วิจัยเชิงสำรวจโดยใช้เทคนิคเหมืองข้อมูล เพื่อค้นหาแผนการเรียนที่เหมาะสมสำหรับการศึกษาต่อระดับมัธยมศึกษาตอนปลาย

ศรารัตน์ วรรณแจ่ม (Sararat Wannajam), ศศิวิมล ตาลเพชร (Sasiwimon Tanphet)
ปิยะพร คมคาย (Piyaporn Komkay), วัชราภรณ์ ศรียานงค์ (Watcharaporn Sriyanong)
สาขาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
sararat.w@kkumail.com, sasiwimon_t@kkumail.com, watcharaporns@kkumail.com, piyapornkk@kkumail.com

บทคัดย่อ

การวิจัยนี้มีจุดประสงค์ เพื่อสร้างตัวแบบจำลองพยากรณ์การเลือกแผนการเรียนที่เหมาะสมในระดับมัธยมศึกษาตอนปลาย และทำการทดลองด้วยโปรแกรม WEKA ในการสร้างแบบจำลองโดยใช้อัลกอริทึม ต้นไม้ตัดสินใจ, ข่ายงานเบย์ และโครงข่ายประสาทเทียม เพื่อหาประสิทธิภาพของแต่ละเทคนิค โดยจะนำผลจากการทดลองที่ได้มาเปรียบเทียบประสิทธิภาพในแต่ละอัลกอริทึม จากตัวแบบดังกล่าวทำให้ทราบว่าตัวแปรที่มีผลต่อการแผนการเรียนในระดับมัธยมศึกษาตอนปลาย ได้แก่ ผลการเรียนเฉลี่ย และเกรดเฉลี่ยแต่ละรายวิชา ผลการวิจัยพบว่าแบบจำลองที่ให้ค่าความถูกต้องมากที่สุดคือ วิธีการต้นไม้ตัดสินใจ C 4.5 ร้อยละ 89.12 ซึ่งสรุปได้ว่ามีตัวแปรที่ได้จากการพัฒนาตัวแบบด้วย.วิธีการต้นไม้ตัดสินใจ C 4.5 มีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้

คำสำคัญ: ต้นไม้ตัดสินใจ, ข่ายงานเบย์, โครงข่ายประสาทเทียม

1. บทนำ

การเรียนรู้ช่วงที่สำคัญช่วงนึง คือช่วงมัธยมศึกษา โดยการศึกษาในขั้นนี้มุ่งเน้นให้ผู้เรียนมีความรู้ทางด้านวิชาการและวิชาชีพให้เหมาะสมกับวัย ความต้องการ ความสนใจ เพื่อให้ผู้เรียนเกิดความเข้าใจและรู้จักเลือกอาชีพที่เป็นประโยชน์แก่ตนเองและสังคม โดยผู้เรียนจบในระดับมัธยมศึกษานี้สามารถไปประกอบอาชีพที่เหมาะสมกับวัยและความสามารถ หรือศึกษาต่อในระดับ อุดมศึกษา ในระดับมัธยมศึกษาตอนต้นผู้เรียนเลือกเรียนกลุ่มวิชาการหรือวิชาชีพตามความถนัดของตนเอง ในระดับมัธยมศึกษาตอนปลายผู้เรียนสามารถเรียนกลุ่มวิชาที่ผู้เรียนชื่นชอบเพื่อยึดเป็นอาชีพได้[2] ดังนั้นการเลือกเรียนในสาขาวิชาที่ตนเองถนัดนั้นจึงเป็นปัญหาสำคัญที่เกิดขึ้นเพราะมีนักเรียนจำนวนไม่น้อยที่ไม่รู้ว่าตนเองชอบหรือถนัดอะไร จึงไม่สามารถเลือกสาขาวิชาที่เหมาะสมกับตนเองและเรียนได้อย่างประสบความเร็จ จนเป็นสาเหตุที่ทำให้นักเรียนขอย้ายสาขา ,ขอพักการเรียน จนกระทั่งถึงการลาออกจากการเรียน เนื่องจากการเรียนในสาขาที่ไม่ถนัด ขาดความสนใจและประสบการณ์ที่จะเรียนรู้ในบทเรียน จึงส่งผลให้สอบไม่ผ่านเกณฑ์ที่กำหนดในแต่ละรายวิชา ผลสัมฤทธิ์ทางการเรียนเฉลี่ยต่ำกว่าเกณฑ์จึงเป็นผลให้สถานศึกษาต้องประกาศพ้นสภาพผู้เรียน นักเรียนไม่สามารถเรียนจบหลักสูตรที่โรงเรียนวางไว้ โดยส่วนใหญ่นักเรียนมักเลือกสาขาโดยใช้ความรู้สึก หรือตามเพื่อนและผู้ปกครอง ซึ่งอาจไม่ได้เกิดจากความต้องการที่แท้จริงและสามารถสร้างความมั่นใจได้ว่านักเรียนจะสามารถเรียนได้ดีและประสบผลสำเร็จ

