19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

การทำนายผลสำเร็จการศึกษาของนักศึกษา โดยระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้
Graduation Successfulness Forecasting by Adaptive Neuro-Fuzzy Inference System

นิภาพร อนุวงศ์[1], ดุษฎี ประเสริฐธิติพงษ์[2] และ วิจักษณ์ ศรีสัจจะเลิศวาจา[3]
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่, เชียงใหม่
E-mail: nipaporn_anuwong@cmu.ac.th[1], dussadee.p@cmu.ac.th[2], wijak.s@cmu.ac.th[3]

บทคัดย่อ

การติดตามความก้าวหน้าของนักศึกษาระดับปริญญาตรี เพื่อลดปัญหาการพ้นสภาพและจบเกินหลักสูตร 4 ปี ด้วยการดูจากเกรดเฉลี่ยสะสมเพียงอย่างเดียวไม่เพียงพอ เนื่องจากยังมีปัจจัยอื่นที่เกี่ยวข้อง และข้อมูลนักศึกษามีขนาดใหญ่ จึงมีการประยุกต์ใช้การทำเหมืองข้อมูลทางการศึกษาและเทคนิคระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ (ANFIS) เพื่อหาปัจจัยที่เกี่ยวข้องและทำนายผลการสำเร็จการศึกษาด้วยการใช้ข้อมูลนักศึกษาภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ เมื่อเปรียบเทียบผลการทำนายระหว่างเกรดเฉลี่ยสะสมกับโมเดล ANFIS พบว่าผลที่ได้จากโมเดลมีค่าความถูกต้องมากกว่าการทำนายจากเกรดเฉลี่ยสะสมในทุกชุดข้อมูลทดสอบที่แบ่งตามชั้นปีและภาคเรียน โดยชุดตัวแปรอินพุตที่มีการแยกหมวดหมู่วิชาศึกษาทั่วไป และมีกลุ่มวิชาที่ได้เกรด S สามารถทำนายผลสำเร็จแต่ละสถานภาพทางการศึกษาได้ดีกว่าชุดตัวแปรอินพุตอื่น

คำสำคัญ: การทำเหมืองข้อมูลทางการศึกษา, ระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้

Abstract

Monitoring the progress of undergraduate student to reduce the retired problem and over finished the course of four years with only use the cumulative grade point average (GPA) is not enough. Due to there are other factors involved and the large student data. Therefore, the educational data mining and adaptive neuro-fuzzy inference systems (ANFIS) technique was applied to find the relevant factors and predict graduated with using student data of the department of computer science, faculty of science, chiang mai university. When comparing the results of predictions between the cumulative GPA and ANFIS model indicated that the model had higher accuracy than GPA predications in all testing data according to divide each where with, the input dataset of general education categories and the subject has got S grade were separated and can predict each educational status better than other input dataset.

Keywords: educational data mining, adaptive neuro-fuzzy inference system

1. บทนำ

1.1 ที่มาและความสำคัญ

ปัญหาการพ้นสภาพหรือการสำเร็จการศึกษาช้ากว่าหลักสูตรของนักศึกษาก่อให้เกิดการสูญเสียต้นทุนด้านการเงินและเวลาของทั้งสถาบันศึกษาและนักศึกษา ทำให้ต้องมีการติดตามความก้าวหน้าและความเสี่ยงต่อปัญหาดังกล่าวด้วยเกรดเฉลี่ยสะสมที่นักวางแผนทางวิชาการใช้เป็นปัจจัยร่วมในการประเมินผล และติดตามความก้าวหน้าของนักศึกษา [1-2] แต่เกรดเฉลี่ยสะสมเพียงปัจจัยเดียวไม่สามารถวิเคราะห์สถานภาพของนักศึกษาในอนาคตได้อย่างเหมาะสม เนื่องจากมีปัจจัยอื่นที่เกี่ยวข้อง เช่น เกรดเฉลี่ยสะสมของกลุ่มวิชาเอก กลุ่มวิชาปัญหาพิเศษ หน่วยกิตรวม ของแต่ละกลุ่มวิชา เป็นต้น ซึ่งข้อมูลนักศึกษาในฐานข้อมูลที่มีการเก็บอย่างต่อเนื่องเป็นจำนวนมากได้มีการนำมาวิเคราะห์ด้วยระบบคอมพิวเตอร์ เช่น การใช้เทคนิคทางสถิติแบบเดิมร่วมกับการจัดการด้านฐานข้อมูล แต่ขาดการจัดการกับชุดข้อมูลที่มีขนาดใหญ่ มีความซับซ้อนและข้อมูลที่ไม่ชัดเจน ทำให้เกิดการปรับปรุงเทคนิคและเครื่องมือสำหรับการวิเคราะห์ในชุดข้อมูลขนาดใหญ่ หนึ่งในเทคนิคนั้นคือ การทำเหมืองข้อมูล (Data mining) ที่ได้รับความสนใจจากนักวิจัยหลายแขนง เช่น การออกแบบฐานข้อมูล สถิติ การจดจำรูปแบบ กลไกการเรียนรู้ เป็นต้น [3] เมื่อนำมาประยุกต์ใช้กับการศึกษาเรียกว่า การทำเหมืองข้อมูลทางการศึกษา (Educational data mining) เพื่อให้ครอบคลุมปัญหาด้านการศึกษามากขึ้น และสามารถแก้ปัญหาที่มีความเฉพาะได้โดยตรง [4] การทำเหมืองข้อมูลทางการศึกษาจะเกี่ยวข้องกับการสกัดสิ่งที่เป็นประโยชน์หรือรูปแบบใหม่มาจากฐานข้อมูลด้านการศึกษาขนาดใหญ่ เพื่อทำให้เกิดความเข้าใจ การปรับปรุงประสิทธิภาพการศึกษา และการประเมินกระบวนการเรียนรู้ของนักศึกษาที่ดีขึ้น [5-6] หนึ่งในวิธีการสร้างโมเดลที่ได้รับความนิยมคือ โครงข่ายประสาทเทียม [7] ที่มีคุณสมบัติการปรับแต่งความรู้ที่ซ่อนอยู่ภายในโครงข่ายด้วยการต่อเชื่อมโยงกัน และมีการส่งผ่านข้อมูลแบบขนานทำให้มีการประมวลผลที่เร็ว แต่ผลลัพธ์ที่ได้ไม่สามารถอธิบายในลักษณะของเหตุและผลได้ ส่วนระบบฟัซซีสามารถอธิบายเหตุและผลที่สอดคล้องกับตรรกะความคิดของมนุษย์ เพราะมีการตีความในรูปถ้า-แล้ว (What-if analysis) แต่ไม่มีกระบวนการเรียนรู้ในการปรับแต่งโครงสร้างของกฎและตัวแปรของระบบ จึงต้องอาศัยผู้เชี่ยวชาญ

