อภินันทนาการ

สำนักวิจัยฯ

สัญญาเลขที่ *R2560c102*

รายงานวิจัยฉบับสมบูรณ์

การเปรียบเทียบประสิทธิภาพแบบจำลองพยากรณ์ความเสี่ยงการสูญเสียเงินกู้กยศ. โดยวิธีฟัซซีนิวรอลเน็ตเวิร์คกับวิธีวิเคราะห์สมการถดถอยเชิงเส้นพหุคูณ

สำนักหอสมุด มหาวิทยาลัยนเรศวร
วันลงทะเบียน 05 ส.ค. 2564
เลขทะเบียน 1034770
เลขเรียกหนังสือ จ QA
9.64
ท651ร
2563

คณะผู้วิจัย สังกัด

ผู้ช่วยศาสตราจารย์เทวิน ธนะวงษ์ ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

สนับสนุนโดยกองทุนวิจัยมหาวิทยาลัยนเรศวร

กิตติกรรมประกาศ

งานวิจัยฉบับนี้ได้รับเงินอุดหนุนงบประมาณรายได้ มหาวิทยาลัยนเรศวร ประจำปีงบประมาณ 2560 ได้สำเร็จลุล่วงด้วยความกรุณาอย่างยิ่งของภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ที่เอื้อเฟื้อทรัพยากรต่าง ๆ ที่นำมาใช้เป็นส่วนหนึ่งของงานวิจัย จนทำให้การศึกษาวิจัยสำเร็จลุล่วงสมบูรณ์ได้ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณคณะผู้บริหาร ตลอดจนเจ้าหน้าที่ฝ่ายงานวิจัย และบุคลากรของภาควิชาวิทยาการคอมพิวเตอร์ฯ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร รวมทั้งหน่วยงานของภาครัฐและภาคเอกชน และหน่วยงานต่าง ๆ ที่ได้ให้ความอนุเคราะห์อำนวยความสะดวกและให้ความร่วมมือ พร้อมทั้งให้คำปรึกษาในส่วนของการดำเนินการด้านเอกสารต่าง ๆ ให้ถูกต้องจนสำเร็จสมบูรณ์

คุณค่าและประโยชน์อันพึงมีจากการศึกษาค้นคว้าฉบับนี้ ผู้วิจัยขออุทิศแด่ผู้มีพระคุณทุก ๆ ท่าน

ผู้จัดทำวิจัย

บทคัดย่อ

งานวิจัยนี้นำเสนอการสร้างแบบจำลองเพื่อสนับสนุนการทำงานวิเคราะห์ความเสี่ยงการสูญเสียเงินกู้ กยศ. โดยการเปรียบเทียบเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค กับวิธีวิเคราะห์สมการถดถอยเชิงเส้นพหุคูณ การพัฒนาเริ่มต้นด้วยการเตรียมข้อมูลนำเข้าจำนวน 296 ตัวอย่าง ซึ่งกำหนดให้ตัวแปร Incomeมีคุณลักษณะเป็นตัวแปรเชิงฟัซซี และมีการแก้ปัญหาข้อมูลเชิงคุณภาพด้วยการทำเป็นตัวแปรดัมมี นอกจากนี้ได้นำเสนอวิธีการแปลงข้อมูลเชิงฟัซซีด้วยการทำ Fuzzy attribute matching ที่คำนวณค่าระหว่างพจน์ภาษาฟัซซี กับพจน์ภาษาฟัซซีและพจน์ภาษาฟัซซีกับค่าคริสป์ แล้วคัดกรองตัวแปรด้วยเทคนิคสหสัมพันธ์เพียร์สันและสมการถดถอยเชิงเส้นพหุคูณ ได้ตัวแปรอิสระจำนวน 8 ตัว และมี RiskForPay เป็นตัวแปรตาม ผลลัพธ์ที่ได้จากงานวิจัยนี้ จะพบว่าโมเดลที่เหมาะสมของฟัซซีนิวรอลเน็ตเวิร์ค มีการแบ่งข้อมูลเรียนรู้และทดสอบแบบ Cross validation fold=5 มีค่าประสิทธิภาพความถูกต้องเท่ากับ 83.33% +/- 6.02% ซึ่งโมเดลมีโครงสร้างรูปแบบ 8-5-6 ค่าโมเมนตัม 0.2 และค่าอัตราการเรียนรู้ 0.3 สำหรับโมเดลแบบพยากรณ์ด้วยสมการถดถอยเชิงพเส้นพหุคูณนั้นมีค่า Root Mean Squared Error: 1.513 +/- 0.000 และค่า Squared Correlation: 0.081 ตามลำดับ

คำสำคัญ: โครงข่ายประสาทเทียม, สมการถดถอยเชิงเส้นพหุคูณ, กองทุนเงินให้กู้ยืมเพื่อการศึกษา(กยศ.), ตรรกศาสตร์คลุมเครือ

Abstract

This research presents the comparing the performance of the prediction model of the risk of losing student loan by Fuzzy Neural Network (FNN) and Multiple Linear Regression Analysis (MLR). We used the preparation of 296 samples, assigned the income variables, which are characterised by Fuzzy attribute and had problem-solving with Dummy variables. Also, we proposed the method of converting fuzzy data by making a fuzzy attribute matching. It calculates the fuzzy linguistic term, the fuzzy language and the Crisp value, and after that have filtered variables with Pearson correlation technique and multiple linear regression which get eight independent variables and RiskForPay is a dependent variable. Results from this research, the appropriate model of Fuzzy Neural Network is a division of learning data, and cross-validation fold is 5 with an accuracy of 83.33% +/- 6.02% which models have 8-5-6 structure, momentum 0.2 and learning rate 0.3. The predictive model with multiple linear regression equations has Root Mean Squared Error: 1.513 +/- 0.000 and Squared Correlation: 0.081, respectively.

Key words: Fuzzy Neural Network, Multiple Linear Regression Analysis, Student Loan, Fuzzy logic

สารบัญเรื่อง

หน้า

บทที่1 บทนำ (Introduction) 7

1.1 ที่มาของปัญหา 7

1.2 วัตถุประสงค์ 8

1.3 ขอบเขตการวิจัย 8

1.4 วิธีการดำเนินการวิจัยและระยะเวลาทำวิจัย 8

1.5 ประโยชน์ที่คาดว่าจะได้รับ 9

บทที่2 งานวิจัยที่เกี่ยวข้องและเอกสารทางทฤษฎี 11

2.1 กองทุนกู้ยืมเพื่อการศึกษา (กยศ.) 11

2.2 ตรรกศาสตร์คลุมเครือ (Fuzzy logic) 13

2.3 สมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) 14

2.4 โครงข่ายประสาทเทียม (Neural Network) 16

2.5 โปรแกรม RapidMiner Studio 17

2.6 การทบทวนวรรณกรรมที่เกี่ยวข้อง 18

บทที่3 วิธีดำเนินการวิจัย 21

3.1 สถาปัตยกรรมของระบบการทำงานในงานวิจัย 21

3.2 การจัดตรียมข้อมูลการกู้ยืมทุน กยศ 21

3.3 การคัดเลือกปัจจัยนำเข้า 24

3.4 การสร้างตัวแปรดัมมีจากข้อมูลเชิงคุณภาพสู่ข้อมูลเชิงปริมาณ 26

3.5 การกำหนดค่าฟัซซีแอททริบิวต์ (Fuzzy attribute values) 28

3.6 การสร้างแบบจำลองด้วยเทคนิคสมการถดถอยเชิงเส้นพหุคูณ (Multiple regression) 29

3.7 การทำฟัซซีแอททริบิวต์แมทชิง (Fuzzy attribute matching) 33

3.7.1 การทำฟัซซีแอททริบิวต์แมทชิงระหว่าง พจน์ภาษาฟัซซี (Fuzzy linguistic term) กับพจน์ภาษาฟัซซี(Fuzzy linguistic term) 33

3.7.2 การทำฟัซซีแอททริบิวต์แมทชิงระหว่าง ค่าคริสป์ (crisp value)กับพจน์ภาษาฟัซซี (fuzzy linguistic term) 34

3.8 หลักการทำงานของ Neural network 35

3.8.1 การสร้างแบบจำลองด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค 36

บทที่4 ผลการทดลอง.....43
4.1 การทดสอบประสิทธิภาพของตัวแบบจำลองฟัซซีโครงข่ายประสาทเทียม
และตัวแบบจำลองสมการถดถอยเชิงเส้นพหุคูณ.....43
4.2 การคำนวณผลลัพธ์ของตัวแบบพยากรณ์.....44

สรุปผลการวิเคราะห์และวิจารณ์.....50
บรรณานุกรม.....51
ภาคผนวก ก.....52
ภาคผนวก ข.....74

มหาวิทยาลัยนเรศวร

บทที่1
บทนำ (Introduction)

1.1 ที่มาของปัญหา

สังคมไทยในยุคปัจจุบัน จะพบว่ามีครอบครัวจำนวนมากที่ตระหนักถึงความสำคัญต่อการให้การศึกษาของบุตรหลานตนเอง โดยเฉพาะอย่างยิ่งการส่งเสริมให้รับการศึกษาต่อคณะต่าง ๆ ในระดับอุดมศึกษา อย่างไรก็ตาม เศรษฐกิจในยุคปัจจุบันที่มีค่าครองชีพที่สูงขึ้น ทำให้มีค่าใช้จ่ายภายในครอบครัวยิ่งสูงขึ้นตามลำดับ ส่งผลให้บางครอบครัวหรือจำนวนหลายๆ ครอบครัวเลือกที่จะยุติการส่งบุตรหลานให้ได้รับการศึกษาที่สูงขึ้น หรืออาจทำให้ลดจำนวนการส่งเสริมให้ศึกษาต่อ เป็นต้น อย่างไรก็ตาม หน่วยงานของภาครัฐได้ตระหนักและเล็งเห็นปัญหาอุปสรรคที่เกิดขึ้นและหลักการที่จะช่วยเหลือนักเรียน นิสิตและนักศึกษาในสถาบันการศึกษาต่าง ๆ ทั่วประเทศ โดยได้กำหนดกฎเกณฑ์ต่าง ๆ เพื่อสร้างเป็นกลไกของการกู้ยืมขึ้น ซึ่งกองทุนเงินให้กู้ยืมเพื่อการศึกษาหรือที่เรียกว่า กยศ. นั้นได้ดำเนินการมาระยะเวลาหนึ่ง จะพบว่าปัญหาอุปสรรคที่เกิดขึ้นของการดำเนินงานนั่นก็คือการขาดวิธีช่วยคัดกรองเด็กที่มีคุณภาพในการชำระหนี้เงินกู้ จึงทำให้เกิดสูญเสียเงินจากการขาดการชำระเงินคืนของบัณฑิตจากสถาบันการศึกษาต่าง ๆ จำนวนมาก ซึ่งผลที่เกิดขึ้นนี้มีผลกระทบต่อการดำเนินการกู้ยืมของนักศึกษาที่เป็นรุ่นน้องของแต่ละสถาบันการศึกษานั้น ๆ ทั้งนี้ปัญหาดังกล่าวนี้มิได้รับการแก้ไขอย่างจริงจังรวมทั้งการหาแนวทางในการประเมินความเสี่ยงที่จะเกิดขึ้นของการสูญเสียเงินจากการกู้ยืมของสถาบันภายใต้คณะต่าง ๆ ยังไม่สามารถดำเนินอย่างเป็นรูปธรรมเช่นกัน ซึ่งกระบวนการนการคัดกรองการกู้ยืมนั้นจะพบว่าเป็นเพียงการซักถามประวัติและดูเอกสารประกอบการพิจารณาที่เป็นข้อมูลพื้นฐานของนิสิต นักศึกษาที่จะกู้ยืมเป็นต้น การจัดสรรจำนวนเงินให้แก่คณะต่าง ๆ นั้นยังมิได้มีหลักเกณฑ์หรือระบบที่จะช่วยประเมินและคาดการณ์ทั้งการได้งานทำของบัณฑิตรวมทั้งข้อมูลพื้นฐานของบัณฑิตที่ดำรงอยู่ในสถานะนิสิต นักศึกษานำมาวิเคราะห์ระดับความสามารถในการชำระเงินกู้ เป็นต้น อย่างไรก็ตาม ข้อมูลและปัจจัยที่เกี่ยวของบัณฑิตที่จัดเก็บไว้ในหน่วยงานกองบริการการศึกษาของสถาบันในระดับอุดมศึกษานั้น หากนำมาใช้หลักการทำเหมืองข้อมูลด้วยการเลือกใช้อัลกอริธึมต่าง ๆ และการใช้หลักการทางคณิตศาสตร์ภายใต้การใช้ปัจจัยของตัวแปรอิสระกับตัวแปรตามที่อยู่นรูปของสมการถดถอยเชิงเส้นพหุคูณเพื่อสร้างแบบจำลองการพยากรณ์ความเสี่ยงที่จะสูญเสียเงินกู้ยืมเพื่อการศึกษาย่อมจะทำให้เกิดความน่าเชื่อถือมากยิ่งขึ้น

ดังนั้น ผู้วิจัยจึงเล็งเห็นความสำคัญที่จะประยุกต์ใช้เทคโนโลยีเพื่อสร้างแบบจำลองในการพยากรณ์และวิเคราะห์ความเสี่ยงของการสูญเสียเงินกู้ พร้อมทั้งการประยุกต์ใช้แบบจำลองที่พัฒนาสร้างแอพพลิเคชันเพื่อให้บุคลากรที่เกี่ยวข้องกับการบริหารจัดการการกู้ยืมเงินเพื่อการศึกษา ได้ใช้เป็นแนวทางช่วยในการตัดสินใจคัดเลือกนิสิต นักศึกษาที่จะกู้ยืมเงินได้อย่างเหมาะสมมากยิ่งขึ้น ทั้งนี้ผลลัพธ์ที่ได้จากการใช้งานของแอพพลิเคชันนั้น จะช่วยคาดการณ์ความสามารถในการจะชำระหนี้ที่จะเกิดขึ้นในอนาคตของผู้กู้ยืมโดยจะสอดคล้องกับเหตุปัจจัยของบัณฑิตรุ่นพี่ที่มีวิธีการชำระหนี้เงินกู้ เพราะฉะนั้น ผู้วิจัยจึงสนใจที่จะนำเสนองานวิจัย ภายใต้หัวข้อเรื่องการเปรียบเทียบประสิทธิภาพแบบจำลองพยากรณ์แนวโน้มสูญเสียการชำระเงินกู้กยศ. โดยวิธีฟัซซีนิวรอลเน็ตเวิร์คกับวิธีวิเคราะห์สมการถดถอยเชิงเส้นพหุคูณ โดยได้พัฒนาโมเดลฟัซซีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและสมการถดถอยเชิงเส้นพหุคูณพร้อมทั้งเปรียบเทียบประสิทธิภาพการใช้งานที่เหมาะสม เพื่อนำไปประยุกต์สำหรับปรับกลยุทธ์ในการจัดการบริหารกองทุนกู้ยืมให้เหมาะสมกับการชำระหนี้คืนของบัณฑิตให้สอดคล้องกับนักศึกษาในคณะต่าง ๆ ของมหาวิทยาลัยนเรศวรให้มีประสิทธิภาพที่ดีขึ้น ทั้งนี้ในการดำเนินการวิจัยจะประกอบไปด้วย ส่วนต่าง ๆ นั่นคือ ส่วน

ของการจัดเตรียมข้อมูลองค์ความรู้ที่เกี่ยวข้องกับปัจจัยพื้นฐานของบัณฑิตขณะดำรงสถานะนิสิตในมหาวิทยาลัยนเรศวรและการประกอบอาชีพในอนาคตภายใต้ลักษณะของ Training data sets และ Testing data sets โดยอาศัยเทคนิคการ Classification ด้วยอัลกอริธึมฟัซซีโครงข่ายประสาทเทียมและการจำลองสมการถดถอยเชิงเส้นพหุคูณ พร้อมทั้งทำการปรับปรุงโมเดลในการเรียนรู้ปัญหาเพื่อดำเนินการพยากรณ์ภายใต้แอพพลิเคชันที่พัฒนาขึ้นสำหรับใช้ปฏิบัติงานได้จริงลักษณะ Responsive web application ทั้งรายบุคคล รายคณะ

1.2 วัตถุประสงค์

1) เพื่อใช้เทคนิคเหมืองข้อมูลและสมการทางคณิตศาสตร์วิเคราะห์และพยากรณ์ความเสี่ยงของสูญเสียเงินจากการชำระหนี้กองทุนเงินให้กู้ยืมเพื่อการศึกษาของนิสิตมหาวิทยาลัยนเรศวร

2) เพื่อเปรียบเทียบประสิทธิภาพการใช้อัลกอริธึมโครงข่ายประสาทเทียมกับสมการถดถอยเชิงเส้นพหุคูณที่เหมาะสมสำหรับวิเคราะห์และพยากรณ์ความเสี่ยงของสูญเสียเงินจากการชำระหนี้กองทุนเงินให้กู้ยืมเพื่อการศึกษา

3) เพื่อพัฒนาแอพพลิเคชันที่ทำงานในลักษณะ Responsive Web Application สำหรับใช้พยากรณ์ผลความเสี่ยงรายบุคคล และรายคณะของการสูญเสียเงินจากการชำระหนี้กองทุนเงินให้กู้ยืมเพื่อการศึกษาได้

1.3 ขอบเขตการวิจัย

1.3.1 ส่วนของขอบเขตแหล่งข้อมูล

- รวบรวมข้อมูลและทำการคัดเลือกตัวแปรด้วยวิธีการจัดเก็บข้อมูลแบบปฐมภูมิและทุติยภูมิ จำนวน 296 ตัวอย่าง
- นักศึกษา 4 คณะ ที่ศึกษาสาขาวิทยาศาสตร์และเทคโนโลยี ได้แก่ คณะวิทยาศาสตร์ คณะวิศวกรรมศาสตร์ คณะเกษตรศาสตร์ และคณะสถาปัตยกรรมศาสตร์ มหาวิทยาลัยนเรศวร โดยจัดเก็บข้อมูลจากนักศึกษาที่เคยศึกษาและกู้ยืม กยศ. และได้สำเร็จการศึกษาแล้ว

1.3.2 ส่วนของการใช้เทคโนโลยีหลัก

- การใช้เทคนิค โครงข่ายประสาทเทียบเรียนรู้แบบแพร่ย้อนกลับ และต้นไม้ตัดสินใจเพื่อสร้างแบบจำลองบนโปรแกรม RapidMiner Studio 6.0
- การใช้เทคนิคตรรกศาสตร์เชิงคลุมเครือ (Fuzzy logic)
- โปรแกรม IBM SPSS Statistics22 สำหรับจัดเก็บข้อมูลและคัดเลือกตัวแปร

1.4 วิธีการดำเนินการวิจัย และ ระยะเวลาทำการวิจัย

ในการทำวิจัยครั้งนี้ จะใช้อุปกรณ์และเครื่องมือภาควิชาวิทยาการคอมพิวเตอร์ฯ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

1) รวบรวมข้อมูลพื้นฐานของบัณฑิตในการสมัครกู้ยืมเงินกยศ ขณะสถานะเป็นนิสิตของมหาวิทยาลัยนเรศวร

2) ดำเนินการจำแนกและวิเคราะห์ความสัมพันธ์ของข้อมูลเพื่อใช้ในการจัดเตรียมข้อมูลนำเข้าสู่ระบบและการจัดเก็บข้อมูลเพิ่มเติมเพื่อความสมบูรณ์ของการดำเนินการวิจัย

3) การสร้างโมเดลสมการถดถอยเชิงเส้นพหุคูณและโมเดลฟัซซีโครงข่ายประสาทเทียมพร้อมทั้งเปรียบเทียบประสิทธิภาพ

4) การทดสอบการทำงานของระบบ และการปรับปรุงแก้ไข

5) เขียนรายงาน นำเสนอผลงานวิจัย จัดพิมพ์รูปเล่มพร้อมส่งผลงานที่เสร็จสมบูรณ์

ตารางแสดงแผนงานและผลงาน (ในแต่ละช่วงเวลา) นับตั้งแต่เริ่มโครงการจนถึงสิ้นสุดโครงการ

กิจกรรม	เดือนที่											
	1	2	3	4	5	6	7	8	9	10	11	12
1.ศึกษาค้นคว้า บทความ และทฤษฎีที่เกี่ยวข้อง	√	√										
2.เก็บรวบรวมข้อมูลภาค สนาม และในหน่วย งานต่าง ๆ ที่เกี่ยวข้อง		√	√	√	√	√	√					
3.จำแนกข้อมูลและวิเคราะห์ข้อมูล และเพิ่มเติมการจัดเก็บข้อมูลเพื่อความสมบูรณ์ของระบบ			√	√	√	√	√	√				
4.สร้างโมเดลของระบบ การสอนข้อมูลและทดสอบการทำงานของโมเดลด้วยข้อมูลโครงข่ายประสาทเทียมและสมการถดถอยเชิงเส้นพหุคูณ					√	√	√	√	√	√		
5. การทดสอบข้อมูลจริงจากตัวแบบจำเหมาะสมสำหรับใช้พยากรณ์ผลโอกาสและการสิ้นสุดการชำระเงินกู้กยศ พร้อมทั้งแก้ไขปรับปรุง						√	√	√	√	√	√	
6. เขียนรายงาน นำเสนอผลงานวิจัย จัดพิมพ์รูปเล่มพร้อมส่งผลงานที่เสร็จสมบูรณ์										√	√	√

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้ประยุกต์ใช้เทคนิควิธีฟัซซีนิวรอลเน็ตเวิร์คกับวิธีวิเคราะห์สมการถดถอยเชิงเส้นพหุคูณสร้างแบบจำลองเพื่อพยากรณ์ความเสี่ยงการสูญเสียเงินกู้ กยศ. จากนิสิตมหาวิทยาลัยนเรศวร

2) ได้แอพพลิเคชันจาการเลือกใช้แบบจำลองที่เหมาะสมช่วยในการคัดกรองนิสิตสำหรับจะได้รับการกู้ยืมเงิน กยศ ได้เหมาะสมที่สุดภายใต้ศักยภาพการชำระหนี้เงินกู้

3) ได้ประยุกต์ใช้เทคโนโลยีให้เกิดประโยชน์และเพิ่มประสิทธิภาพของการบริหารจัดการกองทุนเงินให้กู้ยืมเพื่อของหน่วยงานกลางของมหาวิทยาลัยนเรศวรและเป็นต้นแบบของการนำโปรแกรมคอมพิวเตอร์มาช่วยพยากรณ์ ความสามารถในการชำระเงินกู้ยืมของนิสิตแต่ละบุคคล และคณะ

บทที่2
งานวิจัยที่เกี่ยวข้องและเอกสารทางทฤษฎี

ทฤษฎีที่เกี่ยวข้อง

2.1 กองทุนกู้ยืมเพื่อการศึกษา (กยศ.) [11]

กองทุนเงินให้กู้ยืมเพื่อการศึกษาเป็นกองทุนหมุนเวียน ที่จัดตั้งตามมติคณะรัฐมนตรี เมื่อวันที่ 28 มีนาคม พ.ศ. 2538 และมติคณะรัฐมนตรีเมื่อวันที่ 16 มกราคม พ.ศ. 2539 ให้เริ่มดำเนินการกองทุนในลักษณะเงินหมุนเวียน จึงเริ่มเปิดให้กู้ยืมตั้งแต่ปีการศึกษา 2539 ซึ่งผลการดำเนินงานตั้งแต่ปี 2539 จนถึงปีการศึกษา 2555 รวมผู้กู้ยืมจำนวนทั้งสิ้น 4,183,604 ราย เป็นจำนวนเงินกู้ยืมทั้งสิ้น 422,131.82 ล้านบาท และที่สำเร็จการศึกษาแล้วจำนวน 3 ล้านราย โดยการดำเนินงานได้รับความร่วมมือจากสถานศึกษาทั้งภาครัฐและภาคเอกชนทั่วประเทศกว่า 4,000 แห่ง บมจ ธนาคารกรุงไทย และธนาคารอิสลามแห่งประเทศไทย สำหรับการดำเนินงานกองทุนให้กู้ยืมเพื่อการศึกษานั้น สามารถแสดงรายละเอียด (ดังตารางที่ 2.1) ของแต่ละปีการศึกษาได้ดังต่อไปนี้

ตารางที่ 2.1 แสดงรายละเอียดการกู้ยืมนแต่ละปีการศึกษา
(ที่มา: www.studentloan.or.th)

ปีการศึกษา	จำนวนผู้กู้ยืม (คน)	จำนวนเงินให้กู้ยืม (ล้านบาท)
2551	711,015	25,959.5412
2552	830,012	29,679.6307
2553	892,152	33,527.7848
2554	906,573	35,384.4129
2555	866,142	33,801.3009

สำหรับรายละเอียดของการชำระเงินคืนนั้น กองทุนเงินให้กู้ยืมเพื่อการศึกษาสามารถจำแนกกลุ่มผู้กู้ยืมภายใต้ข้อมูลในปี พ.ศ. 2555 มีรายละเอียดดังต่อไปนี้

- ผู้กู้ยืมที่ครบกำหนดชำระเงินคืนกองทุนฯ จำนวน 2,617,837 คน
- ผู้กู้ยืมกำลังศึกษาในปีการศึกษา 2555 จำนวน 882,888 คน
- ผู้กู้ยืมที่อยู่ระหว่างระยะเวลาปลอดหนี้ 2 ปี จำนวน 489,022 คน
- ผู้กู้ยืมที่ชำระปิดบัญชี จำนวน 193,857 คน

รายละเอียดของจำนวนผู้กู้ยืมและเงินกู้ยืมที่ครบกำหนดชำระหนี้คืนหากแยกตามกลุ่มการชำระ ณ ข้อมูลปี 2555 จะมีรายละเอียดดังนี้ (ดังตารางที่ 2.2)

ตารางที่ 2.2 แสดงจำนวนผู้กู้และเงินกู้ยืมครบกำหนดชำระหนี้คืน

(ที่มา: www.studentloan.or.th)

รายละเอียด	ผู้กู้ยืม (คน)	ร้อยละ	เงินกู้ยืม (ล้านบาท)			
			ได้รับชำระเงินคืน	ร้อยละ	ค้างชำระ	ร้อยละ
1.กลุ่มที่ชำระเงินคืนตามปกติ	839,550	32.07	19,839.33	32.49	-	-
2.กลุ่มที่ชำระเงินคืนบางส่วน	1,203,661	45.98	12,241.65	20.04	16,714.59	27.37
3.กลุ่มที่ถึงแก่ความตาย	25,362	0.97	71.32	0.12	217.34	0.35
4.กลุ่มที่ไม่ติดต่อชำระเงินคืน	549,264	20.98	-	-	11,986.28	19.63
รวมทั้งสิ้น	2,617,837	100.00	32,152.30	52.65	28,918.21	47.35
			61,070.51			100.00

1). ผู้กู้ยืมเงินที่มีหน้าที่ชำระเงินกู้ยืมคืนกองทุน

- ผู้กู้ยืมเงินที่สำเร็จการศึกษาหรือเลิกการศึกษาแล้วเป็นเวลา 2 ปี นับจากปีที่สำเร็จการศึกษาหรือเลิกศึกษา (ครบระยะเวลาปลอดหนี้)

- กรณีผู้กู้ยืมเงินที่ไม่กู้ต่อเนื่องติดต่อกัน 2 ปี และไม่แจ้งสถานภาพให้ผู้บริหารและจัดการเงินให้กู้ยืมทราบ จะถือว่าเป็นผู้ครบกำหนดชำระหนี้และมีหน้าที่ต้องชำระหนี้คืนกองทุน

2). หลักเกณฑ์การชำระหนี้

- ผู้กู้ยืมเงินต้องชำระเงินกู้ยืม พร้อมทั้งดอกเบี้ยหรือประโยชน์อื่นใดร้อยละ 1 ต่อปี โดยต้องคืนให้กองทุนให้เสร็จสิ้นภายใน 15 ปี นับแต่วันที่ต้องเริ่มชำระหนี้

- ให้ผู้กู้ยืมเงินชำระหนี้งวดแรก ภายในวันที่ 5 กรกฎาคม หลังจากครบระยะเวลาปลอดหนี้ โดยให้ชำระหนี้คืนเฉพาะเงินต้นในอัตราร้อยละ 1.5 ของวงเงินที่ได้กู้ยืม

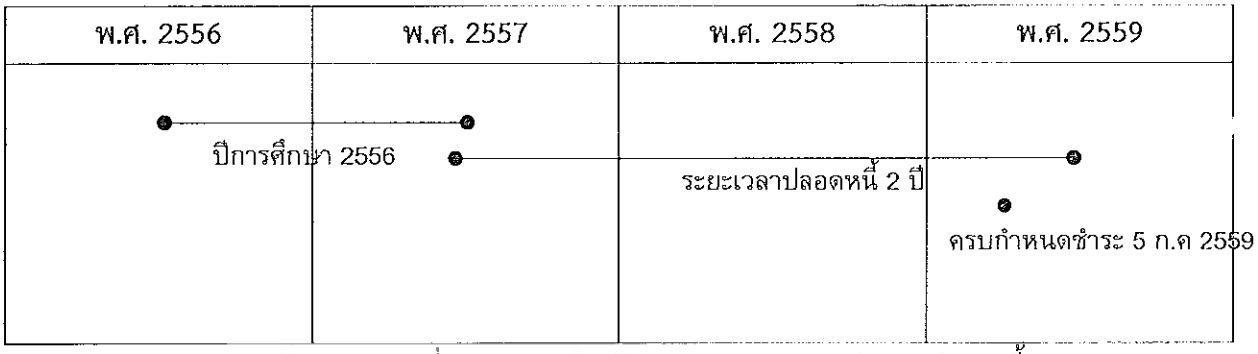
- การชำระหนี้งวดต่อ ๆ ไปให้ผู้กู้ยืมเงินชำระเงินต้นคืน ตามอัตราผ่อนชำระที่กองทุนกำหนด พร้อมด้วยดอกเบี้ยหรือประโยชน์อื่นใด ในอัตราร้อยละ 1 ต่อปีของเงินต้นที่คงค้าง ภายในวันที่ 5 กรกฎาคมของทุกปี