ในงานวิจัยในครั้งนี้ จึงได้ทำการเปรียบเทียบแบบจำลองแต่ละแบบจากการนำข้อมูลของนักเรียนและผลการเรียน เข้ามาทดสอบในแบบจำลองนี้ โดยใช้เทคนิคการทำเหมืองข้อมูลวิเคราะห์ข้อมูลที่มีขนาดใหญ่ เพื่อค้นหาความสัมพันธ์และรูปแบบ หรือกฎที่ซ่อนอยู่ และนำความสัมพันธ์เหล่านี้มาคำนวณเพื่อหาค่าความความถูกต้อง และค่าความคลาดเคลื่อนเพื่อนำไปสู่การได้แบบจำลองที่มีค่าความถูกต้องสูงที่สุด ซึ่งจะช่วยในการนำไปพัฒนาการระบบการพยากรณ์หรือช่วยตัดสินใจในการเลือกสาขาวิชาที่มีโอกาสประสบผลสำเร็จในการเรียน และโรงเรียนก็สามารถนำผลที่ได้ไปประกอบการจัดทำหลักสูตร หรือการแนะแนวการศึกษาเพื่อสร้างความมั่นใจและแนวทางสู่ความสำเร็จในการเรียนการสอนที่มีประสิทธิภาพต่อไป

2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

งานวิจัยฉบับนี้ผู้วิจัยได้นำอัลกอริทึม (Algorithm) ต้นไม้ตัดสินใจ (Decision Tree C4.5) อัลกอริทึมโครงข่ายประสาทเทียม (Artificial Neuron Network), และข่ายงานเบย์ (Bayesian Networks) มาทำการวิเคราะห์เชิงเปรียบเทียบเพื่อให้เห็นถึงประสิทธิภาพของอัลกอริทึมทั้ง 3 แบบดังกล่าวซึ่งทฤษฎีและกระบวนการในการทำเหมืองข้อมูลของแต่ละอัลกอริทึมแสดงรายละเอียดดังต่อไปนี้

2.1 กระบวนการทำเหมืองข้อมูล (Data Mining) [1,4]

คือ การนำเทคนิคการเรียนรู้ของเครื่อง วิธีการทางสถิติ วิธีทางปัญญาประดิษฐ์ หรือวิธีอื่น ๆ มาทำการวิเคราะห์และสกัดความรู้จากข้อมูลที่จัดเก็บไว้ในฐานข้อมูลหรือจัดเก็บไว้ในรูปอื่น ๆ โดย

จุดประสงค์ของการทำเหมืองข้อมูล คือ การวิเคราะห์แนวโน้ม ความสัมพันธ์ กฎหรือรูปแบบของข้อมูล ซึ่งเป็นความรู้ที่ถูกซ่อนอยู่ภายใต้ข้อมูลขนาดใหญ่ และนำสารสนเทศที่ได้มาช่วยในการวางแผนการตัดสินใจการบริหารหรือแก้ปัญหาด้านต่าง ๆ

ประโยชน์หลักของเหมืองข้อมูลคือ การค้นหาความรู้ที่ซ่อนอยู่ในฐานข้อมูลเพื่อให้ได้ซึ่งความรู้มาช่วยประกอบในการตัดสินใจ