ขอสงวนลิขสิทธิ์

เมื่อนำมาผสมผสานกันเป็นการเพิ่มข้อดีและลดข้อเสีย เรียกว่า ระบบนิวโรฟัซซี (Neuro-fuzzy) ระบบผสมนิวโรฟัซซีที่นิยมคือ ระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ (Adaptive neuro-fuzzy inference systems, ANFIS) [8] ที่มีโครงสร้างแบบหลายอินพุตหนึ่งเอาท์พุต ส่วนตัวแปรอินพุตได้มาจากการศึกษาเอกสารข้อบังคับมหาวิทยาลัยเชียงใหม่ว่าด้วยการศึกษาระดับปริญญาตรี เพื่อหาเงื่อนไขของการสำเร็จการศึกษา

วัตถุประสงค์ของการศึกษานี้จึงมุ่งเน้นหาผลการสำเร็จการศึกษาด้วย ANFIS และศึกษาเฉพาะนักศึกษาภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ โดยใช้ข้อมูลลงทะเบียนเรียนของนักศึกษาที่เข้าศึกษาตั้งแต่ปีการศึกษา 2548 ถึง 2551 ในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ปี 2548 ถึง 2551

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูลทางการศึกษา

การทำเหมืองข้อมูลทางการศึกษาเป็นสาขาวิชาที่เกิดขึ้นเพื่อพัฒนาวิธีการสำรวจข้อมูลชนิดใหม่ที่เกี่ยวกับการศึกษา ทำให้เกิดความเข้าใจต่อนักศึกษาและสิ่งที่เกี่ยวข้องกับการศึกษามากขึ้น ซึ่งจะมีศาสตร์ด้านจิตวิทยาเข้ามาเกี่ยวข้อง ทำให้ต่างจากการทำเหมืองข้อมูลทั่วไป [7] และแต่ละปัญหามีความเฉพาะ ทำให้ไม่สามารถประยุกต์ใช้กับประเภทข้อมูลและปัญหาที่เกิดขึ้นได้โดยตรง จึงต้องมีการปรับกระบวนการค้นหาความรู้ กระบวนการค้นหาความรู้ในฐานข้อมูลทางการศึกษาสามารถได้ความได้จากมุมมองที่แตกต่างกัน [4] ขึ้นอยู่กับสภาพแวดล้อมทางการศึกษา และระบบข้อมูล แต่มีขั้นตอนหลักเหมือนกันคือ การเตรียมข้อมูล การสร้างโมเดลด้วยเทคนิคที่เหมาะสมกับงานวิจัย และการประเมินผลลัพธ์

2.2 ระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้

ระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้เป็นการผสมผสานการทำงานระหว่างระบบฟัซซีกับโครงข่ายประสาทเทียม โดยที่ระบบฟัซซีจะใช้สร้างกฎฟัซซีจากการเรียนรู้ของตัวแปรอินพุต และตัวแปรเอาท์พุต ส่วนโครงข่ายประสาทเทียมจะมีการเรียนรู้แบบไปข้างหน้า และแบบย้อนกลับเพื่อปรับพารามิเตอร์ด้านอินพุต และด้านเอาท์พุตของแต่ละกฎฟัซซี ซึ่งนำเสนอโดย Jang [9] ปีค.ศ. 1993

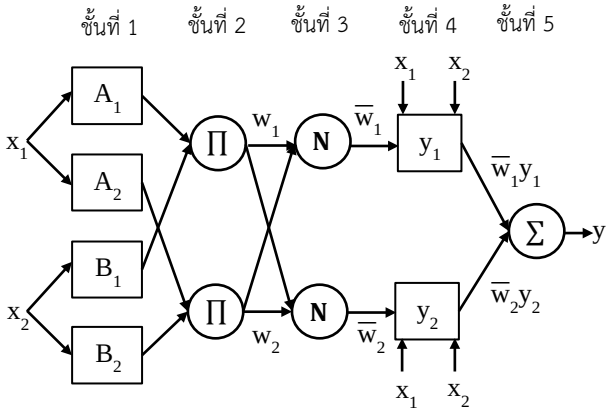
กฎฟัซซีจะถูกสร้างด้วยการอนุมานกฎฟัซซีแบบ Sugeno ที่มีเอาท์พุตสุดท้ายได้เพียงตัวเดียว แบ่งเป็นด้านอินพุตกับด้านเอาท์พุต แต่ละด้านจะมีพารามิเตอร์ที่ต้องปรับให้เหมาะสมกับชุดข้อมูล โดยที่ด้านอินพุตเรียกว่า พารามิเตอร์ข้อตั้ง (Premise parameters) จะอยู่ในฟังก์ชันความเป็นสมาชิก ส่วนด้านเอาท์พุตเรียกว่า พารามิเตอร์ข้อตาม (Consequent parameters) อยู่ในเอาท์พุตเชิงเส้นของแต่ละกฎฟัซซี ตัวอย่างเช่น กรณีที่มีตัวแปรอินพุต 2 ตัว คือ x_1 และ x_2 และมีตัวแปรเอาท์พุตคือ y จะได้กฎอยู่ 2 กฎ ดังนี้

กฎข้อที่ 1 IF x_1 is A_1 and x_2 is B_1 THEN $y_1 = p_1x_1 + q_1x_2 + r_1$

กฎข้อที่ 2 IF x_1 is A_2 and x_2 is B_2 THEN $y_2 = p_2x_1 + q_2x_2 + r_2$

โดยที่ A_1, A_2, B_1, B_2 เป็นฟัซซีเซตของพจน์ภาษาเช่น พอใช้ ปานกลาง ดี p_1, q_1, r_1, p_2, q_2, r_2 เป็นพารามิเตอร์ข้อตามของกฎฟัซซี

