

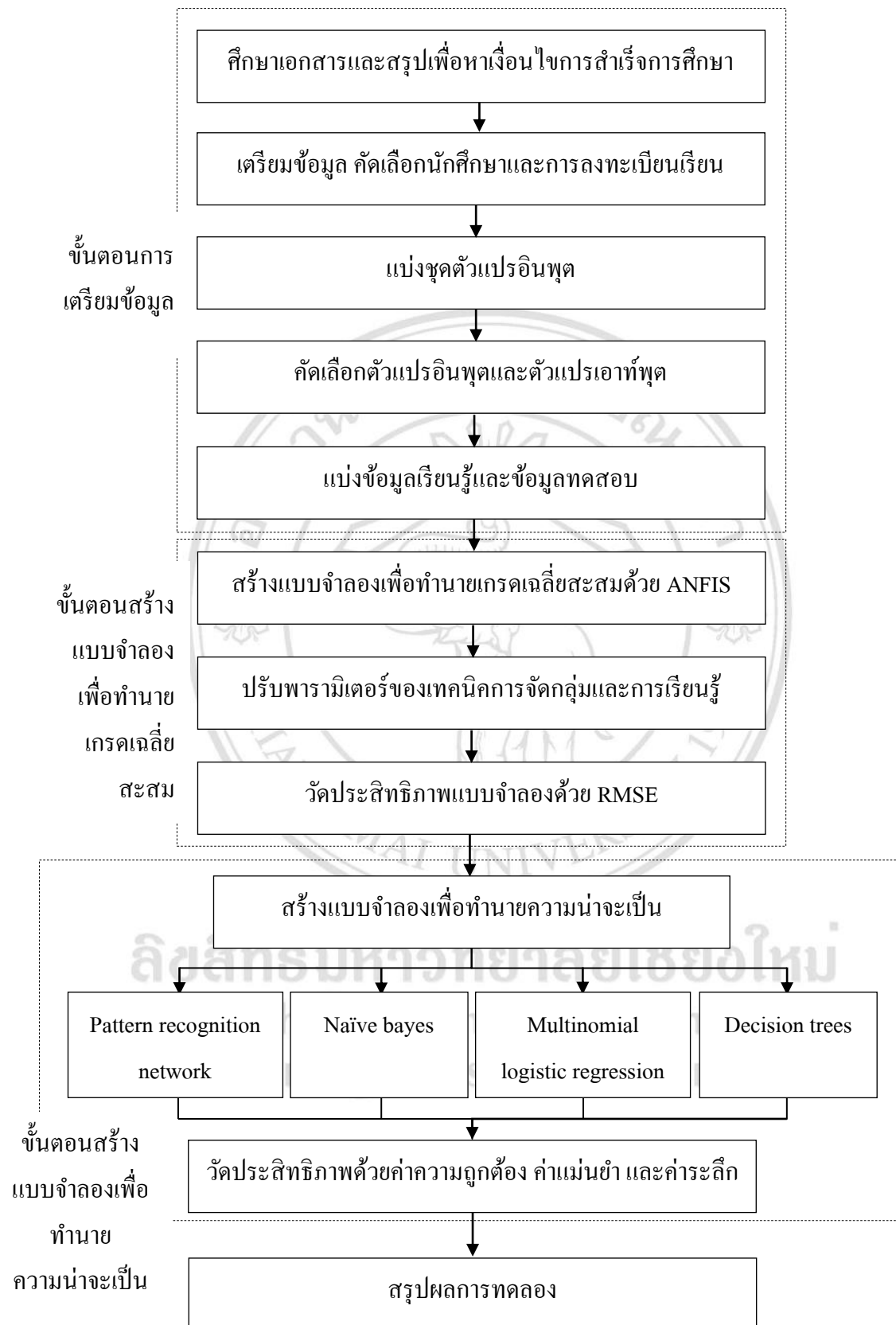
บทที่ 3

วิธีการดำเนินงาน

การทำนายความน่าจะเป็นของการสำเร็จการศึกษาจะใช้หมวดหมู่หรือกลุ่มวิชาในโครงสร้างหลักสูตรเป็นตัวแปรอินพุต มาประยุกต์ใช้กับการทำเหมืองข้อมูลทางการศึกษา และใช้เทคนิคระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้เพื่อทำนายเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชา จากนั้นจะนำเกรดเฉลี่ยสะสมที่ได้มาหาค่าความน่าจะเป็นด้วยเทคนิคโครงข่ายรู้จำแบบ การวิเคราะห์ความถดถอยโลจิสติกพหุกลุ่ม การจำแนกประเภทเบย์อย่างง่าย และตัดไม้ตัดสินใจ มีขั้นตอนการดำเนินงานดังนี้

3.1 ขั้นตอนการดำเนินงาน

จากปัญหาหาการฟื้นฟูสภาพหรือจบเกินหลักสูตร ถ้าหากสามารถทำนายความน่าจะเป็นของการสำเร็จการศึกษาได้ล่วงหน้าก็จะเป็นการเพิ่มโอกาสการจบตามหลักสูตรและสามารถแก้ไขปัญหาดังกล่าวได้ จึงมีการติดตามความก้าวหน้าด้วยการดูเกรดเฉลี่ยสะสมรวม แต่ยังมีปัจจัยอื่นที่เกี่ยวข้องกับการสำเร็จการศึกษาด้วย เช่น เกรดเฉลี่ยสะสมของกลุ่มวิชาเอก หน่วยกิตสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชา จึงต้องมีการศึกษาหาเงื่อนไขการสำเร็จการศึกษาหรือการฟื้นฟูสภาพก่อนเพื่อใช้เป็นตัวแปรอินพุตของการสร้างแบบจำลองการทำนายเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาในภาคเรียนที่ 11 หรือชั้นปี 4 ภาคเรียนที่ 2 ซึ่งจะช่วยให้ทราบถึงหมวดหมู่หรือกลุ่มวิชาที่เป็นปัญหาก่อนล่วงหน้า จากนั้นจะนำเกรดเฉลี่ยสะสมที่ได้มาเป็นตัวแปรอินพุตในแบบจำลองทำนายค่าความน่าจะเป็นของการสำเร็จการศึกษา เพื่อแนะนำความน่าจะเป็นที่จะอยู่ในสถานภาพต่างๆ ให้กับนักศึกษาต่อไป มีขั้นตอนการทำงานตามกระบวนการสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานความรู้ด้านการศึกษาที่ได้มีการศึกษาจากบทก่อนหน้า มีขั้นตอนการดำเนินงานของงานวิจัย ดังนี้ ขั้นตอนการเตรียมข้อมูล ขั้นตอนการจัดกลุ่ม ขั้นตอนการทำนายเกรดเฉลี่ยสะสมภาคเรียนที่ 11 หรือชั้นปีที่ 4 ภาคเรียนที่ 2 ของแต่ละหมวดหมู่หรือกลุ่มวิชา และขั้นตอนการทำนายค่าความน่าจะเป็น ดังภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการดำเนินงานของงานวิจัย

3.2 ขั้นตอนการเตรียมข้อมูล

3.2.1 ศึกษาเอกสารโครงสร้างหลักสูตรของหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ และ เอกสารข้อบังคับมหาวิทยาลัยเชียงใหม่ ว่าด้วยการศึกษาระดับปริญญาตรี พ.ศ. 2550 เพื่อหาเงื่อนไขการสำเร็จการศึกษาและพื้นฐาน สรุปลงเงื่อนไขการสำเร็จการศึกษาได้ว่า เงื่อนไขหลักของการสำเร็จการศึกษา คือ เกรดเฉลี่ยสะสมเฉลี่ยหน่วยกิต และรายวิชาบังคับ ซึ่งจะสำเร็จการศึกษาได้ต้องมีระยะเวลาการศึกษาทั้งหมดมากกว่าหรือเท่ากับ 7 ภาคเรียนปกติ แต่ข้อมูลที่นำมาใช้ในงานวิจัยจะเลือกใช้ระยะเวลาการศึกษามากกว่าหรือเท่ากับ 8 ภาคเรียนปกติสำหรับกรณีสำเร็จการศึกษา ซึ่งยังอยู่ในเงื่อนไขของการสำเร็จการศึกษาตามหลักสูตรคือ 8 ภาคเรียนปกติ ถ้ามากกว่านี้จะมีสถานภาพจบเกินหลักสูตร ส่วนเงื่อนไขหลักของการพื้นฐาน คือ เกรดเฉลี่ยสะสม มีรายละเอียดของเงื่อนไขต่างๆ ดังนี้