- หากผู้กู้ยืมเงินผิดนัดชำระหนี้ ผู้กู้ยืมเงินจะต้องชำระค่าปรับ หรือค่าธรรมเนียมจัดการกรณีผิดนัดชำระหนี้ตามอัตราที่กองทุนกำหนด

- ผู้กู้ยืมเงินที่มีความประสงค์จะขอชำระหนี้คืนก่อนครบกำหนดระยะเวลาชำระหนี้ หรือก่อนครบระยะเวลาปลอดหนี้ 2 ปี สามารถชำระคืนได้โดยไม่เสียดอกเบี้ยหรือประโยชน์อื่นใด

3). วิธีการนับระยะเวลาการครบกำหนดชำระหนี้ของผู้ที่สำเร็จการศึกษา

ตัวอย่าง: ผู้กู้ยืมเงินที่สำเร็จการศึกษาในปีการศึกษา 2556 มีระยะเวลาปลอดหนี้ 2 ปี ผู้กู้ยืมเงินจะครบกำหนดชำระหนี้คืนภายในวันที่ 5 กรกฎาคม 2559 ดังนี้



ภาพที่ 2.1 แสดงการนับระยะเวลาการครบกำหนดชำระหนี้

(ที่มา: www.studentloan.or.th)

4). อัตราการผ่อนชำระ

ผู้กู้ยืมเงินสามารถเลือกผ่อนชำระหนี้เป็นรายปีหรือรายเดือน โดยจำนวนเงินต้นที่ชำระเป็นรายปีหรือรายเดือนรวมตลอดปีหรือต้องไม่ต่ำกว่าอัตราดังนี้

ปีที่ชำระ	จำนวนร้อยละของยอดหนี้ที่กู้	ปีที่ชำระ	จำนวนร้อยละของยอดหนี้ที่กู้ยืม
1	1.5	9	7.0
2	2.5	10	8.0
3	3.0	11	9.0
4	3.5	12	10.0
5	4.0	13	11.0
6	4.5	14	12.0
7	5.0	15	13.0
8	6.0	รวม	100.0

ภาพที่ 2.2 แสดงรายละเอียดอัตราการผ่อนชำระ

(ที่มา: www.studentloan.or.th)

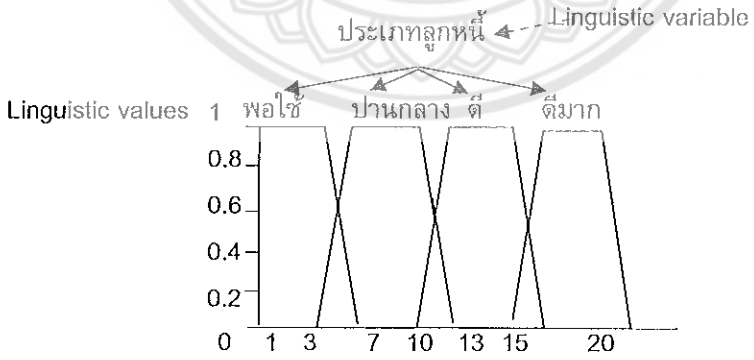
2.2 ตรรกศาสต์คลุมเครือ (Fuzzy logic)[14][15]

จากการอธิบายปัญหา (Problem description) โดยทั่ว ๆ ไปนั้น เราจะพบว่าในบางครั้งปัญหาอุปสรรคที่เกิดขึ้นของการอธิบายปัญหา ก็คือ ความคลุมเครือ (หรือเรียกว่าฟัซซี) ที่เกิดขึ้น ตัวอย่างเช่น “จำนวนเงินกู้สูงมาก” หรือ “จำนวนเงินกู้ประมาณ200,000 บาท” เป็นต้น จากสองประโยคข้างต้นภายใต้คีย์เวิร์ดของคำว่า “สูงมาก” และ “ประมาณ” ถือได้ว่าเป็นลักษณะค่าที่คลุมเครือ

อย่างไรก็ตาม ในงานวิจัยนี้ผู้วิจัยสนใจที่จะนำหลักการทำงานของ ฟัซซีลอจิก (Fuzzy logic) ซึ่งเป็นเครื่องมือที่ช่วยในการตัดสินใจภายใต้ความไม่แน่นอนหรือคลุมเครือของข้อมูลมาประยุกต์ใช้งาน ซึ่งโดยทั่วไปแล้วเราจะพบว่าหลักการทำงานของฟัซซีเซตจะถูกนำมาใช้ในการแทนแนวคิดของภาษา ดังเช่น 'น้อย', 'ปานกลาง' เป็นต้น เราเรียกคำเหล่านี้ว่าค่าของตัวแปรภาษา (Linguistic value) หรือพจน์ภาษา (Linguistic term) ซึ่งเป็นค่าของตัวแปรภาษา (Linguistic variable) แต่ละตัวแปรภาษาถูกนิยามเป็น ห้าทูเพิล (quintuple) คือ (x, T(x), U, g, m) โดยที่ x เป็นชื่อของตัวแปรภาษา T(x) เป็นพจน์ภาษา (Linguistic term) ของตัวแปรภาษา x ในเซตสากล สำหรับ U, g เป็นกฎวากยสัมพันธ์ (Syntactic rule) ที่ใช้ในการสร้างพจน์ภาษา และ m เป็นฟัซซีเซตที่อยู่บนเซตสากล U นั่นคือ m: T$\rightarrow$P(U) ตัวอย่างของตัวแปรภาษา ซึ่งมีตัวแปรภาษา x คือ 'ประเภทลูกหนี้' และT(x) หรือพจน์ภาษา คือ 'ดีมาก', 'ดี', 'ปานกลาง', 'พอใช้' และ 'แย่' (ดังภาพที่ 2.3) นอกจากนี้ ฟังก์ชันความเป็นสมาชิก (Membership function) ซึ่งเป็นฟังก์ชันที่มีการกำหนดระดับความเป็นสมาชิกของตัวแปร เช่น ฟังก์ชันสี่เหลี่ยมคางหมู โดยมีรูปแบบของชนิดของฟังก์ชันความเป็นสมาชิก ดังต่อไปนี้ คือ

1). ฟังก์ชันสี่เหลี่ยมคางหมู (Trapezoidal membership function) เป็นฟังก์ชันสี่เหลี่ยมคางหมูที่มี 4 พารามิเตอร์คือ $\{a, b, c, d\}$

$$Trapezoidal(x; a, b, c, d) = \begin{cases} 0 & \text{เงื่อนไขกรณี } x < a \\ (x-a)/(b-a) & \text{เงื่อนไขกรณี } a \le x < b \\ 1 & \text{เงื่อนไขกรณี } b \le x < c \\ (d-x)/(d-c) & \text{เงื่อนไขกรณี } c \le x < d \\ 0 & \text{เงื่อนไขกรณี } x \ge d \end{cases}$$



ภาพที่ 2.3 แสดงตัวอย่างของตัวแปรภาษา (Linguistic variable)

2.3 สมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) [4][5]

การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นที่ทำหน้าที่พยากรณ์ตั้งแต่ 2 ตัวขึ้นไปกับตัวแปรตาม 1 ตัว เช่น

ต้องการพยากรณ์ผลการเรียนของนักศึกษาแผนกคอมพิวเตอร์(Y) โดยใช้ตัวพยากรณ์ 3 ตัว ประกอบด้วย ความสนใจของผู้เรียน (X1) ความรู้พื้นฐานของผู้เรียน (X2) และคุณภาพการสอนของผู้สอน (X3) เป็นต้น ในการวิเคราะห์การถดถอยพหุคูณนั้นจะต้องหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (Multiple Correlation Coefficient) เพื่อให้ทราบถึงความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นจำนวนทั้ง 3 ตัวกับตัวแปรตามว่ามีความสัมพันธ์กันเช่นใด สำหรับการวิเคราะห์การถดถอยพหุคูณ จะต้องหาสมการถดถอยเพื่อใช้ในการพยากรณ์ของตัวแปรตาม (Y) และหาค่าความคลาดเคลื่อนมาตรฐาน รวมทั้งหาค่าสหสัมพันธ์พหุคูณ (Multiple Correlation) เพื่อหาความสัมพันธ์เชิงเส้นตรงที่เป็นไปได้สูงสุดระหว่างตัวแปรอิสระหรือตัวแปรต้นกับตัวแปรตาม สำหรับสมการการถดถอยเชิงเส้นพหุคูณเพื่อใช้ในการพยากรณ์อยู่ในรูปดังนี้

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k \qquad (1)$$

เมื่อ $\hat{Y}$ แทนค่าของตัวแปรตาม Y ที่ได้จากการทำนายของตัวแปรอิสระ X_i โดยที่ i =1,...,k

b_0 แทนค่าคงตัว (Constant)

b_i แทนสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระ X_i โดยที่ i=1,...,k

สิ่งสำคัญที่ต้องการหาในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

- สัมประสิทธิ์สหสัมพันธ์พหุคูณ
- สมการพยากรณ์ในรูปคะแนนดิบ หรือในรูปคะแนนมาตรฐาน หรือทั้งคู่
- และความคลาดเคลื่อนมาตรฐานในการพยากรณ์

การคัดเลือกตัวแปรเพื่อการพยากรณ์ วิธีคัดเลือกตัวแปรพยากรณ์เข้าสู่สมการถดถอยมีอยู่หลายวิธี เช่น

1). การวิเคราะห์ถดถอยพหุคูณแบบปกติ (Enter Regression)

2). การวิเคราะห์ถดถอยพหุคูณแบบคัดเลือกออก (Remove Regression)

3). การวิเคราะห์ถดถอยพหุคูณแบบเดินหน้า (Forward Regression)

4). การวิเคราะห์ถดถอยพหุคูณแบบถอยหลัง (Backward Regression)

5). การวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได (Stepwise Regression)

ขั้นตอนการวิเคราะห์

1). ตรวจสอบข้อตกลงเบื้องต้น

2). คำนวณค่า r_{xy} ของตัวแปรพยากรณ์กับตัวแปรเกณฑ์

3). คัดเลือกตัวแปรพยากรณ์ที่มีความสัมพันธ์สูงสุดกับตัวแปรเกณฑ์เข้าสมการ และคำนวณค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R)

4). ทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R) ทดสอบว่าตัวแปรพยากรณ์ที่เข้าในสมการยังคงอยู่ในสมการต่อไปได้หรือไม่ด้วยสถิติ F

5). หาค่าน้ำหนักความสำคัญของตัวแปรพยากรณ์ (b) หรือ เบต้า หรือทั้งสองอย่าง เพื่อนำมาใช้ในการเขียนสมการพยากรณ์ และเปรียบเทียบว่าตัวแปรพยากรณ์ตัวใดพยากรณ์ตัวแปรเกณฑ์ได้ดีกว่า

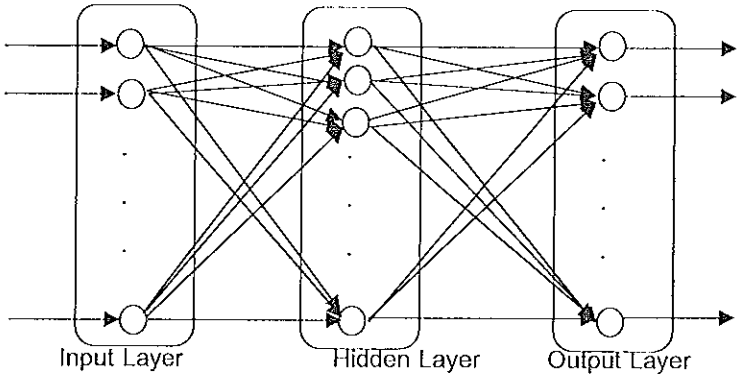
6). ทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอย เพื่อตรวจสอบว่าตัวแปรพยากรณ์สามารถพยากรณ์ตัวแปรเกณฑ์ได้หรือไม่ ด้วยสถิติ t

7). คำนวณค่าความคลาดเคลื่อนมาตรฐานของตัวแปรพยากรณ์ที่เข้าสมการ และคำนวณค่าความคลาดเคลื่อนมาตรฐานของการพยากรณ์

8). คัดเลือกตัวแปรพยากรณ์ที่มีความสัมพันธ์สูงกับตัวแปรเกณฑ์รองลงมาเข้าสมการและทำการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่เปลี่ยนแปลง (Adjusted R Square) ด้วยสถิติ F ถ้า Adjusted R Square ไม่มีนัยสำคัญก็แสดงว่าตัวแปร พยากรณ์ไม่สามารถอยู่ในสมการพยากรณ์ได้ แต่ถ้ามีนัยสำคัญก็ดำเนินการตามข้อ 4, 5, 6 และ 7 และ ดำเนินการต่อไปจนกว่าจะไม่มีตัวแปรพยากรณ์ใดเข้าในสมการ (การดำเนินการตามข้อ 8 เป็นวิธีการวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได)

2.4 โครงข่ายประสาทเทียม (Neural Network) [1][3][13]

แบบจำลองระบบโครงข่ายประสาทเทียม เป็นแบบจำลองทางคณิตศาสตร์ที่ประกอบไปด้วยโครงสร้างการทำงาน และการเชื่อมต่อเลียนแบบระบบการทำงานของเซลล์สมองของมนุษย์ งานวิจัยด้านโครงข่ายประสาทเทียมมีความแพร่หลายอย่างมากที่ใช้กระบวนการแพร่กลับ (Back Propagation Algorithm) โดยทั่วไปแล้วโครงข่ายของแบบจำลองประกอบด้วยชั้นของ node 3 ชั้น ได้แก่ ชั้นข้อมูลด้านเข้า (Input layer) ชั้นแฝง (Hidden layer) และชั้นแสดงผล (Output layer) ข้อมูลจากชั้นข้อมูลด้านเข้าจะถูกคำนวณและส่งผ่านฟังก์ชั่นแปลงรูปจากชั้นแฝงไปยังชั้นแสดงผล โดยหลักการของกระบวนการเรียนรู้ของแบบจำลองคือ การเปลี่ยนแปลงค่าถ่วงน้ำหนักของแต่ละการเชื่อมต่อ เพื่อปรับผลลัพธ์ของแบบจำลองให้ใกล้เคียงกับค่าความจริงมากที่สุดโดยอาศัยการแพร่กลับ (ดังภาพที่ 2.4)



ภาพที่ 2.4 แสดงโครงข่ายประสาทเทียมแบบหลายชั้น

(ที่มา: Holland. 1975. Adaptation in Natural and Artificial System)

– **ชั้นอินพุต (Input Layer)**

- เป็นชั้นที่รอรับข้อมูลเข้าสู่โครงข่ายประสาทเทียม
- ข้อมูลที่นำเข้าสู่โครงข่าย ก็คือข้อมูลที่ผ่านการเตรียมข้อมูลดังที่ได้แสดงไว้ในขั้นตอนการเตรียมข้อมูลของโครงข่ายประสาทเทียม

– **ชั้นซ่อน (Hidden Layer)**

- เป็นชั้นที่เพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูล เพื่อที่จะส่งต่อไปยังชั้นเอาต์พุต

– **ชั้นเอาต์พุต (Output Layer)**

- เป็นชั้นที่คำนวณหาผลลัพธ์ของโครงข่ายประสาทเทียม
- จำนวนโหนดในชั้นเอาต์พุตจะขึ้นอยู่กับจำนวนกลุ่มข้อมูลที่ต้องการจัด โดยมีสมการในการคำนวณดังสมการ

$$Z_k = f(\sum_{j=1}^{m} y_j w_{jk} + \theta_k)$$

โดยที่ Z_k คือ ผลลัพธ์ในชั้นเอาต์พุตโหนดที่ k

y_j คือ ผลลัพธ์ในชั้นซ่อน หรือข้อมูลส่งออกในชั้นซ่อนโหนดที่ j

w_{jk} คือ น้ำหนักบนเส้นเชื่อมระหว่างโหนดที่ j ในชั้นซ่อนและโหนดที่ k ในชั้นเอาท์พุต

θ_k คือ ค่าไบแอสของโหนดที่ k ในชั้นเอาท์พุต

m คือ จำนวนโหนดทั้งหมดของชั้นซ่อน

$f(u)$ คือ Sigmoid Function

ซึ่ง Sigmoid Function หาได้จากสมการที่นำไปคิดดังนี้

$$f(u) = \frac{1}{1+e^{-u}}$$

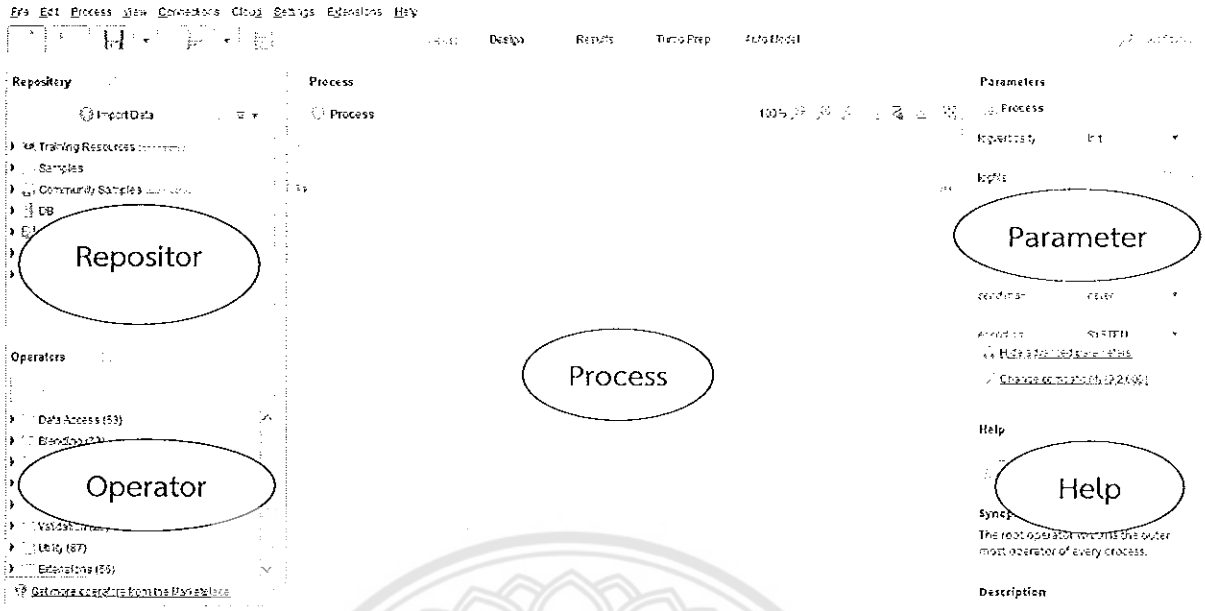
โดยที่ u คือ ผลลัพธ์ที่ได้จาก $\sum_{i=1}^{n} x_i w_{ij} + \theta_j$

e มีค่าประมาณ 2.71828

2.5 โปรแกรม RapidMiner Studio [6]

การพัฒนาโปรแกรมนี้เริ่มขึ้นจากบริษัทที่ชื่อว่า Rapid-I ในประเทศเยอรมนีและเมื่อช่วงปลายปี 2013 โดยที่บริษัทแห่งนี้ได้รับทุนสนับสนุนจำนวนมากจากนักลงทุนในประเทศสหรัฐอเมริกาจึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทน และได้ทำการย้ายที่ตั้งของสำนักงานใหญ่ไปอยู่ที่ประเทศสหรัฐอเมริกา อย่างไรก็ตาม เราสามารถดาวน์โหลดซอฟต์แวร์ RapidMiner Studio ได้จากเว็บไซต์ http://rapidminer.com

สำหับองค์ประกอบต่าง ๆ ที่เกี่ยวข้องกับโปรแกรม RapidMiner Studio จะมีรายละเอียดที่จำนำเสนอคร่าวๆ ดังต่อไปนี้



ภาพที่ 2.5 แสดงองค์ประกอบหน้าต่างหลักของการทำงาน

2.6 การทบทวนวรรณกรรมที่เกี่ยวข้อง

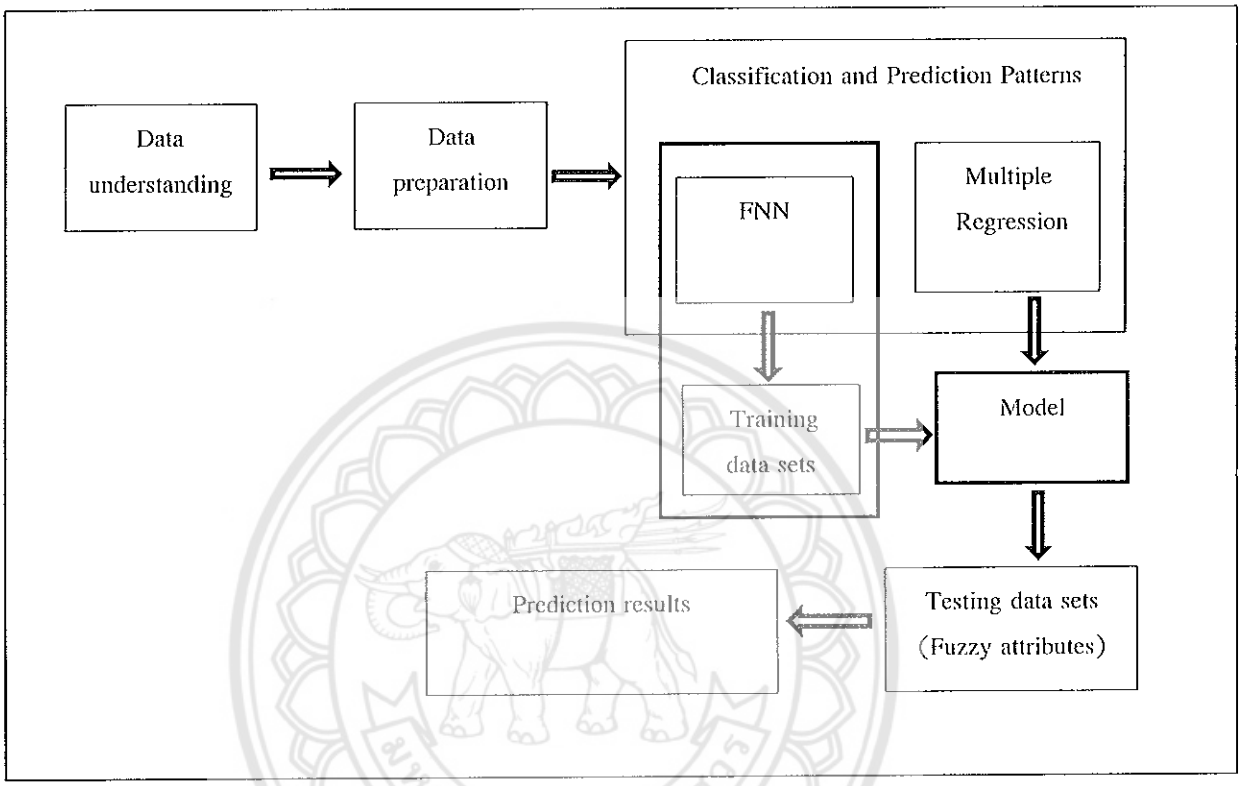
จากมติคณะรัฐมนตรีในคราวประชุมเมื่อวันที่ 16 มกราคม 2539 ได้อนุมัติโครงการจัดตั้งกองทุนเงินให้กู้ยืมเพื่อการศึกษา (กยศ.) ในลักษณะเงินหมุนเวียน โดยได้รับการจัดสรรงบประมาณจากกระทรวงการคลังในปีแรกจำนวน 3,000 ล้านบาทและจัดสรรงบประมาณสมทบให้เป็นรายปีตามความจำเป็น และได้ออกระเบียบว่าด้วยการชำระเงินกู้ยืมคืนกองทุน พ.ศ. 2542 และออกข้อบังคับกองทุนว่าด้วยการดำเนินงาน วิธีการกู้ยืมเงินกองทุนเริ่มบังคับใช้ตั้งแต่ปี พ.ศ2546 ซึ่งจุดประสงค์หลักของการจัดตั้งกองทุนโดยมีวัตถุประสงค์เพื่อให้กู้ยืมแก่นักเรียน หรือนักศึกษาที่ขาดแคลนทุนทรัพย์ ตั้งแต่ระดับมัธยมศึกษาตอนปลายทั้งสายสามัญและสายอาชีพจนถึงระดับปริญญาตรี เพื่อเป็นค่าเล่าเรียน ค่าใช้จ่ายที่เกี่ยวข้องกับการศึกษาและค่าใช้จ่ายที่จำเป็นในการครองชีพระหว่างศึกษา [11] จะพบว่ามีนักเรียน นิสิต นักศึกษาทั่วประเทศที่เข้าร่วมโครงการดังกล่าว โดยแต่ละสถาบันการศึกษาได้ดำเนินการคัดเลือกบุคคลตามเงื่อนไขที่เหมาะสมที่จะกู้ยืมโดยแต่ละสถาบันการศึกษาได้รับจำนวนโค้วต้ายอดจำนวนเงินที่จะจัดสรรในปริมาณที่แตกต่างกัน ทั้งขึ้นอยู่กับเงื่อนไขที่หน่วยงานของกองทุนกำกับไว้ ถึงแม้กองทุนที่ได้จัดตั้งขึ้นนี้มีวัตถุประสงค์หลักเพื่อต้องการช่วยเหลือนิสิต นักศึกษากู้ยืมเงินเพื่อการศึกษาและชำระเงินคืนเมื่อจบการศึกษาและผ่อนชำระตามกฎเกณฑ์ที่ตั้งไว้แล้วนั้น จะพบว่า เมื่อถึงเวลาการชำระเงินกู้จะพบว่ามีบัณฑิตจำนวนมากที่ขาดการชำระหนี้เงินกู้ดังกล่าว ซึ่งจากรายงานในปี พ.ศ. 2549 ลูกหนี้เงินให้กู้ยืมตามบัญชีของกองทุนมียอดคงเหลือจำนวน 217,791.05 ล้านบาท จากจำนวนลูกหนี้ 2,668,563 ราย และหนี้สงสัยจะสูญประมาณ 4.83 ล้านบาท ค่าใช้จ่ายติดตามหนี้ 63,351,039.17 บาท สำหรับรายงานในปี พ.ศ. 2550 ลูกหนี้เงินให้กู้ยืมตามบัญชีของกองทุนมียอดคงเหลือ

