึกษาที่เข้าศึกษาระหว่างปี พ.ศ. 2546-2549 ของมหาวิทยาลัยราชภัฏเลยมีจำนวน 11 แอททริบิวต์ และ 12,865 ชุดข้อมูล โดยได้ทำการทดลองวัดประสิทธิภาพความถูกต้องเปรียบเทียบกับนิวรอลเน็ตเวิร์ก และ C5.0 ซึ่งแบบจำลองแบบคอมพิวเตอร์แมชชีนให้ประสิทธิภาพในการจำแนกข้อมูลสูงที่สุด มีค่าเท่ากับ 75.32 และงานวิจัยเกี่ยวกับการสร้างตัวแบบสำหรับหาปัจจัยที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษาและพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจอย่างอัตโนมัติแบบออนไลน์ (ชุดิมา อุดมะมุณี และประสงค์ ปราณีตพลกรัง, 2553) โดยใช้ข้อมูลด้านการเรียนของกลุ่มตัวอย่างจากนักศึกษาในระดับอุดมศึกษาทั้งภาครัฐและเอกชนจำนวน 9 มหาวิทยาลัย โดยใช้วิธีของข่ายงานเบย์ พบว่าตัวแปรสำคัญที่มีผลต่อการตัดสินใจเลือกสาขาวิชาเรียนของนักศึกษา ได้แก่ เกรดเฉลี่ยวิชาคณิตศาสตร์ ก่อนศึกษา เกรดเฉลี่ยวิชาคณิตศาสตร์ขณะกำลังศึกษา เกรดเฉลี่ยวิชาเขียนโปรแกรม ความรู้ในการพัฒนาระบบคอมพิวเตอร์ และซอฟต์แวร์ ความรู้ด้านวิชาเรียน 1 ความรู้ด้านวิชาเรียน 2 และความถนัด งานวิจัยถัดมาได้ศึกษาอิทธิพลของบริบททางสังคมต่อการตัดสินใจเลือกศึกษาต่อสาขาวิชาวิทยาศาสตร์และเทคโนโลยีของนักเรียนมัธยมศึกษาชั้นปีที่ 6 (ธีรดา ภิฎโญ, 2551) พบว่า ปัจจัยที่มีอิทธิพลมากที่สุดต่อการตัดสินใจเลือกศึกษาต่อสาขาวิชาวิทยาศาสตร์และเทคโนโลยี คือ ปัจจัยเจตคติและความสนใจทางด้านวิทยาศาสตร์และเทคโนโลยี

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงได้นำแนวคิดเกี่ยวกับการทำเหมืองข้อมูล (data mining) มาใช้สำหรับการเรียนรู้คุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม โดยทำการศึกษากับนักศึกษาในกลุ่มโปรแกรมคอมพิวเตอร์ในชั้นปีที่ 3 และปี 4 ซึ่งจะทำการศึกษาจากข้อมูลพื้นฐานและเจตคติของผู้เรียนก่อนเข้าศึกษา เพื่อสร้างเป็นต้นแบบสำหรับการพัฒนาระบบสนับสนุนการตัดสินใจเข้าศึกษาต่อในสาขาวิชาของกลุ่มโปรแกรมคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม เพื่อเป็นทางเลือกสำหรับผู้สนใจจะเข้าศึกษาต่อในระดับปริญญาตรีของกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐมต่อไป

ทฤษฎีที่เกี่ยวข้อง

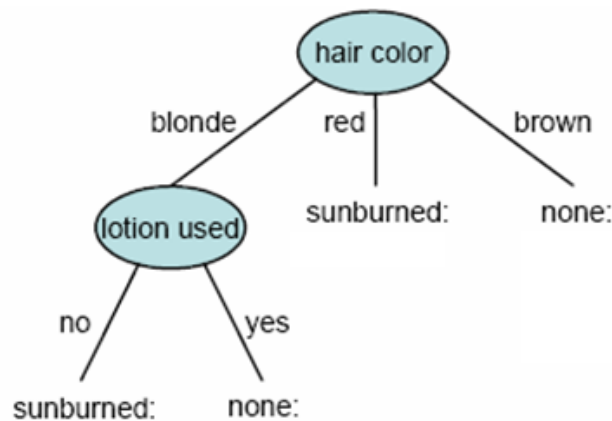
1. การจำแนกประเภทข้อมูล (data classification)

การจำแนกประเภทข้อมูล เป็นการเรียนรู้แบบมีผู้สอน (supervised learning) โดยวัตถุประสงค์เพื่อกำหนดประเภทข้อมูลให้กับข้อมูลใหม่ ๆ ที่ยังไม่รู้ประเภท โดยใช้ลักษณะของชุดข้อมูลที่มีอยู่ซึ่งเป็นชุดข้อมูลที่รู้ประเภทแล้ว เทคนิคพื้นฐานที่นิยมนำมาใช้สำหรับการจำแนกประเภทข้อมูล ได้แก่

1.1 ต้นไม้ตัดสินใจ (decision tree)

ต้นไม้ตัดสินใจ (quinlan, 1986) เป็นตัวแบบทางคณิตศาสตร์ที่ใช้จำแนกประเภทข้อมูล โดยพิจารณาจากคุณลักษณะของข้อมูล ซึ่งค่าของคุณลักษณะเหล่านี้จะอยู่ในรูปของค่าที่ไม่ต่อเนื่อง (discrete value) เช่น เพศ คณะวิชา สีผม เป็นต้น

