

การประยุกต์ใช้วิธีวิเคราะห์ความถดถอยร่วมกับตรรกศาสตร์คลุมเครือเพื่อพยากรณ์

การพัฒนาศาสนิกศึกษาสาขาวิชาเทคโนโลยีสารสนเทศธุรกิจ

Apply regression analysis and fuzzy logic to forecast the students status in curriculum of Business Information

สัจฉาภูมิ ไทยพานิช¹ ศณินทร เสือแพรว² และ พยุง มีสัง²

¹คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ

²คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ABSTRACT – This article presents the prediction techniques for the status of students in the department Information technology on Faculty of Business Administration and information technology Rajamangala University of Technology Suvarnabhumi. Linear regression analysis and fuzzy logic is chosen to generate the predictor model Status of students. Background information of the student in first session of information technology department is academic year 2551 (2008) number 151 people were selected to analyze and create a student status prediction models. Then tested to determine the error in forecasting from background information of the student in first session is academic year 2552(2009) in the same department. The experiment showed that the method presented has the error in the forecast of 0.02 %

KEY WORDS – Regression Analysis; Fuzzy Logicist; Prediction Technique

บทคัดย่อ -- บทความนี้นำเสนอเทคนิคการพยากรณ์การพัฒนาศาสนิกศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ การวิเคราะห์ความถดถอยเชิงเส้น และวิธีการตรรกศาสตร์คลุมเครือถูกเลือกเพื่อสร้างระบบพยากรณ์การพัฒนาศาสนิกศึกษา ข้อมูลปฐมหลังของนักศึกษา ชั้นปีที่ 1 สาขาวิชาเทคโนโลยีสารสนเทศ ปีการศึกษา 2551 จำนวน 151 คนถูกเลือกใช้เพื่อวิเคราะห์และสร้างแบบจำลอง พยากรณ์การพัฒนาศาสนิกศึกษา และทดสอบเพื่อหาความคลาดเคลื่อนในการพยากรณ์ด้วยข้อมูลนักศึกษาชั้นปีที่ 1 ปี การศึกษา 2552 ในสาขาวิชาเดียวกัน ผลการทดสอบพบว่า วิธีการที่นำเสนอมีความคลาดเคลื่อนในการพยากรณ์ เท่ากับ 0.02 เปอร์เซ็นต์

คำสำคัญ -- การวิเคราะห์ความถดถอย; ตรรกศาสตร์คลุมเครือ, เทคนิคการพยากรณ์

1. บทนำ

การพัฒนาศาสนิกศึกษานับว่าเป็นปัจจัยสำคัญในการจัดการ การเรียนการสอน ซึ่งมีผลกระทบโดยตรงต่อระบบงานบริหาร ภายในองค์กร โดยเฉพาะการบริหารจัดการเงินงบประมาณ การ คาดการณ์โอกาสการพัฒนาศาสนิกของนักศึกษาย่อมทำให้เกิด ประโยชน์ในการวางแผนการจัดการเรียนการสอน ตลอดจน การบริหารงบประมาณ ผู้วิจัยตระหนักถึงความสำคัญในปัญหา

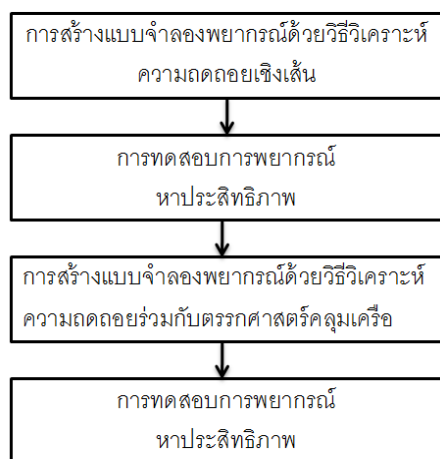
ดังกล่าวจึงมีแนวคิดในการศึกษาและออกแบบวิธีการพยากรณ์ การพัฒนาศาสนิกศึกษาในสาขาวิชาเทคโนโลยีคอมพิวเตอร์ เพื่อ ใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการวางแผนการ จัดการเรียนการสอนและงบประมาณ