จำนวน 230,502.38 ล้านบาท จากจำนวนลูกหนี้ 2,682,300 ราย และหนี้สงสัยจะสูญประมาณ 5.68 ล้านบาท ค่าใช้จ่ายติดตามหนี้ 71,488,928.20 บาท นอกจากนี้ยังมีค่าบริหารจัดการในการดำเนินคดีผู้กู้ที่ถูกฟ้องดำเนินคดีปี 2548 2549 2550 และ 2551 ที่ได้ดำเนินการในปี พ.ศ. 2554 มีค่าเท่ากับ 156,300.00 2,243,007.50 6,876,645.00 และ 29,216,203.50 ตามลำดับ และที่ได้ดำเนินการในปี พ.ศ. 2555 มีค่าเท่ากับ 435,400.00 1,447,800.00 1,884,870.00 และ 10,675,300.00 [11] ตามลำดับ ซึ่งจะพบว่ามีค่าการดำเนินการคดีกับผู้กู้เป็นจำนวนมากยิ่งขึ้น อย่างไรก็ตาม เราจะพบว่าจำนวนเงินของหนี้สงสัยจะสูญยิ่งมีจำนวนมากขึ้นรวมทั้งทั้งค่าใช้จ่ายในการติดตามหนี้ก็สูงเพิ่มขึ้นมากเช่นกัน โดยเฉพาะอย่างยิ่งจากข้อมูลในปี พ.ศ. 2556 จะพบว่า มียอดค้างชำระเงินกู้กว่า 1.485 ล้านราย เป็นวงเงินกู้ 136,237 ล้านบาท จากจำนวนผู้กู้ทั้งหมด 2.15 ล้านราย วงเงินกู้ 194,711 ล้านบาท ถือเป็นยอดเบี้ยวหนี้เกือบ 80 เปอร์เซ็นต์ [7] จากเหตุการณ์ดังกล่าวข้างต้นของการชำระหนี้เงินกู้ที่มีเปอร์เซ็นต์ที่ลดลงนั้น อาจทำให้กองทุนกู้ยืมเพื่อการศึกษาที่ได้จัดตั้งขึ้นนี้ต้องประสบกับปัญหาสภาพคล่องในอนาคต เพราะบัณฑิตที่กู้ยืมเงิน ไม่ผ่อนชำระคืน ในขณะที่กยศ. ต้องปล่อยกู้ให้แก่นักเรียนและนักศึกษารายเก่าจนกว่าสำเร็จการศึกษา ส่วนนักศึกษาใหม่ที่จะได้ก็จะมีจำนวนรายลดน้อยไปด้วยเพราะ กยศ.ไม่มีสภาพคล่องเพียงพอที่จะปล่อยกู้ให้แก่ผู้กู้ได้ครบทุกราย อย่างไรก็ตาม ทางกรมบังคับคดี กระทรวงยุติธรรม น.ส.รื่นวดี สุวรรณมงคล อธิบดีกรมบังคับคดี และ ดร.ทิตติมา วิชัยรัตน์ ผู้จัดการกองทุนเงินให้กู้ยืมเพื่อการศึกษา (กยศ.) ร่วมลงนามบันทึก ข้อตกลงว่าด้วยการจัดโครงการไกล่เกลี่ยลูกหนี้ ตามคำพิพากษาในชั้นบังคับคดี เพื่อเพิ่มช่องทางในการชำระหนี้ให้แก่ลูกหนี้ตามคำพิพากษา โดย ดร.ทิตติมากล่าวว่า รัฐบาลมีนโยบายให้ปล่อยกู้ กยศ.ให้ยาก จ่ายคืนให้ง่าย เพื่อให้เด็กเห็นคุณค่าเงิน และเน้นคัดกรองเด็กที่มีคุณภาพเพื่อไม่ให้เกิดปัญหาในอนาคตนั้น กยศ.ได้จัดโครงการต่างๆ เพื่อรณรงค์ และช่วยผู้ค้างชำระหนี้ไม่ต้องถูกดำเนินคดีตามกฎหมาย โดยโครงการไกล่เกลี่ยลูกหนี้ตามคำพิพากษาในชั้นบังคับคดีครั้งนี้ ก็ถือเป็นการช่วยให้ผู้ค้างชำระ ตั้งแต่ปี 2555 จำนวน 4,000 กว่าราย เพื่อช่วยผู้กู้รู้ตัวและมาชำระหนี้จะได้ไม่ต้องถูกฟ้องร้อง แต่หากเพิกเฉยไม่เข้าสู่ กระบวนการไกล่เกลี่ยหรือติดต่อชำระหนี้ กยศ.ต้องฟ้องร้อง ซึ่งเมื่อคดีถึงที่สุดก็ต้องยึดทรัพย์ตามประมวลกฎหมายพิจารณาความแพ่ง มาตรา 27 กำหนดให้การบังคับคดีภายใน 10 ปี นับแต่วันที่มีคำพิพากษา โดยที่คำพิพากษาหรือคำสั่งนั้นถึงที่สุดแล้ว ซึ่งขณะนี้มีผู้กู้ กยศ.ทั้งหมด 4 ล้านกว่าราย เป็นผู้กู้ครบชำระหนี้ 2.9 ล้านคน ในจำนวนนี้เป็นผู้กู้ปกติไม่ได้ค้างชำระ 2 ล้านคน ส่วนที่เหลือเป็นผู้กู้ค้างชำระ [9] นอกจากนี้ทางกองทุนเงินให้กู้ยืมเพื่อการศึกษา (กยศ.) เปิดเผยว่า "โครงการรณรงค์ชำระเงินคืนกองทุน กยศ." ที่ได้เริ่มดำเนินการมาตั้งแต่เดือนพฤศจิกายน 2556 และสิ้นสุดโครงการในวันที่ 31 มี.ค. 57 ทั้งนี้ กองทุนฯ ได้เชิญชวนผู้ครบกำหนดชำระหนี้ และผู้ที่ค้างชำระหนี้ที่ยังไม่ได้เข้าร่วมโครงการให้เข้าร่วมโครงการเพื่อรับสิทธิพิเศษ ซึ่งแบ่งออกเป็น 1. ผู้ที่ไม่มียอดค้างชำระหากท่านทำการปิดบัญชีทั้งหมด จะได้รับการลดเงินต้น 3.5 เปอร์เซ็นต์ของยอดหนี้ 2. ผู้ที่มียอดหนี้ค้างชำระหากท่านทำการชำระหนี้ให้เป็นยอดหนี้ปกติ จะได้รับการลดเบี้ยปรับให้ทั้งหมด 100 เปอร์เซ็นต์ หรือหากท่านทำการปิดบัญชีทั้งหมด จะได้รับการลดเบี้ยปรับให้ทั้งหมด 100 เปอร์เซ็นต์เต็มแล้ว ยังจะได้รับการลดดอกเบี้ยให้อีก 50 เปอร์เซ็นต์"หลังจากสิ้นสุดระยะเวลาของโครงการ กองทุนฯ จะสรุปข้อมูลของผู้ที่ยังค้างชำระหนี้ โดยเฉพาะอย่างยิ่งในกลุ่มผู้ค้างชำระหนี้ 5 งวด หรือ 4 ปีขึ้นไป กองทุนฯ จะต้องดำเนินการฟ้องร้องดำเนินคดี

และในปีล่าสุดนายรังสรรค์ ศรีวรศาสตร์ ปลัดกระทรวงการคลัง ในฐานะประธานกรรมการกองทุนเงินให้กู้ยืมเพื่อการศึกษา เปิดเผยว่า "คณะกรรมการกองทุนฯ ได้มีมติเห็นชอบให้กองทุนฯ จัดโครงการรณรงค์ชำระหนี้ ระหว่างวันที่ 13 ธันวาคม 2557-30 เมษายน 2558 เพื่อเปิดโอกาสให้ผู้กู้ยืมที่ค้างชำระหนี้ กยศ.ได้แก้ไขปัญหาหนี้ค้างชำระของตนเองให้เป็นปกติ โดยกองทุนได้ออกมาตรการลดเบี้ยปรับ 100% สำหรับผู้ที่ค้างชำระหนี้ ที่มาชำระหนี้ปิดบัญชี หรือลดเบี้ยปรับ 50% สำหรับผู้กู้ที่มาชำระหนี้ค้างให้เป็นปกติ นอกจากนั้นยังมีเงื่อนไขพิเศษสำหรับกลุ่มผู้กู้ที่ชำระหนี้เป็นปกติ หากมีความประสงค์จะปิดบัญชีจะได้ลดหย่อนหนี้ถึง 3% ของเงินต้นที่มาปิดบัญชีเพื่อเป็นรางวัลให้แก่ผู้กู้ที่มีประวัติชำระหนี้ดีผู้กู้ยืมที่มีสิทธิเข้าร่วมโครงการ ได้แก่ ผู้กู้ยืมที่ครบกำหนดชำระหนี้แล้วและเป็นผู้กู้ยืมที่ไม่เคย ทำสัญญาประนีประนอมยอมความในโครงการไกล่เกลี่ยหรือถูกดำเนินคดี ซึ่งมีประมาณ 2.15 ล้านราย [8] นอกจากนี้ทางกองทุนฯ ได้ดำเนินสร้างจิตสำนึกความรับผิดชอบต่อตนเองและสังคมมากขึ้นด้วยการกำหนดเงื่อนไขเพิ่มเติมสำหรับผู้กู้รายเก่าดังเช่น การกำหนดการเข้าร่วมทำประโยชน์ต่อสังคมและสาธารณะ ไม่น้อยกว่า 36 ชั่วโมงต่อปีการศึกษา รวมทั้งการกำหนดให้สถานศึกษาต้องผ่านการรับรองคุณภาพจากสำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา (สมศ.) หลักสูตรที่เปิดสอนนระดับปริญาตรีต้องได้รับการรับทราบหลักสูตรจากสำนักงานคณะกรรมการการอุดมศึกษา (สกอ.) [10] อย่างไรก็ตาม การประยุกต์เทคโนโลยีเพื่อนำมาช่วยการคัดครองการให้เงินกู้ยืมแก่นิสิต นักศึกษานั้น สามารถเลือกใช้อัลกอริธึมทางเหมืองข้อมูลเพื่อสร้างแบบจำลองการพยากรณ์ความเสี่ยงที่จะสูญเสียการชำระเงินกู้ได้ ทั้งนี้จำเป็นที่จะต้องมีการเปรียบเทียบกับสมการทางคณิตศาสตร์ด้านอื่นๆ เพื่อกำหนดแบบจำลองให้เหมาะสมมากที่สุดของงานวิจัย ดังเช่นตัวอย่างของงานวิจัยที่มีการเทคนิคที่กล่าวข้างต้น [2] ใช้สำหรับการพยากรณ์ด้านการครอบครองยานพาหนะของครัวเรือนในประเทศไทยและการพยากรณ์คุณภาพผลไม้ [12] เป็นต้น

บทที่ 3
วิธีดำเนินการวิจัย

3.1 สถาปัตยกรรมของระบบการทำงานในงานวิจัย



ภาพที่ 3.1 แสดงสถาปัตยกรรมระบบ

เราได้กำหนดแนวคิดของงานวิจัยนี้ (ภาพที่ 3.1) เพื่อเปรียบเทียบระหว่างการใช้เทคนิคฟัซซีนิวรอลเน็ตเวิร์ค (Fuzzy Neural Network) กับเทคนิค Multiple regression สำหรับพยากรณ์ผลของความเสี่ยงจากการกู้ยืมเงิน กยศ ของนักศึกษามหาวิทยาลัยนเรศวร เราสามารถแบ่งการทำงานหลักๆ ออกเป็น 3 ส่วน คือ

ส่วนที่ 1 เป็นส่วนการรวบรวมข้อมูลและทำการคัดเลือกตัวแปรด้วยวิธีจัดเก็บข้อมูลแบบปฐมภูมิและทุติยภูมิ แล้วทำการคัดเลือกตัวแปรอิสระนำเข้าที่เหมาะสมด้วยเทคนิคการหาความสัมพันธ์ระหว่างตัวแปรด้วยรูปแบบสัมประสิทธิ์สหสัมพันธ์และสมการถดถอยเชิงเส้นพหุคูณ

ส่วนที่ 2 เป็นการเปรียบเทียบเทคนิคระหว่างฟัซซีนิวรอลเน็ตเวิร์คกับวิธีสมการถดถอยเชิงเส้นเพื่อสร้างตัวแบบจำลองที่เหมาะสมสำหรับในไปใช้สร้างเป็นกฎแล้วฝังรหัสเป็นโค้ดในแอพพลิเคชัน

ส่วนที่ 3 การทำงานของระบบพยากรณ์ความเสี่ยงการสูญเสียเงินกู้ กยศ. ผ่านแบบจำลองพยากรณ์

3.2 การจัดตรียมข้อมูลการกู้ยืมทุน กยศ

งานวิจัยได้ดำเนินการจัดเก็บข้อมูลเบื้องต้นของนิสิตที่ทำการกู้ยืม กยศ. ที่สำเร็จการศึกษาแล้วและมีรายงานการชำระเงิน ในรูปแบบต่างๆ ของแต่ละบุคคลตามแต่ละสาขา ทั้งนี้ ผู้วิจัยเน้นนิสิตที่สำเร็จการศึกษาทางสาขาด้านวิทยาศาสตร์และเทคโนโลยีเป็นหลักของการดำเนินการวิจัยเพื่อใช้ตัวแทนที่แสดงถึงภาพโดยรวมของการนำข้อมูลไปใช้สำหรับสร้างตัวแบบจำลอง ซึ่งขั้นตอนของการจัดเก็บข้อมูลเบื้องต้นอยู่ภายใต้รูปแบบ Microsoft excel(ดังภาพที่ 3.2 รายละเอียดคณะเกษตรศาสตร์ ภาพที่ 3.3 รายละเอียดคณะวิศวกรรมศาสตร์ และ ภาพ 3.4 รายละเอียดคณะวิทยาศาสตร์)

รหัสนิสิต	คณะ	สาขาวิชา	GPAX	การศึกษา	อาชีพบิดา	รวมได้ร่วม	จำนวนที่กู้	ค่าลงทะเบียน	ค่าหนังสือ	ค่าครองชีพ	รวม	ข้อมูลสถานะการกู้ยืมของทุนเงินให้กู้ยืมเพื่อการศึกษา
52162041	เกษตรศาสตร์ทรัพยากร	การประมง	2.41	ไม่ระบุ	ค้าขาย	15000	3	12,000.00	0	12,000.00	24,000.00	ไม่เคยกู้ยืม
52162072	เกษตรศาสตร์ทรัพยากร	การประมง	2.30	รับจ้าง	แม่บ้าน	8000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52162355	เกษตรศาสตร์ทรัพยากร	การประมง	2.13	เกษตรกร	เกษตรกร	7500	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52162034	เกษตรศาสตร์ทรัพยากร	การประมง	2.23	ธุรกิจส่วนตัว	เกษตรกร	5000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52162294	เกษตรศาสตร์ทรัพยากร	การประมง	2.97	เกษตรกร	เกษตรกร	10000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52162027	เกษตรศาสตร์ทรัพยากร	การประมง	2.10	เกษตรกร	เกษตรกร	3000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164182	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - ทรัพยากรดิน	2.39	เกษตรกร	แม่บ้าน	3500	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164359	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - ทรัพยากรดิน	2.95	รับจ้าง	แม่บ้าน	8000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164717	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - ทรัพยากรดิน	3.16	รับจ้าง	รับจ้าง	6000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163566	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.98	เกษตรกร	เกษตรกร	20000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164323	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.23	ค้าขาย	ค้าขาย	5000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163662	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.84	เกษตรกร	เกษตรกร	9000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164518	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.48	เกษตรกร	แม่บ้าน	9000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164007	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.18	เกษตรกร	เกษตรกร	5000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52165645	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.68	เกษตรกร	เกษตรกร	45000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163963	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.68	เกษตรกร	แม่บ้าน	11400	2	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163559	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.66	เกษตรกร	เกษตรกร	10000	2	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163031	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.41	เกษตรกร	เกษตรกร	9000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52163550	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.83	พนักงานบริษัท	เกษตรกร	10000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่เคยกู้ยืม
52164427	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	3.12	เกษตรกร	เกษตรกร	15000	3	12,000.00	0.00	0.00	12,000.00	ไม่เคยกู้ยืม
52164709	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.80	รับจ้าง	รับจ้าง	10000	1	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสา
52163655	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.92	เกษตรกร	เกษตรกร	8000	0	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสา
52164441	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.50	เกษตรกร	เกษตรกร	10000	0	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสา
52164756	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - พืชศาสตร์	2.53	เกษตรกร	เกษตรกร	5000	2	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสา
52165533	เกษตรศาสตร์ทรัพยากร	เกษตรศาสตร์ - สัตวศาสตร์	2.62	เกษตรกร	เกษตรกร	25000	0	12,000.00	0	12,000.00	24,000.00	ไม่เคยกู้ยืม

ภาพที่ 3.2 รายละเอียดคณะเกษตรศาสตร์

ลำดับที่	เลขที่ประจำตัวประชาช	คณะ	สาขาวิชา	GPAX	การศึกษา	อาชีพบิดา	รวมได้ร่วม	จำนวนที่กู้	ค่าลงทะเบียน	ค่าหนังสือ	ค่าครองชีพ	รวม	ข้อมูลสถานะการกู้ยืมของทุนเงินให้กู้ยืมเพื่อการศึกษา
1	1-1003-01532-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.05	ค้าขาย	ค้าขาย	12000	1	15,000.00	0.00	12,000.00	27,000.00	ไม่เคยกู้ยืม
2	1-6003-00149-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.39	รับจ้าง	เกษตรกร	10000	0	15,000.00	0.00	12,000.00	27,000.00	ไม่เคยกู้ยืม
3	1-3004-00149-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.01	ไม่ระบุ	รับจ้าง	35000	0	15,000.00	0.00	12,000.00	27,000.00	ไม่เคยกู้ยืม
4	1-6799-00204-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.20	รับจ้าง	รับจ้าง	10000	1	15,000.00	0.00	12,000.00	27,000.00	ไม่เคยกู้ยืม
5	1-1014-01826-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.41					15,000.00	0.00	0.00	15,000.00	ไม่เคยกู้ยืม
6	1-5399-00321-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.27	รับจ้าง	รับจ้าง	12000	0	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
7	1-6011-00192-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.56	ไม่ระบุ	เกษตรกร	75000	0	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
8	1-6299-00159-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.66	เกษตรกร	เกษตรกร	6000	1	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
9	1-6199-00149-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.42	เกษตรกร	เกษตรกร	6000	1	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
10	1-6502-00093-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.31	เกษตรกร	เกษตรกร	2400	1	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
11	1-6502-00090-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.20	ไม่ระบุ	เกษตรกร	3000	1	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
12	1-6599-00427-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.65	รับจ้าง	รับจ้าง	15000	1	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
13	1-6999-00169-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.15	เกษตรกร	แม่บ้าน	6000	0	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ
14	2-6706-00021-xx-x	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้า	2.80	เกษตรกร	เกษตรกร	5000	0	15,000.00	0.00	12,000.00	27,000.00	เคยกู้ยืม ในระดับมัธยมศึกษาตอนปลาย สายสามัญ

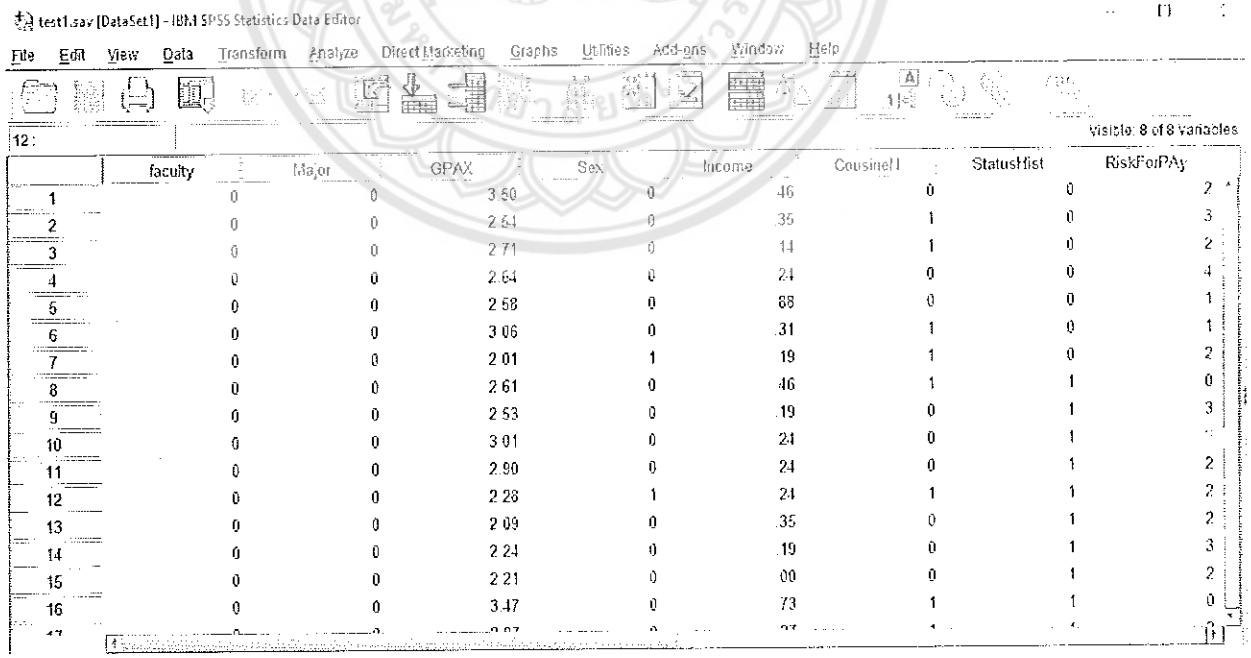
ภาพที่ 3.3 รายละเอียดคณะวิศวกรรมศาสตร์

ลำดับที่	เลขที่ประจำตัวประชาชน	คณะ	สาขาวิชา	GPAX	อาชีพบิดา	อาชีพมารดา	รายได้ต่อเดือน	จำนวนพี่น้อง	ค่าใช้จ่ายในการเรียน	ค่าที่อยู่อาศัยเมื่อมาเรียน	ค่าของใช้	รวม	ข้อมูลลักษณะการกู้ยืมของนักเรียนให้กู้ยืมหรือไม่
1	1-5799-00307-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	3.50	เกษตรกร	เกษตรกร	10000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
2	1-5603-00172-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.54	รับจ้าง	รับจ้าง	8000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
3	1-5359-00280-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.71	รับจ้าง	ไม่ประกอบ	8000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
4	1-6001-00392-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.64	ไม่ประกอบ	รับจ้าง	6000	0	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
5	1-6599-03402-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.58	ข้าราชการ	ทำงาน	176880	0	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
6	1-6702-00157-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	3.06	เกษตรกร	แม่บ้าน	7200	1	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
7	1-6699-00167-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.01	รับจ้าง	รับจ้าง	8000	1	12,000.00	0.00	12,000.00	24,000.00	ไม่กยศ.กู้ยืม
8	1-4017-03119-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.61	เกษตรกร	เกษตรกร	10000	1	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา
9	1-5399-00277-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.53	ไม่ประกอบ	รับจ้าง	8000	0	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา
10	1-6004-00137-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	3.01	เกษตรกร	เกษตรกร	6000	0	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา
11	1-6399-00112-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.90	เกษตรกร	เกษตรกร	6000	0	0.00	0.00	12,000.00	12,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา
12	1-6497-00076-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.28	รับจ้าง	รับจ้าง	6000	1	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา
13	1-6495-00100-xx-x	คณะวิทยาศาสตร์	คณิตศาสตร์	2.09	รับจ้าง	เกษตรกร	8000	0	12,000.00	0.00	12,000.00	24,000.00	เคยกู้ยืม ในระดับ มัธยมศึกษาตอนปลาย สายสา

ภาพที่ 3.4 รายละเอียดคณะวิทยาศาสตร์

การดำเนินการจัดเก็บข้อมูลได้รวบรวมไว้จำนวน 296 ตัวอย่าง จากนิสิต 4 คณะ ในสาขาวิทยาศาสตร์และเทคโนโลยี ซึ่งการดำเนินการได้ทำในลักษณะทุติยภูมิ การจัดเก็บข้อมูลเบื้องต้นของนิสิตที่ทำการกู้ยืม กยศ. ที่สำเร็จการศึกษาแล้วและมีรายงานการชำระเงิน ในรูปแบบต่างๆ ของแต่ละบุคคลตามแต่ละสาขา ทั้งนี้ ผู้วิจัยจำเป็นจะต้องจัดและปรับช่วงค่าของข้อมูลให้มีรูปแบบและค่าใกล้เคียงกันโดยให้มีค่าอยู่ระหว่าง 0-1 ของแอททริบิวต์ Income เพื่อให้อยู่มาตรฐานเดียวกัน ดังสมการ (2)

$$\text{Output} = \frac{\text{input-min}}{\text{Max-min}} \qquad (2)$$



ภาพที่ 3.5 แสดงรายละเอียดการปรับเปลี่ยนระดับข้อมูล

3.3 การคัดเลือกปัจจัยนำเข้า

งานวิจัยนี้ใช้โปรแกรม IBM SPSS Statistics 22 (ดังภาพที่ 3.6) เพื่อคัดเลือกแอททริบิวต์ด้วยการหาความสัมพันธ์ระหว่างตัวแปร ด้วยเทคนิค Pearson's correlation coefficient (ดังสมการ 3) จะเป็นการใช้สัญลักษณ์ r หรือ r_{XY} เพื่อคำนวณค่า r ซึ่งค่า r หรือค่าสัมประสิทธิ์สัมพันธ์ (Correlation Coefficients) นั้น มีระดับความสัมพันธ์ (Strength of association) 0.01-0.0.9, 0.10-0.29, 0.30-0.49, 0.50-0.69, 0.70-0.89 และ 0.90-0.99 มีระดับความสัมพันธ์ เล็กน้อย ต่ำถึงกลาง กลางถึงสูง สูงถึงสูงมาก สูงมาก และ เกือบสมบูรณ์ ตามลำดับ นอกจากนี้การนำตัวแปรอิสระไปใช้งานนั้น ต้องทดสอบว่าความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรอิสระมีค่า r<0.75 หรือไม่ ถ้าน้อยกว่าก็สามารถนำไปใช้ได้ วิธีการนี้เป็นเพียงการทดสอบหาความสัมพันธ์ในรูปแบบ association ไม่ใช่การหาความสัมพันธ์ในลักษณะ causal relationship ซึ่งในงานวิจัยนี้ เราพบว่าความสัมพันธ์ระหว่างตัวแปร GPAX กับตัวแปร Income มีความสัมพันธ์กันน้อยสุดคือ 0.002 ในทิศทางแปรผกผันเล็กน้อย (Trivial) และมีตัวแปร Faculty กับ Major มีความสัมพันธ์กันมากที่สุดคือ 0.956 ในทิศทางแปรผันตามเกือบสมบูรณ์ (Near perfect) และพบว่าความสัมพันธ์ระหว่างตัวแปรอิสระ Faculty และ Major กับ ตัวแปรตาม RiskForPay มีค่าเท่ากับ 0.051 และ -0.003 ตามลำดับ รายละเอียดสามารถแสดงได้ดังตารางที่ 1

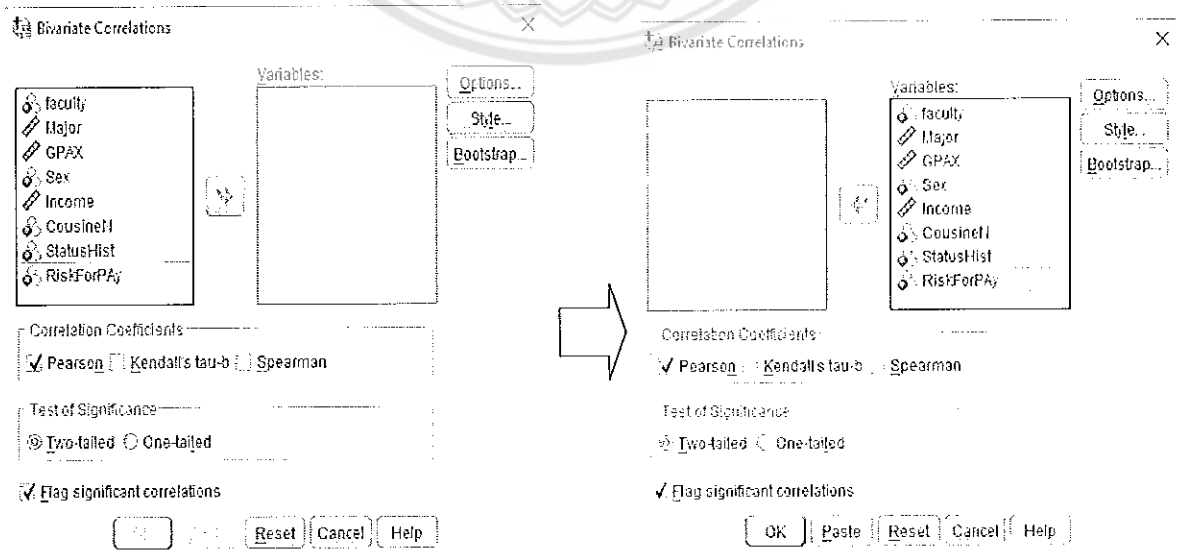
$$r = \frac{(1/N-1)(\sum XY-((\sum X)(\sum Y)/N))}{S_X S_Y} \quad (3)$$

r = ค่าสัมประสิทธิ์สหสัมพันธ์

N = จำนวนตัวอย่าง

S_x = ส่วนเบี่ยงเบนมาตรฐาน x

S_y = ส่วนเบี่ยงเบนมาตรฐาน y



ภาพที่ 3.6 แสดงการใช้งาน IBM SPSS Statistics 22 ตรวตสอบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรอิสระ

จ QA
9.64
ท6515
2563

05 มี.ค. 256
1034770

การใช้งาน Correlation Coefficients ในงานวิจัยนี้ได้ทำการโดยการใช้ Pearson เนื่องมีตัวแปรเชิงคุณภาพน้อยกว่าจำนวนตัวแปรเชิงปริมาณ ซึ่งเมื่อดำเนินการในขั้นต่อไป จะพบว่า เมื่อคลิกปุ่ม OK ผลลัพธ์ของการคำนวณ มีรายละเอียดดังภาพที่ 3.7

Correlations

		faculty	Major	GPAX	Sex	Income	CousineN	StatusHist	RiskForPAy
faculty	Pearson Correlation	1	.956**	.028	-.023	.003	-.023	.383**	.051
	Sig. (2-tailed)		.000	.626	.687	.954	.696	.000	.386
	N	296	296	296	296	296	296	296	296
Major	Pearson Correlation	.956**	1	.022	-.042	.027	-.011	-.363**	-.003
	Sig. (2-tailed)	.000		.711	.474	.649	.857	.000	.958
	N	296	296	296	296	296	296	296	296
GPAX	Pearson Correlation	.028	.022	1	-.109	.002	.045	.176**	-.114*
	Sig. (2-tailed)	.626	.711		.060	.972	.444	.002	.050
	N	296	296	296	296	296	296	296	296
Sex	Pearson Correlation	-.023	-.042	-.109	1	-.005	-.062	-.157**	-.008
	Sig. (2-tailed)	.687	.474	.060		.925	.285	.007	.885
	N	296	296	296	296	296	296	296	296
Income	Pearson Correlation	.003	.027	-.002	-.005	1	.096	-.114*	-.006
	Sig. (2-tailed)	.954	.649	.972	.925		.101	.049	.922
	N	296	296	296	296	296	296	296	296
CousineN	Pearson Correlation	-.023	-.011	.045	-.062	.096	1	.013	.015
	Sig. (2-tailed)	.696	.857	.444	.285	.101		.822	.796
	N	296	296	296	296	296	296	296	296
StatusHist	Pearson Correlation	-.383**	-.363**	.176**	-.157**	-.114*	.013	1	.014
	Sig. (2-tailed)	.000	.000	.002	.007	.049	.822		.817
	N	296	296	296	296	296	296	296	296
RiskForPAy	Pearson Correlation	.051	-.003	-.114*	-.008	-.006	-.015	.014	1
	Sig. (2-tailed)	.386	.958	.050	.885	.922	.796	.817	
	N	296	296	296	296	296	296	296	296

