บทที่ 2
แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

2.1 ระบบปลูกพืชโดยไม่ใช้ดิน (Hydroponics System)

2.1.1 ประวัติระบบปลูกพืชโดยไม่ใช้ดิน

เทคนิคการปลูกพืชโดยไม่ใช้ดินด้วยวิธีการปลูกในน้ำ (Hydroponics) หรือใช้วัสดุปลูกชนิดอื่นที่ไม่ใช้ดิน มนุษย์รู้จักวิธีการปลูกพืชในลักษณะปลูกในน้ำมานานกว่า 300 ปี โดยมีหลักฐานการปลูกพืชด้วยวิธีไฮโรโพนิคส์ที่ “The Parthenon” มหาวิหารอันเลื่องชื่อบนเขาอโครโพลีส (Acropolis) ตั้งแต่สมัยกรีกก่อนคริสศตรรษ 447 ปี และ 600 ปีต่อมาก่อนคริสศตวรรษ ในอาณาจักรบาลิโลเนีย ได้มีการสร้างสวนลอยฟ้า (Hanging Gardens) ขึ้น ถือได้ว่าเป็นสวนไฮโดรโพนิกส์แห่งแรกของโลก (The First Hydroponics Gardens) ซึ่งต่อมาได้มีการบันทึกอักษรภาพ การปลูกพืชในน้ำโดยไม่ใช้ดิน ในช่วงร้อยปีก่อนคริสตกาลของอาณาจักรอียิปต์แห่งลุ่มแม่น้ำไนล์ และในศตวรรษที่ 11 ชนเผ่า Aztecs ของ เม็กซิโก ในอเมริกากลางได้ค้นพบวิธีการปลูกพืชบนผิวน้ำ เรียกว่า “Chinampas” ซึ่งหมายถึงสวนลอยน้ำ (Floating Gardens) การปลูกพืชโดยไม่ใช้ดินเริ่มนานแล้วเช่นกันในทวีปเอเชียโดยในช่วงที่มาโคโปโล (Marco Polo) เดินทางมาถึงประเทศจีน ในช่วงประมาณปี ค.ศ. 1275 เข้าพบเห็นสวนลอยน้ำ (Floating Gardens) ของชาวจีนและได้บันทึกไว้ในบันทึกการเดินทางของเขา แนวคิดการทดลองปลูกพืชโดยใช้สารละลายที่มีธาตุอาหารอยู่ในน้ำ ได้ริเริ่มโดยนับย้อนหลังไปถึงศตวรรษที่ 17 โดยในปี ค.ศ. 1620 ได้เริ่มมีการทดลองโดย วันเฮลมองท์ (Van Helmont) ได้ทดลองปลูกต้นหลิว (Willow) ในท่อที่มีน้ำหนักดินแห้ง 90.7 กิโลกรัม โดยต้นหลิวในช่วงเริ่มปลูกจะมีน้ำหนัก 2.27 กิโลกรัม การปลูกพืชนี้จะให้น้ำฝนกับพืชเพียงอย่างเดียว โดยการปลูกหลิวนี้ใช้ระยะเวลา 5 ปี จะพบว่าต้นหลิวมีน้ำหนักเพิ่มขึ้นเป็น 72.56 กิโลกรัม ในขณะที่ดินมีน้ำหนักหายไปเพียง 56.7 กิโลกรัม เขาได้สรุปผลการทดลองว่า พืชเจริญเติบโตได้เพราะได้สารช่วยการเจริญเติบโตจากน้ำเท่านั้น ซึ่งการสรุปดังกล่าวไม่เป็นที่ยอมรับของนักวิทยาศาสตร์ในรุ่นหลัง เนื่องจากในขณะนั้นองค์ความรู้ด้านธาตุอาหารพืช ยังไม่ได้พัฒนาและรวบรวมเป็นหลักเกณฑ์อย่างเช่นปัจจุบัน เฮลมองท์ยังไม่ได้มีความรู้ว่า พืชยังได้รับก๊าซคาร์บอนไดออกไซด์และก๊าซออกซิเจนจากอากาศ