2.2 ต้นไม้ตัดสินใจ (Decision Tree C4.5) [4]

ต้นไม้ตัดสินใจเป็นอัลกอริทึมที่ใช้ในการจำแนกประเภทนำเสนอโดย Ross Quinlan ใช้หลักการสร้างต้นไม้โดยคัดเลือกแอทริบิวต์ที่สำคัญที่สุดมาเป็นโหนดราก (Root Node) โดยใช้ค่า Gain Ratio ที่สูงที่สุดเป็นโหนดราก (Root Node) และโหนดถัดไปในการหาค่าGain Ratio ต้องทาการหาค่า Split Information และการ Entropy ก่อน ดังนี้

1. สมการ Entropy

$$\text{Entropy(s)} = \sum_{i=1}^{c} - P_i \ log_2 P_i \qquad (1)$$

โดย S คือ แอทริบิวต์ที่นามาวัดค่า

P_i คือ สัดส่วนนของจำนวนสมาชิกในกลุ่มเท่ากับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

2. สมการ Information Gain

$$\text{Gain } (S,A) = \text{Entropy}(S) - \sum_{v \grave{o}\ \text{Values}(A)} \frac{|S_V|}{|S|} \text{Entropy}(S_V) \qquad (2)$$

โดย A คือ แอทริบิวต์ A

$|S_V|$คือ สมาชิกของแอทริบิวต์ A ที่มีค่า $_V$

$|S|$ คือ จำนวนสมาชิกของกลุ่มตัวอย่าง

3. สมการ Split Information

$$Split\ Information\ (S,A) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \qquad (3)$$

4. สมการ Gain Ratio

$$Gain\ Ratio\ (S,A) = \frac{Gain\ (S,A)}{Split\ Information\ (S,A)} \qquad (4)$$

2.3 โครงข่ายประสาทเทียม (Artificial Neural Network) [3]

การทำงานของระบบโครงข่ายประสาทเทียมนั้นเป็นการเลียนแบบการทำงานของสมองมนุษย์ผ่านกลไกของการเรียนรู้โดยการใช้ประโยชน์จากตัวอย่างที่ผ่านมาหลายๆ ตัวอย่างในการฝึกฝนซึ่งองค์ประกอบของโครงข่ายประสาทเทียม ดังรูปที่ 1 ประกอบด้วย

ข้อมูลนำเข้า (Input) หมายถึง ข้อมูลที่ระบบจะนำเข้ามาประมวลผล

น้ำหนัก (Weight) หมายถึง ค่าเฉพาะที่กำหนดให้ข้อมูลนำเข้าแต่ละตัว เพื่อใช้แยกแยะความแตกต่างของข้อมูลนำเข้า

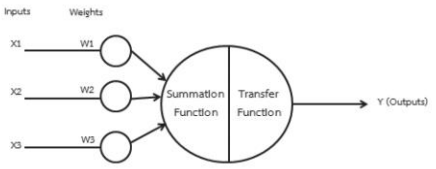
ฟังก์ชันการรวม (Summation Function) เป็นผลรวมของข้อมูลป้อนเข้า และค่าน้ำหนักในแต่ละชั้น(Layer) เพื่อใช้สำหรับสรุปผลความสัมพันธ์ระหว่างข้อมูลนำเข้าทั้งหมด

ฟังก์ชันการแปลง (Transformation Function) เป็นการคำนวณการจำลองการทำงานของโครงข่ายประสาทเทียม โดยจะนำสารสนเทศมาแปลงเพื่อนำไปใช้สำหรับการแสดงผลลัพธ์ ซึ่งฟังก์ชันในองค์ประกอบนี้เรียกว่า “Activation Function” สำหรับฟังก์ชันการแปลงแบ่งได้ 3 ชนิด ได้แก่ Linear, Threshold และ Sigmoid

ข้อมูลส่งออก (Output) คือ ผลลัพธ์ที่เกิดขึ้นจริงแสดงผลให้แก่ผู้ใช้

โครงข่ายประสาทเทียมมีความยืดหยุ่นและมีประสิทธิภาพองค์ประกอบภายในของ โครงข่ายประสาทเทียมประกอบไป

