



## โครงการทางสถิติ 2

เรื่อง การจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ  
คณะวิทยาศาสตร์ โดยใช้วิธีการเรียนรู้แบบรวมกลุ่ม

*Classifying Dropout for Undergraduate Students in The Department of Statistics: Using  
Ensemble Method*

นำเสนอโดย

นายศิริพัฒน์ จานเชื้อ

รหัสประจำตัว 613020198-8

นางสาวยุวลักษณ์ ดวงมะลา

รหัสประจำตัว 613020196-2

อาจารย์ที่ปรึกษา

อาจารย์ ดร.จิตรจิรา ไชยฤทธิ์

โครงการนี้เป็นส่วนหนึ่งของรายวิชา SC 614775 โครงการทางสถิติ 2  
ปีการศึกษา 2564

## สารบัญเนื้อหา

บทที่ 1 .....	1
บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	2
สมมติฐานของการวิจัย .....	2
ขอบเขตของการวิจัย .....	2
ความหมายหรือนิยามคำศัพท์เฉพาะ.....	3
ประโยชน์ที่คาดว่าจะได้รับการวิจัย .....	3
ความเป็นไปได้ในการพัฒนาต่อยอด.....	3
บทที่ 2 .....	4
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	4
ทฤษฎีการพันสภาพ .....	5
การทำเหมืองข้อมูล.....	6
เทคนิคการแก้ปัญหาข้อมูลไม่สมดุลสำหรับการจำแนกประเภท (Imbalanced Data Problem Solving in Classification Technique).....	8
1. วิธีสุ่มเกิน.....	8
2. วิธีสุ่มลด .....	8
3. วิธีผสมผสาน.....	8
4. วิธีสังเคราะห์ข้อมูลเพิ่ม.....	8
เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques) .....	9
การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning).....	12
1. Bagging Method.....	12
2. Boosting Method.....	13
3. แรนดอมฟอร์เรส.....	15

มาตราวัดและการทดสอบประสิทธิภาพแบบจำลอง .....	16
งานวิจัยที่เกี่ยวข้อง .....	18
กรอบแนวคิดงานวิจัย .....	21
บทที่ 3 .....	22
วิธีการดำเนินการวิจัย .....	22
เครื่องมือที่ใช้ในการวิจัย .....	22
กรอบวิธีการดำเนินการวิจัย .....	24
บทที่ 4 .....	43
ผลการวิจัย .....	43
กระบวนการจัดเตรียมข้อมูลและการสำรวจข้อมูลเบื้องต้น .....	44
ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบเดี่ยว .....	51
ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบรวมกลุ่ม .....	56
เปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภท .....	68
ทำนายการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 .....	73
บทที่ 5 .....	76
สรุปผลการวิจัย .....	76
5.1 สรุปผลการวิจัย .....	76
5.2 อภิปรายผลการวิจัย .....	77
5.3 ประโยชน์ของสถิติที่ใช้ในการวิเคราะห์ .....	78
5.4 ข้อเสนอแนะ .....	78
เอกสารอ้างอิง .....	79
ภาคผนวก ก .....	81

## สารบัญภาพ

รูปที่ 1	DECISION TREE .....	10
รูปที่ 2	SUPPORT VECTOR MACHINE .....	11
รูปที่ 3	แสดงโครงสร้างพื้นฐานของการเรียนรู้แบบรวมกลุ่ม .....	12
รูปที่ 4	โครงสร้างวิธีการทำงานแบบ BAGGING .....	13
รูปที่ 5	แสดงโครงสร้างวิธีการทำงานแบบ BOOSTING (ADABOOST).....	14
รูปที่ 6	RANDOM FOREST .....	15
รูปที่ 7	ROC CURVE .....	17
รูปที่ 8	กรอบแนวคิดการวิจัย .....	21
รูปที่ 9	กรอบวิธีการดำเนินงานวิจัยสำหรับการทำแนกประเภท.....	25
รูปที่ 10	ขั้นตอนการแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ DT .....	32
รูปที่ 11	ขั้นตอนการแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ SVM .....	33
รูปที่ 12	ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี BAGGING โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	35
รูปที่ 13	ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี BAGGING โดยใช้ SVM เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	36
รูปที่ 14	ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี BOOSTING โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	38
รูปที่ 15	ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี BOOSTING โดยใช้ SVM เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	39
รูปที่ 16	ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี RANDOM FOREST .....	41
รูปที่ 17	ขั้นตอนการทำนายการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 .....	42
รูปที่ 18	แผนภูมิวงกลมสถานะนักศึกษา .....	44
รูปที่ 19	ปัจจัยที่ส่งผลต่อการฟื้นสภาพของนักศึกษาของการฟื้นสภาพกรณีที่ 1.....	69
รูปที่ 20	ปัจจัยที่ส่งผลต่อการฟื้นสภาพของนักศึกษาของการฟื้นสภาพกรณีที่ 2.....	70
รูปที่ 21	ปัจจัยที่ส่งผลต่อการฟื้นสภาพของนักศึกษาของการฟื้นสภาพกรณีที่ 3.....	72
รูปที่ 22	แผนภูมิวงกลมของการทำนายการฟื้นสภาพนักศึกษา .....	73
รูปที่ 23	แผนภูมิวงกลมของการทำนายการฟื้นสภาพนักศึกษา .....	74
รูปที่ 24	แผนภูมิวงกลมของการทำนายการฟื้นสภาพนักศึกษา .....	75

## สารบัญตาราง

ตาราง 1	CONFUSION MATRIX.....	16
ตาราง 2	เปรียบเทียบวิธีการสร้างแบบจำลอง ความแม่นยำ และปัจจัยที่มีความสำคัญกับการฟื้นฟู ของนักศึกษาระหว่างการศึกษาต่างๆ .....	20
ตาราง 3	ชื่อคุณลักษณะและคำนิยาม .....	26
ตาราง 4	กำหนดพารามิเตอร์ให้กับอัลกอริทึม DECISION TREE (DT).....	31
ตาราง 5	กำหนดพารามิเตอร์ให้กับอัลกอริทึม SUPPORT VECTOR MACHINE (SVM).....	31
ตาราง 6	กำหนดพารามิเตอร์ให้กับอัลกอริทึม BAGGING โดยใช้ DECISION TREE (DT) และ SUPPORT VECTOR MACHINE (SVM) เป็นอัลกอริทึมพื้นฐาน .....	34
ตาราง 7	กำหนดพารามิเตอร์ให้กับอัลกอริทึม BOOSTING โดยใช้ DECISION TREE (DT) และ SUPPORT VECTOR MACHINE (SVM) เป็นอัลกอริทึมพื้นฐาน .....	37
ตาราง 8	กำหนดพารามิเตอร์ให้กับอัลกอริทึม RANDOM FOREST .....	40
ตาราง 9	ชื่อคุณลักษณะและคำนิยามของข้อมูลการฟื้นฟูสภาพกรณีที่ 1 .....	45
ตาราง 10	ชื่อคุณลักษณะและคำนิยามของข้อมูลการฟื้นฟูสภาพกรณีที่ 2.....	47
ตาราง 11	ชื่อคุณลักษณะและคำนิยามของข้อมูลการฟื้นฟูสภาพกรณีที่ 3.....	48
ตาราง 12	ค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึม DT .....	51
ตาราง 13	CONFUSION MATRIX โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 1 .....	52
ตาราง 14	CONFUSION MATRIX โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 2.....	52
ตาราง 15	CONFUSION MATRIX โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 3.....	52
ตาราง 16	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม DT .....	53
ตาราง 17	ค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึม SVM .....	54
ตาราง 18	CONFUSION MATRIX โดยใช้อัลกอริทึม SVM ของการฟื้นฟูสภาพกรณีที่ 1 .....	54
ตาราง 19	CONFUSION MATRIX โดยใช้อัลกอริทึม SVM ของการฟื้นฟูสภาพกรณีที่ 2.....	54
ตาราง 20	CONFUSION MATRIX โดยใช้อัลกอริทึม SVM ของการฟื้นฟูสภาพกรณีที่ 3.....	54
ตาราง 21	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม SVM .....	55
ตาราง 22	ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี BAGGING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการ เรียนรู้ .....	56
ตาราง 23	CONFUSION MATRIX ด้วยวิธี BAGGING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การฟื้นฟูสภาพกรณีที่ 1 .....	56
ตาราง 24	CONFUSION MATRIX ด้วยวิธี BAGGING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การฟื้นฟูสภาพกรณีที่ 4.....	57

ตาราง 25 CONFUSION MATRIX ด้วยวิธี BAGGING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การผันสภาพกรณ์ที่ 3.....	57
ตาราง 26 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธี BAGGING โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	57
ตาราง 27 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี BAGGING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการ เรียนรู้.....	58
ตาราง 28 CONFUSION MATRIX ด้วยวิธีการ BAGGING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 1 .....	59
ตาราง 29 CONFUSION MATRIX ด้วยวิธีการ BAGGING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 2 .....	59
ตาราง 30 CONFUSION MATRIX ด้วยวิธีการ BAGGING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 3 .....	59
ตาราง 31 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ BAGGING โดยใช้ SVM เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	60
ตาราง 32 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี BOOSTING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการ เรียนรู้.....	61
ตาราง 33 CONFUSION MATRIX วิธีการ BOOSTING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การผันสภาพกรณ์ที่ 1 .....	61
ตาราง 34 CONFUSION MATRIX วิธีการ BOOSTING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การผันสภาพกรณ์ที่ 2.....	62
ตาราง 35 CONFUSION MATRIX วิธีการ BOOSTING โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของ การผันสภาพกรณ์ที่ 3.....	62
ตาราง 36 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ BOOSTING โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้.....	62
ตาราง 37 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี BOOSTING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับ การเรียนรู้.....	63
ตาราง 38 CONFUSION MATRIX ด้วยวิธีการ BOOSTING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการ เรียนรู้ ของการผันสภาพกรณ์ที่ 1 .....	63
ตาราง 39 CONFUSION MATRIX ด้วยวิธีการ BOOSTING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการ เรียนรู้ ของการผันสภาพกรณ์ที่ 2 .....	64

ตาราง 40	CONFUSION MATRIX ด้วยวิธีการ BOOSTING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการพันสภาพกรณีที่ 3 .....	64
ตาราง 41	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ BOOSTING โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ .....	64
ตาราง 42	ค่าพารามิเตอร์ที่เหมาะสมโดยใช้อัลกอริทึม RANDOM FOREST.....	65
ตาราง 43	CONFUSION MATRIX ด้วยวิธี RANDOM FOREST ของการพันสภาพกรณีที่ 1 .....	66
ตาราง 44	CONFUSION MATRIX ด้วยวิธี RANDOM FOREST ของการพันสภาพกรณีที่ 2 .....	66
ตาราง 45	CONFUSION MATRIX ด้วยวิธี RANDOM FOREST ของการพันสภาพกรณีที่ 3 .....	66
ตาราง 46	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม RANDOM FOREST	67
ตาราง 47	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการพันสภาพของข้อมูลการพันสภาพกรณีที่ 1.....	68
ตาราง 48	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการพันสภาพของข้อมูลการพันสภาพกรณีที่ 2.....	70
ตาราง 49	การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการพันสภาพของข้อมูลการพันสภาพกรณีที่ 3.....	71
ตาราง 50	รายละเอียดค่าใช้จ่ายในการดำเนินงาน .....	82
ตาราง 51	การดำเนินงานโครงการการวิจัย.....	83

## บทที่ 1

### บทนำ

#### ความเป็นมาและความสำคัญของปัญหา

สถาบันการศึกษามีบทบาทที่สำคัญในการพัฒนาประเทศ การศึกษาเป็นรากฐานที่สำคัญในการสร้างบุคคลให้มีความรู้ความสามารถในการปฏิบัติหน้าที่ และสามารถดำรงชีวิตอยู่ในสังคมอย่างสันติสุข (ชนิดาภา บุญประสม และ จรรย์ แสงราช, 2561) การที่นักศึกษาสามารถเล่าเรียนจนจบหลักสูตรได้นั้น จำเป็นจะต้องมีความรู้ความเข้าใจในวิชาชีพของตน นอกเหนือจากนี้ผู้ที่เกี่ยวข้องกับวงการศึกษาจำเป็นจะต้อง วางแผน ติดตาม และควบคุม ในแต่ละกระบวนการเพื่อส่งเสริมการพัฒนากระบวนการเรียนการสอนให้มีประสิทธิภาพ ตลอดจนช่วยกันหาแนวทางในการป้องกัน และการแก้ปัญหาการพัฒนาศักยภาพของนักศึกษาในระหว่างเรียน

ในสถาบันการศึกษาปัญหาการพัฒนาศักยภาพของนักศึกษาเป็นสิ่งที่จะต้องหาแนวทางในการแก้ไขปัญหา เนื่องจากปัจจุบันมีอัตราการพัฒนาศักยภาพของนักศึกษาในระดับที่เป็นปัญหา จากข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ระหว่างปี 2558 - 2563 มีการพัฒนาศักยภาพของนักศึกษามากถึง 205 คน คิดเป็นร้อยละ 30.19 จากจำนวนทั้งหมด 679 คน และพบว่าอัตราการพัฒนาศักยภาพของนักศึกษาเฉลี่ยแต่ละปีอยู่ที่ 29.24% (สำนักทะเบียนมหาวิทยาลัยขอนแก่น, 2564) ซึ่งมาจากหลายปัจจัย เช่น ปัญหาครอบครัว ปัญหาเศรษฐกิจ ปัญหาการเรียน และปัจจัยที่มาจากตัวนักศึกษาเองเป็นต้น (นนทวัฒน์ ทวีชาติ, อรยา เฟื่องประจักษ์, วิไลรัตน์ ยาทองไชย, และชูศักดิ์ ยาทองไชย, 2563) ปัจจัยเหล่านี้จะส่งผลกระทบต่อคณะและมหาวิทยาลัย ทำให้เสียเวลาในการบริหารจัดการ และเสียทรัพยากรในการลงทุน ส่วนนักศึกษาเสียเวลา และเสียค่าใช้จ่าย ดังนั้นทางคณะและมหาวิทยาลัยควรส่งเสริม และพัฒนากระบวนการเรียนการสอนให้มีประสิทธิภาพ ตลอดจนช่วยกันหาแนวทางในการป้องกันและแก้ไขปัญหา การพัฒนาศักยภาพของนักศึกษา หากนักศึกษามีพัฒนาศักยภาพก่อนจะจบการศึกษาถือว่าเป็นความสูญเสียทางการศึกษา จะส่งผลกระทบต่อด้านเศรษฐกิจของประเทศ และเศรษฐกิจของครอบครัวซึ่งต้องสิ้นเปลืองค่าใช้จ่ายไปโดยไม่ได้ประโยชน์ที่คุ้มค่า

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ดังนั้นการทำเหมืองข้อมูลเป็นการนำเอาข้อมูลมาวิเคราะห์เพื่อให้ได้ความรู้ใหม่ออกมาเพื่อนำไปใช้ประโยชน์ในการตัดสินใจ (นิสานันท์ พลอาสา, 2558) ในช่วงหลายปีที่ผ่านมาได้มีการนำเทคนิคการทำเหมืองข้อมูลมาใช้ในการศึกษาปัจจัย และทำนายการพัฒนาศักยภาพของนักศึกษาอย่างแพร่หลาย เช่น การทำนายการลาออกของนักศึกษา (Tenpipat &



Akkarajitsakul, 2020) และ การใช้แบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มเพื่อทำนายการลาออกของนักศึกษา (Naseem, Chaudhary, Sharma, & Lal, 2020) เป็นต้น

งานวิจัยนี้มีแนวคิดที่จะพัฒนาประสิทธิภาพในการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้การเรียนรู้แบบรวมกลุ่ม เพื่อแก้ไขปัญหาในด้านความถูกต้องและแม่นยำสำหรับการจำแนกการพัฒนาของนักศึกษา โดยได้นำเอาหลักทฤษฎีการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) มาใช้เพื่อช่วยเพิ่มประสิทธิภาพในความถูกต้องในการจำแนกประเภทการพัฒนาของนักศึกษาให้สูงขึ้น และใช้อัลกอริทึมในการตรวจจับการพัฒนาของนักศึกษา โดยมีนำข้อมูลก่อนที่เหตุการณ์การพัฒนาของนักศึกษาจะเกิดขึ้นมาหาแนวโน้มการพัฒนาของนักศึกษา เพื่อที่จะแก้ไขปัญหาและวางกลยุทธ์ในการจัดการการพัฒนาของนักศึกษาได้อย่างทันทั่วทั้ง

### วัตถุประสงค์ของการวิจัย

1. สร้างแบบจำลองการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้วิธีการเรียนรู้แบบเดี่ยวและแบบรวมกลุ่ม
2. เปรียบเทียบประสิทธิภาพการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้วิธีการเรียนรู้แบบเดี่ยวและแบบรวมกลุ่ม

### สมมติฐานของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการตั้งสมมติฐานไว้ คือ การนำเทคนิควิธีการเรียนรู้แบบเดี่ยวและแบบรวมกลุ่มมาใช้สำหรับการจำแนกประเภทในการพัฒนาของนักศึกษา เพื่อที่จะได้แบบจำลองที่มีประสิทธิภาพความถูกต้องและความแม่นยำที่สุด

### ขอบเขตของการวิจัย

1. การวิจัยนี้ใช้ข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้ข้อมูลเฉพาะนักศึกษาปี 1 ในแต่ละปีการศึกษา 6 ปีย้อนหลังตั้งแต่ปี พ.ศ. 2558 - 2564 จำนวน 719 คน
2. ใช้อัลกอริทึม Decision Tree และ Support Vector Machine สำหรับขั้นตอนการจำแนกประเภทแบบเดี่ยว รวมถึงใช้เป็นอัลกอริทึมพื้นฐานการเรียนรู้ร่วมกับวิธีการเรียนรู้แบบรวมกลุ่มแบบ Bagging และ Boosting
3. ใช้วิธีการเรียนรู้แบบรวมกลุ่ม 3 วิธี ได้แก่ Bagging, Boosting, และ Random Forest

### ความหมายหรือนิยามคำศัพท์เฉพาะ

*Ensemble Learning* หมายถึง การเรียนรู้แบบรวมกลุ่ม

*Machine Learning* หมายถึง การเรียนรู้ของเครื่อง

*Data Mining* หมายถึง การทำเหมืองข้อมูล

*Data Classification Techniques* หมายถึง เทคนิคการจำแนกประเภทข้อมูล

*Cross Validation* หมายถึง วิธีการแบ่งข้อมูลเพื่อใช้วัดประสิทธิภาพแบบจำลอง

### ประโยชน์ที่คาดว่าจะได้รับการวิจัย

1. ทราบปัจจัยที่มีผลต่อการพัฒนาสภาพนักศึกษาของนักศึกษาระดับปริญญาตรีสาขาวิชาสถิติ คณะวิทยาศาสตร์
2. ได้แบบจำลองการจำแนกประเภทที่มีประสิทธิภาพความถูกต้องและแม่นยำในการจำแนกประเภทสำหรับการพัฒนาสภาพของนักศึกษา
3. นำข้อมูลสารสนเทศที่ได้จากการจำแนกประเภทสำหรับการพัฒนาสภาพของนักศึกษาไปประยุกต์ใช้ในการปรับปรุงหลักสูตร วางแผน และพัฒนานักศึกษา
4. ใช้แบบจำลองในการลดเวลาปฏิบัติงานของเจ้าหน้าที่ที่เกี่ยวข้องในการประเมิน และวางแผนในการพัฒนาสภาพของนักศึกษา

### ความเป็นไปได้ในการพัฒนาต่อยอด

1. สามารถนำแบบจำลองไปใช้งานในระบบออนไลน์ เพื่อให้ผู้ที่เกี่ยวข้องในการดูแลนักศึกษาในการเตรียมความพร้อม วางกลยุทธ์ และวางแผนรับมือกับการพัฒนาสภาพของนักศึกษาในระดับปริญญาตรี
2. สามารถนำชุดข้อมูลการพัฒนาสภาพของนักศึกษาในระดับคณะ และมหาวิทยาลัยขอนแก่นมาพัฒนาตัวแบบให้สามารถจำแนกประเภทการพัฒนาสภาพของนักศึกษาได้แม่นยำมากขึ้น

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการจำแนกประเภทสำหรับการฟื้นฟูสภาพของนักศึกษา โดยใช้วิธีการเรียนรู้แบบรวมกลุ่ม ทางผู้วิจัยได้ทำการศึกษาทฤษฎีที่เกี่ยวข้อง คีขางานวิจัย บทความ และเอกสารทางวิชาการต่างๆ เพื่อเป็นแนวทางและกรอบในการศึกษา โดยสรุปเป็นหัวข้อต่างๆดังต่อไปนี้

1. ทฤษฎีการฟื้นฟูสภาพ
2. การทำเหมืองข้อมูล (Data Mining)
3. เทคนิคการแก้ปัญหาข้อมูลไม่สมดุลสำหรับการจำแนกประเภท (Imbalanced Data Problem Solving in Classification Technique)
  - 3.1 วิธีสุ่มเกิน (Over sampling)
  - 3.2 วิธีสุ่มลด (Under sampling)
  - 3.3 วิธีผสมผสาน (Hybrid Methods)
  - 3.4 วิธีสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE)
4. เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques)
  - 4.1 ต้นไม้ตัดสินใจ (Decision Tree)
  - 4.2 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)
5. การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)
  - 5.1 Bagging Method
  - 5.2 Boosting Method
  - 5.3 แรนดอมฟอร์เรส (Random Forest)
6. มาตรฐานวัดและการทดสอบประสิทธิภาพแบบจำลอง
7. งานวิจัยที่เกี่ยวข้อง
8. กรอบแนวคิดงานวิจัย

## ทฤษฎีการพัฒนสภาพ

### 1. การพัฒนสภาพ

การพัฒนสภาพหมายถึง การที่นักศึกษาพ้นจากสภาพนักศึกษา โดยมหาวิทยาลัยให้ออกจากสถาบันก่อนเรียนครบตามหลักสูตรที่กำหนดไว้ 2 กรณี ได้แก่ กรณีที่ 1 พ้นจากสภาพนักศึกษาเนื่องจากคะแนนเฉลี่ยสะสมไม่ถึงเกณฑ์ที่มหาวิทยาลัยกำหนด กรณีที่ 2 พ้นจากสภาพนักศึกษาเนื่องจากขาดการติดต่อกับมหาวิทยาลัย (สุกัญญา ทารส, 2562)

### 2. ระเบียบมหาวิทยาลัยขอนแก่น ว่าด้วยการศึกษาชั้นปริญญาตรี พ.ศ. 2562

จากระเบียบมหาวิทยาลัยขอนแก่น ว่าด้วยการศึกษาชั้นปริญญาตรี พ.ศ. 2562 การวัดและประเมินผล และการพัฒนสภาพนักศึกษาในมหาวิทยาลัยขอนแก่น (สำนักทะเบียนมหาวิทยาลัยขอนแก่น, 2562) มีดังนี้

#### 2.1 การวัดและประเมินผล

มหาวิทยาลัยขอนแก่นจัดให้มีการวัดผลแต่ละรายวิชาที่นักศึกษาลงทะเบียน ซึ่งอาจารย์ผู้สอนต้องแจ้งเกณฑ์และเงื่อนไขในการประเมินผลในแต่ละวิชาให้นักศึกษาทราบล่วงหน้า และการประเมินผลในแต่ละรายวิชาจะใช้ระดับคะแนนตัวอักษร ดังนี้

ระดับคะแนนตัวอักษร	ความหมาย	ค่าคะแนนต่อหน่วยกิจ
A	ผลการประเมินขั้นดีเยี่ยม (Excellent)	4.0
B+	ผลการประเมินขั้นดีมาก (Very Good)	3.5
B	ผลการประเมินขั้นดี (Good)	3.0
C+	ผลการประเมินขั้นค่อนข้างดี (Fairly Good)	2.5
C	ผลการประเมินพอใช้ (Fair)	2.0
D+	ผลการประเมินขั้นอ่อน (Poor)	1.5
D	ผลการประเมินขั้นอ่อนมาก (Very Poor)	1.0
F	ผลการประเมินขั้นตก (Fail)	0

และตัวอักษรที่มีความหมายเฉพาะซึ่งแสดงถึงสถานภาพนักศึกษา คือ I P R S T U และ W ตัวอักษรเหล่านี้ไม่มีค่าคะแนนยกเว้น ตัวอักษร T

ตัวอักษร	ความหมาย
I	ยังไม่สมบูรณ์ (Incomplete)
P	กำลังดำเนินการอยู่ (In Progress)
R	ซ้ำชั้น (Repeat)

ตัวอักษร	ความหมาย
S	พอใจ (Satisfactory)
T	รับโอน (Transfer)
U	ไม่พอใจ (Unsatisfactory)
W	การขอถอนรายวิชา (Withdrawal)

## 2.2 การพัฒนสภาพนักศึกษา

จากระเบียบมหาวิทยาลัยขอนแก่น ว่าด้วยการศึกษาชั้นปริญญาตรี พ.ศ. 2562 นักศึกษาจะพัฒนสภาพนักศึกษาก็ต่อเมื่อ

1. ตาย
2. ลาออก
3. ตกออก นักศึกษาจะถูกพิจารณาให้ตกออกในกรณีดังต่อไปนี้
  - 3.1. ระดับคะแนนเฉลี่ยสะสมไม่ถึง 1.50 เมื่อได้ลงทะเบียนเรียนมาแล้ว และมีหน่วยกิตสะสมตั้งแต่ 30-59 หน่วยกิต
  - 3.2. ระดับคะแนนเฉลี่ยสะสมไม่ถึง 1.75 เมื่อได้ลงทะเบียนเรียนมาแล้ว และมีหน่วยกิตสะสมตั้งแต่ 60 หน่วยกิตขึ้นไป
  - 3.3. สำหรับนักศึกษาหลักสูตรแพทยศาสตรบัณฑิต ให้เป็นไปตามหลักเกณฑ์ที่มหาวิทยาลัยกำหนด
4. ถูกสั่งให้พ้นสภาพตามระเบียบข้อบังคับของมหาวิทยาลัย
5. ขาดคุณสมบัติการเข้าเป็นนักศึกษาของมหาวิทยาลัย ตามระเบียบฯ
6. เรียนสำเร็จตามหลักสูตร
7. ไม่ลงทะเบียนให้เสร็จสิ้นภายในเวลาที่มหาวิทยาลัยกำหนดในแต่ละภาคการศึกษา
8. ไม่ชำระค่าธรรมเนียมเพื่อขึ้นโดยต่อทะเบียนภายในระยะเวลาที่มหาวิทยาลัยกำหนด
9. ศึกษาในมหาวิทยาลัยเกินจำนวนสองเท่าของระยะเวลาการศึกษาที่กำหนดไว้ในหลักสูตร
10. ต้องโทษโดยคำพิพากษาถึงที่สุดให้จำคุกเว้นแต่ความผิดลหุโทษ หรือความผิดที่ได้กระทำโดยประมาท
11. โอนไปเป็นนิสิตนักศึกษาของสถาบันอุดมศึกษาอื่น
12. เหตุอื่นตามที่มหาวิทยาลัยกำหนด