**. Correlation is significant at the 0.01 level (2-tailed).

ภาพที่ 3.7 แสดงรายละเอียดการคำนวณค่า Pearson's correlations

3.4 การสร้างตัวแปรดัมมีจากข้อมูลเชิงคุณภาพสู่ข้อมูลเชิงปริมาณ

ในงานวิจัยนี้จะพบว่าข้อมูลนำเข้าของสาขาวิชาและอาชีพของบิดามารดานั้น เป็นข้อมูลเชิงคุณภาพซึ่งไม่สามารถจะใช้เข้าสู่การทำงานเพื่อสร้างตัวแบบจำลองได้ จึงจำเป็นที่จะต้องดำเนินการปรับแต่งข้อมูลให้อยู่ในลักษณะข้อมูลเชิงปริมาณ โดยผู้วิจัยได้กำหนดค่าให้ G คือจำนวนกลุ่มของตัวแปร (ตารางที่ 3.1) แสดงรายละเอียดตัวแปรหุ่นของ Major จำนวน 8 กลุ่ม โดยมีการกระทำของตัวแปรสาขาวิชาเป็นตัวแปรทวิภาค (Dichotomous) ชุดใหม่ ที่มีจำนวน G-1 ตัว โดยมีการเลือกกลุ่ม G8 (นิสิตสาขาเทคโนโลยีสารสนเทศ) เป็นกลุ่มอ้างอิง (Reference group) เพื่อใช้เป็นกลุ่มเปรียบเทียบ และ (ตารางที่ 3.2) แสดงรายละเอียดการจัดทำดัมมีของตัวแปร Faculty โดยแบ่งออกเป็น 4 กลุ่ม โดยมี Archi เป็นกลุ่มอ้างอิง อย่างไรก็ตาม เราจะพบว่า ตัวแปร Faculty เป็นตัวแปรที่มีความเหมาะสมมากกว่าตัวแปร Major จึงจำเป็นจะต้องตัดตัวแปร Major ทิ้งก่อนที่จะดำเนินการทำ Multiple regression ในลำดับถัดไป (ตารางที่ 3.3) ทั้งนี้เราจึงไม่จำเป็นจะต้องทำการ dummy ตัวแปร Major

ตารางที่ 3.1 แสดงตัวอย่าง Dummy variable coding ของตัวแปรของคณะวิทยาศาสตร์ภายใต้สาขาวิชา (Major)

Group	Major	X_1	X_2	X_3	X_4	X_5	X_6	X_7
G1	MATH	1	0	0	0	0	0	0
G2	STAT	0	1	0	0	0	0	0
G3	CHEM	0	0	1	0	0	0	0
G4	BIO	0	0	0	1	0	0	0
G5	CO_2	0	0	0	0	1	0	0
G6	PHYSICAPP	0	0	0	0	0	1	0
G7	PHYSIC	0	0	0	0	0	0	1
(Reference) G8	IT	0	0	0	0	0	0	0

ตารางที่ 3.2 แสดงตัวอย่าง Dummy variable coding ของตัวแปร Faculty

Faculty	Sci	Engi	Aggri
Sci(0)	1	0	0
Engi (1)	0	1	0
Aggri (3)	0	0	1
Archi (2) (Reference group)	0	0	0

ตารางที่ 3.3 แสดงรายละเอียดบางส่วนของข้อมูลนำเข้าทดสอบการทำ Multiple regression

sci	engi	aggri	GPAX	Sex	VeryPoor	Poor	Medium	High	cousineN	statusHist	riskForPay
1	0	0	3.5	0	0	0	1	0	0	0	2
1	0	0	2.54	0	0	0	1	0	1	0	3
1	0	0	2.71	0	1	0	0	0	1	0	2
1	0	0	2.64	0	1	0	0	0	0	0	4
1	0	0	2.58	0	0	0	0	0	0	0	1
1	0	0	3.06	0	0	1	0	0	1	0	1
1	0	0	2.01	1	1	0	0	0	1	0	2
1	0	0	2.61	0	0	0	1	0	1	1	0
1	0	0	2.53	0	1	0	0	0	0	1	3
1	0	0	3.01	0	1	0	0	0	0	1	2
1	0	0	2.9	0	1	0	0	0	0	1	2
1	0	0	2.28	1	1	0	0	0	1	1	2

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.456	8	5.057	2.130	.033[b]
	Residual	681.405	287	2.374		
	Total	721.861	295			

a. Dependent Variable: riskForPay

b. Predictors: (Constant), statusHist, cousineN, engi, Income, GPAX, Sex, aggri, sci

ภาพที่ 3.8 แสดงผลข้อมูลทดสอบด้วยการวิเคราะห์สมการถดถอยพหุคูณ

ตารางที่ 3.4 แสดงค่าสัมประสิทธิ์มาตรฐานของตัวแปร

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	5.448	.833		6.540	.000
Sci	-1.243	.358	-.379	-3.468	.001
Engi	-1.156	.366	-.331	-3.158	.002
Aggri	-1.098	.370	-.318	-2.964	.003
GPAX	-.770	.256	-.185	-3.012	.003
Sex	-.144	.198	-.045	-.729	.016

จากการทำการวิเคราะห์ข้อมูลนำเข้าด้วยวิธี multiple regression แบบวิธี Enter เราจะพบว่า ค่า Standardized Coefficients โดยมีตัวแปรอิสระทั้งหมด 8 ตัวแปร ซึ่งหากพิจารณาค่า Beta จะพบว่า ตัวแปร Sci ซึ่งเป็นตัวแปรอิสระมีผลต่อตัวแปรตาม (riskForPay) มีค่ามากที่สุดในทิศทางผกผันกลับนั่นคือ หากเป็นนิสิตคณะวิทยาศาสตร์จะมีความเสี่ยงต่อการสูญเสียหนี้น้อยที่สุด และนอกจากนี้ยังมีตัวแปร Income มีค่าน้อยสุดในทิศทางแปรผันตรงกับตัวแปรตาม (Dependent variable: riskForPay) แต่อย่างไรก็ตาม ตัวแปรเหล่านี้จะถูกนำไปใช้สร้างตัวแปบบจำลอง

3.5 การกำหนดค่าฟัซซีแอททริบิวต์ (Fuzzy attribute values)

งานวิจัยได้ใช้เทคนิคการทำเหมืองข้อมูลที่ใช้รูปแบบการจำแนกข้อมูล(Data classification) สำหรับฝึกสอนการเรียนรู้ข้อมูล (Training data)และการทดสอบข้อมูล (Testing data) เราได้นำเสนอการใช้เทคนิคโครงข่ายประสาทเทียม (Neural net) ที่มีบางแอททริบิวต์มีลักษณะเป็นแอททริบิวต์เชิงฟัซซี (Fuzzy attribute) ซึ่งเราพบว่า ตัวแปร Income นั้น เป็นตัวแปรที่สามารถแสดงค่าข้อมูลในลักษณะฟัซซีได้ ตัวอย่างเช่น รายได้สูง รายได้น้อย รายได้ 35K (Bahts) เป็นต้น เพราะฉะนั้นเพื่อให้รูปแบบของการรับข้อมูลนำเข้าจากแอททริบิวต์ดังกล่าวนี้ได้ เราจึงนำเสนอรูปแบบข้อมูลที่ใช้ 2 รูปแบบ นั่นคือ รูปแบบที่1 คือ ค่าพจน์ภาษาฟัซซีได้แก่ Very High (VH), High (H), Medium (M), Poor (P) และ Very Poor (VP) ตามลำดับ โดยสอดคล้องกับค่าที่กำหนดไว้ในรูปแบบสี่เหลี่ยมคางหมู (Trapezoidal type) ที่มีการกำหนดขอบเขตของข้อมูลภายใต้ตัวแปรα, β, γ และδ ตามลำดับ (ดังตารางที่ 5) สำหรับรูปแบบที่2 คือ ค่าคริสป์ (Crisp value) เช่น Income เท่ากับ 25.4K เป็นต้น ซึ่งการสร้างแบบจำลองนี้ เราได้กำหนดให้ค่าตัวแปรอิสระอยู่ในรูปแบบของพจน์ภาษาฟัซซี แต่เมื่อผู้ใช้หากต้องการนำเข้าค่าตัวแปรอิสระใดๆ ก็สามารถนำเข้าในรูปแบบพจน์ภาษาฟัซซีหรือค่าคริสป์ ได้เช่นเดียวกันซึ่งระบบสามารถรองรับการทำงานค่าทั้งสองรูปแบบดังกล่าว

ตารางที่ *3.5* แสดงค่าลักษณะพจน์ภาษาเชิงฟัซซีของแอททริบิวต์ $attr_{006}$: "Income"

fuzzy ID	label name	attrName	type	Distribution Name			
				α	β	γ	δ
006	Income	Very High	Trapezoidal	0.70	0.75	0.85	1.00
		High	Trapezoidal	0.50	0.55	0.80	0.90
		Medium	Trapezoidal	0.35	0.40	0.60	0.75
		Poor	Trapezoidal	0.20	0.25	0.45	0.55
		Very Poor	Trapezoidal	0.00	0.15	0.30	0.45

3.6 การสร้างแบบจำลองด้วยเทคนิคสมการถดถอยเชิงเส้นพหุคูณ (Multiple regression)

การสร้างแบบจำลองด้วยเทคนิคสมการถดถอยเชิงเส้นพหุคูณเราใช้ตัวแปรอิสระและตัวแปรตาม เหมือนกับการสร้างตัวแบบจำลองด้วยโครงข่ายประสาทเทียม (Artificial Neural Network) ซึ่งการคัดเลือกตัวแปรนั้นใช้วิธีการคัดเลือกแบบขั้นตอน(Stepwise Selection) เป็นวิธีการคัดเลือกตัวแปรที่ใช้การทำงานร่วมกันระหว่างแบบก้าวหน้าและแบบถอยหลังร่วมกัน สำหรับวิธีการคัดเลือกนั้น เราเริ่มต้นจากทำการเลือกตัวแปรที่มีค่าสัมประสิทธิ์สหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการเป็นลำดับแรก แล้วจึงทำการคัดเลือกตัวแปรที่ไม่ได้อยู่ในสมการ โดยมีข้อกำหนดคือ มีตัวแปรใดที่มีสิทธิ์เข้ามาอยู่ในสมการด้วยวิธีการคัดเลือกแบบก้าวหน้า (Forward Selection) และขณะเดียวกันก็ให้คัดเลือกตัวแปรที่อยู่ในสมการด้วยว่า มีตัวแปรพยากรณ์ที่อยู่ในสมการตัวใดมีโอกาสถูกตัดทิ้งออกจากสมการด้วยวิธีการคัดเลือกแบบถอยหลัง (Backward Selection) โดยเราสามารถทำการคัดเลือกผสมทั้งสองวิธีนี้ในทุกขั้นตอนจนไม่มีตัวแปรใดที่ถูกคัดออกจากสมการและไม่มีตัวแปรใดที่ถูกนำเข้าสมการเราถือว่าจบวิธีการและได้สมการถดถอยที่มีสัมประสิทธิ์การพยากรณ์สูงสุด โดยใช้สูตรสมการในการคำนวณหาสมการถดถอยเชิงเส้น (ดังสมการ 4) เมื่อได้ตัวแปรจากกระบวนการทำดังกล่าว เราสามารถนำไปคำนวณหาประสิทธิภาพการทำงานโดยใช้รูปแบบการทำงานบนโปรแกรม RapidMiner Studio 9.2 ซึ่งสามารถแสดงรายละเอียดขั้นตอนการทดสอบได้ (ดังภาพที่ 3.9)

$$Y' = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k \qquad (4)$$

เมื่อ Y' คือ ค่าพยากรณ์ของตัวแปรตาม (Dependent variable)

a คือ ค่าคงที่ของสมการพยากรณ์ในรูปแบบข้อมูลดิบ

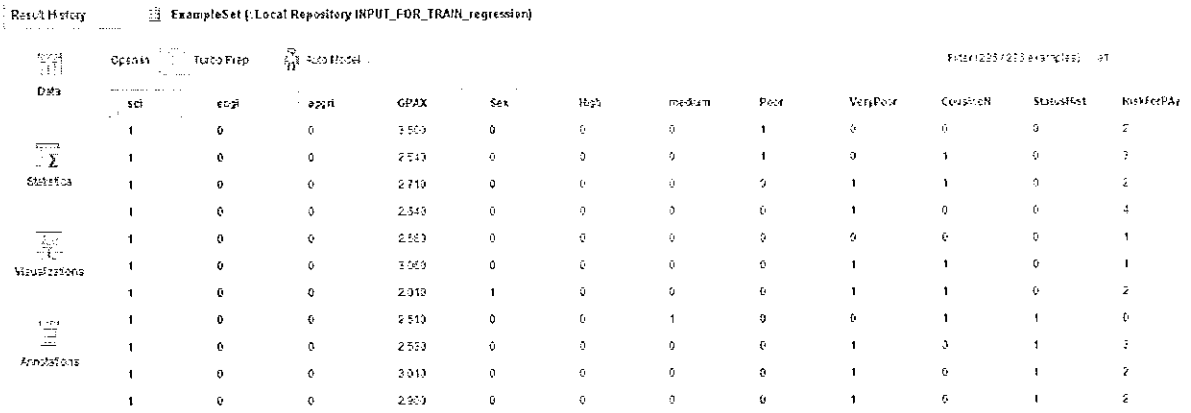
$b_1, b_2, \ldots, b_k$ คือ ค่าน้ำหนักคะแนนหรือสัมประสิทธิ์การถดถอยของตัวแปรพยากรณ์ ตัวที่ 1 ถึงตัวที่ k ตามลำดับ

$X_1, X_2, \ldots, X_k$ คือ ข้อมูลของตัวแปรพยากรณ์ (Predictors: Constant) ตัวที่ 1 ถึงตัวที่ k ตามลำดับ

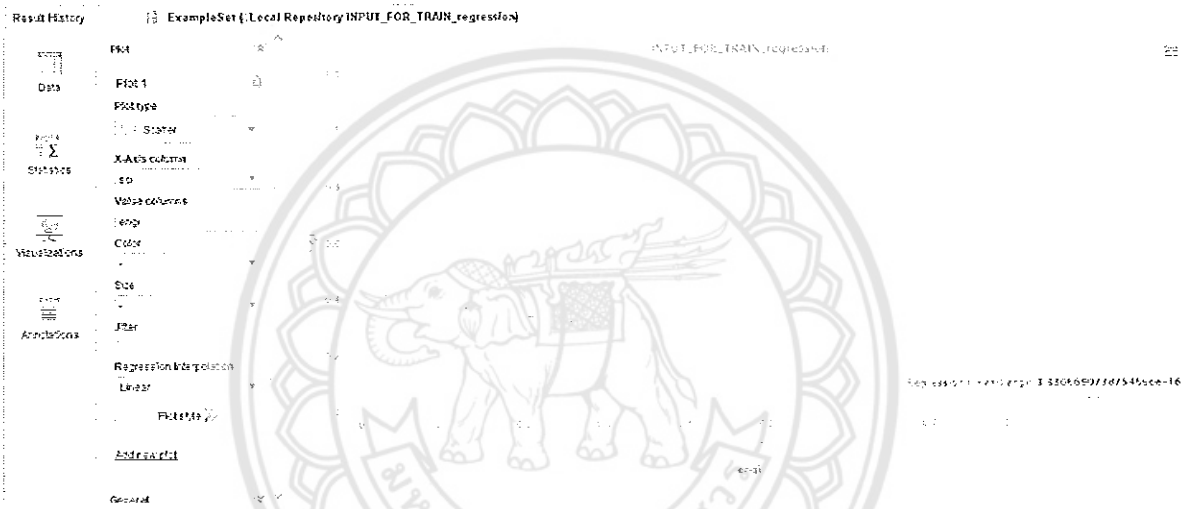
k คือ จำนวนตัวแปรพยากรณ์

การนำข้อมูลเข้าสู่ RapidMiner Studio

เราจะนำข้อมูลชื่อ INPUT_FOR_TRAIN_regression.csv เข้าสู่โปรแกรม RapidMiner ซึ่งได้ทำการกรองข้อมูลเรียบร้อยแล้ว ซึ่งลักษณะของข้อมูลที่นำเข้าสามารถแสดงรายละเอียดได้ดังภาพที่ 3.9 นอกจากนี้เราสามารถตรวจสอบข้อมูลด้วยการแสดงผลจากการทำ Visualizations ของข้อมูลนำเข้าดังภาพที่ 3.10 ตามลำดับ

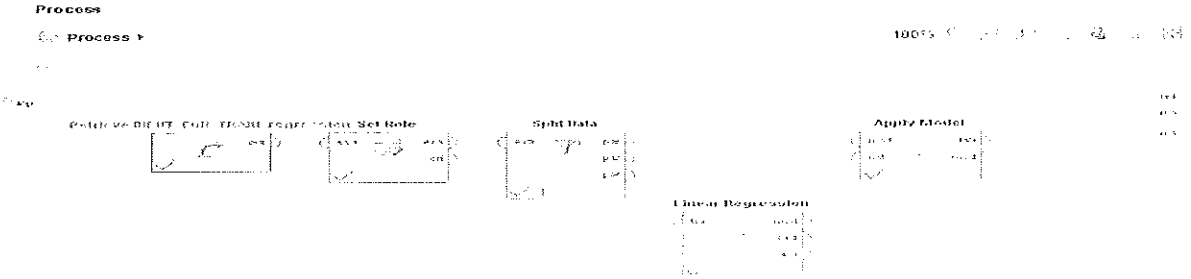


ภาพที่ 3.9 แสดงข้อมูลนำเข้าในโปรแกรม RapidMiner Studio

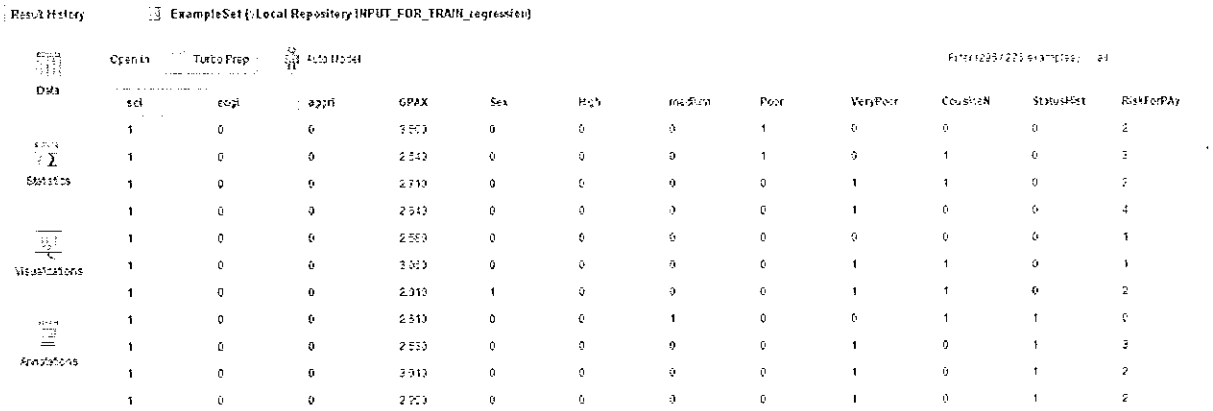


ภาพที่ 3.10 แสดงข้อมูลนำเข้าในรูปแบบ Visualizations รูปแบบLinear regression

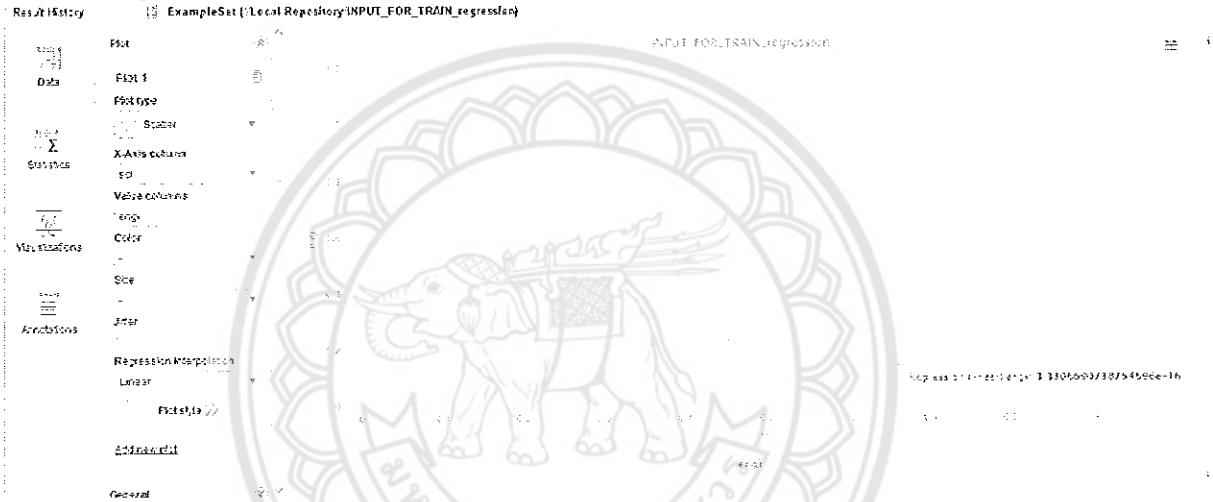
การพัฒนาตัวแบบจำลองของ MLR สามารถดำเนินการด้วย RapidMiner Studio 9 (ดังภาพที่ 3.11) ซึ่งรายละเอียดในการพัฒนาเริ่มต้นจากการรูปแบบการพัฒนาจะมี 2 รูปแบบนั่นคือ รูปแบบที่แรก จะแบ่งข้อมูลสำหรับการจำแนกด้วยลักษณะ split data และรูปแบบที่สองในลักษณะ Validation



ภาพที่ 3.11 แสดงรายละเอียดขั้นตอนการทำ Split data ของ MLR

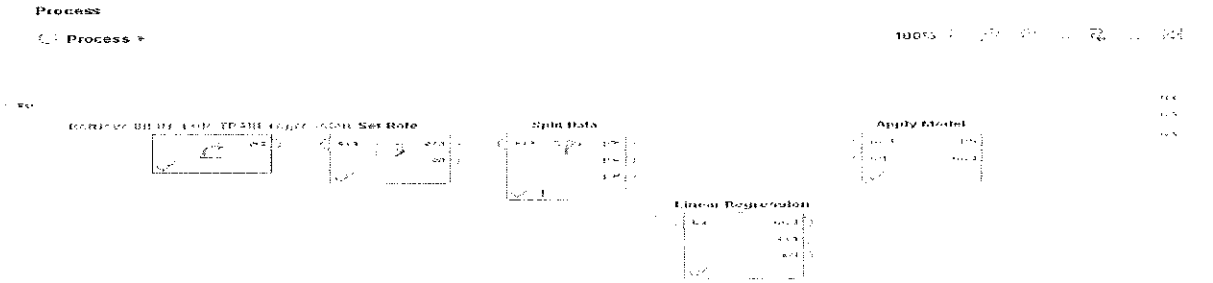


ภาพที่ 3.9 แสดงข้อมูลนำเข้าในโปรแกรม RapidMiner Studio



ภาพที่ 3.10 แสดงข้อมูลนำเข้าในรูปแบบ Visualizations รูปแบบLinear regression

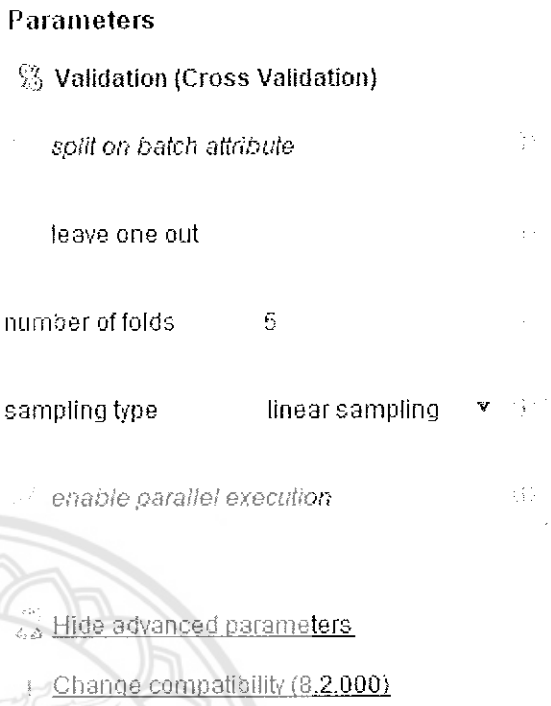
การพัฒนาตัวแบบจำลองของ MLR สามารถดำเนินการด้วย RapidMiner Studio 9 (ดังภาพที่ 3.11) ซึ่งรายละเอียดในการพัฒนาเริ่มต้นจากการรูปแบบการพัฒนาจะมี 2 รูปแบบนั่นคือ รูปแบบที่แรก จะแบ่งข้อมูลสำหรับการจำแนกด้วยลักษณะ split data และรูปแบบที่สองในลักษณะ Validation



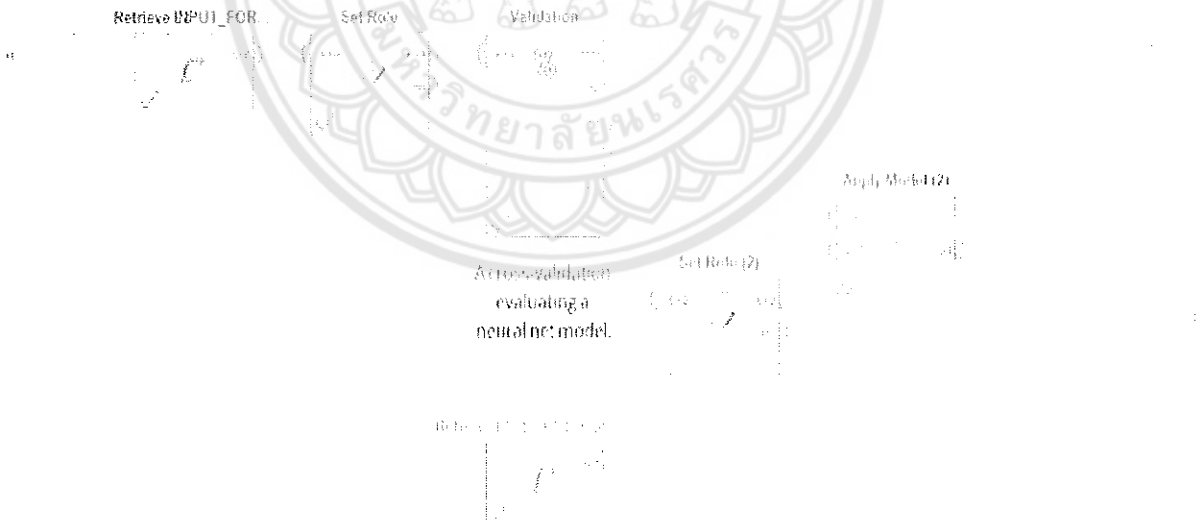
ภาพที่ 3.11 แสดงรายละเอียดขั้นตอนการทำ Split data ของ MLR

สำหรับการทำ Split data (ดังภาพที่ 3.12) จะเป็นการกำหนดให้ Set role คือ RiskForPay (ภาพที่ 3.13)

ภาพที่ 3.12 แสดง set role คือ RiskForPay



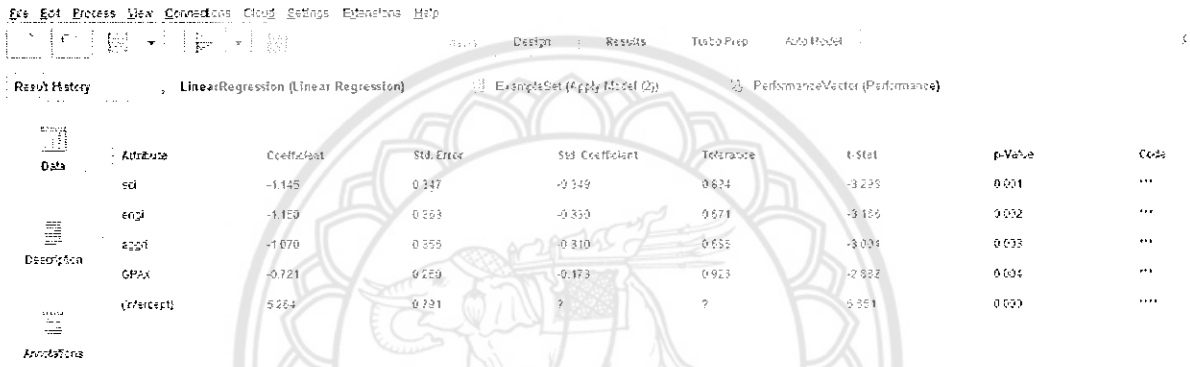
ภาพที่ 3.13 แสดง การแบ่ง cross folds=5



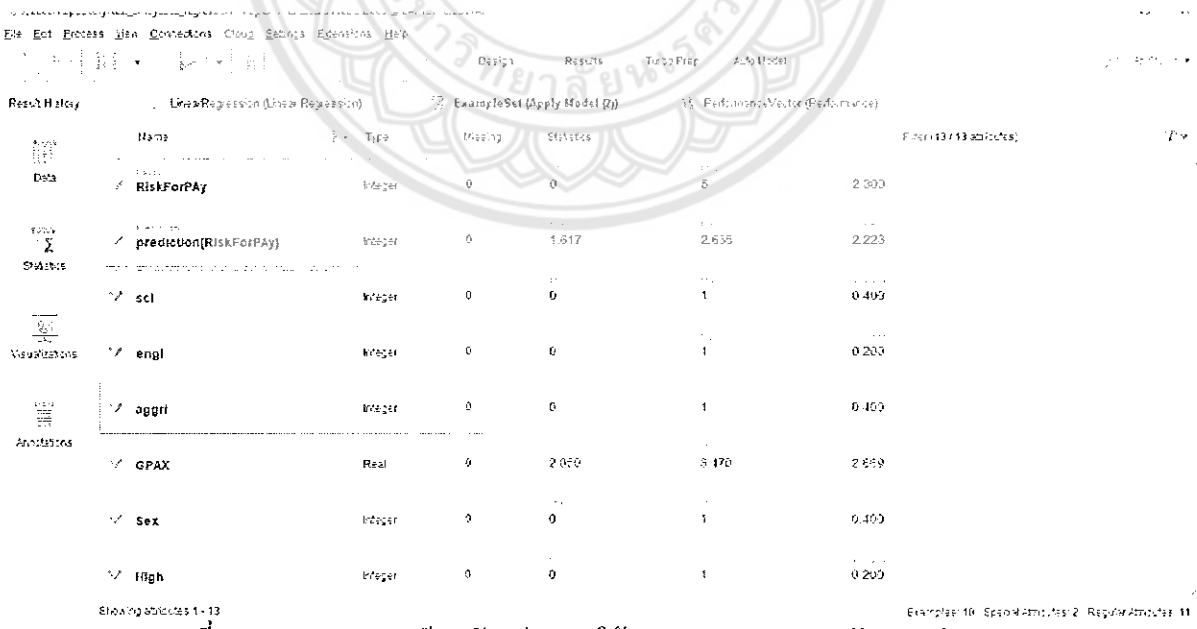
ภาพที่ 3.14 แสดงรายละเอียดลำดับการทำงาน MLR ด้วยการจำแนกข้อมูลแบบ Cross validation fold