- 1) เกรดเฉลี่ยสะสม
 - 1.1) กรณีสำเร็จการศึกษา
 - 1.1.1) เกรดเฉลี่ยสะสมเฉลี่ยในวิชาเอก ไม่น้อยกว่า 2.00
 - 1.1.2) เกรดเฉลี่ยสะสมเฉลี่ยตลอดหลักสูตร ไม่น้อยกว่า 2.00
 - 1.2) กรณีพื้นฐาน
 - 1.2.1) เมื่อเรียนมาแล้วครบสองภาคเรียนปกติ ยังมีเกรดเฉลี่ยสะสมเฉลี่ยทั้งหมดไม่ถึง 1.50
 - 1.2.2) เมื่อเรียนมาแล้วครบสี่ภาคเรียนปกติ ยังมีเกรดเฉลี่ยสะสมเฉลี่ยทั้งหมดไม่ถึง 1.75
 - 1.2.3) เมื่อเรียนมาแล้วสี่ภาคเรียนปกติขึ้นไป ยังมีเกรดเฉลี่ยสะสมเฉลี่ยทั้งหมด ไม่ถึง 1.75 ติดต่อกันถึงสองภาคการศึกษาปกติ
- 2) หน่วยกิตสะสม แบ่งเป็น 2 แผนการเรียน มี 14 เงื่อนไข ดังตารางที่ 3.1

ตารางที่ 3.1 เงื่อนไขการสำเร็จการศึกษาเกี่ยวกับหน่วยกิต

ลำดับ	หมวดหมู่วิชา	หน่วยกิต	
		แผน 1	แผน 2
1	หน่วยกิตรวมตลอดหลักสูตร	130	132
2	หมวดวิชาศึกษาทั่วไป	30	30
3	กลุ่มวิชาสังคมศาสตร์	6	6

ตารางที่ 3.1 เงื่อนไขการสำเร็จการศึกษาเกี่ยวกับหน่วยกิต (ต่อ)

ลำดับ	หมวดหมู่วิชา	หน่วยกิต	
		แผน 1	แผน 2
4	กลุ่มวิชามนุษยศาสตร์	6	6
5	กลุ่มวิชาภาษาและการสื่อสาร	12	12
6	กลุ่มวิชาวิทย์และคณิต	6	6
7	หมวดวิชาเฉพาะ	94	96
8	วิชาแกน	24	24
9	วิชาเอก	55	57
10	วิชาเอกบังคับ	36	36
11	วิชาเอกบังคับเลือก	4	9
12	วิชาเอกเลือก	15	12
13	วิชาโท	15	15
14	หมวดวิชาเลือกเสรี	6	6

3) รายวิชาบังคับ มีทั้งหมด 32 เงื่อนไข ดังตารางที่ 3.2

ตารางที่ 3.2 เงื่อนไขการสำเร็จการศึกษาเกี่ยวกับรายวิชาบังคับ

ลำดับ	รหัสวิชา	หมวดหมู่วิชา
1	001103	ภาษาอังกฤษพื้นฐาน 1
2	001104	ภาษาอังกฤษพื้นฐาน 2
3	001203	ภาษาอังกฤษเพื่อการศึกษา
4	001205	ภาษาอังกฤษสาขาวิชาวิทยาศาสตร์และเทคโนโลยี
5	201110	คณิตศาสตร์บูรณาการ
6	201111	โลกแห่งวิทยาศาสตร์
7	202111	ชีววิทยา 1
8	203111	เคมี 1
9	203115	ปฏิบัติการเคมี 1
10	206116	แคลคูลัส 1
11	206112	แคลคูลัส 2
12	206281	คณิตศาสตร์ดิสครีต
13	207117	ปฏิบัติการฟิสิกส์ 1
14	207187	ฟิสิกส์ 1
15	208263	สถิติเบื้องต้น

ตารางที่ 3.2 เงื่อนไขการสำเร็จการศึกษาเกี่ยวกับรายวิชาบังคับ (ต่อ)

ลำดับ	รหัสวิชา	หมวดหมู่วิชา
16	204111	วิทยาการคอมพิวเตอร์เบื้องต้น
17	204112	การเขียนโปรแกรมคอมพิวเตอร์ 1
18	204211	การเขียนโปรแกรมคอมพิวเตอร์ 2
19	204218	การจัดการระบบคอมพิวเตอร์และการเขียนโปรแกรมระบบ
20	204306	จริยธรรมสำหรับผู้ประกอบวิชาชีพคอมพิวเตอร์
21	204315	การจัดระเบียบของภาษาโปรแกรม
22	204321	ระบบฐานข้อมูล 1
23	204341	ระบบการดำเนินงาน
24	204351	โครงสร้างข้อมูล
25	204361	วิศวกรรมซอฟต์แวร์
26	204391	หลักพื้นฐานการวิจัยด้านวิทยาการคอมพิวเตอร์
27	204433	สถาปัตยกรรมคอมพิวเตอร์
28	204451 หรือ 204452	การออกแบบและการวิเคราะห์อัลกอริทึมหรือทฤษฎีการคำนวณ
29	204497	สัมมนาทางวิทยาการคอมพิวเตอร์
30	204390 หรือ 204498	การฝึกงานคอมพิวเตอร์ หรือ สหกิจศึกษา
31	204399	การค้นคว้าอิสระ 1
32	204499	การค้นคว้าอิสระ 2

สมมติฐานของงานวิจัยนี้ คือ เกรดเฉลี่ยสะสมทั้งหมดเพียงปัจจัยเดียวไม่สามารถวิเคราะห์สถานภาพของนักศึกษาในอนาคตได้อย่างเหมาะสม เนื่องจากมีปัจจัยอื่นที่เกี่ยวข้อง จึงมีการศึกษาเอกสารแล้วพบว่า มี 3 เงื่อนไขหลักคือ เกรดเฉลี่ยสะสม หน่วยกิต และรายวิชาบังคับ แต่เนื่องจากรายวิชาบังคับเป็นเงื่อนไขที่ไม่มีการเปลี่ยนแปลงในแต่ละภาคเรียน เมื่อเรียนผ่านแล้วจะไม่ลงทะเบียนเรียนรายวิชานั้นซ้ำอีก นอกจากได้เกรดเอฟ (F) หรือต้องการรีเกรด แต่เป็นส่วนน้อย จึงตัดเงื่อนไขรายวิชาบังคับออกจากการพิจารณาเป็นข้อมูลอินพุต ดังนั้นจึงใช้เกรดเฉลี่ยสะสมและหน่วยกิตสะสมเป็นข้อมูลที่ใช้ในการทำนาย

เมื่อพิจารณาเฉพาะประโยชน์ของการใช้ข้อมูลอินพุตระหว่างเกรดเฉลี่ยสะสมกับหน่วยกิตสะสม กรณีสถานภาพสำเร็จการศึกษา เกรดเฉลี่ยสะสมกับหน่วยกิตมีความสำคัญในระดับที่เท่ากัน เช่น เมื่อเรียนอยู่ในภาคเรียนที่ 11 หรือชั้นปีที่ 4 ภาคเรียนที่ 2 เกรดเฉลี่ยสะสมของกลุ่มวิชาแกนที่ได้เท่ากับ 2.50 แต่หน่วยกิตสะสมของกลุ่มวิชาแกน