## การทำเหมืองข้อมูล (Data Mining)

### 1. ทฤษฎีการทำเหมืองข้อมูล

ทฤษฎีการทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ดังนั้นการทำเหมืองข้อมูล เป็นการนำเอาข้อมูล

มาวิเคราะห์เพื่อให้ได้ความรู้ใหม่ออกมาเพื่อนำไปใช้ประโยชน์ในการตัดสินใจ (นิสานันท์ พลอาสา, 2558) ในปัจจุบันมีความก้าวหน้าทางเทคโนโลยีทำให้การจัดเก็บข้อมูลเป็นจำนวนมาก เพื่อช่วยวิเคราะห์ปัญหาตัดสินใจ และดำเนินงานในหน่วยงานในปัจจุบัน การวิเคราะห์ข้อมูลเพื่อให้ความรู้ที่เกี่ยวข้องกับข้อมูลได้แก่ ตัวแบบที่แสดงความสัมพันธ์ระหว่างข้อมูล ความรู้เหล่านี้สามารถนำมาใช้ประโยชน์ในการดำเนินงาน และการตัดสินใจภายในองค์กร การทำเหมืองข้อมูลที่สำคัญได้แก่ การคัดเลือกข้อมูล (Selection) การเตรียมข้อมูล (Preprocessing) การแปลงข้อมูล (Transformation) การวิเคราะห์และค้นหารูปแบบข้อมูล (Data Mining) และการแปล/ประเมินผลการวิเคราะห์ข้อมูล (Interpretation/Evaluation)

## 2. ประเภทการทำเหมืองข้อมูล

การทำเหมืองข้อมูลแบบทำนาย (Predictive Mining) คือการนำข้อมูลที่มีอยู่มาใช้ในการทำนายผลข้อมูลในอนาคตที่ไม่ทราบมาก่อน ซึ่งการสร้างแบบจำลองรูปแบบนี้จะเน้นการแบ่งข้อมูลออกเป็นกลุ่มตามคุณสมบัติของข้อมูล ในกรณีที่ข้อมูลไม่ต่อเนื่องจะใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) และในกรณีที่ข้อมูลมีความต่อเนื่องจะใช้เทคนิคการถดถอย (Regression)

การทำเหมืองข้อมูลแบบพรรณนา (Descriptive Mining) คือการนำข้อมูลที่มีอยู่มาศึกษาหาคำอธิบายคุณลักษณะทั่วไปของข้อมูล เพื่อใช้เป็นแนวทางในการตัดสินใจ เช่น เทคนิคการหาความสัมพันธ์ (Association) หรือเทคนิคการจัดกลุ่ม (Clustering)

## 3. การทำเหมืองข้อมูล

Cluster Analytic คือ การจัดกลุ่มข้อมูลซึ่งมีลักษณะคล้ายกับการแบ่งประเภท (Classification) แต่จะไม่เหมือนกันโดยการแบ่งประเภทจะวิเคราะห์ข้อมูลที่กำหนดผลลัพธ์ แต่สำหรับการแบ่งกลุ่มเป็นการวิเคราะห์โดยไม่พิจารณาข้อมูลที่กำหนดผลลัพธ์ แต่จะใช้ขั้นตอนวิธีการจัดกลุ่มเพื่อค้นหากลุ่มที่สามารถยอมรับได้เพื่อจัดเข้ากลุ่ม กล่าวคือ กลุ่มของวัตถุมีการสร้างขึ้นโดยเปรียบเทียบวัตถุที่มีความเหมือนกันจัดเข้ากลุ่มเดียวกัน

Association Rule เป็นการค้นหาความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน ความสำคัญของกฎทำการวัดโดยใช้ข้อมูลสองตัวด้วยกันคือ ค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ หรือเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่ถูกถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่ถูกกฎสามารถนำไปใช้ได้ ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ อัลกอริทึม Apriori

Classification Analytic เป็นการจัดประเภทของ วัตถุประสงค์เพื่อให้สามารถใช้เป็นตัวแทน ทำนายประเภท ซึ่งตัวแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยใช้ข้อมูลที่ระบุผลลัพธ์เรียบร้อยแล้ว รูปแบบของตัวแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น

### เทคนิคการแก้ปัญหาข้อมูลไม่สมดุลสำหรับการจำแนกประเภท (Imbalanced Data Problem Solving in Classification Technique)

ข้อมูลไม่สมดุลของกลุ่มตัวแปรผลลัพธ์ที่นำมาศึกษามีผลต่อความถูกต้องของสมการการทำนาย ซึ่งเป็นปัญหาหลักที่นักวิจัยให้ความสนใจในปัญหาความไม่สมดุลของข้อมูลที่พบได้บ่อยครั้งในข้อมูลจริง เมื่อนำข้อมูลเหล่านี้มาใช้งานทางด้านการเรียนรู้ของเครื่อง (Machine Learning) และการทำเหมืองข้อมูล (Data Mining) จะส่งผลกระทบต่อการเรียนรู้ของอัลกอริทึมในการจำแนกข้อมูลด้วยวิธีการจำแนกข้อมูลแบบปกติที่ให้ความสำคัญกับข้อมูลกลุ่มผลลัพธ์เท่ากันจะทำให้ประสิทธิภาพในการจำแนกประเภทข้อมูลส่วนน้อยมีความถูกต้องน้อยลง (วิชญ์วิสิฐ เกษรสิทธิ์, วิชิต หล่อจิระชุมณฑ์กุล, และจิราวัลย์ จิตรถเวช, 2561) ซึ่งการแก้ปัญหาข้อมูลไม่สมดุลมีใช้เทคนิควิธีดังนี้

#### 1. วิธีสุ่มเกิน (Over Sampling)

วิธีการสุ่มเกินเป็นการเพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อยให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมาก ซึ่งการเพิ่มข้อมูลนั้นจะเพิ่มโดยการสุ่มเลือกจากข้อมูลเดิมในกลุ่มส่วนน้อยโดยใช้วิธีการสุ่มแบบเป็นระบบ

#### 2. วิธีสุ่มลด (Under Sampling)

วิธีสุ่มลดเป็นการลดจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมากให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อยโดยใช้วิธีการสุ่มแบบเป็นระบบ

#### 3. วิธีผสมผสาน (Hybrid Methods)

วิธีผสมผสานเป็นวิธีการที่นำเทคนิควิธีสุ่มเกินและวิธีสุ่มลดมาทำงานร่วมกัน โดยพยายามหาค่ากลางในการชักตัวอย่างให้ได้ตามจำนวนที่อยู่ตรงกลางระหว่างข้อมูลในกลุ่มส่วนมากกับข้อมูลในกลุ่มส่วนน้อย

#### 4. วิธีสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE)

วิธีสังเคราะห์ข้อมูลเพิ่มเป็นเทคนิคการสุ่มตัวอย่างแบบพิเศษของการสุ่มเพิ่ม แทนที่จะสุ่มเพิ่มโดยใช้ข้อมูลเดิมแต่จะทำการสังเคราะห์ข้อมูลขึ้นมาใหม่ จากข้อมูลเดิมที่มีอยู่โดยใช้อัลกอริทึมเพื่อนบ้านที่อยู่ใกล้ที่สุด (K-Nearest Neighbor) ในการขยายขอบเขตการตัดสินใจของตัวแบบ ซึ่งในขั้นตอนการสังเคราะห์ข้อมูลมีขั้นตอนดังนี้คือระบุเพื่อนบ้านที่ใกล้เคียงที่สุด  $k$  ค่าของข้อมูลเดิม สำหรับข้อมูลเดิม  $M$  หาก  $k=1$  ที่มีระยะใกล้เคียงกับข้อมูลเดิมโดย  $l$  คือจำนวนเพื่อนบ้านใกล้เคียงกับจุด  $M$  และสุ่มเลือกจุดระหว่างสองจุดและสร้างกรณีใหม่

ตัวอย่างเช่นสร้างจุด  $m_1(c_1, c_2, \dots, c_n)$  ระหว่าง  $M(a_1, a_2, \dots, a_n)$  และ  $M_1(b_1, b_2, \dots, b_n)$  เมื่อ

$$c_1 = a_1 + (b_1 - a_1) \times \text{rand}('UNIFORM')$$

$$c_2 = a_2 + (b_2 - a_2) \times \text{rand}('UNIFORM')$$

$$c_n = a_n + (b_n - a_n) \times \text{rand}('UNIFORM')$$

โดยที่  $m_1$  คือจุดที่สังเคราะห์ขึ้นมาใหม่ระหว่าง  $M(a_1, a_2, \dots, a_n)$  และ  $M_1(b_1, b_2, \dots, b_n)$   
 $a_1, a_2, \dots, a_n$  คือข้อมูลในค่าสังเกตที่จุด  $M$  และ  $b_1, b_2, \dots, b_n$  คือข้อมูลในค่าสังเกตที่จุด  $M_1$

## เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques)

### 1. ต้นไม้ตัดสินใจ (Decision Tree: DT)

Decision Tree (DT) เป็นหนึ่งในอัลกอริทึมการเรียนรู้ของเครื่องที่เร็วและโดดเด่น ต้นไม้การตัดสินใจจำลองตรรกะการตัดสินใจ เช่น ทดสอบความสอดคล้องผลลัพธ์สำหรับการจัดประเภทรายการข้อมูลให้เป็นโครงสร้างแบบต้นไม้ โหนดของต้นไม้โดยปกติมีหลายระดับโดยที่โหนดแรกหรือบนสุดเรียกว่า รูดโหนด (Root node) โหนดภายใน (Internal Node) ทั้งหมด (เช่น โหนดที่มีโหนดย่อยอย่างน้อยหนึ่งรายการ) เป็นโหนดที่แสดงถึงคุณลักษณะ (Feature) ที่ใช้ในการแบ่งกลุ่มข้อมูล โดยมีรูดโหนด (Root Node) อยู่บนสุดเป็นโครงสร้าง อัลกอริทึมการจำแนกประเภทต้นไม้ตัดสินใจจะแยกตัวไปยังโหนดย่อยที่เหมาะสมขึ้นอยู่กับผลการทดสอบ โดยที่กระบวนการทดสอบและการแตกแขนงจะทำซ้ำจนกว่าจะถึงใบโหนด (Leaf Node) การตัดสินใจต้นไม้ถูกพบว่าง่ายต่อการตีความ และเรียนรู้ได้รวดเร็ว เมื่อสำรวจกฎที่สร้างมาจากต้นไม้ตัดสินใจจะพบว่าเส้นทางจะให้ข้อมูลที่เพียงพอต่อการคาดเดาเกี่ยวกับผลลัพธ์ต่างๆ (Uddin, Khan, & Moni, 2019)

จากภาพที่ 1 แสดงตัวอย่างต้นไม้ตัดสินใจ (Decision Tree) ที่มีตัวแปรแต่ละตัว (C1, C2 และ C3) จะถูกแทนด้วยวงกลม และผลการตัดสินใจ (Class A และ Class B) จะแสดงด้วยสีเหลี่ยม เพื่อที่จะจำแนกตัวอย่างไปยังผลลัพธ์ได้สำเร็จ แต่ละเส้นทางจะมีป้ายกำกับว่า 'จริง' หรือ 'เท็จ' ตามค่าผลลัพธ์จากการทดสอบ สูตรที่ใช้ในการคำนวณหาค่า Information Gain ต้องเริ่มจากการหาค่า Entropy ดังนี้



Entropy:

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

โดย  $S$  คือ Attribute ที่นำมาวัดค่า Entropy

$P_i$  คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม  $i$  กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

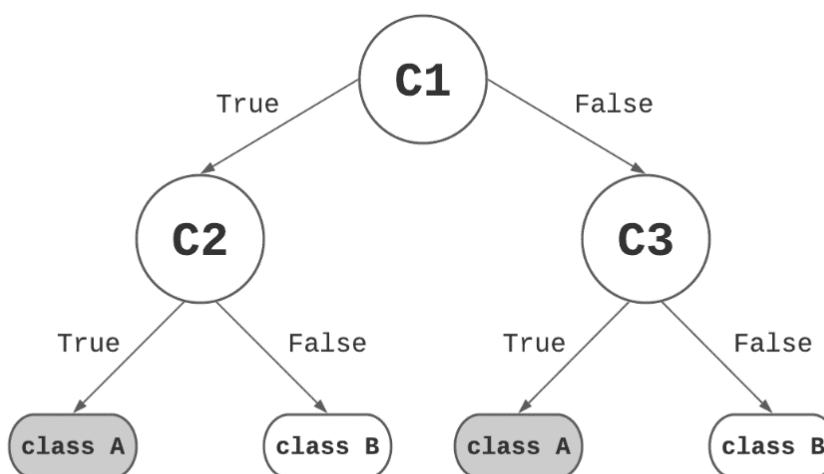
Information Gain:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

โดย  $A$  คือ Attribute  $A$

$|S_v|$  คือ สมาชิกของ Attribute  $A$  ที่มีค่า  $v$

$|S|$  คือ จำนวนสมาชิกของกลุ่มตัวอย่าง



รูปที่ 1 Decision Tree

## 2. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

Support Vector Machine (SVM) เป็นเทคนิคการจำแนกประเภทของข้อมูลที่มีพื้นฐานมาจากทฤษฎีการเรียนรู้ทางสถิติ และสามารถจำแนกได้ทั้งข้อมูลเชิงเส้นและไม่เชิงเส้น ซึ่งคล้ายเทคนิคโครงข่ายประสาทเทียม โดย SVM ใช้หลักการลดค่าความเสี่ยงเชิงโครงสร้างให้ต่ำที่สุด (Structural Risk Minimization) เพื่อลดค่าความผิดพลาดของการทำนาย (Minimized Error) พร้อมกับเพิ่มระยะการแบ่งแยกให้มากที่สุด (Maximized Margin) ระยะขอบ (Margin) คือระยะห่างระหว่างการตัดสินใจไฮเปอร์เพลน (Decision Hyperplane) และตัวอย่างที่ใกล้ที่สุดจะเป็นสมาชิกของผลลัพธ์นั้น

หลักการของ SVM คือการหาสมบัติของสมการเพื่อสร้างเส้นแบ่งแยกประเภทข้อมูลทำการโดยการเลือกเส้นหรือระนาบเพื่อแบ่งแยกประเภทข้อมูลที่เหมาะสมที่สุด (ปติตญา บุญรักษา และ จารีย์ ทองคำ, 2560)

จากภาพที่ 2 Support Vector Machine คือวิธีการทำงาน SVM ได้ระบุไฮเปอร์เพลน (Hyperplane) เป็นเส้นซึ่งเป็นการแยกระหว่างผลลัพธ์ ‘ดาว’ และ ‘วงกลม’ ให้เหมาะสมที่สุด

กำหนดให้  $(x_i, y_i), \dots, (x_n, y_n)$  เมื่อ  $x \in R^m, y \in \{-1, 1\}$  เป็นตัวอย่างที่ใช้สำหรับการสอน โดย

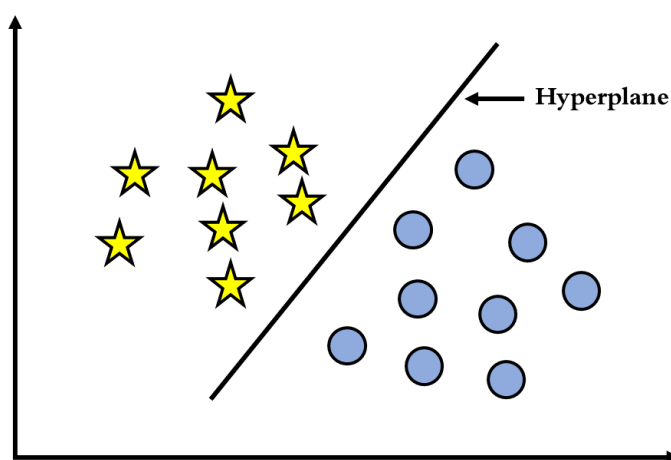
- $n$  คือ จำนวนข้อมูลตัวอย่าง
- $m$  คือ จำนวนมิติข้อมูล
- $x$  คือ ข้อมูลนำเข้า
- $y$  คือ ประเภทข้อมูล ซึ่งประกอบไปด้วย 2 กลุ่ม มีค่า  $+1$  หรือ  $-1$   
( $+1$  = “ข้อมูลบวก” และ  $-1$  = “ข้อมูลลบ”)

การสร้างเส้นระนาบตัดสินใจเพื่อแบ่งแยกกลุ่มผลลัพธ์ของข้อมูลสามารถคำนวณได้ดังนี้

$$(W * x_1) + b > 0 \text{ ถ้า } y_i = +1 \text{ และ } (W * x_2) + b < 0$$

$$\text{ถ้า } y_i = -1$$

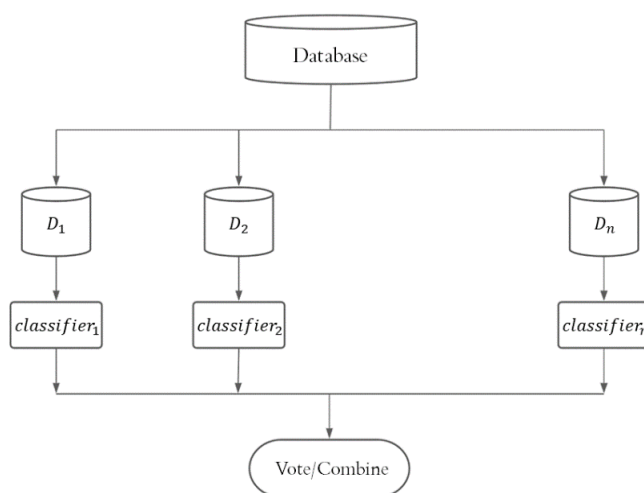
- โดย
- $w$  คือ เวกเตอร์น้ำหนัก
  - $x_1$  คือ เวกเตอร์ข้อมูลที่มีค่าเป็นบวก
  - $x_2$  คือ เวกเตอร์ข้อมูลที่มีค่าเป็นลบ
  - $b$  คือ ค่าความคลาดเคลื่อน (Bias)



รูปที่ 2 Support Vector Machine

## การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)

การเรียนรู้แบบรวมกลุ่มเป็นวิธีการรวมเอากลุ่มของตัวจำแนกประเภทข้อมูลที่สร้างขึ้นหลายๆ ตัว จำแนกและมีความเป็นอิสระต่อกันมาพิจารณาร่วมกัน เพื่อช่วยในการตัดสินใจสำหรับการหาคำตอบโดยใช้วิธีการรวม (Combine) หรือ วิธีการโหวต (Voting) เพื่อให้ได้ผลลัพธ์ในการจำแนกข้อมูลที่มีประสิทธิภาพสูงซึ่งเทคนิคการเรียนรู้แบบรวมกลุ่มมีอยู่หลากหลายวิธีแต่สำหรับวิธีการที่มีประสิทธิภาพได้รับความนิยมได้แก่ วิธี Bagging และ Boosting (ปัทม์ อุปการ, 2560)



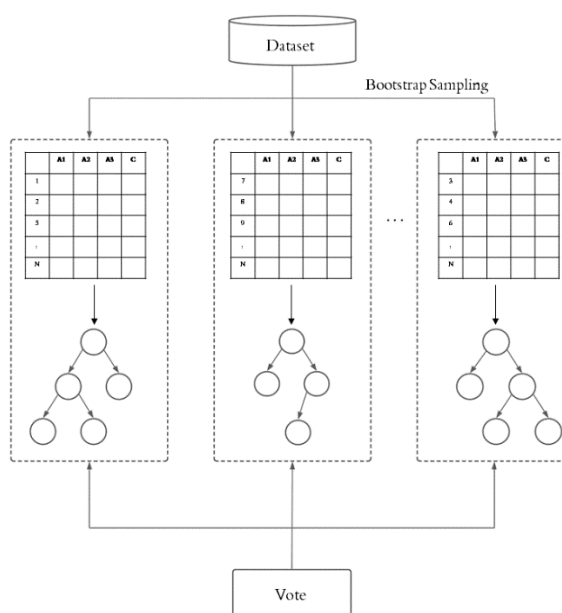
รูปที่ 3 แสดงโครงสร้างพื้นฐานของการเรียนรู้แบบรวมกลุ่ม

จากภาพที่ 3 แสดงโครงสร้างพื้นฐานของแบบจำลองการเรียนรู้แบบรวมกลุ่ม ซึ่งจะประกอบด้วย 3 ขั้นตอนหลัก ได้แก่ ขั้นตอนที่ 1 การสร้างชุดข้อมูลตัวอย่างขึ้นมาหลายๆชุดที่มีลักษณะแตกต่างกันในแต่ละชุด ขั้นตอนที่ 2 สร้างแบบการจำแนกประเภทข้อมูลหลายๆตัวจำแนก เพื่อเรียนรู้ชุดข้อมูลที่สร้างขึ้นในแต่ละชุด และขั้นตอนสุดท้ายเป็นการรวบรวมตัวจำแนกประเภทหลายๆตัวจำแนกที่สร้างขึ้นจากขั้นตอนที่ 2 เพื่อร่วมกันตัดสินใจในการพิจารณาหาคำตอบ โดยใช้วิธีการรวมแบบจำลองหรือการโหวตจากเสียงข้างมาก (Majority Vote) เพื่อให้ได้คำตอบที่ดีที่สุด

### 1. Bagging Method

วิธีการเรียนรู้แบบรวมกลุ่มโดยใช้วิธี Bagging หรือเรียกอีกอย่างว่า Bootstrap Aggregating เป็นวิธีหนึ่งของการเรียนรู้แบบรวมกลุ่มที่มีประสิทธิภาพ โดยส่วนใหญ่จะใช้งานร่วมกับการจำแนกประเภทในรูปแบบต้นไม้ช่วยตัดสินใจแต่ก็สามารถใช้งานร่วมกับอัลกอริทึมการจำแนกประเภทในรูปแบบอื่นได้ ซึ่งกระบวนการทำงานจะสุ่มชุดข้อมูลด้วยตัวอย่างขึ้นมาใหม่โดยใช้วิธีที่เรียกว่า “Bootstrap Sampling” ซึ่งข้อมูลในแต่ละชุดที่ถูกสุ่มขึ้นมาจะถูกเรียนรู้ที่มีลักษณะต่างกันในการสร้างแบบจำลอง

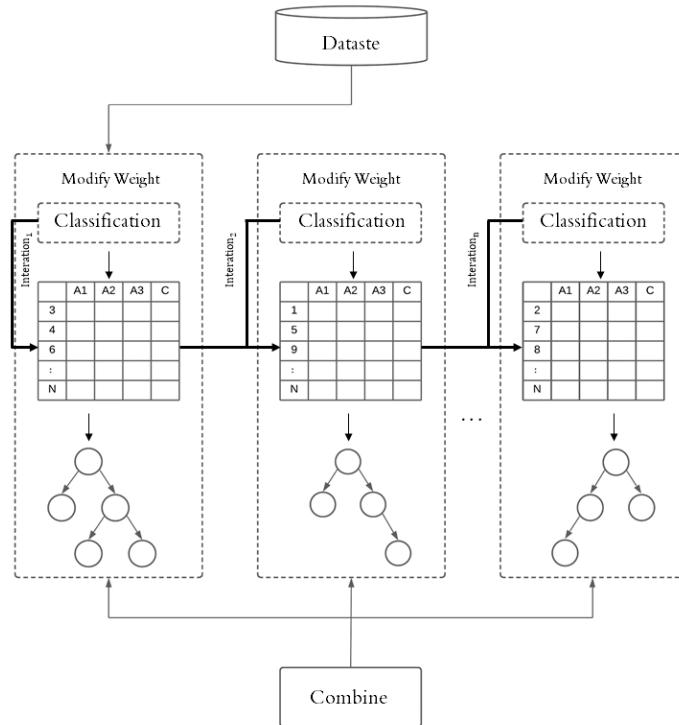
เนื่องจากข้อมูลที่ถูกเรียนรู้มีความแตกต่างกันทำให้เกิดความหลากหลายของแบบจำลอง ผลลัพธ์สุดท้ายจะทำการโหวตเสียงข้างมากจากแบบจำลองที่สร้างขึ้นเหล่านั้นเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดเพียงหนึ่งคำตอบ วิธีการทำงานแบบ Bagging จะมีผลกระทบต่อการจำแนกประเภทข้อมูลที่ไม่เป็นเชิงเส้น (Non-Linear) เมื่อข้อมูลที่ใช้ในการเรียนรู้ เปลี่ยนแปลงเพียงเล็กน้อย สามารถแสดงวิธีการทำงานแบบ Bagging ได้ดังรูปที่ 4



รูปที่ 4 โครงสร้างวิธีการทำงานแบบ Bagging

## 2. Boosting Method

Boosting เป็นวิธีการเรียนรู้แบบรวมกลุ่มอีกหนึ่งวิธีที่นิยมนำมาใช้ในการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่ม ลักษณะจะแตกต่างจากวิธีการของ Bagging ในส่วนของการถ่วงค่าน้ำหนักให้กับข้อมูล ตัวอย่างที่ได้ทำการเรียนรู้โดยเน้นไปที่การหาค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้ข้อมูลเรียกลักษณะนี้ว่า “Weak Learning” และในขั้นตอนสุดท้ายจะใช้วิธีการรวมตัวแบบจำแนกประเภทที่สร้างขึ้นไปหลายๆตัวจำแนกโดยพิจารณาจากค่าเฉลี่ยในการถ่วงน้ำหนัก (Mean Weight) และทำการโหวตเพื่อให้ได้ผลลัพธ์เพียงคำตอบเดียว สำหรับวิธีการของ Boosting ที่นิยมนำมาใช้คือ Adaptive Boosting (AdaBoost) วิธีการของ AdaBoost จะทำการถ่วงค่าน้ำหนักให้กับข้อมูลตัวอย่างที่ถูกเรียนรู้ของแต่ละรอบในการสร้างแบบจำลอง โดยข้อมูลตัวอย่างที่ทำการจำแนกประเภทให้ถูกต้องจะถูกลดค่าน้ำหนักลง ส่วนข้อมูลตัวอย่างที่จำแนกประเภทผิดพลาดจะถูกเพิ่มค่าน้ำหนักให้มีความสำคัญมากขึ้นเพื่อให้ข้อมูลนั้นมีโอกาสถูกเลือกในการเรียนรู้รอบต่อไป