รายละเอียดการสร้างตัวแบบจำลอง MLR ด้วยการแบ่งข้อมูลแบบ Cross validation fold จะมีรายละเอียดๆ ของการทดสอบด้วย folds=5 และ 10 ตามลำดับ (ดังภาพที่ 3.14 และ 3.15 เพื่อทำการเปรียบเทียบระหว่างการจำแนกข้อมูลในหลากหลายรูปแบบ (ดังภาพที่ 3.16-3.18)

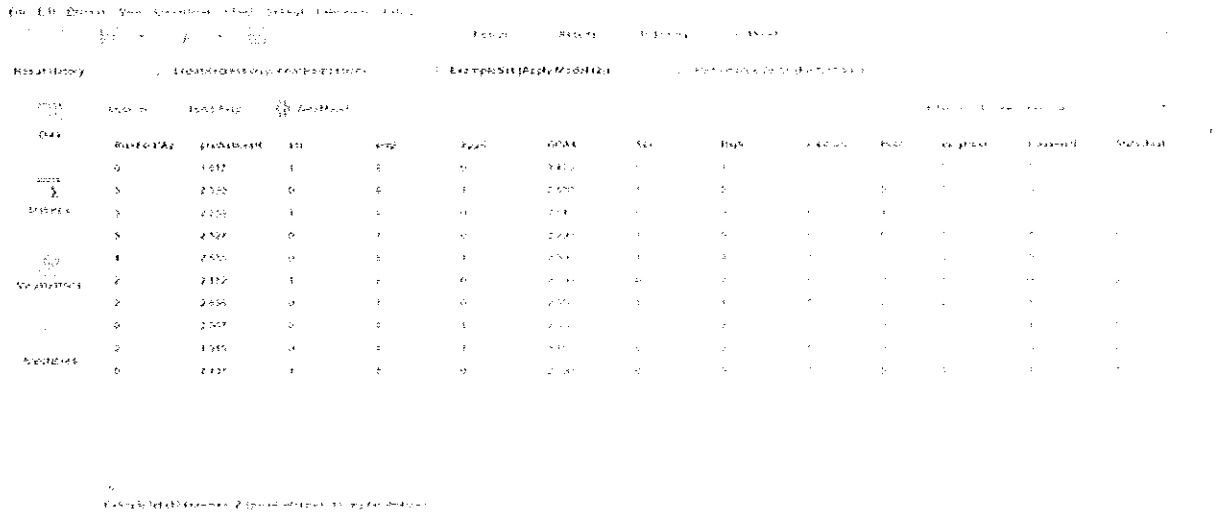
ภาพที่ 3.15 แสดงรายละเอียดลำดับการทำงาน MLR ด้วยการจำแนกข้อมูลแบบ Cross validation fold (ต่อ)



ภาพที่ 3.16 แสดงรายละเอียดตัวอย่างการใช้งาน Data จากตัวแบบทำงาน MLR



ภาพที่ 3.17 แสดงรายละเอียดตัวอย่างการใช้งาน Statistics จากตัวแบบทำงาน MLR



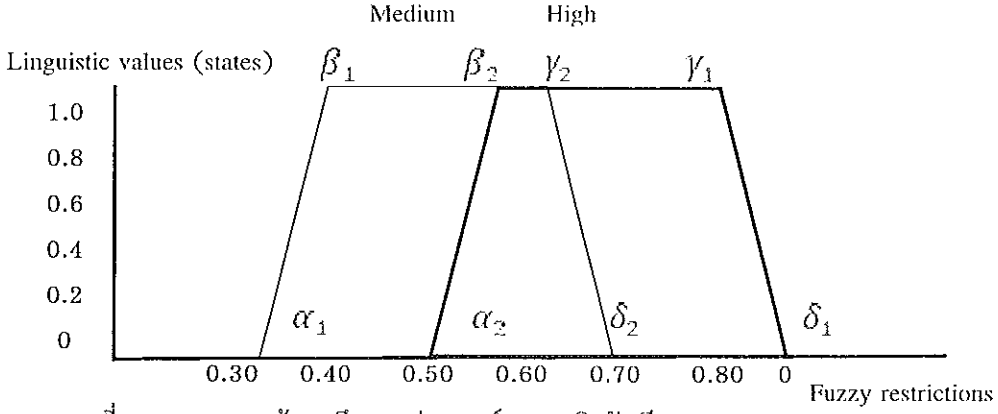
ภาพที่ 3.18 แสดงรายละเอียดตัวอย่างการใช้งาน Apply Model จากตัวแบบทำงาน MLR

3.7 การทำฟัซซีแอททริบิวต์แมทชิง (Fuzzy attribute matching)

3.7.1 การทำฟัซซีแอททริบิวต์แมทชิงระหว่าง พจน์ภาษาฟัซซี (Fuzzy linguistic term) กับพจน์ภาษาฟัซซี(Fuzzy linguistic term)

การทำ Fuzzy attribute matching ระหว่างพจน์ภาษาฟัซซี (Fuzzy linguistic term) กับพจน์ภาษาฟัซซี (Fuzzy linguistic term จะทำการเปรียบเทียบกันระหว่างค่าพจน์ภาษาฟัซซี 2 ค่าใดๆ ในแอทริบิวต์เดียวกัน ซึ่งวิธีการหาค่าดีกรีความคล้ายคลึง ระหว่าง 2 ฟัซซีเซ็ตใดๆ โดยเราจะได้ผลลัพธ์ความคล้ายคลึงของค่าดีกรีระหว่างคู่พจน์ภาษาฟัซซีที่มีค่ามากกว่าหรือเท่ากับค่า t (Threshold)การวัดค่าความคล้ายคลึง (similarity) ของ 2 ฟัซซีเซตใด ๆ ในงานวิจัยนี้เราได้กำหนดรูปแบบการซ้อนทับกันของกราฟฟังก์ชันสมาชิก กำหนดให้ฟัซซีเซตคือ A_1 และ A_2 สอดคล้องกับตำแหน่งกึ่งกลาง (center) c_1และ c_2 และ boundary point เป็น $\alpha_i, \beta_i, \gamma_i, \delta_i \ (i = 1,2)$ โดยที่ $c_i = (\gamma_i + \beta_i)/2$ ซึ่ง สมมติให้ $c_2 > c_1$ มีรายละเอียดดังนี้

<u>กำหนดให้</u> $(\alpha_1 < \alpha_2, \delta_1 > \delta_2, \beta_1 \leq \beta_2, \gamma_2 \leq \gamma_1)$ สำหรับกรณีนี้ ฟัซซีแอททริบิวต์ A_2 ถูกรวมอยู่ใน ฟัซซีแอททริบิวต์ A_1 นั่นคือ $A_2 \subset A_1$ คำนวณได้ (ดังสมการ 5) โดยค่าความคล้ายคลึง (Similarity) ของฟัซซีเซต แสดงได้ (ดังภาพที่ 3.19)



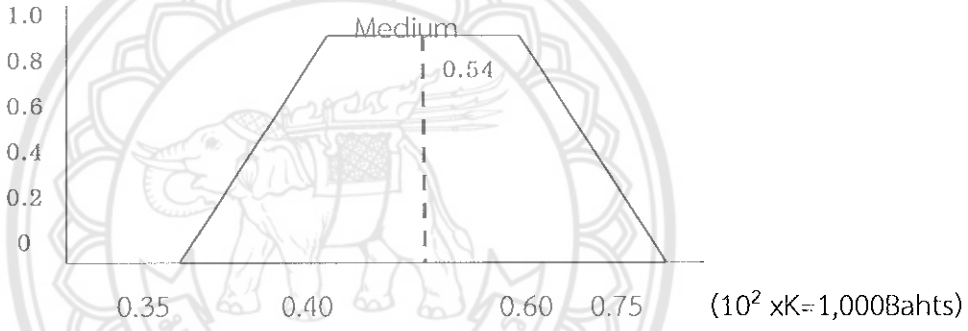
ภาพที่ *3.19* ความคล้ายคลึงระหว่างพจน์ภาษาเชิงฟัซซี "Medium" กับ "High"

$$S = \frac{|A_2|}{|A_1|} = \frac{W_2}{W_1} \quad \text{หรือก็คือ} \quad S = \frac{\delta_2 - \alpha_2 + \gamma_2 - \beta_2}{\delta_1 - \alpha_1 + \gamma_1 - \beta_1} \tag{5}$$

ซึ่ง $W_1 = W_{b_1} + W_{t_1}$, $W_2 = W_{b_2} + W_{t_2}$

3.7.2 การทำฟัซซีแอททริบิวต์แมทชิงระหว่าง ค่าคริสป์ (crisp value)กับพจน์ภาษาฟัซซี (fuzzy linguistic term)

การทำ Fuzzy attribute matching ระหว่างค่าคริสป์ (Crisp value) กับพจน์ภาษาฟัซซี (Fuzzy linguistic term) นั่นคือเป็นการเปรียบเทียบค่าคริสป์กับพจน์ภาษาฟัซซี เพื่อต้องการทราบว่ามีค่าระดับดีกรีความเป็นสมาชิกที่เท่าใด เราสามารถแสดงตัวออย่างการหาค่าคริสป์เพื่อเปรียบเทียบกับค่าพจน์ภาษาฟัซซี โดยกำหนดให้ค่าคริสป์และค่าพจน์ภาษาฟัซซีของแอททริบิวต์ Income คือ 0.54 (10^2 xK (Bahts)) และ 'Medium' ตามลำดับ สามารถแสดงรายละเอียดได้ (ดังภาพที่ 3.20)



ภาพที่ 3.20 แสดงการเปรียบเทียบค่าของฟัวซีแอททริบิต์ Income ระหว่างค่า 'Medium กับ 0.54K

สำหรับการทำแมทชิงระหว่าง "Medium" กับ"0.54" เราใช้ข้อมูลจากค่าฟังก์ชันสมาชิก (membership function) คือ $\alpha, \beta, \gamma, \delta$ เพื่อทำการเปรียบเทียบว่า ค่า "0.54" นั้น มีค่าระดับดีกรีความเป็นสมาชิกเท่าใดของค่าพจน์ภาษาฟัซซี "Medium" ทั้งนี้ค่าของการเปรียบเทียบจะใช้เงื่อนไขมีรายละเอียดคือ (ดังสมการ 6)

$$\mu(x; \alpha, \beta, \gamma, \delta) = \begin{cases} 1 & when \quad \beta \le x \le \gamma \\ 0 & when\, x \le \alpha \; or \; x \ge \delta \\ \frac{\alpha - x}{\alpha - \beta} & when \quad \alpha < x < \beta \\ \frac{\delta - x}{\delta - \gamma} & when \quad \gamma < x < \delta \end{cases} \tag{6}$$

โดยที่ x คือค่าคริสป์ใด ๆ

เราได้ทำการเปรียบเทียบค่าดีกรีความเป็นสมาชิก (membership degree) ระหว่างค่าคริสป์กับค่าพจน์ภาษาฟัซซี หากผลการคำนวณมีค่ามากกว่าหรือเท่ากับค่า threshold ก็จะผ่านเงื่อนไขยอมรับ ดังนั้นผลลัพธ์ที่ได้ เราพบว่าเมื่อใช้ฟัซซีแอททริบิวต์ Income ค่า 0.54 เปรียบเทียบกับค่าความคล้ายคลึงพจน์ภาษาฟัซซี

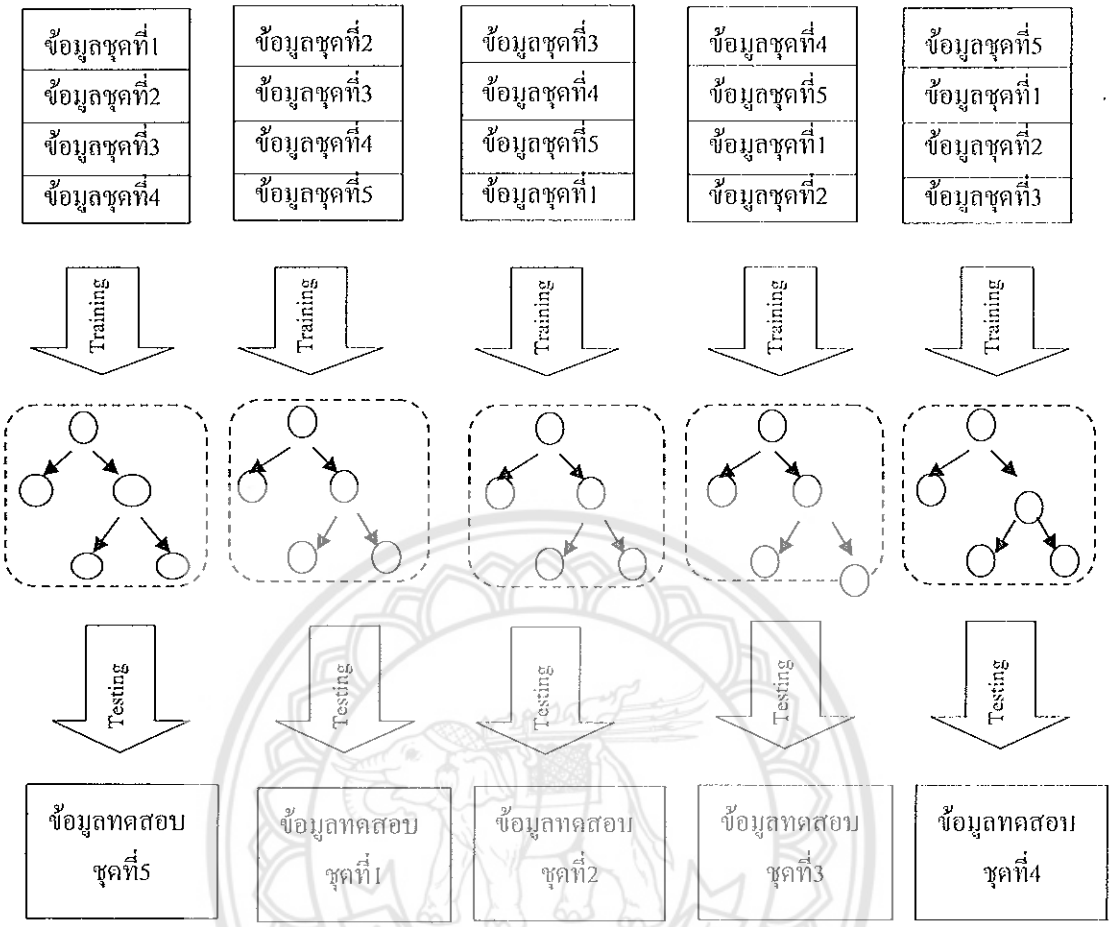
อื่นๆ ได้แก่ Very High, High, Medium, Poor และ Very Poor มีค่าเท่ากับ 0, 0.8, 1, 0.1 และ 0 ตามลำดับ เพราะฉะนั้นค่าฟัซซีแอททริบิวต์ Income ค่า 0.54 ได้ค่าใกล้เคียงมากที่สุดกับพจน์ภาษาฟัซซี "Medium" ซึ่งเราเป็นเหตุผลที่เราเลือกใช้ค่าน้ำหนักของแอทรริบิวต์ Medium เพื่อคำนวณด้วยโครงข่ายประสาทเทียมในงตัวแบบจำลองที่สร้างขึ้นสำหรับใช้พยากรณ์ผล เป็นต้น

3.8 หลักการทำงานของ Neural network

ในงานวิจัยนี้ ได้ดำเนินการเพิ่มจำนวนแอททริบิวต์ด้วยการทำดัมมี่ จากตัวแปรพยากรณ์ที่มีลักษณะเป็นตัวแปรเชิงคุณภาพ โดยเพิ่มจำนวนตัวแปรนำเข้า จาก 6 ตัว เป็น 12 ตัว แล้วผ่านการกลั่นกรองความสัมพันธ์ด้วยเทคนิคสมการสหสัมพันธ์เพียร์สันจะได้แปรพยากรณ์จำนวน 8 ตัว เพื่อเป็นปัจจัยร่วมสำหรับใช้สร้างตัวแบบจำลองพยากรณ์ ดังนั้น ในงานวิจัยนี้ตัวแบบจำลองพยากรณ์สามารถรองรับการนำเข้าข้อมูลเพื่อทดสอบการพยากรณ์ภายใต้ลักษณะข้อมูลนำเข้าได้ทั้ง 2 รูปแบบ นั่นคือ ค่าคริสป์ และค่าพจน์ภาษาฟัซซี ได้ตามลำดับ แล้วเลือกเป็นโหนดตัวแทนเพื่อใช้ดำเนินการคำนวณของการสร้างผลลัพธ์ของการพยากรณ์ ซึ่งรายละเอียดการทำงานของตัวแบบจำลองที่จะต้องดำเนินการเพื่อสร้างตัวแบบจำลองนั้นสามารถแสดงรายละเอียดทั้งรูปแบบของ Fuzzy neural network model และรูปแบบของ Multiple linear regression model ตามลำดับ ภายใต้การดำเนินการเพื่อเลือกรูปแบบการแบ่งข้อมูลเพื่อใช้วัดประสิทธิภาพของตัวแบบจำลอง โดย ผู้วิจัยได้เลือกการแบ่งข้อมูล 2 รูปแบบหลักๆ นั่นคือรูปแบบแรกจะใช้ลักษณะ Split Test (ใช้ข้อมูลตัวอย่างฝึกสอน 80% และ 34% โดย ใช้ข้อมูลตัวอย่างทดสอบ 20% และ 66% ตามลำดับ และรูปแบบที่สองคือวิธี Cross-validation Test ที่ผู้วิจัยจะกำหนดค่า K=5 และ K=10 เท่านั้น เพราะมีความเหมาะสมในจำนวนครั้งที่ทำการทดสอบและเป็นที่นิยม เพื่อใช้ในการวัดประสิทธิภาพของโมเดล ในงานวิจัยนี้ ผู้วิจัยขอยกตัวอย่างหลักการแบ่งข้อมูลของผู้วิจัย กรณีการแบ่งข้อมูลใน ลักษณะ 5-fold cross-validation ซึ่งกำหนดให้ค่า K=5 ด้วยการแบ่งข้อมูลออกเป็น 5 ส่วน จากจำนวนข้อมูลตัวอย่างที่นำเสนอในรายงานนี้ 296 ตัวอย่างมีความเหมาะสมเมื่อแบ่งเป็นส่วนๆ จะได้ส่วนละ 59 ตัวอย่าง โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน

- หลักการวนรอบของการทำงานจากข้อมูลตัวอย่างเพื่อนำมาสร้างโมเดลและใช้เป็นโมเดลทำนายเพื่อหาประสิทธิภาพที่ดีที่สุดของโมเดล นั่นก็คือ ผู้วิจัยกำลังดำเนินการดังนี้
  - รอบที่ 1 ใช้ข้อมูลส่วนที่ 1,2,3 และ 4 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 5
  - รอบที่ 2 ใช้ข้อมูลส่วนที่ 2,3,4 และ 5 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 1
  - รอบที่ 3 ใช้ข้อมูลส่วนที่ 3,4,5 และ 1 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 2
  - รอบที่ 4 ใช้ข้อมูลส่วนที่ 4,5,1 และ 2 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 3
  - รอบที่ 5 ใช้ข้อมูลส่วนที่ 5,1,2 และ 3 สร้างโมเดล และใช้โมเดลทำนายข้อมูลส่วนที่ 4

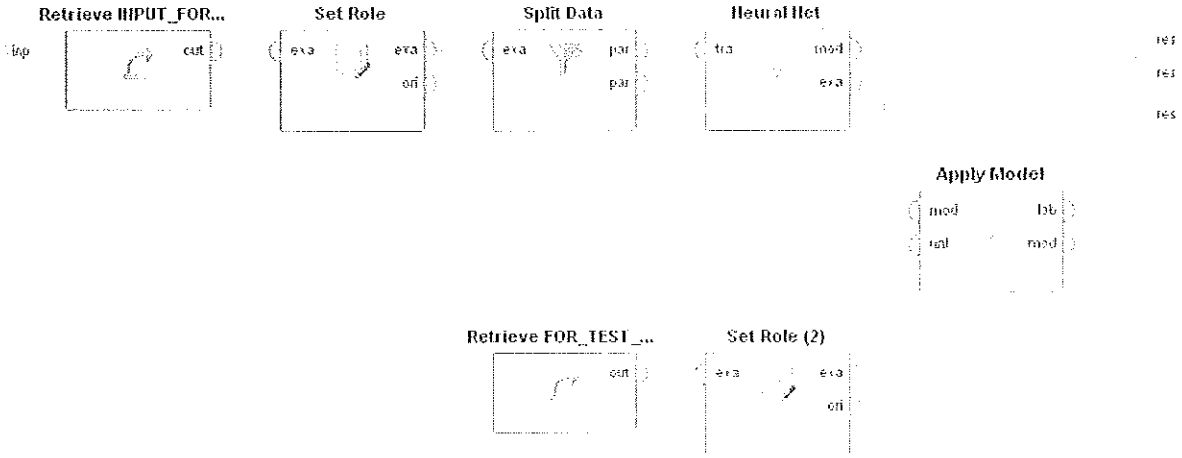
ซึ่งทั้งสองอัลกอริธึมนี้ จะพบว่า การแบ่งข้อมูลสำหรับการเรียนรู้และทดสอบนั้น ในรูปแบบ cross validation fold=5 จะมีความเหมาะสมที่สุด โดยรูปแบบของการดำเนินการเพื่อการแบ่งข้อมูลสำหรับการเรียนรู้นั้นและทดสอบนั้นมีรายละเอียดดังนี้ (ดังภาพที่ 3.21)



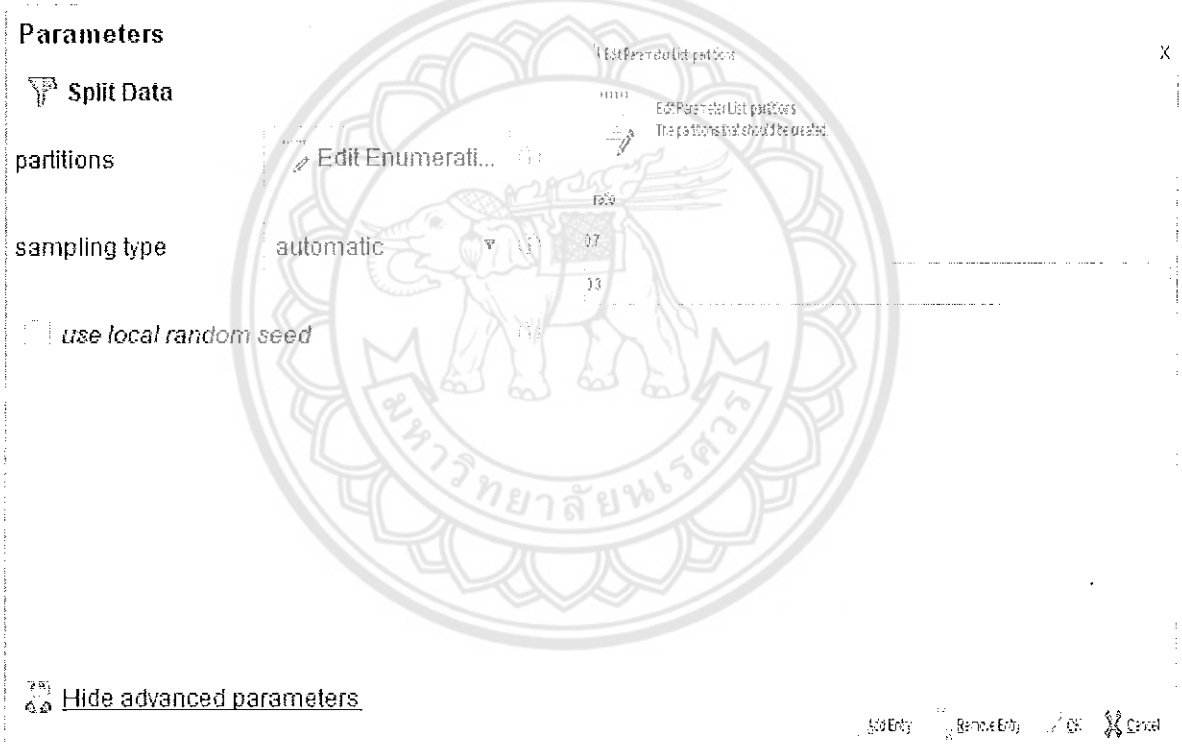
ภาพที่ 3.21 แสดงการแบ่งข้อมูลตัวอย่างใช้ทดสอบประสิทธิภาพตัวแบบจำลองลักษณะ cross validation fold=5

3.8.1 การสร้างแบบจำลองด้วยเทคนิคฟัซซีนิวรอลเน็ตเวิร์ค

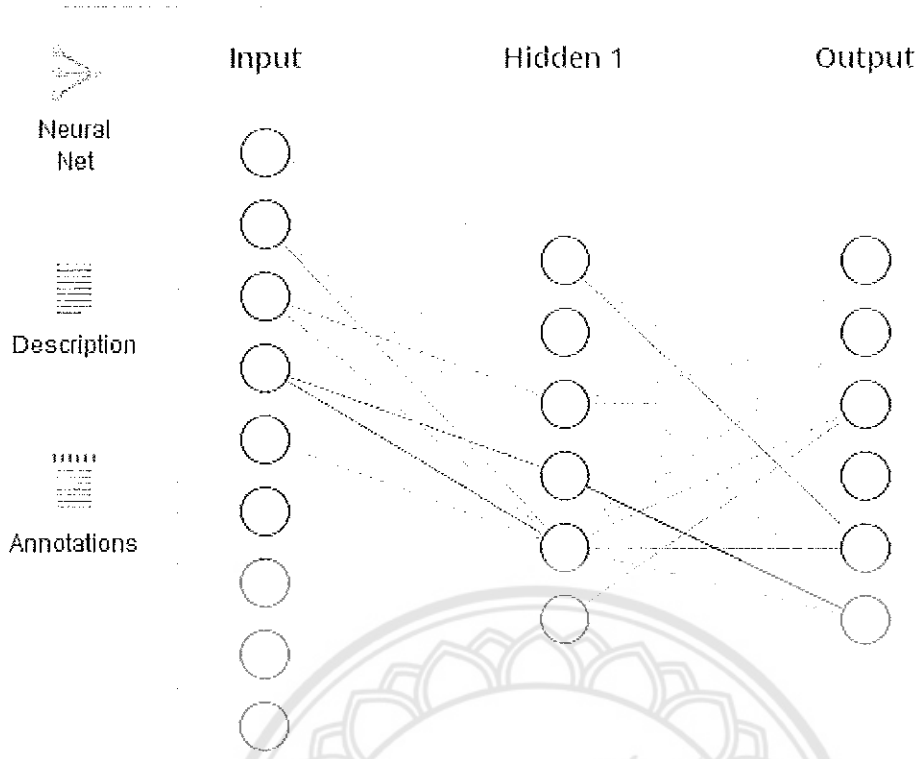
การเลือกรูปแบบการแบ่งข้อมูลเพื่อใช้วัดประสิทธิภาพของตัวแบบจำลอง ในงานวิจัยนี้เราได้เลือกการแบ่งข้อมูล 2 รูปแบบหลักๆ คือรูปแบบ Split Test (ใช้ข้อมูลตัวอย่างฝึกสอน 34% , 70% และ 80% โดยใช้ข้อมูลตัวอย่างทดสอบ 66%, 30% และ20% ตามลำดับ และรูปแบบ Cross validation Test โดยเรากำหนด ค่า Fold=5 และ Fold=10 ตามลำดับ (ดังภาพที่ 3.22)



ภาพที่ 3.22 แสดงการแบ่งข้อมูลแบบ split test สร้างตัวแบบจำลอง neural net



ภาพที่ 3.23 แสดงการแบ่งข้อมูล split test 70:30



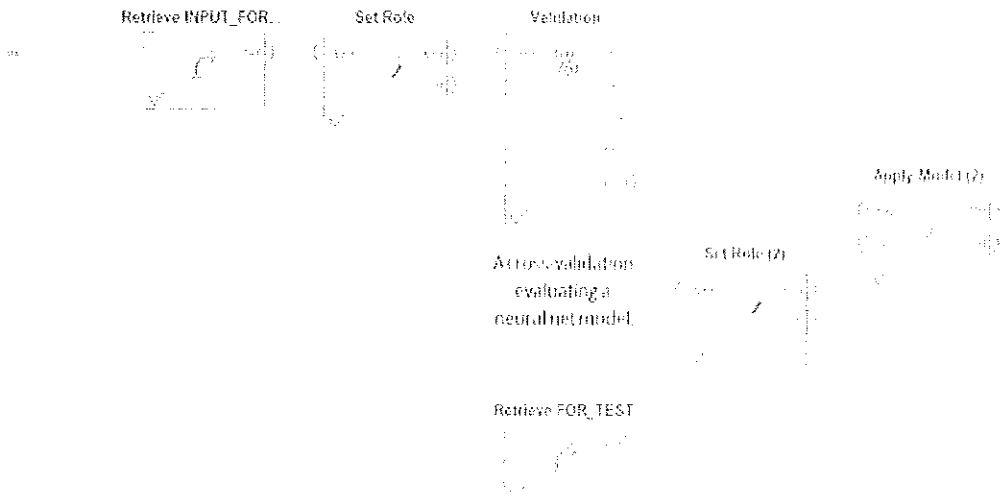
ภาพที่ 3.24 แสดงตัวอย่างการพัฒนาตัวแบบจำลองด้วย 8-5-6 แบบ split test 70:30

