

Forecasting student admissions using back propagation multilayer neural network model

Anyawee Chaiwachiragompol^{1,*}, Sakchan Luangmaneerote¹, Jeeranun Tasuntia¹, Boonlueo Nabumroong¹

ABSTRACT

In this study, the feed forward multilayer neural network model for predicting student admissions was examined and developed. Crucial influencing factors of under-the increase and decrease of the number of students in the Faculty of Agriculture and Technology Rajamangala University of Technology Isan Surin Campus. This research used the historical data of students and were analyzed the characteristics of each student for training the model. An algorithm for supervised learning and the WEKA program were used in the research. There were two stages to the research: an analyzing the appropriate factors affecting students' attendance in step one, and then bringing the information from the first step factors into the model to create learning for the model in step two. In this study, selecting of proper model in the learning of the data is established by structuring the appropriate structure and offering the best validity. This research is divided data in two parts: the first part of this study data 70% was conducted by building learning for the model, the second part consisted of data 30% was analyzed by testing for the model. Moreover, the feed forward multilayer neural network's structural parameters were 8-19-2 and its learning cycles were 500. The result found that the testing on 387 participants showed a mean absolute error (MAE), the root mean square error (RMSE), and the accuracy value was 0.2179, 0.3696, 82.74%, respectively.

Keywords: Mean absolute error, root mean square error, accuracy, neural network

Published Online: 29 August, 2022

ISSN: 2730-3829

A. Chaiwachiragompol^{1,*}

¹Faculty of Agriculture and Technology
Rajamangala University of Technology
Isan, Surin Campus, Thailand
(jiraporn.sy@rmuti.ac.th)

S. Luangmaneerote¹

¹Faculty of Agriculture and Technology
Rajamangala University of Technology
Isan, Surin Campus, Thailand
(Tsakchan@hotmail.com)

J. Tasuntia¹

¹Faculty of Agriculture and Technology
Rajamangala University of Technology
Isan, Surin Campus, Thailand
(jeeranunt.ta@gmail.com)

B. Nabumroong¹

¹Faculty of Agriculture and Technology
Rajamangala University of Technology
Isan, Surin Campus, Thailand
(boonlueo.na@gmail.com)

**Corresponding Author*

Received date: 12 November 2021

Revised date: 23 May 2022

Accepted date: 29 June 2022

การพยากรณ์การเข้าศึกษาต่อของนักศึกษาโดยใช้โครงข่ายประสาทเทียม หลายชั้นแบบแพร่ย้อนกลับ

อัญวิณั ไขวชชะกะมล^{1*} ศักดิ์ชาญ เหลืองมณีโรจน์¹ จิรนนท์ ตะสันเทียะ¹ บุญเหลือ นามารุง¹

บทคัดย่อ

การวิจัยนี้เป็นการศึกษาและพัฒนาตัวแบบโครงข่ายประสาทเทียมหลายชั้นแบบแพร่ย้อนกลับสำหรับการพยากรณ์การเข้าศึกษาต่อของนักศึกษา ภายใต้ปัจจัยที่มีอิทธิพลต่อการเพิ่มและลดของจำนวนนักศึกษา คณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์ การวิจัยนี้ใช้ข้อมูลประวัติของนักศึกษามาทำการวิเคราะห์คุณลักษณะนักศึกษาแต่ละคน เพื่อสร้างการเรียนรู้ให้กับโมเดล การวิจัยนี้ใช้อัลกอริทึมในการเรียนรู้แบบมีผู้สอนและใช้โปรแกรม WEKA โดยการวิจัยมีการดำเนินการ 2 ขั้นตอน ขั้นแรกวิเคราะห์หาปัจจัยที่เหมาะสมที่ส่งผลต่อการเข้าเรียนของนักศึกษา และ ขั้นที่สอง นำข้อมูลของปัจจัยในขั้นตอนแรกเข้าสู่โมเดลเพื่อสร้างการเรียนรู้ให้กับโมเดล สำหรับงานวิจัยนี้การเลือกหาโมเดลที่เหมาะสมกับการเรียนรู้ของข้อมูลพิจารณาจากการกำหนดโครงสร้างที่เหมาะสมและให้ค่าความถูกต้องที่ดีที่สุด โดยการวิจัยได้แบ่งข้อมูลออกเป็น 2 ส่วน คือร้อยละ 70 ของข้อมูลใช้สำหรับการสร้างการเรียนรู้ให้กับโมเดล และร้อยละ 30 ของข้อมูลใช้สำหรับการทดสอบโมเดล นอกจากนี้โครงสร้างของโครงข่ายประสาทเทียมหลายชั้นแบบย้อนกลับมีโครงสร้างแบบ 8-19-2 และจำนวนรอบของการเรียนรู้เท่ากับ 500 รอบ ซึ่งผลของการทดสอบกับข้อมูลจำนวน 387 คน พบว่ามีค่า mean absolute error (MAE), root mean square error (RMSE) และค่าความถูกต้องเท่ากับ 0.2179, 0.3696, 82.74 เปอร์เซ็นต์ ตามลำดับ

คำสำคัญ: คลาดเคลื่อนสัมบูรณ์เฉลี่ย รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความถูกต้อง โครงข่ายประสาทเทียม

Published Online: 29 August, 2022

ISSN: 2730-3829

อัญวิณั ไขวชชะกะมล^{1*}

¹คณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์
(jiraporn.sy@muti.ac.th)

ศักดิ์ชาญ เหลืองมณีโรจน์¹

¹คณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์
(Tsakchan@hotmail.com)

จิรนนท์ ตะสันเทียะ¹

¹คณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์
(jeeranunt.ta@gmail.com)

บุญเหลือ นามารุง¹

¹คณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์
(boonlueo.na@gmail.com)