รูปที่ 5 แสดงโครงสร้างวิธีการทำงานแบบ Boosting (AdaBoost)

สามารถแสดงว่า Boosting แบบ AdaBoost ได้ดังรูปที่ 5 เนื่องจาก AdaBoost เป็น อัลกอริทึมที่มีการปรับปรุงค่าน้ำหนักให้กับข้อมูล ดังนั้นการหาค่าน้ำหนักจะอาศัยค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้ ของข้อมูลเป็นหลัก

$$\varepsilon_i = \sum_{k: c_i(x_k) \neq y_k} D_i(k) \quad (1)$$

เริ่มจากการคำนวณหาความผิดพลาดที่เกิดขึ้นในการเรียนรู้ข้อมูลตามสมการ (1) โดยที่ เป็นการกำหนด  $D_i$  น้ำหนักให้กับชุดข้อมูล  $D_i = 1/m$  เมื่อ  $m$  คือ จำนวนข้อมูลตัวอย่างทั้งหมด และ  $c_i$  เป็นตัวจำแนกประเภทที่จำแนกข้อมูลตามที่  $k$  ผิดพลาด ซึ่ง  $\varepsilon_i$  คือ ผลรวมของค่าความผิดพลาดที่ได้จากตัวจำแนกประเภทที่จำแนกข้อมูลตัวที่  $k$  ผิดพลาดหลังจากนั้นทำการคำนวณค่าน้ำหนักของข้อมูลในสมการ (2)

$$\alpha_i = \frac{1}{2} \ln\left(\frac{1-\varepsilon_i}{\varepsilon_i}\right) \quad (2)$$

ทำการปรับปรุงค่าน้ำหนักข้อมูลตัวอย่าง แต่ละตัวจากสมการ (3) และที่  $Z_i$  คือปัจจัยความเป็นปกติ (Normalization Factor) หาได้จากสมการ (4)

$$D_{i+1}(k) = \frac{D_i(k)}{Z_i} x \begin{cases} e^{-\alpha} & \text{if } c_i(x_k) = y_k \\ e^{\alpha} & \text{if } c_i(x_k) \neq y_k \end{cases} \quad (3)$$

$$Z_i = 2\sqrt{\varepsilon_i(1 - \varepsilon_i)} \quad (4)$$

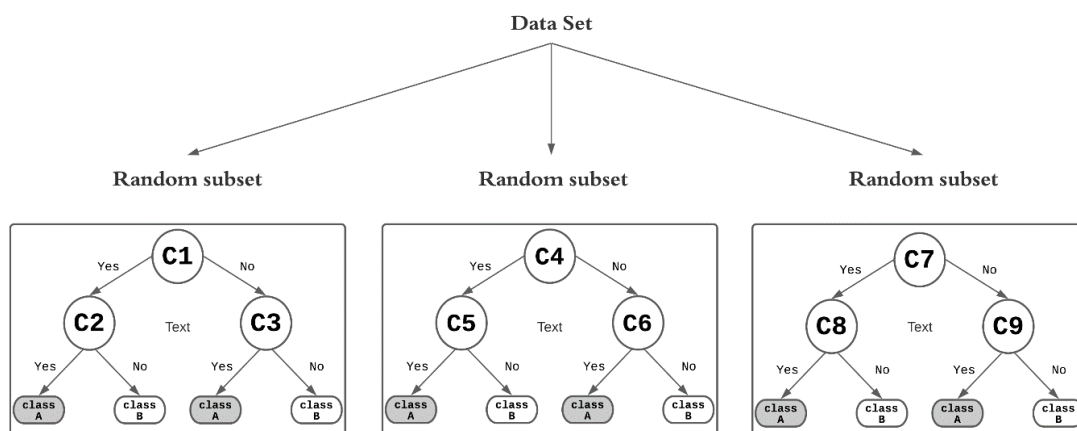
ผลลัพธ์สุดท้ายที่ได้จากกระบวนการจำแนกประเภทข้อมูลด้วยวิธี Boosting แบบ AdaBoost สามารถแสดงได้ดังสมการที่ (5)

$$C(x) = \text{Sign}(\sum_{i=1}^J \alpha_i c_i(x)) \quad (5)$$

โดยที่  $\alpha_i$  คือค่าน้ำหนักของข้อมูลทั้งหมดที่คำนวณได้ในแต่ละรอบจากการคำนวณหาค่าความผิดพลาดในการจำแนกข้อมูล และ  $C_i$  คือตัวจำแนกประเภทข้อมูล

### 3. แรนดอมฟอเรส (Random Forest: RT)

เป็นแบบจำลองที่ถูกจัดให้เป็นอัลกอริทึมการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) และใช้พื้นฐานจากต้นไม้ตัดสินใจเป็นการทำนายแบบชุดของ Decision Tree หลายต้น (Ensemble of Decision Trees) โดยสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับคืน (Random Sampling with Replacement) เพื่อนำมาสร้างเป็นแบบจำลองต้นไม้โดยแต่ละต้นมีลักษณะที่สำคัญ โดยแต่ละแบบจำลองจะมีการทำนายผล ซึ่งผลจากการทำนายของต้นไม้แต่ละต้นจะทำการโหวต เลือกผลการทำนายที่ได้รับการโหวตมากที่สุด รูปที่ 6 Random Forest คือภาพประกอบของ Random Forest ซึ่งประกอบด้วยต้นไม้ตัดสินใจที่แตกต่างกันสามต้น ต้นไม้การตัดสินใจทั้งสามต้นนั้นจะได้รับการฝึกอบรมโดยใช้ชุดย่อยแบบสุ่มของข้อมูลการฝึกอบรม (Train Data) (ธนัท จริยะสมบูรณ์ และ วราภรณ์ วิทยานนท์ 2561)



รูปที่ 6 Random Forest

## มาตรวัดและการทดสอบประสิทธิภาพแบบจำลอง

### 1. การวัดประสิทธิภาพแบบจำลอง

ความสามารถในการวินิจฉัยของตัวจำแนกประเภทถูกกำหนดโดย Confusion Matrix ตารางที่ 1 Confusion Matrix ประกอบไปด้วยตาราง 2X2 มีแนวตั้งคือผลการทำนาย (Prediction class) และแนวนอนคือค่าจริง (Actual class) ภายในตาราง 2X2 (Pangon Gatchalee, 2019) จะประกอบไปด้วย

True Positive (TP) คือความถี่สิ่งที่ทำนายว่า ‘จริง’ และมีค่าเป็น ‘จริง’

True Negative (TN) คือความถี่สิ่งที่ทำนายว่า ‘ไม่จริง’ และมีค่าเป็น ‘ไม่จริง’

False Positive (FP) คือความถี่สิ่งที่ทำนายว่า ‘จริง’ และมีค่าเป็น ‘ไม่จริง’ และ

False Negative (FN) คือความถี่สิ่งที่ทำนายว่า ‘ไม่จริง’ และมีค่าเป็น ‘จริง’

ตาราง 1 Confusion Matrix

		Predicted Class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

โดยทั่วไปมีการวัดประสิทธิภาพของตัวแยกประเภทโดยอิงจาก ตารางที่ 1 Confusion Matrix สามารถนำมาคำนวณเพื่อหาค่าวัดประสิทธิภาพสำหรับตัวจำแนกประเภท ซึ่งประกอบไปด้วย

1.1 ค่าความถูกต้อง (Accuracy) แสดงถึงความถูกต้องในการทำนายในภาพรวมทั้งกลุ่ม Positive และ Negative คำนวณได้ดังนี้

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

1.2 ค่าความครบถ้วน (Recall) คือ อัตราส่วนระหว่างจำนวน positive ที่ถูกจำแนกได้อย่างถูกต้อง คำนวณได้ดังนี้

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

1.3 ค่าความแม่นยำ (Precision) คือ อัตราส่วนความถูกต้องของการทำนายกลุ่ม Positive เมื่อเทียบกับผลการทำนาย Positive ทั้งหมด คำนวณได้ดังนี้

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

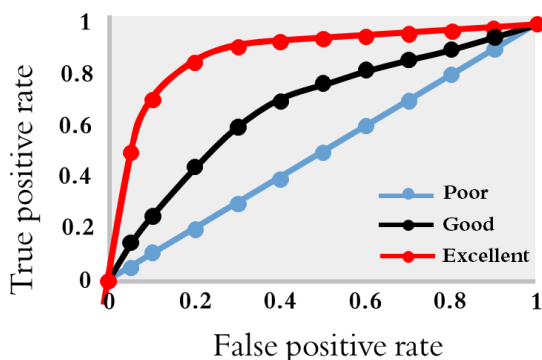
1.4 ค่าความถ่วงดุล (F1 score) เป็นการวัดความถูกต้องโดยการใช้ค่าเฉลี่ยฮาร์โมนิระหว่าง True Positive Rate และ Precision คำนวณได้ดังนี้

$$F_1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

## 2. Receiver Operating Characteristic (ROC) curve

Receiver Operating Characteristic (ROC) curve เป็นเครื่องมือพื้นฐานสำหรับการประเมินการทดสอบวินิจฉัย และสร้างขึ้นโดย True Positive Rate (1-Specificity) เทียบ False Positive Rate (Sensitivity) โดยทั่วไปแล้วเส้นโค้ง ROC (AUC) ยังใช้กำหนดความสามารถในการคาดการณ์ของตัวแยกประเภท รูปที่ 7 ROC curve แสดงการนำเสนอเส้นโค้ง ROC สามเส้นตามชุดข้อมูล ค่า AUC สำหรับเส้นโค้ง ROC สีแดงมีความแม่นยำในการคาดการณ์ที่ดีมาก เส้นโค้ง ROC สีดำมีความแม่นยำในการคาดการณ์ที่ดี และเส้นโค้ง ROC สีฟ้ามีความแม่นยำในการคาดการณ์ที่ต่ำ โดยที่พื้นที่ใต้เส้นโค้ง ROC สีฟ้าคือครึ่งหนึ่งของสี่เหลี่ยมผืนผ้าที่แรงเงา ค่า AUC สำหรับเส้นโค้ง ROC สีฟ้าคือ 0.5

ดังนั้นตัวแยกประเภทที่สร้างเส้นโค้ง ROC สีแดงจะมีความแม่นยำในการคาดการณ์ที่สูงกว่าเมื่อเทียบกับตัวแยกประเภทอื่นที่สร้างเส้นโค้ง ROC สีดำ และสีฟ้า (Uddin et al, 2019)



รูปที่ 7 ROC curve



### 3. การแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง

Holdout คือ การแบ่งชุดข้อมูลออกเป็น 2 ชุดข้อมูลย่อยด้วยวิธีการสุ่ม โดยชุดข้อมูลที่ได้จะเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ ซึ่งโดยปกติชุดข้อมูลฝึกสอนจะมีปริมาณข้อมูลเท่ากับ 2 ใน 3 ของชุดข้อมูลทั้งหมด และชุดข้อมูลทดสอบจะมีปริมาณ 1 ใน 3 ของชุดข้อมูล หลังจากแบ่งชุดข้อมูลแล้วจะนำชุดข้อมูลฝึกสอนจะถูกใช้ในการสร้างตัวจำแนกข้อมูล และชุดข้อมูลทดสอบจะถูกใช้ในการทดสอบตัวจำแนกที่สร้างขึ้น

Cross Validation คือ การแบ่งข้อมูลเพื่อใช้สำหรับวัดประสิทธิภาพแบบจำลอง โดยใช้วิธีการ Cross Validation ซึ่งเป็นวิธี ที่ได้รับความนิยมสำหรับการแบ่งข้อมูล เพื่อประสิทธิภาพแบบจำลอง เนื่องจากผลลัพธ์ที่ได้มีความน่าเชื่อถือ โดยหลักในการแบ่งข้อมูลด้วยวิธีนี้จะเริ่มจากการกำหนดค่า  $k$  หรือการแบ่งชุดข้อมูลออกเป็น  $k$  ส่วนเท่า ๆ กัน

### งานวิจัยที่เกี่ยวข้อง

1. การทำนายผลการลาออกของนักศึกษา โดยทำการแบ่งเป็น ‘ลาออก’ และ ‘ไม่ลาออก’ และทำความเข้าใจข้อมูล แปลงข้อมูล สกัดคุณลักษณะ และปรับสมดุลของข้อมูล ซึ่งเป็นการเตรียมข้อมูลสำหรับขั้นตอนการจำแนกประเภทสำหรับการลาออกของนักศึกษา โดยได้ทำการเปรียบเทียบอัลกอริทึม Decision Tree, Random Forest และ Gradient Boosting ผลการทดลองพบว่า การใช้อัลกอริทึม Gradient Boosting ให้ประสิทธิภาพความแม่นยำดีที่สุดในการจำแนกประเภทการลาออกของนักศึกษา (93%) และคุณลักษณะที่สำคัญได้แก่ ปีการศึกษาของนักเรียน เกรดเฉลี่ยของโรงเรียนมัธยม ช่องทางการรับเข้ามหาวิทยาลัย คณาจารย์ของนักเรียน และเพศ (Tenpipat, & Akkarajitsakul, 2020)

2. การใช้แบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มเพื่อทำนายการลาออกของนักศึกษา โดยทำการปรับขอบเขตข้อมูลโดยใช้วิธี Min-Max Normalization ให้อยู่ในช่วง  $[0,1]$  สำหรับขั้นตอนการจำแนกประเภทของการลาออกของนักศึกษา โดยใช้อัลกอริทึม Random Forest และการใช้ K-Fold cross validation ที่  $K = 5, 10$  ในการวัดประสิทธิภาพของตัวแบบ ผลการทดลองพบว่า การใช้ Random Forest ที่  $K=5, 10$  ให้ประสิทธิภาพความแม่นยำในการจำแนกการลาออกของนักศึกษาเท่ากับ 81.77% และ 80.11% ตามลำดับ (Naseem et al., 2020)

3. ระบบการทำนายในการพัฒนาของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล ได้นำเสนอเทคนิคต้นไม้ตัดสินใจ (Decision Tree) มาช่วยในการสร้างกฎเพื่อจะพัฒนาระบบการทำนายการพัฒนาของนักศึกษา โดยใช้ข้อมูลคุณลักษณะ สาขาวิชาที่ศึกษาในคณะวิทยาศาสตร์ เกรดเฉลี่ยในภาคเรียนที่ 1-6 เกรดเฉลี่ยจากโรงเรียนมัธยม แผนการเรียนที่ศึกษาในโรงเรียนมัธยม ขนาดโรงเรียน สถานะกู้ยืมเพื่อการศึกษา สถานะการพ้น

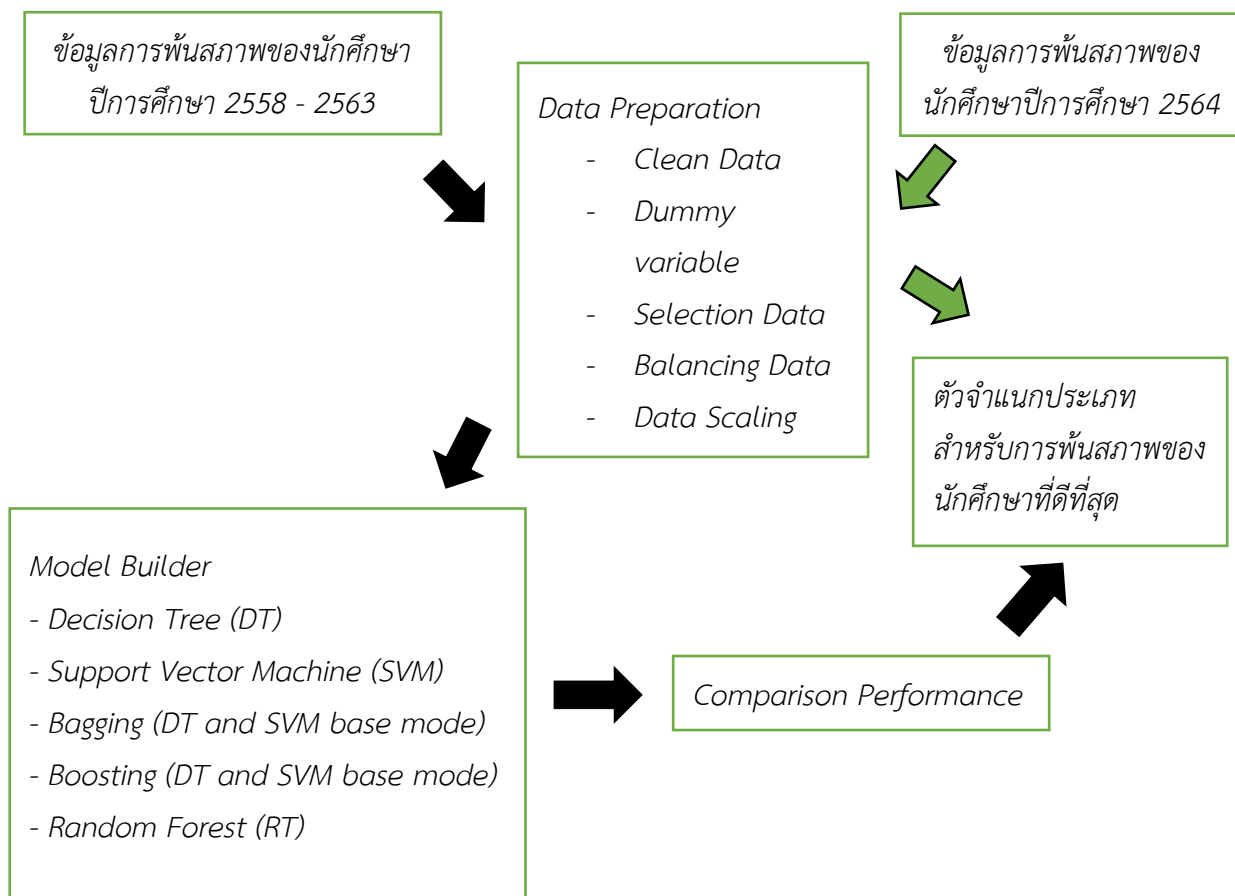
สภาพ โดยรวมมีตัวแปรที่เกี่ยวข้องทั้งหมด 7 ตัวแปร ผลลัพธ์จากงานวิจัยพบว่า รูปแบบการทำนายการฟื้นสภาพด้วยวิธีต้นไม้ตัดสินใจมีจำนวน 32 กฎ ประเมินโดยใช้ 10-Folds Cross Validation มีค่าความถูกต้องเฉลี่ย 95.57% (นนทวัฒน์ ทวีชาติ และคณะ, 2563)

4. การวิเคราะห์ปัจจัยที่มีผลต่อการฟื้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาหลักสูตรวิทยาการคอมพิวเตอร์หลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา ได้มีการนำเสนอเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (Back Propagation Neural Network : BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มาทำการเปรียบเทียบด้วย 10-Fold Cross Validation จากผลลัพธ์การทดลองพบว่า SVM ให้ประสิทธิภาพในการจำแนกประเภทโดยเฉลี่ยสูงสุด 97.75 % โดยแต่ละโมเดลให้ความแม่นยำเฉลี่ยมากกว่า 97% และปัจจัยที่ส่งผลต่อการฟื้นสภาพของนักศึกษามี ผลการเรียนรู้รายวิชาพื้นฐานทางฟิสิกส์ วิชาแพลตฟอร์มเทคโนโลยี ผลการเรียนรู้เฉลี่ยปีที่สอง และผลการเรียนรายวิชาโครงสร้างข้อมูล (ขอและ เกป็น, พิมลพรรณ สีสากัทรพันธุ์, และอัจฉราพร ยกขุน, 2561)

ตาราง 2 เปรียบเทียบวิธีการสร้างแบบจำลอง ความแม่นยำ และปัจจัยที่มีความสำคัญกับการพัฒนาของนักศึกษาระหว่างการศึกษาต่างๆ

ผู้นิพนธ์	Machine Learning Model	Train-test Data	ความแม่นยำ (%)	ปัจจัยการการพัฒนาของนักศึกษา
Tenpipat, & Akkarajitsakul, 2020	1) Decision Tree 2) Random Forest 3) Gradient Boosting	1) 10-Fold CV	Gradient Boosting ให้ประสิทธิภาพความแม่นยำโดยเฉลี่ยดีที่สุด 93%	ปีการศึกษาของนักเรียน, เกรดเฉลี่ย, เกรดเฉลี่ยระดับมัธยมศึกษา, ช่องทางการรับเข้ามหาวิทยาลัย, คณาจารย์ของนักเรียน, และเพศ
Naseem et al., 2020	1) Random Forest	1) 5-Fold CV 2) 10-Fold CV	Random Forest ที่ K=5,10 ให้ประสิทธิภาพความแม่นยำเท่ากับ 81.77% และ 80.11% ตามลำดับ	สถานะการณได้รับทุน, เพศ, อายุ, เกรด, คะแนนงานต่างๆ
นนทวัฒน์ ทวีชาติ และคณะ, 2563	1) Decision Tree	1) 10-Folds CV	วิธีต้นไม้ตัดสินใจมีจำนวน 32 กฎ และมีค่าความถูกต้องเฉลี่ย 95.57%	สาขาวิชา, เกรดเฉลี่ยในภาคเรียนที่ 1-6, เกรดเฉลี่ยจากโรงเรียนมัธยม, แผนการเรียนที่ศึกษาในโรงเรียนมัธยม, ขนาดโรงเรียน, สถานะกู้ยืมเพื่อการศึกษา และสถานะการพัฒนสภาพ
ซอและ เกป็น, พิมพ์พรรณ และคณะ, 2561	1) Decision Tree 2) Back Propagation Neural Network 3) Support Vector Machine	1) 10-Fold Cross Validation	SVM ให้ประสิทธิภาพในการจำแนกประเภทโดยเฉลี่ยสูงที่สุด 97.75 %	ผลการเรียนรายวิชาพื้นฐานทางฟิสิกส์, วิชาแคลคูลัส, เทคโนโลยี, ผลการเรียนเฉลี่ยปีที่สอง และผลการเรียนรายวิชาโครงสร้างข้อมูล

### กรอบแนวคิดงานวิจัย



รูปที่ 8 กรอบแนวคิดการวิจัย

จากรูปที่ 8 แสดงกรอบแนวคิดการวิจัยโดยมีการนำข้อมูลการพัฒนสภาพของนักศึกษาปีการศึกษา 2558 – 2563 และ 2564 มาเข้าสู่กระบวนการในการเตรียมความพร้อมของข้อมูล เพื่อที่จะสร้างตัวแบบในการจำแนกประเภทสำหรับการพัฒนสภาพของนักศึกษา หลังจากนั้นจะทำการเปรียบเทียบแต่ละโมเดลในการเลือกโมเดลที่มีประสิทธิภาพมากที่สุด ผลลัพธ์สุดท้ายจะได้ตัวแบบจำแนกประเภทสำหรับการพัฒนสภาพของนักศึกษาที่ใช้สำหรับการทำนายข้อมูลการพัฒนสภาพของศึกษาปีการศึกษา 2564

### บทที่ 3

#### วิธีการดำเนินการวิจัย

การศึกษาวิจัยเรื่อง “การจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้วิธีการเรียนรู้แบบรวมกลุ่ม” ได้ทำการเปรียบเทียบระหว่างวิธีการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และแบบจำลองการจำแนกแบบรวมกลุ่ม ผู้วิจัยได้ทำการศึกษาแนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง เพื่อพัฒนาประสิทธิภาพการจำแนกประเภท พัฒนาประสิทธิภาพการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้เทคนิควิธีการเรียนรู้แบบรวมกลุ่ม โดยได้กำหนดวิธีการดำเนินงาน วิจัยซึ่งประกอบด้วยเนื้อหาดังต่อไปนี้

1. เครื่องมือที่ใช้ในงานวิจัย
2. กรอบวิธีการดำเนินการวิจัย
  - 2.1 การเตรียมข้อมูลสำหรับการทดลอง
  - 2.2 การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว
  - 2.3 การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม
  - 2.4 เปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภท
  - 2.5 ทำนายการพัฒนาของนักศึกษาปีการศึกษา 2564

#### เครื่องมือที่ใช้ในการวิจัย

1. เทคนิคที่ใช้ในการเหมืองข้อมูล (Data Mining Tool)

การจัดประเภท (Classification) เป็นการจัดประเภทของข้อมูล โดยหาตัวแบบการจำแนกประเภทข้อมูล ซึ่งตัวแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยเป็นกลุ่มข้อมูลที่มีการระบุกลุ่มผลลัพธ์เรียบร้อยแล้ว วัตถุประสงค์เพื่อใช้เป็นตัวแบบในการทำนายข้อมูลที่ไม่เคยเห็นมาก่อน ในการวิจัยครั้งนี้ได้ใช้อัลกอริทึมดังนี้

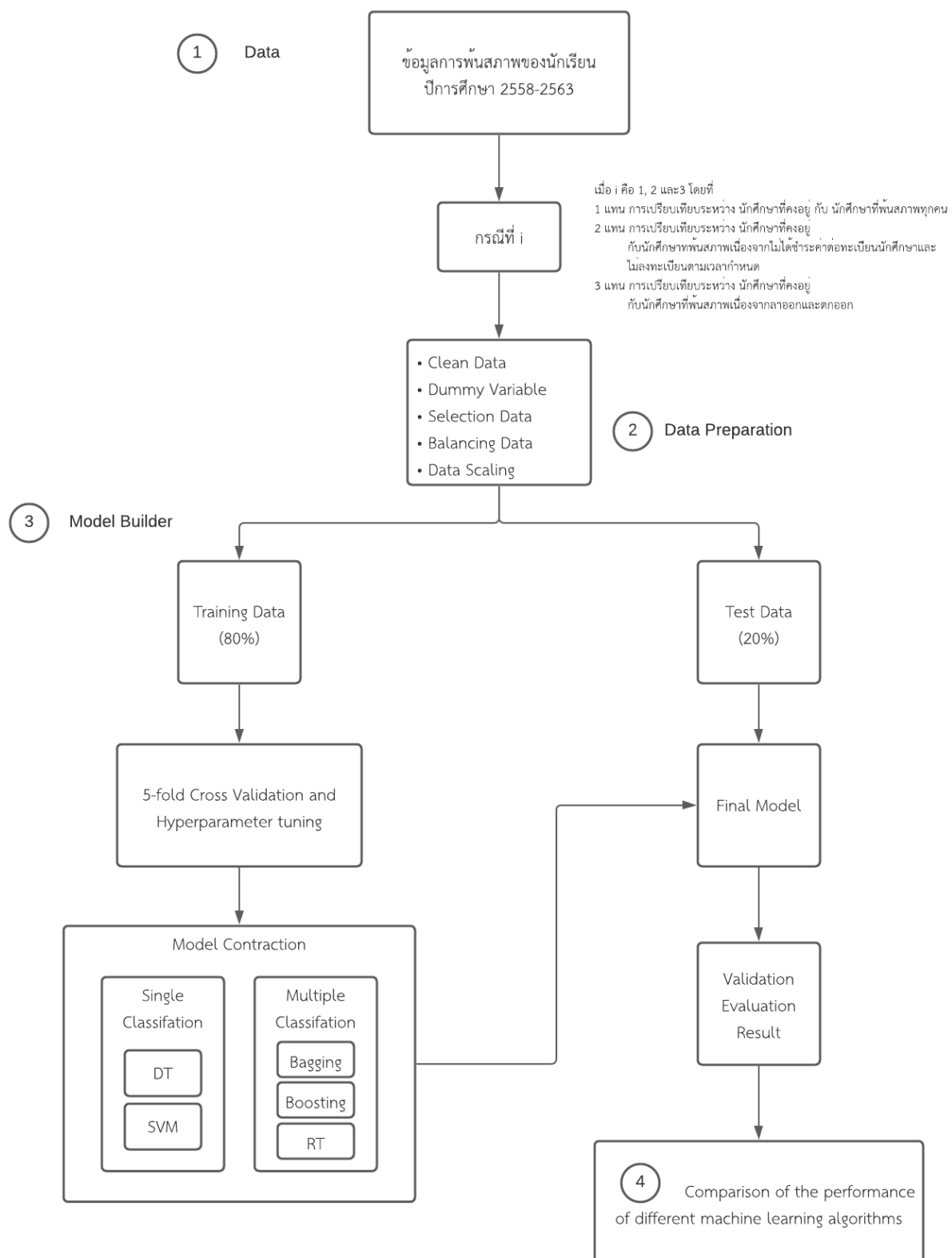
- 1.1 Decision Tree (DT)
- 1.2 Support Vector Machine (SVM)
- 1.3 Bagging (DT base mode)
- 1.4 Bagging (SVM base mode)
- 1.5 Boosting (DT base mode)
- 1.6 Boosting (SVM base mode)
- 1.7 Random Forest (RT)