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

ต่อมาในปี ค.ศ. 1699 นักพฤกษศาสตร์ชาวอังกฤษ มร.วู้ดเวิร์ด (John Woodward) ได้ทดลองการปลูกมิ้นในน้ำ แล้วสรุปว่าถ้ามีการเพิ่มเติมดินเข้าไปในน้ำในสัดส่วนต่างๆ กันจะมีผลทำให้พืชเจริญเติบโตได้ดี โดยปริมาณดินที่เพิ่มลงในน้ำในปริมาณมากจะทำให้พืชเจริญดีที่สุดและดีกว่าการปลูกในน้ำธรรมดาและในน้ำกลั่นตามลำดับ

ความก้าวหน้าทางงานวิจัยและพัฒนายังไม่หยุดยั้งในการจำแนกให้ได้ว่า สารชนิดใดช่วยให้พืชเจริญเติบโตได้ เป็นที่ดำเนินต่อไปอย่างเชื่องช้าและในที่สุดเมื่อองค์ความรู้ด้านเคมีมีการพัฒนาจนถึงช่วงหนึ่ง โดนในปี พ.ศ. 2347 De Saussure ได้ศึกษาถึงชนิดธาตุอาหารที่พืชได้รับจากดินน้ำและอากาศ โดยได้มีการทดลองเพิ่มเติมโดยนักทดลองชาวฝรั่งเศสที่ชื่อ Boussingault ในปี พ.ศ. 2394 ได้ทดลองปลูกพืชในทรายและกรวด และให้สารละลายที่รู้ถึงสัดส่วนของธาตุอาหารพืชชนิดต่างๆ เขาสามารถสรุปว่าพืชประกอบด้วยธาตุไนโตรเจนและวัสดุอื่นๆ อีกมากมาย โดยน้ำจะให้ธาตุไฮโดรเจน อากาศจะให้ธาตุคาร์บอนและออกซิเจนแก่พืชในช่วงศตวรรษที่ 19 ได้มีการค้นพบทฤษฎีของฮิวมัสและทฤษฎีการสลายตัวของปุ๋ยคอก โดยนักวิทยาศาสตร์ชื่อดังชาวเยอรมัน มร.ลีบิค (Justus Von Liebig) ซึ่งมีผลทำให้วิทยาการทางด้านวิทยาศาสตร์ด้านวิทยาศาสตร์เกษตร ได้ก้าวหน้าขึ้นสู่ยุคการศึกษาลึกลงไปในทฤษฎีของธาตุอาหารพืชโดยมีนักวิทยาศาสตร์ชาวเยอรมัน 2 คน คือ Sachs ในปี พ.ศ. 2403 และ Knop ในปี พ.ศ. 2404 ได้ทำการทดลองเรื่องธาตุอาหารพืชโดยการศึกษาในห้องปฏิบัติการ เพื่อหาความสัมพันธ์ของธาตุอาหารพืชกับสรีรวิทยาของพืช โดยในการทดลองถึงการเจริญเติบโตของพืชผักปกติ ในสารละลายที่มีธาตุอาหารที่พืชต้องการหรือจำเป็นต้องใช้ได้แก่ ไนโตรเจน ฟอสฟอรัส โพแทสเซียม แคลเซียม และแมกนีเซียม ซึ่งเป็นธาตุที่พืชต้องการมากและยังได้พบธาตุอาหารพืชที่ต้องการในปริมาณน้อยๆอีก 7 ธาตุ ได้แก่ เหล็ก คลอรีน แมงกานีส โบรอน สังกะสี ทองแดง และโมลิบดีนัม ซึ่งในภายหลังได้มีนักวิทยาศาสตร์อีกจำนวนมากได้คิดค้นสูตรอาหารที่เหมาะสมกับการเจริญเติบโตของพืช โดยสูตรสารละลายธาตุอาหารพืชจำนวนมากยังคงนำมาใช้อยู่ในงานวิจัยในห้องปฏิบัติการในปัจจุบัน