ImprovedNeuralNet

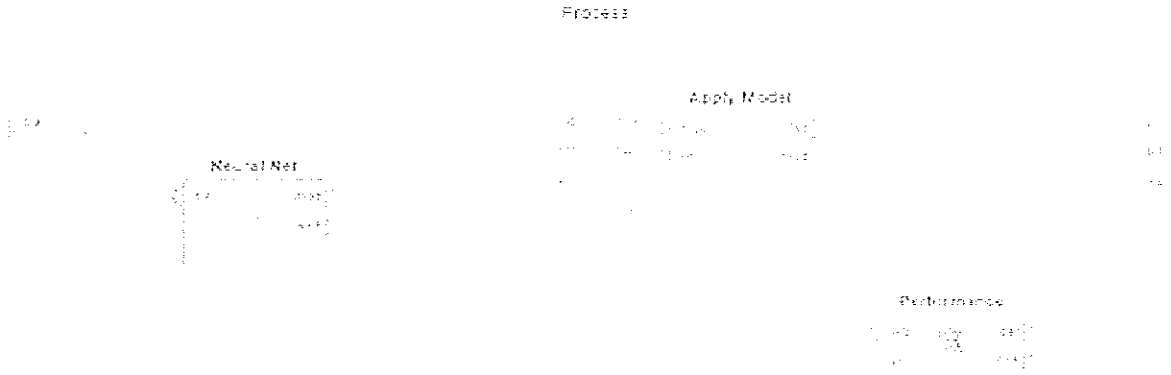
```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: -0.952
engi: -0.089
aggri: -0.538
GPAX: -0.884
Sex: -0.546
Income: 0.318
CousineN: -0.013
StatusHist: 0.115
Bias: -0.497

Node 2 (Sigmoid)
----------------
sci: -0.294
engi: 0.401
aggri: 0.139
GPAX: 0.271
Sex: 0.072
Income: -0.379
CousineN: -0.333
StatusHist: 0.056
Bias: -0.021

.. .

Output
```

```
======

Class 'less' (Sigmoid)
----------------------
Node 1: -0.820
Node 2: -0.152
Node 3: 0.567
Node 4: 0.152
Node 5: -0.893
Threshold: -0.434

Node 5: -1.255
Threshold: -1.271

Class 'verl' (Sigmoid)
----------------------
Node 1: -0.916
Node 2: -0.850
Node 3: -1.041
Node 4: -0.495
Node 5: 0.588
Threshold: -0.732


. . .

Class 'veryhigh' (Sigmoid)
--------------------------
Node 1: 0.253
Node 2: -0.934
Node 3: 0.197
Node 4: -2.301
Node 5: 1.228
Threshold: -0.650
```

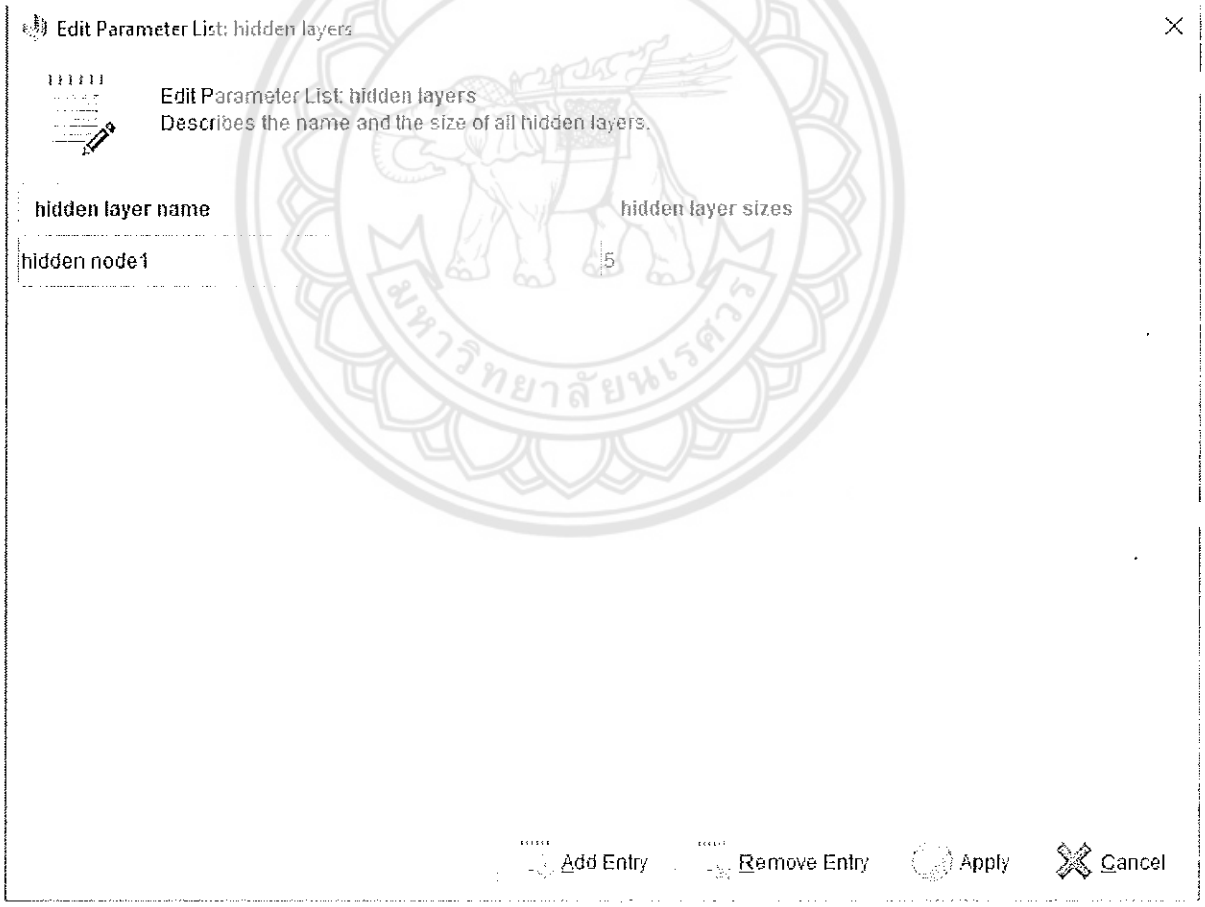
ภาพที่ 3.25 แสดงรายละเอียดบางส่วนของค่าการคำนวณตัวแบบจำลอง



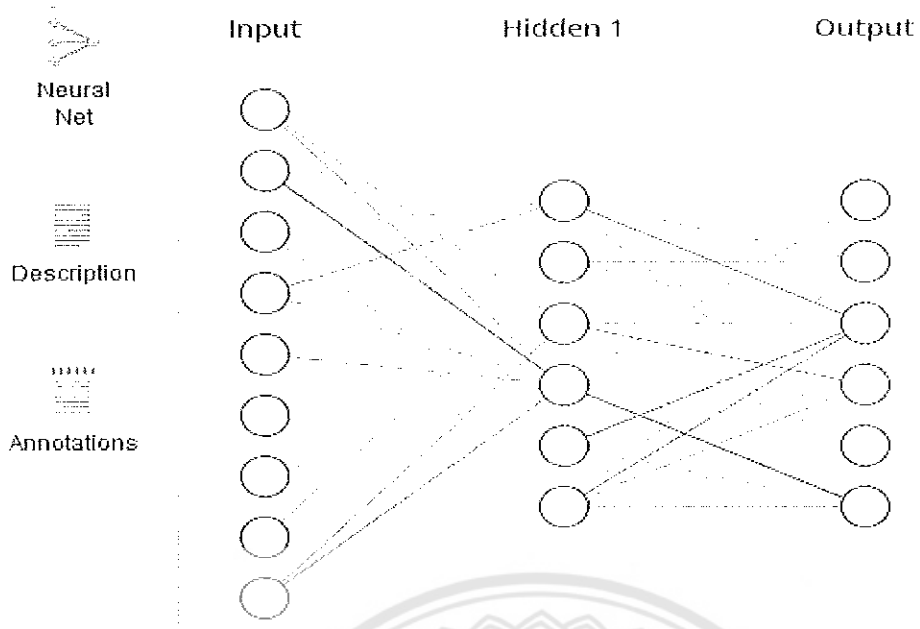
ภาพที่ 3.26 แสดงการพัฒนาตัวแบบจำลองด้วยวิธีการแบ่งข้อมูลแบบ Cross validation fold 5



ภาพที่ 3.27 แสดงการพัฒนาแบบจำลองรูปแบบวิธีแบบ cross validation fold 5 (ต่อ)



ภาพที่ 3.28 แสดงการกำหนด Hidden layer ด้วย Node size=5



ภาพที่ 3.29 แสดงตัวอย่างการพัฒนาตัวแบบจำลองด้วย Cross validation fold=5 hidden size=5

ImprovedNeuralNet

```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: 0.359
engi: -0.460
aggri: 0.772
GPAX: -1.096
Sex: 0.133
Income: -0.109
CousineN: 0.522
StatusHist: -0.102
Bias: -0.184


. . .


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: 0.380
Node 2: -0.637
Node 3: -0.716
Node 4: 0.862
Node 5: 0.169
Threshold: -0.577

Class 'medium' (Sigmoid)
```

```
-------------------------
Node 1: -0.899
Node 2: -0.890
Node 3: 0.004
Node 4: 0.610
Node 5: -0.519
Threshold: -0.825

Class 'high' (Sigmoid)
----------------------
Node 1: -1.342
Node 2: -0.849
Node 3: 0.655
Node 4: -0.788
Node 5: -1.509
Threshold: -1.332

.. .

Class 'veryhigh' (Sigmoid)
--------------------------
Node 1: 0.251
Node 2: 0.555
Node 3: -0.511
Node 4: -1.737
Node 5: -0.897
Threshold: -0.927
```

ภาพที่ 3.30 แสดงรายละเอียดค่าการคำนวณตัวแบบจำลองเพียงบางส่วน

บทที่4

ผลการทดลอง

4.1 การทดสอบประสิทธิภาพของตัวแบบจำลองฟัซซีโครงข่ายประสาทเทียมและตัวแบบจำลองสมการถดถอยเชิงเส้นพหุคูณ

จากการดำเนินวิจัยนี้ พบว่าแบบจำลองตัวพยากรณ์ของนิวรอลเน็ตนั้นคือ อัลกอริธึมโครงข่ายประสาทเทียมมีที่เหมาะสมจะมีโครงสร้างแบบโครงข่ายประสาทเทียมฟัซซี ซึ่งเราใช้ Income=Very Poor (VP) เป็นตัวแปรอ้างอิง ภายการแบ่งข้อมูลในลักษณะ 5-fold cross-validation ค่า K=5 โมเมนตัม 0.2 อัตราการเรียนรู้ 0.3 มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 83.33%, 83.10%, 83.30% และ 82.40% ตามลำดับ มีรายละเอียดค่าน้ำหนักแอททริบิวต์ (weight attributes) ของตัวแบบจำลอง ดังตารางที่ 4.1 และตารางที่ 4.2 ตามลำดับ สำหรับตัวแบบจำลองสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear regression) จะมีรูปแบบสมการที่มีค่าของตัวแปร sci , engi, aggri, GPAX - 0.851 เท่ากับ -0.851,- 0.816,- 0.819 และ - 0.689 และมีค่า Performance vector นั่นคือ ค่า root mean squared error: 1.513 +/- 0.000 และค่า squared correlation: 0.081 ตามลำดับ

ตารางที่ 4.1 ค่าน้ำหนักระหว่างโหนดซ่อนกับโหนดผลลัพธ์

Node 1: 1.436	Node 1: 1.397	Node 1: 1.161
Node 2: -2.927	Node 2: -0.673	Node 2: -0.192
Node 3: 2.875	Node 3: 1.583	Node 3: -0.093
Node 4: 1.326	Node 4: 0.535	Node 4: 0.732
Node 5: -0.870	Node 5: 0.532	Node 5: -0.045
Threshold: -3.151	Threshold: -2.777	Threshold: -2.414

ตารางที่ 4.2 ค่าน้ำหนักระหว่างโหนดข้อมูลนำเข้ากับกับโหนดซ่อน

Attributes	Node1(Sigmoid)	Node2(Sigmoid)	Node3(Sigmoid)	Node 4(Sigmoid)	Node 5(Sigmoid)
sci	1.108	0.381	6.511	-8.500	0.254
engi	-4.994	-0.847	2.169	1.670	6.113
aggri	3.806	-0.314	1.842	-0.077	7.691
GPAX	-2.372	6.976	12.906	8.434	2.353
Sex	-6.841	-0.664	4.639	4.078	4.184
Income	6.783	8.669	9.542	-2.920	-11.275
CousineN	3.972	-5.261	3.651	-9.078	5.077
StatusHist	-0.762	0.561	-1.669	10.417	6.144
Threshold	-6.532	-6.050	-3.864	-4.413	5.753

4.2 การคำนวณผลลัพธ์ของตัวแบบพยากรณ์

การคำนวณเพื่อหาผลลัพธ์ของแบบจำลองนั้น จะแบ่งออกเป็น 2 ขั้นตอน ประกอบไปด้วย

2.1 การแปลงค่าของฟัซซีแอททริบิวต์

การดำเนินการในขั้นตอนนี้มีความจำเป็นอย่างยิ่งที่จะต้องแปลงค่าของฟัซซีแอททริบิวต์ เพื่อนำค่าที่ได้ไปใช้ในการคำนวณการพยากรณ์ของตัวแบบจำลอง จากตารางที่ 4.3 แสดงถึงตัวอย่างเพียงบางส่วนของข้อมูลนำเข้าที่ต้องการทดสอบ เราจะพบว่า ฟัซซีแอททริบิวต์ Income ที่จะต้องคำนวณโดยใช้ (สูตรสมการที่ 4) ดังนั้น การคำนวณค่าของ Input(2) และ Input(3) ด้วย $\mu(x=0.54:\alpha=0.35, \beta=0.40, \gamma=0.60, \delta=0.75)$ และ $\mu(x=0.33:\alpha=0.20, \beta=0.25, \gamma=0.45, \delta=0.55)$ จะพบว่ามีระดับความเป็นสมาชิกใกล้เคียงกับ “Medium” เท่ากับ 1 และ “Poor” เท่ากับ 1 ตามลำดับ

ตารางที่ 4.3 ตัวอย่างการพยากรณ์เงินกู้ กยศ ด้วยเทคนิค TFNN และMLR

	Input (1)	Input(2)	Input(3)	Input(4)	Input(5)
Sci	1	0	0	0	1
Engi	0	0	1	1	0
Aggri	0	1	0	0	0
GPAX	2.55	2.8	3.25	2.78	2.09
Sex	1	1	0	1	0
Income (Fuzzy attribute)	High	0.54 (≅Medium)	0.33(≅Poor)	Low	Medium

CousineN	2	1	1	1	0
StatusHist	1	1	0	0	0
...	...	...	...	...	...

ขั้นตอนการทดสอบแบบจำลองสามารถกระทำได้โดยทดสอบจากข้อมูลนำเข้าจำนวน 5 ตัวอย่างมีรายละเอียดดังนี้

ตัวอย่างของข้อมูลนำเข้า input 1

- ขั้นตอนแรกจะต้องแปลงข้อมูลจากตารางนำเข้า โดยมีรายละเอียดดังนี้ คือ

  range=(max-min)/2 โดย max และ min ได้จากค่ามากสุดและน้อยสุดของแต่ละแอททริบิวต์

  base=(max+min)/2

  norm_attribute= (attribute-base)/range

ดังนั้น แอททริบิวต์ Sci ให้เป็น a1 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a1 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ Engi ให้เป็น a2 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a2 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ Aggri ให้เป็น a3 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a3 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ GPAX ให้เป็น a4 จะได้ คือ

range =(3.58-2/2 =0.79

base =(3.58+2)/2 = 2.79

norm_a4 =(2.55-2.79)/0.79 = -0.30

ดังนั้น แอททริบิวต์ Sex ให้เป็น a5 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a5 =(1-0.50)/0.50 = 1.00

ดังนั้น แอททริบิวต์ Income ให้เป็น a6 จะได้ คือ

range =(1-0)/2 =0.50

base =(1+0)/2=0.50

norm_a6 =(0-0.50)/0.50 = -1.00

ดังนั้น แอททริบิวต์ CousineN ให้เป็น a7 จะได้ คือ

range =(3-0)/2 =1.50

base =(3+0)/2=1.50

norm_a7 =(0-1.50)/1.50 = -1.00

ดังนั้น แอททริบิวต์ statusHist ให้เป็น a8 จะได้ คือ

range =(2-0)/2 =1.00

base =(2+0)/2 =1.00

norm_a8 =(1-1.00)/1.00 = 0.00

จะคำนวณค่าแอททริบิวต์ต่าง ๆ ที่แปลงแล้วมาใส่ในสมการเพื่อส่งไปยัง active function ซึ่งเป็นโหนดใน Hidden layer มีรายละเอียดดังนี้ คือ

Sigmoid node1 =x_1 = 1.00x(1.108)+1.00x(-4.994)-1.00x(3.806)-0.3x(-2.372)+1.00x(-6.841)-1.00x(6.783)-1.00x(3.972)+0.00x(-0.762)-6.532

= -31.108

{f(x)_1) = \frac{1}{1+e^{31.108}} =3.09e-14

Sigmoid node2=x_2=1.00x(0.381)+1.00x(-0.847)-1.00x(-0.314)-0.3x(6.976)+1.00x(-0.664)-1.00x(8.669)-1.00x(-5.261)+0.00x(0.561)-6.050

= -13.97

{f(x)_2) = \frac{1}{1+e^{13.97}} = 8.57e-7

Sigmoid node3=x_3 =1.00x(6.511)+1.00x(2.169)-1.00x(-1.842) 0.3x(12.906)+1.00x(4.639)-1.00x(9.542)-1.00x(3.651)+0.00x(-1.669)-3.864

= -5.77

{f(x)_3) = \frac{1}{1+e^{5.77}} = 0.003

Sigmoid node1 =x_4 = 1.00x(-8.500)+1.00x(1.670)-1.00x(-0.077)-0.3x(8.434)+1.00x(4.078)-
1.00x(-2.920)-1.00x(-9.078)+0.00x(10.417)-4.413
= -20.392

{f(x}_4) = \frac{1}{1+e^{20.392}} =1.39e-9

Sigmoid node1 =x_5 = 1.00x(0.254)+1.00x(6.113)-1.00x(7.691)-0.3x(2.353)+1.00x(4.184)-
1.00x(-11.275)-1.00x(5.077)+0.00x(6.144)+5.753
= 14.105

{f(x}_5) = \frac{1}{1+e^{14.105}} =7.49e-7

หลังจากนั้นจะเข้าสู่โหนดการคำนวณเพื่อนำเสนอผลลัพธ์ ภายใต้การแสดงผลของโหนด 6 โหนด (กรณีเปลี่ยนข้อมูลเชิงปริมาณเป็นข้อมูลเชิงคุณภาพ)การารกำหนดซึ่งมีรายละเอียดนั่นคือ

Sigmoid node Low=x_{Low} = (3.09e-14*(-2.690))+(8.57e-7*5.552)+(0.003*(-3.705))+
(1.39e-9*(-3.638))+(7.49e-7*3.082) 0.451
= -0.462

{f(x}_{Low}) = \frac{1}{1+{e\ }^{0.462}} = 0.387

Sigmoid node Medium =x_{Medium} = (3.09e-14*(-0.688))+(8.57e-7*(-0.614))+
(0.003*(-0.236))+(1.39e-9*(0.330))+(7.49e-7*(-0.868))-1.232
= -1.233

{f(x}_{Medium}) = \frac{1\ }{1+e^{-Medium}} =\frac{1}{1+e^{1.233}} =0.226

Sigmoid node High =x_{High} = (3.09e-14*(-1.209))+(8.57e-7*(-12.160))+
(0.003*(-4.761))+(1.39e-9*(3.916))+(7.49e-7*(-4.417))-2.966
= -2.980

{f(x}_{High}) = \frac{1\ }{1+e^{-High}} =\frac{1}{1+e^{2.980}} = 0.048

Sigmoid node veryLow =x_{veryLow} = (3.09e-14*1.436)+(8.57e-7*(-2.927))+
(0.003*2.875)+(1.39e-9*1.326)+(7.49e-7*(-0.870))-3.151
= -3.142

{f(x}_{veryLow}) = \frac{1\ }{1+e^{-veryLow}} =\frac{1}{1+e^{3.142}} = 0.041

Sigmoid node None =x_{None} = (3.09e-14*(1.397))+(8.57e-7*(-0.673))+
(0.003*(1.583))+(1.39e-9*(0.535))+(7.49e-7*(0.532))-2.777
= -2.772

{f(x}_{None}) = \frac{1\ }{1+e^{-None}} =\frac{1}{1+e^{2.772}} = 0.059

Sigmoid node veryHigh =x_{veryHigh} = (3.09e-14*(-1.161))+(8.57e-7*(-0.192))+
(0.003*(-0.093))+(1.39e-9*(0.732))+(7.49e-7*(-0.045))-2.414

= -2.414

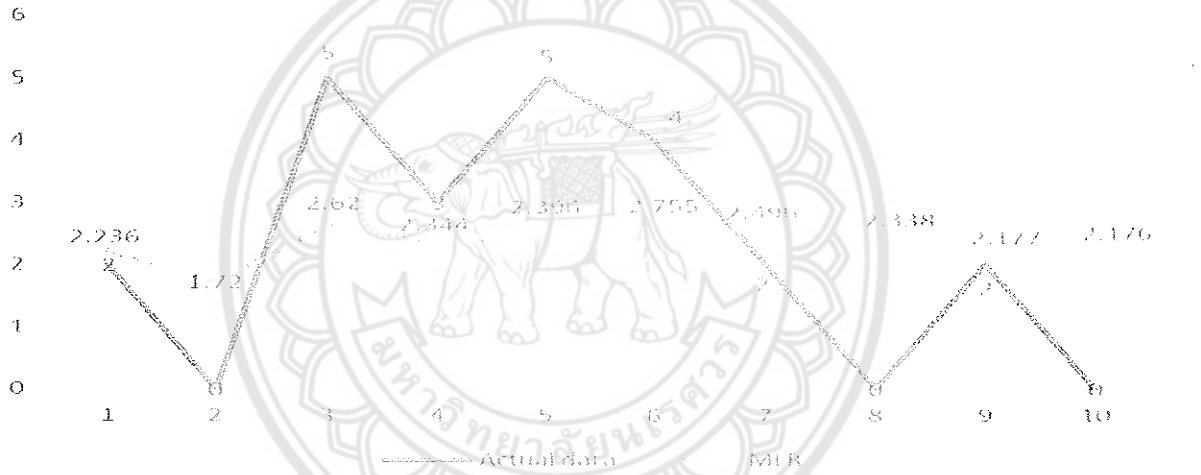
{f(x)_{veryHigh}) = \frac{1\ }{1+e^{-veryHigh}} =\frac{1}{1+e^{2.414}} = 0.082

ดังนั้น จากค่าที่คำนวณสามารถสรุปได้ว่า ข้อมูลตัวอย่างที่นำเข้าสู่ระบบพยากรณ์ด้วยการใช้เทคนิคฟัซซีโครงข่ายประสาทเทียม (Fuzzy neural network) จะพบว่าค่าของแอททริบิวต์ RiskForPay เข้าใกล้ Sigmoid node Low=x_{Low} ซึ่ง กำหนดให้เป็นระดับ "Low" ดังนั้นผลการศึกษาเราจะพบว่า เมื่อทำการเปรียบเทียบการทำงานระหว่างการ FNN กับ MLR นั้น จะพบว่า ทั้งสองวิธีใช้หลักการเดียวกันในการแปลงฟัซซีแอททริบิวต์เพื่อใช้เป็นข้อมูลนำเข้าสำหรับทดสอบ แต่เมื่อดำเนินการทดสอบแล้วจะพบว่าการใช้วิธีของตัวแบบจำลอง FNN จะให้ความแม่นยำมากกว่า โดยสามารถดูการเปรียบเทียบผลที่ได้ดังตารางที่ 4.4 และภาพที่ 4.1 4.2 และ 4.3 ตามลำดับ อย่างไรก็ตามการทำงานที่ได้จากแบบจำลอง FNN นั้นมีความจำเป็นที่จะต้องปรับปรุงการใช้ข้อมูลเชิงฟัซซีให้มีความเหมาะสมและมีความหลากหลายของข้อมูลนำเข้าเพื่อที่จะเพิ่มประสิทธิภาพของความแม่นยำในการพยากรณ์ให้สูงยิ่งขึ้นต่อไป

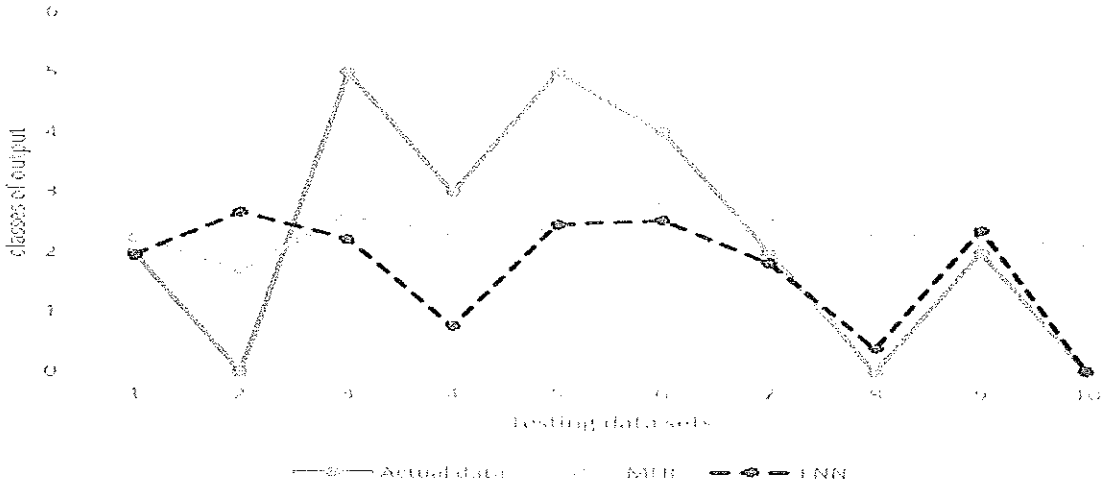
ตารางที่ 4.4 ผลลัพธ์การพยากรณ์ด้วย FNN และ MLR

The FNN model Generated data		The MLR Generated data		The actual data	
Low	1.950	Low	2.236	Low	2
Medium	2.672	Low	1.720	None	0
Low	2.221	Medium	2.620	veryHigh	5
veryLow	0.764	Low	2.344	Medium	3
Medium	2.463	Low	2.396	veryHigh	5
Medium	2.535	Medium	2.755	High	4
Low	1.826	Medium	2.496	Low	2
None	0.399	Low	2.338	None	0
Low	2.378	Low	2.177	Low	2
None	0.022	Low	2.176	None	0

ภาพที่ 4.1 แสดงรายละเอียดการเปรียบเทียบการพยากรณ์ด้วย FNN กับข้อมูลจริงนำเข้า



ภาพที่ 4.2 แสดงรายละเอียดการเปรียบเทียบการพยากรณ์ด้วย MLR กับข้อมูลจริงนำเข้า



ภาพที่ 4.3 ผลการพยากรณ์เปรียบเทียบระหว่างฟัซซีโครงข่ายประสามเทียมกับสมการเชิงเส้นถดถอยเชิงพหุ

สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอการเปรียบเทียบการสร้างแบบจำลองในรูปแบบของ FNN และ MLR ซึ่งวิธีที่ที่นำเสนอนี้เป็นการนำเสนอเริ่มตั้งแต่การจัดการข้อมูลเชิงฟัซซีแอททริบิวต์ รวมทั้งการคัดกรองข้อมูลนำเข้าเพื่อสร้างแบบจำลอง จากแนวทางที่ดำเนินการนั้น เราจะพบว่าแบบจำลองในลักษณะ FNN ที่มีโครงสร้างมีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล มีค่าเท่ากับ 83.33%, 83.10%, 83.30% และ 82.40% ตามลำดับ นั้นมีประสิทธิภาพการทำงานที่สูงกว่าแบบจำลอง MLR ที่มีค่า Performance vector นั่นคือ ค่า root mean squared error: 1.513 +/- 0.000 และค่า squared correlation: 0.081 ตามลำดับ อย่างไรก็ตาม รูปแบบของตัวแบบพยากรณ์ที่ทำงานนั้น เราจะพบว่าการกำหนดค่าระดับของฟัซซีแอททริบิวต์ที่ทำงานภายใต้ Income ซึ่งเป็นตัวเดียวที่กำหนดไว้ในเชิงฟัซซี เป็นการยากที่จะกำหนดให้เกิดความแม่นยำได้ โดยในงานวิจัยนี้อาศัยเพียงพื้นฐานของการกำหนดค่าเฉลี่ยแล้วดำเนินการแบ่งเป็นช่วงระยะซึ่งอาจจะมีผลกระทบต่อความผิดพลาดของการคำนวณได้ จึงจำเป็นอย่างยิ่งที่จะต้องดำเนินการหาวิธีการที่จะกำหนดระดับค่าเชิงฟัซซีดังกล่าวให้มีความน่าเชื่อถือมากยิ่ง รวมทั้งการกำหนดปัจจัยนำเข้าเพื่อใช้สำหรับฝึกสอนข้อมูล และการรวบรวมข้อมูลให้มากยิ่งขึ้นถึงแม้การดำเนินการจัดเก็บข้อมูลเป็นเรื่องยุ่งยากมากในการดำเนินวิจัยที่จะต้องเกี่ยวข้องกับบุคคลากรหลายๆ ฝ่ายทั้งในภาครัฐที่เป็นหน่วยงานของมหาวิทยาลัย หน่วยงานที่รับผิดชอบกองทุน กยศ รวมทั้งภาคเอกชนที่เกี่ยวข้องกับธนาคารต่าง ๆ ทั้งนี้เพื่อให้การวิจัยที่สามารถดำเนินการได้อย่างมีประสิทธิภาพที่สามารถยอมรับและนำไปประยุกต์ใช้งานได้ตามวัตถุประสงค์ของหน่วยงานต่าง ๆ เป็นลำดับต่อไป

บรรณานุกรม

[1] ชนวัฒน์ ศรีสอ้าน, ฐานข้อมูล คลังข้อมูลและเหมืองข้อมูล

[2] ถิรยุทธ ลิมานนท์, อาทิตย์ ศรีแก้ว และคณะ, การเปรียบเทียบประสิทธิภาพแบบจำลองการครอบครองยานพาหนะของ ครัวเรือนในประเทศไทย: โดยวิธีวิเคราะห์สมการถดถอยเชิงเส้นแบบพหุกับวิธีเครือข่ายประสาทเทียมการเรียนรู้แบบแพร่ย้อนกลับ, 3rd A Trans Symposium Student Chapter Session, August, 27,2010 Bangkok Thailand.