### 1.2 โปรแกรมที่ใช้ในการประมวลผล (Software Tools)

ผู้วิจัยได้ใช้โปรแกรมทางคอมพิวเตอร์ Google Colaboratory ในการประมวลผลทางข้อมูล ซึ่งเป็นโปรแกรมที่นักวิทยาศาสตร์ทางข้อมูลนิยมใช้กันเป็นจำนวนมากในการทำเหมืองข้อมูล และจัดการกับข้อมูลที่มีขนาดใหญ่ ในการวิจัยครั้งนี้ผู้วิจัยได้ใช้ Library Pandas ในการจัดการกับข้อมูลที่มีขนาดใหญ่ และ Library Scikit-Learn ในการเลือกใช้อัลกอริทึมต่างๆในการสร้างตัวจำแนกประเภท

### กรอบวิธีการดำเนินการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้เสนอกรอบวิธีการดำเนินงานวิจัย รวมถึงการออกแบบการทดลองโดยแบ่งออกเป็น 3 ส่วนหลักๆ แสดงดังรูปที่ 9 ในส่วนที่ 1 การจัดเตรียมข้อมูลการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2558 – 2564 จำนวน 719 คน โดยได้ทำการแบ่งชุดข้อมูลที่ใช้สำหรับการทดลองออกเป็น 3 กรณี เพื่อใช้สำหรับการทดลองเป็นกระบวนการในการจัดเตรียมข้อมูลก่อนการสร้างโมเดลโดยจะนำข้อมูลมาทำความสะอาดข้อมูล ตัวแปรหุ่น ทำการคัดเลือกข้อมูล ปรับสมดุลของข้อมูล และการปรับปรุงขอบเขตข้อมูล หลังจากผ่านขั้นตอนกระบวนการเตรียมข้อมูลในส่วนที่ 1 ได้นำข้อมูลมาแบ่งเป็นข้อมูลฝึกสอน (Training Data) 80% และข้อมูลทดสอบ (Test Data) 20% จากนั้นนำข้อมูลฝึกสอน (Training Data) แบ่งข้อมูลออกเป็น 5 ชุดโดยใช้วิธี Cross Validation สำหรับการปรับพารามิเตอร์ของโมเดลให้เหมาะสมที่สุดโดยใช้ Grid Search หลังจากนั้นนำข้อมูลที่ได้เตรียมไว้เข้าสู่กระบวนการเรียนรู้การจำแนกประเภท ในขั้นตอนนี้ได้ใช้อัลกอริทึม Decision Tree และ Support Vector Machine เรียนรู้ข้อมูลตัวอย่างในแต่ละชุดเพื่อใช้สำหรับการจำแนก ในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และทำการเปรียบเทียบการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม 3 วิธีได้แก่ Bagging Boosting และ Random Forest โดยใช้อัลกอริทึม Decision Tree และ Support Vector Machine เป็นอัลกอริทึมพื้นฐานใช้งานร่วมกับวิธีการเรียนรู้แบบรวมกลุ่ม 2 วิธี คือ Bagging และ Boosting ส่วนวิธีการของ Random Forest จะใช้ต้นไม้ช่วยตัดสินใจซึ่งเป็นอัลกอริทึมพื้นฐานในการเรียนรู้ของตัวอัลกอริทึมเอง จากนั้นใช้ข้อมูลทดสอบ (Test Data) ที่อัลกอริทึมไม่เคยได้เห็นหรือยังไม่ได้เรียนรู้สำหรับวัดประสิทธิภาพแบบจำลองแต่ละแบบจำลอง เพื่อให้ได้แบบจำลองที่มีความน่าเชื่อถือ และในลำดับสุดท้ายจะดำเนินการเปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภทที่ได้จากการทดลองทั้งหมด ซึ่งจะอธิบายแต่ละส่วนดังนี้



รูปที่ 9 กรอบวิธีการดำเนินงานวิจัยสำหรับการทำแผนประเภท



## 1. การเตรียมข้อมูลสำหรับการสร้างแบบจำลองการจำแนกประเภท

### 1.1 ชุดข้อมูล

ในการวิจัยครั้งนี้ได้รับข้อมูลตัวอย่างการพัฒนาศักยภาพของนักศึกษาระดับปริญญาตรี จากฐานข้อมูล สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น โดยใช้ข้อมูลตั้งแต่ปีพ.ศ. 2558-2563 จำนวน 614 คน และปีพ.ศ. 2564 จำนวน 105 คน แสดงในตารางที่ 3 โดยได้ทำการแบ่งชุดข้อมูลที่ใช้สำหรับการทดลองออกเป็น 3 กลุ่มตัวอย่างดังนี้

#### 1.1.1 ข้อมูลกรณีที่ 1

การพัฒนาศักยภาพของนักศึกษาประกอบไปด้วย พัฒนาศักยภาพเนื่องจากลาออก พัฒนาศักยภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา พัฒนาศักยภาพเนื่องจากตกออก และพัฒนาศักยภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด

#### 1.1.2 ข้อมูลกรณีที่ 2

การพัฒนาศักยภาพของนักศึกษาประกอบไปด้วย พัฒนาศักยภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา และพัฒนาศักยภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด

#### 1.1.3 ข้อมูลกรณีที่ 3

การพัฒนาศักยภาพของนักศึกษาประกอบไปด้วย พัฒนาศักยภาพเนื่องจากลาออก และพัฒนาศักยภาพเนื่องจากตกออก

ตาราง 3 ชื่อคุณลักษณะและคำนิยาม

Attribute Name	Definition
เพศ	0 = หญิง 1 = ชาย
เกรดเฉลี่ยของโรงเรียนมัธยม	คะแนนเฉลี่ยสะสมของโรงเรียนมัธยม
เกรดเฉลี่ยปัจจุบัน	คะแนนเฉลี่ยสะสมปัจจุบัน
ช่องทางการรับเข้ามหาวิทยาลัย	0 = สอบคัดเลือกจากระบบกลาง (Admissions) 1 = สอบคัดเลือกประเภทโควตาภาคตะวันออกเฉียงเหนือ 2 = โควตาทั่วประเทศ 3 = โครงการรับนักเรียนที่เป็นผู้มีคุณธรรมจริยธรรม และบริการสังคม

Attribute Name	Definition
	<p>4 = โครงการร่วมรับนักศึกษาภาคใต้กับมหาวิทยาลัยสงขลานครินทร์ โดยวิธีรับตรง</p> <p>5 = การคัดเลือกโดยวิธีพิเศษ</p> <p>6 = โครงการร่วมรับนักศึกษาภาคเหนือกับมหาวิทยาลัยเชียงใหม่ โดยวิธีรับตรง</p> <p>7 = โครงการมูลนิธิส่งเสริมโอลิมปิกวิชาการฯ (สอวน.)</p>
สถานะครอบครัว	<p>0 = อยู่ด้วยกัน</p> <p>1 = หย่าขาดจากกัน</p> <p>2 = บิดาถึงแก่กรรม</p> <p>3 = แยกกันอยู่เพราะเหตุผลอื่นๆ</p> <p>4 = แยกกันอยู่เพราะความจำเป็นเกี่ยวกับอาชีพ</p> <p>5 = มารดาถึงแก่กรรม</p> <p>6 = บิดามารดาถึงแก่กรรม</p>
สาขา	<p>0 = สารสนเทศสถิติ</p> <p>1 = สถิติ</p> <p>2 = สถิติและวิทยาการข้อมูล สาขาวิชาเอก สารสนเทศสถิติและวิทยาการข้อมูล</p> <p>3 = สถิติและวิทยาการข้อมูล สาขาวิชาเอก สถิติศาสตร์</p>
รายได้ปัจจุบันของมารดา	รายได้ปัจจุบันของมารดาต่อเดือน
รายได้ปัจจุบันของบิดา	รายได้ปัจจุบันของบิดาต่อเดือน
อายุบิดา	อายุบิดาปัจจุบัน
อายุมารดา	อายุมารดาปัจจุบัน
บุตรคนที่	บุตรคนที่
จำนวนพี่น้อง	จำนวนพี่น้องที่กำลังศึกษาอยู่
เกรดรายวิชา ปี 1 เทอม 1	<p>0 = A</p> <p>1 = B+</p> <p>2 = B</p> <p>3 = C+</p>

Attribute Name	Definition
	<p>4 = C</p> <p>5 = D+</p> <p>6 = D</p> <p>7 = F</p> <p>8 = S</p> <p>9 = W</p> <p>10 = S AU</p> <p>11 = U</p>
เกรดรายวิชา ปี 1 เทอม 2	<p>0 = A</p> <p>1 = B+</p> <p>2 = B</p> <p>3 = C+</p> <p>4 = C</p> <p>5 = D+</p> <p>6 = D</p> <p>7 = F</p> <p>8 = S</p> <p>9 = W</p> <p>10 = S AU</p> <p>11 = U</p>
สถานภาพ	<p>0 = นักศึกษาปัจจุบัน สถานะปกติ</p> <p>1 = สำเร็จการศึกษา</p> <p>2 = พ้นสภาพเนื่องจากลาออก</p> <p>3 = พ้นสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา</p> <p>4 = พ้นสภาพเนื่องจากตกออก</p> <p>5 = พ้นสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด</p> <p>6 = ลาพักการเรียน</p> <p>7 = มารายงานตัวแล้ว</p>

## 1.2 การทำความสะอาดข้อมูล (Clean Data)

การทำความสะอาดข้อมูล คือการลดข้อมูลที่ผิดปกติ และข้อมูลที่สูญหาย เพื่อให้ข้อมูลมีความสมบูรณ์และพร้อมสำหรับการทดลอง ซึ่งใช้วิธี KNNImputer ในการประมาณค่าข้อมูลที่สูญหายสามารถประมาณค่าได้โดยใช้จุดข้อมูลที่ใกล้ที่สุดในตัวแปรอื่นๆ และพิจารณาความสัมพันธ์ของข้อมูล ซึ่ง KNNImputer สามารถการประมาณค่าข้อมูลที่สูญหายได้ทั้งตัวแปรที่เป็น numerical และ categorical

1.2.1 ตัวแปรรายได้บิดา รายได้มารดา อายุบิดาอายุมารดา บุตรคนที่ และจำนวนพี่น้องที่มีข้อมูลที่ผิดปกติคือ {“ไม่แน่นอน, ไม่ทราบ, -, 15,000บาท, 40,000 - 50,000”} ในข้อมูลที่มีการกรอกรายได้จะทำการหาค่าเฉลี่ยส่วนข้อมูลที่กรอกรมา 15,000บาท ทำการลบข้อมูลที่เป็นหนังสือ และนอกเหนือจากนี้ข้อมูลที่กรอกร {“ไม่แน่นอน, ไม่ทราบ, -”} แทนเป็นค่าสูญหาย (Missing value)

1.2.2 เนื่องจากตัวแปรรายได้บิดามารดา อายุบิดามารดา บุตรคนที่และจำนวนพี่น้องที่มีข้อมูลที่ค่าสูญหาย (Missing value) ซึ่งจะทำการประมาณค่าโดยใช้ KNNImputer

## 1.3 ตัวแปรหุ่น (Dummy Variable)

ตัวแปรหุ่น (Dummy Variable) คือ ตัวแปรที่ถูกกำหนดให้มีสองค่า (Binary) คือ 0 และ 1 โดยจะทำการแปลงข้อมูลเกรดรายวิชาปี 1 เทอม 1 และเกรดรายวิชาปี 1 เทอม 2 ให้เป็นตัวแปรหุ่น เพื่อต้องการศึกษาเกรดในรายวิชาต่างๆเป็นปัจจัยต่อการพัฒนาหรือไม่

## 1.4 การคัดเลือกข้อมูล (Selection Data)

ในขั้นตอนการการคัดเลือกข้อมูล คือการนำข้อมูลเพศ ช่องทางการรับเข้ามหาวิทยาลัย สถานะครอบครัว สาขา เกรดรายวิชาปี 1 เทอม 1 และเกรดรายวิชาปี 1 เทอม 2 ที่เป็นตัวแปรหุ่น (Dummy Variable) มาทำการคัดเลือกข้อมูลที่มีความสำคัญกับสถานภาพโดยใช้ค่าสถิติ Chi-Square ในการทดสอบ Test of Independence เพื่อที่จะตัดและลดขนาดของข้อมูลหรือคุณลักษณะที่ไม่จำเป็นในการวิเคราะห์

## 1.5 ปรับสมดุลของข้อมูล (Balancing Data)

ในการวิจัยครั้งนี้ได้ใช้วิธีสุ่มเกิน (Over sampling) ในการปรับสมดุลของข้อมูล คือการเพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อยให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมาก ซึ่งการเพิ่มข้อมูลนั้นจะเพิ่มโดยการสุ่มเลือกจากข้อมูลเดิมในกลุ่มส่วนน้อยโดยใช้วิธีการสุ่มแบบเป็นระบบ

### 1.6 การปรับปรุงขอบเขตข้อมูล (Data Scaling)

นำข้อมูลที่ผ่านมากระบวนการคัดเลือกข้อมูล ทำความสะอาดข้อมูล และปรับสมดุลของข้อมูล มาทำการปรับปรุงขอบเขตข้อมูลที่เป็นข้อมูลเชิงปริมาณ โดยใช้วิธี Min-Max Normalization ทำให้ข้อมูลอยู่ในช่วง  $[0,1]$

## 2. การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว

ในขั้นตอนการสร้างแบบจำลองการจำแนกประเภทประเภทเดี่ยว โดยใช้อัลกอริทึม Decision Tree และ Support Vector Machine เพื่อหาประสิทธิภาพความถูกต้องหรือแม่นยำ การสำหรับการจำแนกประเภท ซึ่งจะแสดงรายละเอียดดังต่อไปนี้

### 2.1 การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม Decision Tree (DT) และ Support Vector Machine (SVM)

2.1.1 นำข้อมูลการพันสภาพกรณี 1 2 และ 3 ที่ผ่านกระบวนการเตรียมข้อมูลมาทำการแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลฝึกสอน (Training Data) 80% และข้อมูลทดสอบ (Test Data) 20%

2.1.2 นำข้อมูลฝึกสอน (Training Data) มาทำการแบ่งข้อมูลให้เป็น 5 ส่วน ข้อมูลทุกส่วนจะถูกเรียนรู้และตรวจสอบในการวัดประสิทธิภาพ ด้วยวิธี 5-fold Cross validation และทำการหาค่าพารามิเตอร์ที่เหมาะสมโดยใช้ Grid Search สำหรับอัลกอริทึม Decision Tree (DT) และ Support Vector Machine (SVM) ดังตารางที่ 4 - 5

2.1.3 กำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับอัลกอริทึม Decision Tree (DT) และ Support Vector Machine (SVM) หลังจากนั้นทำการเรียนรู้โดยใช้ข้อมูลทดสอบ (Test Data)

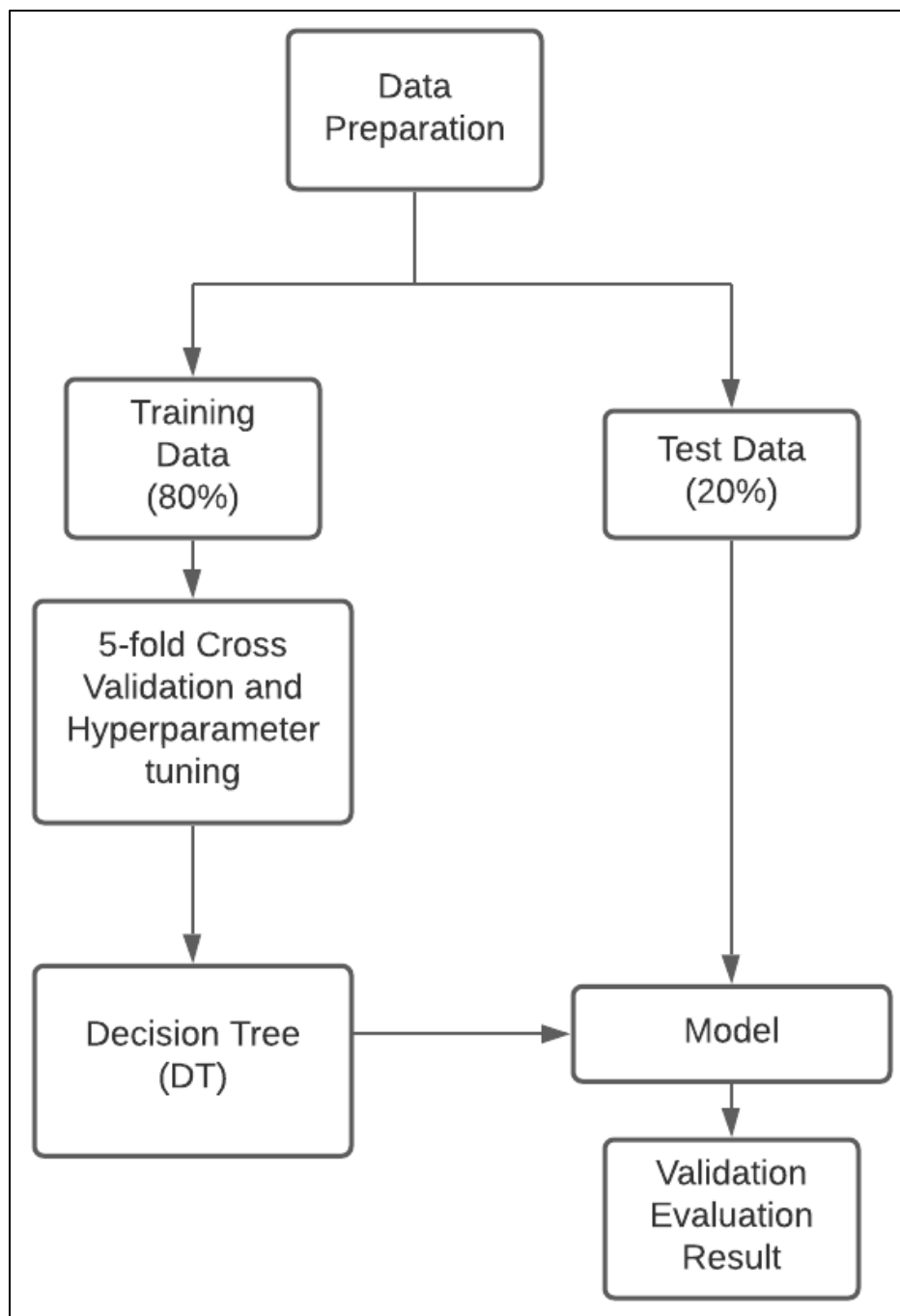
2.1.4 ผลลัพธ์สุดท้ายของการประเมินการตรวจสอบความถูกต้องของกระบวนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม Decision Tree และ Support Vector Machine จะได้ค่าวัดประสิทธิภาพของโมเดล คือ Accuracy, Precision, F1-Measure, recall, และ ROC curve ดังรูปที่ 10 - 11

ตาราง 4 กำหนดพารามิเตอร์ให้กับอัลกอริทึม Decision Tree (DT)

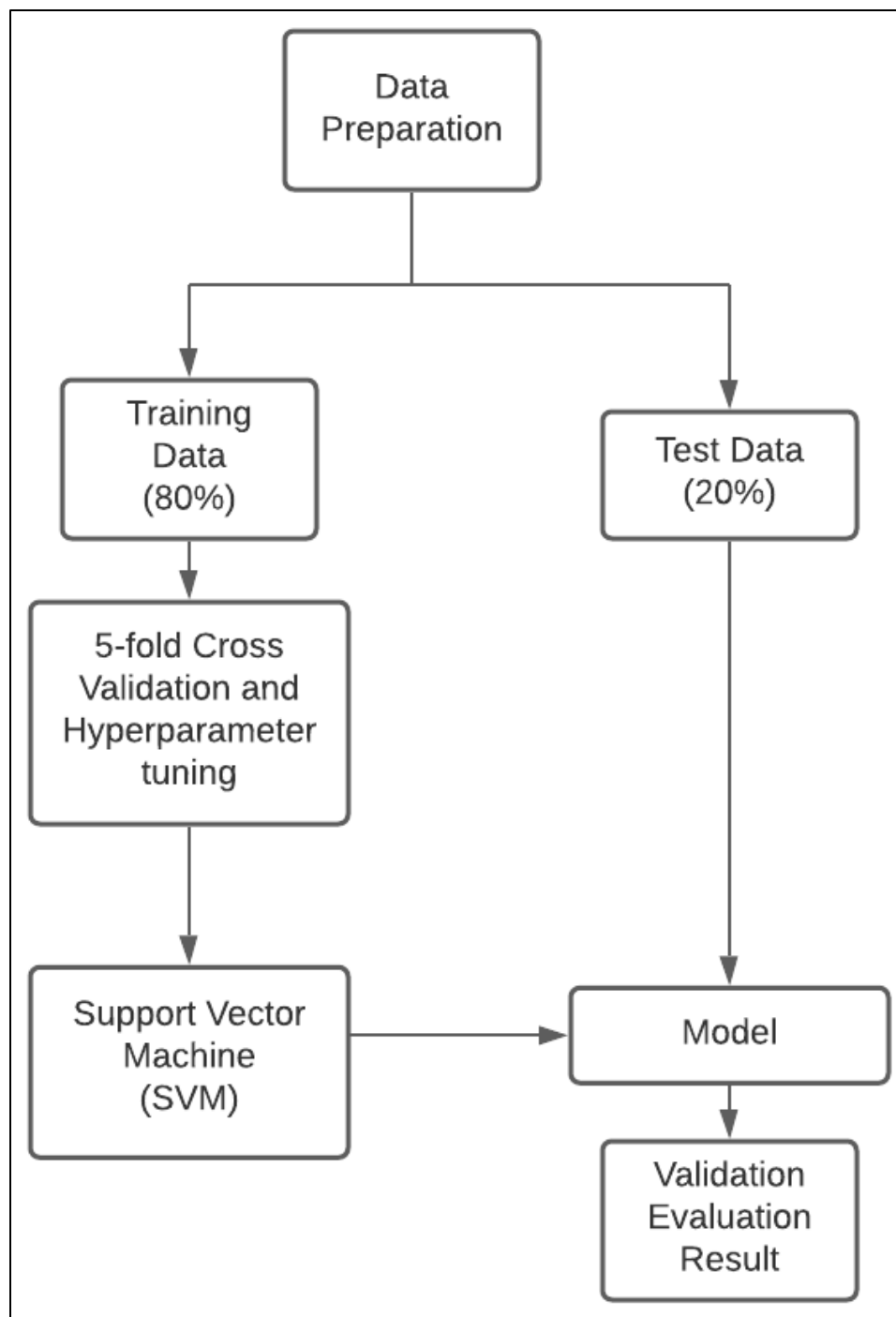
พารามิเตอร์	ค่าพารามิเตอร์
<i>Criterion</i>	<i>Gini, Entropy</i>
<i>Max Depth</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
<i>Min Samples_Leaf</i>	2, 5, 10
<i>Max Features</i>	1, 2, 4, 6, 7, 8, 13, 15, 16, 17, 19

ตาราง 5 กำหนดพารามิเตอร์ให้กับอัลกอริทึม Support Vector Machine (SVM)

พารามิเตอร์	ค่าพารามิเตอร์
<i>C</i>	0.1, 1, 10, 100
<i>Gamma</i>	1, 0.1, 0.01, 0.001
<i>Kernel</i>	<i>Rbf, Poly, Sigmoid</i>



รูปที่ 10 ขั้นตอนการแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ DT



รูปที่ 11 ขั้นตอนการแบ่งจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ SVM



### 3. การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม

ในขั้นตอนวิธีการสร้าง Multiple Classification โดยใช้อัลกอริทึม Decision Tree (DT) และ Support Vector Machine (SVM) เป็นอัลกอริทึมพื้นฐานที่ได้กำหนดพารามิเตอร์ที่เหมาะสมไว้แล้วในขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวใช้งานร่วมกับร่วมกับวิธีการเรียนรู้แบบรวมกลุ่ม 2 วิธี คือ Bagging และ Boosting ส่วนวิธีการของ Random Forest จะใช้ต้นไม้ช่วยตัดสินใจซึ่งเป็นอัลกอริทึมพื้นฐานในการเรียนรู้ของตัวอัลกอริทึมเอง

#### 3.1 การสร้างแบบจำลอง Bagging โดยใช้ Decision Tree (DT) และ Support Vector Machine (SVM) เป็นอัลกอริทึมพื้นฐาน

3.1.1 นำข้อมูลการผันสภาพกรณี 1 2 และ 3 ที่ผ่านกระบวนการเตรียมข้อมูลมาทำการแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลฝึกสอน (Training Data) 80% และข้อมูลทดสอบ (Test Data) 20%