น้อยกว่า 24 หน่วยกิต จึงทำให้ไม่สามารถสำเร็จการศึกษาได้ ในลักษณะเดียวกัน ถ้าเกรดเฉลี่ยสะสมของกลุ่มวิชาแกนน้อยกว่า 2.00 แต่มีหน่วยกิตสะสมของกลุ่มวิชาแกนมากกว่า 24 หน่วยกิต ก็ไม่สามารถสำเร็จการศึกษาได้ กรณีสถานะพ้นสภาพ เมื่อพิจารณาในช่วงระยะเวลาการศึกษาสูงสุด 8 ปี สถานภาพนี้จะขึ้นอยู่กับเกรดเฉลี่ยสะสมทั้งหมดเพียงอย่างเดียว โดยที่ต้องมีเงื่อนไขเกรดเฉลี่ยสะสมตามที่ได้กล่าวไว้ในข้อที่ 1.2 นอกจากนี้เกรดเฉลี่ยสะสมยังสามารถบ่งบอกถึงระดับประสิทธิภาพการศึกษาของนักศึกษาแต่ละคนได้ จึงมีการเลือกใช้ข้อมูลอินพุตเป็นเกรดเฉลี่ยสะสม

เมื่อใช้ข้อมูลอินพุตเป็นเกรดเฉลี่ยสะสมแล้ว และมีสมมติฐานว่า ไม่ได้มีแค่เกรดเฉลี่ยสะสมทั้งหมดเพียงอย่างเดียวที่เป็นปัจจัยของการสำเร็จการศึกษา จึงมีการแยกย่อยเกรดเฉลี่ยสะสมทั้งหมดเป็นเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาตามโครงสร้างหลักสูตร แต่เมื่อนำข้อมูลหน่วยกิตสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชามารวมกัน และคะแนนสะสมที่ได้จากการคูณกันของหน่วยกิตกับเกรดของแต่ละหมวดหมู่หรือกลุ่มวิชามารวมกัน จากนั้นนำมาหารกัน จะได้เป็นเกรดเฉลี่ยสะสมทั้งหมด โดยที่แนวคิดในการแยกย่อยเกรดเฉลี่ยสะสมทั้งหมดเป็นแต่ละหมวดหมู่หรือกลุ่มวิชาแทนมาจากเงื่อนไขการสำเร็จการศึกษาที่ขึ้นอยู่กับเกรดเฉลี่ยสะสมของกลุ่มวิชาเอกด้วย และเงื่อนไขที่เกี่ยวกับหน่วยกิตสะสมที่มีการแบ่งตามโครงสร้างหลักสูตร จึงนำมาประยุกต์รวมกัน ได้ข้อมูลและตัวแปรอินพุตเป็นเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาตามโครงสร้างหลักสูตร และจากบทก่อนหน้านี้นี้มีงานวิจัยที่ใช้เกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาเป็นตัวแปรอินพุตด้วย

3.2.2 เตรียมข้อมูลด้วยการคัดเลือกข้อมูลนักศึกษา คัดเลือกข้อมูลลงทะเบียนเรียน คัดค่าว่างเติมค่าว่างด้วยศูนย์ และจัดข้อมูลให้มีรูปแบบเดียวกัน ซึ่งข้อมูลที่นำมาใช้ในงานวิจัยเป็นข้อมูลนักศึกษาที่เข้าเรียนตั้งแต่ปีการศึกษา 2548 ถึง 2551 ในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ เนื่องจากอยู่ในโครงสร้างหลักสูตรเดียวกัน นักศึกษาต้องมีสถานภาพสำเร็จการศึกษา หรือพ้นสภาพเนื่องจากผลการเรียนต่ำกว่าเกณฑ์เท่านั้น

ข้อมูลที่ใช้เป็นข้อมูลผลการเรียนของนักศึกษาดังแต่ชั้นปีที่ 1 ถึงชั้นปีที่ 2 โดยแบ่งเป็น 2 ชุดข้อมูล คือ ข้อมูลผลการเรียนภาคเรียนที่ 1 ถึงภาคเรียนที่ 4 หรือชั้นปีที่ 2 ภาคเรียนที่ 1 ทั้งหมด 323 คน และข้อมูลผลการเรียนภาคเรียนที่ 1 ถึงภาคเรียนที่ 5 หรือชั้นปีที่ 2 ภาคเรียนที่ 2 ทั้งหมด 319 คน เนื่องจากนักศึกษาบางส่วนได้พ้นสภาพในภาคเรียนที่ 4

จึงทำให้จำนวนนักศึกษาในภาคเรียนที่ 5 ลดลง ส่วนข้อมูลลงทะเบียนเรียนจะเลือกใช้รายวิชาที่ได้รับเกรด A, B+, B, C+, C, D+, D และ F เท่านั้น เนื่องจากรายวิชาที่มีเกรดเหล่านี้เป็นรายวิชาที่ใช้ในการคิดเกรดเฉลี่ยสะสม

จากตารางที่ 3.3 เป็นสรุปจำนวนนักศึกษาที่ลงทะเบียนเรียนของแต่ละหมวดหมู่หรือกลุ่มวิชาที่แบ่งจากหมวดหมู่วิชา จนแยกย่อยมาถึงกลุ่มวิชา ในแต่ละภาคเรียน พบว่าบางหมวดหมู่หรือกลุ่มวิชาในภาคเรียนที่ 3 หรือภาคเรียนฤดูร้อนไม่มีการลงทะเบียนเรียน จึงต้องการทดลองเปรียบเทียบระหว่างมีภาคเรียนฤดูร้อนกับไม่มีภาคเรียนฤดูร้อน และเปรียบเทียบระหว่างเมื่อเรียนถึงภาคเรียนที่ 4 กับภาคเรียนที่ 5 เพื่อหาภาคเรียนที่เหมาะสมต่อการทำนายการสำเร็จการศึกษามากที่สุด

ตารางที่ 3.3 จำนวนนักศึกษาที่ลงทะเบียนเรียนในแต่ละภาคเรียนตามหมวดหมู่หรือกลุ่มวิชา

หมวดหมู่หรือกลุ่มวิชา	ภาคเรียน										
	ชั้นปีที่ 1			ชั้นปีที่ 2			ชั้นปีที่ 3			ชั้นปีที่ 4	
	1	2	3	4	5	6	7	8	9	10	11
หมวดหมู่วิชาศึกษาทั่วไป	323	323	206	311	307	158	72	31	1	21	29
กลุ่มวิชาสังคมศาสตร์	1	134	108	91	114	85	29	15	0	7	19
กลุ่มวิชามนุษยศาสตร์	1	116	131	121	85	80	36	9	1	11	8
กลุ่มวิชาภาษาและสื่อสาร	318	312	13	285	295	11	15	10	0	3	3
กลุ่มวิชาวิทย์และเทคโนโลยี	322	320	0	3	0	0	0	1	0	0	0
หมวดหมู่วิชาเฉพาะ	323	320	177	319	315	139	310	309	9	307	227
กลุ่มวิชาแกน	323	319	159	293	23	65	40	21	9	26	16
กลุ่มวิชาชีววิทยา	317	6	0	1	0	0	0	0	0	0	0
กลุ่มวิชาเคมี	321	1	8	0	0	14	1	0	2	1	0
กลุ่มวิชาคณิตศาสตร์	277	251	121	277	12	26	29	3	5	10	3
กลุ่มวิชาฟิสิกส์	6	314	56	3	3	20	0	0	2	0	5
กลุ่มวิชาสถิติ	0	0	0	229	11	10	16	18	1	16	8
กลุ่มวิชาเอก	312	300	22	318	315	4	309	309	0	306	210
วิชาเอกบังคับ	312	300	22	317	314	0	309	309	0	83	108
วิชาเอกเลือก	0	6	0	7	81	4	261	285	0	297	189
วิชาโท	1	3	15	48	151	87	208	185	1	179	61
หมวดหมู่วิชาเลือกเสรี	3	36	68	103	169	155	92	115	3	102	135