การวิเคราะห์ความถดถอยนับได้ว่าเป็นกรรมวิธีในการ พยากรณ์ที่ได้รับความนิยมในการวิเคราะห์ความสัมพันธ์ของ ข้อมูล ตลอดจนใช้ในการสร้างแบบจำลองเพื่อพยากรณ์

ประมาณค่า [1] การวิเคราะห์ความถดถอยเหมาะกับการพยากรณ์ค่าที่มีช่วงที่หลากหลาย การนำวิธีการดังกล่าวมาใช้พยากรณ์ที่มีรูปแบบไบนารีจึงไม่เหมาะสม ตรรกศาสตร์คลุมเครือเป็นเครื่องมือที่ช่วยในการตัดสินใจภายใต้ความไม่แน่นอนของข้อมูล โดยยอมให้มีความยืดหยุ่นได้ ใช้หลักเหตุผลที่คล้ายการเลียนแบบวิธีความคิดที่ซับซ้อนของมนุษย์[2] ผู้วิจัยมีแนวคิดในการผสมผสานวิธีการทั้งสองเพื่อให้ออกแบบวิธีการพยากรณ์การฟื้นสภาพของนักศึกษาเพื่อประสิทธิภาพที่สูงขึ้นเมื่อเทียบกับการพยากรณ์ด้วยวิธีวิเคราะห์ความถดถอย

2. ขั้นตอนการวิจัย

ในการทำวิจัยได้ออกแบบขั้นตอนการดำเนินงานออกเป็น 4 ส่วน โดยเป็นขั้นตอนในการสร้างแบบจำลองในการพยากรณ์การฟื้นสภาพนักศึกษา และการทดลองเพื่อหาค่าความคลาดเคลื่อนของแบบจำลอง ซึ่งขั้นตอนทั้งสองใช้ข้อมูลจากนักศึกษาชั้นปีที่ 1 ปีการศึกษา 2551 และ 2552 ตามลำดับ โดยแสดงลำดับขั้นตอนการปฏิบัติงานตามผังขั้นตอนในรูปที่ 1



รูปที่ 1 ผังการดำเนินการ

2.1 การสร้างแบบจำลองพยากรณ์ด้วยวิธีวิเคราะห์ความถดถอยเชิงเส้น

การวิเคราะห์ความถดถอย (Regression analysis) [3] เป็นเครื่องมือทางสถิติที่มีการประยุกต์ใช้ในการประมวลผลข้อมูลในงานวิจัยค่อนข้างมาก นักสถิติยุคก่อนได้คิดค้นทฤษฎี

เกี่ยวกับ Regression เอาไว้มากมาย โดยเราจำแนกประเภทของ Regression ออกได้ 2 ประเภทคือ Continuous response และแบบ Discrete response ซึ่งทั้งสองประเภทเหมาะที่จะนำไปใช้กับลักษณะของข้อมูลที่แตกต่างกัน ในการดำเนินการนี้จะนำเสนอการสร้างแบบจำลองด้วยการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่ายและการวิเคราะห์ความถดถอยเชิงเส้นพหุ การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่ายคือ การวิเคราะห์ความสัมพันธ์ของ 2 ตัวแปร คือตัวแปรอิสระหนึ่งตัวและตัวแปรตามหนึ่งตัว ความสัมพันธ์ของตัวแปรทั้งสองเป็นลักษณะเชิงเส้นตรง ถ้ามีตัวแปร 2 ตัวเป็นคู่ๆ เราอาจจะดูว่าตัวแปร 2 ตัวนั้นมีความสัมพันธ์กันในลักษณะใด โดยการสร้างแผนภาพการกระจาย (Scatter Diagram) ระหว่างตัวแปร 2 ตัว ในกรณีที่แผนภาพการกระจายนั้นเป็นเส้นตรง เราสามารถสร้างเป็นความสัมพันธ์ในรูปแบบของสมการได้ ดังสมการที่ 1 ซึ่งสมการดังกล่าวเรียกว่าสมการถดถอยอย่างง่าย [4]