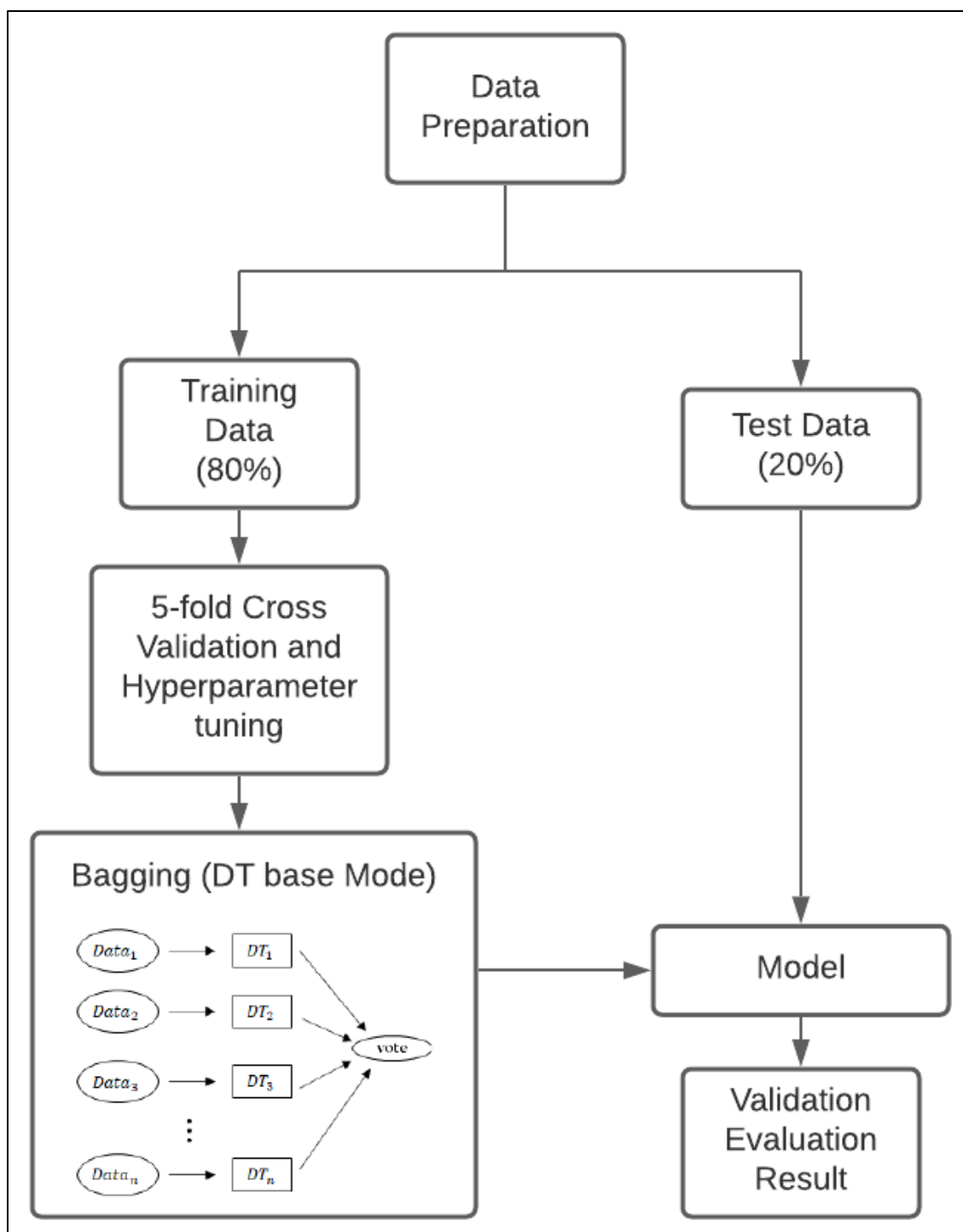
3.1.2 นำข้อมูลฝึกสอน (Training Data) มาทำการแบ่งข้อมูลให้เป็น 5 ส่วน ข้อมูลทุกส่วนจะถูกเรียนรู้และตรวจสอบในการวัดประสิทธิภาพ ด้วยวิธี 5-fold Cross validation และทำการหาค่าพารามิเตอร์ที่เหมาะสมโดยใช้ Grid Search สำหรับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Bagging โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ดังตารางที่ 6

3.1.3 กำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Bagging โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้หลังจากนั้นทำการเรียนรู้โดยใช้ข้อมูลทดสอบ (Test Data)

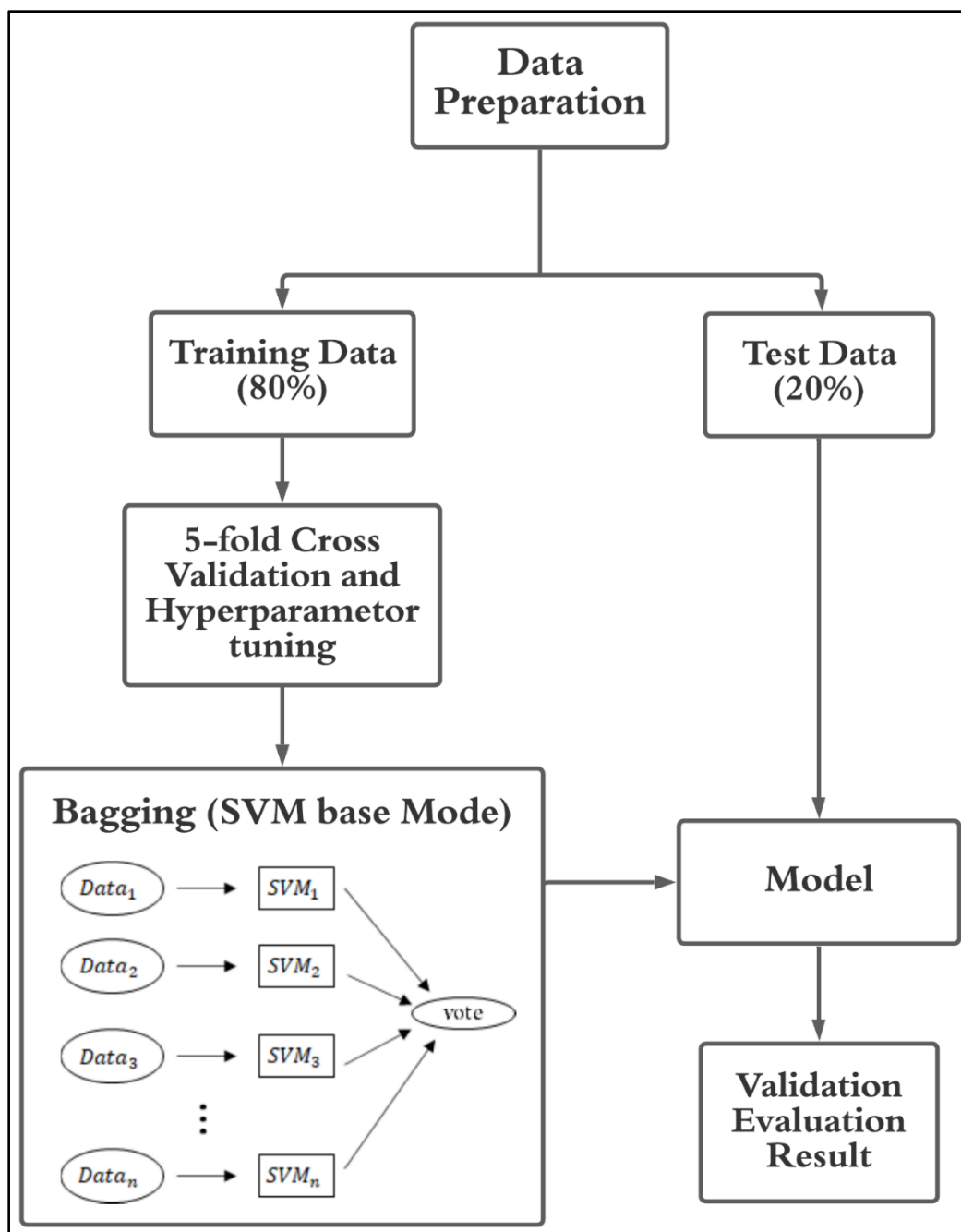
3.1.4 ผลลัพธ์สุดท้ายของการประเมินการตรวจสอบความถูกต้องของกระบวนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Bagging โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้และทำการเรียนรู้ จะได้ค่าวัดประสิทธิภาพของโมเดล คือ Accuracy, Precision, F1-Measure, recall, และ ROC curve ดังรูปที่ 12 - 13

ตาราง 6 กำหนดพารามิเตอร์ให้กับอัลกอริทึม Bagging โดยใช้ Decision Tree (DT) และ Support Vector Machine (SVM) เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ค่าพารามิเตอร์
<i>n_estimators</i>	10, 50, 100, 500
<i>learning_rate</i>	0.0001, 0.001, 0.01, 0.1, 1.0



รูปที่ 12 ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Bagging โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้



รูปที่ 13 ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

### 3.2 การสร้างแบบจำลอง Boosting โดยใช้ Decision Tree (DT) และ Support Vector Machine (SVM) เป็นอัลกอริทึมพื้นฐาน

3.3.1 นำข้อมูลการพันสภาพการณ์ที่ 1 2 และ 3 ที่ผ่านกระบวนการเตรียมข้อมูลมาทำการแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลฝึกสอน (Training Data) 80% และข้อมูลทดสอบ (Test Data) 20%

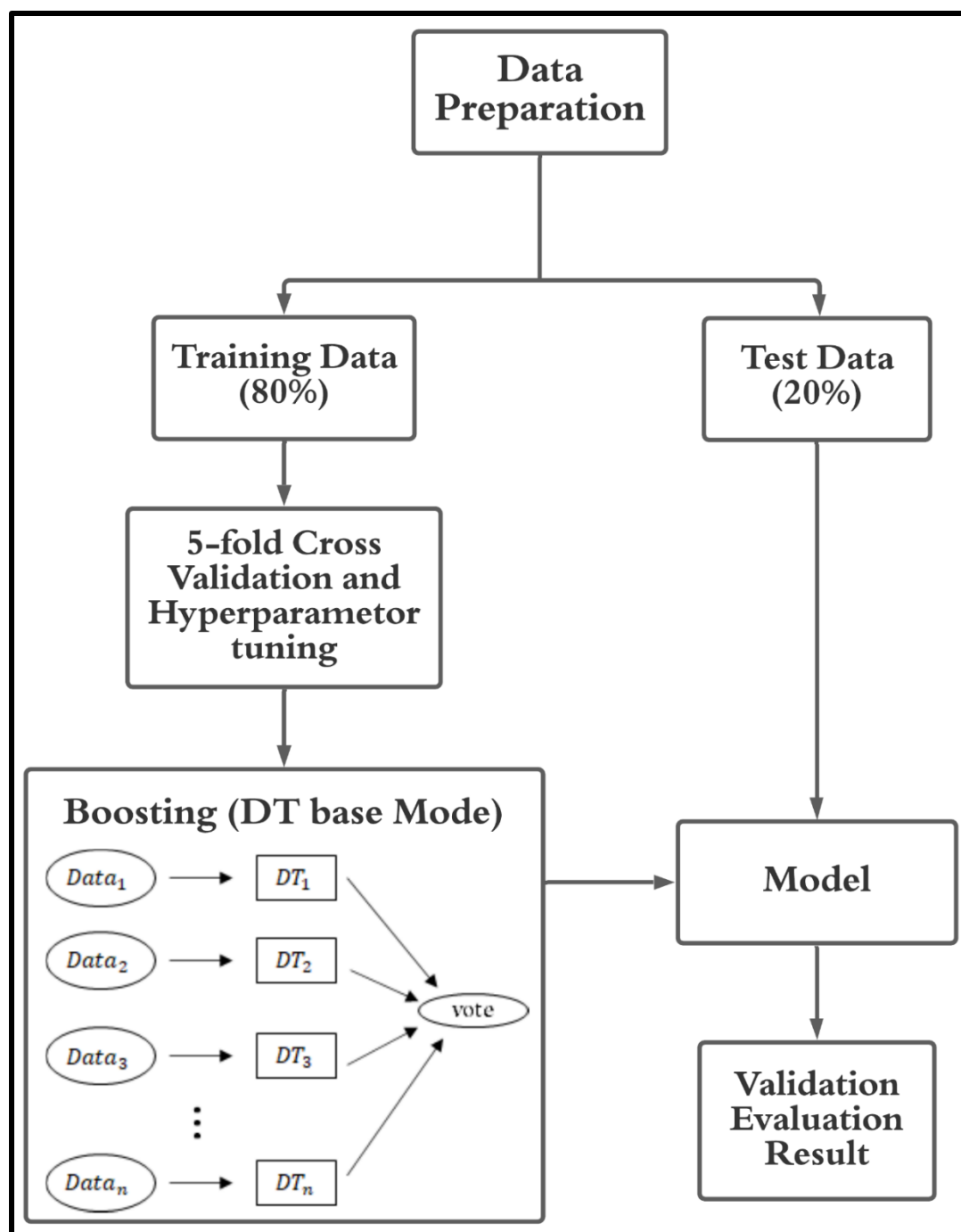
3.3.2 นำข้อมูลฝึกสอน (Training Data) มาทำการแบ่งข้อมูลให้เป็น 5 ส่วน ข้อมูลทุกส่วนจะถูกเรียนรู้และตรวจสอบในการวัดประสิทธิภาพ ด้วยวิธี 5-fold Cross validation และทำการหาค่าพารามิเตอร์ที่เหมาะสมโดยใช้ Grid Search สำหรับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Boosting โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ดังตารางที่ 7

3.3.3 กำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Boosting โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้หลังจากนั้นทำการเรียนรู้โดยใช้ข้อมูลทดสอบ (Test Data)

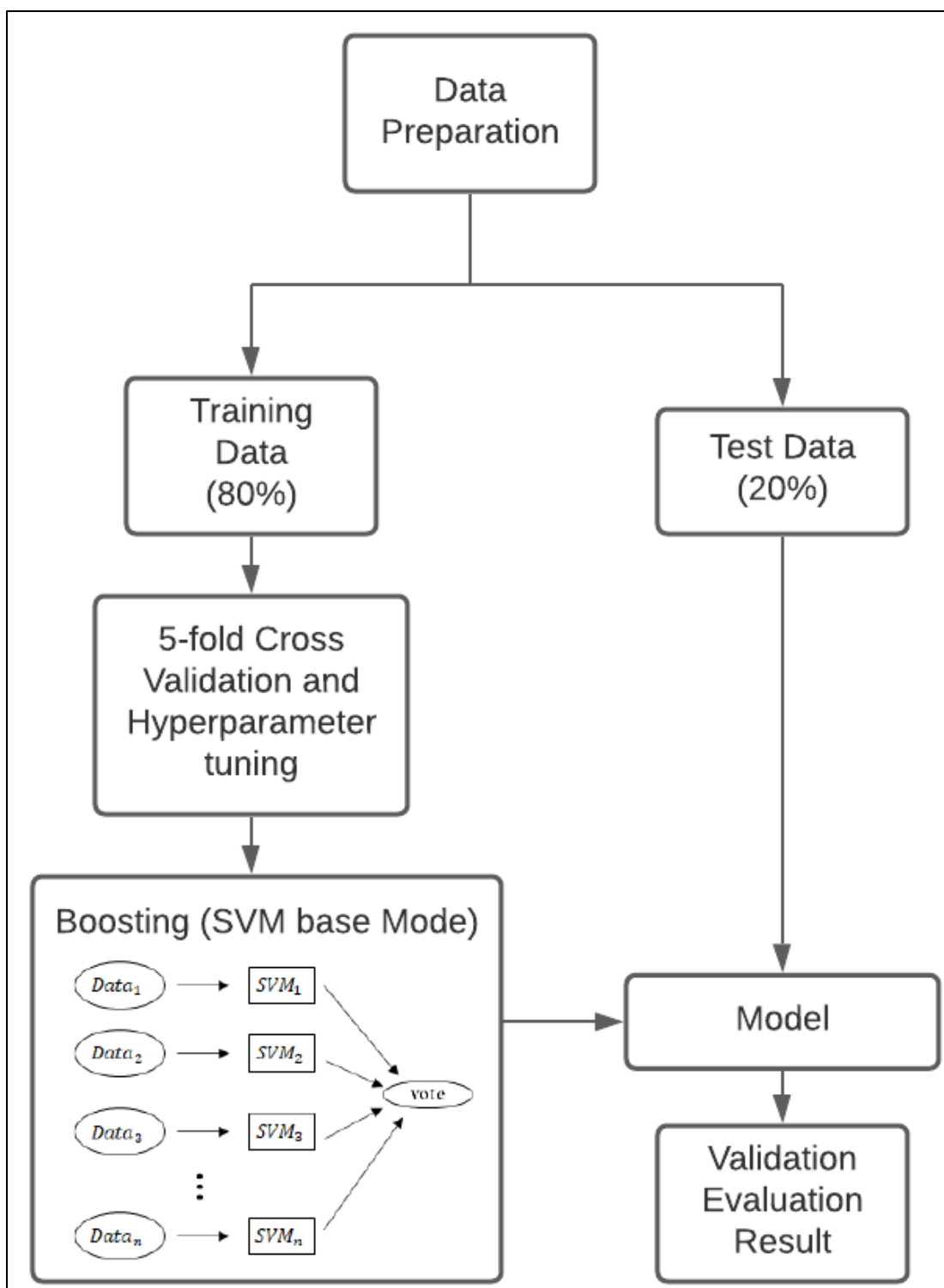
3.3.4 ผลลัพธ์สุดท้ายของการประเมินการตรวจสอบความถูกต้องของกระบวนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Boosting โดยใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้และทำการเรียนรู้ จะได้ค่าวัดประสิทธิภาพของโมเดล คือ Accuracy, Precision, F1-Measure, recall, และ ROC curve ดังรูปที่ 14 - 15

ตาราง 7 กำหนดพารามิเตอร์ให้กับอัลกอริทึม Boosting โดยใช้ Decision Tree (DT) และ Support Vector Machine (SVM) เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ค่าพารามิเตอร์
<i>n_estimators</i>	10, 50, 100, 500
<i>learning_rate</i>	0.0001, 0.001, 0.01, 0.1, 1.0



รูปที่ 14 ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมกลุ่มด้วยวิธี Boosting โดยใช้ DT เป็น อัลกอริทึมพื้นฐานสำหรับการเรียนรู้



รูปที่ 15 ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมกลุ่มด้วยวิธี Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

### 3.3 การสร้างแบบจำลอง Random Forest

3.4.1 นำข้อมูลการพันสภาพกรณ์ที่ 1 2 และ 3 ที่ผ่านกระบวนการเตรียมข้อมูลมาทำการแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลฝึกสอน (Training Data) 80% และข้อมูลทดสอบ (Test Data) 20%

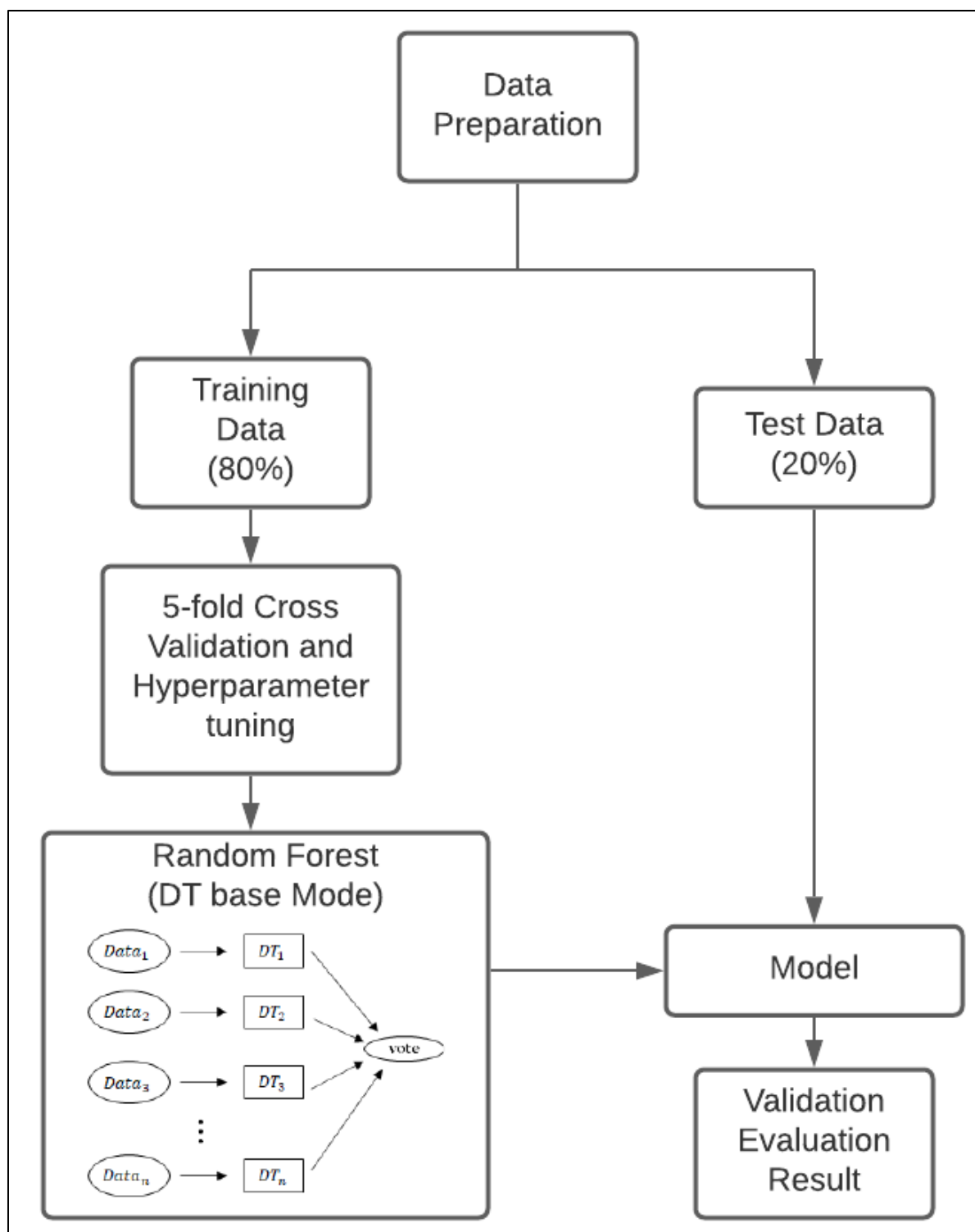
3.4.2 นำข้อมูลฝึกสอน (Training Data) มาทำการแบ่งข้อมูลให้เป็น 5 ส่วน ข้อมูลทุกส่วนจะถูกเรียนรู้และตรวจสอบในการวัดประสิทธิภาพ ด้วยวิธี 5-fold Cross validation และทำการหาค่าพารามิเตอร์ที่เหมาะสมโดยใช้ Grid Search สำหรับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Random Forest ดังตารางที่ 8

3.4.3 กำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Random Forest หลังจากนั้นทำการเรียนรู้โดยใช้ข้อมูลทดสอบ (Test Data)

3.4.4 ผลลัพธ์สุดท้ายของการประเมินการตรวจสอบความถูกต้องของกระบวนการสร้างแบบจำลองการจำแนกประเภทรวมรวมกลุ่มด้วยวิธี Random Forest จะได้ค่าวัดประสิทธิภาพของโมเดล คือ Accuracy, Precision, F1-Measure, recall, และ ROC curve ดังรูปที่ 16

ตาราง 8 กำหนดพารามิเตอร์ให้กับอัลกอริทึม Random Forest

พารามิเตอร์	ค่าพารามิเตอร์
Max_Depth	3, 5, 10, 20, 30
Criterion	Gini , Entropy
Max_Features	Auto , Sqrt, Log2
Min_Samples_Split	2, 5, 10



รูปที่ 16 ขั้นตอนการสร้างแบบจำลองการจำแนกประเภทรวมกลุ่มด้วยวิธี Random Forest

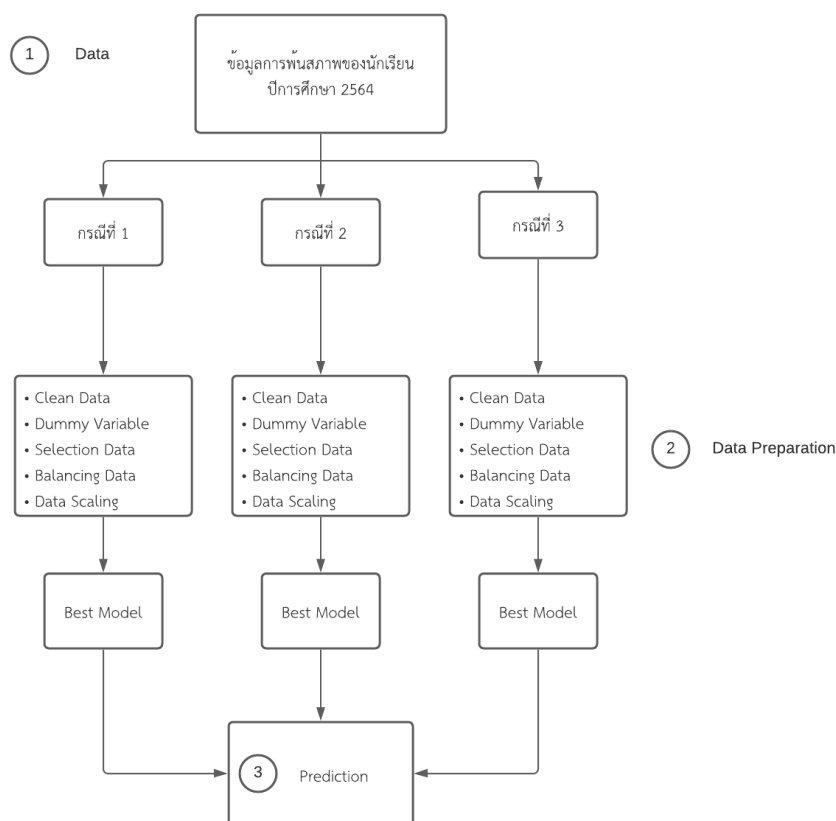


#### 4. เปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภท

ในส่วนนี้จะนำผลลัพธ์จากการวัดประสิทธิภาพของโมเดล คือ Accuracy, Precision, F1-Measure, recall, และ AUC ที่ได้จากการทดลองทั้งหมดนำมาทำการเปรียบเทียบ โดยจะพิจารณาเปรียบเทียบค่า F1-Measure และ AUC แต่ละโมเดลเพื่อที่จะเลือกโมเดลที่ดีที่สุดในการจำแนกประเภทสำหรับการพัฒนาศักยภาพของนักศึกษา