ด้วยโหนดต่างๆ เชื่อมต่อกันโดยในแต่ละโหนดคอยรับค่าข้อมูลนำเข้า และค่าน้ำหนัก เข้ามาวิเคราะห์ ดังสมการที่ (1)



รูปที่ 1 องค์ประกอบของโครงข่ายประสาทเทียม

$$n = \left(\sum_{i=1}^{p} x_i\, w_i\right) + b \qquad (1)$$

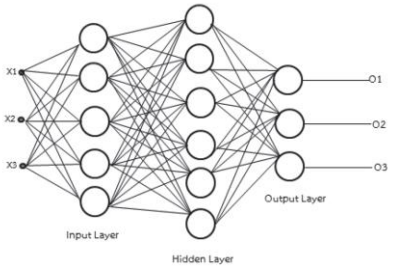
เมื่อ p เป็นจำนวนโหนดของข้อมูลนำเข้า ทั้งหมด, xi คือ ข้อมูลนำเข้าตัวที่ i, wi คือ ค่าน้ำหนักของข้อมูลนำเข้าตัวที่ i และ b คือ ค่าความโน้มเอียง (bias) ที่เพิ่มเข้าไปในผลรวมของ n โหนดนั้น จากนั้นผลรวมของข้อมูลนำเข้าที่ถูกปรับเปลี่ยนค่าน้ำหนักแล้วจะถูกประมวลผลผ่านฟังก์ชันการกระตุ้น เพื่อให้เกิดเป็นข้อมูลส่งออกและส่งไปยังโหนดในชั้นถัดไป ดังสมการที่ (2)

$$y = f(n) = f\left[\left(\sum_{i=1}^{p} x_i\, w_i\right) + b\right] \qquad (2)$$

เมื่อ y คือ ข้อมูลส่งออก n คือ ผลรวมของค่าน้ำหนักข้อมูลนำเข้า และ f คือ ฟังก์ชันการกระตุ้น มีหลายรูปแบบ เช่น linear function, threshold function และ sigmoid function สามารถเลือกได้ว่าจะใช้ฟังก์ชันอะไร ตามความเหมาะสมของงาน

เพอร์เซปตรอนหลายชั้น (Multi-Layer Perceptron)

เพอร์เซ็ปตรอนแบบหลายชั้น (Multi-Layers Perceptron) ใช้สำหรับงานที่มีความซับซ้อนได้เป็นอย่างดีประกอบด้วยชั้นข้อมูลนำเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นข้อมูลส่งออก (output layer) โดยใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) สำหรับการฝึกฝน ดังรูปที่ 2



รูปที่ 2 เพอร์เซ็ปตรอนแบบหลายชั้น

เทคนิควิธีการเรียนรู้แบบแพร่ย้อนกลับมีหลักการทำงาน คือ เป็นการปรับค่าน้ำหนักที่เส้นเชื่อมต่อระหว่างโหนดให้เหมาะสมจนกว่าค่าความผิดพลาดจะอยู่ในช่วงที่ยอมรับได้ ซึ่งกระบวนการส่งค่าย้อนกลับประกอบด้วย

การส่งผ่านไปข้างหน้า (Forward Pass) โดยข้อมูลจะถูกส่งเข้าสู่โครงข่ายประสาทเทียมที่ชั้นข้อมูลนำเข้า และจะส่งผ่านจากชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลส่งออก

ส่วนการส่งผ่านย้อนกลับ (Backward Pass) จะทำการประมวลผลค่าของข้อมูลนำเข้าก่อนทำการตรวจสอบค่าความผิดพลาด (Error) ที่จะนำไปพิจารณาเปรียบกับค่าความผิดพลาดที่ยอมรับได้เนื่องจากเพอร์เซ็ปตรอนแบบหลายชั้นจะมีชั้นซ่อน

2.4 ข่ายงานเบย์ (Bayesian Belief Networks)