ในการเรียนรู้จะแทนความรู้ในรูปแบบโครงสร้างต้นไม้ โดยมีโนดภายใน (inner node) เป็นชื่อของคุณลักษณะต่าง ๆ ของชุดข้อมูลที่นำมาเรียนรู้ แต่ละโนดภายในต้นไม้จะมีกิ่ง (branch) เท่ากับจำนวนค่าที่เป็นไปได้สำหรับคุณลักษณะนั้น ๆ ส่วนโนดใบจะเป็นประเภทที่เป็นไปได้ของข้อมูล ตัวอย่างต้นไม้ตัดสินใจ แสดงดังรูปภาพที่ 1 ซึ่งจากรูปแสดงตัวอย่างต้นไม้ตัดสินใจสำหรับจำแนกประเภทผลลัพธ์ของการอาบแดด โดยคุณลักษณะที่ใช้มี 2 คุณลักษณะ คือ สีผม (hair color) ซึ่งมีค่าที่เป็นไปได้ 3 ค่า (blonde, red และ brown) และ การใช้โลชั่น (lotion used) ซึ่งมีค่าที่เป็นไปได้ 2 ค่า (yes และ no) ส่วนผลลัพธ์ของการจำแนกประเภทจะประกอบด้วย 2 ประเภท คือ ผิวไหม้ (sunburned) และ ไม่เป็นอะไร (none)



รูปภาพที่ 1 ตัวอย่างต้นไม้ตัดสินใจ

ที่มา: บุญเสริม กิจสิริกุล, 2546: 154

จากรูปภาพที่ 1 จะแทนกฎการจำแนกประเภท ดังนี้

- IF <hair color> = blonde and <lotion used> = no THEN sunburned
- IF <hair color> = blonde and <lotion used> = yes THEN none
- IF <hair color> = red THEN sunburned
- IF <hair color> = brown THEN none

ขั้นตอนการเรียนรู้สำหรับต้นไม้ตัดสินใจ แสดงดังรูปภาพที่ 2

ต้นไม้ตัดสินใจ (decision tree)

ข้อมูลนำเข้า: ชุดข้อมูลสอน $x(a_1, a_2, \dots, a_n; class)$

1. ทดสอบคุณลักษณะของข้อมูลด้วยค่า information gain (IG) ซึ่งคำนวณได้จาก

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{1}$$

$$Entropy(S) = \sum_{i=1}^c - p_i \log_2 p_i \tag{2}$$

โดยที่

S คือ เซตของข้อมูลทั้งหมดที่ใช้ในการทดสอบ

S_v คือ เซตย่อยของ S ที่คุณลักษณะ A ของข้อมูลมีค่าเท่ากับ v

$Values(A)$ คือ ค่าที่เป็นไปได้ของคุณลักษณะ A

$Entropy(S)$ คือ ค่าความไร้ระเบียบของข้อมูลในเซต S

c คือ ประเภท (class) ของข้อมูลที่เป็นไปได้

p_i คือ ค่าความน่าจะเป็นของข้อมูลที่อยู่ในประเภท i

2. เลือกคุณลักษณะที่ให้ค่า IG สูงที่สุดมาสร้างเป็นโนดราก (root node)
3. ขณะที่ $Entropy(S_v)$ ของกิ่งใด ๆ ที่ยังไม่เท่ากับ 0 ให้นำคุณลักษณะที่เหลือมาทดสอบและเลือกคุณลักษณะที่ให้ค่า IG สูงที่สุดมาแตกเป็นกิ่งของต้นไม้ต่อไปเรื่อย ๆ

ผลลัพธ์: ต้นไม้ตัดสินใจ

รูปภาพที่ 2 ขั้นตอนการเรียนรู้ของต้นไม้ตัดสินใจ

1.2 ข่ายงานความเชื่อเบย์ (Bayesian belief network)

ข่ายงานเบย์ (Bayes net) (Bouckaert, 2005) เป็นวิธีการเรียนรู้ที่ลดข้อจำกัดของการเรียนรู้เบย์อย่างง่าย ในสมมติฐานของความไม่ขึ้นต่อกันระหว่างคุณลักษณะในวิธีการเรียนรู้เบย์อย่างง่าย ในสมมติฐานของความไม่ขึ้นต่อกันซึ่งในความเป็นจริงพบว่าคุณลักษณะบางตัวจะขึ้นต่อกันบ้างและควรที่จะนำความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดลด้วยจึงใช้ข่ายงานความเชื่อเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปรปริบทของข่ายงานความเชื่อเบย์นิยมใช้คำว่า “ตัวแปร” (variable) แทนคำว่า “คุณลักษณะ” เพื่อให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่ายงานความเชื่อเบย์ให้อยู่ในรูปของโครงสร้างข่ายงานและตารางความน่าจะเป็นมีเงื่อนไข ซึ่งถือว่าเป็นข้อดีของข่ายงานแบบเบย์ ข่ายงานเบย์มีลักษณะที่สำคัญคือสามารถที่จะอธิบายความสัมพันธ์ระหว่างตัวแปรโดยแสดงออกมาในรูปแบบแผนภาพ (graphical model) โดยอาศัยฐานความรู้ก่อนหน้า (prior knowledge) ในการที่จะอธิบายและสร้างข่ายงานเบย์