การพัฒนาเรื่องของธาตุอาหารพืชในสารละลายสำหรับการปลูกพืช เริ่มต้นมีการพัฒนาอย่างมากอีกครั้งในช่วง พ.ศ.2468ในช่วงที่มีการพัฒนาระบบปลูกพืชในโรงเรือนเพื่อแก้ปัญหาการปลูกพืชในบริเวณที่มีปัญหาเรื่องโครงสร้างดินที่ไม่ดี ดินมีความอุดมสมบูรณ์ต่ำ ต้องลงทุนสูง และดินมีปัญหาเรื่องการสะสมของโรค และแมลง โดยการพัฒนาเรื่องสารละลายสำหรับการปลูกพืชในโรงเรือนได้กระทำควบคู่ไปกับการพัฒนารูปแบบของการปลูกพืชในโรงเรือน และความพยายามหาวิธีการปลูกพืชเพื่อเลียนแบบการปลูกพืชที่ใช้ปฏิบัติมา โดยในช่วงปี พ.ศ. 2468 -2478 เป็นช่วงที่มีความพยายามในการพัฒนาระบบการปลูกพืชที่ใช้ในห้องปฏิบัติการมาขยายขนาดให้เป็นระบบการปลูกพืชในโรงเรือนขนาดใหญ่

การปลูกพืชโดยไม่ใช้ดินในยุคใหม่เริ่มต้นประมาณ ปี ค.ศ. 1930 โดยนักวิทยาศาสตร์ชาวอเมริกัน ศาสตราจารย์ดร.เกอริค (William F. Gericke) แห่งมหาวิทยาลัยแคลิฟอร์เนีย โดยเป็นการศึกษาการปลูกพืชในน้ำที่มีความลึกประมาณ 15 เซนติเมตร ในสภาพน้ำนิ่ง โดยทดสอบปลูกมะเขือเทศในน้ำยาผสมธาตุอาหารตามสูตรที่เขาดัดแปลงขึ้น เพื่อพัฒนาเป็นระบบการปลูกที่เป็นการค้าสามารถปลูกได้ต้นยาวถึง 25 ฟุต จนกระทั่งออกดอกและติดผลที่มีขนาดรับประทานได้ ผลงานของ ดร.เกอริค ได้รับการตีพิมพ์เผยแพร่ในหนังสือพิมพ์ ดร.เกอริค เป็นผู้ตั้งชื่อวิธีการนี้ว่า “ไฮโดรโพนิคส์” ซึ่งมีรากศัพท์มาจากคำในภาษากรีก 2 คำ คือ Hydro แปลว่า น้ำ (Water) และ Ponos แปลว่า การทำงานหรือแรงงาน (Labor) รวมกันเป็นการปฏิบัติงานเกี่ยวกับน้ำ (Water-Working) และศาสตราจารย์ท่านนี้เป็นคนแรกที่นำเทคนิคการปลูกพืชแบบนี้ไปประยุกต์ในการปลูกพืช จากการทดลองพบว่าวิธีการปลูกวิธีนี้สามารถปลูกพืชได้เกือบทุกชนิด

2.1.2 ระบบการปลูกพืชโดยไม่ใช้ดิน (Hydroponics System)

เป็นการปลูกพืชที่ไม่มีข้อจำกัดในเรื่องของดินเนื่องจากเราไม่ได้ใช้ดิน เข้ามาเป็นปัจจัยในการเจริญเติบโตของพืช การปลูกพืชโดยไม่ใช้ดินมีหลายรูปแบบ เช่น โดยสารอาหารจะไหลผ่านเป็นแผ่นฟิล์มบางๆ 1 – 3 มิลลิเมตร (NFT: Nutrient Film Technique) และการปลูกแบบรากสัมผัสกับสารอาหารในน้ำลึก 3 – 5 เซนติเมตร (DFT: Deep Flow Technique)