โครงสร้างของ ANFIS แบ่งออกเป็น 5 ชั้น ดังรูปที่ 1 มีการทำงานดังนี้ ชั้นที่ 1 จะแปลงข้อมูลอินพุตเป็นค่าฟัซซีด้วยการหาค่าความเป็นสมาชิกจากฟังก์ชันความเป็นสมาชิก เช่น ฟังก์ชันเกาส์เซียน ฟังก์ชันระฆังคว่ำ ซึ่งจะมีการปรับค่าพารามิเตอร์ข้อตั้งชั้นนี้ จากนั้นชั้นที่ 2 จะคูณค่าความเป็นสมาชิกในแต่ละตัวแปรอินพุต ค่าที่ได้เป็นน้ำหนัก (Firing strength) ของแต่ละกฎฟัซซี ต่อมาชั้นที่ 3 เป็นการปรับค่าน้ำหนักที่ได้จากชั้นที่ 2 ให้เป็นมาตรฐานเดียวกัน (Normalized firing strength) ด้วยการนำค่าน้ำหนักของแต่ละกฎฟัซซี หารด้วยผลรวมของค่าน้ำหนักทั้งหมด จากนั้นชั้นที่ 4 จะคำนวณหาเอาท์พุตของแต่ละกฎฟัซซี โดยนำค่าที่ได้จากชั้นที่ 3 คูณกับเอาท์พุตเชิงเส้นซึ่งจะปรับค่าพารามิเตอร์ข้อตามที่ชั้นนี้ สุดท้ายชั้นที่ 5 เป็นการรวมค่าที่ได้จากชั้นที่ 4 ทุกกฎฟัซซีเข้าด้วยกัน



รูปที่ 1 โครงสร้างของ ANFIS กรณีที่มี 2 ตัวแปรอินพุต และ 1 ตัวแปรเอาท์พุต

ในแต่ละรอบการเรียนรู้ของ ANFIS จะมีการเรียนรู้แบบผสมผสาน (Hybrid Learning) ระหว่างการคำนวณไปข้างหน้าด้วยวิธีปรับตามความลาดชัน (Gradient descent method) กับการคำนวณย้อนกลับด้วยวิธีประเมินกำลังสองน้อยสุด (Least squares estimate) เพื่อปรับค่าพารามิเตอร์ โดยที่การคำนวณไปข้างหน้าเป็นการหาค่าเอาท์พุตและค่าความผิดพลาดด้วยการปรับพารามิเตอร์ข้อตามในชั้นที่ 4 และให้พารามิเตอร์ข้อตั้งมีค่าคงที่ เมื่อได้อัตราค่าความผิดพลาดที่มากกว่ากำหนดจะคำนวณย้อนกลับด้วยการปรับค่าพารามิเตอร์ข้อตั้งในชั้นที่ 1 และให้ค่าพารามิเตอร์ข้อตามมีค่าคงที่

2.3 งานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูลเป็นสายงานของการค้นคืนสารสนเทศและใช้ประโยชน์จากข้อมูลขนาดใหญ่ โดยได้มีการประยุกต์ใช้กับหลายสายงาน เช่น ธุรกิจค้าขาย หุ้น เป็นต้น ในปัจจุบันมีการเพิ่มความน่าสนใจในการใช้งานของการทำเหมืองข้อมูล เพื่อตรวจสอบคำถามทางวิทยาศาสตร์ของงานวิจัยทางการศึกษา เรียกว่า การทำเหมืองข้อมูลทางการศึกษา Baker และ Yacef [7] ได้จำแนกเป้าหมายของงานวิจัยเป็น 4 กลุ่ม คือ 1) เพื่อทำนายพฤติกรรมการเรียนรู้ของนักศึกษาในอนาคตด้วยการสร้างโมเดลนักศึกษาของแต่ละคน โดยใช้ความรู้ของนักศึกษา แรงจูงใจ กระบวนการรู้คิด และทัศนคติ 2) เพื่อค้นหาหรือปรับปรุงโมเดลองค์ประกอบหลักของสถาบันการศึกษา 3) เพื่อศึกษาผลกระทบของการสนับสนุนการเรียนการสอนประเภทต่างๆ จากซอฟต์แวร์การเรียนรู้ และ 4) เพื่อแสดงความก้าวหน้าของความรู้ทางวิทยาศาสตร์ที่เกี่ยวกับการเรียนรู้และผู้เรียนรู้

ผู้ใช้งานวิจัยสามารถใช้ได้

ด้วยการสร้างโมเดลทางคอมพิวเตอร์ ซึ่งเป้าหมายดังกล่าวครอบคลุมทั้งด้านนักศึกษา ด้านสถาบันการศึกษา ด้านซอฟต์แวร์ที่ใช้สนับสนุนการเรียนการสอน และด้านทฤษฎี นอกจากนี้ยังมีการแบ่งเทคนิคที่ใช้ในการสร้างโมเดลเป็น 5 กลุ่มดังนี้ การทำนายล่วงหน้า การจัดกลุ่ม การกลั่นกรองข้อมูลเพื่อให้มนุษย์ตัดสินใจ การทำเหมืองความสัมพันธ์ และการพัฒนาโมเดล เทคนิคการทำนายล่วงหน้าเป็นการพัฒนาโมเดลที่สามารถอนุมานตัวแปรที่ถูกทำนายหนึ่งตัวด้วยการรวมกันของตัวแปรทำนายหลายตัวได้ ส่วนมากจะประยุกต์ใช้กับการตรวจสอบพฤติกรรมของนักศึกษา การพัฒนาโมเดลองค์ประกอบหลักของสถาบันการศึกษา การทำนายและการทำความเข้าใจเกี่ยวกับผลการศึกษาของนักศึกษา โดยมีวิธีสร้างโมเดลดังนี้ การจำแนกประเภท การวิเคราะห์ความถดถอย และการประมาณฟังก์ชันความหนาแน่น ซึ่งวิธีการที่ได้รับความนิยม ได้แก่ ตัดสินใจแบบต้นไม้ การวิเคราะห์ความถดถอยโลจิสติก โครงข่ายประสาทเทียม ซัพพอร์ตเวกเตอร์แมชชีน การวิเคราะห์ความถดถอยเชิงเส้น การประมาณฟังก์ชันเคอร์เนล