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

โดย Y_i เป็นค่าของตัวแปรตาม
 α เป็นค่าของ Y_i เมื่อ X_i เป็นศูนย์
 β เป็นค่า Slope ของสมการ, สัมประสิทธิ์ของความถดถอย
 ε_i เป็นค่าความคลาดเคลื่อนระหว่างค่าจริงกับค่าประมาณของ Y_i

วิธีหาค่า α และค่า β จะประมาณโดยวิธีกำลังสองน้อยที่สุดซึ่งเป็นวิธีที่ได้รับความนิยมกันมาก โดยมีหลักการคือจะทำให้ผลรวมกำลังสองของความแตกต่างระหว่าง Y_i และ $\hat{Y}_i$ ที่ได้จากการประมาณมีค่าต่ำที่สุด ซึ่งสมการประมาณการจะมีรูปแบบ ดังสมการที่ 2 [4]

$$\hat{Y}_i = a + bX_i \quad (2)$$

โดย a เป็นค่า Y-intercept เป็นค่าประมาณของ α
 b คือค่า Slope เป็นค่าประมาณของ β
 $\hat{Y}_i$ คือค่าประมาณการของ Y_i เมื่อตัวแปรอิสระคือ X_i

ในการสร้างแบบจำลองด้วยวิธีวิเคราะห์ความถดถอยเชิงเส้นแบบง่าย ข้อมูลป้อนหลังของนักศึกษาที่เข้าศึกษาชั้นปีที่ 1 ในปีการศึกษา 2551 จำนวน 151 คน ถูกใช้ในการวิเคราะห์เพื่อสร้างแบบจำลอง ข้อมูลนักศึกษามีประกอบด้วยตัวแปรจำนวน 43 ตัวแปร และนำข้อมูลแต่ละตัวแปรมาวิเคราะห์ทางสถิติพบว่า มีตัว

Name of Variable	Model	R-Square (adj)
Sex	$Y = 0.455 - 0.193\text{Sex}$	2.9%
GPA	$Y = 1.40 - 0.503\text{GPA}$	43.1%
Computer math	$Y = 1.01 - 0.242\text{Commath}$	18.8%
OS	$Y = 0.837 - 0.443\text{OS}$	76.0%
Major	$Y = 0.268 + 0.0681\text{Major}$	0.0%
Family Status	$Y = 0.317 + 0.0005\text{Status}$	0.0%
Programming	$Y = 0.775 - 0.308\text{Program}$	24.2%

แปร 7 ตัวมีความสัมพันธ์กับการผันสภาพของนักศึกษาแสดงดังตารางที่ 1

ตาราง 1. แสดงแบบจำลองและค่าสัมประสิทธิ์การตัดสินใจของตัวแปรอิสระ

จากแบบจำลองที่นำเสนอในตารางที่ 1 จะเห็นว่าแบบจำลองที่สร้างจากตัวแปรอิสระ OS มีค่าสัมประสิทธิ์การตัดสินใจสูงที่สุด จึงเลือกแบบจำลองดังกล่าวใช้ในการทดสอบหาประสิทธิภาพของแบบจำลอง โดยใช้ข้อมูลของนักศึกษาจำนวน 49 คน เพื่อทดสอบหาความคลาดเคลื่อนในการพยากรณ์พบว่าแบบจำลองที่นำเสนอมีการพยากรณ์ผิดพลาดจำนวน 7 ตัวอย่างคิดเป็นร้อยละ 14.28%