**Corresponding Author*

Received date: 12 November 2021

Revised date: 23 May 2022

Accepted date: 29 June 2022

1. บทนำ

การส่งเสริมการศึกษาของประเทศไทยได้มีการพัฒนาอย่างต่อเนื่องตั้งแต่อดีตจนถึงปัจจุบัน ทั้งนี้เพื่อสร้างทรัพยากรบุคคลที่มีศักยภาพให้กับประเทศ อย่างไรก็ตามภายใต้สถานการณ์โรคระบาดโควิด-19 (Ngamlamom, 2022) ทำให้มีนักศึกษาจำนวนหนึ่งที่ไม่ได้เข้าศึกษาต่อในระดับที่สูงขึ้นไป จากปัญหาดังกล่าวทำให้นักวิจัยมีแนวคิดที่จะศึกษาถึงปัจจัยที่มีผลต่อการเข้าศึกษาต่อในระดับที่สูงขึ้นไปของนักศึกษา ตลอดจนการสร้างโมเดลการพยากรณ์การเข้าศึกษาต่อของนักศึกษา (Wanon and Muangsang, 2020) ในการศึกษาปัจจัยที่มีผลต่อการเข้ารับการศึกษานักเรียนและนักศึกษาหลาย ๆ แห่ง พบว่าปัจจัยสำคัญที่ส่งผลต่อการเข้ารับการศึกษานักเรียนและนักศึกษา คือ รายได้ของผู้ปกครอง เศรษฐกิจ อาชีพ (Intawee, 2021) จากปัจจัยเหล่านี้ส่งผลทำให้นักเรียนและนักศึกษาไม่ได้รับการศึกษาในระดับที่สูงขึ้นไปเป็นจำนวนมาก จากปัญหาดังกล่าวได้ส่งผลกระทบไปยังสถานศึกษาในระดับสูงหลายแห่ง

ในด้านจำนวนนักศึกษาที่ลดจำนวนลง รวมทั้งคณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์ (PimPa, 2018) ดังนั้นการวิจัยครั้งนี้จึงเป็นการศึกษาปัจจัยที่มีผลต่อการเพิ่มขึ้นหรือลดลงของจำนวนนักศึกษาในสถานศึกษา โดยทำการวิเคราะห์หาปัจจัยที่มีความเกี่ยวข้องกับการเพิ่มขึ้นหรือลดลงของนักศึกษามากที่สุด และทำการสร้างโมเดลพยากรณ์แนวโน้มจำนวนนักศึกษาในแต่ละปี

ในขั้นตอนการศึกษาวิเคราะห์หาปัจจัยที่มีผลต่อการเพิ่มขึ้นหรือลดลงของจำนวนนักศึกษาของคณะเกษตรศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์นั้น เริ่มตั้งแต่ศึกษาวิเคราะห์ข้อมูลพื้นฐานด้านครอบครัว เศรษฐกิจ และปัจจัยอื่น ๆ ที่มีผลต่อจำนวนการเข้าศึกษาของนักศึกษา หลังจากนั้นได้ทำการศึกษาโมเดลการพยากรณ์แบบต่าง ๆ ที่นักวิจัยได้มีการนำไปประยุกต์ใช้ในการแก้ปัญหาต่าง ๆ เช่น การพยากรณ์รูปแบบของการกระทำความผิดซ้ำของเด็กและเยาวชน (Lorpunmanee et al., 2019) การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล (Treepuech, 2019) เป็นต้น จากการศึกษางานวิจัยในโมเดลการพยากรณ์ในรูปแบบต่าง ๆ พบว่าได้มีการนำเอาโมเดลแบบโครงข่ายประสาทเทียมไปประยุกต์ใช้ในการพยากรณ์ข้อมูลที่หลากหลายรูปแบบ ซึ่งงานวิจัยชิ้นนี้ได้มีการประยุกต์ใช้โมเดลแบบโครงข่ายประสาทเทียมหลายชั้นแบบย้อนกลับมาทำการพยากรณ์จำนวนนักศึกษาที่ทำการศึกษาต่อในระดับที่สูงต่อไป (Kanjanasamran wong, et al., 2017)

วัตถุประสงค์

1. เพื่อศึกษาการพยากรณ์ข้อมูลจากตัวแบบโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับเพื่อหาตัวแบบที่เหมาะสมในการเรียนรู้
2. เพื่อเปรียบเทียบการทดสอบตัวแบบโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับด้วยค่าความถูกต้องของคลาดเคลื่อนสัมบูรณ์เฉลี่ยและรากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

2. วิธีการวิจัย

1. แนวคิดและทฤษฎีที่เกี่ยวข้อง

1.1 การทำเหมืองข้อมูล (data mining) โดย Roiger and Geatz (2003) ได้ให้ความหมายของการทำเหมืองข้อมูล (data mining) ว่าเป็นกระบวนการที่ใช้เทคนิคการเรียนรู้ผ่าน ระบบคอมพิวเตอร์มากกว่าหนึ่งเพื่อวิเคราะห์และสกัดความรู้โดยอัตโนมัติจากข้อมูลที่จัดเก็บอยู่ ภายในฐานข้อมูลหนึ่ง ๆ การทำเหมืองข้อมูลมีวัตถุประสงค์เพื่อระบุแนวโน้มและแบบแผนของข้อมูล นอกจากนี้ Witten and Frank (2005) กล่าวว่าการทำเหมืองข้อมูลเป็นกระบวนการของการ ค้นหาแบบแผนในข้อมูล โดยเป็นกระบวนการที่เป็นลักษณะอัตโนมัติหรือกึ่งอัตโนมัติ การค้นหาแบบแผน นั้นมีความสำคัญ จะนำไปสู่ประโยชน์ต่าง ๆ โดยข้อมูลมักถูกนำเสนอในเชิงปริมาณ ส่วน Rahman (2009) กล่าวว่า การทำเหมืองข้อมูลคือ กระบวนการที่ช่วยให้ผู้ใช้สามารถเข้าถึงข้อมูลได้ โดยตรงจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อวิเคราะห์ข้อมูล แบ่งแยกประเภทข้อมูล จำแนกรูปแบบและความสัมพันธ์ของข้อมูลในคลังข้อมูล และนำผลที่ได้ไปใช้ในการตัดสินใจทางธุรกิจ หรือวิเคราะห์ ข้อมูลทางสถิติทำให้เกิดความรู้ใหม่ ๆ โดยผลลัพธ์ที่ได้จะมีหลายรูปแบบ เช่น อยู่ในรูปแบบของ กฎเกณฑ์ (rules) เป็นต้น เช่นเดียวกับ Jabjone (2015) ได้สรุปความหมายของการทำ เหมืองข้อมูลว่าเป็นกระบวนการค้นหาสารสนเทศหรือความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาใช้ในการทำนายหรือสร้างตัวแบบสำหรับจำแนกหน่วย กลุ่ม หมู่ หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของ สารสนเทศในฐานข้อมูลที่ประกอบจากกระบวนการทางสถิติ และการเรียนรู้ผ่านระบบคอมพิวเตอร์ ซึ่งการ ทำเหมืองข้อมูลมีขั้นตอนการดำเนินการหลายขั้นตอน ดังนั้น จึงจำเป็นต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น การจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น (Visutthirat, 2018)

การค้นหาความสัมพันธ์และรูปแบบทั้งหมดซึ่งมีอยู่จริงในฐานข้อมูล แต่ได้ถูกซ่อนไว้อยู่ในข้อมูลจำนวนมาก การทำเหมืองข้อมูลจะทำการสำรวจและวิเคราะห์อย่างอัตโนมัติหรือกึ่งอัตโนมัติ ในปริมาณข้อมูลจำนวนมากให้อยู่ใน

รูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎโดยความสัมพันธ์เหล่านี้แสดงให้เห็นถึงความรู้ต่าง ๆ ที่มีประโยชน์
 ในฐานข้อมูล (Pacharawongsakda, 2014) ดังแสดงใน Figure 1

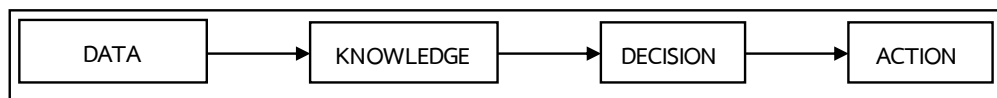


Figure 1 Decision-making and action

1.2 โครงข่ายประสาทเทียม (artificial neural network) หรือข่ายงานประสาท (neural network) เป็นโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ด้วยวัตถุประสงคที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำแบบรูป (pattern recognition) และการอุปมานความรู้ (knowledge deduction) สถาปัตยกรรมโครงข่ายประสาทเทียมแบ่งเป็น 4 แบบ คือโครงข่ายการส่งข้อมูลแบบไม่ย้อนกลับ (feed forward network) โครงข่ายการส่งข้อมูลแบบย้อนกลับ (feedback network) โครงข่ายประสาทเทียมอย่างง่าย (perceptron) และโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (back propagation) (Supabut, et al., 2015) โดยงานวิจัยเลือกแบบโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ดังแสดงใน Figure 2

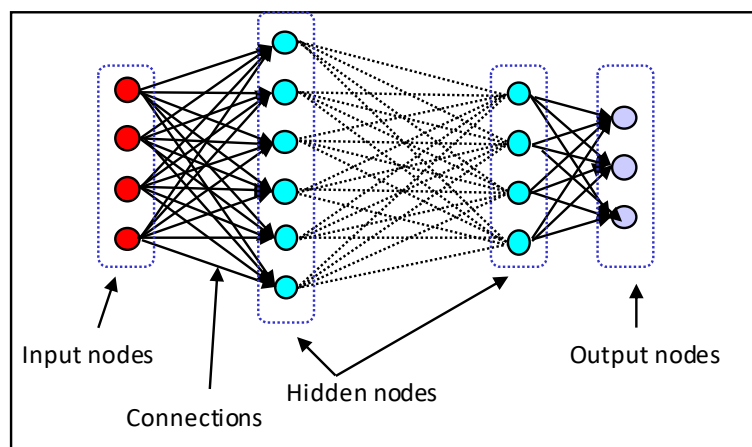


Figure 2 Neural network structure

1.2.1 โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ประกอบด้วยชั้นของข้อมูลเบื้องต้น 3 ชั้น ได้แก่ ชั้นข้อมูลนำเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นผลลัพธ์ (output layer) หรือ back propagation algorithm เป็นอัลกอริทึมที่ใช้ในการเรียนรู้ของเครือข่ายประสาทวิธีหนึ่ง ที่นิยมใช้ในโครงข่ายประสาทเทียมหลายชั้น (multilayer perceptron) เพื่อปรับค่าน้ำหนัก (weight) ในเส้นเชื่อมต่อระหว่างโหนดให้เหมาะสม โดยการปรับค่านี้นี้จะขึ้นกับความแตกต่างของค่าเอาต์พุตที่คำนวณได้กับค่าเอาต์พุตที่ต้องการ (Kijisirikul, 2003) คำนวณซ้ำ ๆ จนกว่าจะได้ผลลัพธ์ที่ใกล้เคียงค่าผลลัพธ์ที่ต้องการซึ่งเป็นการเรียนรู้แบบมีผู้สอนโดยมีขั้นตอนดังต่อไปนี้ ดังแสดงใน Figure 3 แสดงโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

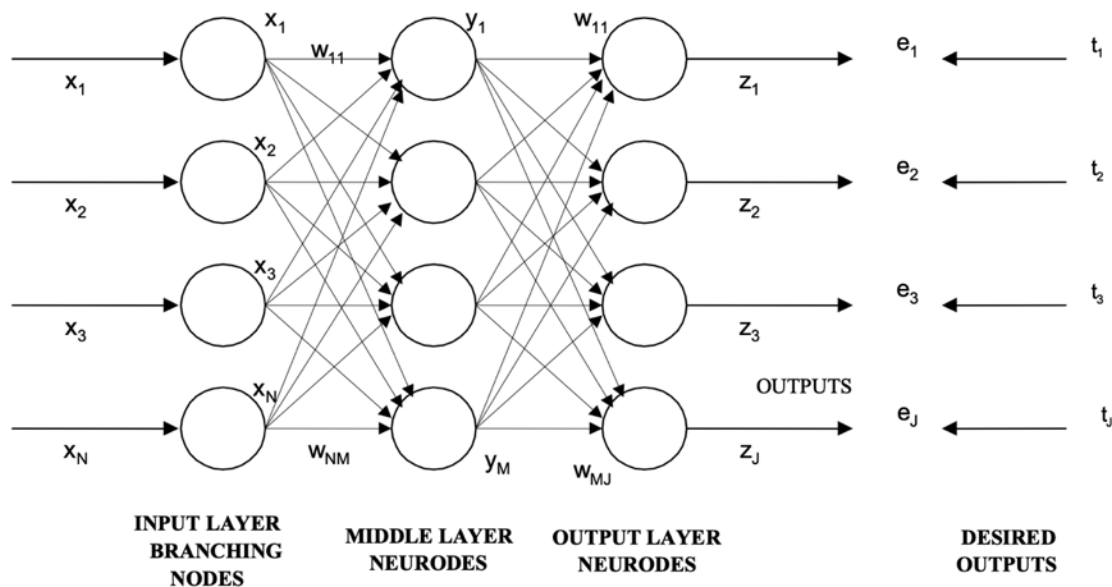


Figure 3 Feed forward neural network (modified from Winston, 1992)

ขั้นตอนการเรียนรู้ของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

1. ค่าค่าเอาต์พุตของชั้นซ่อนก่อนทำการปรับค่า ดังสมการที่ 1

$$S_m = \sum_{n=1}^N X_n * W_{nm} \quad \dots\dots\dots (1)$$

2. ค่าเอาต์พุตของชั้นซ่อนหลังทำการปรับค่า ดังสมการที่ 2

$$y_m = \int (S_m) \quad \dots\dots\dots (2)$$

3. ฟังก์ชันที่ใช้ปรับค่า $f(x)$ ดังสมการที่ 3

$$f(x) = \frac{1}{1+e^{-x}} \quad \dots\dots\dots (3)$$

4. คำนวณค่าเอาต์พุตของโหนดในชั้นเอาต์พุตด้วยสมการที่ 3-4 จากนั้นการปรับค่าผลรวมด้วยฟังก์ชันกระตุ้นฟังก์ชันซิกมอยด์ ดังสมการที่ 3-3 ซึ่งจะได้ค่าของโหนดในชั้นเอาต์พุตที่อยู่ในช่วง $[0, 1]$ สำหรับค่าของผลลัพธ์ในโหนดของชั้นเอาต์พุตนั้น หลังทำการปรับค่าแสดงดังสมการที่ 3-5

$$v_j = \sum_{m=1}^M y_m * w_{mj} \quad \dots\dots\dots (4)$$

5. ค่าเอาต์พุตของชั้นเอาต์พุตหลังทำการปรับค่า ดังสมการที่ 3-5

$$z_j = f(v_j) \quad \dots\dots\dots (5)$$

การหาความผิดพลาดของโหนดในชั้นเอาต์พุตและการทำการปรับค่าน้ำหนักนั้น โดยการนำเอาต์พุตที่คำนวณได้จริงเปรียบเทียบกับเอาต์พุตที่ได้กำหนดไว้ เพื่อหาความผิดพลาดของข้อมูล โดยถ้าค่าผิดพลาดของข้อมูลน้อยกว่าค่าผิดพลาดที่ยอมรับได้แล้วนั้น โครงข่ายประสาทเทียมก็จะทำการรับข้อมูลชุดต่อไปเข้าสู่โครงข่าย โดยถ้าไม่ได้ทำการปรับน้ำหนักแล้วนั้นโครงข่ายประสาทเทียมจะทำการรับข้อมูลแถวถัดไป และจะกลับไปทำขั้นตอนที่ 1 แต่ถ้าเป็นข้อมูลชุดสุดท้ายของข้อมูลแล้วนั้น จะไปตามสมการที่ 9 และ 10

6. ค่าความผิดพลาดในแต่ละแถวของข้อมูล ดังสมการที่ 6

$$e^{(q)} = \frac{1}{2} \sum_{j=1}^J (t_j - z_j)^2 \quad \dots\dots\dots (6)$$

7. การปรับน้ำหนักระหว่างโหนดในชั้นซ่อนและชั้นเอาต์พุต ดังสมการที่ 7

$$w_{mj}^{(r+1)} = w_{mj}^{(r)} + n \{ (t_j^{(q)} - z_j^{(q)}) * [z_j^{(q)} (1 - z_j^{(q)})] * y_m^{(q)} \} \dots\dots\dots (7)$$

8. การปรับน้ำหนักระหว่างโหนดในชั้นอินพุตและชั้นซ่อน มีสมการที่ 8

$$w_{nm}^{(r+1)} = w_{nm}^{(r)} + n \{ \sum_{j=1}^J (t_j^{(q)} - z_j^{(q)}) [z_j^{(q)} (1 - z_j^{(q)})] w_{mj}^{(r)} \} * [y_m^{(q)} (1 - y_m^{(q)})] [x_n^{(q)}] \dots\dots\dots (8)$$

1.3 การหาค่าความถูกต้องทางสถิติ

1.3.1 ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error : MAE) เป็นวิธีที่วัดค่าคลาดเคลื่อนที่เป็นที่นิยมอีกวิธีหนึ่ง ซึ่งจะช่วยบอกถึงขนาดของความคลาดเคลื่อนรวมได้ ในการวัดค่าความถูกต้องด้วยวิธีนี้หากค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก (Kaewwijit, 2016) สามารถหาได้จากสมการที่ 9

$$MAE = \frac{1}{n} * \sum |\bar{y}_i - y_i| \quad \dots\dots\dots (9)$$

1.3.2 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error : RMSE) เป็นวิธีการวัดค่าคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย สามารถหาได้จากสมการที่ 10 (Taesombat, 1996)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad \dots\dots\dots (10)$$

โดยที่ $\bar{y}_i$ แทน ค่าของข้อมูลที่ได้จากการพยากรณ์ , y แทน ค่าของข้อมูลที่เกิดขึ้นจริง และ n แทน จำนวนข้อมูลทั้งหมด

1.3.3 ค่าความถูกต้อง (accuracy) ที่สามารถหาได้จากสมการที่ 11

$$accuracy(Out) = \sum_{i=1}^{Out} \frac{Out_i}{T}, Out_i \in T \quad \dots\dots\dots (11)$$

เมื่อ *accuracy (out)* แสดงค่าเป็นเปอร์เซ็นต์ความถูกต้องจาก, *out_i* ค่าความถูกต้องข้อมูลที่พยากรณ์ได้ และ *T* คือจำนวนข้อมูลที่พยากรณ์ทั้งหมด (Chaiwachiragompol and Suwannata, 2016)

2. ขั้นตอนและวิธีการ

สำหรับการสร้างโมเดลการจำแนกคุณลักษณะของข้อมูล สามารถทำได้หลายแนวทางทางหนึ่งนั้น คือการนำข้อมูลประวัตินักศึกษามาสร้างเป็นโมเดลการจำแนกประเภทข้อมูลโดยใช้ วิธีโครงข่ายประสาทเทียม (neural network) บอกปัจจัยที่สำคัญและคลาส (class) ปลายทางแทนนักศึกษาที่ต้องการจะเข้าเรียนกับนักศึกษาที่ไม่ต้องการจะเข้าเรียน โดยผู้วิจัยได้แบ่งการดำเนินงานออกเป็นขั้นตอนดังนี้

2.1 การเก็บรวบรวมข้อมูล จะทำการเก็บรวบรวมข้อมูลประวัตินักศึกษาคณะเกษตรศาสตร์และเทคโนโลยีมหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์ โดยเก็บรวบรวมเป็นปีการศึกษา ปี พ.ศ. 2560 ได้ขนาดตัวอย่างข้อมูลที่ใช้ในการวิจัยจำนวน 1,192 คน

2.2 การเตรียมข้อมูลทำการคัดเลือกและจัดรูปแบบข้อมูลให้เหมาะสมกับโครงข่ายประสาทเทียม (neural network) ด้วยเทคนิคการทำเหมืองข้อมูล โดยมีวิธีการคัดเลือกข้อมูลดังนี้

2.2.1. การคัดเลือกข้อมูล (data selection) จุดประสงค์หลักคือ การระบุลักษณะข้อมูลที่ต้องการแล้วทำการคัดเลือกข้อมูลที่ต้องการ ซึ่งข้อมูลที่ได้จะแตกต่างกันไปตามจุดประสงค์ของแต่ละปัจจัย

2.2.2. การกลั่นกรองข้อมูล (data cleaning) จุดประสงค์เพื่อมั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นถูกต้องและเหมาะสม เนื่องจากข้อมูลที่ถูกเลือกมาจากการบวนการเลือกข้อมูลนั้นอาจมีข้อมูลไม่ถูกต้อง

2.2.3. การแปลงรูปข้อมูล (data transformation) จุดประสงค์เพื่อแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมจะนำไปวิเคราะห์ตามอัลกอริทึมของการทำเหมืองข้อมูลที่ใช้ เช่น การแบ่งช่วงอายุให้เป็นกลุ่ม ๆ หรือ กำหนดตัวเลขให้กับแต่ละกลุ่มของข้อมูล ดังแสดงใน Table 1 และ Table 2

Table 1 Variables used in model

Variables	Description
sex	sex of student
age	age of student
distance	distance
status	status of parent
income	total income of parent
occupation	occupation of parent
major_of_study	major of study
education_Level	education level

Table 2 Initial variable value

Variables	Value
sex: male, female	0,1
age	real number
distance	real number
status:	
live together	01
separation	02
father remarried	03
mother remarried	04
father and mother remarried	05
divorce	06
parents die	07
income	real number
occupation:	
government service	01
government employee	02
business person	03
farmer	04

Variables	Value
work for hire	05
unemployed	06
others	07
major of study: technology computer, mechanical engineering, and others	1,0,...
education level: bachelor	0
certification	1

เมื่อแทนค่าข้อมูลด้วยตัวเลขและแปลงให้สามารถนำไปใช้ในตัวแบบ โดยผลลัพธ์ของข้อมูลนำเข้าที่ประกอบด้วย เพศ อายุ ระยะทาง สถานภาพของบิดามารดา รายได้รวมของครอบครัว อาชีพ สาขาวิชา ระดับชั้น คือนักศึกษาที่จะเข้าศึกษาและนักศึกษาที่ไม่เข้าศึกษา

3. วิเคราะห์ปัจจัยที่มีอิทธิพลต่อการเข้าศึกษา

ขั้นตอนการทำงานโครงสร้างตัวแบบการฝึกสอนข้อมูล ตั้งแต่ พ.ศ. 2557 ถึง พ.ศ. 2561 มีจำนวนทั้งหมด 1,192 คน ประกอบด้วย เพศ อายุ ระยะทาง สถานภาพของครอบครัว รายได้รวมของครอบครัว อาชีพของผู้ปกครอง สาขา และระดับชั้น

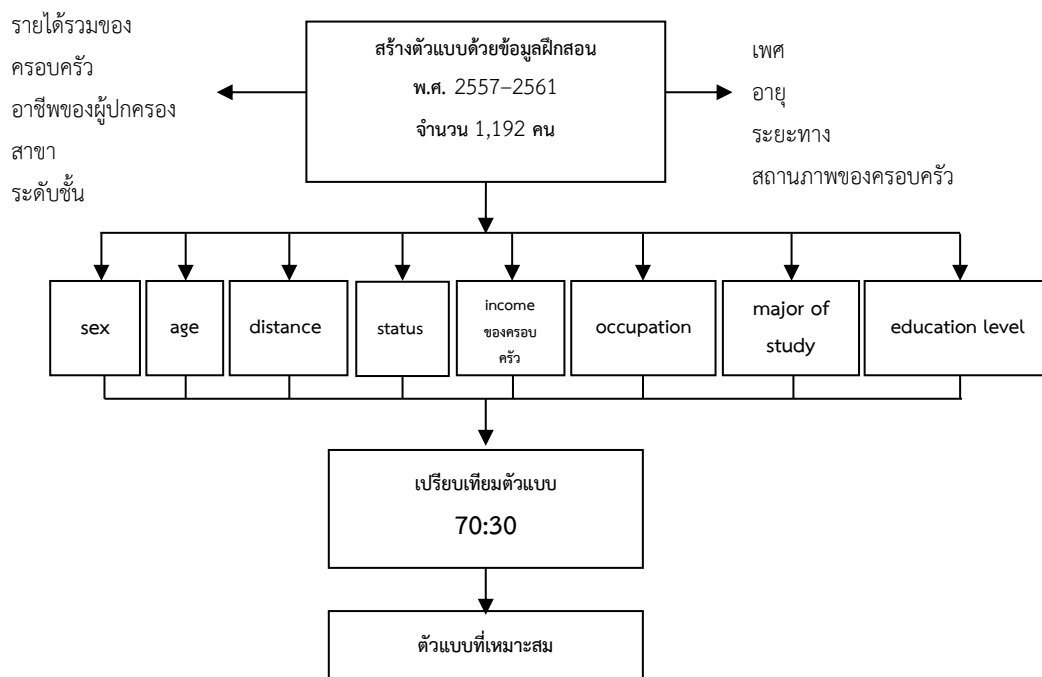


Figure 4 Factor analysis diagram

ขั้นตอนการวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการเข้าศึกษา ดังแสดงใน Figure 4 สามารถอธิบายขั้นตอนการทำงานได้ดังนี้

3.1 กำหนดปัจจัยที่จะทำการฝึกสอน ได้แก่ เพศ อายุ ระยะทาง สถานภาพของครอบครัว รายได้รวมของครอบครัว อาชีพของผู้ปกครอง สาขา ระดับชั้น ตั้งแต่ปี พ.ศ. 2557 ถึง 2561 จำนวน 1,192 คน

3.2 ทำการป้อนอินพุตชุดข้อมูลฝึกสอนให้แก่โครงข่ายและทำการคำนวณเอาต์พุตโดยเริ่มต้นจากข้อมูลแรก

3.3 กำหนดอัตราการเรียนรู้ของโครงข่ายเป็น 70:30 จำนวนการสอน 500 รอบ โดยใช้โครงสร้างแบบจำลองที่ใช้ในการสอน 8-19-2

3.3.1 ทำการป้อนชุดข้อมูลจำนวนทั้งหมดคือ 8 ปัจจัยและดูผลลัพธ์ที่ได้จากค่าความถูกต้อง ค่า MAE และ ค่า RMSE

3.3.2. ทำการลดข้อมูลนำเข้าที่ละปัจจัย และดูผลลัพธ์ที่ได้จากค่าความถูกต้อง ค่า MAE และ ค่า RMSE

3.4 ทำการเปรียบเทียบผลลัพธ์ที่ได้จากอัตราการเรียนรู้ 70:30

3.5 เลือกตัวแบบที่เหมาะสมและมีผลลัพธ์ที่ดีที่สุดขั้นตอนการสอน (training)

3.5.1 นำข้อมูลเข้า (input) ที่มีการแปลงเรียบร้อยแล้ว ขั้นตอนการสอนต้องมีชุดข้อมูลเป้าหมายเพื่อให้ตัวแบบเรียนรู้ค่าเป้าหมาย และต้องทำการแปลงให้อยู่ในรูปแบบเดียวกันเมื่อได้ข้อมูลนำเข้า (input) และชุดข้อมูลเป้าหมายที่แปลงค่าแล้วตามตารางที่ 3 ก็สามารถทำการสอนด้วยตัวแบบ โดยใช้อัตราการเรียนรู้และการทดสอบด้วยอัตราทดสอบ 70:30 โดยมีการกำหนดจำนวนการสอน 500 รอบ โดยโครงสร้างแบบจำลองที่ใช้ในการสอนคือ 8-19-2

4. วิธีการที่ใช้ในการวิจัย

ในการวิเคราะห์ข้อมูลโดยทำเหมืองข้อมูลมีอัลกอริทึมที่ใช้ในการวิเคราะห์มากมาย ซึ่งได้พิจารณาเลือกใช้เทคนิคเครือข่ายประสาทเทียมแบบหลายชั้น (multi-layer neural network) มาทำการสอนด้วยเทคนิควิธีแบ็คพรอพพาเกชัน (back propagation) เพราะเป็นเทคนิคที่นิยมใช้ในการแบ่งประเภทของข้อมูลเนื่องจากความสามารถในการจำลองพฤติกรรมทางกายภาพของระบบที่มีความซับซ้อนจากข้อมูลที่ป้อนให้เรียนรู้ งานวิจัยนี้ได้เลือกใช้โปรแกรม WEKA ในการสร้างแบบจำลองเพื่อการวิเคราะห์ เพราะเป็นโปรแกรมที่มีความสามารถสูง และให้ผลลัพธ์ที่มีประสิทธิภาพจากหลาย ๆ งานวิจัยที่ได้นำโปรแกรม WEKA มาใช้ผลลัพธ์ที่ได้มีประสิทธิภาพและแม่นยำ

3. ผลการวิจัย

ผลจากการดำเนินงานวิจัยเมื่อนำข้อมูลเข้ามีการแปลงค่าเรียบร้อยแล้ว จึงนำข้อมูลเข้าขั้นตอนการสอนต้องมีชุดข้อมูลเป้าหมายเพื่อให้ตัวแบบเรียนรู้ค่าเป้าหมาย โดยการแปลงค่าข้อมูลให้อยู่ในรูปแบบเดียวกัน ใช้อัตราการเรียนรู้และการทดสอบด้วยอัตรา 70:30 โดยมีการกำหนดจำนวนการสอน 500 รอบ โดยโครงสร้างแบบจำลองที่ใช้ในการสอน คือ 8-19-2 จำนวนปัจจัยนำเข้า 8 จำนวนชั้นซ่อน 19 ชั้นและจำนวนชั้นผลลัพธ์ 2 ชั้นเป็นผลลัพธ์ที่ต้องการ YES คือ นักศึกษาที่จะเข้าศึกษาที่จะเข้าศึกษา NO คือ นักศึกษาที่ไม่เข้าศึกษา จำนวนข้อมูลการสอน 805 และจำนวนข้อมูลทดสอบเท่ากับ 387 คนได้ดังแสดงใน Figure 5 และแสดงผลค่าความถูกต้อง ดังแสดงใน Table 3

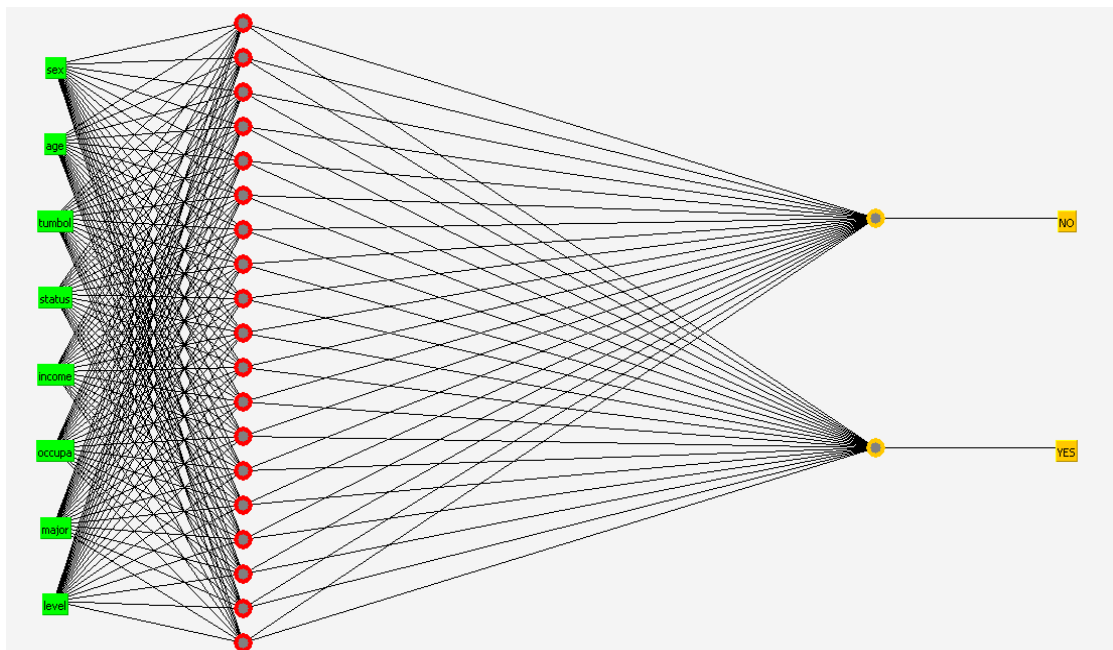


Figure 5 Model structure

Table 3 Factor classification

Classification results	Tested with a 30%	
	Amount of data	Accuracy percentage
accuracy	320	82.74%
error	67	17.52%
MAE	0.2179	
RMSE	0.3696	

จาก Table 3 สรุปผลการทดสอบตัวแบบด้วยการกำหนดอัตราการเรียนรู้ด้วยตัวแบบ 70:30 ใช้โครงสร้าง 8-19-2 ผลที่ได้จากการทดสอบมีความถูกต้อง 82.74 เปอร์เซ็นต์ โดยวัดค่าความคลาดเคลื่อนด้วยวิธี MAE ได้ 0.2179 และ RMSE ได้ 0.3696

4. สรุปผลวิจัยและอภิปรายผล

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อสร้างและทดสอบโมเดล เพื่อหาโมเดลที่เหมาะสมจากค่าความคลาดเคลื่อนวิธีค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) โดยใช้ความสามารถของโครงข่ายประสาทเทียมในการวิเคราะห์ปัจจัย ด้วยเทคนิคแบ็คพรอปพาเกชัน (back propagation neural network) ในอัตราการเรียนรู้ของข้อมูลฝึกสอนและทดสอบ 70:30 ให้ผลการทดสอบตัวแบบเป็นที่ยอมรับได้จากงานวิจัยยังต้องมีสิ่งที่ต้องดำเนินการเพิ่มเติม คือการหาค่าความคลาดเคลื่อนของแต่ละปัจจัยเพื่อนำมาเปรียบเทียบกับตัวแบบที่ได้ออกแบบไว้ การได้โมเดลที่เหมาะสมกับข้อมูลจะสามารถหาปัจจัยที่ส่งผลกระทบต่อการศึกษาต่อ ซึ่งจะได้นำปัจจัยแต่ละตัวมาวิเคราะห์ในลำดับถัดไป และควรมีการเปรียบเทียบกับตัวแบบอื่น ๆ เพื่อหาค่าความถูกต้องของการเรียนรู้ของข้อมูลมากยิ่งขึ้น

การนำความสามารถของโครงข่ายประสาทเทียมมาใช้ในการวิเคราะห์ปัจจัยนั้น สามารถที่จะทำการปรับปรุงการเรียนรู้ได้ด้วยตัวเอง และสามารถปรับรูปแบบต่าง ๆ ของโครงข่ายให้เข้ากับลักษณะของข้อมูลได้ เช่น ชนิดของโครงข่าย ฟังก์ชันการสอน ฟังก์ชันการเรียนรู้ ฟังก์ชันส่งต่อ จำนวนชั้นของโครงข่าย จำนวนโหนดภายในชั้นซ่อนแต่ละชั้น จำนวนรอบการสอน เป็นต้น ซึ่งส่วนประกอบเหล่านี้ไม่มีทฤษฎีที่กำหนดรูปแบบไว้ตายตัว ดังนั้นการนำโครงข่ายประสาทเทียมไปประยุกต์ใช้งานต้องขึ้นอยู่กับลักษณะงานที่ต้องการทำเป็นหลัก ซึ่งในการพัฒนาระบบต่างนั้นหากสามารถที่จะทำการปรับค่าความผิดพลาดเฉลี่ย MAE, RMSE ให้น้อยลงเข้าใกล้ค่าศูนย์ได้มากเท่าใดก็จะส่งผลทำให้ผลลัพธ์ที่ได้มีค่าใกล้เคียงกับค่าจริงมากขึ้นตามไปด้วย

5. กิตติกรรมประกาศ

ขอขอบคุณ คณะเกษตรศาสตร์และเทคโนโลยีที่สนับสนุนทุนวิจัยครั้งนี้ สาขาเทคโนโลยีคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี วิทยาเขตสุรินทร์ บุคลากรทุกภาคส่วนที่มีส่วนร่วมให้การดำเนินการวิจัยสำเร็จลุล่วงไปตามเป้าหมาย

REFERENCES

- Chaiwachiragompol, A. and Suwannata, N. (2016). The features extraction of infants cries by using discrete wavelet transform techniques, international electrical engineering congress. *Procedia Computer Science*, 86, 285–288. <https://doi.org/10.1016/j.procs.2016.05.073>.
- Intawee, T. (2021). Factors in affecting decision making in selecting educational institutions for higher education. *Journal of Educational Innovation and Research*. 5(1), 1–14. [In Thai].
- Jabjone, S. (2015). Data mining. Nakhon Ratchasima: Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University. [In Thai].
- Kaewwijit, T. (2016). The improvement of support vector regression to forecast time series. (Master's Thesis). Nakhon Ratchasima: Suranaree University of Technology. [In Thai].
- Kanjanasamranwong, P., Urawong, K., Mad A Dam, R., and Mad A Dam, S. (2017). The opportunities forecast model in choosing to attend the Prince of Songkhla University of grade 12 students, Songkhla Province. *Journal of Education Naresuan University*. 19(2), 12–24. [In Thai].
- Kijsirikul, B. (2003). Artificial Intelligence. Bangkok: Faculty of Engineering, Chulalongkorn University. [In Thai].
- Lorpunmanee, S., Chimphee, S., Chimphee, W., Nedtharn, W. and Piromnok, N. (2019). The forecasting patterns of juvenile recidivism. *Journal of Science Innovation for Sustainable Development*. 1(1), 80–95. [In Thai].
- Ngamlamom, W. (2022). The human resource management of public sector in the disruption era. *University of the Thai Chamber of Commerce Journal Humanities and Social Sciences*. 42(1), 176–197. [In Thai].
- Pacharawongsakda, E. (2014). An Introduction to data mining techniques. Bangkok: Data cube. [In Thai].
- PimPa, P. (2018). Current Thai studies. *Academic Journal of Mahamakut Buddhist University Roi Et Campus*. 7(1), 242–249. [In Thai].
- Rahman, H. (2009). Data mining applications for empowering knowledge societies. New York: Information Science Reference. P.12.
- Roiger, R. J. and Geatz, M. W. (2003). Data mining: A tutorial-based primer. Boston: Addison Wesley. P.4.
- Supabut, P., Hengprapohm, K., Hengprapohm, S., and Thammasiri, D. (2015). The comparison of prediction efficiency between linear regression and back-propagation neural network method. joint conference on ACTIS & NCOBA 2015. [In Thai].
- Taesombat, S. (1996). Quantitative forecasting techniques. Bangkok: physics-center. [In Thai].
- Treepuech, W. (2019). The analysis of injured and dead people data from accident in Songkran Festival by using data mining techniques. *VRU Research and Development Journal, Science and Technology*. 14(1), 11–20. [In Thai].
- Visutthirat, T. (2018). Using data mining in an analysis and model building of the relationship between social quality and happiness of population in chanthaburi province. (Doctoral dissertation). Bangkok: National Institute of Development Administration (NIDA). [In Thai].

- Wanon, S. and Muangsan, R. (2020). A study and development of forecasting model for the suitability characteristics on the applying major selection by using data mining techniques. *Journal of Management Sciences*. 7(1), 135–152. [In Thai].
- Winston, P. H. (1992). *Artificial Intelligence*. Addison-Wesley Publishing Company. (3rd ed.). Massachusetts. US. 443–469.
- Witten, I. H. and Frank, E. (2005). *Data mining: Practical machine learning tools and technique* (2nd ed.). San Francisco: Morgan Kaufmann Publishers. US. P.5.