ในการปลูกพืชในรูปแบบโดยสารอาหารจะไหลผ่านเป็นแผ่นฟิล์มบางๆ 1 – 3 มิลลิเมตร (NFT: Nutrient Film Technique) และการปลูกแบบรากสัมผัสกับสารอาหารในน้ำลึก 3 – 5 เซนติเมตร (DFT: Deep Flow Technique) จะมีความแตกต่างกัน ประกอบด้วย การหมุนเวียนการปลูกพืช ระยะห่างสำหรับการปลูกพืช รูปแบบโรงเรือนสำหรับการปลูก และการลงทุน

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

2.1.3 ข้อดีข้อเสียระบบปลูกพืชโดยไม่ใช้ดิน

ตารางที่ 2.1 ตารางเปรียบเทียบข้อดีและข้อเสียระบบปลูกพืชโดยไม่ใช้ดิน

ข้อดี	ข้อเสีย
1. สามารถปลูกพืชในบริเวณพื้นที่ที่ดินไม่ดีหรือสภาพแวดล้อมไม่เหมาะสม	1. มีต้นทุนเริ่มต้นการผลิตค่อนข้างสูงเนื่องจากต้องใช้อุปกรณ์ต่างๆ แต่มีศักยภาพในการคืนทุนเร็ว
2. ใช้พื้นที่ปลูกน้อยและสามารถผลิตได้อย่างสม่ำเสมอ	2. ผู้ปลูกต้องมีความชำนาญและมีประสบการณ์ในการดูแล
3. ประหยัดค่าขนส่งสามารถเลือกผลิตในแหล่งรับซื้อ ทำให้มีศักยภาพในเชิงการค้าสูง	3. ต้องการการดูแลควบคุมเสมอ
4. ประหยัดเวลาแรงงานและค่าใช้จ่ายในการเตรียมดิน และกำจัดวัชพืช	4. ถ้าหากไม่มีความรู้และความสามารถในการจัดการที่ดีพออาจทำให้มีปริมาณธาตุอาหารในการผลิตพืชสูง
5. ใช้แรงงานน้อยแต่มีประสิทธิภาพสูง	5. วัสดุบางชนิดสลายยากต้องควบคุม อาจส่งผลต่อสิ่งแวดล้อม
6. สามารถปลูกได้อย่างต่อเนื่องตลอดในพื้นที่เดียวกัน	6. มีข้อจำกัดของชนิดพืชไม่สามารถปลูกทุกชนิดที่สามารถปลูกในดินได้
7. พืชเจริญเติบโตได้เร็ว และให้ผลผลิตเร็วกว่าการปลูกพืชธรรมดา 1-2 สัปดาห์	7. มีการแพร่กระจายของเชื่อโรคทางน้ำได้เร็ว และยากต่อการควบคุม
8. ตัดปัญหาเกี่ยวกับศัตรูพืชเกิดจากดิน	8. เมื่ออุณหภูมิต่ำกว่า 7 ppm ซึ่งส่งผลเสียต่อการปลูกพืช
9. สามารถใช้น้ำและธาตุอาหารพืชอย่างมีประสิทธิภาพ	
10. สามารถควบคุมสภาพแวดล้อมต่างๆที่เกี่ยวข้องกับการเจริญเติบโตได้อย่างถูกต้อง	
11. ให้ผลตอบแทนที่มากกว่าการปลูกพืชในดิน	

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

2.2 ปัจจัยที่ส่งผลต่อการเจริญเติบโต

การเจริญเติบโตของพืช ต้องการปัจจัยหลายประการที่สำคัญ คือ น้ำ แสง ธาตุอาหารต่างๆ พืชเป็นสิ่งมีชีวิตมีการเจริญเติบโตและดำรงชีวิตอยู่ได้ย่อมต้องการ สิ่งแวดล้อมที่เหมาะสม สภาพของสิ่งแวดล้อมต่างๆ ที่มีผลต่อการเจริญเติบโตของพืช ประกอบด้วย

2.2.1 ดิน เป็นปัจจัยสำคัญอันดับแรก ดินที่เหมาะสมต่อการเจริญเติบโตของพืช ต้องเป็นดินที่อุ้มน้ำได้ดี ร่วนซุย มีอินทรีย์วัตถุมาก แต่เมื่อใช้ดินปลูกไปนานๆ ดินอาจเสื่อมสภาพ เช่น หมดแร่ธาตุ จำเป็นต้องมีการปรับปรุงดินให้อุดมสมบูรณ์ ได้แก่ การไถพรวน การใส่ปุ๋ย การปลูกพืชหมุนเวียน เป็นต้น

2.2.2 น้ำ มีความสำคัญต่อการเจริญเติบโตของพืชมาก น้ำช่วยละลายแร่ธาตุอาหารในดิน เพื่อให้รากดูดอาหารไปเลี้ยงส่วนต่างๆ ของลำต้นได้ และยังช่วยให้ดินมีความชุ่มชื้น พืชสดชื่นและการทำงานของกระบวนการต่างๆ ในพืชเป็นไปอย่างปกติ

2.2.3 ธาตุอาหารหรือปุ๋ย เป็นสิ่งที่ช่วยให้พืชเจริญเติบโต ดียิ่งขึ้น ธาตุอาหารที่จำเป็นต่อการเจริญเติบโตของพืชมี 16 ธาตุ แต่ธาตุที่พืชต้องการมากและในดินมักมีไม่เพียงพอ คือ ธาตุไนโตรเจน ฟอสฟอรัส และโพแทสเซียม ธาตุอาหารเหล่านี้จะต้องอยู่ในรูปสารละลายที่พืชนำไปใช้ได้และต้องมีปริมาณ ที่พอเหมาะจึงจะทำให้การเจริญเติบโตของพืชเป็นไปด้วยดี แต่ถ้ามีไม่เพียงพอต้องเพิ่มธาตุอาหารให้แก่พืชในรูปของปุ๋ย

2.2.4 อากาศ ในอากาศมีแก๊สหลายชนิด แต่แก๊สที่พืชต้องการมากคือ แก๊สคาร์บอนไดออกไซด์และแก๊สออกซิเจน ซึ่งใช้ในการสังเคราะห์ด้วยแสงเพื่อสร้างอาหารและหายใจ แก๊สทั้งสองชนิดนี้มีอยู่ในดินด้วย ในการปลูกพืชเราจึงควรทำให้ดินโปร่งร่วนซุยอยู่เสมอ เพื่อให้อาหารที่อยู่ในช่องว่างระหว่างเม็ดดินมีการถ่ายเทได้

2.2.5 แสงสว่างหรือแสงแดด พืชต้องการแสงแดดมาใช้ในการสร้างอาหาร ถ้าขาดแสงแดด พืชจะแคระแกรน ใบจะมีสีเหลืองหรือขาวซีดและตายในที่สุด พืชแต่ละชนิดต้องการแสงไม่เท่ากันพืชบางชนิดต้องการแสงแดดจัด แต่พืชบางชนิดก็ต้องการแสงรำไร

2.2.6 อุณหภูมิ มีส่วนช่วยในการงอกและเจริญเติบโตของพืชเช่นกัน จะเห็นได้ว่าพืชบางชนิดชอบขึ้นในที่มีอากาศหนาวเย็น แต่พืชบางชนิดก็ชอบขึ้นในที่มีอากาศร้อน การนำพืชมาปลูกจึงควรเลือกชนิดที่เหมาะสมกับอุณหภูมิที่เปลี่ยนไปตามฤดูกาล ในแต่ละท้องถิ่นด้วยช่องว่างอากาศ ช่องว่างระหว่างเม็ดดินบางส่วนเป็นที่อยู่ของน้ำ ส่วนที่เหลือจะเป็นที่อยู่ของอากาศ อากาศในดินมีความสำคัญต่อการเจริญเติบโตของพืชมาก เพราะรากพืชต้องการ แก๊สออกซิเจนในการหายใจ ธาตุอาหารที่จำเป็นต่อพืชมี 6 ชนิด ได้แก่ ไนโตรเจน ฟอสฟอรัส โพแทสเซียม แคลเซียม แมกนีเซียม กำมะถัน เหล็ก สังกะสี แมงกานีส ทองแดง โบรอน โมลิบดินัม คลอรีน คาร์บอน ออกซิเจน ไฮโดรเจน ส่วนธาตุอาหารที่พืชต้องการมากกว่า