2. การเลือกคุณลักษณะ (feature selection)

การวิเคราะห์ข้อมูลที่มีมิติจำนวนมากนั้น ถือเป็นเรื่องยากและมีความซับซ้อนสูง ขั้นตอนวิธีสำหรับการเรียนรู้ที่จะนำมาใช้จัดการกับข้อมูลเหล่านี้จะต้องใช้ทรัพยากรทั้งในด้านการคำนวณและการใช้หน่วยความจำจำนวนมาก ในขณะที่ประสิทธิภาพของกระบวนการเรียนรู้อาจลดลงเนื่องจากอาจมีสัญญาณรบกวน (noise) ในข้อมูลที่เกิดจากมิติที่ไม่เกี่ยวข้องกับการวิเคราะห์ข้อมูลนั้น ๆ ดังนั้น เพื่อแก้ปัญหาดังกล่าว มิติของข้อมูลเหล่านี้จะต้องถูกลดจำนวนลงด้วยวิธีการเลือกคุณลักษณะที่สำคัญสำหรับการวิเคราะห์ข้อมูล

การเลือกคุณลักษณะอาจแบ่งประเภทตามวิธีการในการเลือกคุณลักษณะ เป็น 3 ประเภท ได้แก่ Embedded, Wrapper และ Filter (Molina et al., 2002) ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 การเลือกคุณลักษณะแบบ Embedded

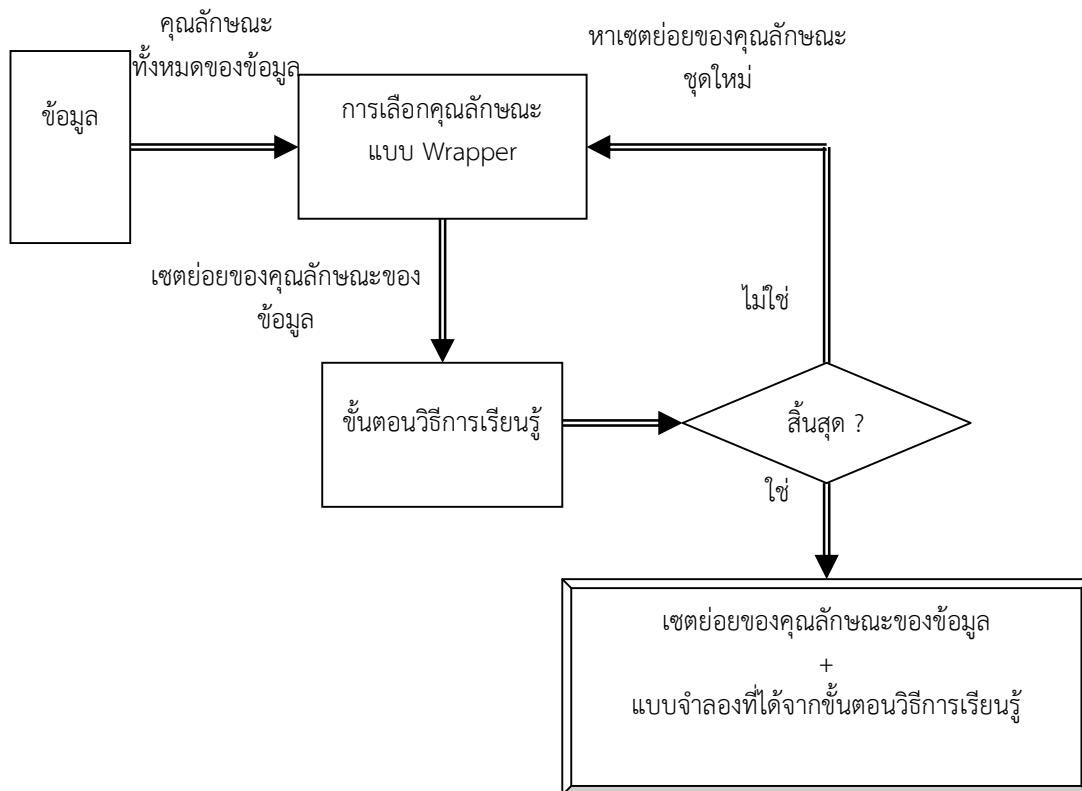
การเลือกคุณลักษณะแบบ Embedded จะเกิดขึ้นโดยขั้นตอนวิธีสำหรับการเรียนรู้เอง ซึ่งขั้นตอนวิธีเหล่านี้ จะมีการเลือกคุณลักษณะที่เหมาะสมสำหรับการสร้างแบบจำลองในการแก้ปัญหาต่าง ๆ โดยไม่จำเป็นต้องเพิ่มขั้นตอนวิธี ในการเลือกคุณลักษณะอื่น ๆ เข้ามาช่วย ตัวอย่างของขั้นตอนวิธีสำหรับการเรียนรู้ที่มีการเลือกคุณลักษณะแบบ Embedded ได้แก่ ต้นไม้ตัดสินใจ และกำหนดการพันธุกรรม