ปัจจุบันเทคนิค ANFIS มีการใช้อย่างกว้างขวางในทางการศึกษาด้านการจำแนกประเภท การประมาณค่า และการทำนายล่วงหน้า [10] Inyang [11] มีการประยุกต์ใช้การทำเหมืองข้อมูลทางการศึกษากับเทคนิค ANFIS เพื่อทำนายเกรดเฉลี่ยสะสม จากนั้นนำมาแบ่งกลุ่มด้วยเทคนิค FCM เพื่อจำแนกกลุ่มความเสี่ยงของนักศึกษาแต่ละคน Osman [12] ใช้เทคนิค ANFIS ทำนายผลการเรียน โดยจะใช้คะแนนสอบต่างๆ เป็นตัวแปรอินพุต แล้วเปรียบเทียบกับวิธีการทางสถิติ Yusoff [13] ทำการเปรียบเทียบโครงข่ายประสาทเทียมกับ ANFIS เพื่อทำนายเกรด ซึ่ง ANFIS ให้ผลที่ดีกว่า

3. วิธีการทดลอง

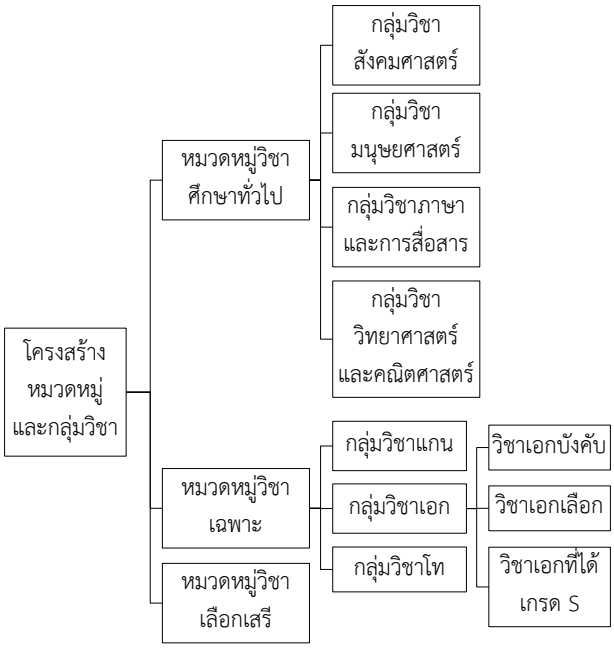
การทำเหมืองข้อมูลทางการศึกษามีกระบวนการค้นหาความรู้ในฐานข้อมูลทางการศึกษาประกอบด้วย ขั้นตอนการเตรียมข้อมูล ขั้นตอนการสร้างโมเดลด้วยเทคนิค ANFIS และขั้นตอนการวัดประสิทธิภาพของโมเดลที่ได้ มีรายละเอียดของแต่ละขั้นตอนดังนี้

3.1 ขั้นตอนการเตรียมข้อมูล

ข้อมูลที่นำมาใช้เป็นข้อมูลนักศึกษาที่เข้าเรียนตั้งแต่ปีการศึกษา 2548 ถึง 2551 ในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ เนื่องจากอยู่ในโครงสร้างหลักสูตรเดียวกัน ซึ่งมีสถานภาพสำเร็จการศึกษาหรือพ้นสภาพเนื่องจากผลการเรียนต่ำกว่าเกณฑ์เท่านั้น มีข้อมูลทั้งหมด 338 คน ส่วนข้อมูลลงทะเบียนเรียนจะเลือกใช้ข้อมูลตั้งแต่ชั้นปีที่ 1 จนถึงชั้นปีสุดท้ายของสถานภาพ โดยมีระยะเวลาเรียนสูงสุด 8 ปี และได้รับเกรด A, B+, B, C+, C, D+, D, F และ S เท่านั้น

จากการศึกษาเอกสารข้อบังคับมหาวิทยาลัยเชียงใหม่ว่าด้วยการศึกษาระดับปริญญาตรี เงื่อนไขที่เกี่ยวข้องกับการสำเร็จการศึกษานอกจากเกรดเฉลี่ยสะสมแล้วมีหน่วยกิตและเกรดเฉลี่ยของหมวดหมู่วิชาและกลุ่มวิชาย่อยตามโครงสร้างหลักสูตร ดังรูปที่ 2 ดังนั้นจึงใช้เป็นตัวแปรอินพุต ส่วนตัวแปรเอาท์พุต คือ ผลสำเร็จการศึกษา โดยแบ่งเป็น 3 กลุ่มคือ กลุ่มพ้นสภาพ กลุ่มจบการศึกษา 4 ปี และกลุ่มจบการศึกษาเกิน 4 ปี

การแบ่งชุดข้อมูลทดสอบที่ประกอบด้วย ข้อมูลเรียนรู้ และข้อมูลทดสอบมี 4 ชุด โดยข้อมูลทดสอบจะเลือกจากชั้นปี และภาคเรียนที่ลงทะเบียนดังนี้ 1) ชั้นปี 2 ภาคเรียน 1 2) ชั้นปี 2 ภาคเรียน 2 3) ชั้นปี 3 ภาคเรียน 1 และ 4) ชั้นปี 3 ภาคเรียน 2 ซึ่งนอกจากชั้นปีและภาคเรียนของข้อมูลทดสอบแล้ว จะเป็นข้อมูลเรียนรู้



รูปที่ 2 โครงสร้างหมวดหมู่และกลุ่มวิชาย่อยที่ใช้เป็นตัวแปรอินพุต

3.2 ขั้นตอนการสร้างโมเดล

ในการศึกษานี้มีขั้นตอนการสร้างโมเดลของ ANFIS [14] ดังนี้

3.2.1 โหลดข้อมูลเรียนรู้ และข้อมูลทดสอบ ซึ่งจะแบ่งเป็น 4 ชุด ตามชั้นปีและภาคเรียนที่ลงทะเบียนเรียนดังที่กล่าวในขั้นตอนเตรียมข้อมูล