บางครั้งเรียกว่า ธาตุอาหารหลัก ได้แก่ ไนโตรเจน ฟอสฟอรัส โพแทสเซียม และธาตุอาหารที่ต้องการรองลงมา เรียกว่า ธาตุอาหารรอง ได้แก่ แคลเซียม
แมกนีเซียม และกำมะถัน

2.3 การทำเหมืองแร่ข้อมูล (Data mining)

การทำเหมืองข้อมูล (Data mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

2.3.1 วิวัฒนาการของการทำเหมืองข้อมูล

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือและป้องกันการสูญหายได้เป็นอย่างดี

- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ในการนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ

- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ

- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

2.3.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines)

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) วิธีการของซัพพอร์ตเวกเตอร์แมชชีน จัดเป็นเทคนิคที่ใช้ในการแก้ปัญหาทางด้านการรู้จำรูปแบบข้อมูลโดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ที่สุด(OptimalSeparating Hyperplane) สำหรับรากฐานเดิมของซัพพอร์ตเวกเตอร์แมชชีน ถูกนำมาใช้กับข้อมูลที่เป็นเชิงเส้นแต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้ในระบบการสอนให้ระบบเรียนรู้ส่วนใหญ่มักเป็นข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งสามารถแก้ปัญหาดังกล่าวด้วยเคอร์เนลที่เป็นการเปลียนแปลงมิติของข้อมูลให้สูงขึ้น เพื่อช่วยในการเรียงตัวของข้อมูลเสียใหม่ที่ เรียกว่า “พื้นที่มิติสูง” (Higher Dimensional Space) เคอร์เนลที่ นิยมใช้มีอยู่ 3 ชนิด ด้วยกันคือ โพลิโนเมียล (Polynomial), เรเดียลเบสิสฟังก์ชัน (Radial Basis Function-RBF) และ ซิกมอยด์ (Sigmoid)

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

2.5 โครงข่ายประสาทเทียม (Neural Network)