ข่ายงานเบย์ (Bayesian Networks) หรือเรียกโดยย่อว่า (Bayes Net) เป็นวิธีที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปรและแสดงออกมาในรูปแบบของโมเดลโดยอาศัยฐานความรู้ก่อนหน้า (Prior Knowledge) ในการที่จะอธิบายและสร้างข่ายงานเบย์ และสามารถอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (Condition Independent) ระหว่างตัวแปรบริบทของข่ายงานความเชื่อเบย์นิยมใช้คำว่า“ตัวแปร”(Variable) แทนคำว่า “คุณสมบัติ” เพื่อทำให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยเราสามารถใส่ความ รู้ก่อนในข่ายงานความเชื่อเบย์ให้อยู่ในรูปของโครงสร้างข่ายงานและตารางความน่าจะเป็นมีเงื่อนไขโดยนิยามความไม่ขึ้นต่อกันอย่างมีเงื่อนไขได้ ดังสมการที่ 1

$$(\forall\, x_i, y_j, z_k)\ \mathrm{P}(\mathrm{X}=x_i \mid Y=y_j, \mathrm{Z}=z_k)=\mathrm{P}(\mathrm{X}=x_i : \mathrm{Z}=z_k) \qquad (1)$$

หรือแสดงในรูปอย่างง่ายได้ดังสมการที่ 2

$$P(X|Y,Z) = P\,(X|Z) \qquad (2)$$

จากสมการหมายความว่าสำหรับ xi,yj,zk ใด ๆ ความน่าจะเป็นที่ X จะมีค่าเป็น xi (X เป็นตัวแปร ส่วน xi คือค่าของตัวแปร) เมื่อรู้ว่า Y มีค่าเป็น Yj และ Z มีค่าเป็น zk จะมีค่าเท่ากับความน่าจะเป็นของ X จะมีค่าเป็น Xi เมื่อรู้ว่า Z มีค่าเป็น Zkในกรณีที่ความน่าจะเป็นทั้งสองเท่ากันเช่นนี้ เราเรียกว่าค่าของ X ไม่ขึ้นกับ Y อย่างมีเงื่อนไขเมื่อรู้ค่าของ Z เราจึงสามารถตัด Yทิ้งไปได้

เทคนิคนี้เป็นวิธีการเรียนรู้ที่ลดข้อจำกัดของการเรียนรู้เบย์อย่างง่ายในสมมติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัติ

ในความเป็นจริงเราพบว่าคุณสมบัติบางตัวจะขึ้นต่อกันบ้างและควรที่จะนำความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดลด้วย

ชุติมา อุตมะมุณีย์ และประสงค์ ประณีตพลกรัง (2553) เป็นการสร้างตัวแบบสำหรับหาปัจจัยที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษา โดยนำข้อมูลของนักศึกษาในระดับอุดมศึกษาของมหาวิทยาลัยรัฐและเอกชน จำนวน 9 มหาวิทยาลัย ใช้ทดสอบแบบจำลองจะทำการทดสอบผลบนพื้นฐานของวิธี K - fold cross validation จากนั้นได้นำผลลัพธ์ที่ดีที่สุดจากเทคนิคข่ายงานเบย์มาสร้างตัวแบบในขั้นสุดท้ายและได้เปรียบเทียบความถูกต้องของตัวแบบกับผลการวิเคราะห์ความถดถอยเชิงพหุคูณ ผลการเปรียบเทียบ การเรียนรู้แบบเบย์ได้ตัวแปรที่มีผลต่อการตัดสินใจแม่นยำสูง 91.35% [4]

ณัฏฐา ผิวมา (2558) ได้จำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษา สาขาคอมพิวเตอร์โดยใช้โครงข่ายประสาทเทียม โดยนำข้อมูลจำนวน 100 คน มีผลการเรียนเฉลี่ยในกลุ่มสาระการเรียนรู้ และคะแนนจากการทำแบบทดสอบบุคลิกภาพที่สอดคล้องกับคุณลักษณะบัณฑิตของสาขาคอมพิวเตอร์มาใช้ในการสร้างแบบจำลองการพยากรณ์ เมื่อนำมาใช้ในการทดลองนี้ทำให้ได้ค่าความถูกต้องอยู่ในเกณฑ์ดี คือ ร้อยละ 75.63 [5]