3.2.2 จัดกลุ่มข้อมูลแบบหยาบจะแบ่งข้อมูลที่ลักษณะคล้ายกันมาอยู่กลุ่มเดียวกัน ซึ่งการจัดกลุ่มมี 2 แบบ คือ การจัดกลุ่มแบบตระแกรง (Grid partitioning) และการจัดกลุ่มแบบลบออก (Subtractive clustering) งานวิจัยนี้เลือกใช้การจัดกลุ่มแบบลบออก เนื่องจากมีจำนวนตัวแปรอินพุตมาก และใช้เวลาน้อย

3.2.3 เลือกฟังก์ชันความเป็นสมาชิก งานวิจัยนี้เลือกฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน (Gaussian) ดังสมการที่ (1)

$$\mu_{A_i}(x)=\exp\left\{-\left(\frac{x-c_i}{a_i}\right)^2\right\} \quad (1)$$

เมื่อ c_i คือค่าเฉลี่ยและ a_i คือค่าเบี่ยงเบนมาตรฐาน

3.2.4 อนุมานกฎฟัซซีด้วยฟังก์ชันการจัดกลุ่มแบบลบออก จะได้จำนวนกฎฟัซซี ค่าพารามิเตอร์ข้อตั้งและข้อตามเริ่มต้น โดยใช้หลักของการหาจุดศูนย์กลางความหนาแน่นของข้อมูล ซึ่งจุดที่ให้ค่ามากที่สุดจะเป็นตัวแทนกลุ่ม วนหาค่าไปเรื่อยๆ จนครบชุดข้อมูล

3.2.5 เลือกอัลกอริทึมของการเรียนรู้ ใช้การเรียนรู้แบบผสมผสานระหว่างการคำนวณไปข้างหน้าด้วยวิธีปรับตามความลาดชัน กับการคำนวณย้อนกลับด้วยวิธีประเมินกำลังสองน้อยสุด

3.2.6 กำหนดจำนวนรอบการเรียนรู้ จะมีรอบเริ่มต้นที่ 50 เพิ่มขึ้นทีละ 50 จนถึงรอบที่ 1000 เมื่อเข้าสู่กระบวนการเรียนรู้ด้วยเทคนิค ANFIS จะได้ค่าเอาท์พุตสุดท้ายเพียงค่าเดียว

ห้ามจำหน่าย

3.2.7 วัดค่าความถูกต้องของการเรียนรู้ในชุดข้อมูลเรียนรู้ด้วยข้อมูลทดสอบ และวัดประสิทธิภาพของโมเดลที่ได้

3.3 ขั้นตอนการวัดประสิทธิภาพโมเดลที่ได้

การทำนายด้วยเกรดเฉลี่ยสะสมมาจากการแบ่งกลุ่มสถานภาพตามเกรดเฉลี่ยสะสมมีเกณฑ์ 3 ช่วงคือ ช่วงที่ 1 มีค่า 0 ถึงไม่เกิน 2.00 จะอยู่ในกลุ่มพ้นสภาพ ช่วงที่ 2 มีค่า 2.00 ถึงไม่เกิน 2.50 จะอยู่ในกลุ่มจบเกิน 4 ปี และช่วงที่ 3 มีค่ามากกว่า 2.50 จะอยู่ในกลุ่มจบ 4 ปี โดยเกณฑ์ของกลุ่มพ้นสภาพมาจากเอกสารข้อบังคับมหาวิทยาลัยเชียงใหม่ว่าด้วยการศึกษาระดับปริญญาตรี ซึ่งค่าความถูกต้องของการทำนายหาได้จากจำนวนนักศึกษาที่มีเกรดเฉลี่ยสะสมอยู่ในกลุ่มแต่ละช่วงหารด้วยจำนวนนักศึกษาทั้งหมด

การทำนายด้วยเทคนิค ANFIS จะใช้สถิติในการวัดประสิทธิภาพของโมเดลที่ได้ ด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่า F-measure โดยที่ค่าความถูกต้องเป็นการวัดความแม่นยำของโมเดลรวมกันทุกกลุ่ม ค่าความแม่นยำเป็นการวัดความแม่นยำของโมเดลที่พิจารณาแยกทีละกลุ่มของเอาท์พุตที่ทำนายได้ ค่าระลึกเป็นการวัดความถูกต้องของโมเดลที่พิจารณาแยกทีละกลุ่มของเอาท์พุตจริง และค่า F-measure เป็นการวัดค่าจากความแม่นยำและค่าระลึก โดยพิจารณาแยกทีละกลุ่ม ดังสมการที่ (2)

$$\text{F-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2)$$

จากนั้นจะนำค่าความถูกต้องที่ได้จากการทำนายด้วยเกรดเฉลี่ยสะสมกับโมเดลที่ได้มาเปรียบเทียบกัน

4. ผลการทดลองและการอภิปราย

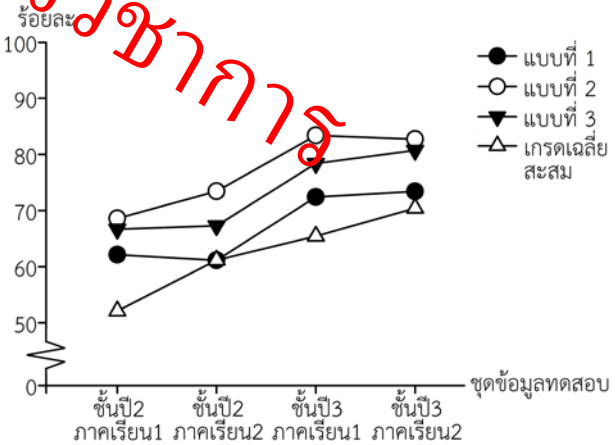
จากการทดลองสร้างโมเดลที่มีตัวแปรอินพุตแตกต่างกันตามโครงสร้างหมวดหมู่และกลุ่มวิชาดังรูปที่ 2พบว่า มีชุดตัวแปรอินพุต 3 แบบที่ให้ค่าความถูกต้องแตกต่างกัน คือ แบบที่ 1 มี 6 ตัวแปร ประกอบด้วย ภาคเรียนที่ลงทะเบียน หมวดหมู่วิชาศึกษาทั่วไป กลุ่มวิชาแกน กลุ่มวิชาเอก กลุ่มวิชาโท หมวดหมู่วิชาเลือกเสรี แบบที่ 2 มี 10 ตัวแปร มีดังนี้ ภาคเรียนที่ลงทะเบียน กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ กลุ่มวิชาแกน กลุ่มวิชาเอก กลุ่มวิชาเอกที่ได้เกรด S กลุ่มวิชาโท หมวดหมู่วิชาเลือกเสรี และแบบที่ 3 มี 11 ตัวแปร ประกอบด้วย ภาคเรียนที่ลงทะเบียน กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ กลุ่มวิชาแกน กลุ่มวิชาเอกบังคับ กลุ่มวิชาเอกเลือก กลุ่มวิชาเอกที่ได้เกรด S กลุ่มวิชาโท หมวดหมู่วิชาเลือกเสรี ดังนั้นจึงใช้ชุดตัวแปรอินพุตทั้ง 3 ชุด มาเปรียบเทียบกับการทำนายจากเกรดเฉลี่ยสะสม จากนั้นจะหาชุดตัวแปรอินพุตที่เหมาะสมกับแต่ละชุดข้อมูลทดสอบต่อไป