#### 5. ทำนายการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2564

หลังจากผ่านกระบวนการเปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภทการพัฒนาศักยภาพของนักศึกษาทั้ง 3 กรณี จากนั้นนำข้อมูลการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2564 ซึ่งทำการแบ่งข้อมูลการพัฒนาศักยภาพของนักศึกษาเป็น 3 กรณี โดยได้ผ่านกระบวนการจัดเตรียมข้อมูล หลังจากนั้นให้โมเดลที่ดีที่สุดสำหรับการการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2558 – 2563 ทั้ง 3 กรณีมาทำนายการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2564 ทั้ง 3 กรณี แสดงดังรูปที่ 17



รูปที่ 17 ขั้นตอนการทำนายการพัฒนาศักยภาพของนักศึกษาปีการศึกษา 2564

## บทที่ 4

### ผลการวิจัย

ในบทนี้ผู้วิจัยได้นำเสนอผลลัพธ์ที่ได้จากกระบวนการจัดเตรียมข้อมูล การสำรวจข้อมูลเบื้องต้น และการทดลองสำหรับการวัดประสิทธิภาพในการจำแนกประเภทการผันสภาพของนักศึกษา รวมไปถึงแสดงการปรับค่าพารามิเตอร์ให้เหมาะสมในอัลกอริทึมแบบเดี่ยวและแบบรวมกลุ่มที่ได้กล่าวมาข้างต้น ซึ่งใช้ข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้ข้อมูลเฉพาะนักศึกษาปี 1 ในแต่ละปีการศึกษา 6 ปีย้อนหลังตั้งแต่ปี พ.ศ. 2558 – 2563 จำนวน 614 คน โดยในการทดลองได้ทำการเปรียบเทียบระหว่างการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม เพื่อเปรียบเทียบประสิทธิภาพการจำแนกประเภทการผันสภาพของนักศึกษาที่ดีที่สุด และใช้ค่าจากตาราง Confusion Matrix ซึ่งเป็นค่าสำหรับใช้วัดประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาทั้งหมด สามารถนำมาคำนวณค่า Accuracy, Precision, F1-Measure, recall, และ ROC curve เพื่อวัดและเปรียบเทียบประสิทธิภาพการจำแนกประเภทการผันสภาพของนักศึกษา จากนั้นนำข้อมูลการผันสภาพของนักศึกษาปีการศึกษา 2564 มาทำนายการผันสภาพของนักศึกษา โดยในงานวิจัยได้ทำการแบ่งผลลัพธ์ของการทดลองออกเป็น 5 ส่วนดังต่อไปนี้

ส่วนที่ 1 การสำรวจข้อมูลเบื้องต้นและกระบวนการจัดเตรียมข้อมูล

ส่วนที่ 2 ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบเดี่ยว

1. อัลกอริทึม DT
2. อัลกอริทึม SVM

ส่วนที่ 3 ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบรวมกลุ่ม

1. วิธีการแบบ Bagging
2. วิธีการแบบ Boosting
3. วิธีการแบบ Random Forest

ส่วนที่ 4 เปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภท

ส่วนที่ 5 ทำนายการผันสภาพของนักศึกษาปีการศึกษา 2564

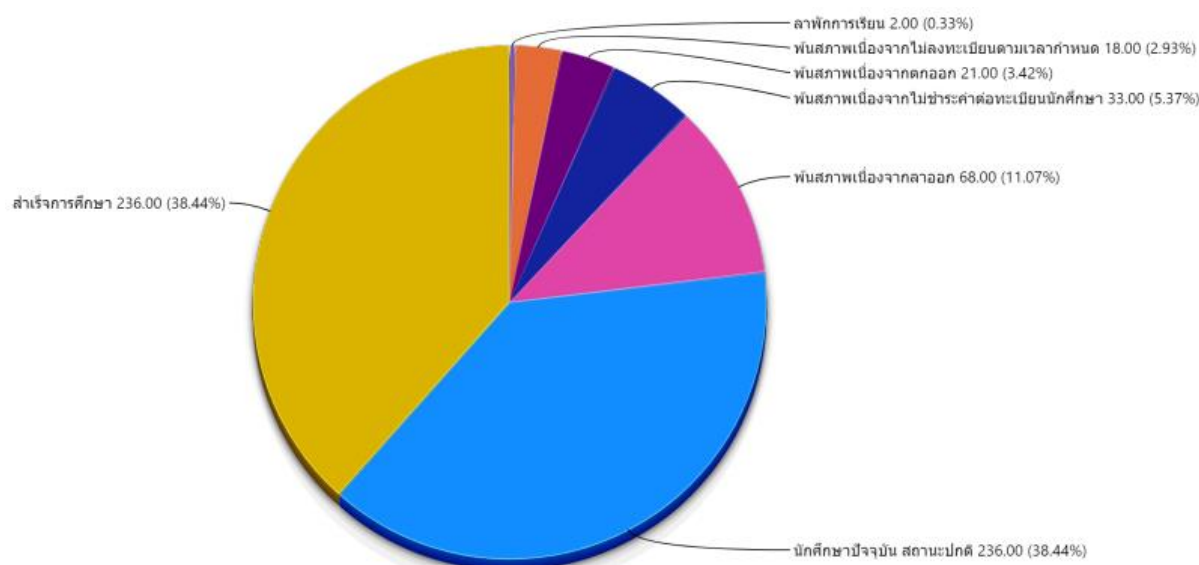
1. ข้อมูลกรณีที่ 1
2. ข้อมูลกรณีที่ 2
3. ข้อมูลกรณีที่ 3

### กระบวนการจัดเตรียมข้อมูลและการสำรวจข้อมูลเบื้องต้น

ผลลัพธ์จากการนำข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้ข้อมูลเฉพาะนักศึกษาปี 1 ในแต่ละปีการศึกษา 6 ปีย้อนหลังตั้งแต่ปี พ.ศ. 2558 – 2563 มีจำนวน 614 คน มาแสดงผลลัพธ์ ได้แก่กระบวนการจัดเตรียมข้อมูล และการสำรวจข้อมูลเบื้องต้น

#### 1. ผลลัพธ์จากการสำรวจข้อมูลเบื้องต้น

ผลลัพธ์จากการสำรวจข้อมูลเบื้องต้นทำการ Data Visualization ข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ 6 ปีย้อนหลังตั้งแต่ปี พ.ศ. 2558 – 2563 มีจำนวน 614 คน



รูปที่ 18 แผนภูมิวงกลมสถานะนักศึกษา

จากรูปที่ 18 แสดงประเภทการพ้นสภาพของนักศึกษาแบ่งออกเป็น พ้นสภาพเนื่องจากลาออก พ้นสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา พ้นสภาพเนื่องจากตกออก พ้นสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด ลาพักการเรียน พบว่า นักศึกษาที่สำเร็จการศึกษามีสัดส่วนอยู่ที่ร้อยละ 38.44% นักศึกษาในปัจจุบันที่มีสถานะปกติมีสัดส่วนอยู่ที่ร้อยละ 38.44% พ้นสภาพเนื่องจากลาออกมีสัดส่วนอยู่ที่ร้อยละ 11.07% พ้นสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษามีสัดส่วนอยู่ที่ร้อยละ 5.37% พ้นสภาพเนื่องจากตกออกมีสัดส่วนอยู่ที่ร้อยละ 3.42% พ้นสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนดมีสัดส่วนอยู่ที่ร้อยละ 2.93% ลาพักการเรียนมีสัดส่วนอยู่ที่ร้อยละ 0.33% จากจำนวนทั้งหมด 614 คน

## 2. ผลลัพธ์จากกระบวนการจัดเตรียมข้อมูล

ผลลัพธ์กระบวนการจัดเตรียมข้อมูลโดยใช้ข้อมูลนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ ตั้งแต่ปี พ.ศ. 2558 – 2563 จำนวน 614 คน และปีพ.ศ. 2564 จำนวน 105 คน ซึ่งได้ทำการแบ่งชุดข้อมูลการพัฒนาศักยภาพของนักศึกษาที่ใช้สำหรับการทดลองออกเป็น 3 กรณี หลังจากนั้นนำข้อมูลแต่ละกรณีมาผ่านกระบวนการทำความสะอาดข้อมูล (Clean Data) ทำการคัดเลือกข้อมูล (Selection Data) ปรับสมดุลข้อมูล (Balancing Data) ปรับปรุงขอบเขตข้อมูล (Data Scaling) และแปลงข้อมูลให้เป็นตัวแปรหุ่น (Dummy) ดังนี้

### 2.1 ข้อมูลการพัฒนาศักยภาพกรณีที่ 1

สถานะการพัฒนาศักยภาพของนักศึกษาประกอบไปด้วย พัฒนาศักยภาพเนื่องจากลาออก พัฒนาศักยภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา พัฒนาศักยภาพเนื่องจากตกออก และพัฒนาศักยภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด โดยมีข้อมูลพ.ศ. 2558 – 2563 มีจำนวน 586 รายการ 39 คุณลักษณะ และมีข้อมูลพ.ศ. 2564 มีจำนวน 103 รายการ 38 คุณลักษณะ แสดงในตาราง 9

ตาราง 9 ชื่อคุณลักษณะและคำนิยามของข้อมูลการพัฒนาศักยภาพกรณีที่ 1

ลำดับ	ตัวแปร	ความหมาย
1	STATUSTEXT	สถานะการพัฒนาศักยภาพ
2	GPA	เกรดเฉลี่ย
3	ENTRYGPA	เกรดเฉลี่ยก่อนเข้ามหาวิทยาลัย
4	STUDENTSEX	เพศ
5	FATHERINCOME	รายได้บิดา
6	YEARFATHER	อายุบิดา
7	MOTHERINCOME	รายได้มารดา
8	YEARMOTHER	อายุมารดา
9	NumberOfSon	บุตรคนที่
10	NumberOfSiblings	จำนวนพี่น้อง
11	BASICS OF STAGE ACTING_F	เกรด F รายวิชา BASICS OF STAGE ACTING
12	BIOLOGY FOR PHYSICAL SCIENCE LABORATORY_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE LABORATORY
13	BIOLOGY FOR PHYSICAL SCIENCE_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE
14	CALCULUS FOR PHYSICAL SCIENCE II_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE II

ลำดับ	ตัวแปร	ความหมาย
15	CALCULUS FOR PHYSICAL SCIENCE I_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE I
16	CALCULUS_F	เกรด F รายวิชา CALCULUS
17	COMPUTER PROGRAMMING I_F	เกรด F รายวิชา COMPUTER PROGRAMMING I
18	ELEMENTARY PHYSICS_F	เกรด F รายวิชา ELEMENTARY PHYSICS
19	ELEMENTARY TO BUSINESS AND ENTREPRENEURSHIP_F	เกรด F รายวิชา ELEMENTARY TO BUSINESS AND ENTREPRENEURSHIP
20	ENERGY AND ENVIRONMENT_F	เกรด F รายวิชา ENERGY AND ENVIRONMENT
21	ENGLISH FOR ACADEMIC PURPOSE I (EAP I)_F	เกรด F รายวิชา ENGLISH FOR ACADEMIC PURPOSE I (EAP I)
22	ENGLISH FOR ACADEMIC PURPOSES I_F	เกรด F รายวิชา ENGLISH FOR ACADEMIC PURPOSES I
23	ENGLISH FOR COMMUNICATION_F	เกรด F รายวิชา ENGLISH FOR COMMUNICATION
24	ENGLISH II_F	เกรด F รายวิชา ENGLISH II
25	ENGLISH I_F	เกรด F รายวิชา ENGLISH I
26	FUNDAMENTAL CHEMISTRY_F	เกรด F รายวิชา FUNDAMENTAL CHEMISTRY
27	GENERAL CHEMISTRY LABORATORY_F	เกรด F รายวิชา GENERAL CHEMISTRY LABORATORY
28	GENERAL MATHEMATICS_F	เกรด F รายวิชา GENERAL MATHEMATICS
29	GENERAL PHYSICS LABORATORY I_F	เกรด F รายวิชา GENERAL PHYSICS LABORATORY I
30	HAPPINESS OF LIFE_F	เกรด F รายวิชา HAPPINESS OF LIFE
31	INFORMATION LITERACY SKILLS_F	เกรด F รายวิชา INFORMATION LITERACY SKILLS
32	INTRODUCTION TO INFORMATION AND COMMUNICATION	เกรด F รายวิชา INTRODUCTION TO INFORMATION AND COMMUNICATION
33	LEARNING SKILLS_F	เกรด F รายวิชา LEARNING SKILLS
34	LOCAL WISDOM_F	เกรด F รายวิชา LOCAL WISDOM
35	MEDITATION FOR LIFE DEVELOPMENT_F	เกรด F รายวิชา MEDITATION FOR LIFE DEVELOPMENT

ลำดับ	ตัวแปร	ความหมาย
36	MULTICULTURALISM_F	เกรด F รายวิชา MULTICULTURALISM
37	OPERATIONS RESEARCH_F	เกรด F รายวิชา OPERATIONS RESEARCH
38	STATISTICAL ANALYSIS I_F	เกรด F รายวิชา STATISTICAL ANALYSIS I
39	STATISTICAL MODEL_F	เกรด F รายวิชา STATISTICAL MODEL

## 2.2 ข้อมูลการฟื้นฟูสภาพกรณีที่ 2

สถานการณ์ฟื้นฟูสภาพของนักศึกษาประกอบไปด้วย ฟื้นฟูสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา และฟื้นฟูสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด โดยมีข้อมูลพ.ศ. 2558 – 2563 มีจำนวน 500 รายการ 27 คุณลักษณะ และมีข้อมูลพ.ศ. 2564 มีจำนวน 99 รายการ 26 คุณลักษณะ แสดงในตาราง 10

ตาราง 10 ชื่อคุณลักษณะและคำนิยามของข้อมูลการฟื้นฟูสภาพกรณีที่ 2

อันดับ	ตัวแปร	ความหมาย
1	STATUSTEXT	สถานการณ์ฟื้นฟูสภาพ
2	GPA	เกรดเฉลี่ย
3	ENTRYGPA	เกรดเฉลี่ยก่อนเข้ามหาวิทยาลัย
4	FATHERINCOME	รายได้บิดา
5	YEARFATHER	อายุบิดา
6	MOTHERINCOME	รายได้มารดา
7	YEARMOTHER	อายุมารดา
8	NumberOfSon	บุตรคนที่
9	NumberOfSiblings	จำนวนพี่น้อง
10	BIOLOGY FOR PHYSICAL SCIENCE LABORATORY_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE LABORATORY
11	BIOLOGY FOR PHYSICAL SCIENCE_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE
12	CALCULUS FOR PHYSICAL SCIENCE I_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE I
13	CALCULUS_F	เกรด F รายวิชา CALCULUS
14	ELEMENTARY PHYSICS_F	เกรด F รายวิชา ELEMENTARY PHYSICS
15	ENGLISH FOR SCIENCES_F	เกรด F รายวิชา ENGLISH FOR SCIENCES

อันดับ	ตัวแปร	ความหมาย
16	ENGLISH FOR ACADEMIC PURPOSES I_F	เกรด F รายวิชา ENGLISH FOR ACADEMIC PURPOSES I
17	ENGLISH FOR COMMUNICATIONS_F	เกรด F รายวิชา ENGLISH FOR COMMUNICATIONS
18	ENGLISH FOR COMMUNICATION_F	เกรด F รายวิชา ENGLISH FOR COMMUNICATION
19	ENGLISH I_F	เกรด F รายวิชา ENGLISH I
20	GENERAL MATHEMATICS_F	เกรด F รายวิชา GENERAL MATHEMATICS
21	GENERAL PHYSICS LABORATORY I_F	เกรด F รายวิชา GENERAL PHYSICS LABORATORY I
22	HAPPINESS OF LIFE_F	เกรด F รายวิชา HAPPINESS OF LIFE
23	INTRODUCTION TO INFORMATION AND COMMUNICATION TECHNOLOGY_F	เกรด F รายวิชา INTRODUCTION TO INFORMATION AND COMMUNICATION TECHNOLOGY
24	MULTICULTURALISM_F	เกรด F รายวิชา MULTICULTURALISM
25	OPERATIONS RESEARCH_F	เกรด F รายวิชา OPERATIONS RESEARCH
26	STATISTICAL ANALYSIS I_F	เกรด F รายวิชา STATISTICAL ANALYSIS I
27	STATISTICAL INFORMATION PROJECT II_F	เกรด F รายวิชา STATISTICAL INFORMATION PROJECT II

### 2.3 ข้อมูลการพัฒนสภาพกรณ์ที่ 3

สถานการณ์พัฒนสภาพของนักศึกษาประกอบไปด้วย พัฒนสภาพเนื่องจากลาออก และพัฒนสภาพเนื่องจากตกออก โดยมีข้อมูลพ.ศ. 2558 – 2563 มีจำนวน 555 รายการ 49 คุณลักษณะ และมีข้อมูลพ.ศ. 2564 มีจำนวน 103 รายการ 48 คุณลักษณะ แสดงในตาราง 11

ตาราง 11 ชื่อคุณลักษณะและคำนิยามของข้อมูลการพัฒนสภาพกรณ์ที่ 3

อันดับ	ตัวแปร	ความหมาย
1	STATUSTEXT	สถานการณ์พัฒนสภาพ
2	GPA	เกรดเฉลี่ย
3	ENTRYGPA	เกรดเฉลี่ยก่อนเข้ามหาวิทยาลัย

อันดับ	ตัวแปร	ความหมาย
4	STUDENTSEX	เพศ
5	FATHERINCOME	รายได้บิดา
6	YEARFATHER	อายุบิดา
7	MOTHERINCOME	รายได้มารดา
8	YEARMOTHER	อายุมารดา
9	NumberOfSon	บุตรคนที่
10	NumberOfSiblings	จำนวนพี่น้อง
11	BASICS OF STAGE ACTING_F	เกรด F รายวิชา BASICS OF STAGE ACTING
12	BIOLOGY FOR PHYSICAL SCIENCE LABORATORY_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE LABORATORY
13	BIOLOGY FOR PHYSICAL SCIENCE_F	เกรด F รายวิชา BIOLOGY FOR PHYSICAL SCIENCE
14	CALCULUS FOR PHYSICAL SCIENCE II_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE II
15	CALCULUS FOR PHYSICAL SCIENCE II_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE II
16	CALCULUS FOR PHYSICAL SCIENCE I_F	เกรด F รายวิชา CALCULUS FOR PHYSICAL SCIENCE I
17	CALCULUS_F	เกรด F รายวิชา CALCULUS
18	COMPUTER PROGRAMMING I_F	เกรด F รายวิชา COMPUTER PROGRAMMING I
19	CREATIVE THINKING AND PROBLEM SOLVING_F	เกรด F รายวิชา CREATIVE THINKING AND PROBLEM SOLVING
20	DATABASE SYSTEMS AND DESIGN LABORATORY_F	เกรด F รายวิชา DATABASE SYSTEMS AND DESIGN LABORATORY
21	DATABASE SYSTEMS AND DESIGN_F	เกรด F รายวิชา DATABASE SYSTEMS AND DESIGN
22	DISCRETE MATHEMATICS AND APPLICATIONS_F	เกรด F รายวิชา DISCRETE MATHEMATICS AND APPLICATIONS
23	ELEMENTARY PHYSICS_F	เกรด F รายวิชา ELEMENTARY PHYSICS
24	ELEMENTARY TO BUSINESS AND ENTREPRENEURSHIP_F	เกรด F รายวิชา ELEMENTARY TO BUSINESS AND ENTREPRENEURSHIP
25	ENERGY AND ENVIRONMENT_F	เกรด F รายวิชา ENERGY AND ENVIRONMENT



อันดับ	ตัวแปร	ความหมาย
26	ENGLISH FOR ACADEMIC PURPOSE I (EAP I) _F	เกรด F รายวิชา ENGLISH FOR ACADEMIC PURPOSE I (EAP I)
27	ENGLISH FOR ACADEMIC PURPOSES I _F	เกรด F รายวิชา ENGLISH FOR ACADEMIC PURPOSES I
28	ENGLISH FOR COMMUNICATION _F	เกรด F รายวิชา ENGLISH FOR COMMUNICATION
29	ENGLISH II _F	เกรด F รายวิชา ENGLISH II
30	ENGLISH IV _F	เกรด F รายวิชา ENGLISH IV
31	ENGLISH I _F	เกรด F รายวิชา ENGLISH I
32	FINACIAL AND ACCOUNTING MANAGEMENT FOR EXECUTIVE _F	เกรด F รายวิชา FINACIAL AND ACCOUNTING MANAGEMENT FOR EXECUTIVE
33	FUNDAMENTAL CHEMISTRY _F	FUNDAMENTAL CHEMISTRY
34	GENERAL CHEMISTRY LABORATORY _F	เกรด F รายวิชา GENERAL CHEMISTRY LABORATORY
35	GENERAL MATHEMATICS _F	เกรด F รายวิชา GENERAL MATHEMATICS
36	GENERAL PHYSICS LABORATORY I _F	เกรด F รายวิชา GENERAL PHYSICS LABORATORY I
37	INFORMATION LITERACY SKILLS _F	เกรด F รายวิชา INFORMATION LITERACY SKILLS
38	INTRODUCTION TO INFORMATION AND COMMUNICATION TECHNOLOGY _F	เกรด F รายวิชา INTRODUCTION TO INFORMATION AND COMMUNICATION TECHNOLOGY
39	LEADERSHIP AND MANAGEMENT _F	เกรด F รายวิชา LEADERSHIP AND MANAGEMENT
40	LEARNING SKILLS _F	เกรด F รายวิชา LEARNING SKILLS
41	LOCAL WISDOM _F	เกรด F รายวิชา LOCAL WISDOM
42	MEDITATION FOR LIFE DEVELOPMENT _F	เกรด F รายวิชา MEDITATION FOR LIFE DEVELOPMENT
43	MULTICULTURALISM _F	เกรด F รายวิชา MULTICULTURALISM
44	OPERATIONS RESEARCH _F	เกรด F รายวิชา OPERATIONS RESEARCH

อันดับ	ตัวแปร	ความหมาย
45	PHYSICAL EDUCATION ACTIVITY (AEROBIC DANCE)_F	เกรด F รายวิชา PHYSICAL EDUCATION ACTIVITY (AEROBIC DANCE)
46	RISK AND INSURANCE_F	เกรด F รายวิชา RISK AND INSURANCE
47	SMALL AND MEDIUM ENTERPRISES MANAGEMENT_F	เกรด F รายวิชา SMALL AND MEDIUM ENTERPRISES MANAGEMENT
48	STATISTICAL ANALYSIS I_F	เกรด F รายวิชา STATISTICAL ANALYSIS I
49	STATISTICAL MODEL_F	เกรด F รายวิชา STATISTICAL MODEL

### ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบเดียว

ผลลัพธ์จากการวิเคราะห์ข้อมูลสำหรับการจำแนกประเภทการฟื้นสภาพของนักศึกษา ในการสร้างแบบจำลองการจำแนกประเภทแบบเดียว โดยได้ทำการเปรียบเทียบ 2 อัลกอริทึม ได้แก่ DT และ SVM

#### 1. ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม DT

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการฟื้นสภาพของนักศึกษาโดยใช้อัลกอริทึม DT ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 12 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้อัตราประสิทธิภาพความถูกต้องของการจำแนกประเภทการฟื้นสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 12 ค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึม DT

พารามิเตอร์	ค่าพารามิเตอร์		
	การฟื้นสภาพกรณีที่ 1	การฟื้นสภาพกรณี ที่ 2	การฟื้นสภาพกรณี ที่ 3
Criterion	Entropy	Gini	Gini
Max Depth	12	18	10
Max Features	34	8	33
Min Samples Leaf	3	2	2

ตาราง 13 Confusion Matrix โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 1

Class	Prediction		
	0	1	
Actual	0	86	8
	1	5	89

ตาราง 14 Confusion Matrix โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 2

Class	Prediction		
Actual	0	1	
	0	92	2
	1	0	94

ตาราง 15 Confusion Matrix โดยใช้อัลกอริทึม DT ของการฟื้นฟูสภาพกรณีที่ 3

Class	Prediction	
	0	1
Actual	0	88
	1	94

จากตารางที่ 13-15 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการฟื้นฟูสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทแบบเดียวโดยใช้อัลกอริทึม DT ดังตารางที่ 16

ตาราง 16 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม DT

Performance	Single Model		
	การพันสภาพกรณี ที่ 1	การพันสภาพกรณี ที่ 2	การพันสภาพกรณี ที่ 3
Accuracy	0.93	0.99	0.97
Precision	0.93	0.99	0.97
Recall	0.93	0.99	0.97
F1-score	0.93	0.99	0.97
AUC	0.96	0.99	0.96

จากตารางที่ 16 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม DT ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 13 - 15 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการพันสภาพกรณีที่ 1 - 3 ดังตารางที่ 12 พบว่า ข้อมูลการพันสภาพกรณีที่ 2 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษาอยู่ที่ 99% ตามด้วยข้อมูลการพันสภาพกรณีที่ 3 และข้อมูลการพันสภาพกรณีที่ 1 โดยมีประสิทธิภาพความถูกต้องอยู่ที่ 97%, และ 93% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณีที่ 2 อยู่ที่ 99% ตามด้วยข้อมูลการพันสภาพกรณีที่ 3 และข้อมูลการพันสภาพกรณีที่ 1 โดยมีค่าความถ่วงดุลอยู่ที่ 97%, และ 93% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณีที่ 2 อยู่ที่ 99% ตามด้วยข้อมูลการพันสภาพกรณีที่ 3 และข้อมูลการพันสภาพกรณีที่ 1 โดยมีค่า AUC เท่ากันอยู่ที่ 96%

## 2. ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม SVM

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการผันสภาพของนักศึกษาโดยใช้อัลกอริทึม SVM ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 17 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการผันสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 17 ค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึม SVM

พารามิเตอร์	ค่าพารามิเตอร์		
	การผันสภาพกรณีที่ 1	การผันสภาพกรณี ที่ 2	การผันสภาพกรณี ที่ 3
C	10	100	100
Gamma	0.1	1	1
Kernel	rbf	rbf	rbf

ตาราง 18 Confusion Matrix โดยใช้อัลกอริทึม SVM ของการผันสภาพกรณีที่ 1

Class	Prediction	
	0	1
Actual	0	89
	1	5
	11	83

ตาราง 19 Confusion Matrix โดยใช้อัลกอริทึม SVM ของการผันสภาพกรณีที่ 2

Class	Prediction	
	0	1
Actual	0	87
	1	94

ตาราง 20 Confusion Matrix โดยใช้อัลกอริทึม SVM ของการผันสภาพกรณีที่ 3

Class	Prediction	
	0	1
Actual	0	85
	1	9
	15	79

จากตารางที่ 18-20 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทแบบเดียวโดยใช้อัลกอริทึม SVM ดังตารางที่ 21