พรรณิภา บุตรเอก, สุรเดช บุญลือ (2557) ได้ทำการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ข้อมูลนักศึกษาระดับปริญญาตรีหลักสูตร 4 ปี สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยนอร์ทกรุงเทพ ปีการศึกษา 2547-2551 จำนวน 138 คน โดยใช้เทคนิคเหมืองข้อมูล ซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองกับแบบพยากรณ์ที่ใช้เทคนิค SVM-PK ให้ความแม่นยำในการพยากรณ์สูงที่สุดที่ร้อยละ 89.13 [6]

เยาวภา ภารสำเร็จ และคณะ (2556) ได้นำเสนอการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลของอัลกอริทึมเหมืองข้อมูล 3 แบบคือ C4.5, Naïve Bayesและ K-Nearest Neighbor ซึ่งอัลกอริทึม C4.5 ในการเปรียบเทียบข้อมูลนักศึกษา จำนวน 4,591 ชุด พบว่าให้ประสิทธิภาพในการจำแนกข้อมูลดีที่สุดคือ ร้อยละ 73.55 [7]

สุธีรา วงศ์อนันทรัพย์ (2559) ได้นำผลการทำแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัดของกลุ่มตัวอย่างมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูลผลการวิเคราะห์แบบประเมินความรู้ก่อนเรียนผลลัพธ์ที่ได้สามารถจำแนกนักศึกษา โดยใช้การวิเคราะห์ผลการทดสอบประเมินความรู้ก่อนเรียน ใช้เทคนิคการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-mean และเทคนิค Naïve Bayes จากการทำนายของชุดข้อมูลสอนให้ร้อยละความถูกต้องเฉลี่ย 79.42 [8]

สุพัฒน์กุล ภัคโชค (2556) การวิจัยนี้มีความมุ่งหมายเพื่อค้นหาปัจจัยที่มีผลต่อการเลือกแผนการเรียน และความสามารถในการศึกษาในแผนการเรียนนั้นได้อย่างประสบความสำเร็จด้วยเทคนิคเหมืองข้อมูล โดยมีการเปรียบเทียบประสิทธิภาพของอัลกอริทึม Decision Rule : Partial Rules กับ ทฤษฎีต้นไม้ตัดสินใจ C 4.5 ผลการวิจัยพบว่าการจำแนกข้อมูลด้วยอัลกอริทึม Decision Rule Partial Rules มีประสิทธิภาพสูงที่สุดให้ค่าความถูกต้องร้อยละ 79.76 [9]

เสกสรรค์ วิลัยลักษณ์ (2558) ได้นำเทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน จำนวน 525 คน ใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่มีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ ค่าความถูกต้องสูงที่สุดที่ ร้อยละ 94.48

อนันต์ ปินะเต และคณะ (2557) ใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมในระดับปริญญาตรี เพื่อการประชาสัมพันธ์ข้อมูลการสมัครให้ผู้สมัครได้เลือกกลุ่มสาขาวิชาได้ตรงตามทักษะ ความรู้ความสามารถของผู้สมัคร การวิจัยพบว่าประสิทธิภาพของการจำแนกข้อมูลแบบวิธีต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุดมีค่าความถูกต้องสูงสุด จาก

ข้อมูลการทดลอง จำนวน 18,879 คน ตามกลุ่มประกันคุณภาพการศึกษาของมหาวิทยาลัย กลุ่ม 1มนุษยศาสตร์และสังคมศาสตร์ ร้อยละ 82.22 กลุ่ม 2วิทยาศาสตร์เทคโนโลยี ร้อยละ 88.26 และกลุ่ม 3 วิทยาศาสตร์สุขภาพ ร้อยละ 87.73 [11]