2.2 การเลือกคุณลักษณะแบบ Wrapper

การเลือกคุณลักษณะแบบ Wrapper จะเป็นขั้นตอนกระบวนการหนึ่งสำหรับคัดเลือกเซตย่อยของคุณลักษณะ ทั้งหมดของข้อมูล โดยจะเน้นที่กระบวนการค้นหาเซตย่อยของคุณลักษณะที่เหมาะสมกับขั้นตอนวิธีการเรียนรู้วิธีใดวิธีหนึ่ง โดยเฉพาะ ดังนั้นวิธีการเลือกคุณลักษณะแบบนี้ ถือได้ว่าเป็นการเพิ่มประสิทธิภาพของขั้นตอนวิธีการเรียนรู้ได้ดีที่สุด แต่ข้อเสียที่สำคัญสำหรับวิธีการนี้ คือ จะต้องใช้เวลาในการเรียนรู้นาน และเซตย่อยของคุณลักษณะที่เลือกได้ จะเหมาะกับ วิธีการเรียนรู้แบบหนึ่ง ซึ่งอาจไม่เหมาะกับวิธีการเรียนรู้แบบอื่น ๆ และปัญหาที่สำคัญอีกประการหนึ่งคือ อาจเกิดปัญหา overfitting ได้ (Kohavi & Sommerfield, 1995) ตัวอย่างของวิธีการเลือกคุณลักษณะแบบ Wrapper เช่น การใช้ขั้นตอนวิธี พันธุกรรม (genetic algorithm) เพื่อค้นหาเซตของคุณลักษณะที่เหมาะสมกับขั้นตอนวิธีการเรียนรู้ใด ๆ (Vafaie & De Jong, 1992) ภาพรวมของการเลือกคุณลักษณะแบบ Wrapper แสดงดังรูปภาพที่ 3

2.3 การเลือกคุณลักษณะแบบ Filter

การเลือกคุณสมบัตินี้แบบ Filter จะเป็นการประเมินประสิทธิภาพของคุณลักษณะของข้อมูลแต่ละตัวว่ามีความ เหมาะสมกับการวิเคราะห์ข้อมูลมากน้อยเพียงใด โดยไม่ขึ้นกับขั้นตอนวิธีของการเรียนรู้แบบใดแบบหนึ่ง การเลือกคุณลักษณะ แบบนี้จะทำการจัดลำดับ (ranking) ตามความสำคัญของคุณลักษณะแต่ละตัว และเลือกคุณลักษณะที่มีระดับความสำคัญ สูงสุดตามจำนวนที่ผู้ใช้ระบุ หรืออาจระบุเป็นค่าขีดแบ่ง (threshold) ของคุณลักษณะที่จะเลือกก็ได้ ข้อดีของการเลือก คุณลักษณะแบบนี้ คือ การประมวลผลที่รวดเร็ว และไม่ขึ้นกับขั้นตอนวิธีการเรียนรู้แบบใด ๆ แต่ข้อเสียที่สำคัญก็คือ จำนวน คุณลักษณะ หรือ ค่าขีดแบ่งที่เหมาะสมนั้น ไม่สามารถทราบได้



รูปภาพที่ 3 ภาพรวมของการเลือกคุณลักษณะแบบ Wrapper

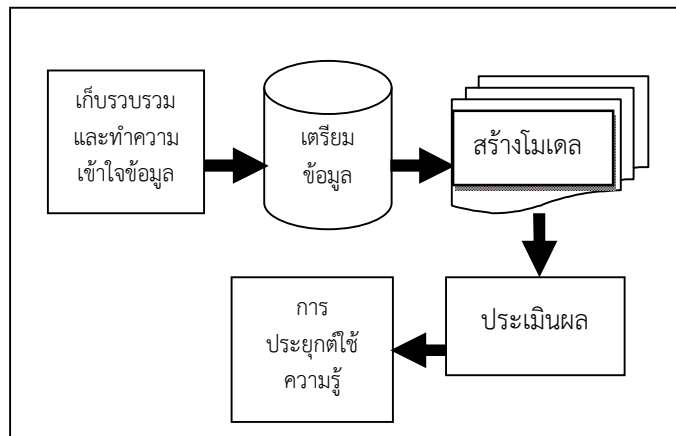
การเลือกคุณลักษณะแบบ Filter ที่เป็นที่นิยมใช้ คือ อัตราส่วนสัญญาณต่อสัญญาณรบกวน (signal-to-noise ratio: SNR) วิธีการทางสถิติเพื่อวัดประสิทธิภาพของคุณลักษณะในการจำแนกประเภทข้อมูลจากข้อมูลกลุ่มหนึ่งออกจากข้อมูลกลุ่มอื่น ๆ (Molina et al., 2002) การคำนวณหาค่า SNR แสดงดังสมการ

$$SNR_F = \left| \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2} \right| \tag{3}$$

โดยที่ μ_1 และ μ_2 คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ 1 และ กลุ่มที่ 2
 σ_1 และ σ_2 คือ ค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลในแต่ละกลุ่ม

วิธีการดำเนินการวิจัย

การดำเนินการวิจัยในที่นี้ ประกอบด้วย การเก็บรวบรวมและทำความเข้าใจข้อมูล การจัดเตรียมและวิเคราะห์ข้อมูล การสร้างโมเดลคุณลักษณะของนักศึกษาต่อผลสัมฤทธิ์ทางการเรียน การประเมินผล และการประยุกต์ใช้ความรู้ แสดงภาพรวมดังรูปภาพที่ 4 โดยมีรายละเอียดดังต่อไปนี้



รูปภาพที่ 4 ภาพรวมขั้นตอนในการวิจัย

1. ข้อมูลที่ใช้ในการวิจัย

ในการวิจัยนี้ได้ทำการเก็บรวบรวมข้อมูลจากนักศึกษาระดับชั้นปีที่ 3 และ ปีที่ 4 กลุ่มโปรแกรมวิชาคอมพิวเตอร์ ซึ่งประกอบด้วย 5 สาขาวิชา ได้แก่ สาขาวิชาวิทยาการคอมพิวเตอร์ สาขาวิชาเทคโนโลยีสารสนเทศ สาขาวิชาเทคโนโลยีคอมพิวเตอร์ สาขาวิชาวิศวกรรมซอฟต์แวร์ และ สาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ จำนวนทั้งสิ้น 144 คน จาก 65 คุณลักษณะ ซึ่งประกอบด้วย 1) ข้อมูลพื้นฐานทั่วไป ได้แก่ เพศ วุฒิการศึกษา ก่อนเข้าศึกษาในระดับปริญญาตรี ข้อมูลเกี่ยวกับครอบครัว และข้อมูลด้านสุขภาพ และ 2) ข้อมูลเกี่ยวกับเจตคติต่อรายวิชาต่าง ๆ ก่อนก่อนเข้าศึกษาต่อในระดับปริญญาตรี ได้แก่ วิชาศิลปะ สังคม วิทยาศาสตร์ ภาษาอังกฤษ การงานอาชีพและเทคโนโลยี และ คณิตศาสตร์

จากข้อมูลทั้ง 144 คน จะถูกแบ่งกลุ่มออกเป็น 3 กลุ่ม ตามค่าคะแนนเฉลี่ย ดังนี้

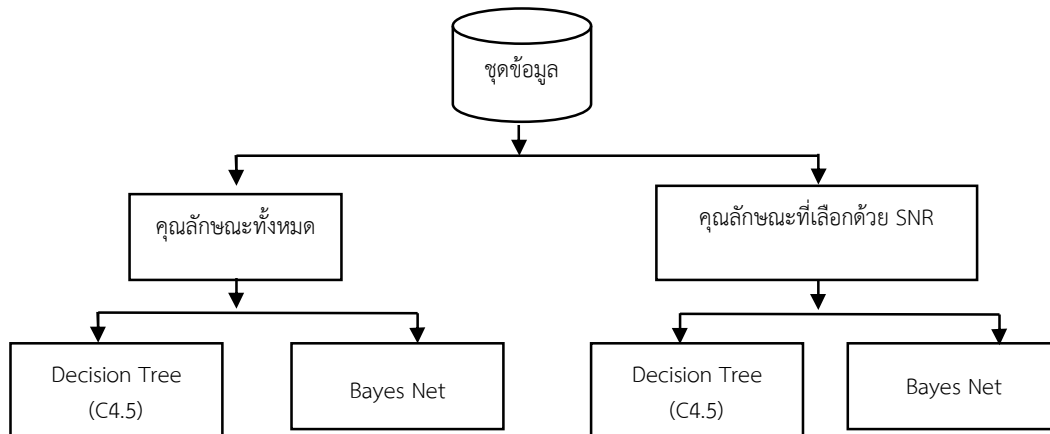
กลุ่มที่ 1 ผลการเรียนดี	ช่วงคะแนนเฉลี่ยระหว่าง 3.01 – 4.00	มีจำนวน 29 คน
กลุ่มที่ 2 ผลการเรียนปานกลาง	ช่วงคะแนนเฉลี่ยระหว่าง 2.51 – 3.00	มีจำนวน 38 คน
กลุ่มที่ 3 ผลการเรียนต่ำ	ช่วงคะแนนเฉลี่ยระหว่าง 0.00 – 2.50	มีจำนวน 77 คน

2. การออกแบบการทดลอง

การออกแบบการทดลองเพื่อสร้างตัวแบบสำหรับการจำแนกประเภทข้อมูลนั้น มีขั้นตอนในการทำงานดังแสดงไว้ในรูปภาพที่ 5 การทดลองดังกล่าวได้ทำการทดลองกับวิธีการ 2 วิธี ได้แก่ ต้นไม้ตัดสินใจ C4.5 และวิธีการข่ายงานความเชื่อเบย์ โดยทำการทดลองกับคุณลักษณะทั้งหมดและการเลือกคุณลักษณะด้วยวิธี SNR

ในการทดลองการเลือกคุณลักษณะด้วย SNR ในขั้นนี้ ได้มีการคัดเลือกคุณลักษณะจากแอททริบิวต์ทั้งหมด 65 แอททริบิวต์ ซึ่งทำการคัดเลือกคุณลักษณะก่อนนำมาสร้างต้นไม้ตัดสินใจ แอททริบิวต์ที่ได้มีจำนวน 22 แอททริบิวต์ ได้แก่ วุฒิการศึกษาของบิดา วุฒิการศึกษาของมารดา ความรับผิดชอบต่อนหน้าที่ได้รับมอบหมาย เจตคติเกี่ยวกับความสุขในการเรียนวิชาศิลปะ เจตคติเกี่ยวกับวิชาศิลปะต่อการดำรงชีวิตหรือการศึกษา เจตคติเกี่ยวกับความชอบในการเรียนวิชาศิลปะ เจตคติเกี่ยวกับความสุขในการเรียนวิชาสังคม เจตคติเกี่ยวกับความชอบในการเรียนวิชาสังคม เจตคติเกี่ยวกับความชอบในการเรียนวิชาวิทยาศาสตร์ ระดับผลการเรียนวิชาวิทยาศาสตร์ เจตคติเกี่ยวกับความสามารถในการทำความเข้าใจในรายวิชาภาษาอังกฤษ เจตคติเกี่ยวกับความจำเป็นสำหรับวิชาภาษาอังกฤษต่อการดำรงชีวิตหรือการศึกษาต่อ เจตคติเกี่ยวกับความชอบในการเรียนภาษาอังกฤษ ระดับผลการเรียนภาษาอังกฤษ เจตคติเกี่ยวกับความสุขในการเรียนวิชาการงานวิชาชีพและเทคโนโลยี เจตคติเกี่ยวกับความสามารถในการประยุกต์ใช้วิชาการงานวิชาชีพและเทคโนโลยีในชีวิตประจำวัน เจตคติเกี่ยวกับความสามารถในการทำความเข้าใจวิชาการงานวิชาชีพและเทคโนโลยีในชีวิตประจำวัน เจตคติเกี่ยวกับความจำเป็นสำหรับวิชาการงานวิชาชีพและเทคโนโลยีต่อการดำรงชีวิตหรือการศึกษาต่อ เจตคติเกี่ยวกับความชอบในการเรียนในรายวิชาการงานวิชาชีพและเทคโนโลยี ระดับผลการเรียนรายวิชาชีพและเทคโนโลยี ระดับผลการเรียนรายวิชาคณิตศาสตร์ เจตคติเกี่ยวกับความชอบในการเรียนในรายวิชาสุขศึกษา และสุดท้ายคือระดับผลการเรียนรายวิชาสุขศึกษา

หลังจากนั้นทำการทดสอบประสิทธิภาพด้วยวิธีการ 10 Fold cross validation ซึ่งในการทดลองนี้ใช้โปรแกรม Weka เป็นเครื่องมือในการทดลอง และสุดท้ายได้นำผลของตัวแบบที่ให้ผลดีที่สุดมาใช้สำหรับเป็นต้นแบบในการนำไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจต่อไป



รูปภาพที่ 5 การออกแบบการทดลอง

ผลการดำเนินงาน

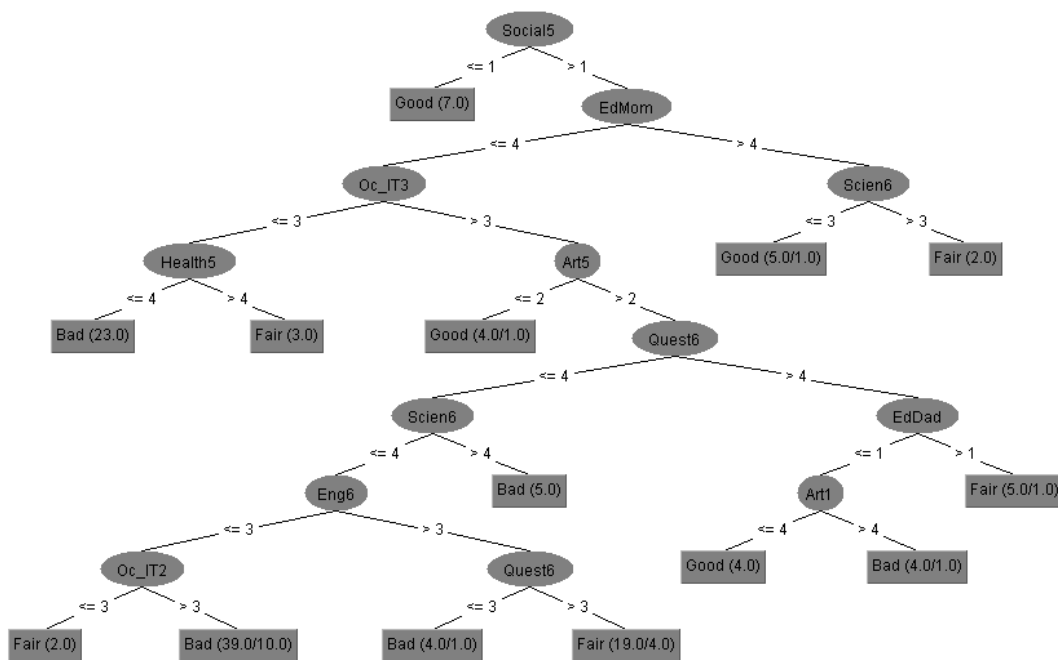
ผลการทดสอบประสิทธิภาพของการจำแนกข้อมูลด้วยวิธีการเครือข่ายความเชื่อเบย์เปรียบเทียบกับต้นไม้ตัดสินใจ พบว่า ต้นไม้ตัดสินใจ ให้ค่าความถูกต้อง (accuracy) ในการจำแนกประเภทข้อมูลได้ดีกว่าโดยให้ค่าความถูกต้องที่ 66.43 ส่วนวิธีการเครือข่ายความเชื่อเบย์ให้ค่าความถูกต้องที่ 55.71