การวิเคราะห์ความถดถอยเชิงเส้นพหุ นั้นเป็นสถิติที่นำมาใช้เพื่อวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระที่มากกว่า 1 ตัวแปร และตัวแปรตาม 1 ตัวแปร โดยตัวแปรอิสระจะเป็นตัวแปรสำหรับทำนายหรือพยากรณ์การผันแปรของตัวแปรตาม ในบางครั้งอาจเรียกตัวแปรอิสระว่า ตัวแปรพยากรณ์ (Predictor Variables) [5] โดยสมการวิเคราะห์ความถดถอยเชิงเส้นพหุสามารถแสดงได้ ดังสมการที่ 3 [5]

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (3)$$

ในการสร้างแบบจำลองด้วยวิธีวิเคราะห์ความถดถอยเชิงเส้นแบบพหุ ได้เลือกใช้ข้อมูลชุดเดียวกันกับการสร้างแบบจำลองด้วยวิธีวิเคราะห์ความถดถอยเชิงเส้นแบบง่าย จากการวิเคราะห์ความถดถอยพหุพบว่าแบบจำลองที่สร้างจากตัวแปรอิสระทั้ง 7 ตัวแปร มีค่าสัมประสิทธิ์การตัดสินใจเท่ากับ

Model	R-Square (adj)
$Y = 1.46 - 0.119\text{Sex} - 0.493\text{GPA}$	44.1%
$Y = 1.07 - 0.121\text{Sex} - 0.232\text{Commath}$	19.6%
$Y = 0.419 - 0.187\text{Sex} + 0.0433\text{Major}$	2.4%
$Y = 0.456 - 0.193\text{Sex} - 0.0015\text{Status}$	2.2%
$Y = 0.816 - 0.0810\text{Sex} - 0.297\text{Program}$	24.3%
...	...
$Y = 1.00 - 0.120\text{Sex} - 0.0393\text{Commath} - 0.439\text{OS} - 0.0195\text{Major} - 0.0141\text{Status} + 0.0369\text{Program}$	76.7%

86.5% โดยแสดงข้อมูลการวิเคราะห์ความถดถอยพหุของแบบจำลองต่างดังตารางที่ 2 และแสดงแบบจำลองที่มีค่าสัมประสิทธิ์การตัดสินใจสูงสุดดังสมการที่ 4

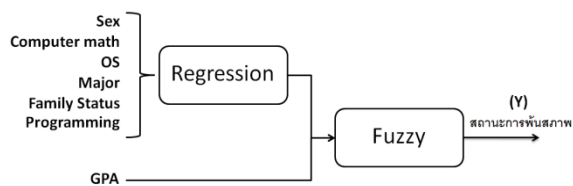
ตารางที่ 2. แสดงแบบจำลองและค่าสัมประสิทธิ์การตัดสินใจของตัวแปรอิสระ

$$\hat{Y} = 0.381 - 0.122\text{sex} + 0.590\text{gpa} - 0.184\text{computer}_{\text{math}} - 0.749\text{os} + 0.0584\text{major} - 0.030\text{family}_{\text{status}} + 0.097\text{programming} \quad (4)$$

จากสมการที่ 4 เมื่อนำแบบจำลองดังกล่าวมาใช้ทดสอบหาความคลาดเคลื่อนในการพยากรณ์การผันสภาพนักศึกษา โดยใช้ข้อมูลของนักศึกษาที่เข้าศึกษาชั้นปีที่ 1 ปีการศึกษา 2552 จำนวน 49 คนพบว่าแบบจำลองที่นำเสนอพยากรณ์ผิดพลาดจำนวนทั้งสิ้น 7 ตัวอย่างคิดเป็นร้อยละ 14.28% เช่นเดียวกับแบบจำลองที่สร้างจากข้อมูล OS

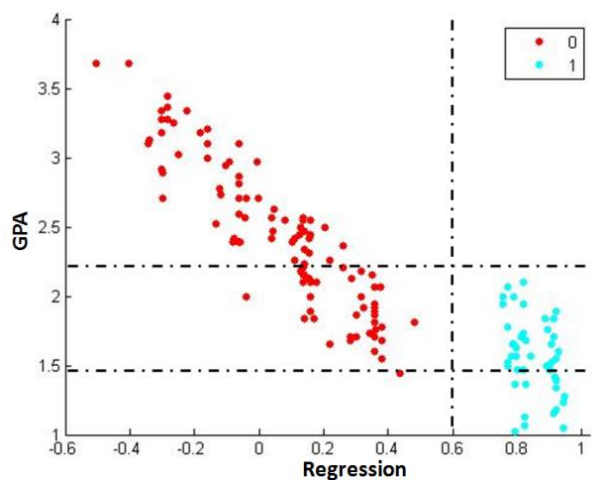
2.2 การสร้างแบบจำลองพยากรณ์ด้วยวิธีวิเคราะห์ความถดถอยเชิงเส้นร่วมกับตรรกศาสตร์คลุมเครือ

การวิเคราะห์ความถดถอยเชิงเส้นแบบง่ายและพบว่ามีความคลาดเคลื่อน 14.28% ผู้วิจัยจึงทำการพัฒนาสร้างแบบจำลองให้มีประสิทธิภาพในการพยากรณ์ที่เพิ่มขึ้น ด้วยการประยุกต์ใช้วิธีการตรรกศาสตร์คลุมเคลือร่วมกับการวิเคราะห์ความถดถอยเชิงเส้นพหุ โดยมีแนวคิดในการนำค่าของตัวแปรตาม (Y) ที่ได้จากการวิเคราะห์ความถดถอยเชิงเส้นพหุและค่า GPA นักศึกษาใช้เป็นอินพุต (Input) เข้าสู่ตรรกศาสตร์คลุมเครือ เพื่อพยากรณ์การผันสภาพของนักศึกษา โดยแสดงไดอะแกรมดังรูปที่ 3



รูปที่ 2 ฟังไดอะแกรมระบบ

ในการออกแบบระบบฟัซซี่ผู้วิจัยได้เลือกสร้างระบบฟัซซี่ด้วยวิธีของ Mamdani โดยกำหนดให้ membership function เป็นแบบ Trapezoidal และ Triangular ในการออกแบบระบบฟัซซี่นั้นต้องทำการวิเคราะห์ลักษณะข้อมูลด้วยการพล็อตแบบกระจาย (scatter plot)



รูปที่ 3 การพล็อตแบบกระจาย

จากนั้นทำการคำนวณค่าทางสถิติค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ยของข้อมูลในแต่ละมิติ และคำนวณค่าสูงสุด ค่าต่ำสุดและค่าเฉลี่ยในแต่ละกลุ่ม เพื่อใช้ในการกำหนดฟัซซี่เซต โดยแสดงค่าในตารางที่ 3 - 5

ตารางที่ 3. แสดงค่าสถิติของข้อมูลทั้งหมดทุกมิติ

ค่าสถิติของข้อมูลทั้งหมดแต่ละมิติ		
	Regression	GPA
MAX	0.946	3.68
AVERAGE	0.315	2.153
MIN	-0.503	1.02

ตารางที่ 4 แสดงค่าสถิติของข้อมูลในกลุ่มที่ 0

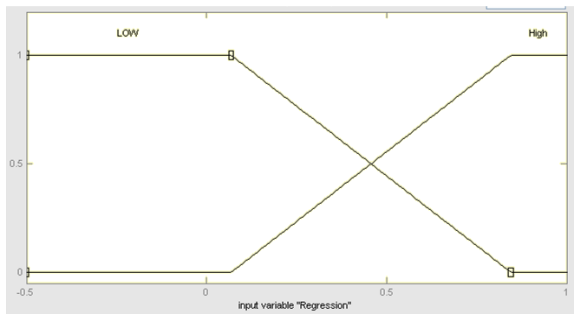
ค่าสถิติของข้อมูลกลุ่มที่ 0		
	Regression	GPA
MAX	0.483	3.68
AVERAGE	0.068	2.428
MIN	-0.503	1.44

ตารางที่ 5 แสดงค่าสถิติของข้อมูลในกลุ่มที่ 1

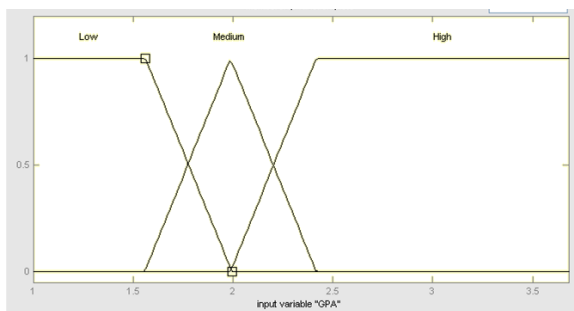
ค่าสถิติของข้อมูลกลุ่มที่ 1		
	Regression	GPA
MAX	0.946	2.1
AVERAGE	0.845	1.563
MIN	0.754	1.02

จากข้อมูลผู้วิจัยได้ทำการสร้างฟัซซี่เซตส่วนของอินพุตจำนวน 2 ตัว คือ Regression และ GPA เมื่อพิจารณาจากรูปที่ 3 ได้ทำการสร้างฟัซซี่เซต Regression ให้มี membership function จำนวน 2 สถานะ และสร้างฟัซซี่เซต GPA ให้มี membership function จำนวน 3 สถานะ ดังรูปที่ 4 และ 5 ตามลำดับ แสดง

กฎฟัซซี่ที่ออกแบบดังสมการที่ 5 แสดงกฎและค่าผลลัพธ์ฟัซซี่
ในรูปกราฟพื้นผิว (Rule Surface) ดังรูปที่ 6

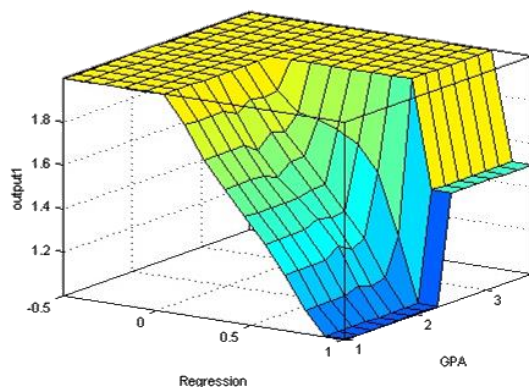


รูปที่ 4 แสดงฟัซซี่เซต Regression



รูปที่ 5 แสดงฟัซซี่เซต GPA

$$Y = \begin{cases} 0 & \text{if Regression} = 0 \text{ and GPA} = 0 ; \text{or} \\ 0 & \text{if Regression} = 0 \text{ and GPA} = 1 ; \text{or} \\ 0 & \text{if Regression} = 0 \text{ and GPA} = 2 ; \text{or} \\ 1 & \text{if Regression} = 1 \text{ and GPA} = 0 ; \text{or} \\ 1 & \text{if Regression} = 1 \text{ and GPA} = 1 ; \text{end} \end{cases} \quad (5)$$



รูปที่ 6 กราฟพื้นผิวของกฎและผลลัพธ์ฟัซซี่

จากแบบจำลองที่สร้างด้วยวิธีการวิเคราะห์ความถดถอยเชิงเส้น
พหุร่วมกับวิธีการตรรกศาสตร์คลุมเครือ เมื่อนำมาทดสอบกับ
ข้อมูลนักศึกษาจำนวน 49 ข้อมูล พบว่ามีประสิทธิภาพในการ
พยากรณ์เพิ่มสูงขึ้นอย่างเห็นได้ชัด โดยผลการทดลองมีการ
พยากรณ์ผิดจำนวน 1 ข้อมูล มีค่าความคลาดเคลื่อนของ
แบบจำลองในการพยากรณ์เท่ากับ 0.02%

3. สรุปและอภิปรายผล

การสร้างแบบจำลองพยากรณ์การฟื้นฟูสภาพของนักศึกษา
สาขาวิชาเทคโนโลยีสารสนเทศที่นำเสนอพบว่าข้อมูลประวัติ
นักศึกษายางข้อมูลมีความสัมพันธ์กับการฟื้นฟูสภาพนักศึกษา
โดยหาด้วยวิธีการวิเคราะห์ความถดถอยเชิงเส้นแบบง่าย เมื่อนำ
ข้อมูลทั้งหมดมาวิเคราะห์ความถดถอยโดยกำหนดให้การฟื้นฟู
สภาพของนักศึกษาเป็นตัวแปรตามผลปรากฏว่ามีข้อมูลที่มีค่า
สัมประสิทธิ์การตัดสินใจของตัวแปรอิสระ R-Square (adj) ที่
แตกต่างกัน และพบว่าตัวแปร Sex, GPA, Computer math, OS,
Family Status, และ Programming มีความสัมพันธ์กับ
สถานะการฟื้นฟูสภาพของนักศึกษา โดยแบบจำลองที่ได้จาก
วิธีการวิเคราะห์ความถดถอยเชิงเส้นพหุมีค่าสัมประสิทธิ์การ
ตัดสินใจสูงที่สุด โดยมีค่าความคลาดเคลื่อน 14.28% เมื่อทำ
การทดสอบด้วยข้อมูลของนักศึกษาจำนวน 49 ชุด และเมื่อทำ
การปรับปรุงวิธีการสร้างแบบจำลองโดยนำวิธีการตรรกศาสตร์
คลุมเครือมารวมใช้งานผลปรากฏว่าวิธีการผสมผสานมีค่า
ความผิดพลาดที่ต่ำกว่าคือ 0.02% ดังแสดงในตารางที่ 6

ตารางที่ 6 แสดงตัวอย่างผลการทดสอบ

Regression	GPA	สถานะที่ได้จาก Fuzzy	สถานะ นักศึกษา
0.13825	2.21	0	0
0.76975	1.78	1	1
0.35775	2.07	0	0
...	...	...	...
0.48315	1.81	1	0
0.78385	1.65	1	1
0.26065	2.21	0	0

จากการปฏิบัติงานพบว่า การสร้างแบบจำลองด้วยวิธีการวิเคราะห์ความถดถอยอาจไม่เหมาะสมกับชนิดข้อมูลที่เป็นกลุ่ม เนื่องจากค่าที่ได้จากการวิเคราะห์ความถดถอยมีช่วงของผลลัพธ์ที่หลากหลายค่า การนำวิธีการตรรกศาสตร์คลุมเครือมาประยุกต์ใช้ร่วมกับวิธีการดังกล่าวทำให้เกิดประสิทธิภาพที่สูงขึ้นในการพยากรณ์

เอกสารอ้างอิง

- [1] D. Dubois, H. Prade, "Fuzzy Sets and Systems: Theory and Applications", Academic Publishers, New York, 1980.
- [2] พยุง มีสัจ. ระบบฟัซซี่และโครงข่ายประสาทเทียม. ศูนย์ผลิตตำราเรียน มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2555.

- [3] บุญใจ ศรีสถิตนรากร. "ระเบียบวิธีวิจัยแนวทางสู่ความสำเร็จ" สำนักพิมพ์แห่งจุฬาลงกรณ์, 2550
- [4] Zimmermann, H. "Fuzzy set theory and its applications". Boston: Kluwer Academic Publishers. ISBN 0-7923-7435-5.
- [5] ฉลอง สืบแก้วสีว. "Regression Analysis", Thailand (Online) Available: [http:// www.statistics.ob.tc/ reg1.htm](http://www.statistics.ob.tc/reg1.htm), 9 July 2012.
- [6] การประมาณความคลาดเคลื่อนมาตรฐานของสมการถดถอย , Available: <http://www.oocities.com/goodyuta/regression.doc>, 11 September 2012.
- [7] Regression Analysis, Available: http://www.nubkk.nu.ac.th/picnews/s_1327074364.chapter%2010%20statistic.pdf, 11 September 2012.