จากการศึกษางานวิจัยข้างต้นนำมาเปรียบเทียบผลลัพธ์ได้ดังตารางที่ 1

ตารางที่ 1 ตารางเปรียบเทียบเทคนิค และผลการทดลอง

บทความ	เทคนิค	ผลการทดลอง (ร้อยละ)
[4]	Naïve Bay	91.35
[5]	Neural Network	75.63
[6]	SVM-PK	89.13
[7]	Decision Tree C4.5	73.55
[8]	K-mean+Naïve bayes	79.42
[9]	Decision Rule : Partial Rules	79.76
[10]	Multi-Layer Perceptron	94.48
[11]	Decision Tree C4.5	กลุ่ม 1 ร้อยละ 82.22 กลุ่ม 2 ร้อยละ 88.26 กลุ่ม 3 ร้อยละ 87.73

จากตารางที่ 1 สำรวจการใช้เทคนิคเหมืองข้อมูลพบว่ามีเทคนิคทั้งหมด 8 เทคนิค คือเนื่องจากผู้วิจัยหลายท่านได้เลือกใช้เทคนิคต้นไม้ตัดสินใจ เทคนิคข่ายงานเบย์ โครงข่ายประสาทเทียมซึ่งได้ค่าความถูกต้องค่อนข้างสูง

ดังนั้นผู้วิจัยจึงได้เลือกใช้เทคนิคดังนี้ ต้นไม้ตัดสินใจ เทคนิคข่ายงานเบย์ โครงข่ายประสาทเทียมมาทำการเปรียบเทียบและเลือกเทคนิคที่ให้ค่าความถูกต้องที่ดีที่สุดเพื่อนำไปเพื่อสร้างตัวแบบจำลองพยากรณ์การเลือกแผนการเรียนที่เหมาะสมในระดับมัธยมศึกษาตอนปลาย

ตารางที่ 2 ผลการทดสอบแบบจำลองพยากรณ์

Algorithm	Accuracy(%)	Precision	Recall	F-Measure
Decision Tree	89.12 %	0.894	0.891	0.891
Neural Network	88.22 %	0.883	0.882	0.882
Naïve Bay	69.21 %	0.769	0.692	0.703

3. ผลการดำเนินงาน

จากตารางเปรียบเทียบโมเดล C4.5 (J48) , Naïve Bayes และ Artificial Neural Network พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้นจากการ ใช้อัลกอริธึม C4.5 (J48) มีค่าความถูกต้องเท่ากับ 89.12 % ค่าที่สามารถสืบค้นคำตอบสูงสุด (Precision) มีค่าเท่ากับ 0.894 ค่าที่ได้จากการตรวจพบข้อมูล หรือค่าการเรียกคืน (Recall) มีค่าเท่ากับ 0.891 ค่าความถูกต้องโดยรวม (F-measure) มีค่าเท่ากับ 0.891 มากกว่าโมเดลที่ถูกสร้างขึ้นจากการใช้อัลกอริธึม Artificial Neural Network ซึ่งมีค่าความถูกต้องเท่ากับ 88.22 % ลองลงมามาจะเป็นโมเดลที่ถูกสร้างขึ้นจากการใช้อัลกอริทึม Naïve bayes ซึ่งมีค่าความถูกต้องเท่ากับ 69.21 % ด้วยเหตุนี้จึงนำเทคนิคต้นไม้ตัดสินใจ (Decision Tree) มาใช้ในการวิจัยในครั้งนี้เพื่อแสดงรูปแบบของแบบจำลองสำหรับพยากรณ์การเลือกแผนการเรียนที่เหมาะสมสำหรับการศึกษาต่อระดับมัธยมศึกษาตอนปลาย

4. สรุป

ในการจัดทำวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบโมเดลที่มีความเหมาะสมและถูกต้อง สำหรับพยากรณ์แผนการเรียนที่เหมาะสมสำหรับการศึกษาต่อระดับมัธยมศึกษาตอนปลาย ซึ่งใช้วิธีการหาอัลกอริทึมที่เหมาะสมโดยการเปรียบเทียบระหว่างอัลกอริทึมของต้นไม้ตัดสินใจ (Decision Tree 4.5),โครงข่ายประสาทเทียม (Artificial Neural Network) , ข่ายงานเบย์ (Bayesian Networks) เนื่องจากเป็นอัลกอริทึมที่นิยมนำมาใช้ในการจัดหมวดหมู่ของข้อมูลที่ทราบผลลัพธ์แน่นอน

ผลจากการวิเคราะห์ด้วยโปรแกรม WEKA พบว่า ปัจจัยที่มีผลต่อการเลือกสาขาวิชาเรียนทั้งหมด 9 ตัวแปรหลัก ประกอบด้วย 1. เกรดเฉลี่ยตามแผนการเรียนในระดับมัธยมศึกษาตอนปลาย 2. เกรดเฉลี่ยม.ตอนต้น วิชาภาษาไทย 3. เกรดเฉลี่ยม.ตอนต้น วิชาคณิตศาสตร์ 4. เกรดเฉลี่ยม.ตอนต้น วิชาวิทยาศาสตร์ 5. เกรดเฉลี่ยม.ตอนต้น วิชาสังคมศึกษา 6. เกรดเฉลี่ยม.ตอนต้น วิชาสุขศึกษา 7. เกรดเฉลี่ยม.ตอนต้น วิชาศิลปะ 8. เกรดเฉลี่ยม.ตอนต้น วิชาการงานอาชีพ 9. เกรดเฉลี่ยม.ตอนต้น วิชาภาษาต่างประเทศ มาทดสอบด้วย อธิบายการเปลี่ยนแปลงของตัวแปรตามได้ร้อยละ 89.12

ดังนั้นผู้วิจัยจึงเลือกใช้โมเดลจากอัลกอริทึมของต้นไม้ตัดสินใจ (Decision Tree 4.5) ในการนำไปพยากรณ์แผนการเรียนที่เหมาะสมสำหรับการศึกษาต่อระดับมัธยมศึกษาตอนปลาย เพราะมีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้

เอกสารอ้างอิง

[1] การทำเหมืองข้อมูล, สืบค้นเมื่อ 27 ตุลาคม 2559, จาก https://th.wikipedia.org/wiki/การทำเหมืองข้อมูล

[2] กรธัช อยู่สุข (2014), แรงจูงใจและปัจจัยในการตัดสินใจศึกษาต่อในระดับบัณฑิตศึกษา (Motivations and Factors in Decision Making for Post Graduate Study), สืบค้นเมื่อ 27 ตุลาคม 2559, จาก https://www.gotoknow.org/posts/581868

[3] มัธยมศึกษา, สืบค้น เมื่อ 27 ตุลาคม 2559, จาก https://th.wikipedia.org/wiki/มัธยมศึกษา

[4] ชุติมา อุตมะมุณีย์ และประสงค์ ประณีตพลกรัง, "การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา," วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยสารสนเทศ สถาบันวิทยาการสารสนเทศ มหาวิทยาลัยศรีปทุม, 2553.

[5] ณัฏฐา ผิวมา, "การพัฒนาแบบจำลองพยากรณ์แนวโน้มการสมัครงานให้ตรงกับวุฒิการศึกษา สาขาคอมพิวเตอร์โดยใช้โครงข่ายประสาทเทียม," วารสารปัญญาภิวัฒน์ ปีที่ 7 ฉบับที่ 2 คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนดุสิต, 2558.

[6] พรรณิภา บุตรเอก, สุรเดช บุญลือ, "การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน," สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, มหาวิทยาลัย-นอร์ท กรุงเทพ, 2557

[7] เยาวภา ภารสำเร็จ และคณะ, "การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา," สาขาวิชา เทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม, 2556.

[8] สุธีรา วงศ์อนันทรัพย์ และคณะ, "การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา" วารสารสังคมศาสตร์ ปีที่ 5 ฉบับที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์ คณะบริหารธุกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก, 2559.

[9] สุพัฒน์กุล ภัคโชค, "ตัวแบบการเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลายโดยการพิจารณาผลการเรียนรายวิชาหลักด้วยเทคนิคเหมืองข้อมูล," สาขาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์, 2556.

[10] เสกสรรค์ วิลัยลักษณ์ และคณะ, "การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา," วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศและการสื่อสารมหาวิทยาลัยสุโขทัยธรรมาธิราช, 2558.

[11] อนันต์ ปินะเต และคณะ, "การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี," วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม, 2557.