**ตาราง 21** การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม SVM

Performance	Single Model		
	การผันสภาพกรณี ที่ 1	การผันสภาพกรณี ที่ 2	การผันสภาพกรณี ที่ 3
Accuracy	0.91	0.96	0.87
Precision	0.91	0.96	0.87
Recall	0.91	0.96	0.87
F1-score	0.91	0.96	0.87
AUC	0.94	0.98	0.93

จากตารางที่ 21 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม SVM ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 18 - 20 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการผันสภาพกรณีที่ 1 - 3 ดังตารางที่ 17 พบว่า ข้อมูลการผันสภาพกรณีที่ 2 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษาอยู่ที่ 96% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีประสิทธิภาพความถูกต้องอยู่ที่ 91%, 87% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 96% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีค่าความถ่วงดุลอยู่ที่ 91%, 87% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 98% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีค่า AUC อยู่ที่ 94%, 93%

## ผลลัพธ์อัลกอริทึมสำหรับการจำแนกประเภทแบบรวมกลุ่ม

ผลลัพธ์จากการวิเคราะห์ข้อมูลสำหรับการจำแนกประเภทการพัฒนสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม โดยได้ทำการเปรียบเทียบ 3 วิธีได้แก่วิธี Bagging, Boosting, และ Random Forest

### 1. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Bagging

ผลการทดลองด้วยวิธี Bagging ซึ่งใช้อัลกอริทึมแบบเดี่ยวเป็นฐานในการสร้างโมเดล สำหรับการวิเคราะห์ประสิทธิภาพความถูกต้องในการจำแนกประเภทการพัฒนสภาพของนักศึกษา โดยแบ่งออกเป็น 2 รูปแบบดังต่อไปนี้

#### 1.1 ผลลัพธ์จากการจำแนกด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการพัฒนสภาพของนักศึกษาด้วยใช้วิธีการ Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 22 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการพัฒนสภาพของนักศึกษามีความแม่นยำมากที่สุดทั้ง 3 กลุ่มตัวอย่าง

ตาราง 22 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

พารามิเตอร์	ค่าพารามิเตอร์		
	การพัฒนสภาพกรณีที่ 1	การพัฒนสภาพกรณี ที่ 2	การพัฒนสภาพกรณี ที่ 3
n estimators	500	100	10
Learning Rate	0.1	0.0001	1.0

ตาราง 23 Confusion Matrix ด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการพัฒนสภาพกรณีที่ 1

Class	Prediction	
	0	1
Actual	0	90    4
	1	0    94

ตาราง 24 Confusion Matrix ด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการผันสภาพกรณ์ที่ 4

Class	Prediction	
	0	1
Actual	0	93
	1	94

ตาราง 25 Confusion Matrix ด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการผันสภาพกรณ์ที่ 3

Class	Prediction	
	0	1
Actual	0	90
	1	92

จากตารางที่ 23 - 25 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ดังตารางที่ 9

ตาราง 26 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

Performance	Bagging Model (DT Base Model)		
	การผันสภาพกรณ์ที่ 1	การผันสภาพกรณ์ที่ 2	การผันสภาพกรณ์ที่ 3
Accuracy	0.98	0.99	0.97
Precision	0.98	0.99	0.97
Recall	0.98	0.99	0.97
F1-score	0.98	0.99	0.97
AUC	0.99	1.00	0.99



จากตารางที่ 26 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธี Bagging โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 23 - 25 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการผันสภาพกรณีที่ 1 - 3 ดังตารางที่ 22 พบว่า ข้อมูลการผันสภาพกรณีที่ 2 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษาอยู่ที่ 99% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีประสิทธิภาพความถูกต้องอยู่ที่ 98%, 97% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 99% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีค่าความถ่วงดุลอยู่ที่ 98%, 97% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 100% ตามด้วยข้อมูลการผันสภาพกรณีที่ 1 และข้อมูลการผันสภาพกรณีที่ 3 โดยมีค่า AUC เท่ากันอยู่ที่ 99%

## 1.2 ผลลัพธ์จากการจำแนกด้วยวิธี Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการผันสภาพของนักศึกษาด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 27 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการผันสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 27 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

พารามิเตอร์	ค่าพารามิเตอร์		
	การผันสภาพกรณีที่ 1	การผันสภาพกรณี ที่ 2	การผันสภาพกรณี ที่ 3
n estimators	500	100	10
Learning Rate	0.1	0.0001	1.0

ตาราง 28 Confusion Matrix ด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 1

Class	Prediction	
	0	1
Actual	0	84 10
	1	11 83

ตาราง 29 Confusion Matrix ด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 2

Class	Prediction	
	0	1
Actual	0	87
	1	94

ตาราง 30 Confusion Matrix ด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 3

Class	Prediction		
	0	1	
Actual	0	86	8
	1	13	81

จากตารางที่ 27 - 30 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ดังตารางที่ 31

**ตาราง 31** การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

Performance	Bagging Model (SVM Base Model)		
	การพันสภาพกรณ์ ที่ 1	การพันสภาพกรณ์ ที่ 2	การพันสภาพกรณ์ ที่ 3
Accuracy	0.89	0.96	0.89
Precision	0.89	0.96	0.89
Recall	0.89	0.96	0.89
F1-score	0.89	0.96	0.89
AUC	0.94	0.98	0.95

จากตารางที่ 31 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ Bagging โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 27 – 30 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการพันสภาพกรณ์ที่ 1 – 3 ดังตารางที่ 27 พบว่า ข้อมูลการพันสภาพกรณ์ที่ 2 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษาอยู่ที่ 96% ตามด้วยข้อมูลการพันสภาพกรณ์ที่ 3 และข้อมูลการพันสภาพกรณ์ที่ 1 โดยมีประสิทธิภาพความถูกต้องเท่ากันอยู่ที่ 89% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณ์ที่ 2 อยู่ที่ 96% ตามด้วยข้อมูลการพันสภาพกรณ์ที่ 3 และข้อมูลการพันสภาพกรณ์ที่ 1 โดยมีค่าความถ่วงดุลเท่ากันอยู่ที่ 89% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณ์ที่ 2 อยู่ที่ 98% ตามด้วยข้อมูลการพันสภาพกรณ์ที่ 3 และข้อมูลการพันสภาพกรณ์ที่ 1 โดยมีค่าความถ่วงดุลอยู่ที่ 95%, และ 94%

## 2. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Boosting

ผลการทดลองด้วยวิธี Boosting ซึ่งใช้อัลกอริทึมแบบเดี่ยวเป็นฐานในการสร้างโมเดล สำหรับการวิเคราะห์ประสิทธิภาพความถูกต้องในการจำแนกประเภทการพ้นสภาพของนักศึกษา โดยแบ่งออกเป็น 2 รูปแบบดังต่อไปนี้

### 2.1 ผลลัพธ์จากการจำแนกด้วยวิธี Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการพ้นสภาพของนักศึกษาด้วยวิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 32 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการพ้นสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 32 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

พารามิเตอร์	ค่าพารามิเตอร์		
	การพ้นสภาพกรณีที่ 1	การพ้นสภาพกรณี ที่ 2	การพ้นสภาพกรณี ที่ 3
n estimators	500	100	100
Learning Rate	1.0	1.0	1.0

ตาราง 33 Confusion Matrix วิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการพ้นสภาพกรณีที่ 1

Class	Prediction	
	0	1
Actual	0	93
	1	94

ตาราง 34 Confusion Matrix วิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการผันสภาพกรณ์ที่ 2

Class	Prediction	
	0	1
Actual	0	93
	1	94

ตาราง 35 Confusion Matrix วิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ของการผันสภาพกรณ์ที่ 3

Class	Prediction	
	0	1
Actual	0	93
	1	94

จากตารางที่ 33 - 35 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทด้วยวิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ดังตารางที่ 36

ตาราง 36 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

Performance	Boosting Model (DT Base Model)		
	การผันสภาพกรณ์ ที่ 1	การผันสภาพกรณ์ ที่ 2	การผันสภาพกรณ์ ที่ 3
Accuracy	0.99	0.99	0.99
Precision	0.99	0.99	0.99
Recall	0.99	0.99	0.99
F1-score	0.99	0.99	0.99
AUC	0.99	0.98	0.95

จากตารางที่ 36 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธี Boosting โดยใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 33 - 35 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการฝึกสภาพกรณี 1 - 3 ดังตารางที่ 32 พบว่า ข้อมูลการฝึกสภาพกรณี 1 - 3 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) ในการจำแนกประเภทการฝึกสภาพของนักศึกษาเท่ากันอยู่ที่ 99% ในข้อมูลการฝึกสภาพกรณี 1 - 3 มีค่าความถ่วงดุล (F1-Score) ในการจำแนกประเภทการฝึกสภาพของนักศึกษาเท่ากันอยู่ที่ 99% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการฝึกสภาพของนักศึกษาคือ ข้อมูลการฝึกสภาพกรณี 1 อยู่ที่ 99% ตามด้วยข้อมูลการฝึกสภาพกรณี 2 และข้อมูลการฝึกสภาพกรณี 3 โดยมีค่าความถ่วงดุลอยู่ที่ 98%, และ 95%

## 2.2 ผลลัพธ์จากการจำแนกด้วยวิธี Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการฝึกสภาพของนักศึกษาด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการทดลอง โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 37 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการฝึกสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 37 ค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

พารามิเตอร์	ค่าพารามิเตอร์		
	การฝึกสภาพกรณี 1	การฝึกสภาพกรณี ที่ 2	การฝึกสภาพกรณี ที่ 3
n estimators	500	50	50
Learning Rate	0.01	1.0	1.0

ตาราง 38 Confusion Matrix ด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการฝึกสภาพกรณี 1

Class	Prediction	
	0	1
Actual	0	55    39
	1	9      85

ตาราง 39 Confusion Matrix ด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 2

Class	Prediction	
	0	1
Actual	0	89
	1	9

ตาราง 40 Confusion Matrix ด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ของการผันสภาพกรณ์ที่ 3

Class	Prediction	
	0	1
Actual	0	<i>81</i>
	1	<i>10</i>

จากตารางที่ 38 - 40 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการผันสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ดังตารางที่ 41

ตาราง 41 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธีการ Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

Performance	Boosting Model (SVM Base Model)		
	การผันสภาพ กรณ์ที่ 1	การผันสภาพกรณ์ที่ 2	การผันสภาพกรณ์ ที่ 3
Accuracy	0.74	0.93	0.88
Precision	0.74	0.93	0.88
Recall	0.74	0.93	0.88
F1-score	0.74	0.93	0.88
AUC	0.86	0.98	0.95

จากตารางที่ 41 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทด้วยวิธี Boosting โดยใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 38 - 40 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการผันสภาพกรณีที่ 1 - 3 ดังตารางที่ 37 พบว่า ข้อมูลการผันสภาพกรณีที่ 2 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษาอยู่ที่ 93% ตามด้วยข้อมูลการผันสภาพกรณีที่ 3 และข้อมูลการผันสภาพกรณีที่ 1 โดยมีประสิทธิภาพความถูกต้องอยู่ที่ 88%, 74% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 93% ตามด้วยข้อมูลการผันสภาพกรณีที่ 3 และข้อมูลการผันสภาพกรณีที่ 1 โดยมีค่าความถ่วงดุลอยู่ที่ 88%, 74% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการผันสภาพของนักศึกษา คือ ข้อมูลการผันสภาพกรณีที่ 2 อยู่ที่ 98% ตามด้วยข้อมูลการผันสภาพกรณีที่ 3 และข้อมูลการผันสภาพกรณีที่ 1 โดยมีค่า ROC อยู่ที่ 95%, 86%

### 3. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Random Forest

ผลลัพธ์การวิเคราะห์ข้อมูลในการจำแนกประเภทการผันสภาพของนักศึกษาด้วยวิธี Random Forest โดยได้กำหนดค่าพารามิเตอร์ดังตารางที่ 42 ซึ่งเป็นค่าพารามิเตอร์ที่เหมาะสมที่จะทำให้ประสิทธิภาพความถูกต้องของการจำแนกประเภทการผันสภาพของนักศึกษามีความแม่นยำมากที่สุด

ตาราง 42 ค่าพารามิเตอร์ที่เหมาะสมโดยใช้อัลกอริทึม Random Forest

พารามิเตอร์	ค่าพารามิเตอร์		
	การผันสภาพกรณีที่ 1	การผันสภาพกรณี ที่ 2	การผันสภาพกรณี ที่ 3
criterion	Gini	Gini	Entropy
Max Depth	30	20	30
Max Features	Log2	Log2	auto
Min Samples Split	2	2	2



ตาราง 43 Confusion Matrix ด้วยวิธี Random Forest ของการฟื้นฟูสภาพกรณีที่ 1

Class	Prediction	
	0	1
Actual	0	93
	1	94

ตาราง 44 Confusion Matrix ด้วยวิธี Random Forest ของการฟื้นฟูสภาพกรณีที่ 2

Class	Prediction	
	0	1
Actual	0	93
	1	94

ตาราง 45 Confusion Matrix ด้วยวิธี Random Forest ของการฟื้นฟูสภาพกรณีที่ 3

Class	Prediction	
	0	1
Actual	0	93
	1	94

จากตารางที่ 43 - 45 Confusion Matrix สามารถที่จะนำมาคำนวณค่าวัดประสิทธิภาพความถูกต้องของแบบจำลองสำหรับการจำแนกประเภทการฟื้นฟูสภาพของนักศึกษาในการสร้างแบบจำลองการจำแนกประเภทด้วยวิธี Random Forest ดังตารางที่ 46

ตาราง 46 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม Random Forest

Performance	Random Forest		
	การพันสภาพ กรณีที่ 1	การพันสภาพกรณีที่ 2	การพันสภาพกรณี ที่ 3
Accuracy	0.99	0.99	0.99
Precision	0.99	0.99	0.99
Recall	0.99	0.99	0.99
F1-score	0.99	0.99	0.99
AUC	0.99	1.00	0.99

จากตารางที่ 46 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภท โดยใช้อัลกอริทึม Random Forest ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตารางที่ 38 - 40 Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมของข้อมูลการพันสภาพกรณีที่ 1 - 3 ดังตารางที่ 42 พบว่า ข้อมูลการพันสภาพกรณีที่ 1 - 3 มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษาอยู่ที่ 99% ค่าความถ่วงดุล (F1-Score) สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณีที่ 1 - 3 อยู่ที่ 99% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC สูงสุดในการจำแนกประเภทการพันสภาพของนักศึกษา คือ ข้อมูลการพันสภาพกรณีที่ 2 อยู่ที่ 100% ตามด้วยข้อมูลการพันสภาพกรณีที่ 1 และข้อมูลการพันสภาพกรณีที่ 3 โดยมีค่า AUC เท่ากันอยู่ที่ 99%

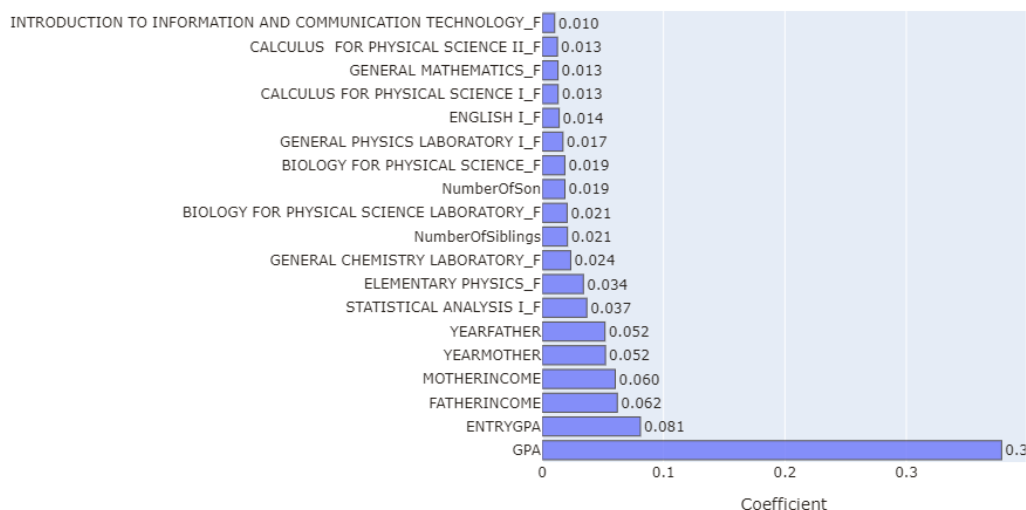
## เปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภท

### 1. ข้อมูลการพัฒนสภาพกรณ์ที่ 1

ข้อมูลการพัฒนสภาพกรณ์ที่ 1 โดยสถานะการพัฒนสภาพของนักศึกษาประกอบไปด้วย พัฒนสภาพเนื่องจากลาออก พัฒนสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา พัฒนสภาพเนื่องจากตกออก และพัฒนสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด จากการทดลองในการสร้างการจำแนกประเภทแบบเดี่ยว และการจำแนกประเภทแบบรวมกลุ่ม พบว่า การใช้การจำแนกประเภทแบบรวมกลุ่มที่ใช้ DT เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ให้ประสิทธิภาพมากกว่าการใช้การจำแนกประเภทแบบเดี่ยวโดยใช้ อัลกอริทึม DT ในทางตรงกันข้ามการใช้การจำแนกประเภทแบบรวมกลุ่มที่ใช้ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ให้ประสิทธิภาพไม่แตกต่างจากการใช้การจำแนกประเภทแบบเดี่ยวโดยใช้ อัลกอริทึม SVM อัลกอริทึมที่มีประสิทธิภาพในการจำแนกประเภทการพัฒนสภาพดีที่สุด คือ Random Forest และ Boosting Model (SVM Base Model) ซึ่งให้ค่าประสิทธิภาพไม่แตกต่างกันโดยมีค่าความถูกต้อง (Accuracy) อยู่ที่ 99% ค่าความแม่นยำ (Precision) อยู่ที่ 99% ค่าความครบถ้วน (Recall) อยู่ที่ 99% ค่าความถ่วงดุล (f1-score) อยู่ที่ 99% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC อยู่ที่ 99% สำหรับการจำแนกประเภทการพัฒนสภาพของนักศึกษา แสดงในตารางที่ 47

ตาราง 47 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการพัฒนสภาพของข้อมูลการพัฒนสภาพกรณ์ที่ 1

Performance	Single Model		Ensemble Model				
	DT	SVM	Bagging		Boosting		Random Forest
			DT	SVM	DT	SVM	
Accuracy	0.93	0.91	0.98	0.89	<u>0.99</u>	0.74	<u>0.99</u>
Precision	0.93	0.91	0.98	0.89	<u>0.99</u>	0.74	<u>0.99</u>
Recall	0.93	0.91	0.98	0.89	<u>0.99</u>	0.74	<u>0.99</u>
F1-score	0.93	0.91	0.98	0.89	<u>0.99</u>	0.74	<u>0.99</u>
AUC	0.96	0.94	0.99	0.94	<u>0.99</u>	0.86	<u>0.99</u>



รูปที่ 19 ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาของการพัฒนากรณีที่ 1

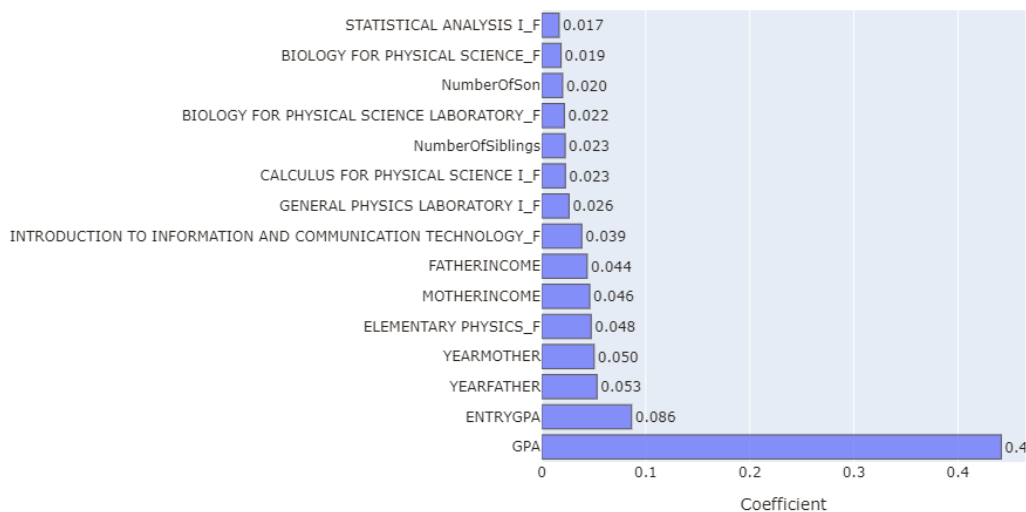
จากการทดลองในการสร้างตัวจำแนกประเภทการพัฒนาของนักศึกษาของการพัฒนากรณีที่ 1 โดยใช้ อัลกอริทึม Random Forest พบว่า ปัจจัยที่สำคัญต่อการพัฒนาของนักศึกษาประกอบด้วย เกรดเฉลี่ย, เกรดเฉลี่ยระดับมัธยม, รายได้บิดา, รายได้มารดา, อายุมารดา, อายุบิดา, เกรด F รายวิชา STATISTICAL ANALYSIS I, ELEMENTARY PHYSICS, GENERAL CHEMISTRY LABORATORY, จำนวนพี่น้อง ตามลำดับ แสดงในรูปที่ 19

## 2. ข้อมูลการพัฒนากรณีที่ 2

ข้อมูลการพัฒนากรณีที่ 2 โดยสถานะการพัฒนาของนักศึกษาประกอบไปด้วย พัฒนาการ เนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา และพัฒนาเนื่องจากไม่ลงทะเบียนตามเวลากำหนด จากการทดลองในการสร้างการจำแนกประเภทแบบเดี่ยว และการจำแนกประเภทแบบรวมกลุ่ม พบว่า การใช้การจำแนกประเภทแบบรวมกลุ่มที่ใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ให้ประสิทธิภาพไม่แตกต่างจากการใช้การจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม DT และ SVM อัลกอริทึมที่มีประสิทธิภาพในการจำแนกประเภทการพัฒนาที่ดีที่สุด คือ Random Forest ซึ่งให้ค่าประสิทธิภาพมีความถูกต้อง (Accuracy) อยู่ที่ 99% ค่าความแม่นยำ (Precision) อยู่ที่ 99% ค่าความครบถ้วน (Recall) อยู่ที่ 99% ค่าความถ่วงดุล (f1-score) อยู่ที่ 99% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC อยู่ที่ 100% สำหรับการจำแนกประเภทการพัฒนาของนักศึกษา แสดงในตารางที่ 48

ตาราง 48 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการผันสภาพของข้อมูลการผันสภาพ  
กรณีที่ 2

Performance	Single Model		Ensemble Model				
	DT	SVM	Bagging		Boosting		Random Forest
			DT	SVM	DT	SVM	
Accuracy	0.99	0.96	<u>0.99</u>	0.96	0.99	0.92	<u>0.99</u>
Precision	0.99	0.96	<u>0.99</u>	0.96	0.99	0.92	<u>0.99</u>
Recall	0.99	0.96	<u>0.99</u>	0.96	0.99	0.92	<u>0.99</u>
F1-score	0.99	0.96	<u>0.99</u>	0.96	0.99	0.92	<u>0.99</u>
AUC	0.99	0.98	0.99	0.98	0.98	0.98	<u>1.00</u>



รูปที่ 20 ปัจจัยที่ส่งผลต่อการผันสภาพของนักศึกษาของการผันสภาพกรณีที่ 2

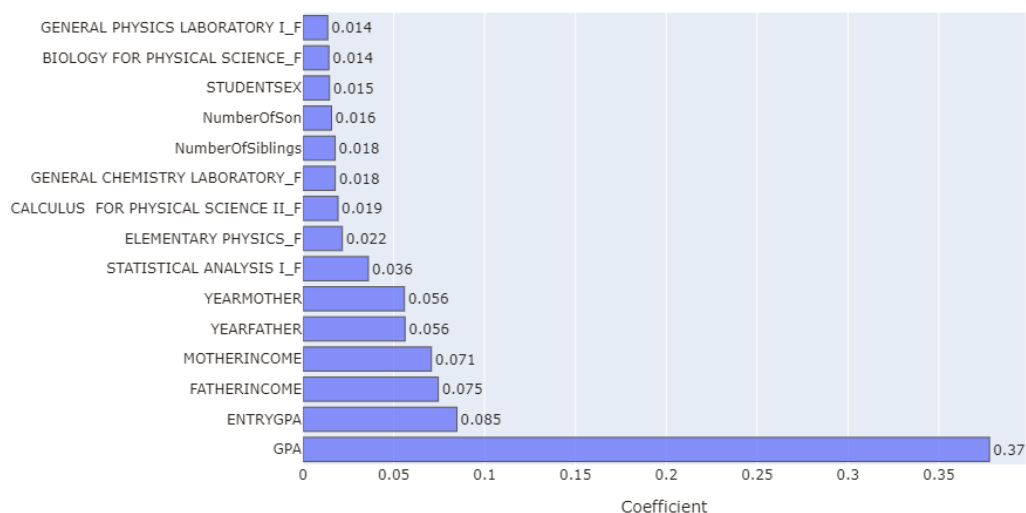
จากการทดลองในการสร้างตัวจำแนกประเภทการผันสภาพของการผันสภาพกรณีที่ 2 โดยใช้  
อัลกอริทึม Random Forest พบว่า ปัจจัยที่สำคัญต่อการผันสภาพของนักศึกษาประกอบด้วย เกรดเฉลี่ย  
, เกรดเฉลี่ยระดับมัธยม, อายุบิดา, อายุมารดา, เกรด F รายวิชา ELEMENTARY PHYSICS, รายได้มารดา,  
รายได้บิดา, เกรด F รายวิชา INTRODUCTION TO INFORMATION AND COMMUNICATION  
TECHNOLOGY, GENERAL PHYSICAL LABORATORY, CALCULUS FOR PHYSICAL SCIENCE I,  
จำนวนพี่น้อง ตามลำดับ แสดงในรูปที่ 20

### 3. ข้อมูลการฝึกสภาพกรณีที่ 3

ข้อมูลการฝึกสภาพกรณีที่ 3 โดยสถานะการฝึกสภาพของนักศึกษาประกอบไปด้วย ฝึกสภาพเนื่องจากลาออก และฝึกสภาพเนื่องจากตกออก จากการทดลองในการสร้างการจำแนกประเภทแบบเดี่ยว และการจำแนกประเภทแบบรวมกลุ่ม พบว่า การใช้การจำแนกประเภทแบบรวมกลุ่มที่ใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ให้ประสิทธิภาพไม่แตกต่างจากการใช้การจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม DT และ SVM อัลกอริทึมที่มีประสิทธิภาพในการจำแนกประเภทการฝึกสภาพที่ดีที่สุด คือ Random Forest ซึ่งให้ประสิทธิภาพค่าความถูกต้อง (Accuracy) อยู่ที่ 99% ค่าความแม่นยำ (Precision) อยู่ที่ 99% ค่าความครบถ้วน (Recall) อยู่ที่ 99% ค่าความถ่วงดุล (f1-score) อยู่ที่ 99% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC อยู่ที่ 99% สำหรับการจำแนกประเภทการฝึกสภาพของนักศึกษา แสดงในตารางที่ 49