กรณีที่ใช้ข้อมูลถึงภาคเรียนที่ 4 ข้อมูลที่ใช้ในการทำนายเป็นข้อมูลเกรดเฉลี่ยสะสมของภาคเรียนที่ 1, 2, 3, และ 4 เมื่อพิจารณาจำนวนนักศึกษาที่มีการลงทะเบียนในแต่ละหมวดหมู่หรือกลุ่มวิชาตั้งแต่ภาคเรียนที่ 5 จนถึงภาคเรียนที่ 11 พบว่า ยังมีการลงทะเบียนเพิ่มเติมจากภาคเรียนที่ 4 อีก ยกเว้นกลุ่มวิชาชีววิทยาที่ไม่มีนักศึกษาลงทะเบียนเพิ่ม และกลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ที่มีนักศึกษาลงทะเบียนเพิ่มเพียงหนึ่งคน แสดงให้เห็นว่า เมื่อใช้ข้อมูลถึงภาคเรียนที่ 4 ทำนายเกรดเฉลี่ยสะสมของภาคเรียนที่ 11 จะมีนักศึกษาบางส่วนที่มีเกรดเฉลี่ยสะสมไม่แตกต่างจากเดิม แต่ในหมวดหมู่หรือกลุ่มวิชาอื่น ก็มีนักศึกษาที่มีการลงทะเบียนเพิ่ม ทำให้เกรดเฉลี่ยสะสมเปลี่ยนแปลงไป จึงเป็นที่มาของการทำนายเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาในภาคเรียนที่ 11 หรือชั้นปีที่ 4 ภาคเรียนที่ 2 ซึ่งเป็นภาคเรียนที่จะมีสถานภาพจบตามหลักสูตรหรือจบเกินหลักสูตร และการใช้ข้อมูลถึงภาคเรียนที่ 5 ก็มีลักษณะที่คล้ายคลึงกัน

- 3.2.3 แบ่งชุดตัวแปรอินพุต เพื่อหาชุดตัวแปรอินพุตที่เหมาะสมกับข้อมูลในภาคเรียนต่างๆ โดยที่จะมีการแบ่งตามโครงสร้างหลักสูตร ดังภาพที่ 3.2 การแบ่งหมวดหมู่หรือกลุ่มวิชาเป็นการแยกย่อยมาจากหมวดหมู่วิชาใหญ่ 3 กลุ่มคือ หมวดหมู่วิชาศึกษาทั่วไป หมวดหมู่วิชาเฉพาะ และหมวดหมู่เลือกเสรี หมวดหมู่วิชาศึกษาทั่วไปแบ่งเป็น 4 กลุ่มย่อยคือ กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาภาษาและการสื่อสาร และกลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ หมวดหมู่วิชาเฉพาะแบ่งเป็น 3 กลุ่มย่อยคือ กลุ่มวิชาแกน กลุ่มวิชาเอก และกลุ่มวิชาโท

กลุ่มวิชาแกนเป็นกลุ่มวิชาที่มีรายวิชาหลากหลายสาขาวิชา และมีการลงทะเบียนมากในชั้นปีที่ 1 กับ 2 จึงทำการแยกย่อยออกเป็นกลุ่มตามสาขาวิชาได้ 5 กลุ่ม คือ กลุ่มวิชาชีววิทยา กลุ่มวิชาเคมี กลุ่มวิชาคณิตศาสตร์ กลุ่มวิชาฟิสิกส์ และกลุ่มวิชาสถิติ กลุ่มวิชาเอกแบ่งเป็น 2 กลุ่ม คือ กลุ่มวิชาเอกบังคับ และกลุ่มวิชาเอกเลือก แต่เนื่องจากกลุ่มวิชาเอกเลือกยังไม่มี การลงทะเบียนมากในชั้นปีที่ 1 กับ 2 จึงไม่พิจารณาแยกกลุ่มวิชาเอก

ปัญหาหลักของการแบ่งแยกย่อยหมวดหมู่หรือกลุ่มวิชาคือ ข้อมูลการลงทะเบียนในแต่ละภาคเรียนที่ไม่สม่ำเสมอ ตัวอย่างเช่น กลุ่มวิชาสังคมศาสตร์ที่มีการลงทะเบียนในภาคเรียนที่ 1 แค่หนึ่งคน ถ้าหากเลือกเป็นตัวแปรอินพุต ข้อมูลจะไม่เพียงพอต่อการเรียนรู้ จึงต้องมีการรวมหมวดหมู่หรือกลุ่มวิชา ข้อมูลที่ได้จากการรวมจะมีเพิ่มขึ้น ทำให้มีข้อมูลที่เพียงพอ และมีข้อมูลการลงทะเบียนที่เหมาะสมกับทุกภาคเรียนที่ใช้เป็นตัวแปรอินพุต ได้ชุดตัวแปรอินพุตมีทั้งหมด 26 ชุด ดังตารางที่ 3.4



ภาพที่ 3.2 โครงสร้างหมวดหมู่และกลุ่มวิชาย่อยที่ใช้เป็นตัวแปรอินพุต

3.2.3 คัดเลือกตัวแปรอินพุตและตัวแปรเอาต์พุต การคัดเลือกตัวแปรอินพุตมาจากการแยกย่อยและรวมหมวดหมู่หรือกลุ่มวิชาได้ชุดตัวแปรอินพุตทั้งหมด 26 ชุด จากนั้นนำหมวดหมู่หรือกลุ่มวิชาในแต่ละชุดตัวแปรอินพุต มาเป็นตัวแปรอินพุตมีทั้งหมด 29 ตัวแปร ดังตารางที่ 3.5

ตัวอย่างเช่น ชุดตัวแปรอินพุตที่ 1 แบ่งเป็น 3 หมวดหมู่คือ หมวดหมู่ศึกษาทั่วไป หมวดหมู่เฉพาะ และหมวดหมู่เลือกเสรี ได้ตัวแปรอินพุต 3 ตัวแปร คือ ตัวแปรหมวดหมู่ศึกษาทั่วไป ตัวแปรหมวดหมู่เฉพาะ และตัวแปรหมวดหมู่เลือกเสรี ชุดตัวแปรอินพุตที่ 2 แบ่งเป็น 5 หมวดหมู่หรือกลุ่มวิชา คือ หมวดหมู่ศึกษาทั่วไป กลุ่มวิชาแกน กลุ่มวิชาเอก กลุ่มวิชาโท และหมวดหมู่เลือกเสรี จะได้ตัวแปรอินพุตเพิ่มอีก 3 ตัวแปร คือ ตัวแปรกลุ่มวิชาแกน ตัวแปรกลุ่มวิชาเอก และตัวแปรกลุ่มวิชาโท สรุปจากตัวอย่าง 2 ชุดตัวแปรอินพุตได้ตัวแปรอินพุตทั้งหมด 6 ตัวแปร คือ ตัวแปรหมวดหมู่ศึกษาทั่วไป ตัวแปรหมวดหมู่เฉพาะ ตัวแปรหมวดหมู่เลือกเสรี ตัวแปรกลุ่มวิชาแกน ตัวแปรกลุ่มวิชาเอก และตัวแปรกลุ่มวิชาโท

ตารางที่ 3.4 ตัวแปรอินพุตที่อยู่ในแต่ละชุดตัวแปรอินพุต

ชุดตัวแปร อินพุต	หมวดหมู่วิชาศึกษาทั่วไป				หมวดหมู่วิชาเฉพาะ									หมวดหมู่ วิชาเลือก เสรี
	กลุ่มวิชา สังคมศาสตร์	กลุ่มวิชา มนุษยศาสตร์	กลุ่ม ภาษา สื่อสาร	กลุ่มวิชา วิทย์และ คณิต	กลุ่มวิชาแกน					กลุ่มวิชาเอก		วิชาโท		
					กลุ่มวิชา ชีววิทยา	กลุ่มวิชา เคมี	กลุ่มวิชา คณิต	กลุ่มวิชา ฟิสิกส์	กลุ่มวิชา สถิติ	วิชาเอก บังคับ	วิชาเอก เลือก			
1	1				2									3
2	1				2						3		4	5
3	1	2	3	4	5									6
4	1	2	3	4	5					6		7	8	
5	1		2	3	4					5		6		
6	1		2	3	4					5		1		
7	1				2					3	4	5	6	
8	1	2	3	4	5					6	7	8	9	
9	1	2	3	4	5	6	7	8	9	10	11	12	13	
10	1				2					3		4		
11	1				2		3			4		5		
12	1				2					3		1		
13	1				2		3			4		1		
14	1		2	3						4		1		