การสร้างโมเดล ANFIS ของแต่ละชุดตัวแปรอินพุตในชุดข้อมูลทดสอบที่แตกต่างกัน พบว่า ในทุกชุดข้อมูลทดสอบ ชุดตัวแปรอินพุตแบบที่ 2 ให้ค่าความถูกต้องมากที่สุด และมีรอบการเรียนรู้ที่ให้ค่าความถูกต้องที่ดีที่สุดแตกต่างกันดังนี้ ชุดข้อมูลทดสอบชั้นปี 2 ภาคเรียน 1 ในรอบการเรียนรู้ที่ 500 ชุดข้อมูลทดสอบชั้นปี 2 ภาคเรียน 2 ในรอบการเรียนรู้ที่ 400 ชุดข้อมูลทดสอบชั้นปี 3 ภาคเรียน 1 ในรอบการเรียนรู้ที่ 900 และชุดข้อมูลทดสอบชั้นปี 3 ภาคเรียน 2 ในรอบการเรียนรู้ที่ 1000

ตารางที่ 1 เปรียบเทียบค่าความถูกต้องของผลการทำนายจากข้อมูลทดสอบระหว่างการดูจากเกรดเฉลี่ยสะสมกับการใช้เทคนิค ANFIS

ชุดข้อมูลทดสอบ		เกรดเฉลี่ยสะสม	ชุดตัวแปรอินพุต		
			แบบที่ 1	แบบที่ 2	แบบที่ 3
ชั้นปี 2	ภาคเรียน 1	52.10	62.14	68.61	66.67
	ภาคเรียน 2	61.17	61.17	73.46	67.31
ชั้นปี 3	ภาคเรียน 1	65.45	72.43	83.39	78.41
	ภาคเรียน 2	70.43	73.42	82.72	80.73

เมื่อเปรียบเทียบกับการทำนายด้วยเกรดเฉลี่ยสะสม พบว่า การทำนายด้วย ANFIS ให้ค่าความถูกต้องมากกว่าการทำนายด้วยเกรดเฉลี่ยสะสมทุกชุดข้อมูลทดสอบ แสดงว่า เกรดเฉลี่ยสะสมเพียงปัจจัยเดียวไม่สามารถทำนายผลการสำเร็จการศึกษาในอนาคตได้อย่างเหมาะสมเนื่องจากปัจจัยอื่นสามารถส่งผลต่อการสำเร็จการศึกษาได้ดีกว่า ดังผลการทดลองตารางที่ 1 ที่ชุดตัวแปรอินพุตแบบที่ 2 ประกอบด้วย ภาคเรียนที่ลงทะเบียน กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ กลุ่มวิชาแกน กลุ่มวิชาเอก กลุ่มวิชาเอกที่ได้เกรด S กลุ่มวิชาโท และหมวดหมู่วิชาเลือกเสรี ให้ค่าความถูกต้องมากกว่า



รูปที่ 3 กราฟเปรียบเทียบค่าความถูกต้องของผลการทำนายจากข้อมูลทดสอบระหว่างเกรดเฉลี่ยสะสมกับการใช้เทคนิค ANFIS

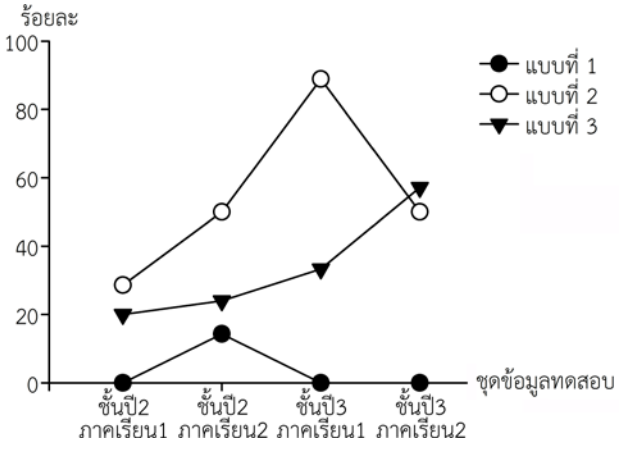
จากแนวโน้มค่าความถูกต้องของแต่ละชุดข้อมูลทดสอบในรูปที่ 3 จะมีค่าเพิ่มขึ้นเรื่อยๆ ซึ่งมีเพียงชุดตัวแปรอินพุตแบบที่ 2 ที่มีค่าลดลงในชุดข้อมูลทดสอบชั้นปี 3 ภาคเรียน 2 เมื่อพิจารณาค่า F-measure ของโมเดลที่ได้จากตารางที่ 2 ที่มีการพิจารณาแยกทีละกลุ่มสถานภาพ กลุ่มพ้นสภาพมีแนวโน้มดังรูปที่ 4 พบว่า ในชุดข้อทดสอบชั้นปี 2 ภาคเรียน 1 ชั้นปี 2 ภาคเรียน 2 และชั้นปี 3 ภาคเรียน 1 ชุดตัวแปรอินพุตแบบที่ 2 ให้ค่ามากที่สุด ซึ่งชุดตัวแปรอินพุตแบบที่ 1 มีการรวมหมวดหมู่วิชาศึกษาทั่วไปเข้า

ผลสัมฤทธิ์วิชาการ

ด้วยกัน และไม่มีกลุ่มวิชาที่ได้เกรด S ต่างจากแบบที่ 2 ที่มีการแยกเป็นกลุ่มวิชาย่อย ดังนี้ กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ และมีกลุ่มวิชาที่ได้เกรด S แสดงว่า กลุ่มวิชาย่อยในหมวดหมู่ศึกษาทั่วไปและกลุ่มวิชาที่ได้เกรด S มีอิทธิพลต่อกลุ่มพันสภาพในชั้นปี 2 และชั้นปี 3 ภาคเรียน 1 ส่วนแบบที่ 3 ที่มีการแยกกลุ่มวิชาเอกเป็นกลุ่มวิชาเอกเลือก และกลุ่มวิชาเอกบังคับ แตกต่างจากแบบที่ 2 ที่มีการรวมกันเป็นกลุ่มวิชาเอกอย่างเดียว แสดงว่า การรวมกลุ่มวิชาเอกมีอิทธิพลกับกลุ่มพันสภาพในชั้นปี 2 และชั้นปี 3 ภาคเรียน 1 มากกว่าการแยกกลุ่มวิชา เนื่องมาจากข้อมูลการลงทะเบียนในวิชาเอกย่อยของภาคเรียนนั้น ในชุดข้อมูลทดสอบไม่เพียงพอในทางตรงข้ามชุดข้อทดสอบชั้นปี 3 ภาคเรียน 2 ชุดตัวแปรอินพุตแบบที่ 3 มีค่ามากที่สุด แสดงว่า ในชั้นปี 3 ภาคเรียน 2 กลุ่มวิชาที่มีการแยกกลุ่มกันมีอิทธิพลต่อกลุ่มพันสภาพ และข้อมูลการลงทะเบียนเรียนของกลุ่มวิชาเอกย่อยเพียงพอต่อการทดสอบ

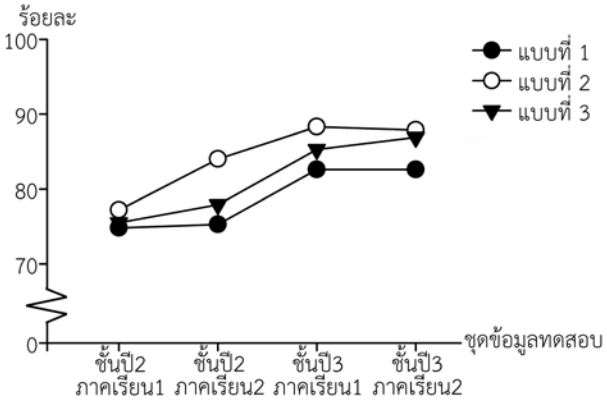
ตารางที่ 2 ค่า F-measure ในข้อมูลทดสอบของแต่ละชุดตัวแปรอินพุต

ชุดข้อมูลทดสอบ	ชุดตัวแปรอินพุต	สถานภาพ พ้นสภาพ	จบ 4 ปี	จบเกิน 4 ปี
ชั้นปี 2 ภาคเรียน 1	แบบที่ 1	0.00	74.83	35.93
	แบบที่ 2	28.57	77.19	58.82
	แบบที่ 3	20.00	75.90	56.38
ชั้นปี 2 ภาคเรียน 2	แบบที่ 1	14.29	75.27	18.71
	แบบที่ 2	50.00	84.01	57.32
	แบบที่ 3	24.00	77.91	51.25
ชั้นปี 3 ภาคเรียน 1	แบบที่ 1	0.00	82.60	35.29
	แบบที่ 2	88.89	88.31	72.09
	แบบที่ 3	33.33	85.30	66.67
ชั้นปี 3 ภาคเรียน 2	แบบที่ 1	0.00	82.61	45.26
	แบบที่ 2	50.00	87.86	73.33
	แบบที่ 3	57.14	86.89	68.51



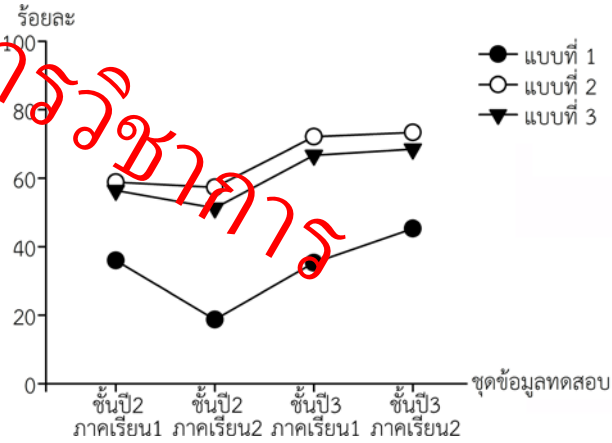
รูปที่ 4 กราฟเปรียบเทียบค่า F-measure ของกลุ่มพันสภาพ

ค่า F-measure ของโมเดลที่ได้ในกลุ่มจบ 4 ปีมีแนวโน้ม ดังรูปที่ 5 พบว่า ชุดตัวแปรอินพุตแบบที่ 2 ให้ค่ามากที่สุดในชุดข้อทดสอบชั้นปี 2 ภาคเรียน 1 ชั้นปี 2 ภาคเรียน 2 ชั้นปี 3 ภาคเรียน 1 และชั้นปี 3 ภาคเรียน 2



รูปที่ 5 กราฟเปรียบเทียบค่า F-measure ของกลุ่มจบ 4 ปี

ค่า F-measure ของโมเดลที่ได้ในกลุ่มจบเกิน 4 ปีมีแนวโน้ม ดังรูปที่ 6 พบว่า ชุดตัวแปรอินพุตแบบที่ 2 และแบบที่ 3 มีค่าในทิศทางเดียวกัน แต่ชุดตัวแปรอินพุตแบบที่ 2 จะให้ค่ามากที่สุดในทุกชุดข้อมูลทดสอบ และชั้นปี 3 ภาคเรียน 2 สามารถทำนายสถานภาพจบเกิน 4 ปีได้ดีที่สุด เนื่องจากเป็นชั้นปีที่มีข้อมูลการลงทะเบียนที่เพียงพอมากที่สุด



รูปที่ 6 กราฟเปรียบเทียบค่า F-measure ของกลุ่มจบเกิน 4 ปี

ดังนั้นจากการวิเคราะห์แนวโน้มในรูปที่ 3 ค่าความถูกต้องของชุดตัวแปรอินพุตแบบที่ 2 จากชุดข้อมูลทดสอบชั้นปี 3 ภาคเรียน 2 มีค่าลดลง เนื่องจากค่า F-measure ของกลุ่มพันสภาพ และจบ 4 ปี มีค่าลดลง (ตารางที่ 2) จึงทำให้ค่าความถูกต้องลดลง

5. สรุปผลการทดลอง

บทความนี้เป็นการทำนายผลการสำเร็จการศึกษาด้วย ANFIS ซึ่งมาจากปัญหาที่ว่า เกรดเฉลี่ยสะสมไม่สามารถทำนายผลการสำเร็จการศึกษาได้อย่างเหมาะสม จึงมีการศึกษาเงื่อนไขการจบเพื่อหาปัจจัยที่เกี่ยวข้องเช่น เกรดเฉลี่ยกลุ่มวิชาเอก หน่วยกิตในหมวดหมู่วิชา เป็นต้น

ห้องสมุดงานวิชาการ

และได้ตัวแปรอินพุตเป็นหมวดหมู่วิชาตามโครงสร้างหลักสูตร จากนั้นทำการทดลองหาตัวแปรอินพุตด้วยการแบ่งรายวิชาเป็นหมวดหมู่วิชา และกลุ่มวิชาย่อย ซึ่งข้อมูลที่ใช้มีทั้งเกรดเฉลี่ย และหน่วยกิต จนได้ชุดตัวแปรอินพุตที่ให้ค่าความถูกต้องที่แตกต่างกัน 3 แบบ ดังที่กล่าวไว้

จากปัญหาเบื้องต้น จึงนำผลการทำนายมาเปรียบเทียบกับผล การทำนายจากเกรดเฉลี่ยสะสม การทำนายด้วย ANFIS มีความถูกต้องมากกว่าในทุกชุดข้อมูลทดสอบในแต่ละชั้นปีและภาคเรียน แสดงให้เห็นว่าเกรดเฉลี่ยสะสมเพียงอย่างเดียวไม่สามารถทำนายผลการสำเร็จการศึกษาได้อย่างถูกต้องมากนัก ซึ่งยังมีปัจจัยอื่นที่เกี่ยวข้อง โดยจากการทดลองมีปัจจัยที่เกี่ยวข้องคือ ภาคเรียนที่ลงทะเบียน และเกรดเฉลี่ยของกลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ กลุ่มวิชาแกน กลุ่มวิชาเอก กลุ่มวิชาเอกที่ได้เกรด S กลุ่มวิชาโท หมวดหมู่วิชาเลือกเสรี และจากการทดลองชุดตัวแปรอินพุตแบบที่ 2 สามารถจำแนกกลุ่มสถานภาพและสามารถทำนายผลการสำเร็จการศึกษาในแต่ละสถานภาพได้ดีที่สุดในทุกชุดข้อมูลทดสอบ

การใช้เทคนิค ANFIS ในการทำนายผลนี้ควรมีการวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการสำเร็จการศึกษาก่อน เนื่องเทคนิคมีข้อจำกัดด้านตัวแปรที่ใช้เป็นตัวอินพุต ซึ่งถ้ามีตัวแปรอินพุตจำนวนมากจะใช้เวลาในการประมวลผลมาก นอกจากนี้จะทำการหาเทคนิคที่ช่วยในการจัดกลุ่มแบบหยาบก่อนที่จะนำเข้าการเรียนรู้ ด้วยเทคนิค ANFIS เพื่อลดจำนวนกฎฟัซซี เวลาในการเรียนรู้ และเพิ่มความถูกต้องของโมเดลอีกด้วย

ขอสงวนลิขสิทธิ์

เอกสารอ้างอิง

[1] M. Shovon and M. Haque, “An Approach of Improving Student’s Academic Performance by using K-means clustering algorithm and Decision tree,” *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 8, 2012.

[2] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, “Application of k Means Clustering algorithm for prediction of Students Academic Performance,” *International Journal of Computer Science and Information Security,* vol. 7, pp. 292–295, 2010.

[3] G. Piatetsky-Shapiro, “Data Mining and Knowledge Discovery 1996 to 2005: Overcoming the Hype and Moving from “University” to “Business” and “Analytics”,” *Data Mining and Knowledge Discovery,* vol. 15, no. 1, 2007.

[4] C. Romero and S. Ventura, “Data mining in education,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 3, 2013.

[5] A. Y. Chan, K. O. Chow and K. S. Cheung, “Online Course Refinement through Association Rule Mining,” *Journal of Educational Technology Systems,* vol 36, no. 4, 2008.

[6] C. Romero, S. Ventura, and E. García, “Data mining in course management systems: Moodle case study and tutorial,” *Computers & Education*, vol. 51, no. 1, 2008.

[7] R. S. J. D. Baker and K. Yacef, “The state of educaitonal data mining in 2009: A review and future visions,” *J. Educ. Data Min.*, vol. 1, 2009.

[8] พยุง มีสัจ, *ระบบฟัซซีและโครงข่ายประสาทเทียม.* เอกสารประกอบการสอน, คณะเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2552.

[9] J. S. R. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 3, 1993.

[10] R. Yayar, M. Hekim, V. Yilmaz and F. Bakirci, “A comparison of ANFIS and ARIMA techniques in the forecasting of electric energy consumption of Tokat province in Turkey,” *Journal of Economic and Social Studies,* vol. 1, no. 2, 2011.

[11] U. G. Inyang and E. E. Joshua, “Fuzzy Clustering of Students’ Data Repository for At-Risks Students Identification and Monitoring,” *Comput. Inf. Sci.*, 2013.

[12] O. Taylan and B. Karagözoğlu, “An adaptive neuro-fuzzy model for prediction of student’s academic performance,” *Comput. Ind. Eng.*, vol. 57, 2009.

[13] N. Yusof, N. A. M. Zin, N. M. Yassin, and P. Samsuri, “Evaluation of student’s performance and learning efficiency based on ANFIS,” *SoCPaR 2009 - Soft Comput. Pattern Recognit.*, 2009.

[14] A. Azizi, A. Y. B. Ali, and L. W. Ping, “Model development and comparative study of bayesian and ANFIS inferences for uncertain variables of production line in tile industry,” *WSEAS Trans. Syst.*, vol. 11, no. 1, 2012.