[3] ธนาวุฒิ ประกอบผล,โครงข่ายประสาทเทียม,วารสาร มฉก. วิชาการ ปีที่12 ฉบับที่24 มกราคม-มิถุนายน 2552

[4] นพพร ธนะชัยขันธ์, สถิติเบื้องต้นสำหรับการทำวิจัยฉบับเสริมการวิเคราะห์ข้อมูลด้วยโปรแกรม Excel, บริษัทวิทยพัฒน์ จำกัด, มีนาคม 2555

[5] วิรัชช พานิชวงศ์, การวิเคราะห์การถดถอย, ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2549

[6] เอกสิทธิ์ พัชรวงศ์ศักดาIntroduction to Data Mining with RapidMiner Studio 7แหล่งข้อมูลจาก http://dataminingtrend.com/2014/wp-content/uploads/2014/02/RM7_chapter1.pdf

[7] สืบค้นประจำวันที่ 20 กรกฎาคม พ.ศ. 2558 ที่มา: ASTVผู้จัดการรายวัน ประจำวันที่ 4 กันยายน 2556 คอลัมน์ "เด็กกู้ต้องอ่าน! มหากาพย์ชักดาบเงินกู้กยศ. หมดเวลาสนุกแล้วสิ"

[8] สืบค้นประจำวันที่ 5 กรกฎาคม 2558 ที่มา: ประชาชาติธุรกิจออนไลน์ ประจำวันที่ 27 ธันวาคม 2557

[9] สืบค้นประจำวันที่ 10 สิงหาคม 2558 ที่มา: ไทยรัฐออนไลน์ ประจำวันที่ 5 ก.พ. 2558

[10] ข่าวประชาสัมพันธ์กองทุนเงินให้กู้ยืมเพื่อการศึกษา (กยศ.), "หลักเกณฑ์การคัดกรองผู้กู้ยืมและสถานศึกษา เพื่อ เตรียมพร้อมก่อนใช้จริงปีการศึกษา 2559", ประจำวันที่ 10 กรกฎาคม 2558

[11] สืบค้นประจำวันที่ 12 กรกฎาคม พ.ศ. 2558 ที่มา: www.studentloan.or.th

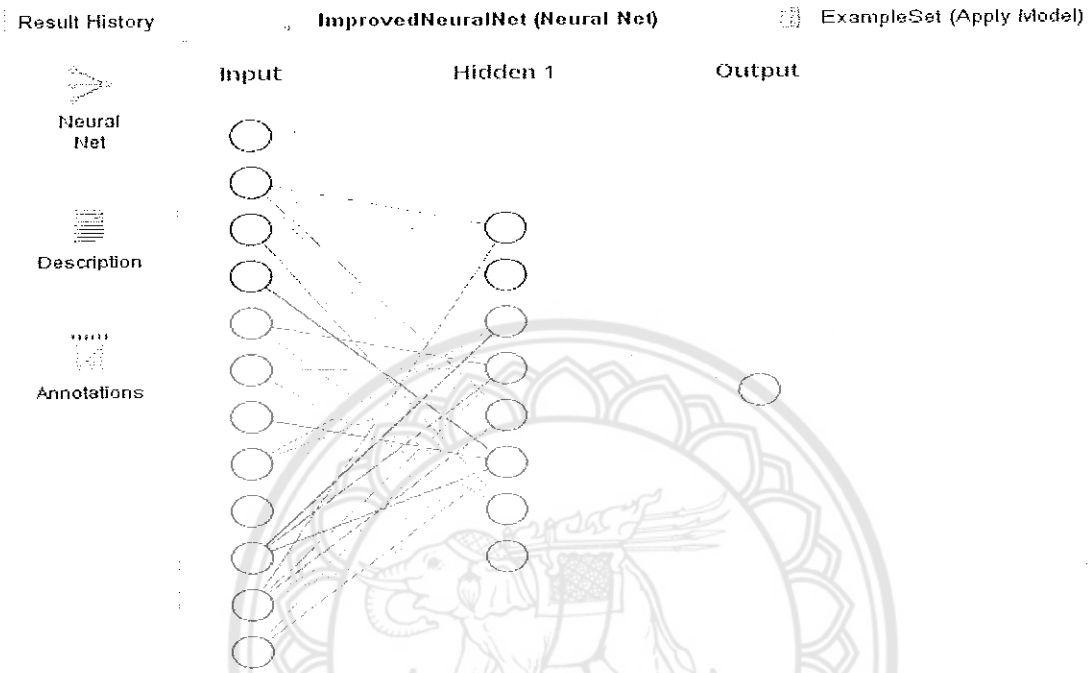
[12] Kim S. C., Herlina A. R. and Ruzairi A. R., A Comparison of Principal Component Regression and Artificial Neural Network in Fruits Quality Prediction, 7th International Colloquium on Signal Processing and its Applications,2011 IEEE.

[13] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Addison-Wesley, 2006.

[14] Chittayasothorn S., Toward Fuzzy Temporal Databases with Temporal Fuzzy Linguistic Terms, The Second International Conference on the Applications of Digital Information and Web Technologies (ICADIWT2009) August 4-6,2009 IEEE.

[15] Riordan D., Hansen B. K., A Fuzzy Case-Based System for Weather Prediction, Engineering Intelligent Systems, September, 2002.

ก-1 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=5 รูปแบบ 11-7-1 และผลลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์



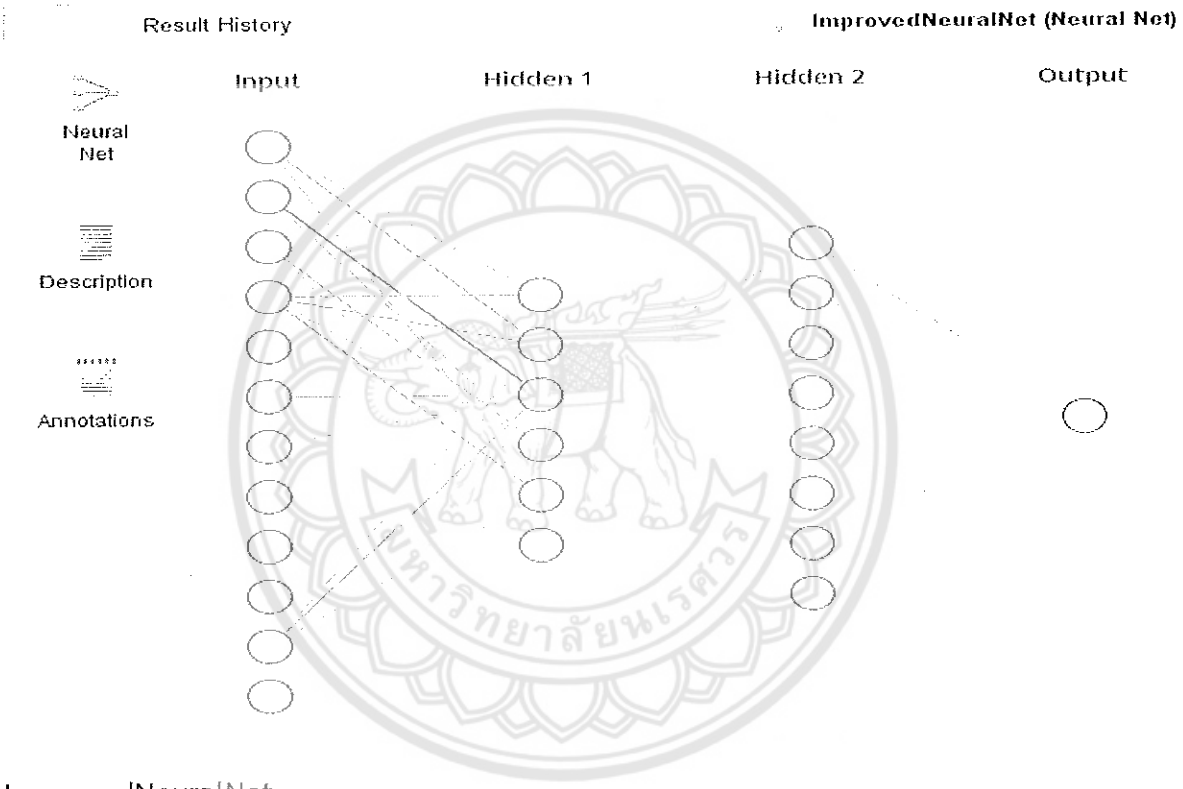
ImprovedNeuralNet

```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: 0.116
engi: 2.007
aggri: 0.998
GPAX: 0.633
Sex: 0.564
High: 0.443
medium: -0.453
Poor: 0.964
VeryPoor: 0.394
CousineN: 0.361
StatusHist: 2.282
Bias: -1.017

Node 2 (Sigmoid)
----------------
sci: -0.706
engi: 0.828
aggri: 0.415
GPAX: 0.278
```

```
Sex: -0.360
High: 0.358
medium: -0.237
Poor: 0.809
VeryPoor: 0.599
CousineN: 0.414
StatusHist: -0.094
Bias: -0.720

Node 3 (Sigmoid)
----------------
sci: -0.642
engi: 0.168
aggri: 0.160
GPAX: 0.500
Sex: -0.172
High: 0.329
medium: -0.436
Poor: 1.899
VeryPoor: -0.800
CousineN: 3.395
StatusHist: 0.490
Bias: 0.284

Node 4 (Sigmoid)
----------------
sci: -0.934
engi: 0.941
aggri: -0.635
GPAX: 1.121
Sex: -2.605
High: -1.463
medium: 1.330
Poor: 1.933
VeryPoor: -0.569
CousineN: 2.824
StatusHist: -1.941
Bias: -1.410

Node 5 (Sigmoid)
----------------
sci: 0.247
engi: 1.857
aggri: 0.791
GPAX: -0.291
Sex: -0.083
High: 0.453
medium: -0.639
Poor: 1.257
VeryPoor: 0.376
CousineN: 0.340
StatusHist: 2.446
Bias: -0.946


.. .

Output
======

Regression (Linear)
```

```
--------------------
Node 1: -1.241
Node 2: 0.679
Node 3: 1.552
Node 4: -0.992
Node 5: -1.279
Node 6: 1.192
Node 7: -1.038
Threshold: -0.172
```

ก-2 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=5 รูปแบบ 11-5-7-1 และผลลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์



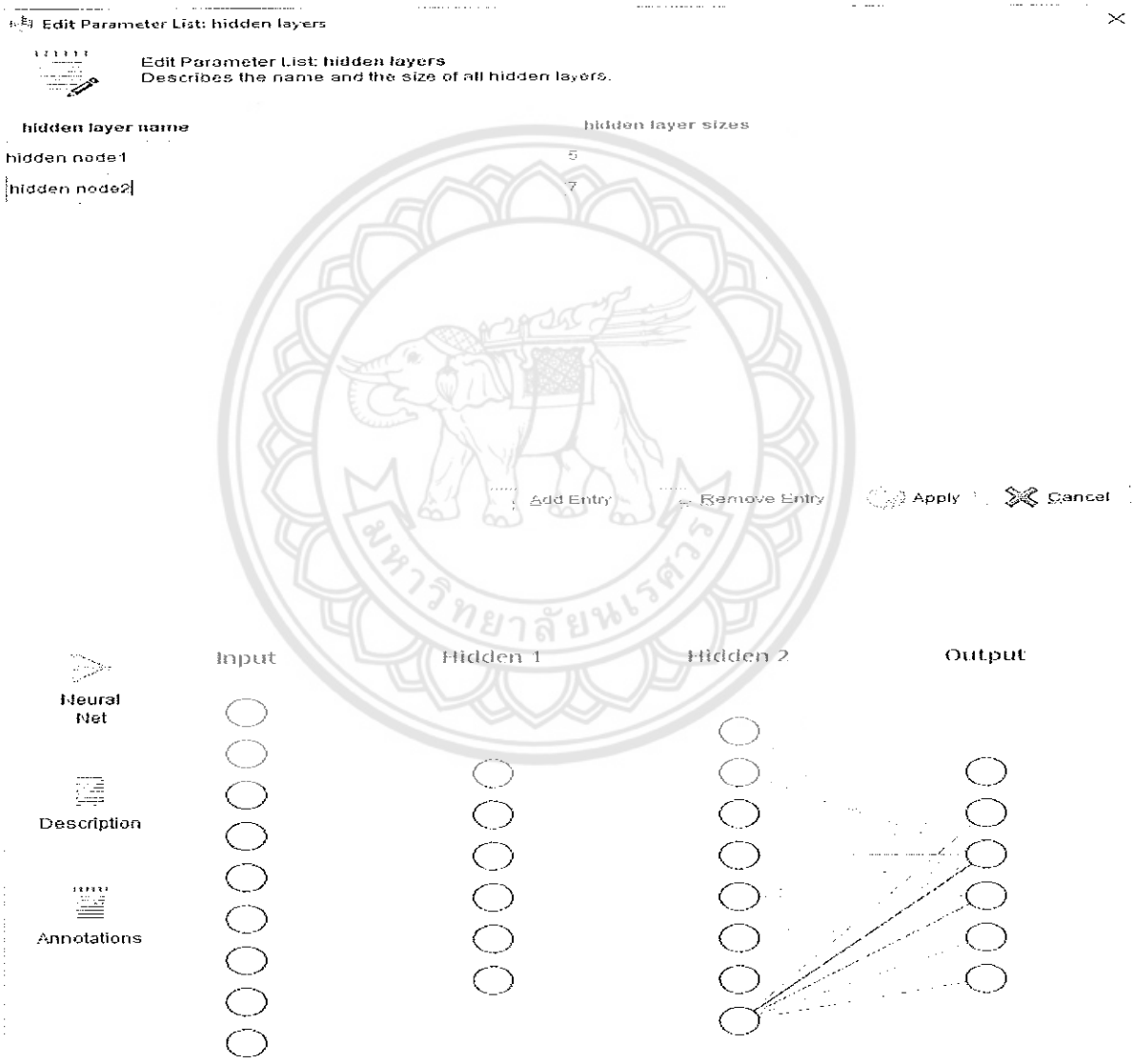
ImprovedNeuralNet

```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: -1.114
engi: -0.322
aggri: -0.734
GPAX: -1.501
Sex: -0.311
High: 0.136
medium: -0.216
Poor: 0.173
VeryPoor: 0.162
CousineN: -0.117
StatusHist: 0.710
Bias: -0.554
```

```
Node 2 (Sigmoid)
----------------
sci: -1.628
engi: 0.052
aggri: -0.543
GPAX: -1.712
Sex: -0.278
High: -0.211
medium: -0.511
Poor: -0.250
VeryPoor: 0.383
CousineN: -0.516
StatusHist: 1.121
Bias: -0.321

Node 3 (Sigmoid)
----------------
sci: -0.625
engi: -2.470
aggri: -0.351
GPAX: -0.794
Sex: 0.004
High: -1.057
medium: 1.264
Poor: -0.436
VeryPoor: -0.207
CousineN: -0.231
StatusHist: -1.626
Bias: 0.280

.. .

Hidden 2
========

Node 1 (Sigmoid)
----------------
Node 1: -0.493
Node 2: -0.626
Node 3: -1.089
Node 4: -0.674
Node 5: -0.552
Bias: -0.542

Node 2 (Sigmoid)
----------------
Node 1: -0.517
Node 2: -0.665
Node 3: -0.994
Node 4: -0.638
Node 5: -0.538
Bias: -0.644

.. .

Output
======

Regression (Linear)
-------------------
```

```
Node 1: -1.138
Node 2: -1.085
Node 3: -1.105
Node 4: -1.091
Node 5: -1.124
Node 6: -1.114
Node 7: -1.135
Threshold: 0.387
```

ก-3 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=5 รูปแบบ 8-5-7-6 และผลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์



ImprovedNeuralNet

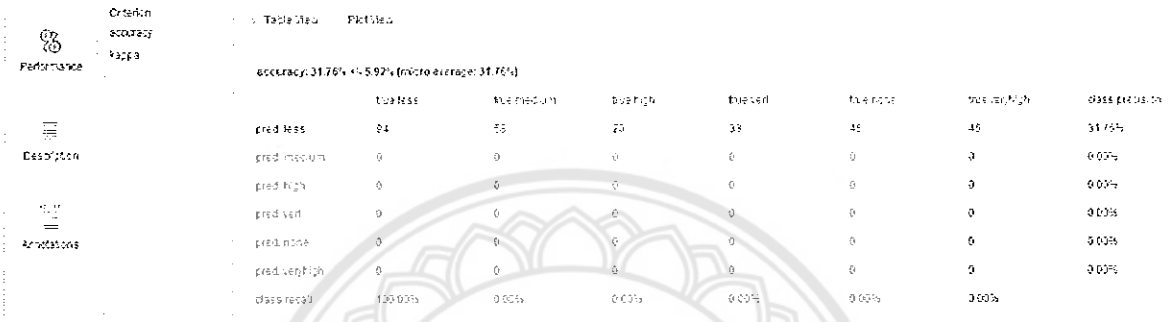
```
Hidden 1
========

Node 1 (Sigmoid)
```

```
----------------
sci: 0.006
engi: 0.002
aggri: 0.027
GPAX: 0.007
Sex: 0.004
Income: -0.004
CousineN: -0.017
StatusHist: 0.003
Bias: 0.077

Node 2 (Sigmoid)
----------------
sci: 0.008
engi: 0.029
aggri: 0.031
GPAX: 0.054
Sex: -0.018
Income: 0.034
CousineN: -0.022
StatusHist: -0.035
Bias: 0.024

Node 3 (Sigmoid)
----------------
sci: -0.042
engi: 0.021
aggri: 0.051
GPAX: 0.013
Sex: 0.046
Income: 0.062
CousineN: -0.073
StatusHist: -0.011
Bias: 0.057

Node 4 (Sigmoid)
----------------
sci: 0.015
engi: 0.027
aggri: -0.033
GPAX: -0.014
Sex: -0.021
Income: 0.028
CousineN: 0.016
StatusHist: -0.005
Bias: 0.014

Node 5 (Sigmoid)
----------------
sci: -0.014
engi: 0.034
aggri: -0.018
GPAX: 0.019
Sex: 0.004
Income: -0.008
CousineN: 0.010
StatusHist: -0.065
Bias: 0.007


Hidden 2
```

```
========

Node 1 (Sigmoid)
----------------
Node 1: -0.208
Node 2: -0.192
Node 3: -0.201
Node 4: -0.136
Node 5: -0.178
Bias: -0.382

Node 2 (Sigmoid)
----------------
Node 1: -0.183
Node 2: -0.156
Node 3: -0.160
Node 4: -0.178
Node 5: -0.156
Bias: -0.379

Node 3 (Sigmoid)
----------------
Node 1: -0.181
Node 2: -0.187
Node 3: -0.235
Node 4: -0.136
Node 5: -0.168
Bias: -0.372

Node 4 (Sigmoid)
----------------
Node 1: -0.166
Node 2: -0.156
Node 3: -0.171
Node 4: -0.194
Node 5: -0.215
Bias: -0.382

Node 5 (Sigmoid)
----------------
Node 1: -0.216
Node 2: -0.163
Node 3: -0.212
Node 4: -0.183
Node 5: -0.175
Bias: -0.310

Node 6 (Sigmoid)
----------------
Node 1: -0.135
Node 2: -0.164
Node 3: -0.226
Node 4: -0.166
Node 5: -0.140
Bias: -0.371

Node 7 (Sigmoid)
----------------
Node 1: -0.197
Node 2: -0.180
Node 3: -0.214
```

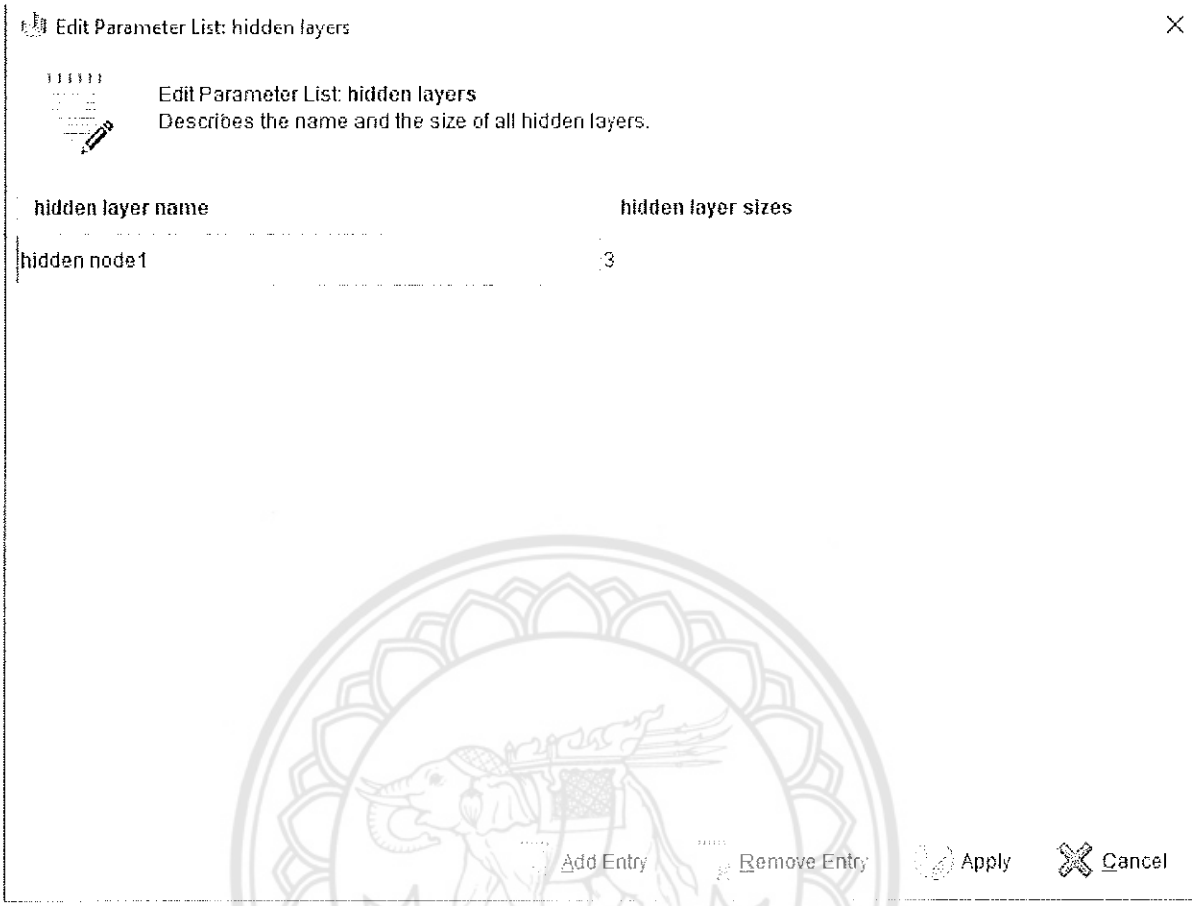
```
Node 4: -0.128
Node 5: -0.197
Bias: -0.377


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: -0.198
Node 2: -0.176
Node 3: -0.193
Node 4: -0.237
Node 5: -0.252
Node 6: -0.236
Node 7: -0.175
Threshold: -0.389

Class 'medium' (Sigmoid)
------------------------
Node 1: -0.361
Node 2: -0.325
Node 3: -0.302
Node 4: -0.360
Node 5: -0.365
Node 6: -0.356
Node 7: -0.340
Threshold: -0.713

Class 'high' (Sigmoid)
----------------------
Node 1: -0.579
Node 2: -0.669
Node 3: -0.601
Node 4: -0.603
Node 5: -0.664
Node 6: -0.622
Node 7: -0.615
Threshold: -1.275

Class 'verl' (Sigmoid)
----------------------
Node 1: -0.513
Node 2: -0.467
Node 3: -0.518
Node 4: -0.470
Node 5: -0.477
Node 6: -0.537
Node 7: -0.469
Threshold: -1.024

Class 'none' (Sigmoid)
----------------------
Node 1: -0.387
Node 2: -0.387
Node 3: -0.441
Node 4: -0.398
Node 5: -0.431
Node 6: -0.427
Node 7: -0.440
```

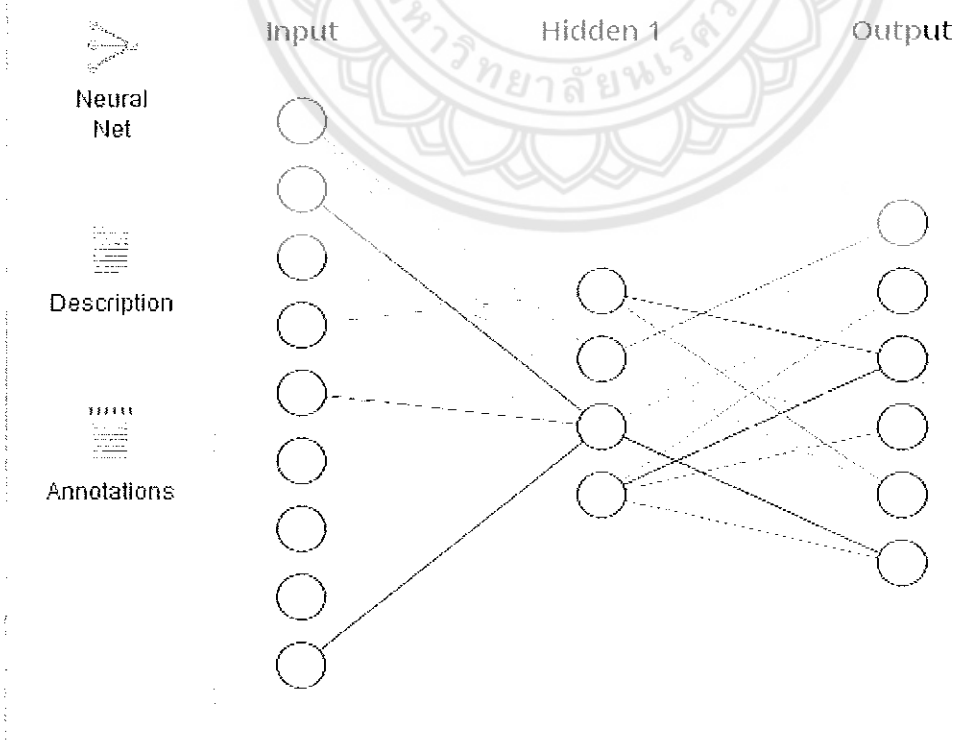
```
Threshold: -0.812

Class 'veryhigh' (Sigmoid)
------------------------------
Node 1: -0.396
Node 2: -0.382
Node 3: -0.360
Node 4: -0.364
Node 5: -0.402
Node 6: -0.430
Node 7: -0.365
Threshold: -0.857
```



PerformanceVector

```
PerformanceVector:
accuracy: 31.76% +/- 5.92% (micro average: 31.76%)
ConfusionMatrix:
True:   less    medium  high    verl    none    veryhigh
less:   94      53      20      33      45      46
medium: 0       0       0       0       0       0
high:   0       0       0       0       0       0
verl:   0       0       0       0       0       0
none:   0       0       0       0       0       0
veryhigh:       0       0       0       0       0       0
kappa: 0.000 +/- 0.000 (micro average: 0.000)
ConfusionMatrix:
True:   less    medium  high    verl    none    veryhigh
less:   94      53      20      33      45      46
medium: 0       0       0       0       0       0
high:   0       0       0       0       0       0
verl:   0       0       0       0       0       0
none:   0       0       0       0       0       0
veryhigh:       0       0       0       0       0       0
```


Edit Parameter List: hidden layers
Edit Parameter List: hidden layers
Describes the name and the size of all hidden layers.
hidden layer name
hidden layer sizes
hidden node1
3
Add Entry
Remove Entry
Apply
Cancel


Input
Hidden 1
Output
Neural Net
Description
Annotations

ImprovedNeuralNet

```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: 0.373
engi: -0.638
aggri: 0.439
GPAX: -1.007
Sex: 0.479
Income: 0.015
CousineN: 0.485
StatusHist: -0.074
Bias: 0.096

Node 2 (Sigmoid)
----------------
sci: -0.871
engi: -0.299
aggri: -0.862
GPAX: -0.297
Sex: -0.288
Income: 0.213
CousineN: -0.007
StatusHist: 0.203
Bias: -0.421

Node 3 (Sigmoid)
----------------
sci: 0.766
engi: 1.525
aggri: -0.171
GPAX: 0.771
Sex: -1.404
Income: -0.061
CousineN: -0.657
StatusHist: -0.047
Bias: -1.601


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: 0.115
Node 2: -1.329
Node 3: 0.729
Threshold: -0.370

Class 'medium' (Sigmoid)
------------------------
Node 1: -0.890
Node 2: -0.306
Node 3: 0.955
Threshold: -1.010

Class 'high' (Sigmoid)
```

```
-----------------------
Node 1: -1.592
Node 2: 0.050
Node 3: -0.404
Threshold: -1.802

Class 'verl' (Sigmoid)
-----------------------
Node 1: -0.350
Node 2: -0.889
Node 3: -0.781
Threshold: -1.270

Class 'none' (Sigmoid)
----------------------
Node 1: -1.155
Node 2: -0.935
Node 3: -0.822
Threshold: -0.583

Class 'veryhigh' (Sigmoid)
--------------------------
Node 1: -0.022
Node 2: 0.084
Node 3: -1.806
Threshold: -1.308
```

Performance

Description

Annotations

PerformanceVector

PerformanceVector:
accuracy: 20.26% +/- 3.93% (micro average: 20.27%)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh
less:	39	32	12	16	24	30
medium:	40	18	6	13	19	11
high:	0	0	0	0	0	0
verl:	0	0	0	0	0	0
none:	2	2	0	2	0	2
veryhigh:	13	6	2	2	2	3

kappa: -0.050 +/- 0.042 (micro average: -0.068)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh
less:	39	32	12	16	24	30
medium:	40	18	6	13	19	11
high:	0	0	0	0	0	0
verl:	0	0	0	0	0	0
none:	2	2	0	2	0	2
veryhigh:	13	6	2	2	2	3

Performance
Criterion
accuracy
kappa

Table View Plot View

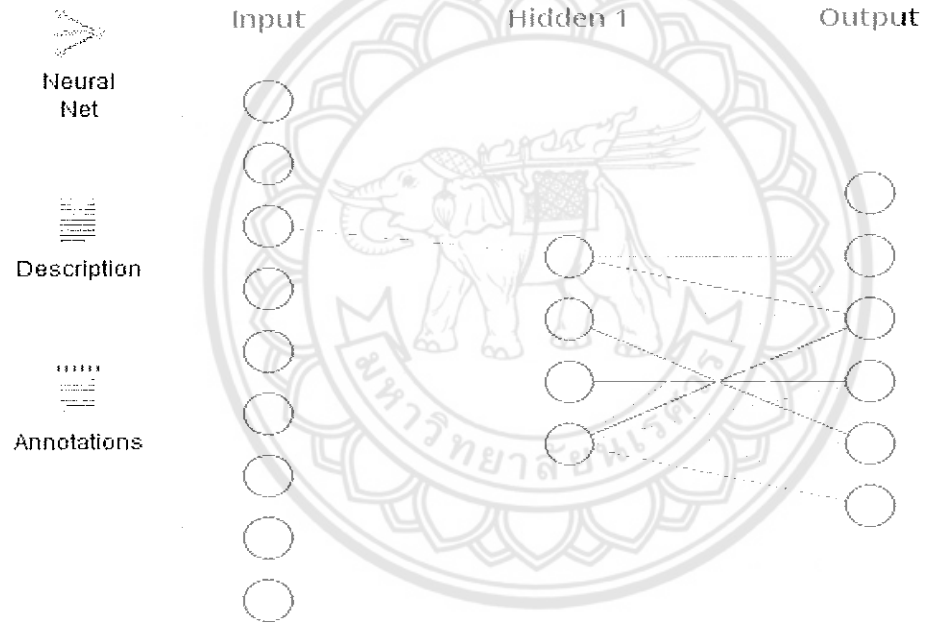
accuracy: 20.26% +/- 3.83% (micro average: 20.27%)

	true less	true medium	true high	true very	true none	true very high	class precision
pred. less	39	32	12	18	24	33	25.48%
pred. medium	43	18	6	13	19	11	15.62%
pred. high	0	0	0	0	0	0	0.00%
pred. very	0	0	0	0	0	0	0.00%
pred. none	2	2	0	2	0	2	0.00%
pred. very high	13	6	2	2	2	3	10.71%
class recall	41.49%	31.03%	0.00%	0.00%	0.00%	6.52%	

Description

Annotations

ก-4 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=10 รูปแบบ 8-3-6 และผลลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์



Performance
Criterion
accuracy
kappa

Table View Plot View

accuracy: 29.41% +/- 9.09% (micro average: 29.35%)

	true less	true medium	true high	true very	true none	true very high	class precision
pred. less	81	51	17	30	42	45	39.23%
pred. medium	13	6	5	2	2	0	26.09%
pred. high	0	0	0	0	0	0	0.00%
pred. very	0	0	0	0	0	0	0.00%
pred. none	0	0	0	0	0	0	0.00%
pred. very high	3	1	0	1	1	0	0.00%
class recall	83.17%	10.34%	0.00%	0.00%	0.00%	0.00%	

Description

Annotations

Performance

PerformanceVector

Description

Annotations

PerformanceVector:
accuracy: 29.41% +/- 9.00% (micro average: 29.39%)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh	
less:	81	51	17	30	42	46	
medium:	10	6	3	2	2	0	
high:	0	0	0	0	0	0	
verl:	0	0	0	0	0	0	
none:	0	0	0	0	0	0	
veryhigh:	3	1	0	1	1		0

kappa: -0.005 +/- 0.030 (micro average: -0.016)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh	
less:	81	51	17	30	42	46	
medium:	10	6	3	2	2	0	
high:	0	0	0	0	0	0	
verl:	0	0	0	0	0	0	
none:	0	0	0	0	0	0	
veryhigh:	3	1	0	1	1		0

ImprovedNeuralNet

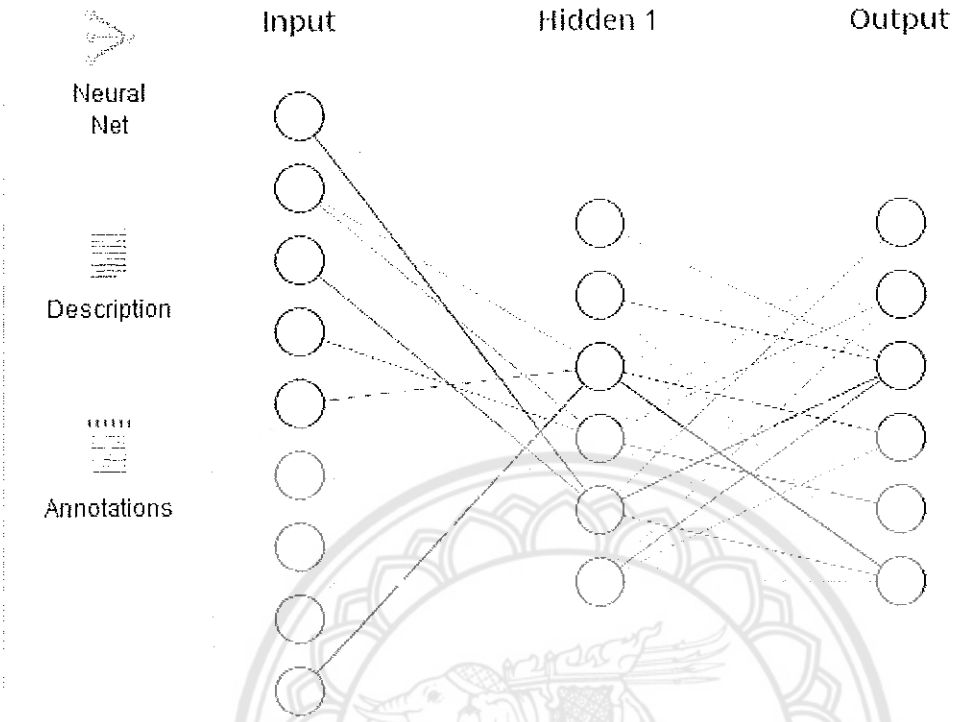
Hidden 1
========

Node 1 (Sigmoid)
----
sci: 0.265
engi: -0.101
aggri: 1.117
GPAX: -0.380
Sex: 0.338
Income: 0.006
CousineN: 0.402
StatusHist: 0.293
Bias: 0.036

Node 2 (Sigmoid)
----
sci: -0.201
engi: -0.612
aggri: -0.629
GPAX: -0.606
Sex: 0.184
Income: 0.092
CousineN: 0.395
StatusHist: -0.272
Bias: -0.582

```
Node 3 (Sigmoid)
----------------
sci: -0.658
engi: 0.437
aggri: -0.180
GPAX: 0.536
Sex: -0.401
Income: -0.140
CousineN: -0.608
StatusHist: 0.539
Bias: -0.347


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: 0.116
Node 2: -0.730
Node 3: -0.689
Threshold: -0.156

Class 'medium' (Sigmoid)
------------------------
Node 1: -1.019
Node 2: -0.647
Node 3: 0.261
Threshold: -0.926

Class 'high' (Sigmoid)
----------------------
Node 1: -1.529
Node 2: -0.601
Node 3: 0.031
Threshold: -1.855

Class 'verl' (Sigmoid)
----------------------
Node 1: -0.733
Node 2: 0.019
Node 3: -1.549
Threshold: -1.090

Class 'none' (Sigmoid)
----------------------
Node 1: 0.028
Node 2: -1.541
Node 3: -0.260
Threshold: -0.887

Class 'veryhigh' (Sigmoid)
--------------------------
Node 1: 0.355
Node 2: -0.352
Node 3: -0.838
Threshold: -1.253
```

ก-5 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=10 รูปแบบ 8-5-6 และผลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์



Performance
Confusion matrix
accuracy
kappa

Descriptors

Annotations

Table View Plot View

accuracy: 27.26% +/- 7.16% (micro average: 27.35%)

	true less	true medium	true high	true very high	true none	true less,high	class precision
pred. less	75	45	15	22	33	42	30.24%
pred. medium	12	5	1	3	0	3	17.65%
pred. high	0	0	0	0	0	0	0.00%
pred. very high	0	1	0	0	0	0	0.00%
pred. none	1	1	0	0	0	0	0.00%
pred. very,high	6	6	1	1	2	1	5.88%
class recall	79.79%	9.62%	0.00%	0.00%	0.00%	2.17%	

Performance

PerformanceVector

Description

Annotations

PerformanceVector:
accuracy: 27.36% +/- 7.16% (micro average: 27.36%)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh	
less:	75	45	18	29	39	42	
medium:	12	5	1	3	4	3	
high:	0	0	0	0	0	0	
verl:	0	1	0	0	0	0	
none:	1	1	0	0	0	0	
veryhigh:	6	6	1	1	2		1

kappa: -0.006 +/- 0.037 (micro average: -0.030)
ConfusionMatrix:

True:	less	medium	high	verl	none	veryhigh	
less:	75	45	18	29	39	42	
medium:	12	5	1	3	4	3	
high:	0	0	0	0	0	0	
verl:	0	1	0	0	0	0	
none:	1	1	0	0	0	0	
veryhigh:	6	6	1	1	2		1

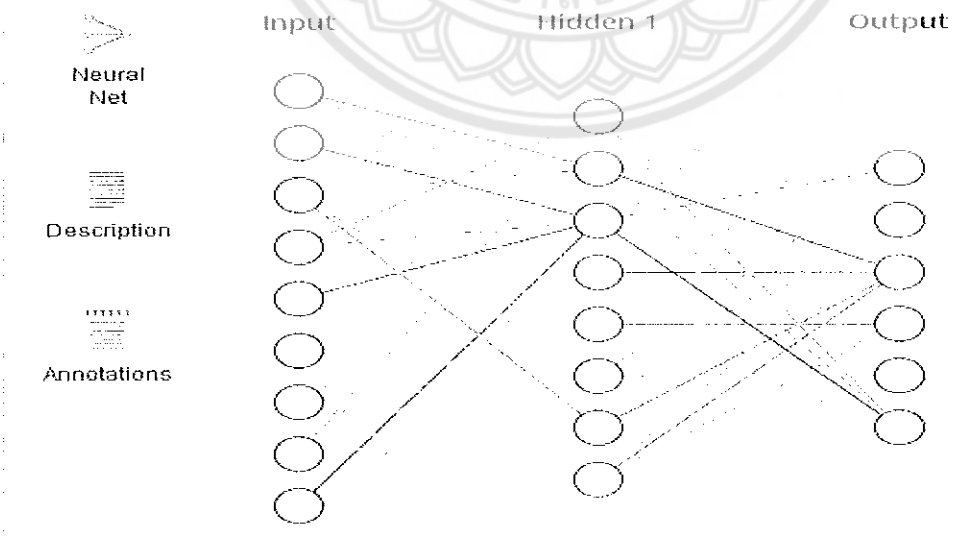
ImprovedNeuralNet

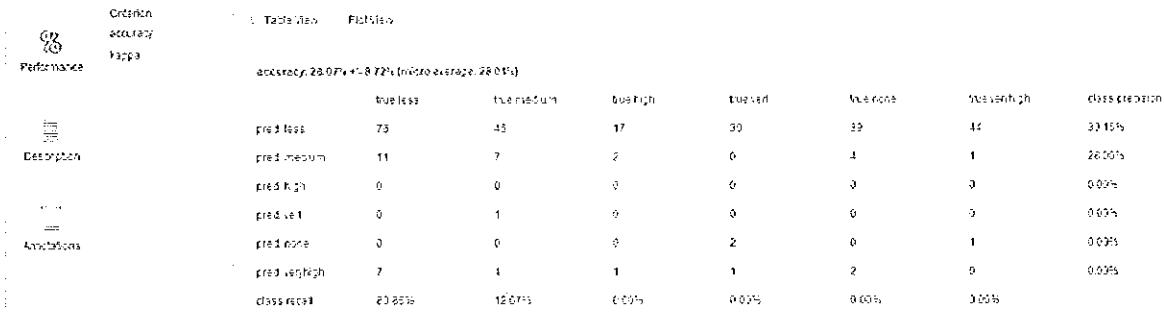
```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: -0.228
engi: 0.324
aggri: 0.644
GPAX: 0.111
Sex: 0.378
Income: 0.175
CousineN: 0.208
StatusHist: 0.473
Bias: -0.108

Node 2 (Sigmoid)
----------------
sci: 0.548
engi: 0.344
aggri: 0.427
GPAX: 0.176
Sex: 0.050
Income: 0.069
CousineN: -0.066
StatusHist: -0.566
Bias: -0.170

Node 3 (Sigmoid)
```

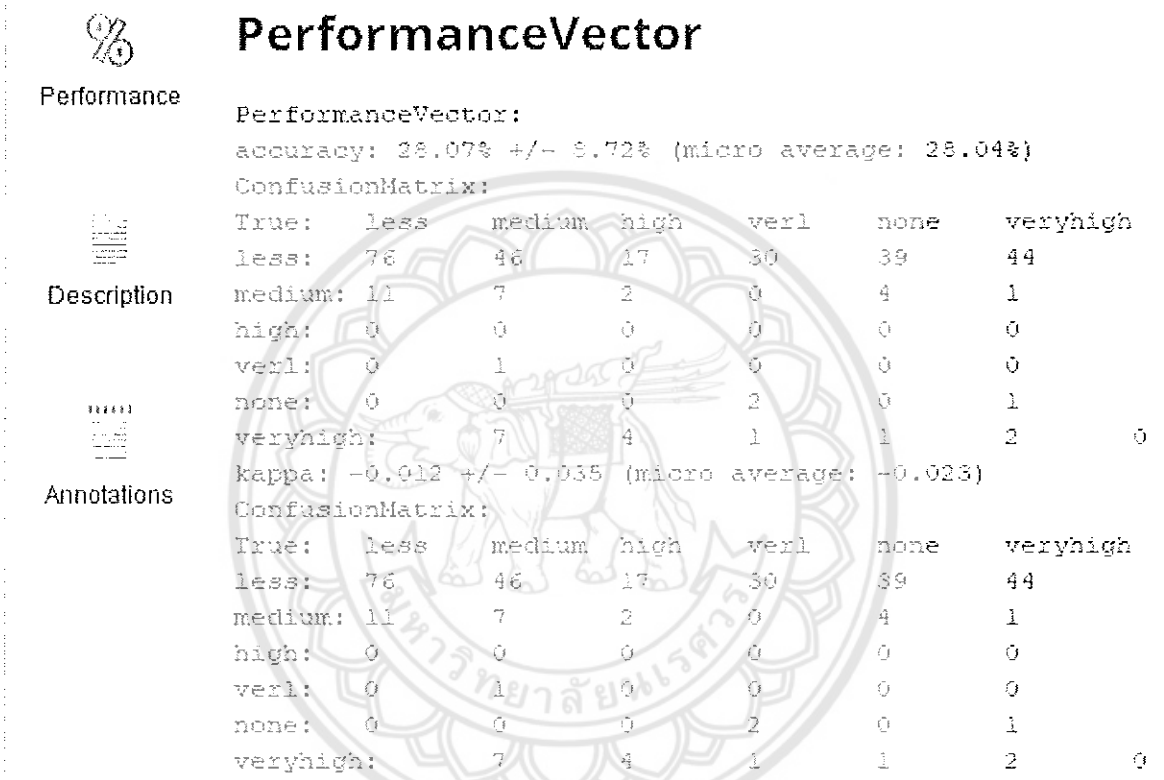
```
-----------------
sci: -0.127
engi: 0.941
aggri: -0.574
GPAX: 0.498
Sex: -1.217
Income: -0.112
CousineN: -0.513
StatusHist: -0.066
Bias: -1.643

Node 4 (Sigmoid)
-----------------
sci: -0.462
engi: -0.973
aggri: -0.476
GPAX: -1.109
Sex: 0.363
Income: -0.056
CousineN: 0.212
StatusHist: 0.234
Bias: -0.169

Node 5 (Sigmoid)
-----------------
sci: 1.372
engi: 0.377
aggri: 1.206
GPAX: 0.593
Sex: 0.365
Income: -0.624
CousineN: -0.080
StatusHist: 0.031
Bias: 0.026


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: -0.088
Node 2: 0.222
Node 3: 0.018
Node 4: -0.772
Node 5: 0.980
Threshold: -0.628

Class 'medium' (Sigmoid)
------------------------
Node 1: -0.738
Node 2: -0.152
Node 3: 0.754
Node 4: -0.971
Node 5: -0.266
Threshold: -0.877

Class 'high' (Sigmoid)
----------------------
Node 1: -0.944
Node 2: -1.183
```

```
Node 3: 0.339
Node 4: -0.141
Node 5: -1.574
Threshold: -1.304

Class 'verl' (Sigmoid)
----------------------
Node 1: -0.674
Node 2: -0.113
Node 3: -1.313
Node 4: -0.509
Node 5: -0.124
Threshold: -1.071

Class 'none' (Sigmoid)
----------------------
Node 1: 0.343
Node 2: -0.373
Node 3: -0.797
Node 4: -1.394
Node 5: 0.248
Threshold: -0.768

Class 'veryhigh' (Sigmoid)
--------------------------
Node 1: -0.114
Node 2: -0.702
Node 3: -1.656
Node 4: 0.645
Node 5: -1.267
Threshold: -0.825
```

ก-6 : แสดงตัวอย่างการพัฒนาตัวแบบจำลองในลักษณะ cross validation fold=10 รูปแบบ 8-7-6 และผลลัพธ์ของค่าการคำนวนของตัวแบบพยากรณ์







PerformanceVector

```
PerformanceVector:
accuracy: 28.07% +/- 8.72% (micro average: 28.04%)
ConfusionMatrix:
True:    less    medium  high    verl    none    veryhigh
less:    76      46      17      30      39      44
medium:  11      7       2       0       4       1
high:    0       0       0       0       0       0
verl:    0       1       0       0       0       0
none:    0       0       0       2       0       1
veryhigh:        7       4       1       1       2       0
kappa: -0.012 +/- 0.035 (micro average: -0.023)
ConfusionMatrix:
True:    less    medium  high    verl    none    veryhigh
less:    76      46      17      30      39      44
medium:  11      7       2       0       4       1
high:    0       0       0       0       0       0
verl:    0       1       0       0       0       0
none:    0       0       0       2       0       1
veryhigh:        7       4       1       1       2       0
```

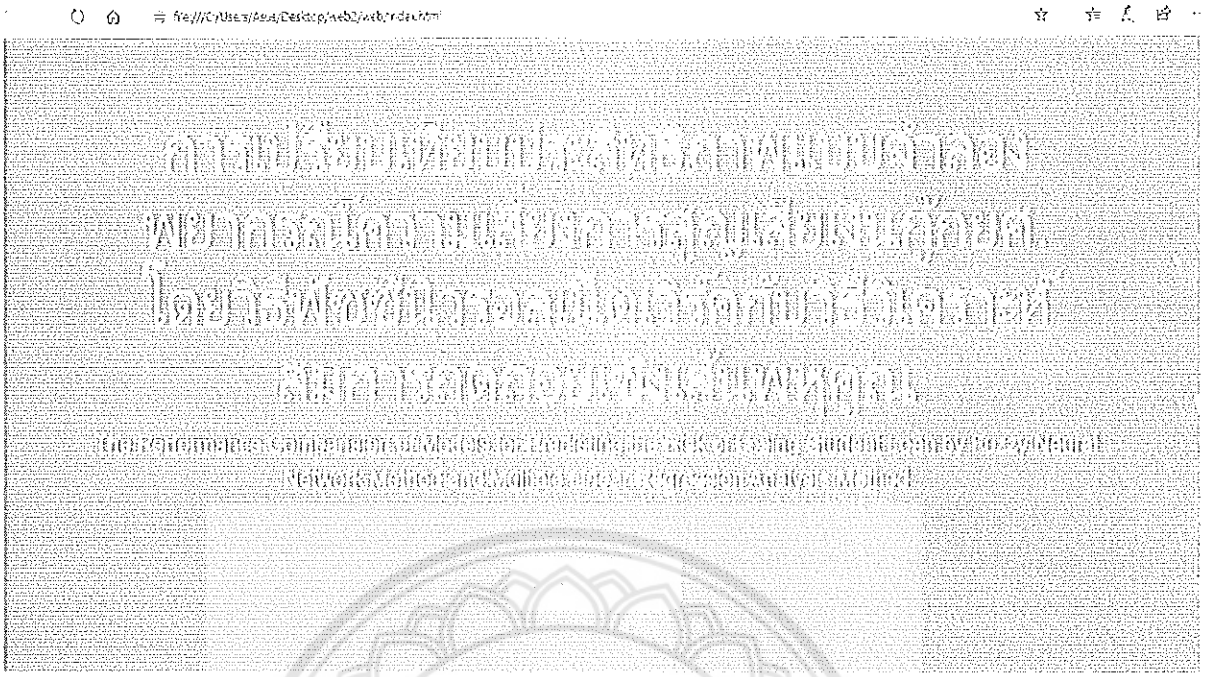
ImprovedNeuralNet

```
Hidden 1
========

Node 1 (Sigmoid)
----------------
sci: -0.362
engi: -0.630
aggri: -0.026
GPAX: -1.084
Sex: -0.039
Income: 0.569
CousineN: 0.405
StatusHist: 0.769
Bias: -0.282

Node 2 (Sigmoid)
```

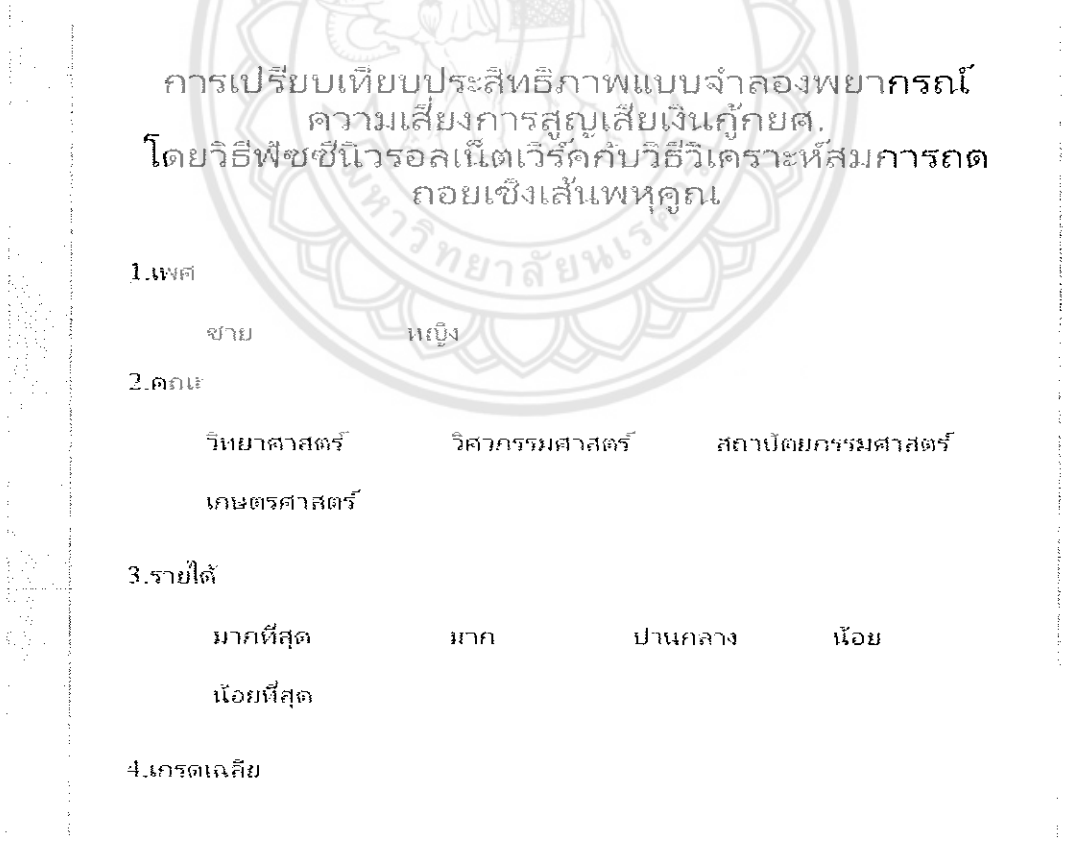
```
-----------------
sci: 1.216
engi: 0.357
aggri: 0.932
GPAX: -0.009
Sex: 0.615
Income: -0.795
CousineN: 0.216
StatusHist: -0.700
Bias: -0.371

Node 3 (Sigmoid)
-----------------
sci: 0.839
engi: 1.466
aggri: -0.325
GPAX: 1.023
Sex: -1.659
Income: 0.154
CousineN: -0.676
StatusHist: 0.500
Bias: -1.745

Node 4 (Sigmoid)
-----------------
sci: 0.279
engi: 0.172
aggri: 0.574
GPAX: -0.053
Sex: 0.358
Income: 0.352
CousineN: 0.359
StatusHist: 0.332
Bias: -0.022

Node 5 (Sigmoid)
-----------------
sci: -0.656
engi: 0.253
aggri: -0.067
GPAX: 0.277
Sex: -0.322
Income: -0.235
CousineN: -0.521
StatusHist: 0.027
Bias: -0.674

Node 6 (Sigmoid)
-----------------
sci: -0.257
engi: -0.281
aggri: -0.840
GPAX: -0.300
Sex: -0.030
Income: -0.172
CousineN: 0.227
StatusHist: 0.064
Bias: -0.521

Node 7 (Sigmoid)
-----------------
```

```
sci: 0.182
engi: 0.382
aggri: 1.135
GPAX: 0.233
Sex: 0.289
Income: 0.083
CousineN: -0.095
StatusHist: -0.093
Bias: -0.135


Output
======

Class 'less' (Sigmoid)
----------------------
Node 1: -0.294
Node 2: 0.752
Node 3: 1.157
Node 4: 0.147
Node 5: -0.628
Node 6: -0.880
Node 7: 0.474
Threshold: -0.692

Class 'medium' (Sigmoid)
------------------------
Node 1: -0.924
Node 2: -0.623
Node 3: 0.665
Node 4: -0.522
Node 5: 0.119
Node 6: -0.577
Node 7: -0.274
Threshold: -0.702

Class 'high' (Sigmoid)
----------------------
Node 1: -0.481
Node 2: -1.641
Node 3: -0.321
Node 4: -1.360
Node 5: 0.264
Node 6: 0.266
Node 7: -1.267
Threshold: -1.302

Class 'verl' (Sigmoid)
----------------------
Node 1: -0.535
Node 2: 0.392
Node 3: -0.773
Node 4: -0.302
Node 5: -1.126
Node 6: -0.229
Node 7: -0.774
Threshold: -0.825

Class 'none' (Sigmoid)
----------------------
Node 1: -0.780
```

```
Node 2: -0.549
Node 3: -0.389
Node 4: 0.116
Node 5: -0.292
Node 6: -0.812
Node 7: 0.177
Threshold: -0.694

Class 'veryhigh' (Sigmoid)
---------------------------
Node 1: 1.010
Node 2: -0.918
Node 3: -1.668
Node 4: -0.070
Node 5: -0.698
Node 6: -0.222
Node 7: -0.527
Threshold: -0.742
```



ภาพผนวก ข-1 แสดงหน้าจอหลักการทำงานของโปรแกรม



ภาพผนวก ข-2 แสดงหน้าจอการกรอกข้อมูลเพื่อใช้พยากรณ์ของโปรแกรม

ผลการพยากรณ์

ระดับความเสี่ยงไม่ชำระหนี้กองทุน กยศ
"ต่ำ"

ภาพผนวกข-3 แสดงตัวอย่างการทดสอบ

```
<html class=" -webkit-">
  <head>…</head>
  <body>
    <div id="form-main">
      <div id="form-div">
        <form class="form" action="index.html" method="Post">
          <center>…</center>
            <br />
            <br />
            <center></center>
            <img class="form-control" alt="image1" src="img/t2.jpg" />
            <br />
            <center></center>
            <br />
          <center>
            <font color="red" size="6">ระดับความเสี่ยงไม่ชำระหนี้กองทุน…</font>
            </center>
            <br />
          <div class="submit">
              <input id="button-blue" type="submit" value="คลิกดูข้อมูลอีกครั้ง" />
              <div class="ease"></div>
            </div>
          </form>
        </div>
      </div>
    </body>
  </html>
```

ภาคผนวก ข-4 แสดงตัวอย่างโค้ดบางส่วนของการทำงาน