ตารางที่ 3.4 ตัวแปรอินพุตที่อยู่ในแต่ละชุดตัวแปรอินพุต(ต่อ)

ชุดตัวแปร อินพุต	หมวดหมู่วิชาศึกษาทั่วไป				หมวดหมู่วิชาเฉพาะ								หมวดหมู่ วิชาเลือก เสรี
					กลุ่มวิชาแกน					กลุ่มวิชาเอก		วิชาโท	
	กลุ่มวิชา สังคมศาสตร์	กลุ่มวิชา มนุษยศาสตร์	กลุ่ม ภาษา สื่อสาร	กลุ่มวิชา วิทย์และ คณิต									
15	1		2	3	4					5		1	
16	1		2	3	4					5		6	7
17	1		2	3	4					5		6	1
18	1					2			3	4		1	
19	1					2				3		1	
20	1					2			1	3		1	
21	1		2	3	4		5	6		7		1	
22	1		2	3	4		5	6	5	7		1	
23	1				2		3	4	5	6		1	
24	1				2		3	4		5		1	
25	1					2	3			4		1	
26	1					2	3	4	1	5		1	

ตัวแปรอินพุตเป็นเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาในแต่ละภาคเรียน โดยจะมีการเปรียบเทียบชุดตัวแปรอินพุตที่มีข้อมูลถึงภาคเรียนที่ 4 กับชุดตัวแปรอินพุตที่มีข้อมูลถึงภาคเรียนที่ 5 เพื่อหาภาคเรียนที่เหมาะสมกับการทำนายเกรดเฉลี่ยสะสมของภาคเรียนที่ 11 มากที่สุด ตัวอย่างเช่น ตัวแปรหมวดหมู่วิชาศึกษาทั่วไปที่มีข้อมูลถึงภาคเรียนที่ 4 คือ เกรดเฉลี่ยสะสมของหมวดหมู่วิชาศึกษาทั่วไปภาคเรียนที่ 1 ภาคเรียนที่ 2 ภาคเรียนที่ 3 และภาคเรียนที่ 4 ส่วนข้อมูลถึงภาคเรียนที่ 5 คือ การเพิ่มภาคเรียนที่ 5 เข้าไปในชุดตัวแปรอินพุต และมีการเปรียบเทียบการมีภาคเรียนถูกรื้อถอนกับไม่มีภาคเรียนถูกรื้อถอนสามารถดูตัวอย่างข้อมูลนำเข้า และตัวแปรอินพุตได้ในภาคผนวก ก

ตารางที่ 3.5 หมวดหมู่หรือกลุ่มวิชาที่มาจากแต่ละชุดตัวแปรอินพุต

ลำดับ	หมวดหมู่หรือกลุ่มวิชา
1	หมวดวิชาศึกษาทั่วไป
2	กลุ่มวิชาสังคมศาสตร์ และกลุ่มวิชามนุษยศาสตร์
3	กลุ่มวิชาสังคมศาสตร์
4	กลุ่มวิชามนุษยศาสตร์
5	กลุ่มวิชาภาษาและการสื่อสาร
6	กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์
7	กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ และหมวดหมู่วิชาเลือกเสรี
8	หมวดวิชาศึกษาทั่วไป กลุ่มวิชาโท และหมวดหมู่วิชาเลือกเสรี
9	กลุ่มวิชาสังคมศาสตร์ กลุ่มวิชามนุษยศาสตร์ กลุ่มวิชาโท และหมวดหมู่เลือกเสรี
10	หมวดวิชาศึกษาทั่วไป กลุ่มวิชาชีววิทยา กลุ่มวิชาโท และหมวดหมู่เลือกเสรี
11	หมวดหมู่ศึกษาทั่วไป กลุ่มชีววิทยา กลุ่มวิชาสถิติ กลุ่มวิชาโท หมวดหมู่เลือกเสรี
12	กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ และกลุ่มวิชาแกน
13	หมวดวิชาเฉพาะ
14	กลุ่มวิชาแกน
15	กลุ่มวิชาชีววิทยา
16	กลุ่มวิชาเคมี
17	กลุ่มวิชาคณิตศาสตร์
18	กลุ่มวิชาฟิสิกส์
19	กลุ่มวิชาสถิติ
20	กลุ่มวิชาชีววิทยา และกลุ่มวิชาเคมี
21	กลุ่มวิชาคณิตศาสตร์ กลุ่มวิชาฟิสิกส์ และกลุ่มวิชาสถิติ
22	กลุ่มวิชาคณิตศาสตร์ และกลุ่มวิชาสถิติ

ตารางที่ 3.5 หมวดหมู่หรือกลุ่มวิชาที่มาจากแต่ละชุดตัวแปรอินพุต (ต่อ)

ลำดับ	หมวดหมู่หรือกลุ่มวิชา
23	กลุ่มวิชาฟิสิกส์ และกลุ่มวิชาสถิติ
24	กลุ่มวิชาเคมี กลุ่มวิชาคณิตศาสตร์ และกลุ่มวิชาฟิสิกส์
25	กลุ่มวิชาเคมี กลุ่มวิชาคณิตศาสตร์ กลุ่มวิชาฟิสิกส์ และกลุ่มวิชาสถิติ
26	กลุ่มวิชาเอก
27	กลุ่มวิชาโท
28	หมวดหมู่วิชาเลือกเสรี
29	กลุ่มวิชาโท และหมวดหมู่วิชาเลือกเสรี

จากตารางที่ 3.6 แสดงค่าเฉลี่ยของเกรดเฉลี่ยสะสม ลำดับตัวแปรอินพุตอ้างอิงจากตารางที่ 3.5 และตัวเลขที่มีตัวหนา แสดงให้เห็นว่า หลังจากภาคเรียนที่ 4 หรือ 5 บางหมวดหมู่หรือกลุ่มวิชามีค่าเฉลี่ยของเกรดเฉลี่ยสะสมที่เปลี่ยนแปลงไปเมื่อเรียนอยู่ภาคเรียนที่ 11

ตารางที่ 3.6 ค่าเฉลี่ยของเกรดเฉลี่ยสะสมในแต่ละตัวแปรอินพุต

ลำดับ ตัวแปรอินพุต	ภาคเรียนที่										
	1	2	3	4	5	6	7	8	9	10	11
1	2.76	2.67	2.69	2.71	2.72	2.72	2.73	2.73	2.73	2.73	2.73
2	4.00	2.76	2.76	2.71	2.72	2.74	2.73	2.74	2.74	2.74	2.74
3	4.00	2.86	2.81	2.77	2.75	2.76	2.76	2.76	2.76	2.76	2.76
4	4.00	2.64	2.73	2.70	2.70	2.72	2.73	2.73	2.73	2.73	2.73
5	2.82	2.72	2.73	2.77	2.78	2.78	2.78	2.79	2.79	2.79	2.79
6	2.76	2.62	2.62	2.62	2.62	2.62	2.62	2.61	2.61	2.61	2.61
7	2.74	2.83	2.84	2.80	2.86	2.90	2.92	2.94	2.94	2.95	2.97
8	2.76	2.69	2.72	2.75	2.80	2.83	2.82	2.84	2.84	2.84	2.85
9	3.06	2.84	2.84	2.81	2.87	2.91	2.89	2.91	2.91	2.90	2.92
10	2.48	2.55	2.60	2.65	2.72	2.75	2.75	2.77	2.77	2.78	2.79
11	2.48	2.55	2.60	2.63	2.70	2.73	2.73	2.75	2.75	2.76	2.77
12	2.23	2.19	2.19	2.22	2.21	2.22	2.22	2.22	2.22	2.21	2.21
13	2.13	2.10	2.12	2.15	2.22	2.24	2.29	2.37	2.37	2.42	2.45
14	2.09	2.04	2.05	2.11	2.11	2.12	2.12	2.12	2.12	2.12	2.11
15	2.06	2.06	2.06	2.05	2.05	2.05	2.05	2.05	2.05	2.05	2.05
16	2.09	2.09	2.10	2.10	2.10	2.13	2.13	2.13	2.13	2.13	2.13
17	2.34	1.99	2.05	2.09	2.09	2.12	2.13	2.13	2.13	2.13	2.13