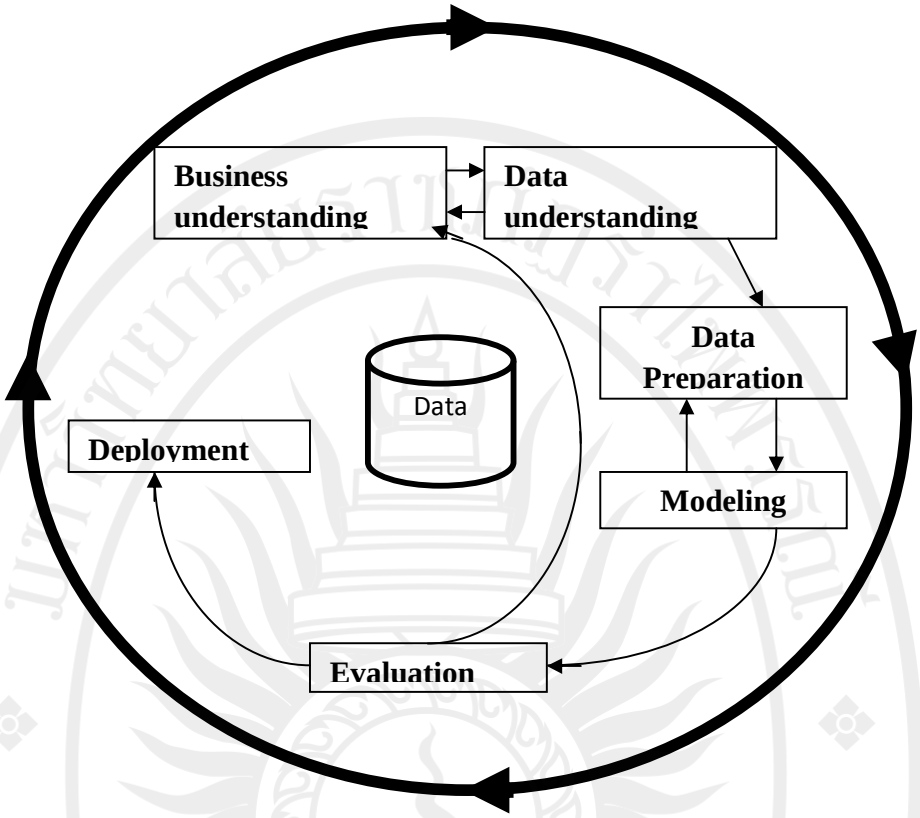
การเรียนรู้ของโครงข่ายแบบ Perceptron จะมีลักษณะของโครงสร้างเป็นชั้น ซึ่งข้อมูลที่จะนำเข้ามาเรียนรู้นั้นจะถูกใส่เข้าไปในชั้นแรกสุดซึ่งเป็นชั้นรับข้อมูล (Input Layer) และเมื่อผ่านการคำนวณจากชั้นแรกนี้แล้วผลลัพธ์ จะถูกส่งต่อไป ยังชั้นกลางของโครงข่าย (Hidden Layer) ซึ่งในแต่ละหน่วยของชั้นนี้ จะรับข้อมูลจากทุกหน่วยในชั้นก่อนหน้ามาคำนวณ แล้วส่งต่อไปยังชั้นถัดไป และเมื่อข้อมูลถูกส่งกันมาจนถึงชั้นสุดท้าย (Output Layer) ก็จะได้ผลลัพธ์ออกมาจากระบบ ซึ่งการส่งผ่านข้อมูลต่อๆ กันไปจะเรียกว่าเป็นการส่งต่อข้อมูลแบบการป้อนไปข้างหน้า (Feed Forward) หลังจากนั้นจะต้องมีการตรวจสอบผลลัพธ์ที่ ได้จากระบบว่ามีความคลาดเคลื่อนจากเป้าหมายมากน้อย หากความคลาดเคลื่อนจากเป้าหมายมากเกินไปจะต้องมีการนำค่าความคลาดเคลื่อนนี้ไปปรับน้ำหนักการเรียนรู้ใหม่ (Weight) แบบแพร่ย้อนกลับ (Back-Propagation) ซึ่งจะเป็นการปรับน้ำหนักความคลาดเคลื่อนจากชั้นผลลัพธ์ไปยังชั้นก่อนหน้า และทำการปรับน้ำหนักไปเรื่อยๆจนกระทั่งถึงชันรับข้อมูล ซึ่งกระบวนการเรียนรู้แบบนี้ต้องอาศัยการทำซ้ำหลายรอบจน จะได้ผลลัพธ์ตามที กำหนดหรือได้ค่าความคลาดเคลื่อนที่น้อยจนพอยอมรับได้ซึ่งจำนวนรอบ จะขึ้นอยู่กับความยากง่ายของปัญหา ขนาดของข้อมูลรวมไปถึงจำนวนชั้นของโครงสร้างของโครงข่ายที่จะสร้าง

2.6 The Cross Industry Standard Process for Data Mining

Process model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks and relationships between these tasks. At this description level, it is not possible to identify all relationships. Essentially, relationships could exist between any data mining tasks depending on the goals, the background and interest of the user and most importantly on the data.

The life cycle of a data mining project consists of six phases. Figure 2.1 shows the phases of a data mining process. The sequence of the phases is not rigid. Moving back and forth between different phases is always required. It depends on the outcome of each phase which phase or which particular task of a phase, has to be performed next. The arrows indicate the most important and frequent dependencies between phases.