เพื่อยืนยันประสิทธิภาพของการจำแนกข้อมูลด้วยวิธีการเครือข่ายความเชื่อเบย์กับต้นไม้ตัดสินใจ จึงได้มีการทดลองด้วยการเลือกคุณลักษณะด้วยวิธี SNR สำหรับวิธีการเครือข่ายความเชื่อเบย์กับต้นไม้ตัดสินใจ ผลการทดลองที่ได้ยังคงยืนยันด้วยค่าความถูกต้องในการจำแนกข้อมูลว่าวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องที่ 67.85 ซึ่งมากกว่าวิธีการเครือข่ายความเชื่อเบย์ที่ให้ค่าความถูกต้องที่น้อยกว่าคือ 57.14 ดังแสดงในตารางที่ 1

ตารางที่ 1 เปรียบเทียบประสิทธิภาพของการทำงานของตัวแบบ

Algorithm	Accuracy	Precision	Recall	F-Measure
BayesNet	55.71	0.400	0.557	0.461
DecisionTree(C4.5)	66.43	0.669	0.664	0.666
SNR-BayesNet	57.14	0.509	0.571	0.477
SNR-DecisionTree (C4.5)	67.85*	0.667*	0.679*	0.718*

จากตารางที่ 1 พบว่า การใช้เทคนิคต้นไม้ตัดสินใจแบบ C4.5 ด้วยการเลือกคุณลักษณะแบบ SNR นั้น ให้ประสิทธิภาพในการทำงานได้ดีที่สุด ซึ่งสามารถแสดงในลักษณะต้นไม้ตัดสินใจดังในรูปภาพที่ 6



รูปภาพที่ 6 ตัวแบบต้นไม้ตัดสินใจที่ได้จากการทดลองด้วยเทคนิคต้นไม้ตัดสินใจ (C4.5)

จากตัวแบบที่ได้ในรูปภาพที่ 6 สามารถสรุปเป็นกฎสำหรับการตัดสินใจได้เป็น 14 ข้อ แสดงดังรูปภาพที่ 7

- 1) IF **Social5** <= 1 Then "Good"
- 2) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 <= 3 And Health5 <= 4 Then "Bad"
- 3) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 <= 3 And Health5 > 4 Then "Fair"
- 4) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 **And Art5 <= 2 Then "Good"**
- 5) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 <= 4 And Scien6 <= 4 And Eng6 <= 3 And Oc_IT2 <= 3 Then "Fair"
- 6) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 <= 4 And Scien6 <= 4 And Eng6 <= 3 And Oc_IT2 > 3 Then "Bad"
- 7) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 <= 4 And Scien6 <= 4 And Eng6 > 3 And Quest6 <= 3 Then "Bad"
- 8) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 <= 4 And Scien6 <= 4 And Eng6 > 3 And Quest6 > 3 Then "Fair"
- 9) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 <= 4 And Scien6 > 4 Then "Bad"
- 10) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 > 4 And EdDad <= 1 And Art1 <= 4 Then "Good"
- 11) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 > 4 And EdDad <= 1 And Art1 > 4 Then "Bad"
- 12) IF Social5 > 1 And EdMom <= 4 And Oc_IT3 > 3 And Art5 > 2 And Quest6 > 4 And EdDad > 1 Then "Fair"
- 13) IF Social5 > 1 And EdMom > 4 And Scien6 <= 3 Then "Good"
- 14) IF Social5 > 1 And EdMom > 4 And Scien6 > 3 Then "Fair"

รูปภาพที่ 7 กฎการตัดสินใจที่ได้จากต้นไม้ตัดสินใจ

ในรูปภาพที่ 7 จากการวิเคราะห์ทัศนคติเบื้องต้น พบว่า ผู้ที่เรียนในสาขาวิชาทางคอมพิวเตอร์ได้ตั้งนั้น จะมีทัศนคติที่ไม่ดีต่อรายวิชาสังคมศึกษา และวิชาศิลปะ เมื่อได้ถูกตั้งคำถามแล้ว จึงได้นำไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจสำหรับการเลือกศึกษาต่อระดับอุดมศึกษาในสาขาวิชาของกลุ่มโปรแกรมวิชาคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม แสดงดังรูปภาพที่ 8 ซึ่งผลการทดลองการใช้งานระบบ ผู้ใช้งานมีความพึงพอใจอยู่ในระดับดี



รูปภาพที่ 8 หน้าจอระบบสนับสนุนการตัดสินใจเลือกเรียนในสาขาวิชาคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม

สรุปผลการทดลอง

ในงานวิจัยนี้ได้ทำการทดลองสร้างตัวแบบสำหรับระบบสนับสนุนการตัดสินใจเลือกศึกษาในระดับปริญญาตรี กลุ่มโปรแกรมคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม ซึ่งผู้วิจัยได้ทำการศึกษาโดยใช้ข้อมูลจากนักศึกษา กลุ่มโปรแกรมวิชาคอมพิวเตอร์ ระดับชั้นปี 3 และ ปี 4 ปีการศึกษา 2556 จำนวน 144 คน จาก 65 คุณลักษณะ โดยเน้นที่ทัศนคติของผู้เรียนก่อนเข้าศึกษาต่อในระดับปริญญาตรี

จากผลการทดลองทำให้ทราบถึงคุณลักษณะสำคัญที่มีผลต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม มีจำนวน 22 แอททริบิวท์ ประกอบด้วย วุฒิการศึกษาของบิดา วุฒิการศึกษาของมารดา ความรับผิดชอบต่อนหน้าที่ที่ได้รับมอบหมาย เจตคติเกี่ยวกับความสุขในการเรียนวิชาศิลปะ เจตคติเกี่ยวกับวิชาศิลปะต่อการดำรงชีวิตหรือการศึกษา เจตคติเกี่ยวกับความชอบในการเรียนวิชาศิลปะ เจตคติเกี่ยวกับความสุขในการเรียนวิชาสังคม เจตคติเกี่ยวกับความชอบในการเรียนวิชาสังคม เจตคติเกี่ยวกับความชอบในการเรียนวิชาวิทยาศาสตร์ ระดับผลการเรียนวิชาวิทยาศาสตร์ เจตคติเกี่ยวกับความสามารถในการทำความเข้าใจในรายวิชาภาษาอังกฤษ เจตคติเกี่ยวกับความจำเป็นสำหรับวิชาภาษาอังกฤษต่อการดำรงชีวิตหรือการศึกษาต่อ เจตคติเกี่ยวกับความชอบในการเรียนภาษาอังกฤษ ระดับผลการเรียนภาษาอังกฤษ เจตคติเกี่ยวกับความสุขในการเรียนวิชาการงานวิชาชีพและเทคโนโลยี เจตคติเกี่ยวกับความสามารถในการประยุกต์ใช้วิชาการงานวิชาชีพและเทคโนโลยีในชีวิตประจำวัน เจตคติเกี่ยวกับความสามารถในการทำความเข้าใจวิชาการงานวิชาชีพและเทคโนโลยีในชีวิตประจำวัน เจตคติเกี่ยวกับความจำเป็นสำหรับวิชาการงานวิชาชีพและเทคโนโลยีต่อการดำรงชีวิตหรือการศึกษาต่อ เจตคติเกี่ยวกับความชอบในการเรียนในรายวิชาการงานวิชาชีพและ

เทคโนโลยี ระดับผลการเรียนรายวิชาซีพีและเทคโนโลยี ระดับผลการเรียนรายวิชาคณิตศาสตร์ เจตคติเกี่ยวกับความชอบในการเรียนในรายวิชาสุศึกษา และสุดท้าย คือ ระดับผลการเรียนรายวิชาสุศึกษา

โดยผลการทดลอง พบว่า วิธีการต้นไม้ตัดสินใจให้ประสิทธิภาพในการจำแนกประเภทได้ดีเมื่อเปรียบเทียบกับตัวแบบที่สร้างด้วยวิธีการเครือข่ายความเชื่อเบย์ เมื่อนำตัวแบบดังกล่าวมาพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกเรียนต่อในสาขาวิชาคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม และทำการประเมินโดยผู้ใช้งาน ผลการทดลองพบว่า ผู้ใช้งานมีความพึงพอใจต่อในการใช้งานระบบอยู่ในระดับดี

ข้อเสนอแนะ

เพื่อให้การทดลองถูกต้องสมบูรณ์มากยิ่งขึ้นควรทำการทดลองกับชุดข้อมูลที่มีขนาดใหญ่ และการทดลองกับหลากหลายสาขา หรือทำการศึกษาค้นคว้าข้อมูลกับต่างสถาบันและทำการทดลองเปรียบเทียบกับเทคนิคทางเหมืองข้อมูลที่หลากหลายมากขึ้น ควรเพื่อเทคนิคสำหรับการคัดเลือกคุณลักษณะสำคัญที่ที่หลากหลายมากขึ้นเพื่อเพิ่มประสิทธิภาพของการทำงานของโมเดล ซึ่งจะทำให้การจำแนกข้อมูลที่ได้มีความถูกต้องแม่นยำมากยิ่งขึ้น

เอกสารอ้างอิง

- ชุตินา อุดมะมุณี และประสงค์ ปราณีตพลกรัง. (2553). การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. *Journal of Information Science and Technology*, 1 (2, กรกฎาคม-ธันวาคม), 39-48.
- ธีรดา ภิญญา. (2551). อิทธิพลของบริบททางสังคมต่อการตัดสินใจเลือกศึกษาต่อสาขาวิชาวิทยาศาสตร์และเทคโนโลยีของนักเรียนมัธยมศึกษาชั้นปีที่ 6. *Songklanakarin Journal of Social Sciences and Humanities*, 14 (2, เมษายน-มิถุนายน), 207-224.
- บุญเสริม กิจสิริกุล. (2546). เอกสารประกอบการสอนวิชา 2116054 ปัญญาประดิษฐ์. กรุงเทพฯ: ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ภัทรพงศ์ พงศ์ภัทรกานต์. (2553). การวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาระดับปริญญาตรี โดยใช้คอมพิวเตอร์แมชชีน. ใน *การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 6*, 491-496.
- Bouckaert, R. R. (2005). *Bayesian Network Classifiers in Weka*. Hamilton: Department of Computing Science, University of Waikato, New Zealand.
- Kohavi R. & Sommerfield D. (1995). Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology. *Proceeding of the First International Conference on Knowledge Discovery and Data Mining*, 192 - 197.
- Molina, L. C., Belanche, L. & Netbot, A. (2002). Feature Selection Algorithms: a survey an experimental evaluation. *Proceeding of the IEEE International Conference on Data Mining*, 306-313.
- Quinlan, J. Ross. (1986). Introduction of Decision Tree. *Machine Learning*, 1, 81-106.
- Vafaie H. & De Jong K. (1992). Genetic Algorithms as a Tool for Feature Selection in Machine Learning. *Proceeding of the 4th International Conference on Tools with Artificial Intelligence*, 200-203.