ตารางที่ 3.6 ค่าเฉลี่ยของเกรดเฉลี่ยสะสมในแต่ละตัวแปรอินพุต (ต่อ)

ลำดับตัวแปรอินพุต	ภาคเรียนที่										
	1	2	3	4	5	6	7	8	9	10	11
18	2.19	2.12	2.12	2.11	2.12	2.10	2.10	2.10	2.10	2.10	2.11
19	0.00	0.00	0.00	2.56	2.54	2.55	2.51	2.46	2.46	2.40	2.39
20	2.07	2.07	2.07	2.07	2.07	2.09	2.09	2.09	2.09	2.09	2.09
21	2.34	2.01	2.05	2.13	2.13	2.14	2.14	2.14	2.14	2.14	2.14
22	2.34	1.99	2.05	2.13	2.13	2.16	2.17	2.17	2.17	2.17	2.16
23	2.19	2.12	2.12	2.20	2.19	2.19	2.19	2.19	2.19	2.18	2.17
24	2.10	2.03	2.05	2.09	2.09	2.10	2.11	2.11	2.11	2.11	2.11
25	2.10	2.03	2.05	2.12	2.12	2.13	2.14	2.13	2.13	2.13	2.13
26	2.39	2.32	2.36	2.20	2.26	2.26	2.31	2.41	2.41	2.47	2.51
27	4.00	3.67	3.14	3.01	3.00	3.01	2.85	2.87	2.87	2.86	2.86
28	2.60	3.51	3.48	3.38	3.39	3.36	3.36	3.36	3.36	3.31	3.30
29	2.95	3.53	3.41	3.26	3.20	3.21	3.08	3.08	3.08	3.05	3.06

การคัดเลือกตัวแปรเอาต์พุตจะมีรูปแบบการแบ่งกลุ่มเอาต์พุตตามตารางที่ 3.7 ซึ่งตัวเลขในตารางแทนลำดับกลุ่มที่เป็นข้อมูลนำเข้าของตัวแปรเอาต์พุต มีการแบ่งดังนี้

- 1) การแบ่งจากสถานภาพตามเกณฑ์หลักสูตร 4 ปี คือ แบบที่ 1 แบ่ง 3 กลุ่ม มี กลุ่มพื้นฐาน กลุ่มจบเกินหลักสูตร และกลุ่มจบตามหลักสูตร
- 2) การแบ่งจากสถานภาพกับชั้นปีที่สำเร็จการศึกษา โดยจะนับภาคเรียนที่ 1, 2, และ 3 รวมกันเป็น 1 ชั้นปี มี 5 รูปแบบ ดังนี้
 - 2.1) แบบที่ 2 แบ่ง 3 กลุ่ม คือ กลุ่มพื้นฐาน กลุ่มจบปี 4 และกลุ่มจบมากกว่าปี 4
 - 2.2) แบบที่ 3 แบ่ง 6 กลุ่มคือ กลุ่มพื้นฐาน กลุ่มจบปี 4 กลุ่มจบปี 5 กลุ่มจบปี 6 กลุ่มจบปี 7 และกลุ่มจบปี 8
 - 2.3) แบบที่ 4 แบ่ง 6 กลุ่ม คือ กลุ่มพื้นฐาน กลุ่มจบปี 8 กลุ่มจบปี 7 กลุ่มจบปี 6 กลุ่มจบปี 5 และกลุ่มจบปี 4
 - 2.4) แบบที่ 5 แบ่ง 4 กลุ่ม คือ กลุ่มพื้นฐาน กลุ่มจบปี 4 กลุ่มจบปี 5 และกลุ่มจบมากกว่าปี 5
 - 2.5) แบบที่ 6 แบ่ง 4 กลุ่มคือ กลุ่มพื้นฐาน กลุ่มจบมากกว่าปี 5 กลุ่มจบปี 5 และกลุ่มจบปี 4

ตารางที่ 3.7 การแบ่งตัวแปรเอาต์พุตตามชั้นปีที่สามารถศึกษาและฝึกสภาพ

แบบที่	พื้นสภาพ	จบปี 4		จบปี 5	จบปี 6	จบปี 7	จบปี 8	
		ภาคเรียน 2	ภาคเรียน 3					
1	1	3	2					
2	1	3			2			
3	1	2			3	4	5	6
4	1	6			5	4	3	2
5	1	2			3	4		
6	1	4			3	2		

การแบ่งกลุ่มเอาต์พุตของแบบที่ 1 กับแบบที่ 2 แตกต่างกันที่กลุ่มจบตามหลักสูตร คือ การจบเมื่ออยู่ภาคเรียนที่ 11 หรือชั้นปีที่ 4 ภาคเรียนที่ 2 ไม่ได้รวมนักศึกษาที่จบชั้นปีที่ 4 ภาคเรียนที่ 3 ซึ่งนักศึกษาที่จบภาคเรียนนี้จะอยู่กลุ่มจบเกินหลักสูตร และจากตารางที่ 3.8 มีจำนวนนักศึกษาที่จบในชั้นปีที่ 4 ภาคเรียนที่ 3 ทั้งหมด 67 คน และมีนักศึกษา 54 คน ที่ลงทะเบียนเฉพาะรายวิชา การค้นคว้าอิสระ 2 รหัสวิชา 204499 สำหรับนักศึกษาแผน 1 หรือรายวิชา สหกิจศึกษา รหัสวิชา 204489 สำหรับนักศึกษาแผน 2 แสดงว่า นักศึกษาที่จบชั้นปีที่ 4 ภาคเรียนที่ 2 และจบชั้นปีที่ 4 ภาคเรียนที่ 3 ส่วนมากไม่มีความแตกต่างของการลงทะเบียนหมวดหมู่หรือกลุ่มวิชาอื่นนอกจากรายวิชาดังกล่าว ซึ่งไม่น่ามาคิดเกรดเฉลี่ยสะสม จึงมีการแบ่งตัวแปรเอาต์พุตเพิ่มอีก 1 แบบ คือ กลุ่มฝึกสภาพ กลุ่มจบมากกว่าปี 4 และกลุ่มจบปี 4

การแบ่งกลุ่มเอาต์พุตของแบบที่ 3 กับ 4 แตกต่างกันที่ข้อมูลนำเข้าของตัวแปรเอาต์พุตแบบที่ 3 จะเรียงตามชั้นปีที่จบจากน้อยไปหามาก ส่วนแบบที่ 4 จะเรียงตามชั้นปีที่จบจากมากไปหาน้อย การแบ่งกลุ่มเอาต์พุตของแบบที่ 5 กับ 6 ก็แตกต่างในลักษณะเดียวกัน

เนื่องจากแบบที่ 3 หรือ 4 มีจำนวนนักศึกษาที่จบการศึกษาในชั้นปีที่ 6, 7, และ 8 น้อย ดังตารางที่ 3.8 ไม่เพียงพอต่อการเรียนรู้และทดสอบ จึงมีการรวมชั้นปีที่ 6, 7, และ 8 เข้าด้วยกัน เพิ่มรูปแบบเอาต์พุตเป็น แบบที่ 5 หรือ 6