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี



ภาพที่ 2.1 Phases of the CRISP-DM reference model

The outer circle in Figure 2.1 symbolizes the cyclical nature of data mining itself. Data mining is not over once a solution is deployed. The lessons learned during the process and from the deployed solution can trigger new, often more focused business questions. Subsequent data mining processes will benefit from the experiences of previous ones.

In the following, we outline each phase briefly:

1. Business understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

2. Data understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

3. Data preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

4. Modeling

In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary.

5. Evaluation

At this stage in the project you have built a model (or models) that appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี

6. Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. It often involves applying “live” models within an organization’s decision making processes, for example in real-time personalization of Web pages or repeated scoring of marketing databases. However, depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models.

2.7 งานวิจัยที่เกี่ยวข้อง

บุษรา เสนอการนำวิธี GA มาใช้คัดเลือกหาลักษณะที่เหมาะสม จากนั้นจึงนำลักษณะที่เลือกได้ มาฝึกสอนกับข้อมูลชุดเรียนรู้โดยใช้วิธีทางระบบเครือข่ายประสาท (GA/ANN) เพื่อทำนายผลสำเร็จการศึกษาของนักศึกษา 2 ผลลัพธ์ คือ จบ และ ไม่จบ ซึ่งผลการทำนายให้ค่าที่ถูกต้องมากกว่าการทำนายโดยไม่ใช้ GA ร่วม (ANN) และลดจำนวนตัวแปรลงเหลือเพียง 22 ค่า จากเดิม 38 ตัวแปร

นิดา มาลินี ชัยวุฒิ และประเสริฐ เสนอหลักการของ Data Mining โดยเฉพาะการทำ Classification โดยใช้โมเดลต้นไม้ตัดสินใจโดยได้แสดงขั้นตอนการสร้างต้นไม้ตัดสินใจ (Decision Tree) จาก Algorithm ได้แก่ ID3, C5.0 และ CART ทำให้เข้าใจถึงการหาเงื่อนไขที่เหมาะสม เพื่อสร้างต้นไม้ตัดสินใจ

Hyontai Sug เสนอวิธีหาความแม่นยำในการตัดสินใจจากเงื่อนไขของ Attribute วัดความสามารถของโมเดลจัดกลุ่มหรือจำแนกกลุ่ม (Clustering and Classification) ทดลองใช้ Conflict Dataset จากการ Random Data และทดสอบกับ Confidence 5 ค่าดังนี้ (0.25, 0.15, 0.01, 0.001) แล้วเปรียบเทียบ Size, Error Rate ผลการทดลองจาก Classification แสดงว่า Attribute ในการตัดสินใจมีส่วนต่อการสร้างโมเดลที่ให้ความถูกต้องและแม่นยำ

Lawrence o. Hall, Kevin w. Bowyer, Robert E. Banfield, Steven Echrich and Richard Collins กล่าวถึงงานวิจัยที่ทดสอบเทคนิค Pruning สำหรับ C4.5 คือ Error Based Pruning

(EBP) ผลที่ได้คือต้นไม้ที่มีขนาดใหญ่แต่ความถูกต้องไม่เพิ่มขึ้น ทั้งนี้เพราะงานวิจัยส่วนใหญ่ให้ค่าพารามิเตอร์ Certainty Factor (CF) =0.25 ซึ่งเป็นคำแนะนำว่าให้ผลดีที่สุดกับหลายปัญหา งานวิจัยนี้จึงทำการทดลองกับข้อมูลหลาย Dataset และกำหนด CF ต่างกันกับจำนวนข้อมูลที่น้อยและมากต่างกัน เพื่อทดลองว่ามีผลต่อขนาดของต้นไม้และความถูกต้องหรือไม่ และเมื่อทดสอบเปรียบเทียบกับ Reduced Error Pruning (REP) พบว่าค่าความถูกต้องที่ได้มีค่าใกล้เคียงกัน

ลิขสิทธิ์ของมหาวิทยาลัยราชภัฏรำไพพรรณี