ตาราง 49 การวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทการฝึกสภาพของข้อมูลการฝึกสภาพกรณีที่ 3

Performance	Single Model		Ensemble Model				
	DT	SVM	Bagging		Boosting		Random Forest
			DT	SVM	DT	SVM	
Accuracy	0.97	0.87	0.97	0.89	<u>0.99</u>	0.88	<u>0.99</u>
Precision	0.97	0.87	0.97	0.89	<u>0.99</u>	0.88	<u>0.99</u>
Recall	0.97	0.87	0.97	0.89	<u>0.99</u>	0.88	<u>0.99</u>
F1-score	0.97	0.87	0.97	0.89	<u>0.99</u>	0.88	<u>0.99</u>
AUC	0.96	0.93	<u>0.99</u>	0.89	0.95	0.95	<u>0.99</u>



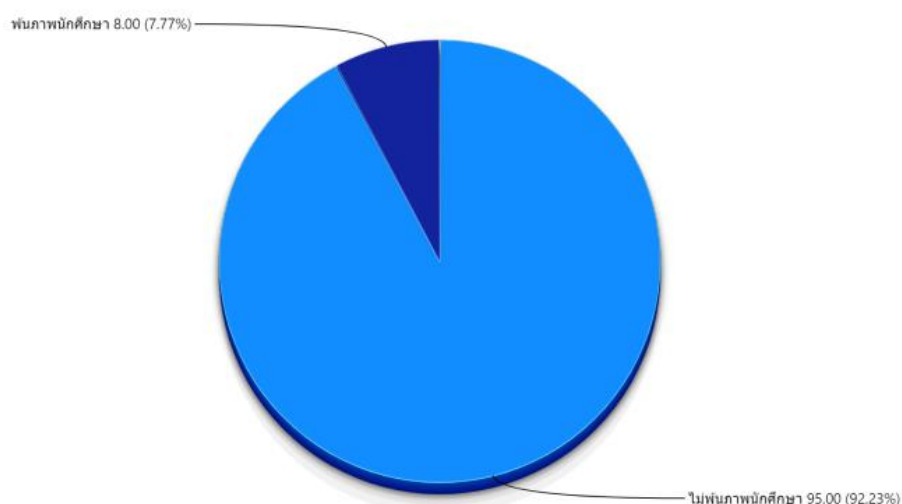
รูปที่ 21 ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาของการพัฒนากรณีที่ 3

จากการทดลองในการสร้างตัวจำแนกประเภทการพัฒนาของนักศึกษากรณีที่ 3 โดยใช้ อัลกอริทึม Random Forest พบว่า ปัจจัยที่สำคัญต่อการพัฒนาของนักศึกษาประกอบด้วย เกรดเฉลี่ย , เกรดเฉลี่ยระดับมัธยม, รายได้บิดา, รายได้มารดา, อายุบิดา, อายุมารดา, เกรด F รายวิชา STATISTICAL ANALYSIS I, ELEMENTARY PHYSICS, CALCULUS FOR PHYSICAL SCIENCE II, GENERAL CHEMISTRY LABORATORY ตามลำดับ แสดงในรูปที่ 21

### ทำนายการฟื้นสภาพของนักศึกษาปีการศึกษา 2564

ผลลัพธ์จากการเลือกแบบจำลองที่ดีที่สุดโดยใช้อัลกอริทึม Random Forest สำหรับการเรียนรู้ข้อมูลการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 โดยทำการแบ่งชุดข้อมูลออกเป็น 3 กลุ่มตัวอย่าง เพื่อทำการจำแนกประเภทหรือทำนายการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 ดังนี้

#### 1. ข้อมูลการฟื้นสภาพกรณีที่ 1

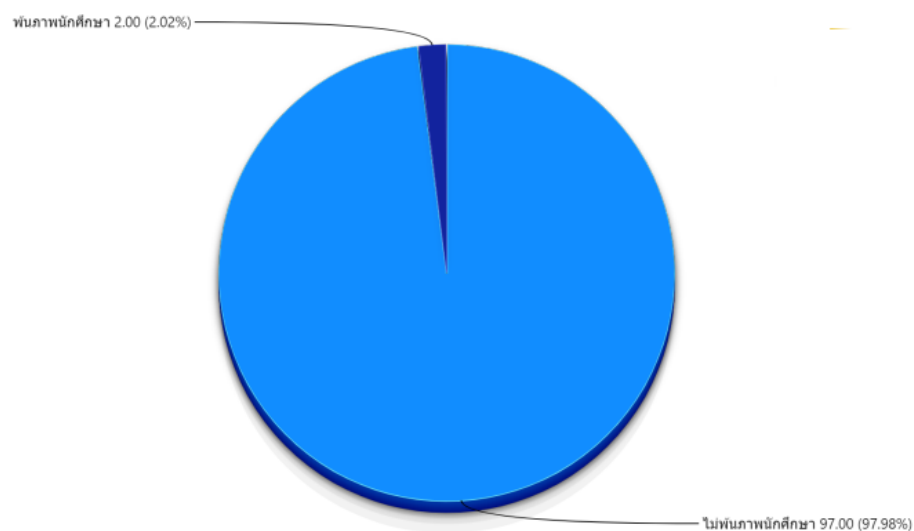


รูปที่ 22 แผนภูมิวงกลมของการทำนายการฟื้นสภาพนักศึกษา

ผลลัพธ์จากการใช้งานใช้อัลกอริทึม Random Forest สำหรับการเรียนรู้ข้อมูลการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 ของข้อมูลการฟื้นสภาพกรณีที่ 1 มีจำนวน 103 คน โดยที่สถานะการฟื้นสภาพของนักศึกษาประกอบไปด้วย ฟื้นสภาพเนื่องจากลาออก ฟื้นสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา ฟื้นสภาพเนื่องจากตกออก และฟื้นสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด พบว่านักศึกษาที่ถูกทำนายว่าไม่ฟื้นสภาพมีจำนวน 95 คนคิดเป็น 92.23% และนักศึกษาที่ถูกทำนายว่าฟื้นสภาพนักศึกษามีจำนวน 8 คนคิดเป็น 7.77% แสดงในรูปที่ 22



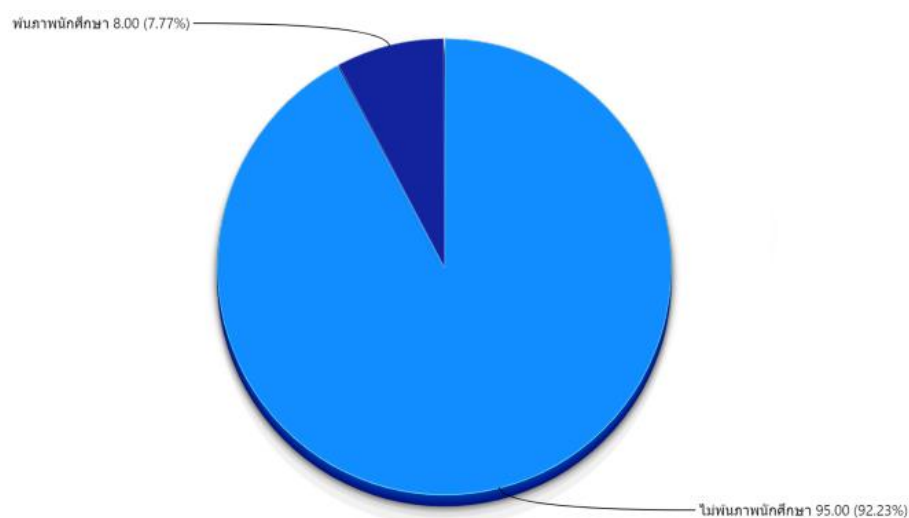
## 2.ข้อมูลการฟื้นสภาพกรณีที่ 2



รูปที่ 23 แผนภูมิวงกลมของการทำนายการฟื้นสภาพนักศึกษา

ผลลัพธ์จากการใช้งานใช้อัลกอริทึม Random Forest สำหรับการเรียนรู้ข้อมูลการฟื้นสภาพของนักศึกษาปีการศึกษา 2564 ของข้อมูลการฟื้นสภาพกรณีที่ 1 มีจำนวน 9 คน โดยที่สถานะการฟื้นสภาพของนักศึกษาประกอบไปด้วย ฟื้นสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา และฟื้นสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด พบว่า นักศึกษาที่ถูกทำนายว่าไม่ฟื้นสภาพมีจำนวน 97 คนคิดเป็น 97.98% และนักศึกษาที่ถูกทำนายว่าฟื้นสภาพนักศึกษามีจำนวน 2 คนคิดเป็น 2.02% แสดงในรูปที่ 23

### 3.ข้อมูลการพัฒนสภาพกรณ์ที่ 3



รูปที่ 24 แผนภูมิวงกลมของการทำนายการพัฒนสภาพนักศึกษา

ผลลัพธ์จากการใช้งานใช้อัลกอริทึม Random Forest สำหรับการเรียนรู้ข้อมูลการพัฒนสภาพของนักศึกษาปีการศึกษา 2564 ของข้อมูลการพัฒนสภาพกรณ์ที่ 1 มีจำนวน 103 คน โดยที่สถานะการพัฒนสภาพของนักศึกษาประกอบไปด้วย พัฒนสภาพเนื่องจากลาออก และพัฒนสภาพเนื่องจากตกรอก พบว่านักศึกษาที่ถูกทำนายว่าไม่พัฒนสภาพมีจำนวน 95 คนคิดเป็น 92.23% และนักศึกษาที่ถูกทำนายว่าพัฒนสภาพนักศึกษามีจำนวน 8 คนคิดเป็น 7.77% แสดงในรูปที่ 24

## บทที่ 5

### สรุปผลการวิจัย

#### 5.1 สรุปผลการวิจัย

ในงานวิจัยครั้งนี้มีจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพความถูกต้องของการจำแนกประเภทการพัฒนสภาพของนักศึกษาด้วยวิธีการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว (Single Classification) กับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม (Ensemble Classification) เพื่อหาประสิทธิภาพ ความถูกต้อง และความแม่นยำ สำหรับการจำแนก ประเภทการพัฒนสภาพของนักศึกษาที่ดีที่สุด โดยใช้ข้อมูลของนักศึกษาระดับปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ โดยใช้ข้อมูลเฉพาะนักศึกษาปี 1 ในแต่ละปีการศึกษา 6 ปีย้อนหลังตั้งแต่ปี พ.ศ. 2558 – 2563 มีจำนวน 614 คน โดยได้ทำการแบ่งชุดข้อมูลที่ใช้สำหรับการทดลองออกเป็น 3 กลุ่มตัวอย่าง จากนั้นนำข้อมูลแต่ละกลุ่มตัวอย่างมาผ่านกระบวนการทำความสะอาดข้อมูล (Clean Data) ตัวแปรหุ่น (Dummy Variable) ทำการคัดเลือกข้อมูล (Selection Data) ปรับสมดุลข้อมูล (Balancing Data) และปรับปรุงขอบเขตข้อมูล (Data Scaling) ก่อนการเรียนรู้หรือสร้างตัวจำแนกประเภท จากผลการวิจัย พบว่า วิธีการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม (Ensemble Model) มีประสิทธิภาพการจำแนกประเภทการพัฒนสภาพของนักศึกษาได้ดีกว่าวิธีการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว (Single Model) ซึ่งมีความสอดคล้องตรงตามสมมติฐานที่ตั้งไว้ จากการเปรียบเทียบประสิทธิภาพการจำแนกประเภทการพัฒนสภาพของนักศึกษา พบว่า การใช้อัลกอริทึม Random Forest สำหรับการเรียนรู้ให้ประสิทธิภาพในการจำแนกทั้ง 3 กลุ่มตัวอย่างดีที่สุด โดยประสิทธิภาพทั้ง 3 กลุ่มตัวอย่างให้ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) ค่าความถ่วงดุล (F1-Score) มากกว่า 98% และค่า AUC พื้นที่ใต้เส้นโค้ง ROC มากกว่า 99% สำหรับการจำแนกประเภทการพัฒนสภาพของนักศึกษา

ปัจจัยที่ส่งผลต่อการพัฒนสภาพของนักศึกษาของข้อมูลการพัฒนสภาพกรณีที่ 1 โดยสถานะการพัฒนสภาพของนักศึกษาประกอบไปด้วย พัฒนสภาพเนื่องจากลาออก พัฒนสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา พัฒนสภาพเนื่องจากตกออก และพัฒนสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด พบว่าปัจจัยที่สำคัญต่อการพัฒนสภาพของนักศึกษาประกอบด้วย เกรดเฉลี่ย, เกรดเฉลี่ยระดับมัธยม, รายได้บิดา, รายได้มารดา, อายุมารดา, อายุบิดา, เกรด F รายวิชา STATISTICAL ANALYSIS I, ELEMENTARY PHYSICS, GENERAL CHEMISTRY LABORATORY, จำนวนพี่น้อง ตามลำดับ

ปัจจัยที่ส่งผลต่อการพัฒนสภาพของนักศึกษาของข้อมูลการพัฒนสภาพกรณีที่ 2 โดยสถานะการพัฒนสภาพของนักศึกษาประกอบไปด้วย พัฒนสภาพเนื่องจากไม่ชำระค่าต่อทะเบียนนักศึกษา และพัฒนสภาพเนื่องจากไม่ลงทะเบียนตามเวลากำหนด พบว่าปัจจัยที่สำคัญต่อการพัฒนสภาพของนักศึกษาประกอบด้วย

เกรดเฉลี่ย, เกรดเฉลี่ยระดับมัธยม, อายุบิดา, อายุมารดา, เกรด F รายวิชา ELEMENTARY PHYSICS, รายได้มารดา, รายได้บิดา, เกรด F รายวิชา INTRODUCTION TO INFORMATION AND COMMUNICATION TECHNOLOGY, GENERAL PHYSICAL LABORATORY, CALCULUS FOR PHYSICAL SCIENCE I, จำนวนพี่น้อง ตามลำดับ

ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาของข้อมูลการพัฒนากรณีที่ 3 โดยสถานะการพัฒนาของนักศึกษาประกอบไปด้วย พัฒนาการเนื่องจากลาออก และพัฒนาการเนื่องจากตกรอก พบว่าปัจจัยที่สำคัญต่อการพัฒนาของนักศึกษาประกอบด้วย เกรดเฉลี่ย, เกรดเฉลี่ยระดับมัธยม, รายได้บิดา, รายได้มารดา, อายุบิดา, อายุมารดา, เกรด F รายวิชา STATISTICAL ANALYSIS I, ELEMENTARY PHYSICS, CALCULUS FOR PHYSICAL SCIENCE II, GENERAL CHEMISTRY LABORATORY ตามลำดับ

จากการทำนายการพัฒนาของนักศึกษาศึกษาปีการศึกษา 2564 ของข้อมูลการพัฒนากรณีที่ 1, 2, 3 โดยใช้อัลกอริทึม Random Forest พบว่านักศึกษาที่ถูกทำนายว่าไม่พัฒนาคิดเป็น 91.23%, 97.98%, 92.23% และนักศึกษาที่ถูกทำนายว่าพัฒนาพัฒนานักศึกษาคิดเป็น 7.77%, 2.02%, 7.77% จากจำนวน 103, 99, 103 คนตามลำดับ

## 5.2 อภิปรายผลการวิจัย

การวิจัยเรื่องการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชา สถิติ คณะวิทยาศาสตร์ โดยใช้วิธีการเรียนรู้แบบรวมกลุ่มสามารถอภิปรายผลการวิจัยได้ดังนี้

สำหรับการเปรียบเทียบประสิทธิภาพการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม DT และ SVM และการจำแนกประเภทแบบรวมกลุ่มที่ใช้ DT และ SVM เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ พบว่า การใช้เทคนิคการจำแนกประเภทแบบรวมกลุ่มให้ประสิทธิภาพความแม่นยำมากกว่าการจำแนกประเภทแบบเดี่ยว และอัลกอริทึม Random Forest ให้ประสิทธิภาพความแม่นยำมากที่สุดสำหรับการจำแนกการพัฒนาของนักศึกษา ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ และงานวิจัยของ (ปัทม์ อุปการ, 2560) ที่กล่าวไว้ว่าการใช้อัลกอริทึม Random Forest สามารถเพิ่มเติมในส่วนของฟังก์ชันการทำงานแบบ สุ่มเลือกคุณลักษณะ ของข้อมูลที่ใช้ในการวิเคราะห์เข้ามา ทำให้ลดค่าสหสัมพันธ์ (Correlation) ของคุณลักษณะลงในการสร้างต้นไม้แต่ละต้นที่มีความเป็นอิสระต่อกัน จึงทำให้ต้นไม้แต่ละต้นที่ถูกสร้างขึ้น เพื่อใช้สำหรับการจำแนกประเภทข้อมูลมีโครงสร้างต้นไม้ที่มีขนาดเล็ก ซึ่งจะทำงานได้เร็วและให้ประสิทธิภาพที่ดีกว่าวิธีการ แบบอื่นๆ

สำหรับการศึกษาปัจจัยของการพัฒนากรณีที่ 1, 2, 3 โดยใช้อัลกอริทึม Random Forest พบว่า เกรดเฉลี่ย เกรดเฉลี่ยก่อนเข้ามหาวิทยาลัย รายได้บิดามารดา อายุของบิดามารดา และการติด F ในรายวิชาต่างๆเป็นปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษา ซึ่งมีปัจจัยบางตัวที่สอดคล้องกับงานวิจัย

ของ (Tenpipat, & Akkarajitsakul, 2020) (Naseem et al., 2020) (นนทวัฒน์ ทวีชาติ และคณะ, 2563) ที่ได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษา

### 5.3 ประโยชน์ของสถิติที่ใช้ในการวิเคราะห์

การวิจัยเรื่องการจำแนกประเภทสำหรับการพัฒนาของนักศึกษาระดับปริญญาตรี สาขาวิชา สถิติ คณะวิทยาศาสตร์ ซึ่งตามวัตถุประสงค์ได้ทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทแบบ เดี่ยว (Single Classification) และแบบรวมกลุ่ม (Ensemble Classification) พบว่าการใช้อัลกอริทึม Random Forest จำแนกการพัฒนาของนักศึกษาได้ดีที่สุด และสามารถทราบปัจจัยที่สำคัญต่อการ ภาพของนักศึกษาเพื่อที่จะทำการหาแนวทางในการป้องกัน แก้ไขปัญหาการ และวางกลยุทธ์ในการ จัดการการพัฒนาของนักศึกษาได้อย่างทัน่วงที่

### 5.4 ข้อเสนอแนะ

5.4.1 นำข้อมูลการพัฒนาของนักศึกษาแต่ละคณะให้อัลกอริทึมเรียนรู้ เพื่อที่จะเพิ่ม ประสิทธิภาพในการทำนายการพัฒนาของนักศึกษาได้แม่นยำมากขึ้น

5.4.2 สามารถนำอัลกอริทึมไปใช้ในการพัฒนาเว็บแอปพลิเคชัน (web application) ให้สามารถ ใช้งานได้จริง ในการทำนายการพัฒนาของนักศึกษา

5.4.3 การนำอัลกอริทึมการตรวจจับค่าผิดปกติ (Outlier Detection Algorithms) ต่างๆมา ประยุกต์ใช้สำหรับการลบค่าที่ผิดปกติออกจากข้อมูล เช่น Isolation Forest, Minimum Covariance Determinant, Local Outlier Factor, และ One-Class SVM

5.4.4 งานวิจัยนี้ผู้วิจัยได้ทำการปรับขอบเขตข้อมูลโดยใช้วิธีการ Min-Max Normalization นอกเหนือจากนี้สามารถนำการทำ Feature Engineering ต่างๆมาประยุกต์ใช้สำหรับกระบวนการ จัดเตรียมข้อมูลอาจทำให้ประสิทธิภาพในการจำแนกการพัฒนาของนักศึกษาเพิ่มสูงขึ้นไม่มากนักน้อย เช่น Principal Component Analysis (PCA), K-Means, Encoding, และ Transform Data

5.4.5 งานวิจัยนี้มุ่งเน้นที่การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มโดยใช้วิธี Bagging Boosting, และ Random Forest ซึ่งแบบจำลองการจำแนกประเภทแบบรวมกลุ่มมีวิธีการอื่นๆ อีกมากมาย ตัวอย่างเช่น Stacking, Blending, และ Voting นอกจากนี้ยังมีโมเดลที่ถูกพัฒนามาจาก วิธีการ Bagging และ Boosting เช่น XGBoost, GBM, Light GBM, CatBoost เป็นต้น

5.4.6 การนำอัลกอริทึมโครงข่ายประสาทเทียม (Neuron Network) หรือ การเรียนรู้เชิงลึก (Deep Learning) มาสร้างแบบจำลองการจำแนกการพัฒนาของนักศึกษา อาจทำให้ประสิทธิภาพใน การจำแนกการพัฒนาของนักศึกษาเพิ่มสูงขึ้นไม่มากนักน้อย เช่น Artificial Neural Network (ANN) และ Convolutional Neural Network (CNN)

### เอกสารอ้างอิง

- ชนิดาภา บุญประสม และจรัญ แสนราช. (2561). การวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษาในระดับปริญญาตรี โดยใช้เทคนิควิธีการทำเหมืองข้อมูล. *Technical Education Journal King Mongkut's University of Technology North Bangkok*, 9(1), 142-151.
- ชอและ เกป็น, พิมลพรรณ สีสากัทรพันธ์, และอัจฉราพร ยกขุน. (2561). การวิเคราะห์ปัจจัยที่มีผลต่อการพัฒนาของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษา หลักสูตรวิทยาการคอมพิวเตอร์หลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา. *Veridian E-Journal, Science and Technology Silpakorn University*, 5(4), 96-110.
- ธันท์ จระณะสมบุรณ์ และ วราภรณ์ วิทยานนท์. (2561). การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่อง (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ
- นนทวัฒน์ ทวีชาติ, อรยา เพ็งประจัญ, วิไลรัตน์ ยาทองไชย, และชูศักดิ์ ยาทองไชย. (2563). ระบบทำนายการพัฒนาของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองคอม. *Journal of Science and Technology Buriram Rajabhat*, 4(1), 47-60.
- นิสานันท์ พลอาสา. (2558). การสร้างแบบจำลองการขายผลิตภัณฑ์และพยากรณ์ยอดขายประกันชีวิต โดยเทคนิคการทำเหมืองข้อมูล กรณีศึกษา บริษัทประกันชีวิตแห่งหนึ่ง วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์
- ปัทมธนา บุญรักษา และ จาริ ทองคำ. (2560). การเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดอุบัติเหตุทางถนน โดยใช้เทคนิคอนุกรมเวลา. *วารสารวิชาการ การจัดการเทคโนโลยีสารสนเทศและนวัตกรรม*, 4(2), 40-46.
- ปัทมัท อุปการ. (2560). การปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้วิธีการเรียนรู้แบบรวมกลุ่ม (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). พิษณุโลก: มหาวิทยาลัยนเรศวร
- วิชญ์วิสิฐ เกษรสิทธิ์, ดร.วิชิต หล่อจิระชุมท์กุล, และดร.จิราวัลย์ จิตรถเวช. (2561). การแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลสำหรับการจำแนกผู้ป่วยโรคเบาหวาน. *KKU Research Journal (Graduate Study)*, 18(3), 11-21.
- สุกัญญา ทารส. (2562). ปัจจัยจำแนกการออกกลางคันของนิสิตปริญญาตรี มหาวิทยาลัยมหาสารคาม วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). มหาสารคาม: มหาวิทยาลัยมหาสารคาม
- สำนักทะเบียนมหาวิทยาลัยขอนแก่น. (2562). ระเบียบมหาวิทยาลัยขอนแก่นว่าด้วย การศึกษาชั้น

- ปริญญาตรี พ.ศ. ๒๕๖๑. สืบค้น 4 ตุลาคม 2564. จาก/  
[https://home.kku.ac.th/meeting/Document/KKU\\_R2562-bachelorgegree.pdf](https://home.kku.ac.th/meeting/Document/KKU_R2562-bachelorgegree.pdf)  
 สำนักทะเบียนมหาวิทยาลัยขอนแก่น. (2564). ทะเบียนรายชื่อนักศึกษา จากระบบของสำนักทะเบียน  
 มหาวิทยาลัยขอนแก่น. สืบค้น 4 ตุลาคม 2564. จาก <https://reg.kku.ac.th/>
- Naseem, M., Chaudhary, K., Sharma, B., & Lal, A., G. (2019). Using Ensemble Decision  
 Tree Model to Predict Student Dropout in Computing Science. *IEEE Asia-  
 Pacific Conference on Computer Science and Data Engineering (CSDE)*.  
 doi: 10.1109/CSDE48274.2019.9162389
- Gatchalee, P. (2019). Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการ  
 ทำนาย ในMachine learning. สืบค้น 15 กรกฎาคม 2564,  
 จาก [https://medium.com/@pagongatchalee/confusion-matrix-เครื่องมือสำคัญใน  
 การประเมินผลลัพธ์ของการทำนาย-ในmachine-learning-fba6e3f9508c](https://medium.com/@pagongatchalee/confusion-matrix-เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย-ในmachine-learning-fba6e3f9508c)
- Tenpipat, W., & Akkarajitsakul, K. (2020). Student Dropout Prediction:  
 A KMUTT Case Study. 2020 1st International Conference on Big Data  
 analytics and Practices (IBDAP). doi: 10.1109/IBDAP50342.2020.9245457
- Uddin, S., Khan, A., & Hossain, M. (2019). Comparing different supervised machine  
 learning algorithms for disease prediction. *BMC Med Inform Decis Mak*,  
 19(1), 1-16. doi: 10.1186/s12911-019-1004-8

ภาคผนวก ก

ตารางการดำเนินโครงการวิจัย  
และค่าใช้จ่ายในการดำเนินโครงการ





การวิจัยครั้งนี้มีงบประมาณค่าใช้จ่ายในการดำเนินงานดังนี้

ตาราง 51 การดำเนินงานโครงการวิจัย

รายการ	จำนวนเงิน
จัดทำเอกสารต่าง ๆ ในงานวิจัย	0
จัดทำรูปเล่มฉบับสมบูรณ์	0
จัดทำโปสเตอร์	0
รวมเป็นเงิน	0