ตารางที่ 3.8 จำนวนนักศึกษาในแต่ละสถานภาพ

สถานภาพ	ก่อนปี 4	ปี 4		ปี 5	ปี 6	ปี 7	ปี 8	รวม
		ภาคเรียนที่ 2	ภาคเรียนที่ 3					
ฝึกสภาพ	13	1	0	5	4	3	1	27
จบการศึกษา	0	136	67	69	14	5	5	296

3.2.5 การแบ่งข้อมูลเรียนรู้และทดสอบ เทคนิคที่ใช้ในการแบ่งข้อมูลมี 2 แบบคือ การแบ่งข้อมูลจากการสุ่มแบบร้อยละ (Hold out) กำหนดจำนวนข้อมูลทดสอบเป็นร้อยละแล้ว จะทำการสุ่มตามร้อยละที่กำหนด สัดส่วนข้อมูลเรียนรู้และข้อมูลทดสอบสำหรับเทคนิคแบบสุ่มร้อยละคือ 80 ต่อ 20 และ 90 ต่อ 10 และการแบ่งข้อมูลแบบ k กลุ่ม (K-fold cross validation) คือการแบ่งข้อมูลออกเป็น k ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นข้อมูลทดสอบ ที่เหลือเป็นข้อมูลเรียนรู้ ทำวนซ้ำต่อไปจนครบ k รอบ ในงานวิจัยนี้มีการแบ่งข้อมูลเป็น 5 กลุ่มที่มีข้อมูลเรียนรู้ 4 กลุ่ม ข้อมูลทดสอบ 1 กลุ่ม กับ 10 กลุ่มที่มีข้อมูลเรียนรู้ 9 กลุ่ม ข้อมูลทดสอบ 1 กลุ่ม เพื่อเปรียบเทียบหาเทคนิคและสัดส่วนการแบ่งข้อมูลที่เหมาะสมกับชุดข้อมูล

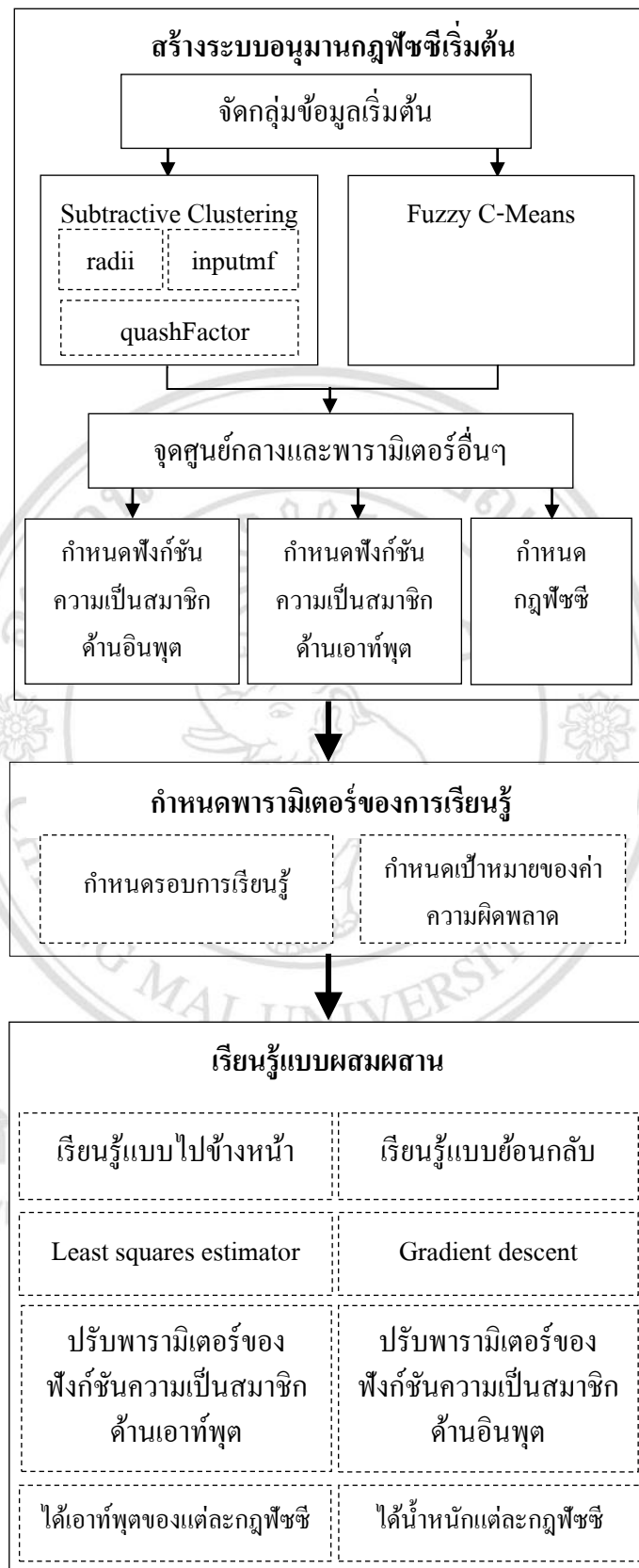
3.3 ขั้นตอนสร้างแบบจำลองเพื่อทำนายเกรดเฉลี่ยสะสม

ขั้นตอนการสร้างแบบจำลองทำนายเกรดเฉลี่ยสะสมด้วยระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ ประกอบด้วย 3 ขั้นตอนคือ ขั้นตอนการสร้างระบบอนุมานกฎฟัซซี ขั้นตอนการเรียนรู้แบบผสมผสาน และขั้นตอนการวัดประสิทธิภาพของแบบจำลอง มีผังแผนภาพที่ 3.3

3.3.1 การสร้างระบบอนุมานกฎฟัซซี เป็นสร้างกฎฟัซซีเริ่มต้น โดยจะมีการปรับพารามิเตอร์การจัดกลุ่ม เพื่อให้ค่าความคลาดเคลื่อนกำลังสองน้อยสุดมีค่าน้อยที่สุด

เริ่มจากการจัดกลุ่มก่อนการเรียนรู้ เพื่อกำหนดจำนวนกฎฟัซซี ค่าจุดศูนย์กลางเริ่มต้น และพารามิเตอร์ตัวอื่นที่เกี่ยวข้องกับฟังก์ชันความเป็นสมาชิกด้านอินพุต งานวิจัยนี้เลือกใช้เทคนิคการจัดกลุ่มแบบลบบอก และฟัซซีซิมีน เนื่องจากข้อมูลที่นำเข้าเป็นข้อมูลเกรดเฉลี่ยที่ไม่มีสามารถจัดกลุ่มข้อมูลได้ชัดเจน จึงต้องนำเทคนิคที่ใช้ระบบฟัซซีเข้ามาช่วยแก้ไขปัญหา จากนั้นเปรียบเทียบประสิทธิภาพของเทคนิคด้วยค่า RMSE ต่อมาจะปรับพารามิเตอร์รัศมีระหว่างกลุ่ม (radii) พารามิเตอร์ค่าที่ใช้ลดปัญหาระยะขอบของกลุ่มที่อยู่ใกล้กัน (qushFactor) และฟังก์ชันความเป็นสมาชิกด้านอินพุต (inputmf)

ค่าพารามิเตอร์ radii ที่ใช้ในการปรับมี 4 ค่า คือ 0.3, 0.4, 0.5, และ 0.6 โดยที่ค่า 0.5 เป็นค่าเริ่มต้น ส่วนพารามิเตอร์ inputmf ที่เลือกใช้มีทั้งหมด 7 ชนิด ดังนี้ ฟังก์ชันเกาส์เซียน (gaussmf) ฟังก์ชันเกาส์เซียนแบบรวมกัน (gauss2mf) ฟังก์ชันระฆังคว่ำ (gbellmf) ฟังก์ชันซิกมอยด์ (sigmf) ฟังก์ชันความแตกต่างระหว่างสองฟังก์ชันซิกมอยด์ (dsigmf) ฟังก์ชันการคูณกันระหว่างสองฟังก์ชันซิกมอยด์ (psigmf) ฟังก์ชันตัวเอส (smf) และค่าพารามิเตอร์ qushFactor มี 3 ค่า คือ 1.00, 1.25, และ 1.50 ค่า 1.25 เป็นค่าเริ่มต้น



ภาพที่ 3.3 ขั้นตอนสร้างแบบจำลองทำนายเกรดเฉลี่ยสะสมด้วยระบบอนุมานนิวโรฟัซซี่แบบปรับตัวได้

3.3.2 การสร้างแบบจำลองด้วยระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้ กฎฟัซซีที่ได้จะแบ่งเป็นส่วนอินพุตและส่วนเอาต์พุต ซึ่งจะต้องมีการเรียนรู้แบบไปข้างหน้าและแบบย้อนกลับเพื่อปรับพารามิเตอร์ การเรียนรู้ของระบบอนุมานนิวโรฟัซซีแบบปรับตัวได้มีพารามิเตอร์ที่เกี่ยวข้องคือ จำนวนรอบการเรียนรู้ และค่าเป้าหมายของความผิดพลาด

3.3.3 การวัดค่าความคลาดเคลื่อน เป็นการวัดประสิทธิภาพของแบบจำลองที่ได้ ถ้ามีความคลาดเคลื่อนน้อยแสดงว่าเหมาะสมกับข้อมูลที่นำเข้า เทคนิคที่เลือกใช้เรียกว่า ความคลาดเคลื่อนกำลังสองน้อยที่สุด สมการที่ (3.1)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (T_n - O_n)^2}{\sum_{i=1}^n T_n}} \quad (3.1)$$

เมื่อ T_n เป็นเอาต์พุตเป้าหมายที่ต้องการ

O_n เป็นเอาต์พุตที่ได้จากแบบจำลอง

3.4 ขั้นตอนการสร้างแบบจำลองเพื่อทำนายความน่าจะเป็น

3.4.1 สร้างแบบจำลองทำนายความน่าจะเป็นจะใช้เทคนิค 4 เทคนิคดังนี้ โครงข่ายการรู้จำแบบการจำแนกเบย์อย่างง่าย ต้นไม้ตัดสินใจ และการถดถอยโลจิสติกพหุกลุ่ม เพื่อทำนายความน่าจะเป็นของการสำเร็จการศึกษา จากนั้นจะมีการเปรียบเทียบเทคนิคที่มีการทำนายได้ถูกต้องมากที่สุด โดยมีตัวแปรอินพุตเป็นเกรดเฉลี่ยสะสมของแต่ละหมวดหมู่หรือกลุ่มวิชาในภาคเรียนที่ 11 หรือชั้นปีที่ 4 ภาคเรียนที่ 2

3.4.2 การวัดประสิทธิภาพของแบบจำลองที่ได้จะวัดจากความคลาดเคลื่อนที่หาได้จากค่าความคลาดเคลื่อนกำลังสองน้อยที่สุด ดังสมการที่ (3.1) และความแม่นยำ ด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความเหวี่ยง (F-measure) ที่หาได้จากตารางคอนฟิวชันเมตริกซ์ (Confusion matrix) มีรูปแบบดังตารางที่ 3.9 โดยจะเลือกเอาต์พุตที่มีค่าความน่าจะเป็นมากที่สุด

ตารางที่ 3.9 ตารางคอนฟิวชันเมตริกซ์

Actual	Predicted	
	Positive	Negative
Positive	True positive	False negative
Negative	False positive	True negative

จากตารางคอนฟิวชันเมตริกซ์สามารถหาค่าต่างๆ ได้จากการนำค่าที่แท้จริง (Actual) และค่าทำนาย (Predicted) มาสรุปเพื่อหาจำนวนข้อมูลที่มีการจำแนกได้ถูกต้องและไม่ถูกต้อง โดยที่ค่าความถูกต้องเชิงบวก (True positive, TP) เป็นจำนวนข้อมูลที่ค่าความจริงอยู่กลุ่มนั้น และค่าทำนายสามารถจำแนกได้กลุ่มนั้น ค่าความถูกต้องเชิงลบ (True negative, TN) เป็นจำนวนข้อมูลที่ค่าความจริงไม่อยู่กลุ่มนั้น และค่าทำนายไม่สามารถจำแนกได้กลุ่มนั้น ค่าความไม่ถูกต้องเชิงบวก (False positive, FP) เป็นจำนวนข้อมูลที่ค่าความจริงไม่อยู่กลุ่มนั้น แต่ค่าทำนายสามารถจำแนกได้กลุ่มนั้น และค่าความไม่ถูกต้องเชิงลบ (False negative, FN) เป็นจำนวนข้อมูลที่ค่าความจริงอยู่กลุ่มนั้น และค่าทำนายไม่สามารถจำแนกได้กลุ่มนั้น ซึ่งค่าในตารางทั้งหมดสามารถนำมาใช้หาค่าต่างๆ ได้ ดังนี้

ค่าความถูกต้องเป็นการวัดความแม่นยำของแบบจำลองรวมกันทุกกลุ่ม ดังสมการที่ (3.4)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

ค่าความแม่นยำเป็นการวัดความแม่นยำของแบบจำลองที่พิจารณาแยกทีละกลุ่มของเอาต์พุตที่ทำนายได้ ดังสมการที่ (3.5)

$$Precision = \frac{TP}{TP + FP} \quad (3.5)$$

ค่าระลึกเป็นการวัดความถูกต้องของแบบจำลองที่พิจารณาแยกทีละกลุ่มเอาต์พุตจริง ดังสมการที่ (3.6)

$$Recall = \frac{TP}{TP + FN} \quad (3.6)$$

ค่าความเหวี่ยงเป็นการวัดค่าเฉลี่ยของความแม่นยำและค่าระลึกโดยพิจารณาแยกทีละกลุ่ม ดังสมการที่ (3.7)

$$F - measure = \frac{(1 + \beta)^2 \times Recall \times Precision}{\beta^2 \times Recall + Precision} \quad (3.7)$$

เมื่อ β คือค่าที่ใช้ควบคุมน้ำหนักให้กับค่าระลึกและค่าความแม่นยำ มีค่าตั้งแต่ 0 ถึง α

การวัดประสิทธิภาพของแบบจำลองที่มีเอาต์พุตที่ได้เป็นค่าความน่าจะเป็น สามารถหาได้จากค่าความคลาดเคลื่อนของเอาต์พุตที่มีค่าแบบไม่ต่อเนื่อง มีค่าเป็น 0 หรือ 1 กับ

ค่าความน่าจะเป็นที่รวมกันแล้วต้องเท่ากับ 1 เรียกเทคนิคการวัดประสิทธิภาพนี้ว่า คะแนนไบรเออ (Brier score) ซึ่งชุดผลลัพธ์ความเป็นไปได้สามารถเป็นได้ทั้งแบบ ไบนารี (Binary) หรือแบบกลุ่ม คิดค้นโดย Glenn W. Brier ในปี ค.ศ. 1950 คล้ายกับการหาค่าความคลาดเคลื่อนระหว่างค่าความน่าจะเป็นกับค่าเอาต์พุตที่เป็นตัวเลขแบบไม่ต่อเนื่อง ดังสมการที่ (3.8)

$$BS = \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^R (f_t - o_t)^2 \quad (3.8)$$

เมื่อ f_t คือค่าความน่าจะเป็นที่ได้จากการทำนาย

o_t คือค่าผลลัพธ์ที่ถูกแทนค่าเป็น 0 หรือ 1

N คือจำนวนข้อมูลทั้งหมด

R คือจำนวนกลุ่มเอาต์พุต เช่น $R=3$ เมื่อมีกลุ่มพันสภาพ จบตามหลักสูตร และ จบเกินหลักสูตร

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved