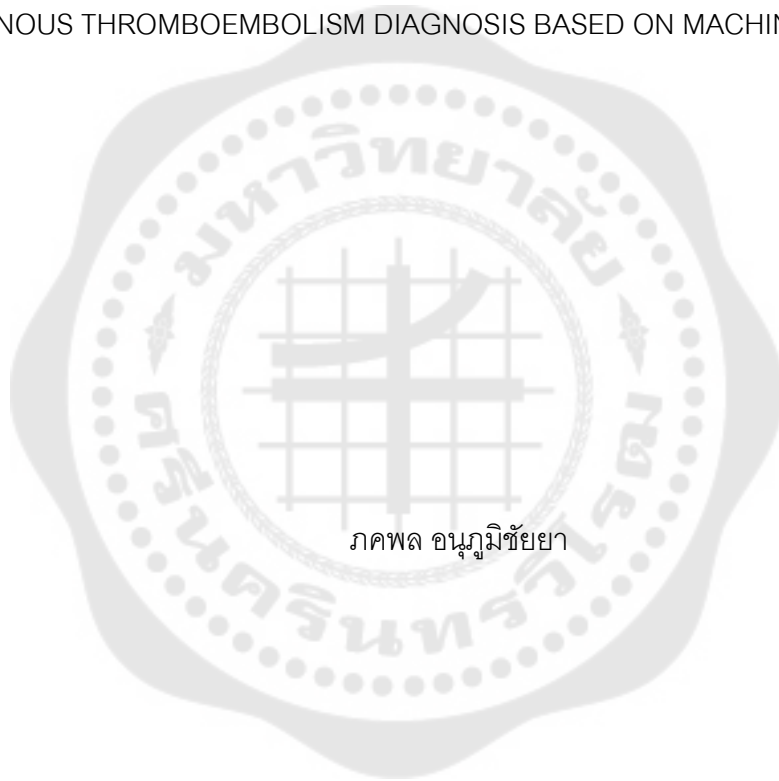




การวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง
VENOUS THROMBOEMBOLISM DIAGNOSIS BASED ON MACHINE LEARNING



ภาคพล อนุภูมิชัยยา

การวินิจฉัยภาวะลิ้มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2564
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

VENOUS THROMBOEMBOLISM DIAGNOSIS BASED ON MACHINE LEARNING



PAKAPOL ANUPOOMCHAIYA

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2021

Copyright of Srinakharinwirot University

ปริญญาบัตร

เรื่อง

การวินิจฉัยภาวะล้มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง

ของ

ภาคพล อนุภูมิซัยยา

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าปริญญาบัตร

ที่ปรึกษาหลัก

(อาจารย์ ดร.วีระ สอึ้ง)

ประธาน

(รศ. ดร.สุพัฒนา เอื้อทวิเกียรติ)

กรรมการ

(อาจารย์ ดร.ศุภร คนธภักดี)

ชื่อเรื่อง	การวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง
ผู้วิจัย	ภคพล อนุภูมิชัยยา
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2564
อาจารย์ที่ปรึกษา	อาจารย์ ดร. วีระ สอิ่ง

การเกิดลิ่มเลือดอุดตันเป็นสาเหตุหลักของลิ่มเลือดที่ขัดขวางการไหลเวียนของเลือดทั้งในหลอดเลือดแดงหรือหลอดเลือดดำ ดังนั้นการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ (Venous Thromboembolism - VTE) จึงเป็นสาเหตุที่ร้ายแรงที่สุดของโรคหัวใจและหลอดเลือด นอกจากนี้โรคนี้ยังเป็นสาเหตุการเสียชีวิตอันดับต้นๆ ในประเทศไทย เนื่องจากขาดความระมัดระวังและความเข้าใจ ในงานวิจัยนี้ขอเสนอแบบจำลองการวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่องที่มีประสิทธิภาพในการทำนายปัจจัยเสี่ยงที่สำคัญสำหรับ VTE เก็บข้อมูลจากหอผู้ป่วยที่โรงพยาบาลจุฬาลงกรณ์ ประเทศไทย ข้อมูลประกอบด้วย 1,290 แถวและ 65 คอลัมน์ งานวิจัยนี้จะแยกข้อมูลที่เตรียมไว้ออกเป็นชุดข้อมูลการฝึกอบรมและชุดข้อมูลการทดสอบด้วยอัตราส่วน 75:25, 70:30 และ 67:33 และเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งสามแบบ ประกอบด้วย Adaptive Boosting (AdaBoost), Random Forest (RF) และ Extreme Gradient Boosting (XGBoost) เพื่อหาแบบจำลองที่ดีที่สุด จากผลการทดลอง แบบจำลอง Random Forest โดยใช้พารามิเตอร์ Class Weight แบบคำนวณ ใช้วิธีการแก้ปัญหาความไม่สมดุลของข้อมูลด้วยวิธี Oversampling กับพารามิเตอร์ sampling_strategy เท่ากับ 0.50 เป็นแบบจำลองที่มีประสิทธิภาพที่ดีที่สุด โดยความแม่นยำในการทำนายอยู่ที่ 99.61% ดังนั้น Random Forest จะเป็นวิธีที่จะช่วยให้แพทย์พิจารณาความเสี่ยงของการเกิดลิ่มเลือดอุดตันหลอดเลือดดำ นอกจากนี้ยังสามารถใช้เพื่อคาดการณ์การเกิด VTE โดยพิจารณาจากปัจจัยเสี่ยงที่สำคัญร่วมกัน

คำสำคัญ : ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ, ปัญหาความไม่สมดุลของข้อมูล, หลักการเรียนรู้ของเครื่อง, วิธีการแก้ปัญหาความไม่สมดุลของข้อมูล

Title	VENOUS THROMBOEMBOLISM DIAGNOSIS BASED ON MACHINE LEARNING
Author	PAKAPOL ANUPOOMCHAIYA
Degree	MASTER OF SCIENCE
Academic Year	2021
Thesis Advisor	Vera Sa-ing , Ph.D.

Thrombosis is the main cause of blood clots that obstruct the flow of blood in an artery or venous thrombosis. Thus, venous thromboembolism (VTE) is the most serious cause of cardiovascular disease. Furthermore, this disease is a leading cause of death in Thailand because of a lack of caution and understanding. In this research, an automatic diagnosis model was proposed by using effective machine learning to predict the important risk factors for VTE. The raw data were collected from the medical ward at King Chulalongkorn Memorial Hospital, Thailand. Before the analysis, this data consisted of 1,290 rows and 65 columns that were analyzed, solved, and transformed into prepared data. By resampling algorithms to import into each model, this research splits the prepared data into the training dataset and the testing dataset with a ratio of 75:25, 70:30, and 67:33. In these experiments, this research compared the effectiveness of the three machine learning models, consisting of Adaptive Boosting (AdaBoost), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) to find the best diagnostic model. According to the experimental results, the Random Forest model was computed by the class weight and oversampled by the sampling strategy with 0.50 technique is the most efficient model to represent the prediction accuracy of 99.61%. Therefore, the Random Forest and the proposed setting will assist medical doctors in determining the risk of symptomatic venous thromboembolism. In addition, the proposed model can be used to forecast the likelihood of VTE based on a combination of the important risk factors.

Keyword : VENOUS THROMBOEMBOLISM, MACHINE LEARNING, IMBALANCED DATA, RESAMPLING DATA METHOD

กิตติกรรมประกาศ

ปริญญาานิพนธ์นี้สำเร็จลุล่วงได้ด้วยความสามารถ ความช่วยเหลือ และความเอาใจใส่ตลอดจนการให้คำแนะนำและข้อคิดเห็นที่เป็นประโยชน์อย่างยิ่งสำหรับการปรับแก้ไขข้อบกพร่องจากคณะกรรมการผู้ควบคุมปริญญาานิพนธ์ ผู้วิจัยขอขอบคุณ อาจารย์ ดร.วีระ สอึ้ง ที่ได้กรุณาให้คำแนะนำ ให้ความช่วยเหลือ ติดตามการทำปริญญาานิพนธ์อย่างใกล้ชิด และชี้แนะแนวทางสิ่งที่เป็นประโยชน์ต่อการทำวิจัยในครั้งนี้

ขอขอบคุณ รองศาสตราจารย์ ดร.สุพัฒนา เอื้อทิวเกียรติ ประธานผู้ทรงคุณวุฒิ ในการสอบปากเปล่าปริญญาานิพนธ์ที่กรุณาให้คำแนะนำและข้อเสนอแนะเกี่ยวกับงานวิจัย เพื่อนำความรู้ที่ได้ไปปรับใช้ในอนาคต

ขอขอบคุณ ศาสตราจารย์ ดร.นายแพทย์พลภัทร โรจน์นครินทร์ ที่สนับสนุนให้ข้อมูลในการทำวิจัยครั้งนี้

ขอขอบคุณ คณาจารย์ทุกท่านและกรรมการบริหารหลักสูตร สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ให้ทั้งความรู้ ประสบการณ์อันเป็นประโยชน์ต่อการศึกษาศึกษา

ขอขอบคุณเพื่อนนิสิตปริญญาโท รหัส 63 สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ทุกคนที่ได้ให้ความช่วยเหลือและเกื้อกูลกัน

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ที่รู้จักทุกท่าน ที่ให้ความช่วยเหลือ และเป็นกำลังใจให้กับผู้วิจัยตลอดมา

ขอขอบคุณครูบาอาจารย์ทุกท่าน ที่เคยอบรมสั่งสอน ให้ความรู้กับผู้วิจัย

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณ บิดา-มารดา และครอบครัว ที่ให้กำลังใจจนสำเร็จการศึกษา

ภคพล อนุภูมิชัยยา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ผ
สารบัญรูปภาพ	ฝ
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและความเป็นมาของงานวิจัย.....	1
1.2 วัตถุประสงค์ของงานวิจัย	4
1.3 ขอบเขตของงานวิจัย	4
1.4 ขั้นตอนการทำงานวิจัย.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย	5
บทที่ 2 วรรณกรรม และงานวิจัยที่เกี่ยวข้อง.....	6
2.1 ทฤษฎีเกี่ยวกับภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ	6
2.1.1 ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำในชั้นลึกที่ขา (Deep Vein Thromboembolism: DVT)	6
2.1.2 ภาวะลิ่มเลือดอุดตันในปอด (Pulmonary Embolism: PE)	7
2.1.3 อาการของภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน.....	8
2.1.4 การวินิจฉัยภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน	8
2.1.5 การรักษาภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน.....	10
2.2 ข้อมูลที่สำคัญต่อการวิเคราะห์ภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ	11

2.3 ทฤษฎีการวิเคราะห์ข้อมูลด้วยหลักการทางคอมพิวเตอร์.....	12
2.3.1 การจำแนกประเภทข้อมูล (Classification)	12
2.3.1.1 ข้อมูลเชิงคุณภาพ	12
2.3.1.2 ข้อมูลเชิงปริมาณ.....	12
2.3.2 การทำความเข้าใจกับข้อมูล (Data understanding)	13
2.3.3 การเตรียมข้อมูล (Data Preparation)	13
2.3.4 การสร้างโมเดล (Modeling).....	13
2.3.5 การประเมินผล (Evaluation)	14
2.4 งานวิจัยที่ศึกษาเกี่ยวกับปัจจัยการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ	14
2.4.1 งานวิจัยเรื่อง Risk Factors for symptomatic venous thromboembolism in Thai hospitalized medical patients (Rojnuckarin et al., 2011)	14
2.4.2 งานวิจัยเรื่อง Predictive and associative models to identify hospitalized medical patients at risk for VTE (Spyropoulos et al., 2011)	15
2.5 งานวิจัยที่ศึกษาเกี่ยวกับการวิเคราะห์ความเสี่ยงของภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ	17
2.5.1 งานวิจัยเรื่อง Automatic Diagnosis of Venous Thromboembolism Risk based on Machine Learning (Sukperm et al., 2021)	17
2.5.2 งานวิจัยเรื่อง A Machine Learning Approach to Predict Deep Venous Thrombosis Among Hospitalized Patients (Ryan et al., 2021)	20
2.5.3 งานวิจัยเรื่อง Machine learning to predict venous thrombosis in acutely ill medical patients (Nafee et al., 2020).....	21
บทที่ 3 วิธีการดำเนินการวิจัย.....	26
3.1 กระบวนการทำงานของแบบจำลอง	26
3.2 การเก็บรวบรวมข้อมูล (Data Acquisition).....	27

4.1.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	53
4.1.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	54
4.1.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	55
4.1.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	56
4.1.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	56
4.1.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	57
4.1.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	58
4.2 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Default Parameters และ Balanced Bagging Classifier	59
4.2.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	59
4.2.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	59
4.2.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	60
4.2.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	61
4.2.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	62
4.2.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	62
4.2.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	63
4.2.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	64
4.2.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	65
4.2.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	65

4.2.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	66
4.2.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	67
4.3 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Tuning Hyperparameter	68
4.3.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	68
4.3.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	68
4.3.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	69
4.3.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	70
4.3.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	71
4.3.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	71
4.3.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	72
4.3.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	73
4.3.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	74
4.3.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	74
4.3.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	75
4.3.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	76
4.4 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier.....	77
4.4.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	77
4.4.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	77

4.4.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	78
4.4.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	79
4.4.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	80
4.4.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	80
4.4.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	81
4.4.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	82
4.4.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	83
4.4.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	83
4.4.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	84
4.4.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	85
4.5 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters	86
4.5.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	86
4.5.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	86
4.5.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	87
4.5.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	88
4.5.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	89
4.5.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	89
4.5.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	90
4.5.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	91

4.5.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	92
4.5.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	92
4.5.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	93
4.5.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	94
4.6 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Balanced Class Weight	95
4.6.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	95
4.6.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	95
4.6.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	96
4.6.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	97
4.6.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	98
4.6.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	98
4.6.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	99
4.6.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	100
4.6.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	101
4.6.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	101
4.6.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	102
4.6.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	103
4.7 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Ratio Class Weight.....	104

4.7.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	104
4.7.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	104
4.7.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	105
4.7.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	106
4.7.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	107
4.7.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	107
4.7.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	108
4.7.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	109
4.7.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	110
4.7.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	110
4.7.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	111
4.7.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	112
4.8 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Balanced Bagging Classifier.....	113
4.8.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	113
4.8.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	113
4.8.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	114
4.8.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	115
4.8.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	116
4.8.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	116

4.8.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	117
4.8.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	118
4.8.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	119
4.8.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	119
4.8.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	120
4.8.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	121
4.9 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters.....	122
4.9.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	122
4.9.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	122
4.9.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	123
4.9.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	124
4.9.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	125
4.9.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	125
4.9.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	126
4.9.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	127
4.9.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	128
4.9.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	128
4.9.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	129
4.9.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	130

4.10 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight	131
4.10.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	131
4.10.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	131
4.10.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification	132
4.10.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree	133
4.10.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	134
4.10.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	134
4.10.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification	135
4.10.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree	136
4.10.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	137
4.10.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	137
4.10.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification	138
4.10.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree	139
4.11 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight	140
4.11.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	140
4.11.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	140
4.11.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification	141
4.11.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree	142

4.11.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	143
4.11.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	143
4.11.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	144
4.11.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	145
4.11.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	146
4.11.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	146
4.11.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	147
4.11.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	148
4.12 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier.....	149
4.12.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	149
4.12.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	149
4.12.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	150
4.12.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	151
4.12.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	152
4.12.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	152
4.12.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	153
4.12.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	154
4.12.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	155
4.12.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	155

4.12.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	156
4.12.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	157
4.13 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters.....	158
4.13.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	158
4.13.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	158
4.13.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	159
4.13.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	160
4.13.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	161
4.13.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	161
4.13.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	162
4.13.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	163
4.13.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	164
4.13.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	164
4.13.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	165
4.13.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	166
4.14 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ scaled_pos_weight.....	167
4.14.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	167
4.14.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	167

4.14.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	168
4.14.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	169
4.14.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	170
4.14.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	170
4.14.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	171
4.14.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	172
4.14.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	173
4.14.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	173
4.14.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	174
4.14.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	175
4.15 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ sample_weight.....	176
4.15.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	176
4.15.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	176
4.15.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	177
4.15.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	178
4.15.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	179
4.15.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	179
4.15.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	180

4.15.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	181
4.15.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	182
4.15.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	182
4.15.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	183
4.15.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	184
4.16 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ Balanced Bagging Classifier	185
4.16.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	185
4.16.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	185
4.16.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	186
4.16.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	187
4.16.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	188
4.16.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	188
4.16.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	189
4.16.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	190
4.16.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	191
4.16.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	191
4.16.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	192
4.16.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	193
4.17 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters	194

4.17.1	ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	194
4.17.1.1	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	194
4.17.1.2	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	195
4.17.1.3	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	196
4.17.2	ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	197
4.17.2.1	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	197
4.17.2.2	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	198
4.17.2.3	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	199
4.17.3	ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	200
4.17.3.1	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	200
4.17.3.2	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	201
4.17.3.3	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	202
4.18	ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight.....	203
4.18.1	ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	203
4.18.1.1	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	203
4.18.1.2	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	204
4.18.1.3	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	205
4.18.2	ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	206
4.18.2.1	ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	206

4.18.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	207
4.18.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	208
4.18.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	209
4.18.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	209
4.18.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	210
4.18.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	211
4.19 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ sample_weight.....	212
4.19.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	212
4.19.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	212
4.19.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	213
4.19.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	214
4.19.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	215
4.19.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	215
4.19.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	216
4.19.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	217
4.19.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	218
4.19.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	218
4.19.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	219

4.19.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	220
4.20 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier.....	221
4.20.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25	221
4.20.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	221
4.20.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	222
4.20.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	223
4.20.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30	224
4.20.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	224
4.20.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	225
4.20.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	226
4.20.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33	227
4.20.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold	227
4.20.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification.....	228
4.20.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree.....	229
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	247
5.1 สรุปผลการวิจัย.....	247
5.2 อภิปรายผลการวิจัย	249
5.3 ข้อเสนอแนะ	250
บรรณานุกรม	251
ประวัติผู้เขียน.....	255



สารบัญตาราง

	หน้า
ตาราง 1 แสดงคำอธิบายข้อมูลของแต่ละตัวแปร.....	28
ตาราง 2 แสดงจำนวนข้อมูลก่อนและหลังทำ Oversampling.....	41
ตาราง 3 แสดงจำนวนข้อมูลก่อนและหลังทำ Oversampling กับ sampling_strategy เท่ากับ 0.5	41
ตาราง 4 แสดงจำนวนข้อมูลก่อนและหลังทำ Undersampling	42
ตาราง 5 แสดงจำนวนข้อมูลก่อนและหลังทำ SMOTE	43



สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1 ระบบหมุนเวียนโลหิต	1
ภาพประกอบ 2 Deep-Vein Thrombosis (DVT)	2
ภาพประกอบ 3 Pulmonary Embolism (PE)	3
ภาพประกอบ 4 แสดงภาวะหลอดเลือดดำชั้นลึกอุดตันที่ขา (Deep Vein Thrombosis: DVT)	7
ภาพประกอบ 5 แสดงภาวะลิ่มเลือดอุดตันที่ปอด (Pulmonary Embolism: PE)	8
ภาพประกอบ 6 แสดงปัจจัยเสี่ยงที่มีนัยสำคัญ	15
ภาพประกอบ 7 แสดงรายงานลักษณะของผู้ป่วยก่อนหน้า	16
ภาพประกอบ 8 แสดงอาการลิ่มเลือดอุดตันในหลอดเลือดดำหลังเข้ารับการรักษาในโรงพยาบาล	16
ภาพประกอบ 9 แสดงปัจจัยที่มีนัยสำคัญที่เกี่ยวข้องกับความเสี่ยงของลิ่มเลือดอุดตันในหลอดเลือดดำ	17
ภาพประกอบ 10 แสดงจำนวนผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตัน VTE positive(1) และไม่เป็นโรคลิ่มเลือดอุดตัน VTE negative(0)	18
ภาพประกอบ 11 Label Encoder	19
ภาพประกอบ 12 Decision Tree	20
ภาพประกอบ 13 DVT SHAP Plot.....	21
ภาพประกอบ 14 ลักษณะพื้นฐานและผลลัพธ์ของผู้ป่วยที่ลงทะเบียนในการทดลอง APEX.....	22
ภาพประกอบ 15 แสดงการเปรียบเทียบ ROC curves ของ Super learner model (ML) และ Reduced model (rML) เทียบกับ IMPROVE score	23
ภาพประกอบ 16 แสดงการกระจายของความเสี่ยงที่ทำนายด้วย rML.....	23
ภาพประกอบ 17 แสดงอัตราการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำในแต่ละปัจจัย.....	24
ภาพประกอบ 18 แสดงแผนภาพกระบวนการทำงานของแบบจำลอง.....	27

ภาพประกอบ 19 แสดงจำนวนผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตัน (VTE Positive) และไม่เป็นโรคลิ่มเลือดอุดตัน (VTE Negative)..... 35

ภาพประกอบ 20 แสดงผลลัพธ์การแปลงข้อมูลให้เป็นตัวเลข ด้วย One Hot Encoding..... 36

ภาพประกอบ 21 ตัวอย่างการแสดงความสัมพันธ์ของปัจจัยเสี่ยงที่ทำให้เกิดภาวะลิ่มเลือดอุดตันในหลอดเลือดดำของการใช้ Feature Selection แบบ Threshold & Correlation 39

ภาพประกอบ 22 แสดงข้อมูลของทั้งสองคลาสที่ไม่สมดุลกัน (Imbalance) 40

ภาพประกอบ 23 แสดงลักษณะการทำงานของ Bagging 44

ภาพประกอบ 24 แสดงลักษณะการทำงานของ Boosting..... 45

ภาพประกอบ 25 Confusion Matrix..... 46

ภาพประกอบ 26 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 50

ภาพประกอบ 27 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 50

ภาพประกอบ 28 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 51

ภาพประกอบ 29 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 51

ภาพประกอบ 30 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree 52

ภาพประกอบ 31 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 52

ภาพประกอบ 32 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold..... 53

ภาพประกอบ 33 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 53

ภาพประกอบ 34 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	54
ภาพประกอบ 35 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	54
ภาพประกอบ 36 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree	55
ภาพประกอบ 37 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	55
ภาพประกอบ 38 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	56
ภาพประกอบ 39 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	56
ภาพประกอบ 40 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	57
ภาพประกอบ 41 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	57
ภาพประกอบ 42 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree	58
ภาพประกอบ 43 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	58
ภาพประกอบ 44 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold	59
ภาพประกอบ 45 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	59

ภาพประกอบ 46 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	60
ภาพประกอบ 47 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	60
ภาพประกอบ 48 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	61
ภาพประกอบ 49 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	61
ภาพประกอบ 50 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold	62
ภาพประกอบ 51 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	62
ภาพประกอบ 52 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	63
ภาพประกอบ 53 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	63
ภาพประกอบ 54 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	64
ภาพประกอบ 55 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	64

ภาพประกอบ 56 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold	65
ภาพประกอบ 57 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	65
ภาพประกอบ 58 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	66
ภาพประกอบ 59 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	66
ภาพประกอบ 60 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	67
ภาพประกอบ 61 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	67
ภาพประกอบ 62 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	68
ภาพประกอบ 63 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	68
ภาพประกอบ 64 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification ...	69
ภาพประกอบ 65 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	69
ภาพประกอบ 66 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree	70

ภาพประกอบ 67 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	70
ภาพประกอบ 68 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	71
ภาพประกอบ 69 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	71
ภาพประกอบ 70 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification ..	72
ภาพประกอบ 71 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	72
ภาพประกอบ 72 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree	73
ภาพประกอบ 73 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	73
ภาพประกอบ 74 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	74
ภาพประกอบ 75 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	74
ภาพประกอบ 76 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification ..	75
ภาพประกอบ 77 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	75
ภาพประกอบ 78 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree	76
ภาพประกอบ 79 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	76

ภาพประกอบ 80 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Threshold..... 77

ภาพประกอบ 81 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 77

ภาพประกอบ 82 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Mutual Information Classification 78

ภาพประกอบ 83 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 78

ภาพประกอบ 84 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 79

ภาพประกอบ 85 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 79

ภาพประกอบ 86 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี
Threshold..... 80

ภาพประกอบ 87 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 80

ภาพประกอบ 88 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี
Mutual Information Classification 81

ภาพประกอบ 89 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 81

ภาพประกอบ 90 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 82

ภาพประกอบ 91 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 82

ภาพประกอบ 92 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี
Threshold..... 83

ภาพประกอบ 93 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 83

ภาพประกอบ 94 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี
Mutual Information Classification 84

ภาพประกอบ 95 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 84

ภาพประกอบ 96 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 85

ภาพประกอบ 97 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 85

ภาพประกอบ 98 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง
ข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 86

ภาพประกอบ 99 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 86

ภาพประกอบ 100 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง
ข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification .. 87

ภาพประกอบ 101 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	87
ภาพประกอบ 102 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree	88
ภาพประกอบ 103 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	88
ภาพประกอบ 104 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	89
ภาพประกอบ 105 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	89
ภาพประกอบ 106 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification ..	90
ภาพประกอบ 107 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	90
ภาพประกอบ 108 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree	91
ภาพประกอบ 109 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	91
ภาพประกอบ 110 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	92
ภาพประกอบ 111 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	92
ภาพประกอบ 112 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification ..	93
ภาพประกอบ 113 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	93

ภาพประกอบ 114 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree	94
ภาพประกอบ 115 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	94
ภาพประกอบ 116 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	95
ภาพประกอบ 117 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	95
ภาพประกอบ 118 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	96
ภาพประกอบ 119 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	96
ภาพประกอบ 120 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	97
ภาพประกอบ 121 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	97
ภาพประกอบ 122 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	98
ภาพประกอบ 123 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	98
ภาพประกอบ 124 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	99

ภาพประกอบ 125 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	99
ภาพประกอบ 126 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	100
ภาพประกอบ 127 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	100
ภาพประกอบ 128 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	101
ภาพประกอบ 129 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	101
ภาพประกอบ 130 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	102
ภาพประกอบ 131 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	102
ภาพประกอบ 132 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	103
ภาพประกอบ 133 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	103
ภาพประกอบ 134 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold ..	104
ภาพประกอบ 135 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	104

ภาพประกอบ 136 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	105
ภาพประกอบ 137 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	105
ภาพประกอบ 138 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree	106
ภาพประกอบ 139 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	106
ภาพประกอบ 140 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold ..	107
ภาพประกอบ 141 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	107
ภาพประกอบ 142 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	108
ภาพประกอบ 143 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	108
ภาพประกอบ 144 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree	109
ภาพประกอบ 145 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	109
ภาพประกอบ 146 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold ..	110

ภาพประกอบ 147 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 110

ภาพประกอบ 148 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio
Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual
Information Classification 111

ภาพประกอบ 149 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 111

ภาพประกอบ 150 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio
Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree
..... 112

ภาพประกอบ 151 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 112

ภาพประกอบ 152 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Threshold..... 113

ภาพประกอบ 153 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 113

ภาพประกอบ 154 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Mutual Information Classification 114

ภาพประกอบ 155 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 114

ภาพประกอบ 156 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 115

ภาพประกอบ 157 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 115

ภาพประกอบ 158 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	116
ภาพประกอบ 159 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	116
ภาพประกอบ 160 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	117
ภาพประกอบ 161 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	117
ภาพประกอบ 162 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	118
ภาพประกอบ 163 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	118
ภาพประกอบ 164 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	119
ภาพประกอบ 165 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	119
ภาพประกอบ 166 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	120
ภาพประกอบ 167 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	120

ภาพประกอบ 168 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 121

ภาพประกอบ 169 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 121

ภาพประกอบ 170 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold 122

ภาพประกอบ 171 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 122

ภาพประกอบ 172 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification
..... 123

ภาพประกอบ 173 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 123

ภาพประกอบ 174 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree 124

ภาพประกอบ 175 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 124

ภาพประกอบ 176 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold 125

ภาพประกอบ 177 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 125

ภาพประกอบ 178 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification
..... 126

ภาพประกอบ 179 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 126

ภาพประกอบ 180 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree 127

ภาพประกอบ 181 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 127

ภาพประกอบ 182 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold 128

ภาพประกอบ 183 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 128

ภาพประกอบ 184 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification
..... 129

ภาพประกอบ 185 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 129

ภาพประกอบ 186 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดย
แบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree 130

ภาพประกอบ 187 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 130

ภาพประกอบ 188 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ
Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี
Threshold..... 131

ภาพประกอบ 189 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 131

ภาพประกอบ 190 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	132
ภาพประกอบ 191 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	132
ภาพประกอบ 192 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	133
ภาพประกอบ 193 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	133
ภาพประกอบ 194 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	134
ภาพประกอบ 195 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	134
ภาพประกอบ 196 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	135
ภาพประกอบ 197 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	135
ภาพประกอบ 198 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	136
ภาพประกอบ 199 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	136

ภาพประกอบ 200 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold..... 137

ภาพประกอบ 201 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 137

ภาพประกอบ 202 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 138

ภาพประกอบ 203 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 138

ภาพประกอบ 204 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 139

ภาพประกอบ 205 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 139

ภาพประกอบ 206 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 140

ภาพประกอบ 207 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 140

ภาพประกอบ 208 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 141

ภาพประกอบ 209 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 141

ภาพประกอบ 210 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	142
ภาพประกอบ 211 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	142
ภาพประกอบ 212 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	143
ภาพประกอบ 213 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	143
ภาพประกอบ 214 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	144
ภาพประกอบ 215 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	144
ภาพประกอบ 216 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	145
ภาพประกอบ 217 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	145
ภาพประกอบ 218 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	146
ภาพประกอบ 219 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	146

ภาพประกอบ 220 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	147
ภาพประกอบ 221 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	147
ภาพประกอบ 222 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	148
ภาพประกอบ 223 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	148
ภาพประกอบ 224 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	149
ภาพประกอบ 225 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	149
ภาพประกอบ 226 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	150
ภาพประกอบ 227 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	150
ภาพประกอบ 228 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	151
ภาพประกอบ 229 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	151

ภาพประกอบ 230 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	152
ภาพประกอบ 231 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	152
ภาพประกอบ 232 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	153
ภาพประกอบ 233 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	153
ภาพประกอบ 234 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	154
ภาพประกอบ 235 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	154
ภาพประกอบ 236 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	155
ภาพประกอบ 237 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	155
ภาพประกอบ 238 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	156
ภาพประกอบ 239 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	156

ภาพประกอบ 240 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 157

ภาพประกอบ 241 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 157

ภาพประกอบ 242 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 158

ภาพประกอบ 243 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 158

ภาพประกอบ 244 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 159

ภาพประกอบ 245 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 159

ภาพประกอบ 246 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree 160

ภาพประกอบ 247 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 160

ภาพประกอบ 248 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold..... 161

ภาพประกอบ 249 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 161

ภาพประกอบ 250 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 162

ภาพประกอบ 251 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 162

ภาพประกอบ 252 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree 163

ภาพประกอบ 253 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 163

ภาพประกอบ 254 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold..... 164

ภาพประกอบ 255 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 164

ภาพประกอบ 256 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 165

ภาพประกอบ 257 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 165

ภาพประกอบ 258 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree 166

ภาพประกอบ 259 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 166

ภาพประกอบ 260 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 167

ภาพประกอบ 261 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 167

ภาพประกอบ 262 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 168

ภาพประกอบ 263 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 168

ภาพประกอบ 264 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 169

ภาพประกอบ 265 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 169

ภาพประกอบ 266 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold..... 170

ภาพประกอบ 267 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 170

ภาพประกอบ 268 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 171

ภาพประกอบ 269 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 171

ภาพประกอบ 270 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 172

ภาพประกอบ 271 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 172

ภาพประกอบ 272 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold..... 173

ภาพประกอบ 273 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 173

ภาพประกอบ 274 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 174

ภาพประกอบ 275 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 174

ภาพประกอบ 276 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 175

ภาพประกอบ 277 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 175

ภาพประกอบ 278 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold..... 176

ภาพประกอบ 279 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 176

ภาพประกอบ 280 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 177

ภาพประกอบ 281 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 177

ภาพประกอบ 282 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 178

ภาพประกอบ 283 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 178

ภาพประกอบ 284 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold..... 179

ภาพประกอบ 285 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	179
ภาพประกอบ 286 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	180
ภาพประกอบ 287 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	180
ภาพประกอบ 288 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	181
ภาพประกอบ 289 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	181
ภาพประกอบ 290 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold	182
ภาพประกอบ 291 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	182
ภาพประกอบ 292 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	183
ภาพประกอบ 293 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	183
ภาพประกอบ 294 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	184
ภาพประกอบ 295 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	184

ภาพประกอบ 296 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold	185
ภาพประกอบ 297 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	185
ภาพประกอบ 298 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	186
ภาพประกอบ 299 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	186
ภาพประกอบ 300 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	187
ภาพประกอบ 301 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	187
ภาพประกอบ 302 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold	188
ภาพประกอบ 303 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	188
ภาพประกอบ 304 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	189
ภาพประกอบ 305 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	189

ภาพประกอบ 306 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	190
ภาพประกอบ 307 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	190
ภาพประกอบ 308 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold	191
ภาพประกอบ 309 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	191
ภาพประกอบ 310 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	192
ภาพประกอบ 311 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	192
ภาพประกอบ 312 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	193
ภาพประกอบ 313 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	193
ภาพประกอบ 314 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	194
ภาพประกอบ 315 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	194
ภาพประกอบ 316 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	195

ภาพประกอบ 317 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	195
ภาพประกอบ 318 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree	196
ภาพประกอบ 319 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	196
ภาพประกอบ 320 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	197
ภาพประกอบ 321 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	197
ภาพประกอบ 322 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	198
ภาพประกอบ 323 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	198
ภาพประกอบ 324 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree	199
ภาพประกอบ 325 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	199
ภาพประกอบ 326 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	200
ภาพประกอบ 327 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	200
ภาพประกอบ 328 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่ง ข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	201
ภาพประกอบ 329 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	201

ภาพประกอบ 330 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree	202
ภาพประกอบ 331 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	202
ภาพประกอบ 332 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	203
ภาพประกอบ 333 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	203
ภาพประกอบ 334 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	204
ภาพประกอบ 335 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	204
ภาพประกอบ 336 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	205
ภาพประกอบ 337 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	205
ภาพประกอบ 338 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	206
ภาพประกอบ 339 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด.....	206
ภาพประกอบ 340 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	207

ภาพประกอบ 341 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	207
ภาพประกอบ 342 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	208
ภาพประกอบ 343 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	208
ภาพประกอบ 344 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	209
ภาพประกอบ 345 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	209
ภาพประกอบ 346 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	210
ภาพประกอบ 347 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	210
ภาพประกอบ 348 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	211
ภาพประกอบ 349 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	211
ภาพประกอบ 350 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	212
ภาพประกอบ 351 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	212

ภาพประกอบ 352 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 213

ภาพประกอบ 353 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 213

ภาพประกอบ 354 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 214

ภาพประกอบ 355 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 214

ภาพประกอบ 356 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold..... 215

ภาพประกอบ 357 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 215

ภาพประกอบ 358 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 216

ภาพประกอบ 359 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 216

ภาพประกอบ 360 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree..... 217

ภาพประกอบ 361 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด..... 217

ภาพประกอบ 362 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold..... 218

ภาพประกอบ 363 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	218
ภาพประกอบ 364 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	219
ภาพประกอบ 365 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	219
ภาพประกอบ 366 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	220
ภาพประกอบ 367 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	220
ภาพประกอบ 368 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold.....	221
ภาพประกอบ 369 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	221
ภาพประกอบ 370 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	222
ภาพประกอบ 371 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	222
ภาพประกอบ 372 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	223
ภาพประกอบ 373 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	223

ภาพประกอบ 374 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold.....	224
ภาพประกอบ 375 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	224
ภาพประกอบ 376 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	225
ภาพประกอบ 377 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	225
ภาพประกอบ 378 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree.....	226
ภาพประกอบ 379 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	226
ภาพประกอบ 380 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold.....	227
ภาพประกอบ 381 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	227
ภาพประกอบ 382 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification	228
ภาพประกอบ 383 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล ของข้อมูลทั้งหมด.....	228

ภาพประกอบ 384 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ
Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี
Decision Tree..... 229

ภาพประกอบ 385 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุล
ของข้อมูลทั้งหมด..... 229

ภาพประกอบ 386 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Threshold และ Oversampling..... 230

ภาพประกอบ 387 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5..... 231

ภาพประกอบ 388 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Mutual Information Classification และ Oversampling 231

ภาพประกอบ 389 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่
มีค่าเท่ากับ 0.5 232

ภาพประกอบ 390 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Decision Tree และ Oversampling 232

ภาพประกอบ 391 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี
Decision Tree และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5... 233

ภาพประกอบ 392 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Threshold และ Oversampling 233

ภาพประกอบ 393 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling 234

ภาพประกอบ 394 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy
ที่มีค่าเท่ากับ 0.5..... 234

ภาพประกอบ 395 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์
sampling_strategy ที่มีค่าเท่ากับ 0.5..... 235

ภาพประกอบ 396 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Mutual Information Classification และ Oversampling 235

ภาพประกอบ 397 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling 236

ภาพประกอบ 398 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับ
พารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 236

ภาพประกอบ 399 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับ
พารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 237

ภาพประกอบ 400 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Decision Tree และ Oversampling 237

ภาพประกอบ 401 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling 238

ภาพประกอบ 402 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้
Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์
sampling_strategy ที่มีค่าเท่ากับ 0.5..... 238

ภาพประกอบ 403 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters
ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์
sampling_strategy ที่มีค่าเท่ากับ 0.5..... 239

ภาพประกอบ 404 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default
Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling 239

ภาพประกอบ 405 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning
Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling 240

ภาพประกอบ 406 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5..... 240

ภาพประกอบ 407 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5..... 241

ภาพประกอบ 408 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling 241

ภาพประกอบ 409 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling 242

ภาพประกอบ 410 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 242

ภาพประกอบ 411 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 243

ภาพประกอบ 412 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling 243

ภาพประกอบ 413 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling..... 244

ภาพประกอบ 414 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5..... 244

ภาพประกอบ 415 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับ พารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.5 245

ภาพประกอบ 416 แผนภูมิแท่งของแบบจำลอง 12 ตัวที่ให้ผลลัพธ์ที่ดีที่สุด 246

ภาพประกอบ 417 ผลลัพธ์ Confusion Matrix ที่ได้จาก Random Forest กับ Default Parameters กับพารามิเตอร์ `Ratio Class Weight` โดยใช้อัตราส่วนของข้อมูลเท่ากับ 75:25 และ ใช้ Feature Selection ด้วยวิธี Mutual Information Classification 248

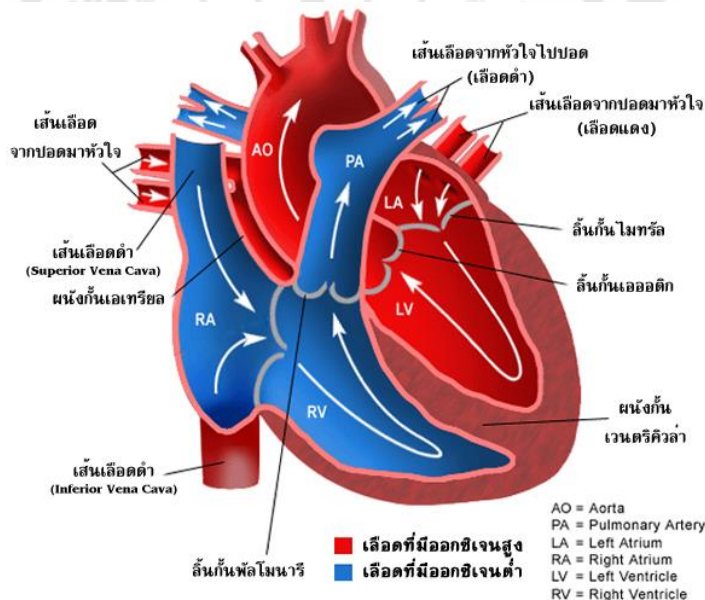


บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของงานวิจัย

ภายในร่างกายมนุษย์ จะมีการทำงานของระบบภายในร่างกายต่างๆ ที่สัมพันธ์กัน เพื่อให้มนุษย์สามารถดำรงชีวิตได้อย่างปกติ ซึ่งระบบภายในร่างกาย สามารถแบ่งออกได้เป็น 10 ระบบ เช่น ระบบประสาท, ระบบย่อยอาหาร และระบบหมุนเวียนโลหิต เป็นต้น (คำอินทร์) โดยระบบหมุนเวียนโลหิต (Circulatory System) จะมีหน้าที่ลำเลียงสารอาหาร เช่น กรดอะมิโน อิเล็กโทรไลต์ ออกซิเจน คาร์บอนไดออกไซด์ ฮอรโมน และเม็ดเลือดเข้าและออกเซลล์ในร่างกาย เพื่อหล่อเลี้ยงและต่อสู้โรค รักษาอุณหภูมิและปรับ pH ของร่างกาย ผ่านทางหลอดเลือด ซึ่งในหลอดเลือดของมนุษย์ จะประกอบด้วยเม็ดเลือดแดง เม็ดเลือดขาว เกล็ดเลือด และน้ำเหลืองไหลเวียนอยู่ โดยมีผนังหลอดเลือดเป็นเสมือนท่อลำเลียง แต่เมื่อไรก็ตามที่เกิดบาดแผล เลือดจะแข็งตัว ร่างกายจะช่วยห้ามการแข็งตัวของเลือดโดยใช้สารโปรตีนตัวหนึ่ง ดังนั้น หากโปรตีนในร่างกายทำงานผิดปกติ ทำให้การห้ามการแข็งตัวของเลือดทำงานน้อยลง เลือดจะไหลเวียนผิดปกติตามไปด้วย ทำให้เกิดลิ่มเลือดอุดตันขึ้นมาได้ ซึ่งหากปล่อยปะละเลย ก็อาจเป็นอันตรายถึงแก่ชีวิต (โรงพยาบาลพญาไท, 2021)



ภาพประกอบ 1 ระบบหมุนเวียนโลหิต

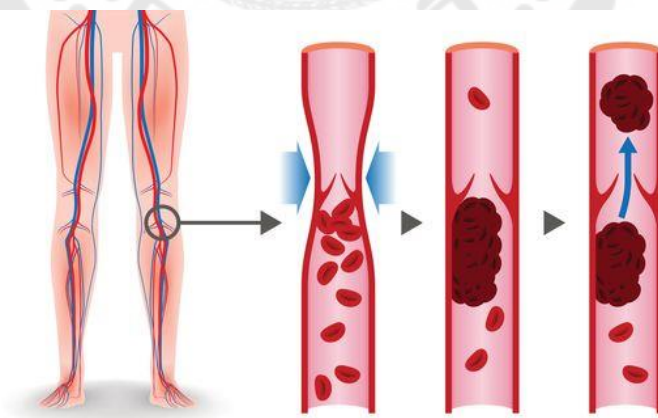
ที่มา: (คำอินทร์)

โรคลิ่มเลือดอุดตัน (Thrombosis) เกิดจากลิ่มเลือดอุดตันหรือขัดขวางการไหลเวียนของเลือดในหลอดเลือดแดง (Arterial Thrombosis) หรือหลอดเลือดดำ (Venous Thrombosis) หากลิ่มเลือดส่วนหนึ่งส่วนใดหลุดออกจากผนังหลอดเลือดแล้วลอยไปตามกระแสเลือดจะเรียกว่า เอ็มโบไล (Emboli) ซึ่งเป็นปัจจัยเสี่ยงที่ทำให้เกิดภาวะลิ่มเลือดอุดตันของหลอดเลือดดำ ซึ่งเป็นโรคในกลุ่มของโรคหัวใจและหลอดเลือดที่เป็นสาเหตุสำคัญทำให้ผู้ป่วยเสียชีวิต (วงการแพทย์)

สาเหตุที่ทำให้เกิดลิ่มเลือดอุดตันผนังหลอดเลือดผิดปกติ เช่น

- ร่างกายมีบาดแผลจากของมีคมบาด, การผ่าตัด หรือ การทำหัตถการ ซึ่งส่งผลทำให้ผนังหลอดเลือดเสียหาย เกิดการแข็งตัวของเลือดผิดปกติ
- เกิดจากสารโปรตีนการห้ามการแข็งตัวของเลือดในร่างกายบกพร่อง เช่น ร่างกายขาดสารโปรตีนที่ทำหน้าที่ยับยั้งการแข็งตัวของเลือด ก็มีโอกาสที่ทำให้เกิดลิ่มเลือดอุดตันได้
- การส่งต่อผ่านทางพันธุกรรม โดยการขาดสารโปรตีนที่เป็นสารห้ามการแข็งตัวของเลือด ก็มีโอกาสขาดสารโปรตีนชนิดเดียวกัน ซึ่งมีความเสี่ยงที่จะเกิดลิ่มเลือดอุดตันได้มากขึ้น
- มีประวัติการใช้ยาเสพติด ยาฮอร์โมนเอสโตรเจน หรือกลุ่มผู้ป่วยโรคมะเร็ง ก็ถือว่าอยู่ในกลุ่มเสี่ยงที่อาจทำให้เกิดภาวะลิ่มเลือดอุดตันได้ง่ายกว่าคนปกติทั่วไป

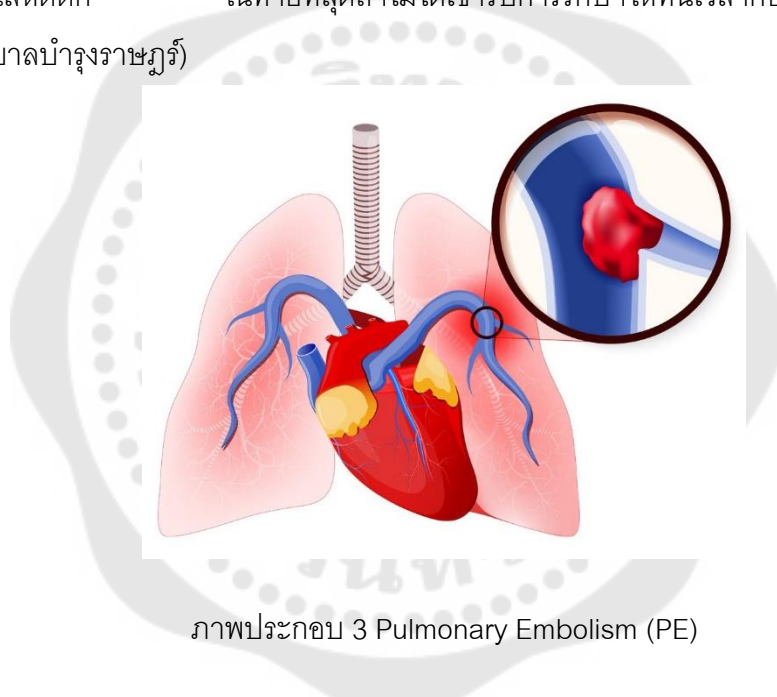
(โรงพยาบาลพญาไท, 2021)



ภาพประกอบ 2 Deep-Vein Thrombosis (DVT)

ที่มา: (Days, 2020)

ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ (Venous Thromboembolism: VTE) นั้นจะเกิดขึ้นที่ขาบ่อยที่สุด การเกิดลิ่มเลือดอุดตันที่หลอดเลือดดำที่ขานั้น จะเรียกว่า Deep Vein Thrombosis (DVT) โดยจะพบลิ่มเลือดอุดตันในหลอดเลือดดำมากกว่าหลอดเลือดแดง เนื่องจากหัวใจจะบีบตัวส่งเลือดแรงผ่านหลอดเลือดแดงตลอดเวลา ไม่หยุดนิ่ง ตรงข้ามกับหลอดเลือดดำที่มักมีการหยุดนิ่งมากกว่าจึงทำให้มีโอกาสเกิดลิ่มเลือดอุดตันได้ง่ายกว่า โดยอาการของลิ่มเลือดอุดตันที่หลอดเลือดดำที่ขานั้น จะมีอาการขาบวม โต รู้สึกกดแล้วเจ็บ หรือเดินแล้วจะรู้สึกปวดบวมมากขึ้น อาการจะเกิดขึ้นกะทันหัน แล้วหลังจากนั้น 2-3 วัน ก็จะพบอาการปวดบวม ซึ่งเกิดจากลิ่มเลือดอุดตันที่ปอด (Pulmonary Embolism: PE) จะมีอาการเหนื่อยง่าย แน่นหน้าอกกะทันหัน ความดันโลหิตตก ในท้ายที่สุดถ้าไม่ได้เข้ารับการรักษาให้ทันเวลา ก็อาจทำให้เสียชีวิตได้ (โรงพยาบาลบำรุงราษฎร์)



ภาพประกอบ 3 Pulmonary Embolism (PE)

ที่มา: (Health)

เราสามารถลดความเสี่ยงในการเกิดโรคลิ่มเลือดอุดตันได้ด้วยการดูแลตัวเองให้มีสุขภาพดี หลีกเลี่ยงการใช้ยาเสพติด ยาคุมกำเนิด รวมถึงยาแก้ปวดไมเกรนบางตัว ซึ่งมีผลทำให้เกิดลิ่มเลือดอุดตันได้ เพราะมีฤทธิ์ทำให้หลอดเลือดหดตัวแบบเฉียบพลัน ทำให้การไหลเวียนของเลือดผิดปกติ จนเกิดตะกอนกลายเป็นลิ่มเลือดในที่สุด นอกจากนี้ การควบคุมน้ำหนักให้ได้อยู่ในเกณฑ์มาตรฐาน ก็ถือเป็นการลดความเสี่ยงที่จะเกิดภาวะลิ่มเลือดอุดตันได้ด้วยเช่นกัน เนื่องจากน้ำหนักตัวที่มากกว่าค่ามาตรฐานของดัชนีมวลกาย (Body Mass Index: BMI) ส่งผลให้เกิดลิ่มเลือดอุดตันได้ง่ายขึ้น (โรงพยาบาลพญาไท, 2021)

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อสร้างแบบจำลองที่เหมาะสมกับการทำนายความเสี่ยงในการเกิดโรคลิ่มเลือดในหลอดเลือดดำไปใช้ประกอบในการตัดสินใจเบื้องต้นของแพทย์ได้

1.3 ขอบเขตของงานวิจัย

1.3.1 ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลของผู้ป่วยที่เข้ารับการรักษาในหอผู้ป่วย แผนกอายุรศาสตร์ของโรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย ปี พ.ศ. 2552

1.3.2 กระบวนการและผลของงานวิจัยนี้ จัดทำขึ้นโดยนักวิจัยที่ไม่ใช่บุคลากรทางการแพทย์ จึงไม่สามารถแปลคำศัพท์ทางการแพทย์ที่เป็นภาษาอังกฤษได้อย่างถูกต้อง จัดหมวดหมู่ของกลุ่มโรคและอาการของผู้ป่วยได้ แต่ได้รับคำแนะนำและความรู้จากบุคลากรทางการแพทย์เพื่อนำมาใช้ประกอบการทำการวิจัยเท่านั้น

1.4 ขั้นตอนการทำงานวิจัย

1.4.1 ศึกษางานวิจัยเกี่ยวกับ ภาวะการเกิดโรคลิ่มเลือดอุดตันในหลอดเลือดดำจากบทความวิจัยเรื่อง Risk factors for symptomatic venous thromboembolism in Thai hospitalized medical patients (Rojnuckarin et al., 2011) โดยข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้มาจากผู้ป่วยที่เข้ารับการรักษาในหอผู้ป่วยแผนกอายุรศาสตร์ของโรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย ปี พ.ศ. 2552

1.4.2 ศึกษางานวิจัยเรื่อง แบบจำลองการวินิจฉัยอัตโนมัติสำหรับความเสี่ยงต่อการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำตามอาการโดยอาศัยการเรียนรู้ของเครื่อง หรือ Automatic diagnosis model for risk of symptomatic venous thromboembolism based on machine learning (Sukperm, Rojnuckarin, Akkawat, & Sa-Ing, 2021) เพื่อทำงานวิจัยต่อยอดโดยการปรับปรุงการใช้เทคนิคการเรียนรู้ของเครื่อง และเพิ่มการใช้เทคนิคการเรียนรู้แบบรวมของเครื่องเพื่อประเมินประสิทธิภาพของแบบจำลองเชิงทำนายสำหรับการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ และสามารถนำผลลัพธ์ที่ได้ไปเป็นข้อมูลประกอบการตัดสินใจในการรักษาของแพทย์ได้

1.4.3 การนำเข้าข้อมูล การเตรียมข้อมูล การแปลงข้อมูล และการจัดการกับข้อมูลที่มีปัญหาเพื่อให้ข้อมูลนั้นสามารถนำมาใช้ประกอบการวิเคราะห์ได้

1.4.4 การวิเคราะห์ข้อมูลเชิงสำรวจ จากปัจจัยที่ทำให้เกิดโรคลิ่มเลือดอุดตันในหลอดเลือดดำที่ได้รับมาจากข้อมูลทางการแพทย์ ซึ่งประกอบด้วย 1,290 แถว และ 65 คอลัมน์ โดย

ข้อมูลทางการแพทย์นี้ได้รับการอนุมัติจากคณะกรรมการจริยธรรมของคณะแพทยศาสตร์
จุฬาลงกรณ์มหาวิทยาลัยแล้ว

1.4.5 สร้างแบบจำลองเชิงทำนายจากการเรียนรู้แบบรวมของเครื่อง และปรับค่าพารามิเตอร์ให้เหมาะสม เพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพมากยิ่งขึ้น ซึ่งวัดผลโดยใช้หลักการทางคณิตศาสตร์

1.4.6 เปรียบเทียบแบบจำลองเชิงทำนายการเรียนรู้ของเครื่อง กับแบบจำลองเชิงทำนายจากการเรียนรู้แบบรวมของเครื่อง และสรุปผลการทดลองเพื่อนำเสนอแบบจำลองที่มีประสิทธิภาพในการทำนายความเสี่ยงของการเกิดโรคล้มเลือดอุดตันในหลอดเลือดดำ

1.4.7 วิเคราะห์และอภิปรายผลการทดลอง แบบจำลองที่เลือกใช้ในการทำนายความเสี่ยงของการเกิดโรคล้มเลือดอุดตันในหลอดเลือดดำ เพื่อนำมาปรับปรุงหรือเสนอแนะข้อจำกัดของแบบจำลอง

1.5 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1.5.1 สามารถสร้างแบบจำลองเชิงทำนายจากการเรียนรู้ของเครื่อง เพื่อให้ผลวินิจฉัยเบื้องต้นให้กับแพทย์ได้อย่างรวดเร็ว และช่วยลดต้นทุนการวินิจฉัยของแพทย์

1.5.2 สามารถจำแนกความเสี่ยงกับผู้ป่วยโดยไม่ต้องทำการทดสอบภายในร่างกาย (Invasive Test)

1.5.3 เพื่อให้ผลลัพธ์ที่ได้จากแบบจำลองเชิงทำนายจากการเรียนรู้ของเครื่อง มีความแม่นยำมากขึ้น จากแบบจำลองก่อนหน้า

1.5.4 เพื่อให้แบบจำลองเชิงทำนายจากการเรียนรู้ของเครื่องเป็นที่ยอมรับจากแพทย์มากยิ่งขึ้น

บทที่ 2

วรรณกรรม และงานวิจัยที่เกี่ยวข้อง

สำหรับการศึกษางานวิจัยในครั้งนี้ ผู้วิจัยได้ทำการศึกษาเอกสารและผลงานวิจัยที่เกี่ยวข้องกับการทำนายภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ และการพัฒนาแบบจำลองการเรียนรู้ของเครื่อง โดยได้นำเสนอต่อหัวข้อต่อไปนี้

2.1 ทฤษฎีเกี่ยวกับภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

2.2 ข้อมูลที่สำคัญต่อการวิเคราะห์ภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

2.3 ทฤษฎีการวิเคราะห์ข้อมูลด้วยหลักการทางคอมพิวเตอร์

2.4 งานวิจัยที่ศึกษาเกี่ยวกับปัจจัยการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

2.5 งานวิจัยที่ศึกษาเกี่ยวกับการวิเคราะห์ความเสี่ยงของภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

2.1 ทฤษฎีเกี่ยวกับภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ

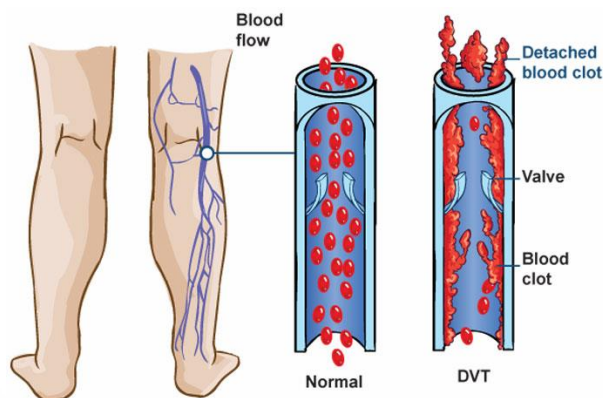
ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ คือ การขัดขวางการไหลเวียนของเลือดในหลอดเลือดดำ ทำให้เลือดไหลเวียนไม่สะดวก ซึ่งเกิดการจับตัวกันเป็นลิ่มเลือด มักจะเกิดขึ้นบริเวณขาหรือแขน ส่งผลทำให้รู้สึกปวดบวมแดง ถ้าหากลิ่มเลือดอุดตันที่ปอดก็จะทำให้รู้สึกหายใจถี่ เจ็บหน้าอก ไอปนเลือด เวียนศีรษะ หมดสติ และหัวใจเต้นผิดจังหวะ ซึ่งเป็นอาการที่ร้ายแรงอาจทำให้เสียชีวิตได้ ซึ่งภาวะลิ่มเลือดอุดตันในหลอดเลือดดำแบ่งออกเป็น 2 ประเภท

2.1.1 ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำในชั้นลึกที่ขา (Deep Vein

Thromboembolism: DVT)

DVT คือการเกิดลิ่มเลือดในหลอดเลือดดำในชั้นลึก มักเกิดขึ้นที่ขาหรือกระดูกเชิงกราน และบางส่วนประมาณ 4-10% จะเกิดขึ้นที่แขน (Palta, Saroa, & Palta, 2014) กลไกที่อยู่เบื้องหลังในการสร้าง DVT มักเกี่ยวข้องกับหลายๆ ปัจจัยเช่น อัตราการไหลเวียนของเลือดที่ลดลง การเปลี่ยนแปลงของผนังหลอดเลือด และการอักเสบ นอกจากนี้ยังมีปัจจัยเสี่ยงที่จะทำให้เกิด DVT ได้แก่ การผ่าตัด อายุที่มากขึ้น มะเร็ง โรคอ้วน การติดเชื้อ โรคอักเสบ กลุ่มอาการแอนโทฟอสโฟลิปิด ประวัติส่วนตัวและประวัติครอบครัวที่เป็น VTE การขาดการเคลื่อนไหว การคุมกำเนิดด้วยฮอร์โมน การตั้งครรภ์ ปัจจัยทางพันธุกรรมรวมถึงกรุ๊ปเลือดที่ไม่ใช่กรุ๊ปเลือด O ข้อบกพร่องของ antithrombin โปรตีน C และโปรตีน S และการกลายพันธุ์ของปัจจัย V Leiden และ

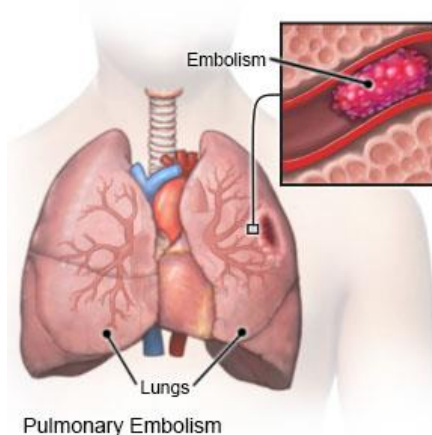
prothrombin G20210A โดยรวมแล้ว มีการระบุปัจจัยเสี่ยงทางพันธุกรรมหลายสิบปัจจัย (Crous-Bou, Harrington, & Kabrhel, 2016) ดังภาพประกอบ 4



ภาพประกอบ 4 แสดงภาวะหลอดเลือดดำชั้นลึกอุดตันที่ขา (Deep Vein Thrombosis: DVT)

2.1.2 ภาวะลิ่มเลือดอุดตันในปอด (Pulmonary Embolism: PE)

PE เป็นภาวะแทรกซ้อนที่เกิดจากการที่มีลิ่มเลือดเกิดขึ้นในหลอดเลือดดำ ซึ่งลิ่มเลือดดังกล่าวหากเกิดขึ้นแล้วมีโอกาสสูงที่จะหลุดเข้าสู่หลอดเลือดดำ Inferior หรือ Superior Vena Cava ก่อนผ่านเข้าหัวใจห้องขวา และหลุดไปอุดตันที่หลอดเลือดที่ปอด ซึ่งทำให้เลือดดำไม่สามารถแลกเปลี่ยนก๊าซออกซิเจนเกิดภาวะออกซิเจนพร่อง (Hypoxia) และหากก้อนลิ่มเลือดมีขนาดใหญ่จะทำให้มีการเพิ่มแรงเสียดทานในเลือดหลอดเลือดปอด (Pulmonary Vascular Resistance) ทำให้ความดันในหัวใจห้องขวาสูงขึ้น และมีการเคลื่อน (Shift) ของผนังกันหัวใจห้องล่างไปทางหัวใจห้องซ้ายล่าง ซึ่งผลดังกล่าวร่วมกับปริมาณเลือดที่ผ่านเนื้อปอดมาสู่หัวใจห้องซ้ายก็จะลดลง ทำให้ Cardiac Output ลดลง ผู้ป่วยจะมีความดันต่ำลง เกิดอาการช็อก และเสียชีวิตในที่สุด ดังภาพประกอบ 5



ภาพประกอบ 5 แสดงภาวะลิ่มเลือดอุดตันที่ปอด (Pulmonary Embolism: PE)

ที่มา: (Drugs.com, 2022)

2.1.3 อาการของภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน

อาการของภาวะลิ่มเลือดในหลอดเลือดดำอุดตันขึ้นอยู่กับตำแหน่งของอวัยวะที่มีลิ่มเลือดอุดตัน ผู้ป่วยที่เป็นลิ่มเลือดอุดตันแบบ DVT บางรายอาจไม่แสดงอาการผิดปกติให้เห็น บางรายอาจมีอาการปวด บวมแดง และรู้สึกร้อนที่ขา หรือบริเวณที่มีลิ่มเลือดอุดตัน ส่วนผู้ป่วยภาวะลิ่มเลือดอุดตันแบบ PE อาจจะมีอาการหายใจไม่อิ่ม ไบปนเลือด หัวใจเต้นผิดปกติ รู้สึกเจ็บหน้าอก เวียนศีรษะ หรือหมดสติ

2.1.4 การวินิจฉัยภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน

ในขั้นแรกแพทย์จะสอบถามอาการผู้ป่วยที่เป็นภาวะนี้ และตรวจร่างกายเบื้องต้น เนื่องจากผู้ป่วยอาจมีอาการบวมแดงตามผิวหนัง ถ้าหากประเมินแล้วว่าผู้ป่วยมีแนวโน้มเป็น VTE อาจทำการตรวจเลือดเพื่อวินิจฉัยเพิ่มเติม เช่น วัดปริมาณเม็ดเลือด วัดปริมาณเกล็ดเลือด สารดีดิมเมอร์ (D-Dimer) หรือสารชนิดต่างๆ ในเลือด เป็นต้น เนื่องจากเลือดของผู้ป่วยมักมีปริมาณสารดังกล่าวจะมีค่าที่ผิดปกติ อย่างไรก็ตาม การวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำแบบ PE และ DVT นั้นแตกต่างกัน

การวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำแบบ DVT สามารถทำได้ดังนี้

- การอัลตราซาวด์ เป็นการใช้คลื่นเสียงความถี่สูง เพื่อจำลองภาพอวัยวะภายใน เพื่อตรวจหาลิ่มเลือดที่อุดตันในหลอดเลือดดำ โดยเฉพาะบริเวณขา

- การตรวจวินิจฉัยหูดเลือด แพทย์จะฉีดสารทึบรังสีเข้าในหลอดเลือดดำ จากนั้นเอกซเรย์เพื่อตรวจสอบความผิดปกติของหลอดเลือดดำบริเวณขา เท้า หรือข้อเท้า

การวินิจฉัยภาวะลิ่มเลือดอุดตันในหลอดเลือดดำแบบ PE สามารถทำได้ดังนี้

- การเอกซเรย์ปอด คือการใช้รังสี เพื่อตรวจสอบดูความผิดปกติบริเวณปอด
- การตรวจคลื่นไฟฟ้าหัวใจ เพื่อวัดอัตราการการทำงานของกระแสไฟฟ้าในหัวใจ
- การฉีดสีในหลอดเลือดปอด โดยการผ่าตัดเปิดแผลเล็กๆ และใส่เครื่องมือพิเศษ เข้าไปในทางหลอดเลือดดำ จากนั้นจึงฉีดสีเพื่อให้เห็นลักษณะของหลอดเลือดดำ ในปอดได้อย่างชัดเจน
- การเอกซเรย์คอมพิวเตอร์ (Computerized Tomography Scan: CT Scan) เพื่อสร้างภาพตัดขวางของปอด ช่วยให้เห็นลักษณะของหลอดเลือดในปอดได้อย่างละเอียด
- การสร้างภาพด้วยสนามแม่เหล็กไฟฟ้า (Magnetic Resonance Imaging: MRI) เพื่อสร้างภาพของหลอดเลือดดำ และอวัยวะภายในช่องอก ทำให้วินิจฉัยได้อย่างแม่นยำยิ่งขึ้น

ที่มา: (พบแพทย์, 2016)

นอกจากนี้ มีงานวิจัยที่เกี่ยวข้องกับการวินิจฉัยปัจจัยความเสี่ยงต่อการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำโรงพยาบาลในประเทศไทย โดยใช้ข้อมูลของผู้ป่วยที่เข้ารับการรักษาหูดผู้ป่วยใน แผนกอายุรศาสตร์ โรงพยาบาลจุฬาลงกรณ์ ในปี 2009 เป็นเวลา 3 วันหรือมากกว่านั้น โดยประเมินจากปัจจัยต่างๆ ได้แก่ ภาวะหัวใจล้มเหลว, โรคทางเดินหายใจเรื้อรังที่มีอาการเฉียบพลัน, อัมพฤกษ์หรืออัมพาต, เป็นมะเร็งเม็ดเลือดและไม่เป็นมะเร็งเม็ดเลือด, โรคภูมิแพ้ตัวเอง, การหายใจผิดปกติที่ต้องใช้เครื่องช่วยหายใจหรือการช่วยหายใจแบบไม่ต้องผ่าตัด, กระดูกหัก, โรคไขข้ออักเสบที่ขา, ประวัติก่อนเป็น VTE, ประวัติครอบครัวที่เป็น VTE, ประวัติก่อนเป็นมะเร็ง, เส้นเลือดขอด, การใช้ Oestrogen, ประวัติการอุดตัน, จำนวนเกล็ดเลือด $> 600 \times 10^9/l$, ดัชนีมวลกาย $> 30 \text{ kg/m}^2$ และกลุ่มอาการของโรคไต แบบฟอร์มการประเมินความเสี่ยงจะถูกแนบมาในแฟ้มประวัติผู้ป่วย แต่ละปัจจัยความเสี่ยงที่เกิดขึ้นจะถูกประเมินโดยแพทย์ และประเมินอีกครั้ง

หลังจากออกจากโรงพยาบาล กรณีที่ไม่มีปัจจัยความเสี่ยงข้างต้น หรือได้รับยาละลายลิ่มเลือดระหว่างอยู่หอผู้ป่วยจะได้รับการยกเว้นนำมาวิเคราะห์ (Rojnuckarin et al., 2011)

ปัจจัยความเสี่ยงของ VTE ถูกแสดงเป็นความเสี่ยงสัมพัทธ์ 95% ของช่วงความเชื่อมั่น (CI) และ P-value ที่ได้จากการ Chi Square Test จากนั้นนำปัจจัยที่มีนัยสำคัญมาวิเคราะห์แบบหลายตัวแปร โดยการวิเคราะห์ทางสถิติทั้งหมดจะถูกคำนวณโดยโปรแกรม SPSS 16.0 สำหรับ Windows จากวิเคราะห์หลายตัวแปร มีเพียงโรคภูมิแพ้ตัวเอง ก้อนเนื้ออก ประวัติครอบครัวที่เป็น VTE เส้นเลือดอุดตัน และการใช้ Oestrogen (Rojnuckarin et al., 2011)

2.1.5 การรักษาลิ่มเลือดในหลอดเลือดดำอุดตัน

วิธีการรักษา VTE ขึ้นอยู่กับความรุนแรงของอาการของผู้ป่วยเป็นหลัก โดยในกรณีที่ผู้ป่วยเป็นลิ่มเลือดอุดตันแบบ PE ผู้ป่วยส่วนใหญ่จะเข้ามารักษาด้วยอาการที่เป็นภาวะฉุกเฉิน แพทย์จึงจำเป็นต้องกักชีวิตผู้ป่วยเป็นอันดับแรก จากนั้นจึงประเมินความรุนแรงของอาการและสุขภาพของผู้ป่วย โดยการรักษามักจะให้ยาต้านการแข็งตัวของเลือด ซึ่งมีฤทธิ์ช่วยให้เลือดแข็งตัวช้าลง แบ่งออกเป็น 2 ชนิด คือ ยาวาร์ฟาริน (Warfarin) และยาเฮพาริน (Heparin) อาจเป็นยารับประทานแบบชนิดเม็ดหรือแบบยาชนิดฉีด ขึ้นอยู่กับดุลยพินิจของแพทย์ ในกรณีที่ผู้ป่วยมีอาการความดันโลหิตต่ำหรือช็อก แพทย์อาจให้สารน้ำร่วมกับยาละลายลิ่มเลือดซึ่งจะช่วยละลายลิ่มเลือดที่อวัยวะโดยตรง แต่ผู้ป่วยที่มีสภาพร่างกายไม่เหมาะแก่การใช้ยาอาจจำเป็นต้องผ่าตัดเอาลิ่มเลือดออก

สำหรับผู้ป่วยที่เป็นลิ่มเลือดอุดตันแบบ DVT มักมีอาการไม่รุนแรง จึงรักษาตามอาการได้ โดยปกติแพทย์จะให้ยาต้านการแข็งตัวของเลือด และควรใช้ยาติดต่อกัน 3-6 เดือน หรืออาจนานกว่านั้นถ้าเคยมีประวัติเป็นลิ่มเลือดมาก่อน ทั้งนี้ยาชนิดนี้มีการข้างเคียงที่ทำให้มีเลือดออกได้ง่าย แพทย์อาจต้องตรวจเลือดผู้ป่วยเป็นระยะๆ เพื่อประเมินปริมาณยาที่ผู้ป่วยได้รับนั้นมากหรือน้อยเกินไป บางกรณีแพทย์อาจพิจารณาการให้ยาละลายลิ่มเลือดร่วมด้วย

นอกจากใช้ยาแล้ว ยังมีอีกหลายวิธี เช่น การวางตะแกรงกรองลิ่มเลือดในหลอดเลือดดำใหญ่เพื่อดักจับลิ่มเลือดที่กำลังเคลื่อนตัวไปยังปอด หรือการใส่ถุงกรองซัพพอร์ทเพื่อกระตุ้นให้เลือดไหลเวียนได้ดีขึ้น ซึ่งจะช่วยป้องกันการเกิดลิ่มเลือด

ที่มา: (พบแพทย์, 2016)

2.2 ข้อมูลที่สำคัญต่อการวิเคราะห์ภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

ผลการศึกษาจากองค์การอนามัยโลก (World Health Organization: WHO) พบว่า สาเหตุการเสียชีวิตและทุพพลภาพที่เกิดจากผู้ป่วยเข้ารับการรักษาที่โรงพยาบาล โดยนอนอยู่กับเตียง ไม่ขยับหรือเคลื่อนไหวร่างกาย ไม่ลุกขึ้นเดินเป็นเวลานาน แพทย์จำเป็นต้องแนะนำให้บริหารขา เพื่อให้เลือดได้ไหลเวียน ถึงแม้ว่าภาวะลิ่มเลือดอุดตันจะสามารถป้องกันได้ แต่ก็พบว่าเป็นปัจจัยเสี่ยงที่ทำให้เกิดภาวะหัวใจล้มเหลว ภาวะหลอดเลือดสมองตีบ ซึ่งเป็นกลุ่มของโรคหัวใจ และหลอดเลือด และเป็นสาเหตุสำคัญที่ทำให้ผู้ป่วยเสียชีวิตและทุพพลภาพมากที่สุดในโลก (มหาวิทยาลัยมหิดล, 2020)

สาเหตุของภาวะการเกิดลิ่มเลือดอุดตันที่พบในงานวิจัยต่างๆ ได้มีการศึกษาปัจจัยความเสี่ยงทั้งที่เป็นปัจจัยภายในและปัจจัยภายนอก ตัวอย่างเช่น ปัจจัยภายใน เช่น อายุ เชื้อชาติ พันธุกรรม มะเร็ง โรคหัวใจ เป็นต้น สำหรับปัจจัยภายนอก เช่น อาหาร สิ่งแวดล้อม การเดินทาง เป็นต้น โดยผู้ที่มีความเสี่ยงหลายปัจจัยร่วมกัน ก็มีโอกาสเพิ่มความเสี่ยงต่อการเกิดลิ่มเลือดอุดตันมากขึ้นเช่นกัน

อายุ ในผู้สูงอายุภาวะการเกิดลิ่มเลือดอุดตันจะชัดเจนมากขึ้น โดยเฉพาะในอายุที่มากกว่า 50 ปี

เพศ จะมีผลขึ้นอยู่กับช่วงอายุของผู้ป่วยด้วย บางการศึกษาพบว่าภาวะการเกิดลิ่มเลือดอุดตันระหว่างเพศชาย และเพศหญิงจะไม่แตกต่างกันมาก ในขณะที่การศึกษาอื่นที่ขัดแย้งพบว่าอาจมีผลจากปัจจัยอื่นๆ ร่วมอีกด้วย

เชื้อชาติ ในการศึกษาภาวะการเกิดลิ่มเลือดอุดตันประชากรในสหรัฐอเมริกา ที่อยู่ในรัฐเดียวกัน หรือภูมิภาคเดียวกัน พบว่าจะมีภาวะการเกิดสูงสุดกับชาวแอฟริกันอเมริกัน ตามด้วยชาวคอเคเซียน อีสพานิกส์ ตามลำดับ แต่สำหรับชาวเอเชียจะมีภาวะการเกิดที่ต่ำกว่าประมาณ 4-5 เท่า ดังนั้นความแตกต่างทางเชื้อชาติจึงมีความสัมพันธ์กับปัจจัยทางพันธุกรรมซึ่งมีส่วนสำคัญต่อการเกิดภาวะลิ่มเลือดอุดตัน

พันธุกรรม มีการศึกษาจำนวนมากที่ศึกษาการเกิดลิ่มเลือดอุดตันกับชาวตะวันตก ซึ่งแสดงให้เห็นถึงปัจจัยทางพันธุกรรมที่มีผลต่อการเพิ่มความเสี่ยงต่อการเกิดภาวะลิ่มเลือดอุดตันที่เกิดครั้งแรกและเกิดซ้ำๆ โดยปัจจัยทางพันธุกรรมนี้ส่งผลให้ผู้ป่วยมีภาวะเลือดแข็งตัวง่ายผิดปกติ ซึ่งแบ่งได้ 2 กลุ่มคือ

กลุ่มที่ 1 การขาดสารต้านการแข็งตัวของเลือด พบบ่อยในชาวเอเชีย รวมทั้งคนไทยแต่พบน้อยในชาวตะวันตก

กลุ่มที่ 2 การเพิ่มขึ้นของปริมาณหรือการทำงานของ Coagulation Factor เป็นความผิดปกติที่พบบ่อยที่สุดในชาวตะวันตกแต่พบน้อยมากในคนไทย ซึ่งความผิดปกติในกลุ่มนี้มีความเสี่ยงต่อภาวะการเกิดลิ่มเลือดอุดตัน แต่น้อยกว่ากลุ่มที่ 1

ผู้ป่วยที่เป็นมะเร็ง พบว่าผู้ป่วยที่เป็นมะเร็งมีภาวะการเกิดลิ่มเลือดอุดตันสูงกว่าประชากรปกติมาก ซึ่งมะเร็งจะเพิ่มความเสี่ยงสูงถึง 7 เท่า โดยมะเร็งที่มีความเสี่ยงต่อการเกิดลิ่มเลือดอุดตัน ได้แก่ มะเร็งกระเพาะ มะเร็งสมอง มะเร็งตับอ่อน มะเร็งรังไข่ มะเร็งโลหิต มะเร็งปอด มะเร็งทางเดินอาหาร และมะเร็งตับ สำหรับปัจจัยที่เพิ่มความเสี่ยงต่อการเกิดลิ่มเลือดอุดตันในผู้ป่วยที่เป็นมะเร็ง ได้แก่ ยาเคมีบำบัด การให้ฮอร์โมนบำบัดในผู้ป่วยมะเร็งเต้านมแต่การผ่าตัดและการฉายแสงจะไม่เพิ่มความเสี่ยงในผู้ป่วยที่เป็นมะเร็ง

ภาวะเจ็บป่วยทางอายุรกรรม พบว่าส่วนใหญ่ผู้ป่วยที่นอนรักษาตัวอยู่ในโรงพยาบาล ในหอผู้ป่วยอายุรกรรมทั่วไป จะมีความเสี่ยงต่อภาวะการเกิดลิ่มเลือดอุดตันเพิ่มขึ้น โดยไม่ได้รับยาต้านการแข็งตัวของเลือดประมาณร้อยละ 5-45 ในการป้องกัน ภาวะที่ทำให้เพิ่มความเสี่ยงต่อภาวะการเกิดลิ่มเลือดอุดตันเพิ่มขึ้น ได้แก่ มะเร็ง อัมพาต ภาวะหัวใจล้มเหลว โรคหัวใจขาดเลือด เบาหวาน ความอ้วน ไขมันในเลือดสูง ภาวะติดเตียง เป็นต้น

การผ่าตัดและการได้รับบาดเจ็บ การผ่าตัดเป็นปัจจัยเสี่ยงสูงที่สุดโดยมีอัตราเสี่ยงสูงถึง 21.7 เท่า และการได้รับบาดเจ็บมีอัตราเสี่ยงสูงรองลงมา อยู่ที่ 12.7 เท่า โดยขึ้นอยู่กับชนิดของการผ่าตัดและการบาดเจ็บ

ที่มา: (Rojnuckarin et al., 2011)

2.3 ทฤษฎีการวิเคราะห์ข้อมูลด้วยหลักการทางคอมพิวเตอร์

2.3.1 การจำแนกประเภทข้อมูล (Classification)

เมื่อจำแนกประเภทของข้อมูลตามลักษณะของข้อมูลจะจำแนกเป็น 2 ลักษณะใหญ่ๆ คือ ข้อมูลเชิงคุณภาพ (Qualitative Data) และข้อมูลเชิงปริมาณ (Quantitative Data)

2.3.1.1 ข้อมูลเชิงคุณภาพ คือ ข้อมูลที่แสดงถึงสถานภาพ คุณลักษณะ หรือคุณสมบัติ เช่น เพศ เชื้อชาติ สถานภาพสมรส ศาสนา กลุ่มเลือด เป็นต้น

2.3.1.2 ข้อมูลเชิงปริมาณ คือ ข้อมูลที่อยู่ในรูปตัวเลข (Numerical Data) ที่แสดงถึงปริมาณ กรณีที่เป็นค่าไม่ต่อเนื่อง (Discrete) จะเป็นจำนวนเต็มหรือจำนวนนับ เช่น จำนวนนิสิตปริญญาโทในมหาวิทยาลัย จำนวนพนักงานในบริษัท เป็นต้น หรือถ้าเป็นค่าที่

ต่อเนื่อง (Continuous) จะเป็นค่าที่มีจุดทศนิยมได้ เช่น ความสูง อัตราดอกเบี้ย อุณหภูมิ ในแต่ละวัน เป็นต้น

ที่มา: (สำนักงานสถิติแห่งชาติ)

2.3.2 การทำความเข้าใจกับข้อมูล (Data understanding)

เพื่อเป็นการศึกษาข้อมูล ผู้วิจัยจะต้องทำความเข้าใจในข้อมูลแต่ละตัวแปรที่จะนำมาใช้งานนั้น แต่ละตัวแปรคืออะไร ค่าของแต่ละตัวแปรมีความหมายว่าอะไรบ้าง แล้วตรวจสอบความถูกต้องของข้อมูล เพื่อพิจารณาว่าข้อมูลนั้นจำเป็นหรือไม่ที่จะนำมาวิเคราะห์ในงานวิจัยนี้

2.3.3 การเตรียมข้อมูล (Data Preparation)

เพื่อเตรียมข้อมูลทั้งหมดให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้ จะแบ่งออกเป็น 3 ขั้นตอนดังนี้

2.3.3.1 Data Selection เป็นขั้นตอนการเลือกตัวแปรที่มีความสำคัญ และมีประโยชน์เพื่อนำไปวิเคราะห์ใช้ในงานวิจัยนี้

2.3.3.2 Data Cleansing เป็นขั้นตอนตรวจสอบข้อมูลในแต่ละตัวแปรว่า ข้อมูลที่ผิดปกติ ข้อมูลที่มีค่าว่าง หรือข้อมูลที่สะกดไม่ถูกต้อง ซึ่งถ้าพบข้อมูลดังกล่าวจะต้องทำการแก้ไข หรือเพิ่มข้อมูลให้ครบถ้วน เพื่อนำไปวิเคราะห์ในงานวิจัยต่อไปได้อย่างเหมาะสม

2.3.3.3 Data Transformation เป็นขั้นตอนการแปลงข้อมูลของแต่ละตัวแปรให้อยู่ในรูปแบบที่ง่ายต่อการทำงานของเครื่องคอมพิวเตอร์และในการวิเคราะห์

2.3.4 การสร้างโมเดล (Modeling)

เป็นขั้นตอนการสร้างแบบจำลองการเรียนรู้ของเครื่อง โดยจะนำข้อมูลทั้งหมดมาแบ่งกลุ่มเป็น Training Data หรือ ข้อมูลที่ใช้สำหรับการเรียนรู้ของแบบจำลอง กับ Testing Data หรือ ข้อมูลที่ใช้ในการทดสอบกับแบบจำลอง แล้วตรวจสอบความถูกต้องโดยการปรับจูนพารามิเตอร์ของแบบจำลองเพื่อให้ได้ค่าที่มีความถูกต้องในระดับที่น่าพอใจ

2.3.5 การประเมินผล (Evaluation)

เป็นขั้นตอนประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง เพื่อใช้ในการเปรียบเทียบประสิทธิภาพและหาว่าแบบจำลองตัวไหนที่ดีที่สุด เพื่อนำแบบจำลองนั้นนำไปใช้งานต่อได้

2.4 งานวิจัยที่ศึกษาเกี่ยวกับปัจจัยการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

ผู้วิจัยนี้ได้ทำการทบทวนวรรณกรรมและศึกษาค้นคว้างานวิจัย ที่เกี่ยวข้องกับปัจจัยการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ โดยเลือกรายละเอียดที่สำคัญในงานวิจัยบางส่วน ดังมีรายละเอียดต่อไปนี้

2.4.1 งานวิจัยเรื่อง Risk Factors for symptomatic venous thromboembolism in Thai hospitalized medical patients (Rojnuckarin et al., 2011)

ในประเทศไทยการป้องกันการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำกับผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลนั้นยังล้มเหลว และยังไม่มีความรู้ของเครื่องสำหรับการทำนายภาวะเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ การศึกษานี้จึงเน้นไปที่การวิเคราะห์และการจำแนกผู้ป่วยที่มีความเสี่ยงสูงออกมาได้

ในปี พ.ศ.2552 ผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลจุฬาลงกรณ์นั้น ได้รับการประเมินวิเคราะห์หาปัจจัยเสี่ยงที่ทำให้เกิดลิ่มเลือดอุดตันในหลอดเลือดดำ โดยกลุ่มผู้ป่วยที่มีความเสี่ยงสูงนั้นจะถูกเฝ้าติดตามตั้งแต่มีอาการของลิ่มเลือดอุดตัน จนถึง 6 สัปดาห์หลังจากออกจากโรงพยาบาล ซึ่งไม่มีการป้องกันด้วยยาเฮปาริน (Heparin) จากจำนวนผู้ป่วยที่มีความเสี่ยงสูง 1,290 ราย โดย 27 รายจะเป็นผู้ป่วยที่เป็นลิ่มเลือดอุดตันในหลอดเลือดดำ และมีจำนวน 25.9% ได้รับการวินิจฉัยหลังจากออกจากโรงพยาบาล สำหรับกรณีที่เป็นลิ่มเลือดอุดตันในหลอดเลือดดำที่เกิดขึ้นในโรงพยาบาลจะนานกว่า

ปัจจัยเสี่ยงที่มีนัยสำคัญในงานวิจัยนี้ได้แก่ โรคแพภูมิตัวเอง ก้อนเนื้องอก มีประวัติครอบครัวเป็นลิ่มเลือดอุดตัน เส้นเลือดขาด และการใช้ยาเอสโตเจน โดยมีความเสี่ยงสัมพัทธ์เท่ากับ 11.8, 4.7, 120.3, 40.1 และ 17.1 ตามลำดับ ดังภาพประกอบ 6

Risk factors	Relative risk (95% CI)	P-value
Autoimmune disease	11.83 (3.89–35.97)	<0.001
Solid tumours	4.66 (1.84–11.79)	0.001
Family history of VTE	120.28 (6.89–2101)	0.001
Varicose vein	40.09 (3.75–429.27)	0.002
Oestrogen	17.08 (1.18–248.20)	0.038
95% CI, 95% confidence interval.		

ภาพประกอบ 6 แสดงปัจจัยเสี่ยงที่มีนัยสำคัญ

ในทางตรงกันข้าม ก่อนหน้านี้รายงานคะแนนลิ้มเลือดอุดตันในหลอดเลือดดำของผู้ป่วยที่เป็นชาวตะวันตก จะไม่สามารถจำแนกความเสี่ยงของลิ้มเลือดอุดตันในหลอดเลือดดำได้ แต่รายงานคะแนนทั้งหมดจะทำนายว่ามีอัตราการเสียชีวิตสูงขึ้น

สรุปได้ว่า ลิ้มเลือดอุดตันในหลอดเลือดดำเป็นเรื่องปกติสำหรับผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล โดยผู้ป่วยที่เป็นโรคแพ้ภูมิตัวเองและผู้ป่วยที่มีเนื้องอกจะมีความอ่อนไหวต่อลิ้มเลือดอุดตันในหลอดเลือดดำสูง ดังนั้นจึงจำเป็นต้องมีวิธีการป้องกันโรคในกลุ่มเหล่านี้

2.4.2 งานวิจัยเรื่อง Predictive and associative models to identify hospitalized medical patients at risk for VTE (Spyropoulos et al., 2011)

งานวิจัยนี้จะประเมินอุบัติการณ์ของลิ้มเลือดอุดตันในหลอดเลือดดำ โดยใช้คะแนนประเมินความเสี่ยงลิ้มเลือดอุดตันในหลอดเลือดดำที่เข้ารับการรักษา และคะแนนลิ้มเลือดอุดตันในหลอดเลือดดำที่อยู่ระหว่างการรักษาในโรงพยาบาล (International Medical Prevention Registry on Venous Thromboembolism: IMPROVE) โดยวิเคราะห์ข้อมูลจากผู้ป่วยจำนวน 15,156 รายใน 12 ประเทศและโรงพยาบาล 52 แห่ง เพื่อหาอุบัติการณ์สะสมของลิ้มเลือดอุดตันในหลอดเลือดดำในช่วงเวลา 3 เดือน จากการเข้ารับการรักษาในการติดตามผล 3 เดือนถูกประเมินโดยใช้วิธี Kaplan-Meier แต่เนื่องจากระยะเวลาพักรักษาตัวในโรงพยาบาลไม่เท่ากัน และไม่ใช่ผู้ป่วยทุกรายที่มีข้อมูล 3 เดือน จึงใช้การวิเคราะห์การถดถอยพหุคูณของ Cox เพื่อระบุปัจจัยเสี่ยงที่จะเกิดลิ้มเลือดอุดตันในหลอดเลือดดำที่เป็นอิสระ

ประชากรในงานวิจัยของ (Tapson et al., 2007) จำแนกได้เป็น เพศหญิง 50% มัธยมฐานของอายุอยู่ที่ 68 ปี มัธยมฐานของน้ำหนักตัวอยู่ที่ 69 กิโลกรัม ระยะเวลาพักรักษาตัวในโรงพยาบาลเฉลี่ย 7 วัน ดังภาพประกอบ 7 ภาวะทางการแพทย์ที่พบบ่อยในกลุ่มข้อมูล ได้แก่ โรค

ติดเชื้อ (32%) การหายใจล้มเหลว (19%) และมะเร็ง (12%) ผู้ป่วยที่เป็นโรคหัวใจคิดเป็น 37% ของจำนวนประชากรที่ศึกษา (โรคหัวใจขาดเลือด 12% ภาวะหัวใจล้มเหลว 11% ภาวะหัวใจอื่นๆ 14%) และผู้ป่วยโรคหลอดเลือดสมองคิดเป็น 6% และ 4% ของผู้ป่วยที่มีประวัติลิ้มเลือดอุดตันในหลอดเลือดดำก่อนหน้า และ 2% เป็นอัมพาตครึ่งท่อนล่าง จากผู้ป่วยทั้งหมด 13,172 ราย (87%)

Characteristic	Patients (N = 15,156)
Women, %	50
Age, y	68 (52-79)
Weight (No. = 10,433), kg	69 (59-81)
Length of hospital stay, d	7 (5-13)
Immobile for ≥ 7 d, including days immediately prior to admission (No. = 15,125), %	19
Time immobile (No. = 1,169), d preadmission	3 (1-15)

ภาพประกอบ 7 แสดงรายงานลักษณะของผู้ป่วยก่อนหน้า

ผู้ป่วยส่วนใหญ่ (98 จาก 143, 69%) มีอาการลิ้มเลือดอุดตันในหลอดเลือดดำภายใน 30 วันหลังจากเข้ารับการรักษาในโรงพยาบาล (ค่ามัธยฐาน 16 วัน) และแบ่งเป็นผู้ป่วยที่เกิดอาการขณะเข้ารับการรักษาอยู่ในโรงพยาบาลจำนวน 79 ราย (55%) ผู้ป่วยที่เกิดอาการหลังจากออกจากโรงพยาบาลจำนวน 64 ราย (45%) ตามลำดับ ดังภาพประกอบ 8

Days After Hospital Admission	In-hospital VTE (n = 79)	Postdischarge VTE (n = 64)	All VTE (N = 143)
1-7	42 (53)	0	42 (29)
8-30	32 (41)	24 (38)	56 (39)
31-60	5 (6)	23 (36)	28 (20)
61-91	0	17 (27)	17 (12)

ภาพประกอบ 8 แสดงอาการลิ้มเลือดอุดตันในหลอดเลือดดำหลังเข้ารับการรักษาในโรงพยาบาล

แบบจำลองการถดถอยพหุคูณของ Cox กับผู้ป่วย 15,156 รายที่มีข้อมูลครบถ้วน แสดงปัจจัยที่มีนัยสำคัญที่เกี่ยวข้องกับความเสี่ยงของลิ้มเลือดอุดตันในหลอดเลือดดำในช่วง 3 เดือนหลังเข้ารับการรักษาในโรงพยาบาลมากที่สุดคือ ประวัติการเป็นลิ้มเลือดอุดตันในหลอดเลือดดำ ภาวะเกล็ดเลือดต่ำ มะเร็ง และอายุที่มากกว่า 60 ปี ดังภาพประกอบ 10

VTE Risk Factor	HR (95% CI)	χ^2	P Value	Points
Previous VTE	5.0 (3.3-7.8)	53	<.001	3
Known thrombophilia	5.2 (1.3-21.5)	5.2	.02	3
Cancer	2.0 (1.3-3.1)	11	.001	1
Age > 60 y	1.8 (1.2-2.7)	8.5	.004	1

ภาพประกอบ 9 แสดงปัจจัยที่มีนัยสำคัญที่เกี่ยวข้องกับความเสี่ยงของลิ่มเลือดอุดตันในหลอดเลือดดำ

ปัจจัยความเสี่ยงทั้งสี่ประการที่เข้ารับการรักษาในโรงพยาบาลสามารถทำนายความเสี่ยงของลิ่มเลือดอุดตันในหลอดเลือดดำในผู้ป่วยที่เข้ารับรักษาในโรงพยาบาลได้ และคะแนน IMPROVE ที่ได้จากปัจจัยทางคลินิก 7 ประการระหว่างการรักษาในโรงพยาบาลอาจช่วยให้เข้าใจความเสี่ยงของลิ่มเลือดอุดตันในหลอดเลือดดำที่มีอาการมากขึ้น

2.5 งานวิจัยที่ศึกษาเกี่ยวกับการวิเคราะห์ความเสี่ยงของภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ

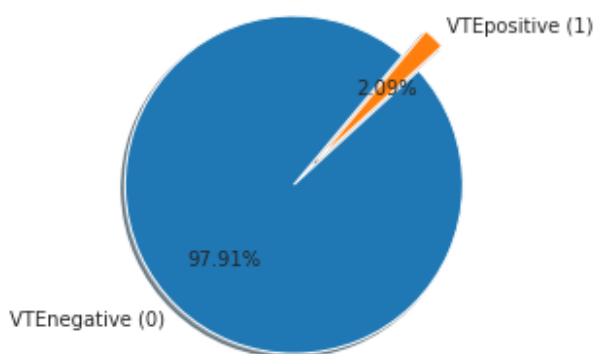
ผู้วิจัยนี้ได้ทำการทบทวนวรรณกรรมและศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับการวินิจฉัยและวิเคราะห์ความเสี่ยงของภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำ ซึ่งเลือกรายละเอียดที่สำคัญในงานวิจัยเพียงบางส่วน โดยมีรายละเอียดต่อไปนี้

2.5.1 งานวิจัยเรื่อง Automatic Diagnosis of Venous Thromboembolism Risk based on Machine Learning (Sukperm et al., 2021)

งานวิจัยนี้ สร้างแบบจำลองการวินิจฉัยความเสี่ยงต่อการเกิดโรคลิ่มเลือดอุดตันในหลอดเลือดดำอัตโนมัติโดยอาศัยหลักการการเรียนรู้ของเครื่อง และเปรียบเทียบประสิทธิภาพแบบจำลองของทั้ง 3 อัลกอริทึมได้แก่ Decision Tree, Logistic Regression และ Neural Network รวมถึงการจัดการปัญหาที่เกี่ยวข้องกับข้อมูลต่างๆ เช่น ความไม่สมดุลของข้อมูล การจัดการกับข้อมูลที่เป็นค่าว่าง และการแปลงข้อมูล

ข้อมูลงานวิจัยนี้มาจากผู้ป่วยที่เข้ารับการรักษาในหอผู้ป่วยแผนกอายุรศาสตร์ ของโรงพยาบาลจุฬาลงกรณ์ ภายในปี พ.ศ.2552 ที่ได้รับการอนุมัติจากคณะกรรมการจริยธรรมของคณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยแล้ว ซึ่งข้อมูลนี้ประกอบด้วยจำนวนผู้ป่วยทั้งหมด 1,290 ราย และปัจจัยต่างๆ ที่ใช้ในการวิเคราะห์หาความเสี่ยงของการเกิดภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ จำนวน 65 ปัจจัย ซึ่งพบว่าผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ (VTE

Positive) มีจำนวน 27 ราย คิดเป็น 2.09% ของข้อมูลทั้งหมด และผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ (VTE Negative) มีจำนวน 1,263 ราย คิดเป็น 97.91% ของข้อมูลทั้งหมด ดังภาพประกอบ 10



ภาพประกอบ 10 แสดงจำนวนผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตัน VTE positive(1) และไม่เป็นโรคลิ่มเลือดอุดตัน VTE negative(0)

สำหรับการจัดการปัญหาที่เกี่ยวข้องกับข้อมูลต่างๆ งานวิจัยนี้ได้จัดการข้อมูลที่เป็นค่าว่าง โดยใช้หลักการเติมศูนย์ (Fill missing values with zero) เนื่องจากข้อมูลคนไข้แต่ละรายไม่สามารถแทนที่กันได้ จึงใช้หลักการเติมค่าว่างด้วยตัวเลขศูนย์ และข้อมูลในบางปัจจัยนั้น ชนิดของข้อมูลเป็นตัวอักษรจะต้องแปลงข้อมูลตัวอักษรให้เป็นตัวเลข ด้วยวิธีการแปลงข้อมูลแบบ Label Encoder ซึ่งแสดงดังภาพประกอบ 11

FinalDx	Comorbid	Procedure	Death	CauseofDeath	Autopsy	DischargeDate
SLE	Myelopathy	Insertion of endotracheal tube. Mechanical ven...	0	0	0	4/8/2009
Guillain-Barre syndrome	Delirium, not superimposed on dementia, so des...	Insertion of endotracheal tube. Mechanical ven...	1	1	1	1/6/2009
Other sstreptococcal septicaemia	Nephrotic syndrome, focal and segmental glomer...	Computerized axial tomography of abdomen	0	0	0	5/28/2009
Cellulitis	DM, Bullous pemphigoid	Computerized axial tomography of abdomen and h...	0	0	0	7/14/2009
Secondary malignant neoplasm of bone and bone ...	Radiotherapy session, Malignant neoplasm, Obst...	Computerized axial tomography of abdomen and h...	0	0	0	11/22/2008

FinalDx	Comorbid	Procedure	Death	CauseofDeath	Autopsy	DischargeDate
528	668	383	0	0	0	255
239	316	383	1	1	1	28
468	738	132	0	0	0	276
140	304	146	0	0	0	322
536	861	146	0	0	0	105

ภาพประกอบ 11 Label Encoder

เนื่องจากจำนวนผู้ป่วยที่เป็นโรคหลอดเลือดอุดตันในหลอดเลือดดำ (VTE Positive) มีจำนวน 27 ราย คิดเป็น 2.09% ของข้อมูลทั้งหมด ซึ่งจะมีผลกับการสร้างแบบจำลอง ทำให้เกิด Bias ขึ้นกับข้อมูล เพราะอัลกอริทึมจะไม่สามารถตรวจพบข้อมูลในคลาสที่น้อยกว่าได้ การจัดการกับปัญหาข้อมูลที่ไม่สมดุลนั้น เป็นวิธีการเปลี่ยนแปลงจำนวนข้อมูล โครงสร้างของข้อมูล และคุณสมบัติของข้อมูลเดิม ซึ่งอาจทำให้เกิดการสูญเสียคุณสมบัติเดิมของข้อมูลเดิม เพื่อที่จะเพิ่มประสิทธิภาพการทำนายของแบบจำลองให้มีประสิทธิภาพที่ดีขึ้นจึงจำเป็นต้องจัดการและแก้ปัญหาคข้อมูลไม่สมดุล โดยจะใช้ทั้งหมด 5 วิธีได้แก่ Oversampling, Undersampling, Synthetic Minority Oversampling TEchnique (SMOTE), Class Weight และ Ensemble Sampling

สรุปผลการทดลองพบว่าแบบจำลอง Decision Tree และใช้วิธีการปรับความสมดุลของข้อมูลด้วย Class Weight ซึ่งจะให้ค่า Sensitivity เท่ากับ 0.25 (25%) ค่า Precision เท่ากับ 0.22 (22%) และ ค่า F1-score เท่ากับ 0.24 (24%) ดังภาพประกอบ 12

Balanced Method	Evaluate Model	Decision Tree	Logistic Regression	Neural Network
Class Weight	Sensitivity	0.25	0.75	-
	Precision	0.22	0.06	-
	F1-score	0.24	0.12	-

ภาพประกอบ 12 Decision Tree

2.5.2 งานวิจัยเรื่อง A Machine Learning Approach to Predict Deep Venous Thrombosis Among Hospitalized Patients (Ryan et al., 2021)

งานวิจัยนี้ได้แสดงถึงปัจจัยเสี่ยงทางพันธุกรรมที่ส่งผลโดยตรงกับการเกิดโรค และปัจจัยที่ได้มาที่มีความเสี่ยงสูงที่จะเป็นลิ่มเลือดอุดตันในหลอดเลือดดำได้แก่ การไม่เคลื่อนไหวเป็นเวลานาน การนอนติดเตียง การผ่าตัด และการไหลเวียนโลหิตบกพร่อง

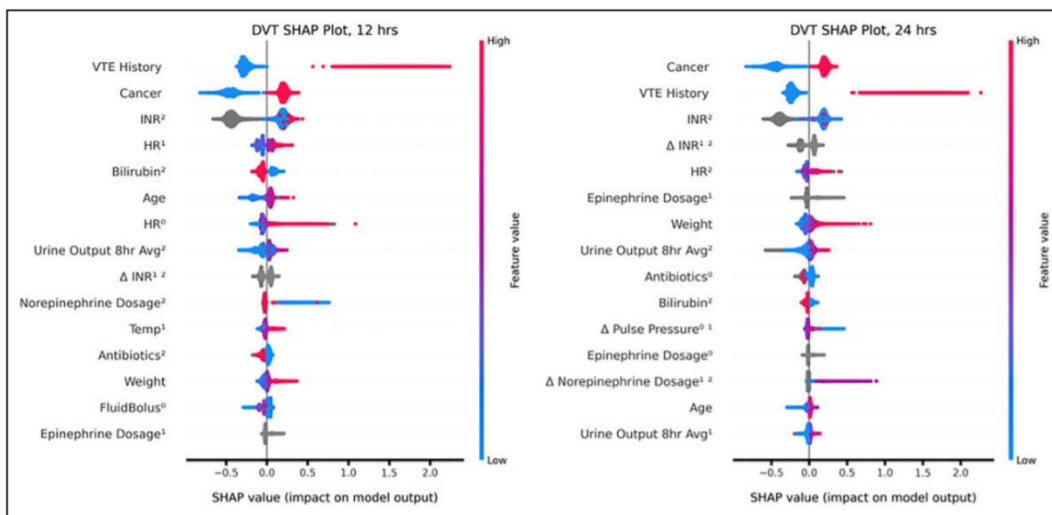
นอกจากนี้ ยังมีการทดสอบสำหรับการจำแนกความเสี่ยงในผู้ป่วยที่สงสัยว่าจะเป็นลิ่มเลือดอุดตัน โดยใช้คะแนน Wells ซึ่งเหมาะสำหรับประเมินความเสี่ยงต่ำกว่าผู้ป่วยที่เป็น DVT ต่ำ และคะแนน IMPROVE (International Medical Prevention Registry on Venous Thromboembolism) ซึ่งใช้สำหรับประเมินความเสี่ยงของ VTE โดยคะแนน Well จะแปรผันตามอายุและผู้ป่วยที่มีประวัติ VTE

งานวิจัยนี้ได้ใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบ Extreme Gradient Boosting จำนวน 2 ตัว โดยตัวที่ 1 ใช้สำหรับทำนายความเสี่ยงก่อนมีอาการ 12 ชั่วโมง ตัวที่ 2 ใช้สำหรับทำนายความเสี่ยงก่อนมีอาการ 24 ชั่วโมง

ผลที่ได้จากการทำ Feature Importance จะมีความคล้ายคลึงกันระหว่างแบบจำลองของ 12 ชั่วโมงและ 24 ชั่วโมง คือ มะเร็ง, ประวัติการเป็น VTE, International Normalized Ratio (INR), การเปลี่ยนแปลงของ INR และ อัตราการเต้นของหัวใจ เป็น 5 อันดับแรก โดยมะเร็งเป็นปัจจัยเสี่ยงที่กระตุ้นทำให้เกิด VTE ที่สำคัญที่สุด โรคอ้วนหรือน้ำหนักตัวที่

เพิ่มขึ้นแสดงให้เห็นว่า มีความสัมพันธ์ที่จะเพิ่มความเสี่ยงสำหรับ VTE มากขึ้น ดังภาพประกอบ

13



ภาพประกอบ 13 DVT SHAP Plot

สำหรับการทำนายความเสี่ยงก่อนเริ่มอาการทั้งแบบ 12 ชั่วโมงและแบบ 24 ชั่วโมง สามารถจำแนกความเสี่ยงและป้องกันการทำให้ Invasive Test กับผู้ป่วยที่ไม่จำเป็นได้ และสามารถเข้ายาด้านการแข็งตัวของเลือดได้ตรงเป้าหมายมากยิ่งขึ้น

2.5.3 งานวิจัยเรื่อง Machine learning to predict venous thrombosis in acutely ill medical patients (Nafee et al., 2020)

Ensemble Learning เป็นการเรียนรู้ประเภทหนึ่งที่รวมการคาดคะเนของ Predictors ต่างๆ ซึ่งในงานวิจัยนี้จะใช้ 10-Fold Cross Validation รวมถึง Elastic Net, Extreme Gradient Boosting, Random Forest , Bayesian Logistic Regression และ Simple Classification Tree โดยแบบจำลองตัวที่หนึ่ง (ML) จะใช้ตัวแปรทั้งหมด 68 ตัว ซึ่งประกอบด้วยตัวแปรพื้นฐานทั้งหมดที่มีอยู่ในชุดข้อมูลการทดลอง APEX สำหรับแบบจำลองตัวที่สอง (rML) จะใช้ตัวแปร 16 ตัวที่มีความสำคัญก่อนเป็นปัจจัยเสี่ยงสำหรับลิ่มเลือดอุดตันในหลอดเลือดดำ ซึ่งตัวแปรทั้ง 16 ตัวนี้เป็นส่วนหนึ่งของคะแนน IMPROVE

ประสิทธิภาพการทำนายถูกเปรียบเทียบด้วยการประเมินทางสถิติแบบ c-statistic และ bootstrapped test และใช้การทดสอบของ Hosmer-Lemeshow เพื่อทดสอบนัยสำคัญทาง

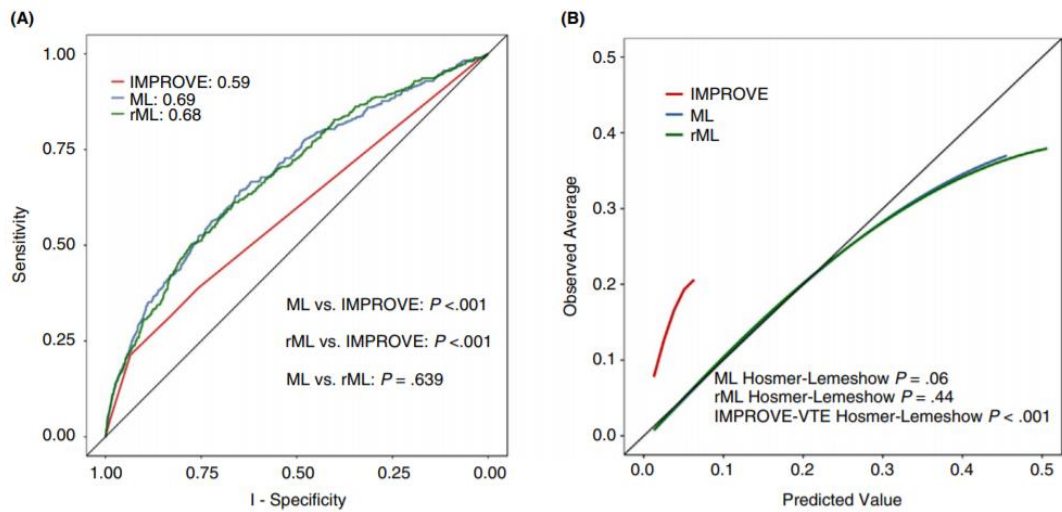
สถิติ ถ้าผลการทดสอบ Hosmer-Lemeshow ที่ไม่มีนัยสำคัญ ($P > .05$) แสดงให้เห็นถึงผลการทดสอบเทียบที่ดี แต่ถ้าผลการทดสอบที่สำคัญ ($P < .05$) แสดงให้เห็นถึงผลการทดสอบที่ไม่ดี

ลักษณะพื้นฐานและผลลัพธ์ของผู้ป่วยที่ลงทะเบียนในการทดลอง APEX แสดงไว้ดังภาพประกอบ 14

	Overall (N = 6459)	No VTE event (N = 6052)	VTE event (N = 407)	P value
Age, y, mean (SD)	76.31 (8.28)	76.26 (8.21)	77.05 (9.22)	.06
Male (%)	2924 (45.3)	2732 (45.1)	192 (47.2)	.46
Weight, kg, mean (SD)	80.55 (19.20)	80.62 (19.24)	79.46 (18.66)	.24
BMI, kg/m ² , median (IQR)	28.40 (24.90-33.20)	28.40 (24.90-33.20)	28.40 (24.35-32.40)	.20
Duration of hospitalization, days, median (IQR)	10.00 (8.00-14.00)	10.00 (8.00-14.00)	11.00 (8.00-15.00)	<.001
Creatinine clearance, mL/min, mean (SD)	71.23 (32.92)	71.41 (32.89)	68.56 (33.29)	.09
Race, n (%)				
White	6063 (95.5)	5686 (95.6)	377 (94.5)	.003
Black	116 (1.8)	106 (1.8)	10 (2.5)	
Asian	13 (0.2)	11 (0.2)	2 (0.5)	
American Indian	7 (0.1)	7 (0.1)	0 (0.0)	
Pacific Islander	1 (0.0)	0 (0.0)	1 (0.3)	
Multiple	44 (0.7)	43 (0.7)	1 (0.3)	
Other	104 (1.6)	96 (1.6)	8 (2.0)	
Using strong P-gp inhibitor, n (%)	1161 (18.0)	1091 (18.0)	70 (17.2)	.72
D-dimer, median (IQR)	1.24 (0.65-2.25)	1.20 (0.63-2.15)	2.05 (1.04-3.78)	<.001
History of cancer, n (%)	759 (11.8)	698 (11.5)	61 (15.0)	.04
History of thrombosis, n (%)	512 (7.9)	420 (6.9)	92 (22.6)	<.001
Chronic heart failure, n (%)	1470 (22.8)	1376 (22.7)	94 (23.1)	.91
Acute infectious disease, n (%)	1009 (15.6)	940 (15.5)	69 (17.0)	.49
Severe varicosities, n (%)	1201 (18.6)	1128 (18.6)	73 (17.9)	.77
Hormone replacement, n (%)	59 (0.9)	53 (0.9)	6 (1.5)	.43
Inherited or acquired thrombophilia (%)	7 (0.1)	5 (0.1)	2 (0.5)	.10

ภาพประกอบ 14 ลักษณะพื้นฐานและผลลัพธ์ของผู้ป่วยที่ลงทะเบียนในการทดลอง APEX

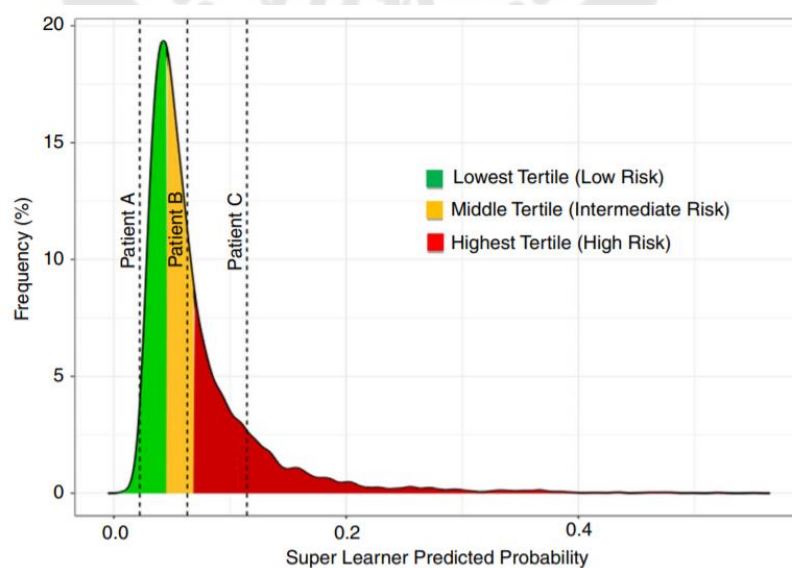
แบบจำลอง ML มีค่า c-statistic เท่ากับ 0.69 และแบบจำลอง rML มีค่า c-statistic เท่ากับ 0.68 โดยทั้งคู่แสดงให้เห็นถึงความสามารถที่เหนือกว่าคะแนน IMPROVE ซึ่งมีค่า c-statistic เท่ากับ 0.59 ($P < .001$) แบบจำลองทั้งสองมีการปฏิบัติที่คล้ายกัน โดยมีค่า P เท่ากับ .28 ดังภาพประกอบ 15 (รูป A)



ภาพประกอบ 15 แสดงการเปรียบเทียบ ROC curves ของ Super learner model (ML) และ Reduced model (rML) เทียบกับ IMPROVE score

รูป B แบบจำลองทั้งสองแสดงให้เห็นถึงผลที่ดีจากการทดสอบโดยใช้ Hosmer-Lemeshow ($P = .06$ และ $P = .44$ ตามลำดับ) ในทางตรงกันข้าม คะแนน IMPROVE ได้ผลที่ไม่ดีนัก (Hosmer-Lemeshow, $P < .001$) ดังภาพประกอบ 15 (B)

ความถี่ของความเสี่ยงที่คาดการณ์ไว้ซึ่งเกิดจากแบบจำลอง rML ถูกพลอตไว้ ดังภาพประกอบ 16 โดยแสดงลักษณะและผลลัพธ์ของผู้ป่วยในแต่ละ Tertile



ภาพประกอบ 16 แสดงการกระจายของความเสี่ยงที่ทำนายด้วย rML

ความเสี่ยง Tertile ต่ำสุดรวมความเสี่ยงที่คาดการณ์ไว้ระหว่าง $0 \leq 3.3\%$ โดยมี ความเสี่ยงมัธยฐาน 2.7% ความเสี่ยง Tertile ที่สองแสดงถึงความเสี่ยงระดับกลางตั้งแต่ $>3.3\% - <5\%$ โดยมีค่ามัธยฐาน 5% ความเสี่ยง Tertile ที่สามที่คาดการณ์ไว้ $\geq 5\%$ โดยมีค่ามัธยฐาน 7.4% ดังภาพประกอบ 17

	Lowest tertile (N = 2103)	Middle tertile (N = 2135)	Highest tertile (N = 2221)	P value
Treatment with betrixaban, n (%)	1031 (49.0)	1036 (48.5)	1140 (51.2)	.15
Treatment with enoxaparin, n (%)	1073 (51.0)	1099 (51.5)	1081 (48.8)	.15
Primary outcome event, n (%)	58 (2.8)	109 (5.1)	240 (10.8)	<.001
Age, y, mean (SD)	73.83 (7.24)	77.23 (7.40)	77.78 (9.40)	<.001
Male (%)	988 (47.0)	967 (45.3)	969 (43.6)	.09
Weight, kg, mean (SD)	85.17 (20.56)	78.53 (17.48)	78.11 (18.68)	<.001
BMI, kg/m ² , median (IQR)	29.60 (25.90-35.50)	28.10 (24.60-32.10)	27.50 (24.20-32.00)	<.001
Duration of hospitalization, days, median (IQR)	9.00 (7.00-13.00)	10.00 (7.75-14.00)	11.00 (8.00-15.00)	<.001
Creatinine clearance, mL/min, mean (SD)	75.84 (33.15)	69.93 (31.32)	68.08 (33.71)	<.001
Race (%)				
White	1996 (95.7)	1992 (95.1)	2075 (95.7)	.73
Black	37 (1.8)	35 (1.7)	44 (2.0)	
Asian	4 (0.2)	7 (0.3)	2 (0.1)	
American Indian	2 (0.1)	3 (0.1)	2 (0.1)	
Pacific Islander	0 (0.0)	1 (0.0)	0 (0.0)	
Multiple	14 (0.7)	18 (0.9)	12 (0.6)	
Other	32 (1.5)	39 (1.9)	33 (1.5)	
Use of strong P-gp inhibitor, n (%)	383 (18.2)	352 (16.5)	426 (19.2)	.07
D-dimer, median (IQR)	0.69 (0.39-1.11)	1.24 (0.76-1.83)	2.70 (1.52-4.37)	<.001
History of cancer, n (%)	217 (10.3)	275 (12.9)	267 (12.0)	.031
History of thrombosis, n (%)	2 (0.1)	13 (0.6)	497 (22.4)	<.001
Chronic heart failure, n (%)	500 (23.8)	467 (21.9)	503 (22.6)	.33
Acute infectious disease, n (%)	313 (14.9)	356 (16.7)	340 (15.3)	.24
Severe varicosities, n (%)	475 (22.6)	312 (14.6)	414 (18.6)	<.001
Hormone replacement, n (%)	4 (0.2)	14 (0.7)	41 (1.8)	<.001
Inherited or acquired thrombophilia, n (%)	2 (0.1)	2 (0.1)	3 (0.1)	.89

ภาพประกอบ 17 แสดงอัตราการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำในแต่ละปัจจัย

อัตราการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำที่สังเกตพบใน Tertile คือ 2.5% ใน Tertile ต่ำสุด, 4.8% ในกลุ่ม Tertile ที่สอง และ 11.4% ในกลุ่ม Tertile ที่สาม อัตราส่วนที่เปรียบเทียบของลิ่มเลือดอุดตันในหลอดเลือดดำ ใน Tertile ที่สองและสาม เมื่อเทียบกับ Tertile ที่ต่ำที่สุด คือ 1.97 และ 5.04 ตามลำดับ แสดงว่าผู้ป่วยที่คาดการณ์ว่าจะอยู่ในระยะที่สองและสาม ของความเสี่ยงลิ่มเลือดอุดตันในหลอดเลือดดำ จะอยู่ที่ 2 และ 5 เท่าเทียบกับ Tertile ต่ำสุด

สรุปได้ว่า การวิเคราะห์ของแบบจำลองการเรียนรู้ของเครื่องสำหรับการทำนายลิ้ม
เลือดในหลอดเลือดดำในผู้ป่วยที่เป็นโรคเฉียบพลัน แสดงให้เห็นประสิทธิภาพที่ดีขึ้นเมื่อเทียบกับ
คะแนน IMPROVE โดยมีค่า c-statistic ที่สูงขึ้น



บทที่ 3

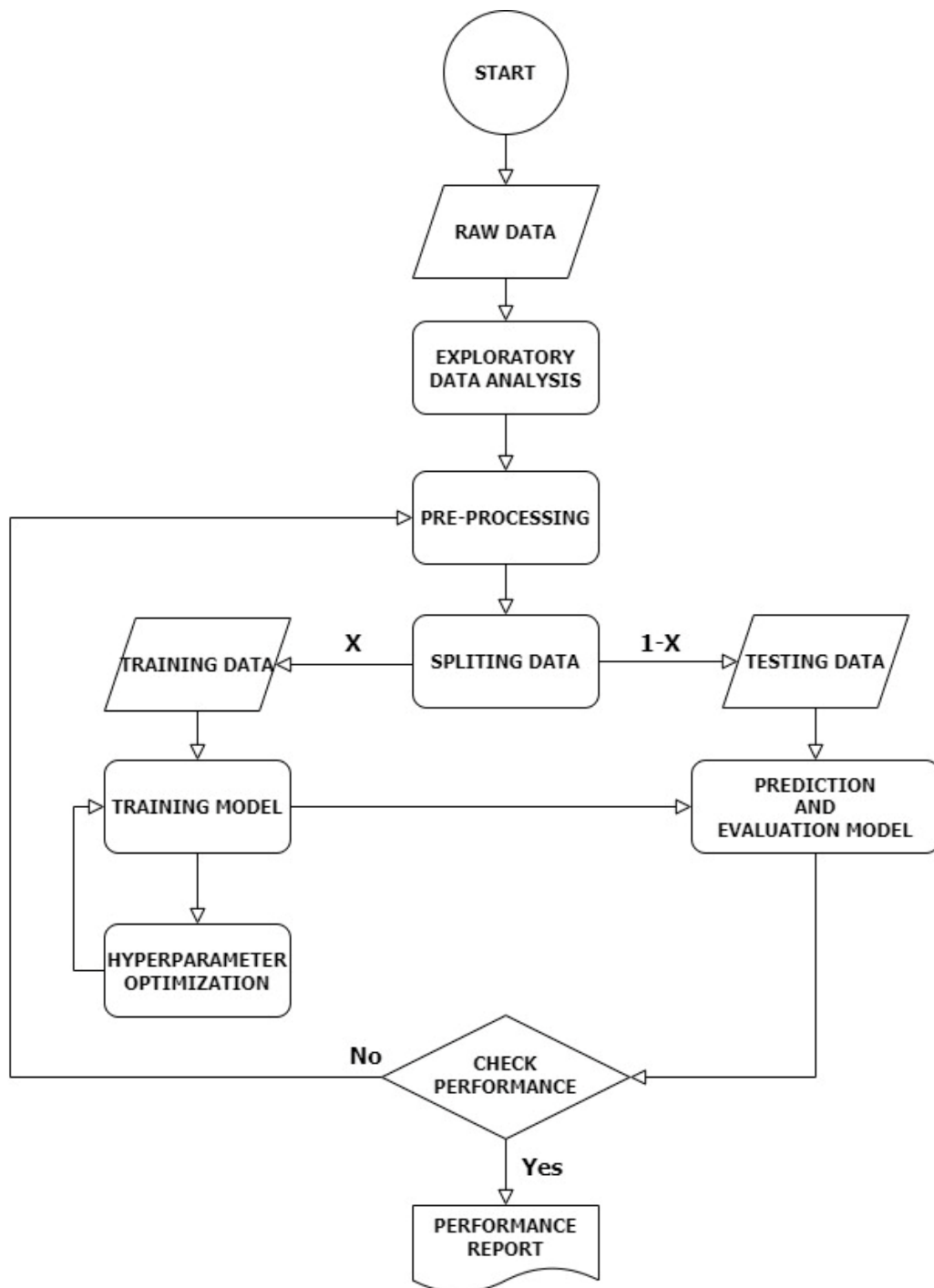
วิธีการดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

- 3.1 กระบวนการทำงานของแบบจำลอง
- 3.2 การรวบรวมข้อมูล (Data Acquisition)
- 3.3 การคัดกรองข้อมูล (Data Filtering)
- 3.4 การเตรียมข้อมูล (Data Preparation)
- 3.5 การสร้างแบบจำลอง (Modeling)
- 3.6 การประเมินผล (Evaluation)

3.1 กระบวนการทำงานของแบบจำลอง

จากรูปกระบวนการทำงานของแบบจำลอง จะอธิบายถึงการทำงานของแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพในการทำงานของแต่ละอัลกอริทึม โดยเริ่มต้นจากการนำเข้าข้อมูล จากนั้นทำการสำรวจข้อมูล (Exploratory Data Analysis: EDA) เพื่อใช้วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรต่างๆ หรือปัจจัยเสี่ยงที่จะทำให้เกิดล้มเหลวจุดเริ่มต้น แล้วทำการตรวจสอบความถูกต้องของข้อมูล การแปลงข้อมูลให้เหมาะสมที่ใช้สำหรับการประมวลผล การแบ่งข้อมูลเพื่อใช้เป็นชุดข้อมูลในการฝึกฝน และชุดข้อมูลในการทดสอบ เมื่อมีความพร้อมก็จะนำข้อมูลเข้าแบบจำลองการเรียนรู้ของเครื่องในการวิเคราะห์ต่อไป หลังจากนั้นปรับค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดที่ทำให้แบบจำลองการเรียนรู้ของเครื่องทำนายผลลัพธ์ได้อย่างมีประสิทธิภาพ และแม่นยำมากที่สุด โดยการสร้างแบบจำลองการเรียนรู้ของเครื่อง ประกอบด้วย 3 อัลกอริทึม คือ AdaBoost, Extreme Gradient Boosting และ Random Forest สุดท้ายทำการประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องทั้ง 3 อัลกอริทึมกับชุดข้อมูลที่ใช้ในการทดสอบว่าแบบจำลองการเรียนรู้ของเครื่องใดที่มีประสิทธิภาพดีที่สุด ดังภาพประกอบ 18



ภาพประกอบ 18 แสดงแผนภาพกระบวนการทำงานของแบบจำลอง

3.2 การเก็บรวบรวมข้อมูล (Data Acquisition)

งานวิจัยนี้เป็นข้อมูลของผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลจุฬาลงกรณ์ ซึ่งข้อมูลดังกล่าวจะอยู่ในไฟล์ดิจิทัลนามสกุล csv ซึ่งแสดงคำอธิบายข้อมูลของแต่ละตัวแปรดังตารางที่ 1

ตาราง 1 แสดงคำอธิบายข้อมูลของแต่ละตัวแปร

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
1	No	เลขที่
2	Number	เลขที่
3	Ward	หอผู้ป่วย
4	GPrivateICU	General ward สามัญ Private ward พิเศษ ICU ward
5	Age	อายุ
6	Date	วันที่เก็บข้อมูล
7	Sex	เพศ
8	Weight	น้ำหนัก
9	Height	ส่วนสูง

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
10	BMI	ดัชนีมวลกาย
11	Race	สัญชาติ
12	Immobilized	การไม่ขยับเคลื่อนไหว
13	CHF	โรคหัวใจวาย
14	Paraparesis	ขาอ่อนแรง 2 ข้าง
15	Cancer	โรคมะเร็ง
16	CAType	ประเภทของมะเร็ง
17	Lymphoma	มะเร็งต่อมน้ำเหลือง
18	Hemato	มะเร็งโลหิตวิทยา

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
19	Stage	ระยะของโรคมะเร็ง
20	Respirator	ใช้เครื่องช่วยหายใจ
21	COPD	โรคปอดอุดกั้นเรื้อรัง
22	CompressionFx	กระดูกสันหลังหักแบบยุบตัว
23	Infection	การติดเชื้อ
24	Hemiparesis	การอ่อนแรงครึ่งซีก
25	BedRidden	ภาวะติดเตียง
26	SLE	โรค Systemic Lupus Erythematosus
27	Arthritis	ข้ออักเสบ
28	Surgery	การผ่าตัด

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
29	SType	ประเภทของการผ่าตัด
30	HxVTE	ประวัติเป็น VTE
31	FamilyVTE	ครอบครัวมีประวัติ VTE
32	Hxcancer	ประวัติเป็นมะเร็ง
33	Varicose	เส้นเลือดขอดที่ขา
34	Estrogen	ได้รับยาฮอร์โมนเพศหญิง เช่น ยาคุมกำเนิด
35	Thrombophilia	ภาวะเลือดแข็งตัวง่าย
36	Ttype	ชนิดของภาวะเลือดแข็งตัวง่าย
37	Platelet	ระดับเกล็ดเลือด
38	Fat	โรคอ้วน

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
39	Nephrotic	โรคไตที่มีโปรตีนรั่ว
40	@4Risk	ปัจจัยที่มีความเสี่ยงมากกว่า 4 ข้อ
41	Prophylaxis	การได้รับยาป้องกัน VTE
42	OtherIndi	การได้รับยาต้านการแข็งตัวของเลือด
43	IndiType	ข้อบ่งชี้ในการได้ยาต้านการแข็งตัวของเลือด
44	LMWH	Low Molecular Weight Heparin
45	UFH	Unfractionated Heparin
46	Warfarin	ยา Warfarin
47	Autoimmune	โรคแพ้ภูมิตัวเอง
48	FinalDx	การวินิจฉัยโรคสุดท้าย

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
49	Comorbid	โรคที่เป็นร่วม
50	Complications	ภาวะแทรกซ้อน
51	Procedure	หัตถการที่ทำ
52	Death	เสียชีวิตหรือไม่
53	DeadDate	วันที่เสียชีวิต
54	CauseofDeath	สาเหตุการเสียชีวิต 1
55	CauseOfDeath1	สาเหตุการเสียชีวิต 2
56	Autopsy	ได้รับการตรวจศพ
57	DischargeDate	วันที่ออกจากโรงพยาบาล
58	LOS	จำนวนวันที่พักในโรงพยาบาล

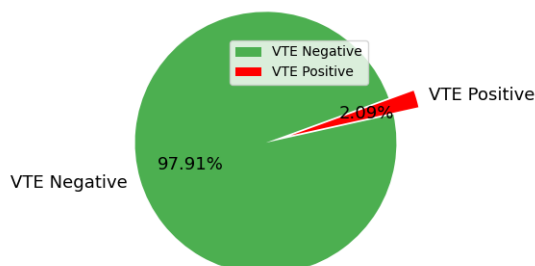
ตาราง 1 (ต่อ)

ลำดับ	ตัวแปรข้อมูล	คำอธิบายข้อมูล
59	DateEval1	วันที่ตรวจหลัง Discharge 1
60	DateEval2	วันที่ตรวจหลัง Discharge 2
61	OPDdeath	เสียชีวิตหลังจากกลับบ้าน
62	OPDdeadDate	วันที่เสียชีวิตหลังจากกลับบ้าน
63	OPDCOD	สาเหตุการเสียชีวิต
64	totalDeath	จำนวนผู้ป่วยที่เสียชีวิตทั้งหมด
65	VTEpositive	ภาวะลิ่มเลือดอุดตัน

3.3 การคัดกรองข้อมูล (Data Filtering)

ข้อมูลในงานวิจัยนี้ แพทย์เป็นผู้เก็บรวบรวมข้อมูลนี้จากผู้ป่วยที่เข้ารับการรักษาในหอผู้ป่วยใน แผนกอายุรศาสตร์ โรงพยาบาลจุฬาลงกรณ์ ปี ค.ศ. 2009 ซึ่งได้รับการอนุมัติจากคณะกรรมการจริยธรรมของคณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยแล้ว ซึ่งข้อมูลนี้จะประกอบด้วยข้อมูลผู้ป่วยทั้งหมด 1,290 รายการ และปัจจัยต่างๆ ที่นำมาใช้พิจารณาเพื่อหาความเสี่ยงของการเกิดภาวะลิ่มเลือดอุดตันในหลอดเลือดดำทั้งหมด 65 คอลัมน์ ซึ่งได้รวมคอลัมน์ที่เป็นคำตอบ (Class Label) ไว้แล้ว

นำข้อมูลทั้งหมดมาวิเคราะห์ออกมาในรูปแบบ Pie Chart พบว่าผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ VTE Positive มีอยู่ 2.09% และผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ VTE Negative มีอยู่ 97.91% ของข้อมูลทั้งหมด ดังภาพประกอบ 19



ภาพประกอบ 19 แสดงจำนวนผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตัน (VTE Positive) และไม่เป็นโรคลิ่มเลือดอุดตัน (VTE Negative)

สำหรับการเตรียมข้อมูลเพื่อนำไปฝึกฝนกับแบบจำลองการเรียนรู้ของเครื่อง ที่ได้มาจากตัวแปรของภาวะการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำนั้น จะมีวิธีการจัดการกับข้อมูลเพื่อนำไปสร้างแบบจำลองการเรียนรู้ของเครื่อง ดังต่อไปนี้

3.4.1 การจัดการค่าว่าง (Missing Values)

ข้อมูลที่เป็นค่าว่าง จะจัดการกับข้อมูลที่เป็นค่าว่างเหล่านี้ด้วยหลักการเติมศูนย์ (Fill missing values with zero) เนื่องจากข้อมูลทางการแพทย์ไม่สามารถระบุข้อมูลแทนกันได้ จึงจำเป็นต้องใช้หลักการเติมศูนย์แทนที่ค่าว่าง

3.4.2 การแปลงข้อมูล (Encoding Values)

เนื่องจากข้อมูลที่นำมาวิเคราะห์นั้น มีตัวแปรบางตัวที่มีชนิดของข้อมูลเป็นตัวอักษร ซึ่งจะไม่มีประสิทธิภาพในการวิเคราะห์ข้อมูลได้อย่างแม่นยำ ดังนั้นผู้วิจัยจึงจำเป็นต้องแปลงข้อมูลชนิดดังกล่าวให้เป็นตัวเลข ด้วยวิธีการแปลงข้อมูลแบบ One Hot Encoding ซึ่งได้ผลลัพธ์ดังภาพประกอบ 20

Color	Red	Yellow	Green
Red	1	0	0
Red	1	0	0
Yellow	0	1	0
Green	0	0	1
Yellow	0	1	0

ภาพประกอบ 20 แสดงผลลัพธ์การแปลงข้อมูลให้เป็นตัวเลข ด้วย One Hot Encoding

ที่มา: (Praharsa)

3.4.3 การคัดเลือกคุณลักษณะ (Feature Selection)

ในการจำแนกประเภทข้อมูล (Classification) จะพบว่าแอตทริบิวต์ (Attribute) หรือคุณลักษณะ (Feature) มีจำนวนมาก และข้อความต่างๆ ที่ใช้เป็นคุณลักษณะก็มีเป็นจำนวนมาก บางตัวก็ไม่ได้มีความสำคัญในการแบ่งแยกคลาส (Class) ดังนั้นจึงจำเป็นต้องทำการคัดเลือกคุณลักษณะที่สำคัญมาใช้งาน ขั้นตอนนี้เรียกว่า Feature Selection ซึ่งสามารถแบ่งได้เป็น 3 กลุ่มใหญ่ดังนี้

1. **Filter Approach** เป็นการคัดเลือกคุณลักษณะโดยการคำนวณหาค่าน้ำหนัก ซึ่งอาจจะเป็นค่าความสัมพันธ์ระหว่างแต่ละคุณลักษณะกับคลาสต่างๆ โดยคุณลักษณะที่ได้จะเรียงลำดับตามค่าความสำคัญที่คำนวณได้ของแต่ละคุณลักษณะ คุณลักษณะที่มีค่าความสำคัญมากจะถูกนำมาใช้งานต่อไป เทคนิคที่ใช้ในการคำนวณหาค่าความสำคัญของคุณลักษณะต่างๆ มีหลายวิธี เช่น Information Gain, Chi-Square หรือ Correlation ซึ่งวิธีการ Filter นี้จะแตกต่างจากวิธีการ Wrapper ที่ไม่มีการสร้างโมเดลในการคัดเลือกคุณลักษณะ

ข้อดีของวิธีนี้ คือ คำนวณง่าย รวดเร็ว หลีกเลี่ยงการเกิดโอเวอร์ฟิตติ้ง (Overfitting) เพราะวิธีนี้ไม่นำผลทดสอบประสิทธิภาพการเรียนรู้ของเครื่องมาพิจารณาไปด้วย ซึ่งคุณลักษณะที่ถูกคัดเลือกจะไม่ถูกไบแอส (Bias) สำหรับข้อเสียของวิธีนี้ คือ คุณลักษณะที่ถูกคัดเลือกจะเป็นคุณลักษณะที่เป็นอิสระต่อกัน เพราะขั้นตอนการคำนวณค่าความสำคัญจะพิจารณาความสัมพันธ์ระหว่างคุณลักษณะนั้นๆ กับข้อมูลเอาท์พุตเท่านั้น ไม่ได้พิจารณาความสัมพันธ์ระหว่างคุณลักษณะกันเอง ดังนั้นคุณลักษณะที่ถูกเลือกนั้นอาจมีความสัมพันธ์กัน หรือไม่มีความสัมพันธ์

กันก็ได้ จึงส่งผลทำให้ค่าความถูกต้องของการเรียนรู้ลดลง ดังนั้นวิธีการนี้จึงเหมาะสมกับการวิเคราะห์ข้อมูลที่มีจำนวนมิติไม่เยอะมาก (Pimpim, 2019)

โดยในงานวิจัยนี้จะใช้วิธี Threshold & Correlation และ Mutual Information Classification เป็นวิธีการคัดเลือกคุณลักษณะใน Filter Approach

Threshold & Correlation ในงานวิจัยนี้จะคัดเลือกคุณลักษณะโดยดูจากจำนวนของข้อมูลในแต่ละคุณลักษณะที่มีค่ามากกว่า 10% ขึ้นไป โดยผลลัพธ์ที่ได้จะนำมาสร้าง Correlation เพื่อตรวจสอบความสัมพันธ์ระหว่างตัวแปรตามตัวหนึ่งกับตัวแปรอิสระหลายๆ ตัว ว่าตัวแปรอิสระเหล่านี้จะต้องไม่มีความสัมพันธ์กัน หรือหากมีความสัมพันธ์กันก็ต้องมีความสัมพันธ์กันไม่สูงมากนัก แต่ในทางปฏิบัตินั้น บางครั้งจะพบว่าตัวแปรอิสระมีความสัมพันธ์กันสูง ในกรณีที่ตัวแปรอิสระเพียง 2 ตัว มีความสัมพันธ์กันสูงจะเรียกว่า Collinearity และในกรณีที่ตัวแปรอิสระมากกว่า 2 ตัว มีความสัมพันธ์กันสูงจะเรียกว่า Multicollinearity ซึ่งทั้งสองเหตุการณ์อาจส่งผลให้แบบจำลองที่ใช้ในการพยากรณ์ตัวแปรมีความคลาดเคลื่อนได้ ดังนั้นในการแก้ปัญหา Collinearity นั้น ผู้วิจัยได้พิจารณาขจัดตัวแปรออกหนึ่งตัว เนื่องจากตัวแปรดังกล่าวสามารถใช้ทดแทนกับตัวแปรอีกตัวหนึ่งได้

สำหรับวิธี Mutual Information Classification นั้น เป็นวิธีการวัดโดยใช้ Entropy ภายใต้เงื่อนไขของค่าเป้าหมาย (Target Value) ซึ่งคะแนน (Score) ที่ได้นั้นจะอยู่ในช่วง 0 ถึง 1 ถ้าคะแนนมีค่าสูงๆ หมายถึงว่าตัวแปรนั้นมีความสัมพันธ์กันสูงกับตัวแปรเป้าหมาย ซึ่งเหมาะที่จะนำมาใช้ในเป็นปัจจัยในการวิเคราะห์ของข้อมูล โดยจะเลือกคุณลักษณะที่มีค่าพารามิเตอร์ของ Score มากกว่า 0.002 ขึ้นไป

2. Wrapper Approach เป็นวิธีที่ถูกพัฒนามาเพื่อแก้ไขวิธี Filter เป็นการคัดเลือกคุณลักษณะด้วยการสร้างโมเดล (Classification Model) ขึ้นมา และวัดประสิทธิภาพการทำงานของโมเดลที่ทำให้โมเดลมีประสิทธิภาพมากที่สุดมาใช้งาน เช่น โมเดลที่ให้ค่าความถูกต้อง (Accuracy) มากที่สุด การคัดเลือกคุณลักษณะด้วยวิธีการนี้แบ่งย่อยได้เป็น 2 แบบคือ

Forward Selection เป็นการสร้างโมเดลโดยการเพิ่มทีละ 1 คุณลักษณะเข้าไปในโมเดล ถ้าคุณลักษณะที่ใส่นั้น สามารถเพิ่มให้ประสิทธิภาพที่ดีได้ก็จะเก็บไว้ และเลือกคุณลักษณะตัวอื่นๆ มาเพิ่มไปเรื่อยๆ จนกว่าประสิทธิภาพของโมเดลไม่ได้ดีขึ้นจึงจะหยุดการทำงาน

Backward Elimination เป็นการสร้างโมเดลโดยใช้คุณลักษณะที่มีอยู่ทั้งหมดก่อน หลังจากนั้นจะค่อยๆ ตัด (Eliminate) คุณลักษณะที่ไม่สำคัญทิ้งไปทีละคุณลักษณะ ถ้าประสิทธิภาพดีขึ้นก็ตัดคุณลักษณะอื่นๆ ต่อไป

ที่มา: (Eakasit Pacharawongsakda, 2015)

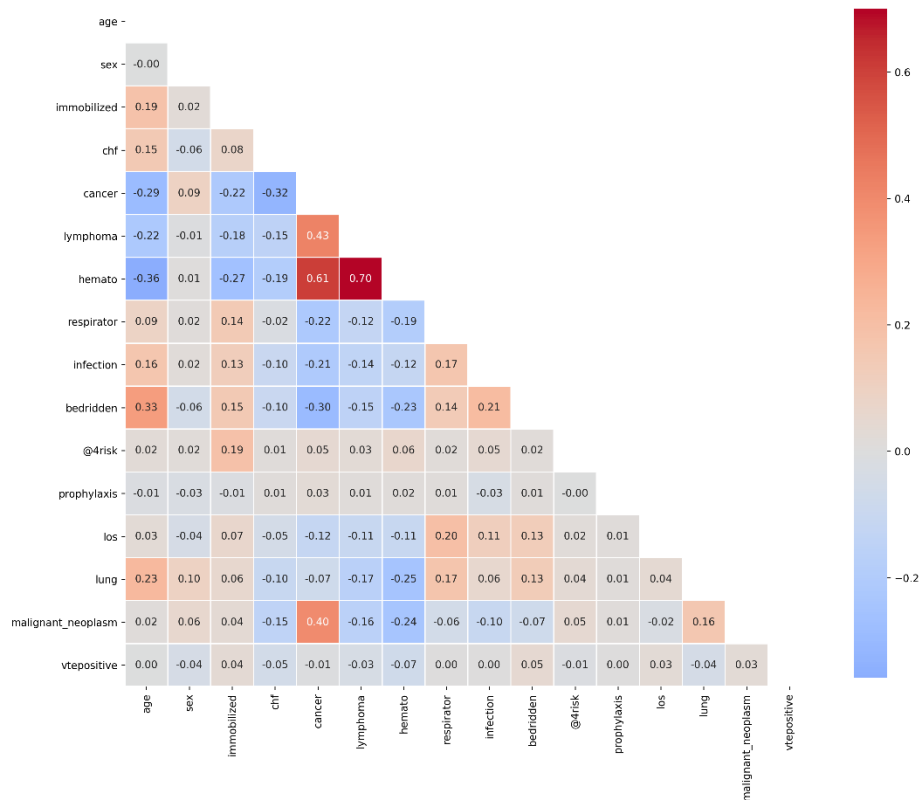
3. Embedded Approach วิธีเลือกคุณลักษณะวิธีการฝังตัวเป็นวิธีถูกออกแบบมาเพื่อแก้ไขข้อเสียของวิธี Filter และวิธี Wrapper ที่ใช้เวลาในการประมวลผลที่นานกว่า การเลือกคุณลักษณะของวิธีนี้มีข้อดีคือ มีการค้นหาเซตของคุณลักษณะทั้ง Global space และ Local space จึงทำให้มีประสิทธิภาพในการค้นหา แต่อย่างไรก็ตามวิธีฝังตัวมีข้อเสีย คือ การเลือกเซตคุณลักษณะที่ไม่มีความยืดหยุ่น เนื่องจากขึ้นอยู่กับอัลกอริทึมในการจำแนกกลุ่มของข้อมูล

ที่มา: (Pimpim, 2019)

โดยในงานวิจัยนี้จะใช้วิธี Decision Tree เป็นวิธีการคัดเลือกคุณลักษณะใน Embedded Approach โดยจะเลือกคุณลักษณะที่มีค่าพารามิเตอร์ของ Importance มากกว่า 0.001 ขึ้นไป

3.4.4 การสำรวจข้อมูล (Exploratory Data Analysis: EDA)

เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ว่ามีความสัมพันธ์กันมากน้อยเพียงใด และมีความสัมพันธ์กันในทิศทางใดของข้อมูลที่เป็นปัจจัยทำให้เกิดภาวะล้มเลือดอุดตันในหลอดเลือดดำทั้งหมด โดยใช้ Heatmap แสดงผล ดังภาพประกอบ 21



ภาพประกอบ 21 ตัวอย่างการแสดงความสัมพันธ์ของปัจจัยเสี่ยงที่ทำให้เกิดภาวะลิ่มเลือดอุดตัน
ในหลอดเลือดดำของการใช้ Feature Selection แบบ Threshold & Correlation

3.4.5 การแบ่งข้อมูลสำหรับการฝึกฝน และทดสอบ (Data Split)

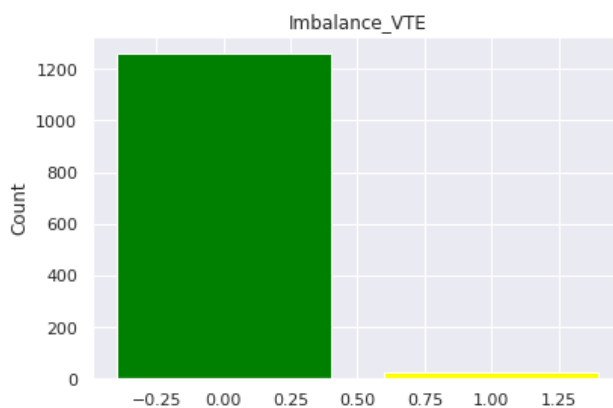
เมื่อจัดการกับข้อมูลเสร็จเรียบร้อยแล้ว และพร้อมที่จะนำไปสร้างแบบจำลองการเรียนรู้ของเครื่อง จะต้องแบ่งข้อมูลออกเป็น ข้อมูลที่ใช้สำหรับการฝึกฝน (Training Data) และข้อมูลที่ใช้สำหรับการทดสอบ (Testing Data) โดยจะแบ่งข้อมูลด้วยอัตราส่วน 75:25, 70:30 และ 67:33 ซึ่งข้อมูลที่ใช้สำหรับการทดสอบที่ได้ในแต่ละชุดข้อมูลนั้น จะถูกใช้ทดสอบกับแบบจำลองทั้งหมดด้วยชุดข้อมูลเดียวกัน และใช้เพียงครั้งเดียว

3.4.6 การจัดการกับข้อมูลที่ไม่สมดุลกัน (Handling Imbalanced Data)

เนื่องจากข้อมูลของงานวิจัยนี้เป็นแบบ Binary Classification ซึ่งจะใช้จำแนกโรคลิ่มเลือดอุดตันในหลอดเลือดดำ เป็น 2 ประเภท โดยจัดกลุ่มได้เป็นดังนี้ (ภาพประกอบ 22)

0 คือ ผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ (VTE Positive)

1 คือ ผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำ (VTE Negative)



ภาพประกอบ 22 แสดงข้อมูลของทั้งสองคลาสที่ไม่สมดุลกัน (Imbalance)

ที่มา: (Sukperm et al., 2021)

ในกรณีที่เกิดปัญหาของข้อมูลที่ไม่สมดุลกัน (Imbalanced Data) คือ ผู้ป่วยที่ไม่เป็นโรคล้มเลือดอุดตันในหลอดเลือดดำมีจำนวนมากกว่าผู้ป่วยที่เป็นโรคล้มเลือดอุดตันในหลอดเลือดดำอยู่เป็นจำนวนมาก จะทำให้เกิดการ Bias ขึ้นกับข้อมูล และทำให้แบบจำลองการเรียนรู้ของเครื่องไม่มีความแม่นยำหรืออาจจะทำให้แบบจำลองการเรียนรู้ของเครื่องทำนายคลาดเคลื่อนได้

ดังนั้นจะใช้วิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูล ทั้งหมด 5 วิธี ดังต่อไปนี้

1. **วิธีการสุ่มเกิน (Oversampling)** เป็นวิธีการเพิ่มจำนวนข้อมูลของกลุ่มข้อมูลที่มีจำนวนน้อยกว่า ให้มีจำนวนใกล้เคียงหรือเท่ากับกลุ่มข้อมูลที่มีจำนวนมากกว่า

ตัวอย่าง วิธีทำการสุ่มเกิน (Oversampling) กับข้อมูลที่แบ่งสำหรับการฝึกฝน และทดสอบในอัตราส่วน 70:30 แล้ว จะได้ข้อมูลที่ใช้ในการฝึกฝนทั้งหมด 903 ซึ่งจะแสดงจำนวนก่อนและหลังการทำวิธีการสุ่มเกิน (Oversampling) ดังตาราง 2

ตาราง 2 แสดงจำนวนข้อมูลก่อนและหลังทำ Oversampling

Oversampling	VTE Negative (0)	VTE Positive (1)
จำนวนก่อนทำ	903	19
จำนวนหลังทำ	903	903

ในงานวิจัยนี้ยังมีการเรียกใช้พารามิเตอร์ `sampling_strategy` ในวิธีการสุ่มเกิน (Oversampling) โดยจะใช้ค่าเท่ากับ 0.5

ตัวอย่าง วิธีทำการสุ่มเกิน (Oversampling) ด้วยพารามิเตอร์ `sampling_strategy` เท่ากับ 0.5 กับข้อมูลที่แบ่งสำหรับการฝึกฝน และทดสอบในอัตราส่วน 70:30 แล้ว จะได้ข้อมูลที่ใช้ในการฝึกฝนทั้งหมด 903 ซึ่งจะแสดงจำนวนก่อนและหลังการทำวิธีการสุ่มเกิน (Oversampling) ด้วยพารามิเตอร์ `sampling_strategy` เท่ากับ 0.5 ดังตาราง 3

ตาราง 3 แสดงจำนวนข้อมูลก่อนและหลังทำ Oversampling กับ `sampling_strategy` เท่ากับ 0.5

Oversampling with <code>sampling_strategy = 0.5</code>	VTE Negative (0)	VTE Positive (1)
จำนวนก่อนทำ	903	19
จำนวนหลังทำ	903	451

2. วิธีการสุ่มลด (Undersampling) เป็นวิธีการลดจำนวนข้อมูลของกลุ่มข้อมูลที่มีจำนวนมากกว่า ให้มีจำนวนใกล้เคียงหรือเท่ากับกลุ่มข้อมูลที่มีจำนวนน้อยกว่า

ตัวอย่าง วิธีทำการสุ่มลด (Undersampling) กับข้อมูลที่แบ่งสำหรับการฝึกฝน และทดสอบในอัตราส่วน 70:30 แล้ว จะได้ข้อมูลที่ใช้ในการฝึกฝนทั้งหมด 19 ซึ่งจะแสดงจำนวนก่อนและหลังการทำวิธีการสุ่มลด (Undersampling) ดังตาราง 4

ตาราง 4 แสดงจำนวนข้อมูลก่อนและหลังทำ Undersampling

Undersampling	VTE Negative (0)	VTE Positive (1)
จำนวนก่อนทำ	903	19
จำนวนหลังทำ	19	19

3. วิธีสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) เป็นเทคนิคการสุ่มตัวอย่างแบบสุ่มเพิ่ม แทนที่จะสุ่มเพิ่มจากข้อมูลเดิม แต่จะทำการสังเคราะห์ข้อมูลขึ้นมาใหม่จากข้อมูลเดิม ซึ่งมีหลักการสุ่มคือ เป็นการเพิ่มจำนวนข้อมูลกลุ่มน้อยให้มีจำนวนเพิ่มขึ้น (Chawla, 2002) โดยการเพิ่มข้อมูลในกลุ่มน้อยนั้นทำให้การกระจายของกลุ่มข้อมูลมีความสมดุลมากขึ้น โดยทำการสุ่มค่าข้อมูลที่อยู่ในกลุ่มข้อมูลน้อยขึ้นมา 1 ค่า หลังจากนั้นพิจารณาค่าข้อมูลใกล้เคียงอีกจำนวน K ค่า (K-nearest neighbor) แล้วคำนวณค่าระยะทาง (Euclidean distance) ระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียงแต่ละค่า เพื่อหาค่าระยะทางที่น้อยที่สุด ระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียง จากนั้นจึงสร้าง ข้อมูลเทียมระหว่างค่าข้อมูลที่สุ่มกับค่าข้อมูลใกล้เคียงตัวที่ให้ค่าระยะทางที่น้อยที่สุด (Haibo & Garcia, 2009)

ตัวอย่าง วิธีทำการสังเคราะห์ข้อมูลเพิ่ม (SMOTE) กับข้อมูลที่แบ่งสำหรับการฝึกฝน และทดสอบในอัตราส่วน 70:30 แล้ว จะได้ข้อมูลที่ใช้ในการฝึกฝนทั้งหมด 903 ซึ่งจะแสดงจำนวนก่อนและหลังการทำวิธีสังเคราะห์ข้อมูลเพิ่ม (SMOTE) ดังตาราง 5

ตาราง 5 แสดงจำนวนข้อมูลก่อนและหลังทำ SMOTE

SMOTE	VTE Negative (0)	VTE Positive (1)
จำนวนก่อนทำ	903	19
จำนวนหลังทำ	903	903

4. **วิธี Class Weight** เป็นวิธีการให้น้ำหนักในแต่ละตัวแปร โดยจะให้ความสำคัญมากกับกลุ่มข้อมูลที่มีจำนวนน้อยกว่ากลุ่มข้อมูลที่มีจำนวนมาก เพื่อจะให้ผลลัพธ์ที่ได้มานั้นมีความเท่าเทียมกันทุกกลุ่ม

5. **วิธี Ensemble Sampling** เป็นวิธีที่ใช้ Balanced Bagging Classifier คือ ความเป็นไปได้ของการสุ่มตัวอย่างของกลุ่มข้อมูลสำหรับการฝึกฝน (Training Data) ที่จะถูกทำให้กระจายไปตามกลุ่มที่อยู่ใกล้เคียงกัน ทำให้ข้อมูลมีความสมดุลขึ้น

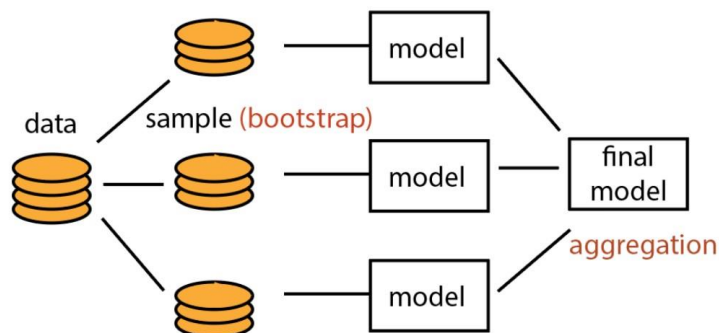
3.5 การสร้างแบบจำลอง (Modeling)

ในงานวิจัยนี้ ขอเสนอวิธีการสร้างแบบจำลองโดยการใช้การเรียนรู้ของเครื่องหลายๆ ครั้ง (Ensemble Learning) เพื่อเพิ่มประสิทธิภาพของแบบจำลอง ซึ่งเทคนิคที่ใช้ มีรายละเอียดดังนี้

3.5.1 อัลกอริทึม Random Forest

เป็นการสร้างแบบจำลองออกมาหลายๆ แบบจำลอง โดยใช้การสุ่มข้อมูลตัวอย่างจากข้อมูลฝึกสอนออกมาเป็นหลายๆ ชุด สำหรับวิธีการสุ่มข้อมูลออกมาเป็นวิธีสุ่มแบบแทนที่ (Random with Replacement) หมายความว่าข้อมูลที่เราใช้ยังคงเดิมไม่ได้ลดลงหลังจากการสุ่ม ซึ่งสามารถสุ่มข้อมูลหลายๆ รอบเพื่อให้ได้ Classifier หลายๆ ตัว เวลาทำนายจะใช้ Classifier ทุกตัวที่สร้างขึ้นมาเพื่อทำนายชุดข้อมูลใหม่ที่พบ โดยลักษณะการทำงานนี้เราเรียกว่า Bagging หรือ Bootstrap Aggregation (Mahapatra, 2022) ดังภาพประกอบ 23

Bagging (Bootstrap aggregation)



ภาพประกอบ 23 แสดงลักษณะการทำงานของ Bagging

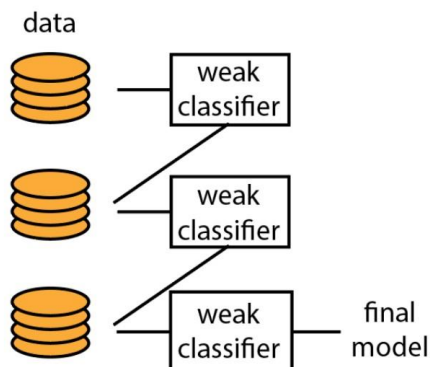
ที่มา: (TITIPATA)

Bagging เป็นพื้นฐานของ Random Forest Classifier ซึ่งเป็นแบบจำลองที่นำต้นไม้ตัดสินใจ (Decision Tree) หลายๆ ต้นมาฝึกสอนร่วมกัน โดยมีจำนวนตั้งแต่ 10 เป็นต้นไป โดยที่ต้นไม้ตัดสินใจ (Decision Tree) แต่ละต้นจะได้รับข้อมูลที่เป็น Subset ของ Feature และข้อมูลเป็นแบบสุ่มทั้งหมด จากนั้นต้นไม้ตัดสินใจ (Decision Tree) แต่ละต้นจะทำนายและเลือกคำตอบจากค่าที่ทำนายที่ได้รับความนิยมมากที่สุด

3.5.2 อัลกอริทึม AdaBoost

เป็นการนำ Classifier ที่มีความแม่นยำต่ำ (Weak Classifier) มาทำนายข้อมูล จากนั้นจะให้ Classifier ที่มีความแม่นยำต่ำตัวใหม่มาแก้ไข Error โดยผลรวมของ Classifier จะเกิดเป็น Classifier ใหม่ขึ้น และจะทำแบบนี้ไปเรื่อยๆ จนแบบจำลองที่ได้ไม่มีค่าคลาดเคลื่อนเกิดขึ้น ซึ่งจะได้แบบจำลองที่ดีที่สุด โดยลักษณะการทำงานนี้จะเรียกว่า Boosting ดังภาพประกอบ 24

Boosting



ภาพประกอบ 24 แสดงลักษณะการทำงานของ Boosting

ที่มา: (TITIPATA)

3.5.3 อัลกอริทึม Extreme Gradient Boosting

เป็นเทคนิคที่พัฒนามาจาก Gradient Boosting ซึ่ง Extreme Gradient Boosting (XGBoost) เป็นแบบจำลองที่นำเอาต้นไม้ตัดสินใจ (Decision Tree) มาฝึกสอนต่อกันหลายๆ ต้น โดยที่ต้นไม้ตัดสินใจ (Decision Tree) แต่ละต้นจะเรียนรู้ความผิดพลาดของต้นก่อนหน้า ซึ่งทำให้ความแม่นยำในการทำนายดีขึ้นเรื่อยๆ จนเมื่อการเรียนรู้ของต้นไม้ตัดสินใจ (Decision Tree) มีความลึกมากพอ แบบจำลองจะหยุดเรียนรู้จนไม่เหลือความผิดพลาดจากต้นไม้ตัดสินใจ (Decision Tree) ต้นก่อนหน้าแล้ว (Tseng, 2018)

3.6 การประเมินผล (Evaluation)

งานวิจัยนี้เป็นการทำนายผลลัพธ์ที่เป็น Classification สำหรับทางการแพทย์ จึงจำเป็นต้องมีการประเมินผลที่แม่นยำที่สุด โดยผู้วิจัยจะใช้ตาราง Confusion Matrix สำหรับในการประเมินผลของแบบจำลองการเรียนรู้ของเครื่อง ดังภาพประกอบ 25

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

ภาพประกอบ 25 Confusion Matrix

ที่มา: (boom626, 2019)

การแปลความหมายของ Confusion Matrix ในงานวิจัยนี้ มีความหมายดังต่อไปนี้

True Positive (TP) คือ ผลตรวจทางการแพทย์พบว่า เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ และการเรียนรู้ของเครื่อง (Machine Learning) ทำนายผลว่า เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ ซึ่งสรุปว่า ทำนายถูกต้อง

True Negative (TN) คือ ผลตรวจทางการแพทย์พบว่า ไม่เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ และการเรียนรู้ของเครื่อง (Machine Learning) ทำนายผลว่า ไม่เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ ซึ่งสรุปว่า ทำนายถูกต้อง

False Positive (FP) คือ ผลตรวจทางการแพทย์พบว่า ไม่เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ แต่การเรียนรู้ของเครื่อง (Machine Learning) ทำนายผลว่า เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ ซึ่งสรุปว่า ทำนายผิด

False Negative (FN) คือ ผลตรวจทางการแพทย์พบว่า เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ แต่การเรียนรู้ของเครื่อง (Machine Learning) ทำนายผลว่า ไม่เป็นโรคลิ้มเลือดอุดตันในหลอดเลือดดำ ซึ่งสรุปว่า ทำนายผิด

3.6.1 Accuracy (ความถูกต้อง)

Accuracy คือ เปอร์เซนต์ความถูกต้องของจำนวนที่ทำนายถูกต้อง

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

3.6.2 Sensitivity (ความไว)

เป็นเครื่องมือที่เหมาะสมสำหรับคัดกรองโรค (Screening Test) หรืออาจกล่าวได้ว่า สำหรับเป็นเครื่องมือ Rule Out โรคที่สงสัย เมื่อผลตรวจเป็นลบ เนื่องจากเมื่อ Sensitivity สูงจะทำให้ค่า False Negative Rate (FNR) ต่ำ ดังนั้น เมื่อผลการทดสอบออกมาเป็นลบ ก็มักจะเป็นลบจริงๆ หรือกล่าวได้ว่า มีโอกาสเป็นโรคต่ำมาก ตัวอย่างเช่น ถ้าผู้ป่วยตรวจ Pap smear แล้วได้ผลเป็นลบ ดังนั้นจะหมายความว่า ผู้ป่วยมีโอกาสที่จะเป็นโรคมะเร็งปากมดลูกน้อยมาก

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3.6.3 Specificity (ความจำเพาะ)

เป็นเครื่องมือที่เหมาะสมสำหรับวินิจฉัยโรค (Diagnostic Test) หรืออาจกล่าวได้ว่า เป็นเครื่องมือ Rule In โรคที่สงสัย เมื่อผลตรวจเป็นบวก เนื่องจากเมื่อ Specificity สูงจะทำให้ค่า False Positive Rate (FPR) ต่ำ ดังนั้น เมื่อผลการทดสอบออกมาเป็นบวก ก็มักจะเป็นบวกจริง หรือกล่าวได้ว่า มีโอกาสเป็นโรคจริงสูงมาก ตัวอย่างเช่น ถ้าผู้ป่วยตรวจเพื่อหาลิ้มเลือดอุดตัน แล้วได้ผลเป็นบวก ดังนั้นจะหมายความว่า ผู้ป่วยมีโอกาสที่จะเป็นลิ้มเลือดอุดตันสูงมาก

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

จากการอ่านตาราง Confusion Matrix สามารถสรุปได้ว่า ถ้า True Negative (TN) และ True Positive (TP) มีค่าสูงยิ่งดี ถือว่าแบบจำลองการเรียนรู้ของเครื่องทำนายถูกต้องและแม่นยำ ส่วน False Negative (FN) และ False Positive (FP) คือความผิดพลาดในการทำนาย จึงควรมีค่าต่ำๆ ดังนั้นการวัดผลในทางการแพทย์ในทางปฏิบัติควรเลือกใช้การตรวจที่มีค่า Sensitivity หรือ Specificity สูงๆ ในการตรวจคัดกรองผู้ป่วยเพื่อทำการรักษา

บทที่ 4

ผลดำเนินการวิจัย

งานวิจัยนี้ คือ การวิจัยภาวะล้มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง สร้างขึ้นเพื่อจำลองการวิจัยภาวะการเกิดล้มเลือดอุดตันในหลอดเลือดดำโดยอาศัยหลักการเรียนรู้ของเครื่อง โดยผู้วิจัยได้ดำเนินการวิจัยโดยการศึกษาและค้นคว้าในขั้นตอนและกระบวนการต่างๆ เพื่อใช้ในการสำรวจข้อมูล การเตรียมข้อมูล เพื่อใช้สำหรับการสร้างแบบจำลอง โดยการแบ่งข้อมูลสำหรับการฝึกฝน (Training data) และข้อมูลสำหรับทดสอบ (Testing data) ด้วย 3 อัตราส่วนได้แก่ 75:25, 70:30 และ 67:33 ของจำนวนข้อมูลทั้งหมดจำนวน 1,290 รายการ ซึ่งจะได้อัตราส่วนของข้อมูลที่แบ่งเป็น 967:323, 903:387 และ 864:426 ตามลำดับ

จากนั้นนำข้อมูลที่ได้ไปสร้างข้อมูลเพื่อให้เกิดความสมดุลของ Class Label โดยมีวิธีทำทั้งหมด 4 วิธี ได้แก่ Undersampling, Oversampling และ Synthetic Minority Oversampling TEchnique (SMOTE) สำหรับ Oversampling นั้น นอกจากที่ทดสอบด้วยค่ามาตรฐาน หรือการใช้พารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 1 แล้ว ยังได้ทำการทดสอบเพิ่มเติมด้วยการใช้พารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.50 (Over with Half) ด้วย

นอกจากนี้ในอัลกอริทึมของ XGBoost มีพารามิเตอร์ 2 ตัวที่ใช้สร้างความสมดุลของข้อมูลคือ `scale_pos_weight` (scaled_pos_weight) และ `sample_weight` (sample_weight) และในอัลกอริทึมของ Random Forest ยังมีอีก 2 พารามิเตอร์ที่ใช้สร้างความสมดุลของข้อมูลเช่นกัน คือ `balanced_class_weight` และ `ratio_class_weight`

สำหรับการคัดเลือกคุณลักษณะ (Feature Selection) ที่ใช้งานวิจัยนี้มีทั้งหมด 3 วิธี คือ Threshold & Correlation, Mutual Information Classification และ Decision Tree ซึ่งผลลัพธ์ที่ได้มีดังต่อไปนี้

คุณลักษณะที่ได้จากวิธี Threshold & Correlation มีทั้งหมด 15 ตัว ประกอบด้วย age, sex, immobilized, chf, cancer, lymphoma, hemato, respirator, infection, bedridden, @4risk, prophylaxis, los, lung และ malignant_neoplasm

คุณลักษณะที่ได้จากวิธี Mutual Information Classification มีทั้งหมด 112 ตัว ประกอบด้วย abdomen, abscess of liver, abscess, acidosis, acute bronchitis, acute cholecystitis, acute cystitis, acute gastritis, acute pancreatitis, acute pulmonary, acute respiratory failure, adjustment disorders, adrenal gland, age, alcohol acute intoxication,

alcohol cirrhosis, alcohol withdrawal, alcoholic, alcoholic cardiomyopathy, alveolar hypoventilation, alzheimer, amyloidosis, anaemia, anal canal, ankle, anterior mediastinum, anterior wall, antidiuretic hormone, aortic valve, arf, arthritis, asthma, asthmatic, atopic dermatitis, atrial fibrillation, autoimmune, autopsy, avian influenza, bacteria, bacterial intestinal infection, bacterial peritonitis, bacterial pneumonia, b-all, bedridden, behavioral disorder, bile duct, biliary, bladder, bleeding, blood, body, bone, bone marrow, brain, breast, bronchiectasis, bronchiolitis, bronchopneumonia, bronchus, burkitt, caecum, calculus of bile duct, calculus of gallbladder, cancer, carbuncle, cardiac arrest, cardiomyopathy, cellulitis, central nervous, cerebellum, cerebral, cerebral meninges, cerebrospinal fluid leakage, cerebrovascular, cervical port, chf, cholangitis, cholestatic jaundice, chronic bronchiectasis, chronic cholecystitis, chronic myeloproliferative, chronic red cell, chronic respiratory failure, compressionfx, copd, diverticular disease, estrogen, familyvte, fat, hemato, hemiparesis, hxcancer, hxvte, immobilized, infection, infective myositis, intrahepatic bile duct, los, lymphoma, nephrotic, niddm, paraparesis, platelet, prophylaxis, respirator, sex, sle, stage, surgery, thrombophilia, varicose และ @4risk

คุณลักษณะที่ได้จากวิธี Decision Tree มีทั้งหมด 10 ตัว ประกอบด้วย adeno ca, burkitt s lymphoma, heart, myeloid, bedridden, stomach, multiple myeloma, pneumonia, nhl และ nephritis

ในการประเมินประสิทธิภาพเพื่อเปรียบเทียบแบบจำลองของทั้ง 3 อัลกอริทึม AdaBoost, Random Forest และ XGBoost ซึ่งแต่ละอัลกอริทึมจะมีการใช้พารามิเตอร์ที่แตกต่างกัน รวมถึงยังมีการใช้ Balanced Bagging Classifier ร่วมกับทั้ง 3 อัลกอริทึม จะประเมินประสิทธิภาพโดยการใช้ตาราง Confusion Matrix เพื่อให้บรรลุวัตถุประสงค์ของการทำวิจัยที่ได้กำหนดไว้ ผลลัพธ์ที่ได้อธิบายไว้ ดังภาพประกอบ 26 - 385

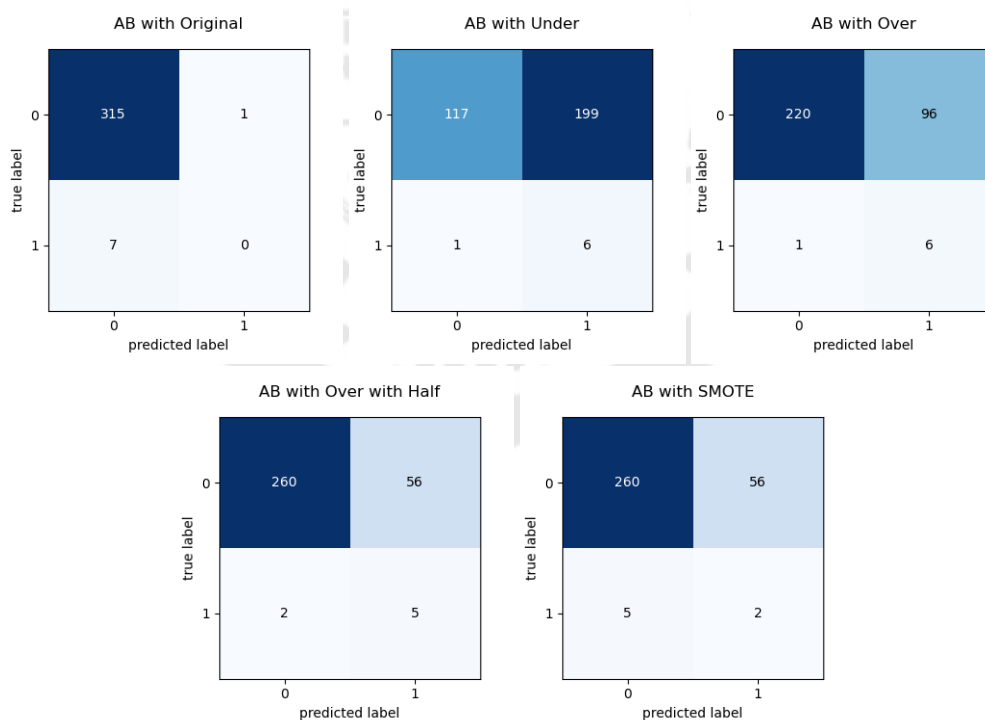
4.1 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Default Parameters

4.1.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.1.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.3808	0.3703	0.8571	0.0293	0.0566
Over	0.6997	0.6962	0.8571	0.0588	0.1101
Over with Half	0.8204	0.8228	0.7143	0.0820	0.1471
SMOTE	0.8111	0.8228	0.2857	0.0345	0.0615

ภาพประกอบ 26 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

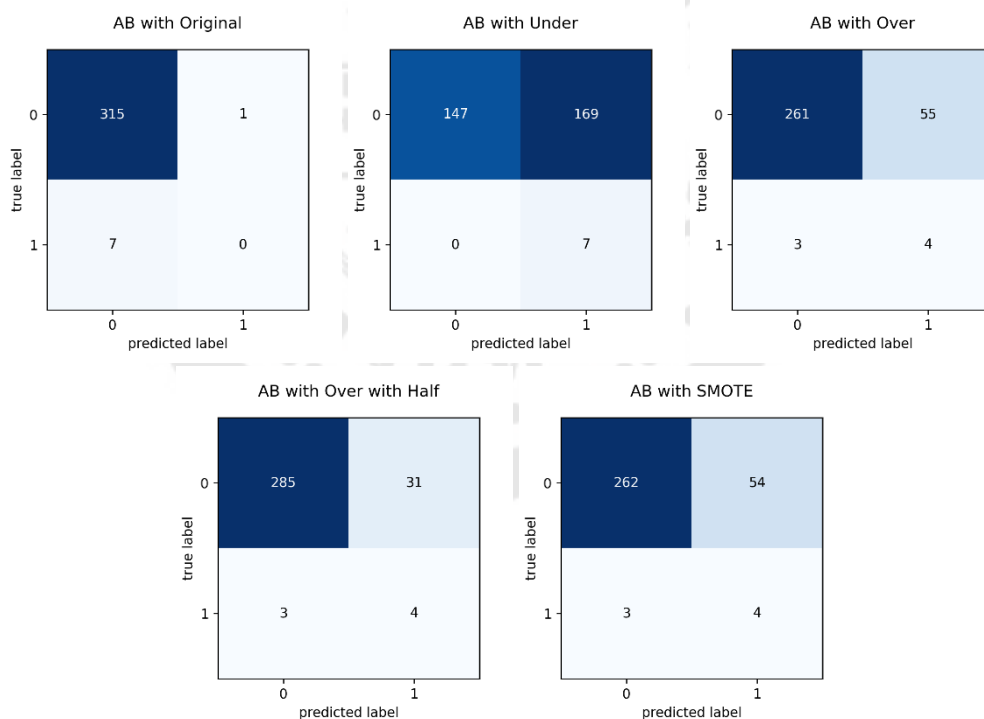


ภาพประกอบ 27 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.4768	0.4652	1.0000	0.0398	0.0765
Over	0.8204	0.8259	0.5714	0.0678	0.1212
Over with Half	0.8947	0.9019	0.5714	0.1143	0.1905
SMOTE	0.8235	0.8291	0.5714	0.0690	0.1231

ภาพประกอบ 28 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

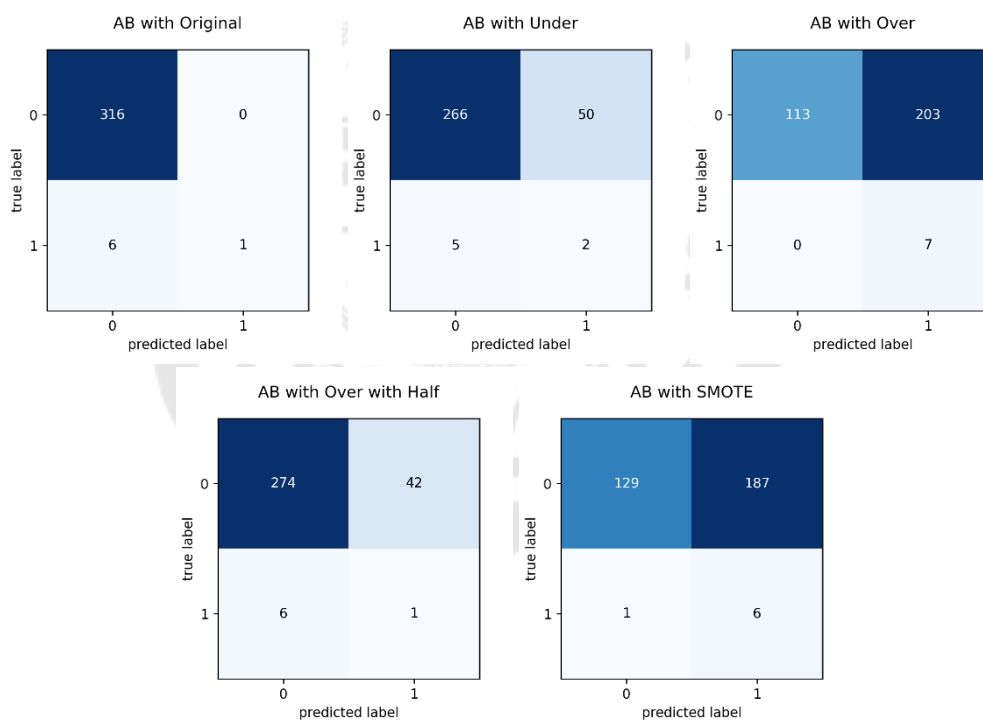


ภาพประกอบ 29 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9814	1.0000	0.1429	1.0000	0.2500
Under	0.8297	0.8418	0.2857	0.0385	0.0678
Over	0.3715	0.3576	1.0000	0.0333	0.0645
Over with Half	0.8514	0.8671	0.1429	0.0233	0.0400
SMOTE	0.4180	0.4082	0.8571	0.0311	0.0600

ภาพประกอบ 30 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



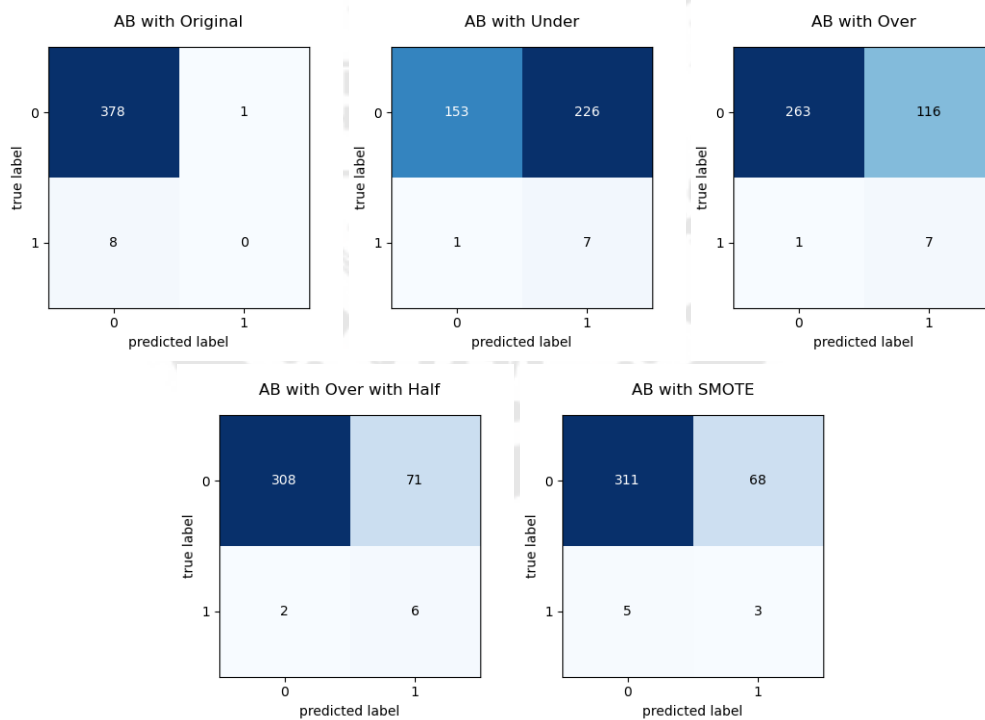
ภาพประกอบ 31 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.1.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9767	0.9974	0.0000	0.0000	0.0000
Under	0.4134	0.4037	0.8750	0.0300	0.0581
Over	0.6977	0.6939	0.8750	0.0569	0.1069
Over with Half	0.8114	0.8127	0.7500	0.0779	0.1412
SMOTE	0.8114	0.8206	0.3750	0.0423	0.0759

ภาพประกอบ 32 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

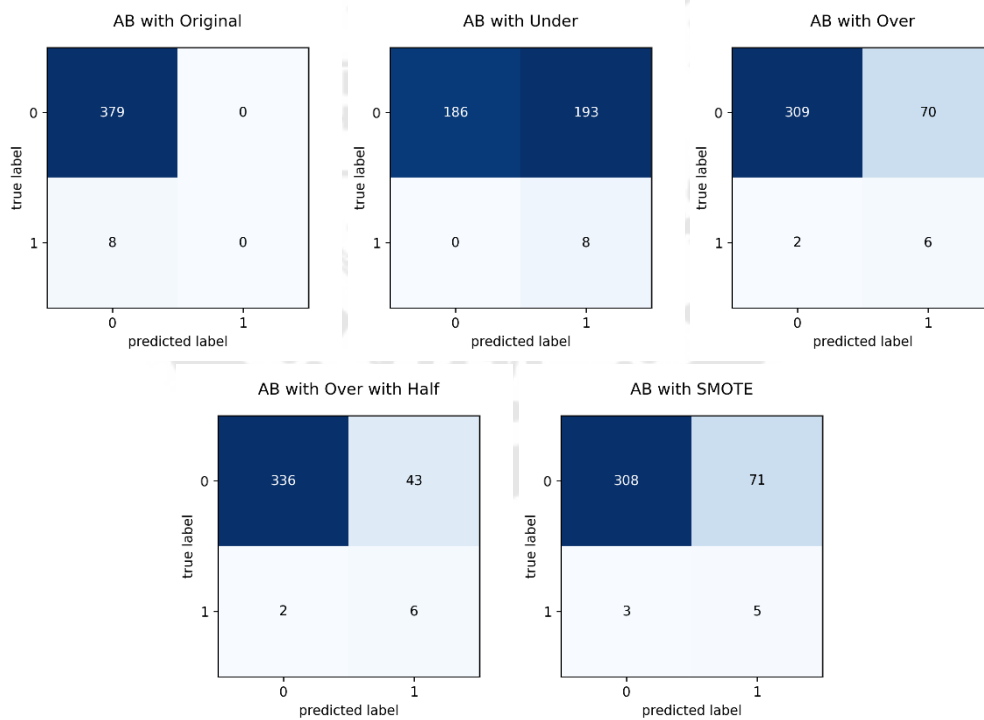


ภาพประกอบ 33 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.5013	0.4908	1.0000	0.0398	0.0766
Over	0.8140	0.8153	0.7500	0.0789	0.1429
Over with Half	0.8837	0.8865	0.7500	0.1224	0.2105
SMOTE	0.8088	0.8127	0.6250	0.0658	0.1190

ภาพประกอบ 34 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

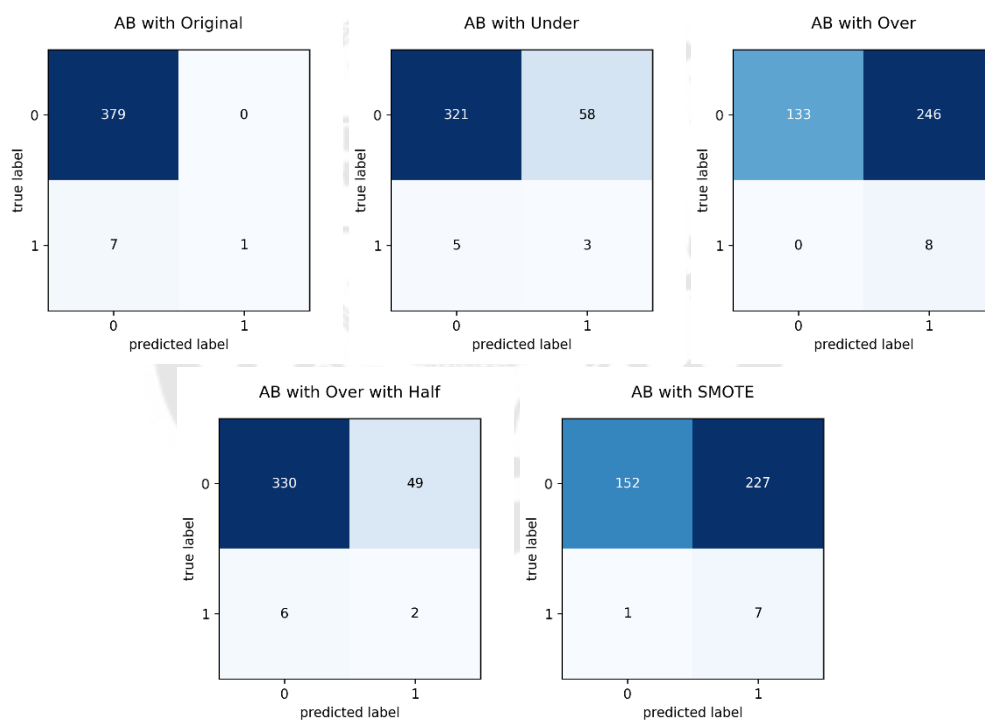


ภาพประกอบ 35 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9819	1.0000	0.1250	1.0000	0.2222
Under	0.8372	0.8470	0.3750	0.0492	0.0870
Over	0.3643	0.3509	1.0000	0.0315	0.0611
Over with Half	0.8579	0.8707	0.2500	0.0392	0.0678
SMOTE	0.4109	0.4011	0.8750	0.0299	0.0579

ภาพประกอบ 36 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



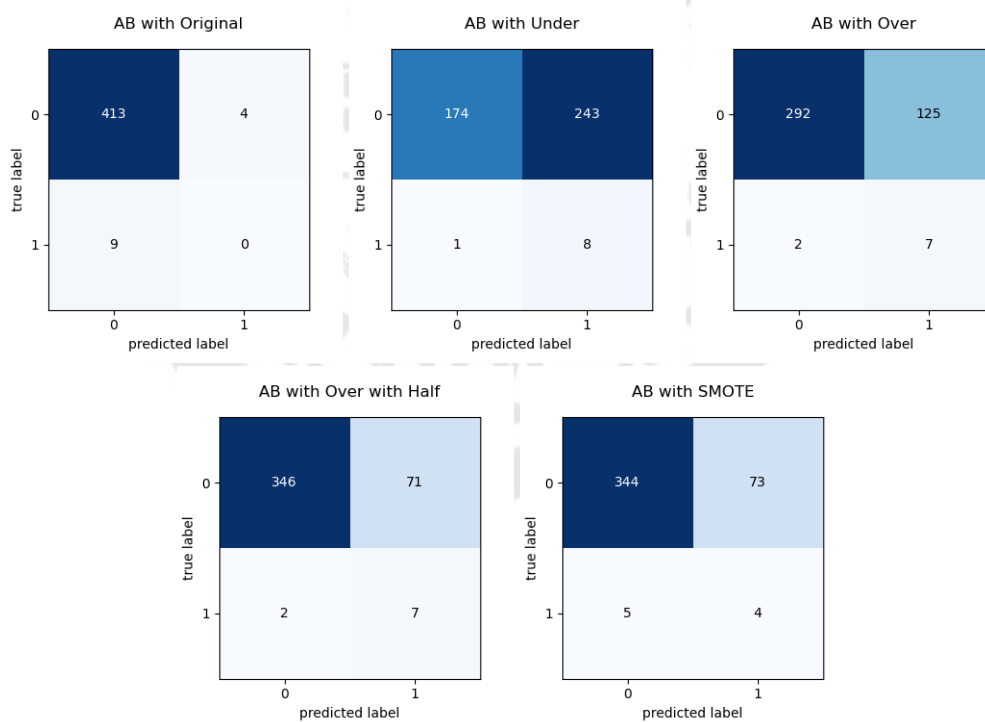
ภาพประกอบ 37 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.1.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9695	0.9904	0.0000	0.0000	0.0000
Under	0.4272	0.4173	0.8889	0.0319	0.0615
Over	0.7019	0.7002	0.7778	0.0530	0.0993
Over with Half	0.8286	0.8297	0.7778	0.0897	0.1609
SMOTE	0.8169	0.8249	0.4444	0.0519	0.0930

ภาพประกอบ 38 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

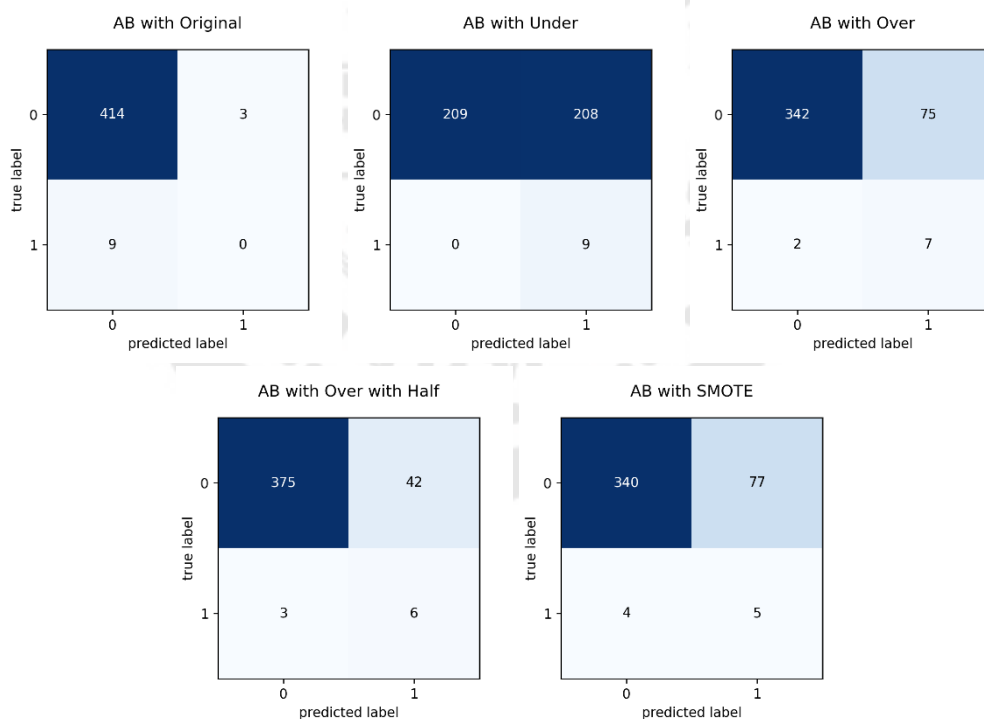


ภาพประกอบ 39 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9718	0.9928	0.0000	0.0000	0.0000
Under	0.5117	0.5012	1.0000	0.0415	0.0796
Over	0.8192	0.8201	0.7778	0.0854	0.1538
Over with Half	0.8944	0.8993	0.6667	0.1250	0.2105
SMOTE	0.8099	0.8153	0.5556	0.0610	0.1099

ภาพประกอบ 40 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

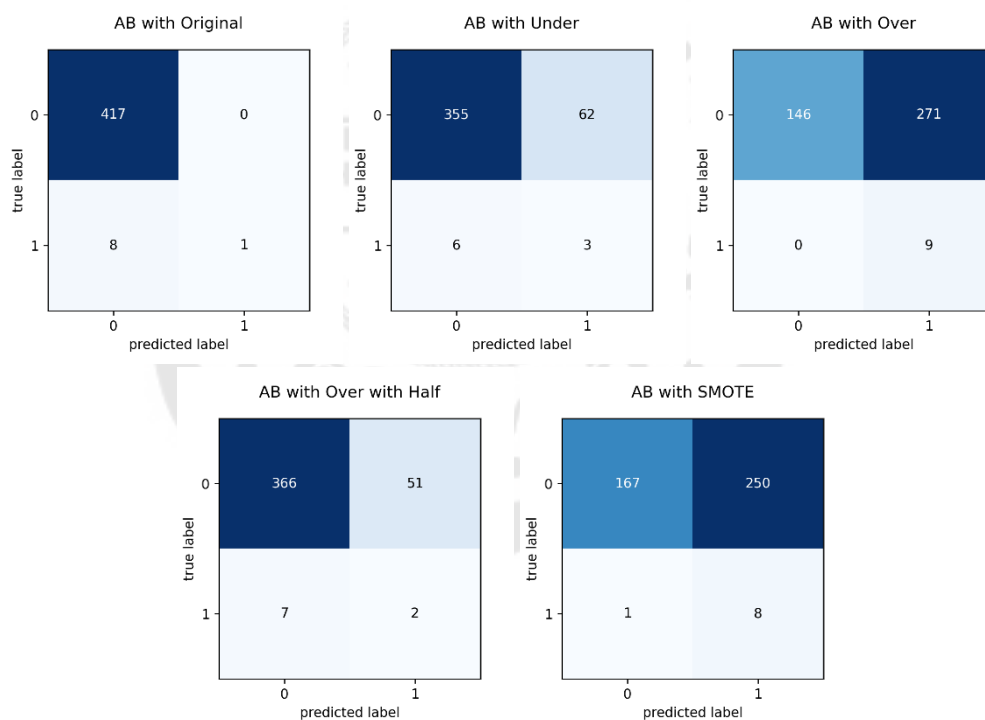


ภาพประกอบ 41 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.1.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9812	1.0000	0.1111	1.0000	0.2000
Under	0.8404	0.8513	0.3333	0.0462	0.0811
Over	0.3638	0.3501	1.0000	0.0321	0.0623
Over with Half	0.8638	0.8777	0.2222	0.0377	0.0645
SMOTE	0.4108	0.4005	0.8889	0.0310	0.0599

ภาพประกอบ 42 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 43 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

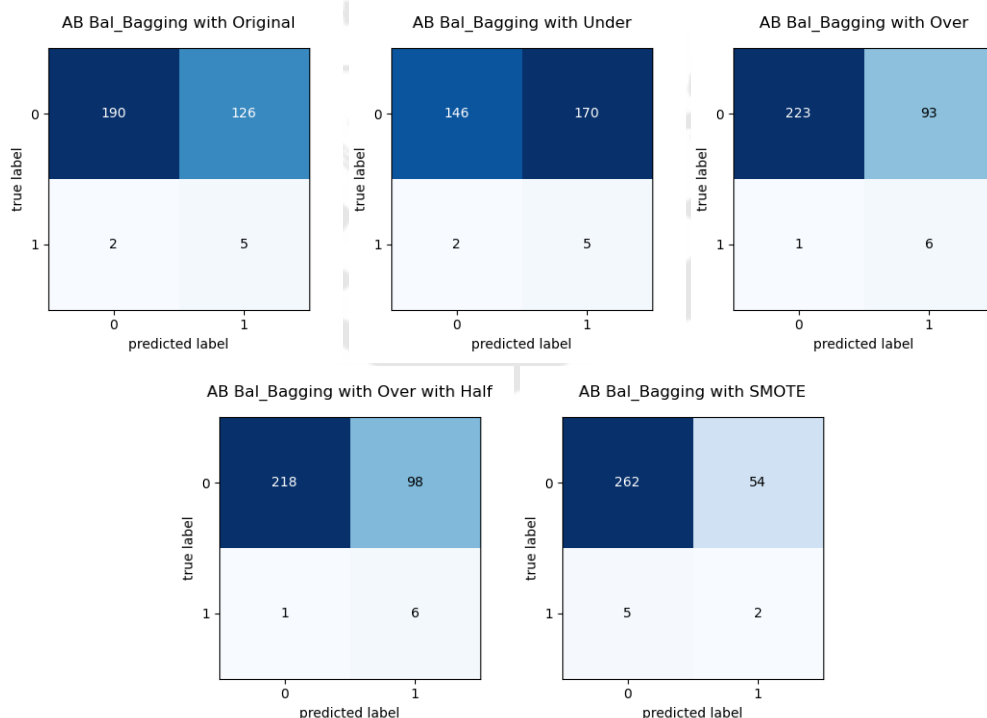
4.2 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Default Parameters และ Balanced Bagging Classifier

4.2.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.2.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6037	0.6013	0.7143	0.0382	0.0725
Under	0.4675	0.4620	0.7143	0.0286	0.0549
Over	0.7090	0.7057	0.8571	0.0606	0.1132
Over with Half	0.6935	0.6899	0.8571	0.0577	0.1081
SMOTE	0.8173	0.8291	0.2857	0.0357	0.0635

ภาพประกอบ 44 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

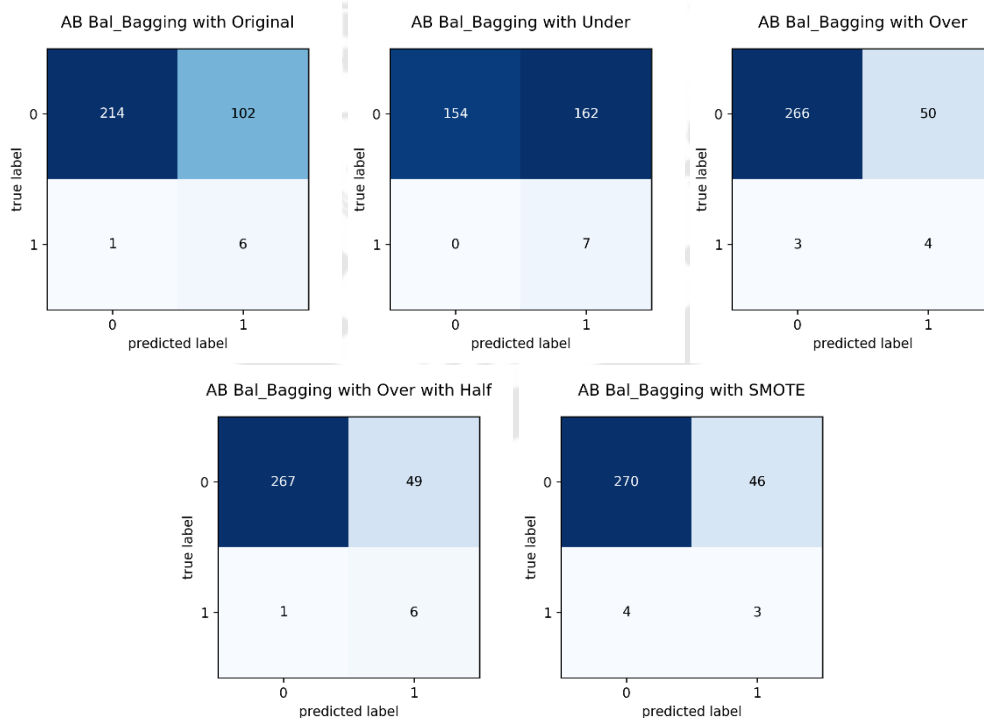


ภาพประกอบ 45 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6811	0.6772	0.8571	0.0556	0.1043
Under	0.4985	0.4873	1.0000	0.0414	0.0795
Over	0.8359	0.8418	0.5714	0.0741	0.1311
Over with Half	0.8452	0.8449	0.8571	0.1091	0.1935
SMOTE	0.8452	0.8544	0.4286	0.0612	0.1071

ภาพประกอบ 46 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

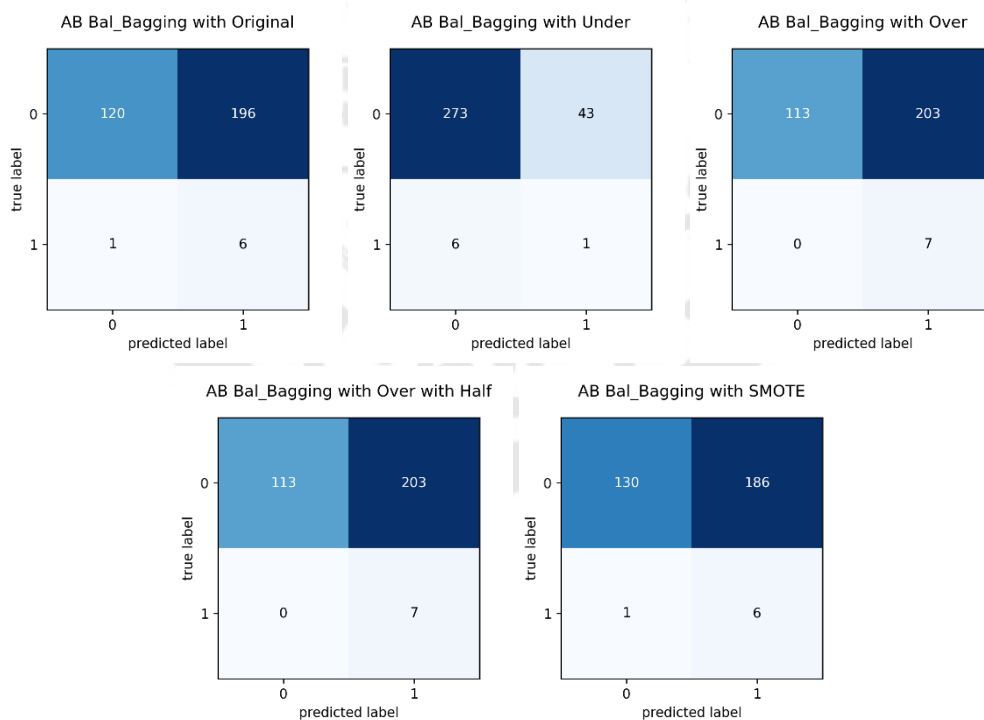


ภาพประกอบ 47 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3901	0.3797	0.8571	0.0297	0.0574
Under	0.8483	0.8639	0.1429	0.0227	0.0392
Over	0.3715	0.3576	1.0000	0.0333	0.0645
Over with Half	0.3715	0.3576	1.0000	0.0333	0.0645
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 48 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



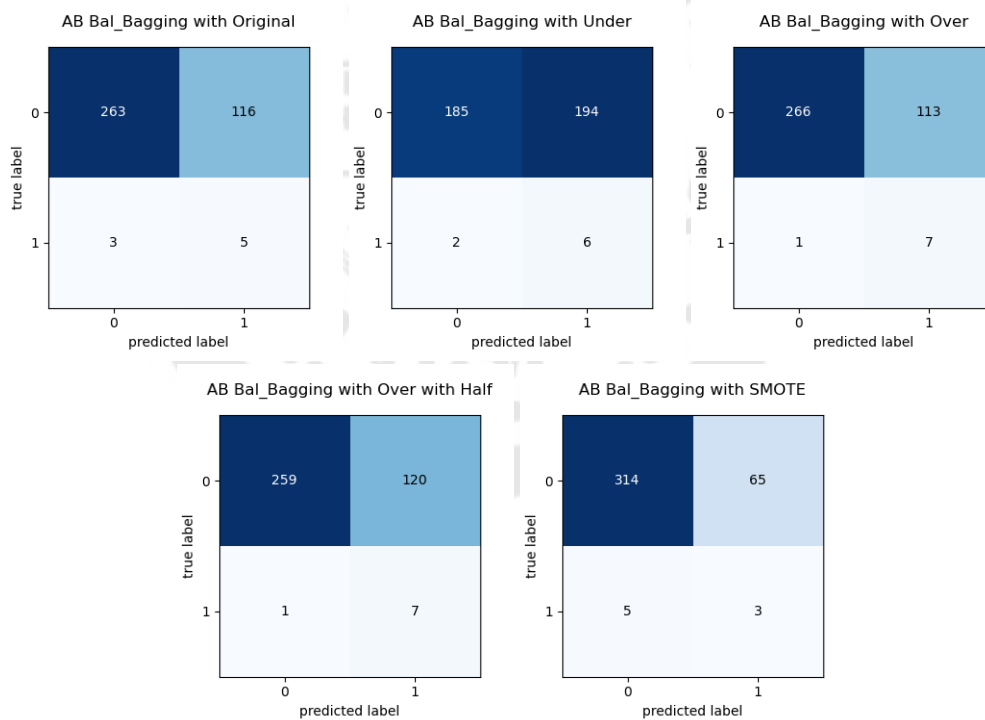
ภาพประกอบ 49 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.2.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6925	0.6939	0.6250	0.0413	0.0775
Under	0.4935	0.4881	0.7500	0.0300	0.0577
Over	0.7054	0.7018	0.8750	0.0583	0.1094
Over with Half	0.6873	0.6834	0.8750	0.0551	0.1037
SMOTE	0.8191	0.8285	0.3750	0.0441	0.0789

ภาพประกอบ 50 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

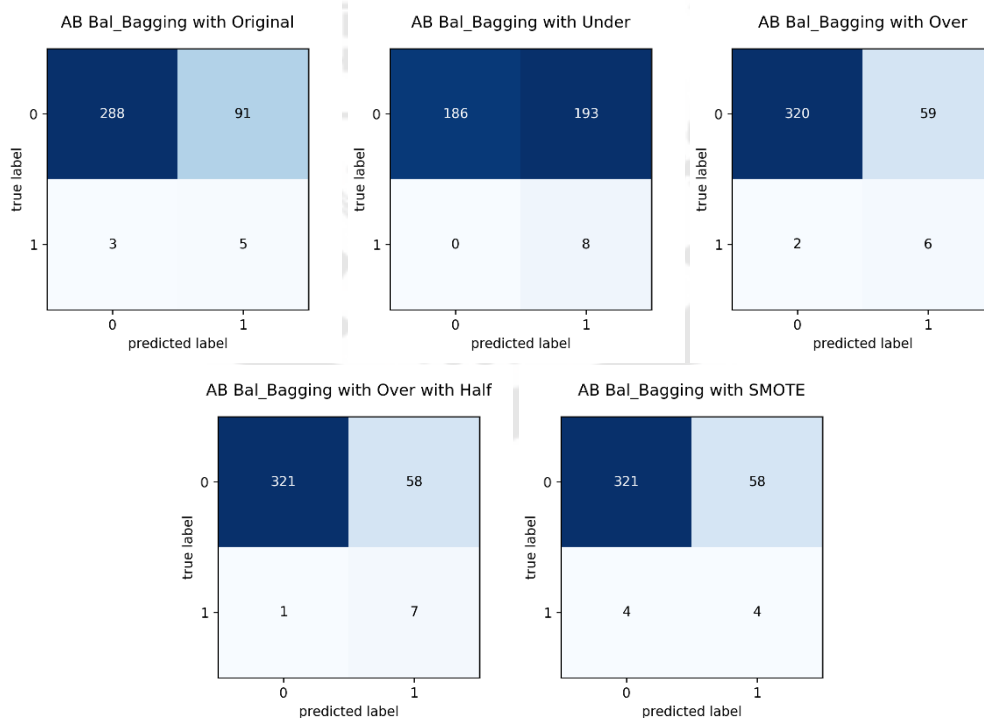


ภาพประกอบ 51 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7571	0.7599	0.6250	0.0521	0.0962
Under	0.5013	0.4908	1.0000	0.0398	0.0766
Over	0.8424	0.8443	0.7500	0.0923	0.1644
Over with Half	0.8475	0.8470	0.8750	0.1077	0.1918
SMOTE	0.8398	0.8470	0.5000	0.0645	0.1143

ภาพประกอบ 52 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

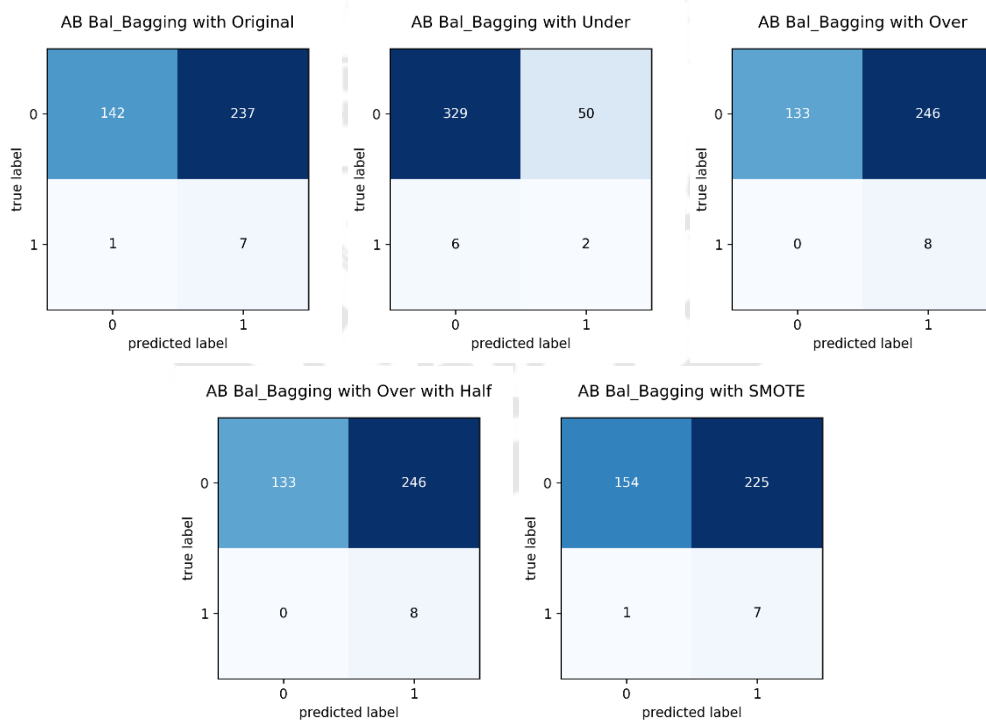


ภาพประกอบ 53 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3850	0.3747	0.8750	0.0287	0.0556
Under	0.8553	0.8681	0.2500	0.0385	0.0667
Over	0.3643	0.3509	1.0000	0.0315	0.0611
Over with Half	0.3643	0.3509	1.0000	0.0315	0.0611
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 54 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



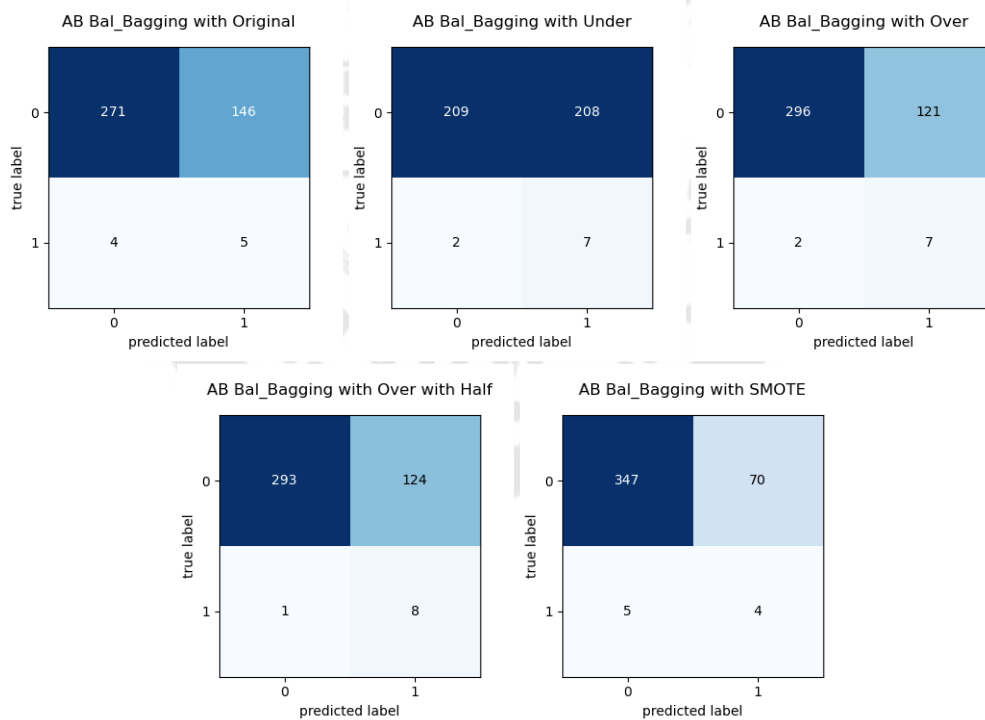
ภาพประกอบ 55 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.2.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6479	0.6499	0.5556	0.0331	0.0625
Under	0.5070	0.5012	0.7778	0.0326	0.0625
Over	0.7113	0.7098	0.7778	0.0547	0.1022
Over with Half	0.7066	0.7026	0.8889	0.0606	0.1135
SMOTE	0.8239	0.8321	0.4444	0.0541	0.0964

ภาพประกอบ 56 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

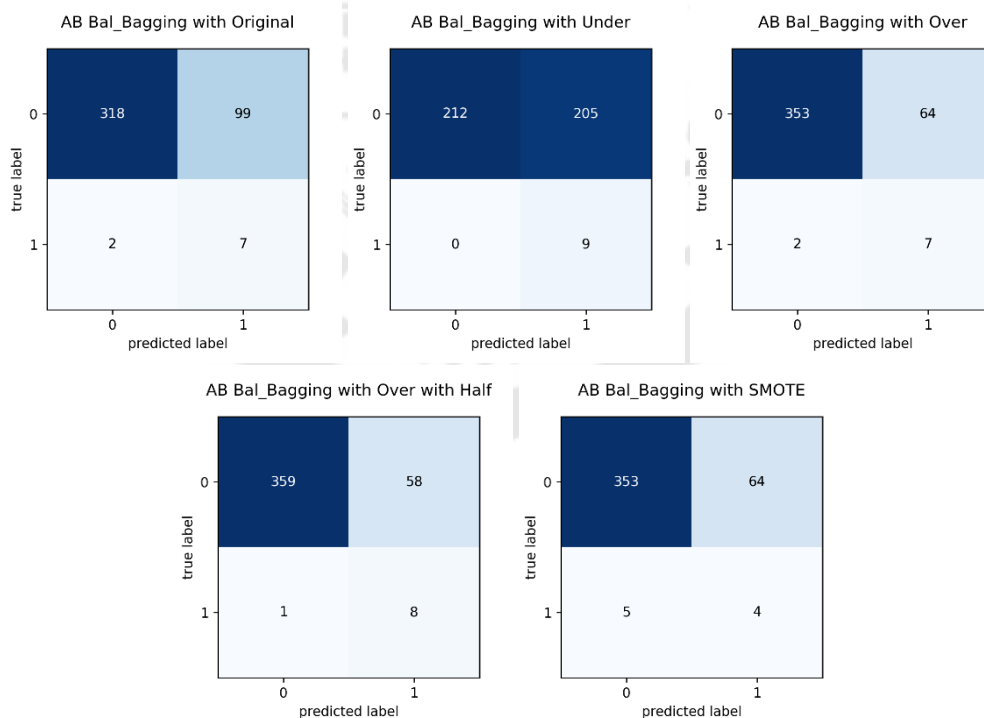


ภาพประกอบ 57 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7629	0.7626	0.7778	0.0660	0.1217
Under	0.5188	0.5084	1.0000	0.0421	0.0807
Over	0.8451	0.8465	0.7778	0.0986	0.1750
Over with Half	0.8615	0.8609	0.8889	0.1212	0.2133
SMOTE	0.8380	0.8465	0.4444	0.0588	0.1039

ภาพประกอบ 58 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

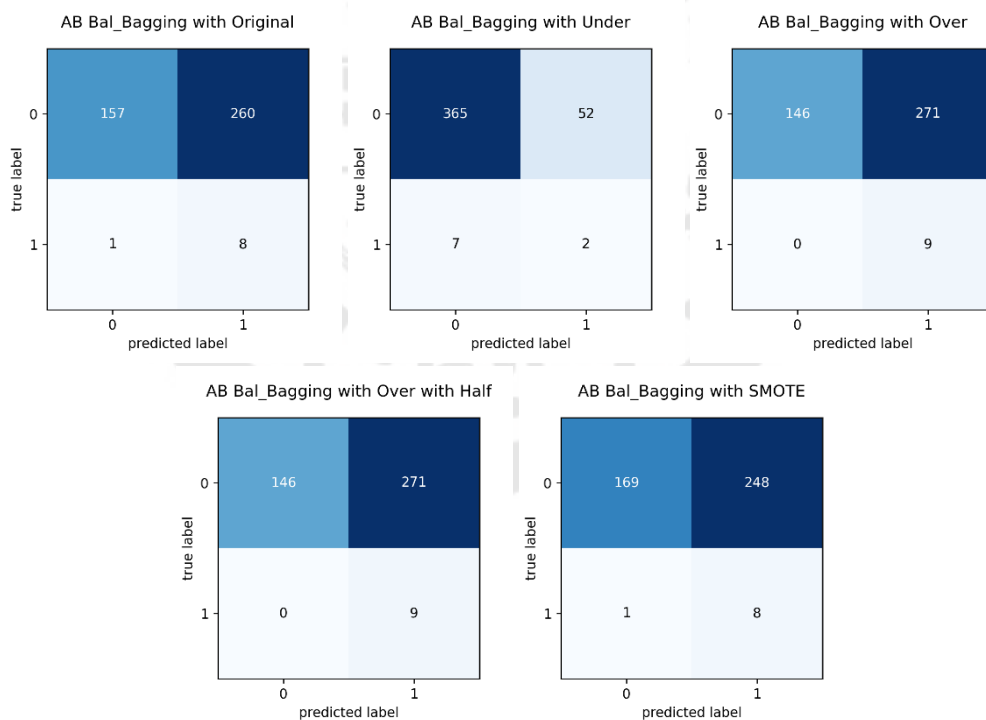


ภาพประกอบ 59 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.2.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3873	0.3765	0.8889	0.0299	0.0578
Under	0.8615	0.8753	0.2222	0.0370	0.0635
Over	0.3638	0.3501	1.0000	0.0321	0.0623
Over with Half	0.3638	0.3501	1.0000	0.0321	0.0623
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 60 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 61 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

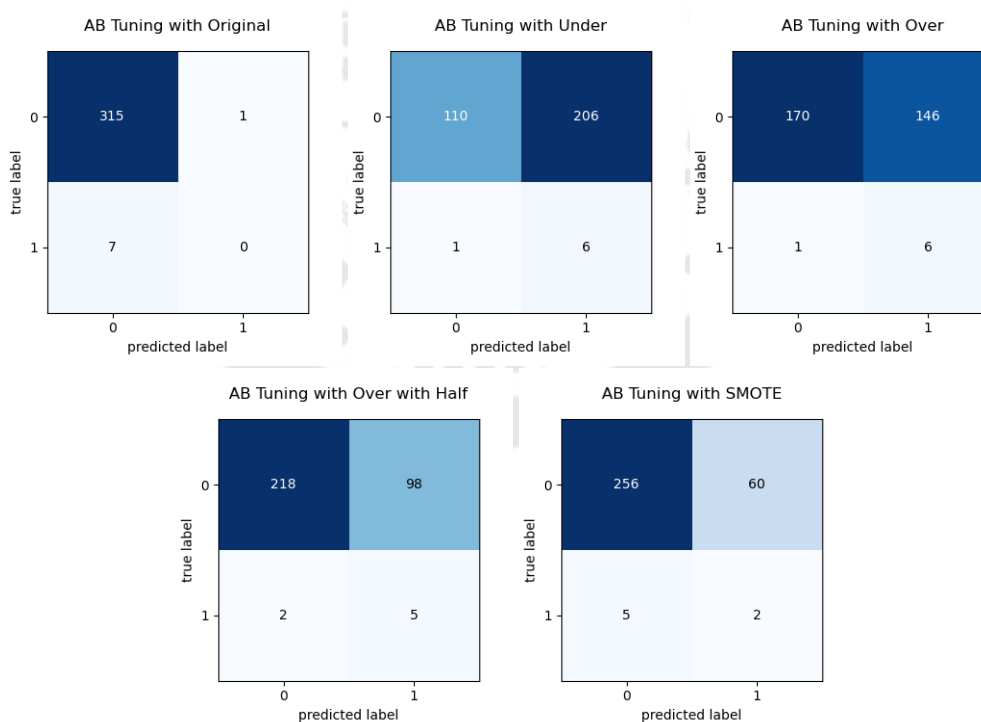
4.3 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Tuning Hyperparameter

4.3.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.3.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.3591	0.3481	0.8571	0.0283	0.0548
Over	0.5449	0.5380	0.8571	0.0395	0.0755
Over with Half	0.6904	0.6899	0.7143	0.0485	0.0909
SMOTE	0.7988	0.8101	0.2857	0.0323	0.0580

ภาพประกอบ 62 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

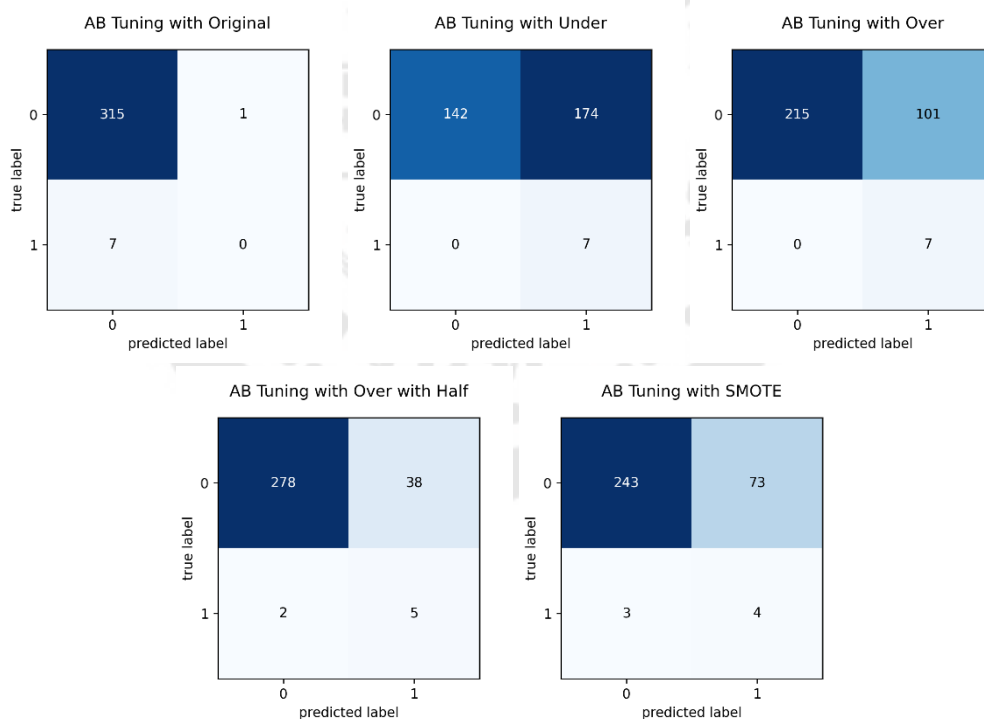


ภาพประกอบ 63 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.4613	0.4494	1.0000	0.0387	0.0745
Over	0.6873	0.6804	1.0000	0.0648	0.1217
Over with Half	0.8762	0.8797	0.7143	0.1163	0.2000
SMOTE	0.7647	0.7690	0.5714	0.0519	0.0952

ภาพประกอบ 64 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

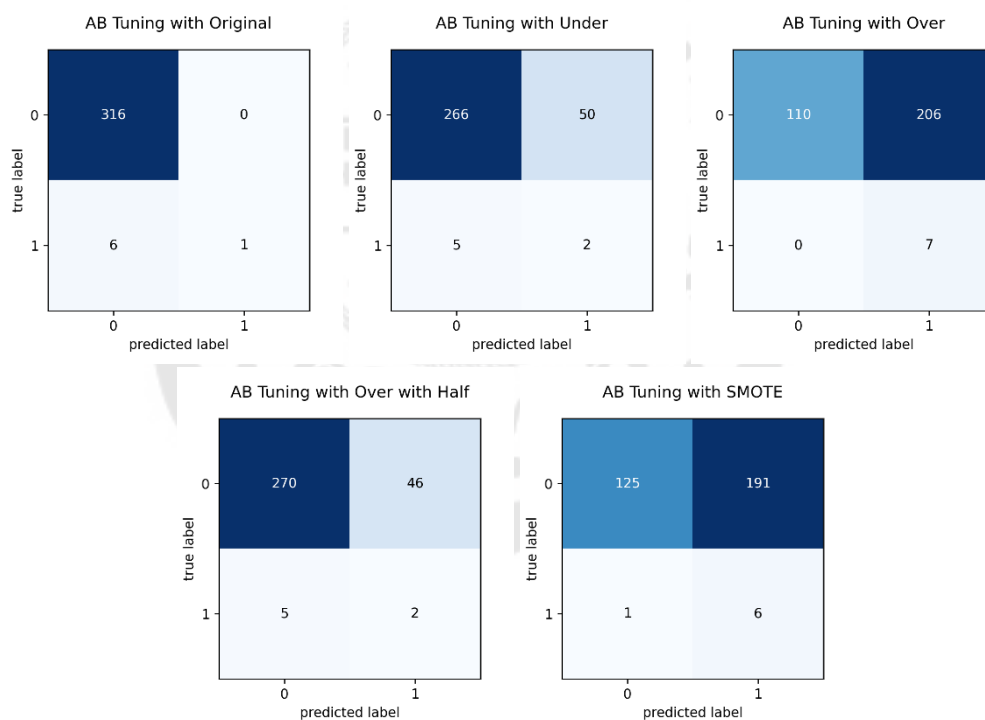


ภาพประกอบ 65 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9814	1.0000	0.1429	1.0000	0.2500
Under	0.8297	0.8418	0.2857	0.0385	0.0678
Over	0.3622	0.3481	1.0000	0.0329	0.0636
Over with Half	0.8421	0.8544	0.2857	0.0417	0.0727
SMOTE	0.4056	0.3956	0.8571	0.0305	0.0588

ภาพประกอบ 66 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



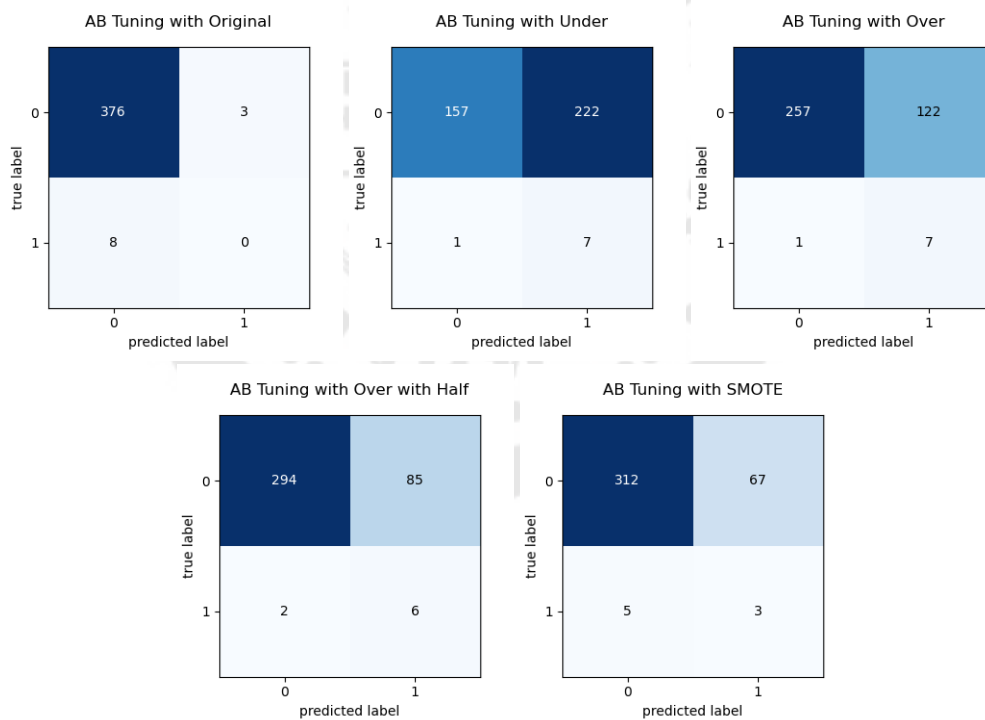
ภาพประกอบ 67 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.3.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9716	0.9921	0.0000	0.0000	0.0000
Under	0.4238	0.4142	0.8750	0.0306	0.0591
Over	0.6822	0.6781	0.8750	0.0543	0.1022
Over with Half	0.7752	0.7757	0.7500	0.0659	0.1212
SMOTE	0.8140	0.8232	0.3750	0.0429	0.0769

ภาพประกอบ 68 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

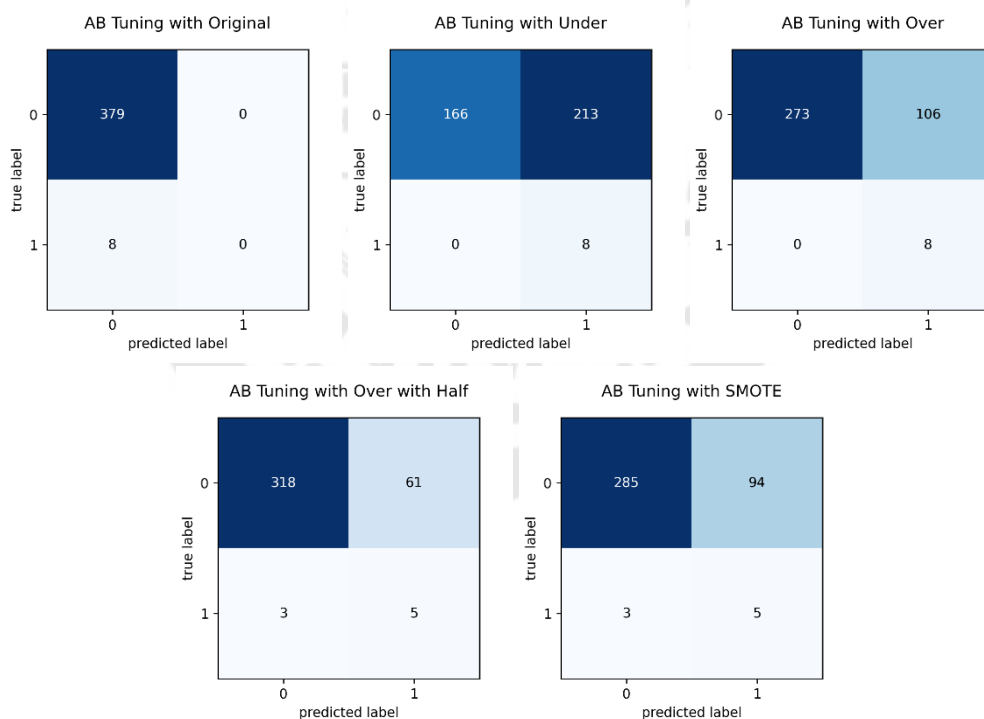


ภาพประกอบ 69 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4496	0.4380	1.0000	0.0362	0.0699
Over	0.7261	0.7203	1.0000	0.0702	0.1311
Over with Half	0.8346	0.8391	0.6250	0.0758	0.1351
SMOTE	0.7494	0.7520	0.6250	0.0505	0.0935

ภาพประกอบ 70 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

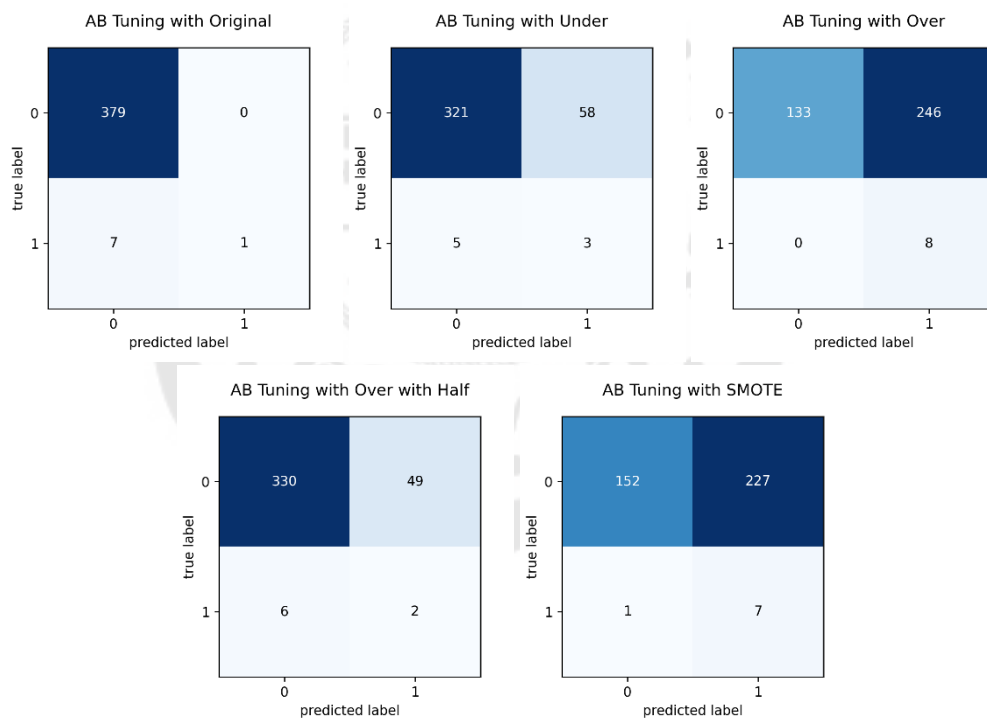


ภาพประกอบ 71 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9819	1.0000	0.1250	1.0000	0.2222
Under	0.8372	0.8470	0.3750	0.0492	0.0870
Over	0.3643	0.3509	1.0000	0.0315	0.0611
Over with Half	0.8579	0.8707	0.2500	0.0392	0.0678
SMOTE	0.4109	0.4011	0.8750	0.0299	0.0579

ภาพประกอบ 72 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



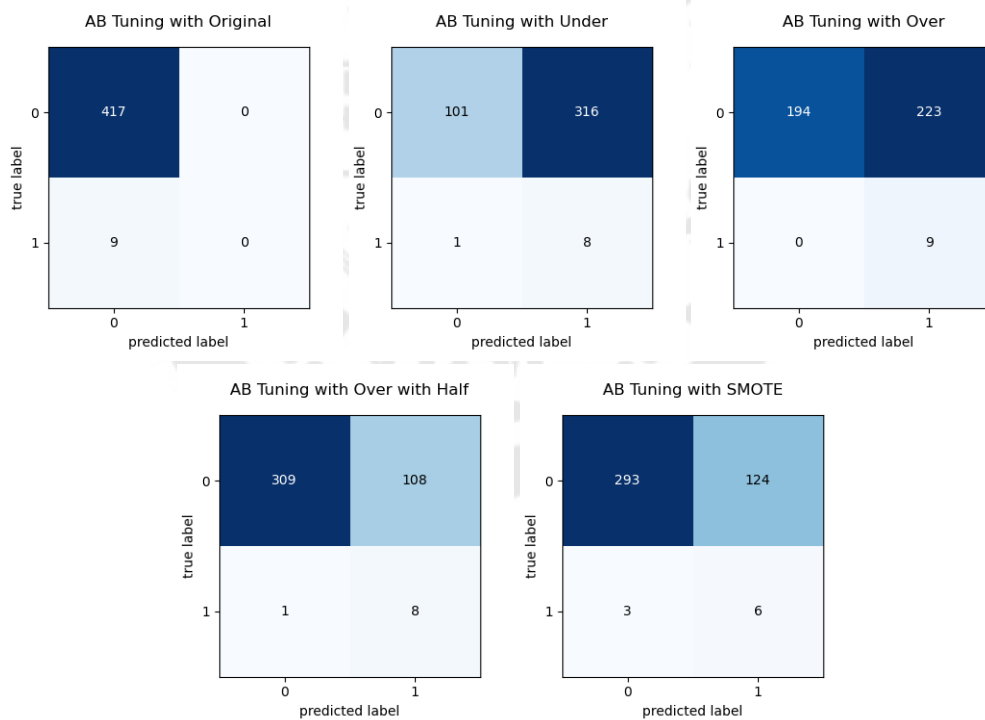
ภาพประกอบ 73 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.3.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.2559	0.2422	0.8889	0.0247	0.0480
Over	0.4765	0.4652	1.0000	0.0388	0.0747
Over with Half	0.7441	0.7410	0.8889	0.0690	0.1280
SMOTE	0.7019	0.7026	0.6667	0.0462	0.0863

ภาพประกอบ 74 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

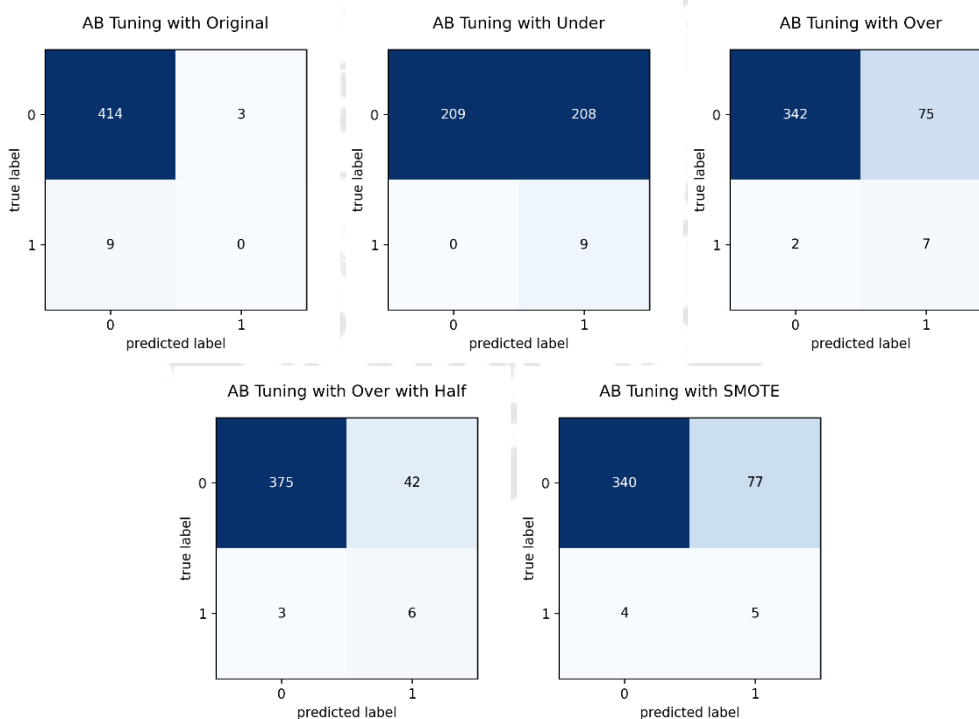


ภาพประกอบ 75 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9718	0.9928	0.0000	0.0000	0.0000
Under	0.5117	0.5012	1.0000	0.0415	0.0796
Over	0.8192	0.8201	0.7778	0.0854	0.1538
Over with Half	0.8944	0.8993	0.6667	0.1250	0.2105
SMOTE	0.8099	0.8153	0.5556	0.0610	0.1099

ภาพประกอบ 76 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

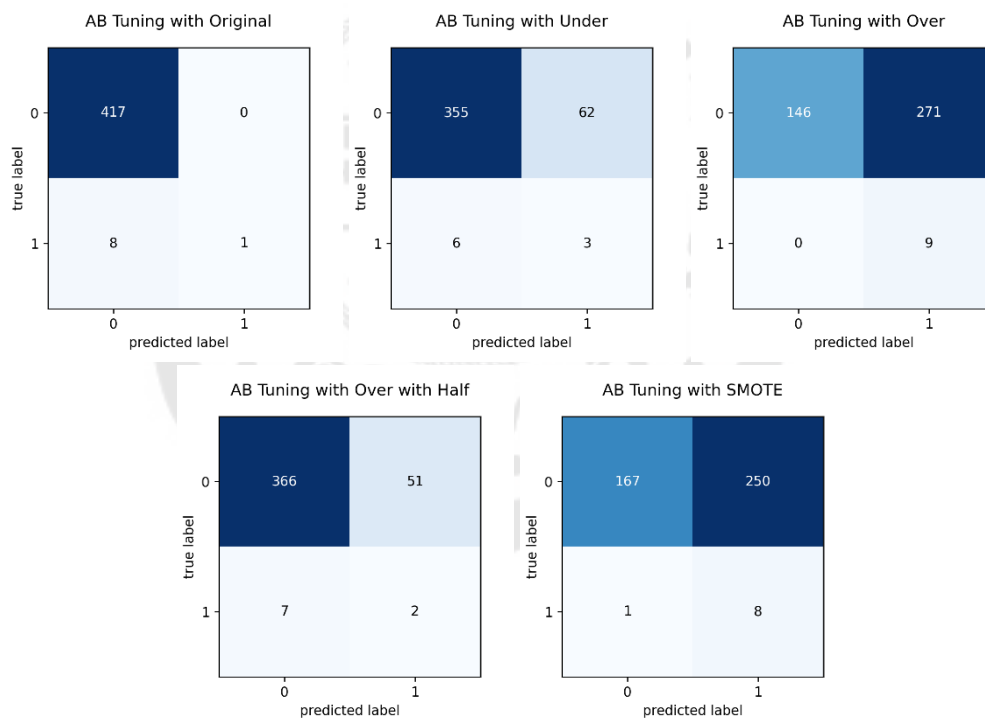


ภาพประกอบ 77 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.3.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9812	1.0000	0.1111	1.0000	0.2000
Under	0.8404	0.8513	0.3333	0.0462	0.0811
Over	0.3638	0.3501	1.0000	0.0321	0.0623
Over with Half	0.8638	0.8777	0.2222	0.0377	0.0645
SMOTE	0.4108	0.4005	0.8889	0.0310	0.0599

ภาพประกอบ 78 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 79 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier

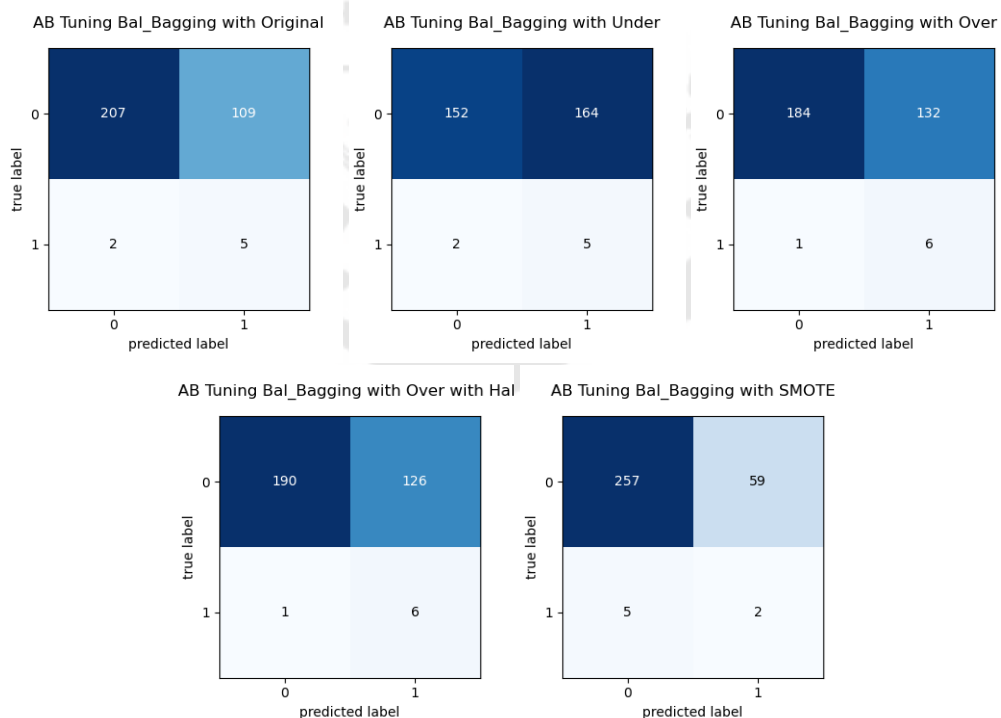
4.4.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.4.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6563	0.6551	0.7143	0.0439	0.0826
Under	0.4861	0.4810	0.7143	0.0296	0.0568
Over	0.5882	0.5823	0.8571	0.0435	0.0828
Over with Half	0.6068	0.6013	0.8571	0.0455	0.0863
SMOTE	0.8019	0.8133	0.2857	0.0328	0.0588

ภาพประกอบ 80 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

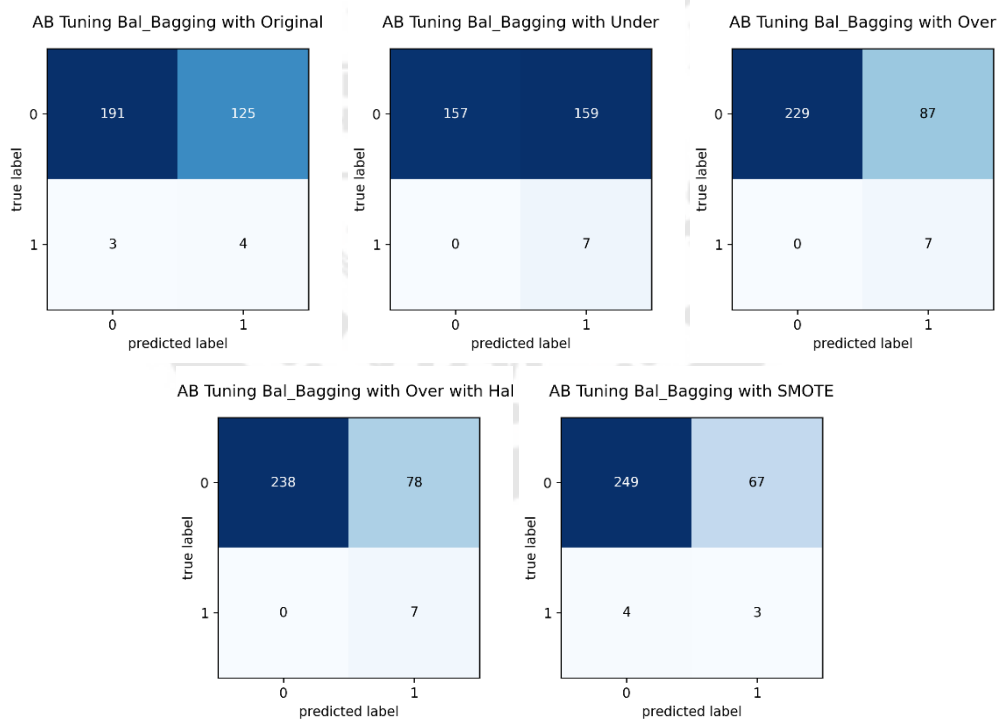


ภาพประกอบ 81 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6037	0.6044	0.5714	0.0310	0.0588
Under	0.5077	0.4968	1.0000	0.0422	0.0809
Over	0.7307	0.7247	1.0000	0.0745	0.1386
Over with Half	0.7585	0.7532	1.0000	0.0824	0.1522
SMOTE	0.7802	0.7880	0.4286	0.0429	0.0779

ภาพประกอบ 82 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

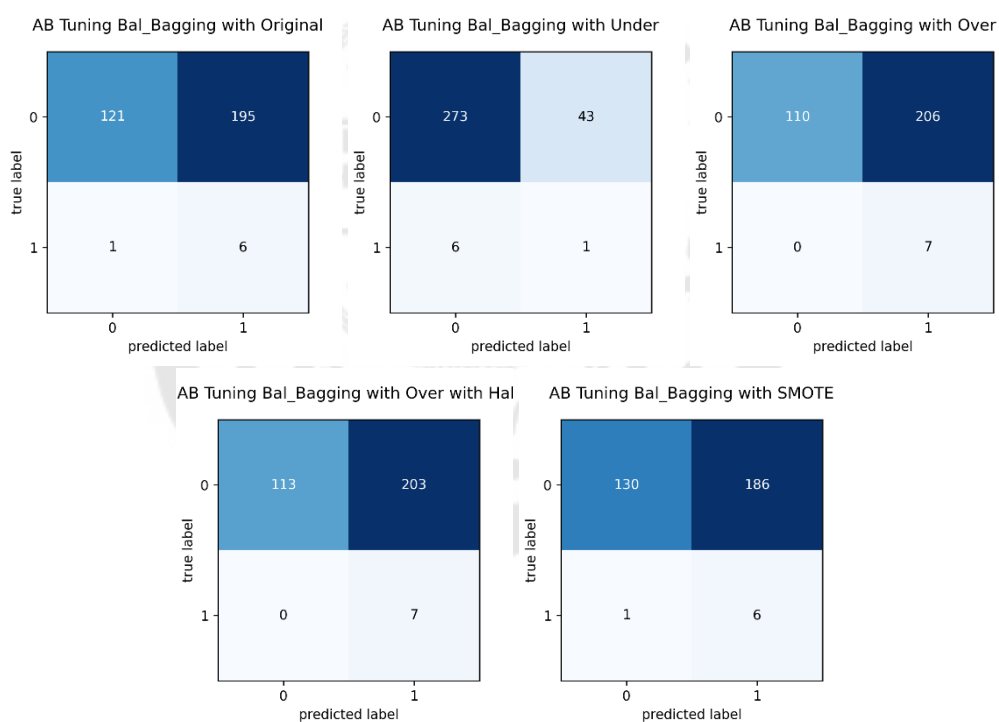


ภาพประกอบ 83 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3932	0.3829	0.8571	0.0299	0.0577
Under	0.8483	0.8639	0.1429	0.0227	0.0392
Over	0.3622	0.3481	1.0000	0.0329	0.0636
Over with Half	0.3715	0.3576	1.0000	0.0333	0.0645
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 84 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



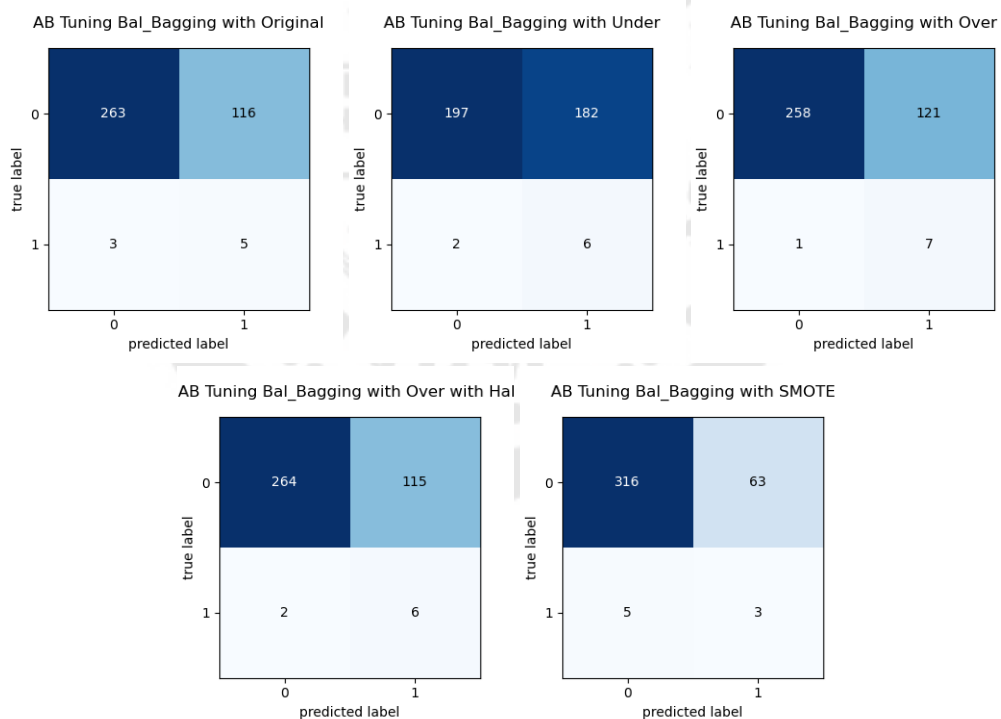
ภาพประกอบ 85 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.4.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6925	0.6939	0.6250	0.0413	0.0775
Under	0.5245	0.5198	0.7500	0.0319	0.0612
Over	0.6848	0.6807	0.8750	0.0547	0.1029
Over with Half	0.6977	0.6966	0.7500	0.0496	0.0930
SMOTE	0.8243	0.8338	0.3750	0.0455	0.0811

ภาพประกอบ 86 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

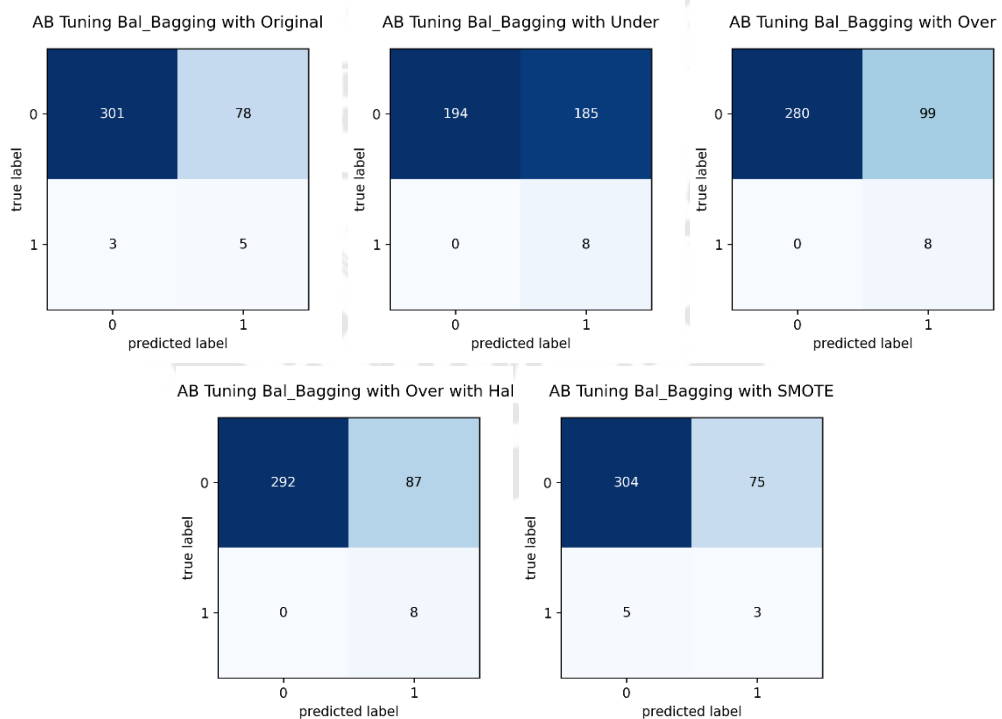


ภาพประกอบ 87 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7907	0.7942	0.6250	0.0602	0.1099
Under	0.5220	0.5119	1.0000	0.0415	0.0796
Over	0.7442	0.7388	1.0000	0.0748	0.1391
Over with Half	0.7752	0.7704	1.0000	0.0842	0.1553
SMOTE	0.7933	0.8021	0.3750	0.0385	0.0698

ภาพประกอบ 88 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



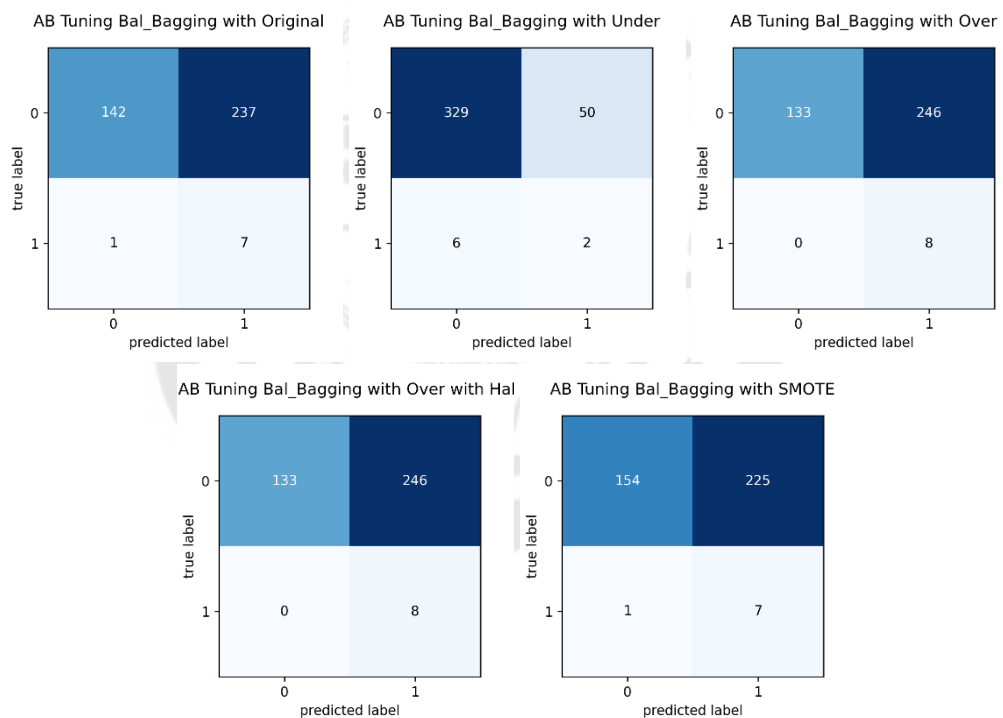
ภาพประกอบ 89 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3850	0.3747	0.8750	0.0287	0.0556
Under	0.8553	0.8681	0.2500	0.0385	0.0667
Over	0.3643	0.3509	1.0000	0.0315	0.0611
Over with Half	0.3643	0.3509	1.0000	0.0315	0.0611
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 90 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



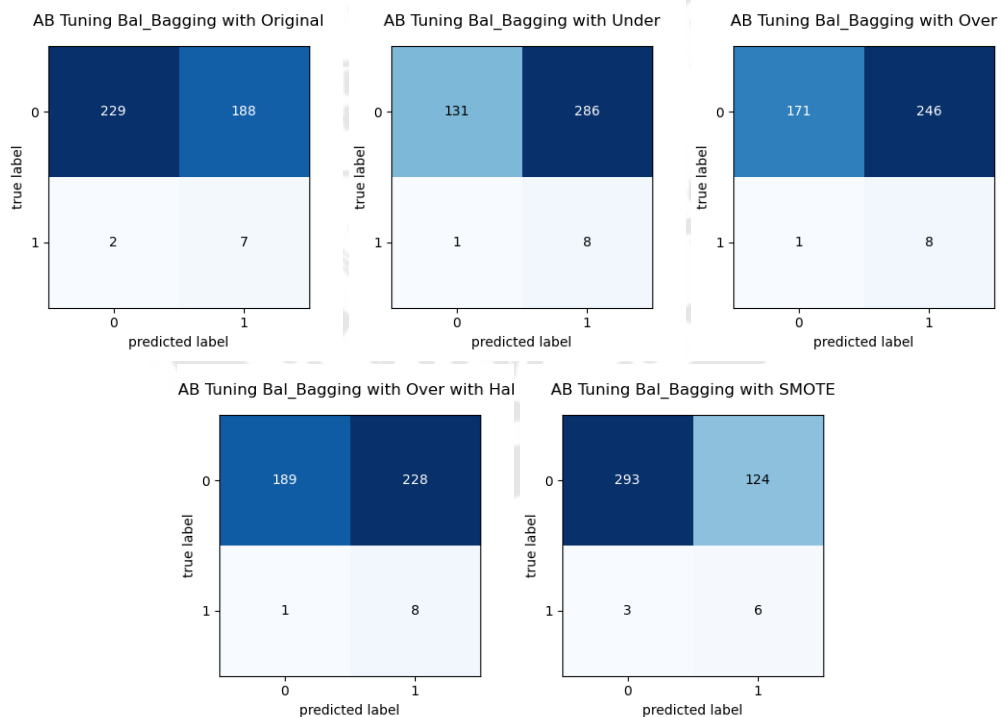
ภาพประกอบ 91 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.4.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.5540	0.5492	0.7778	0.0359	0.0686
Under	0.3263	0.3141	0.8889	0.0272	0.0528
Over	0.4202	0.4101	0.8889	0.0315	0.0608
Over with Half	0.4624	0.4532	0.8889	0.0339	0.0653
SMOTE	0.7019	0.7026	0.6667	0.0462	0.0863

ภาพประกอบ 92 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

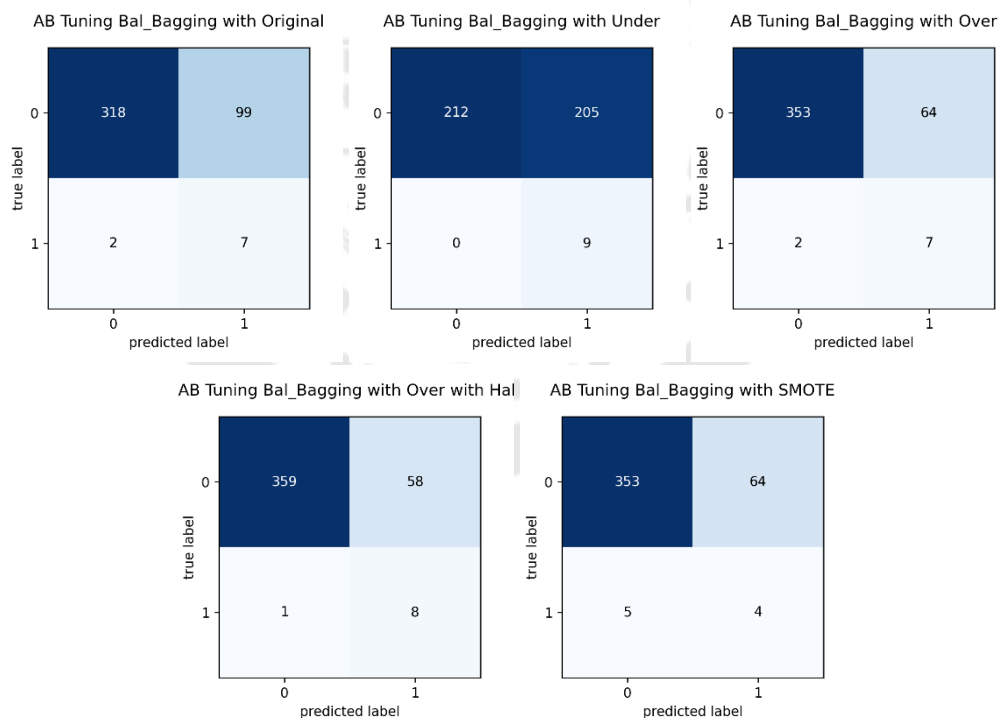


ภาพประกอบ 93 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7629	0.7626	0.7778	0.0660	0.1217
Under	0.5188	0.5084	1.0000	0.0421	0.0807
Over	0.8451	0.8465	0.7778	0.0986	0.1750
Over with Half	0.8615	0.8609	0.8889	0.1212	0.2133
SMOTE	0.8380	0.8465	0.4444	0.0588	0.1039

ภาพประกอบ 94 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



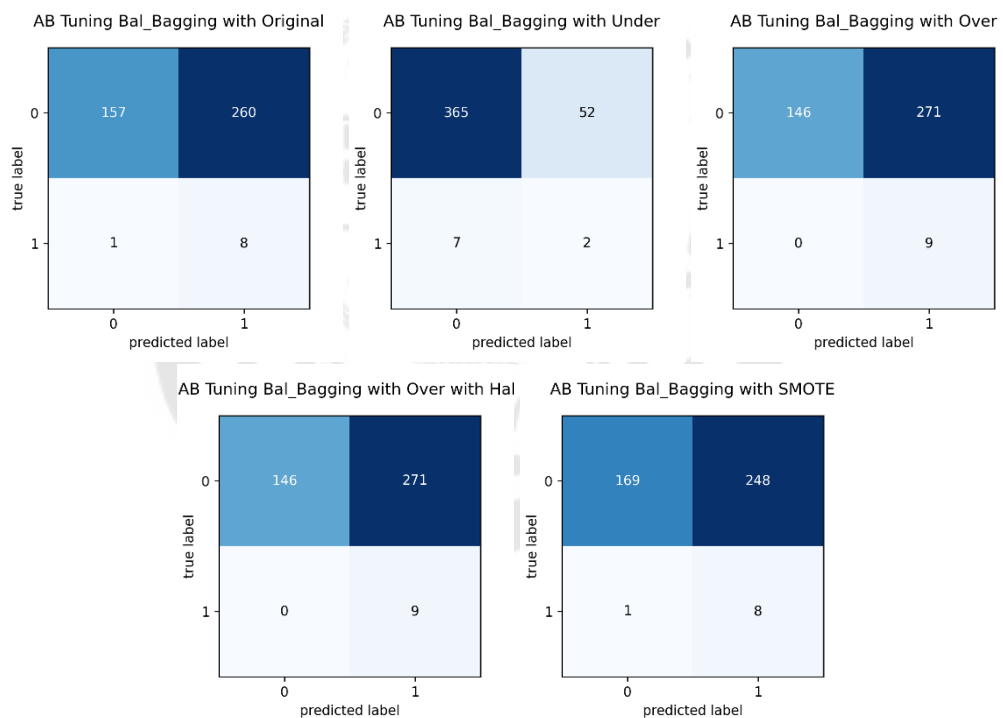
ภาพประกอบ 95 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.4.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3873	0.3765	0.8889	0.0299	0.0578
Under	0.8615	0.8753	0.2222	0.0370	0.0635
Over	0.3638	0.3501	1.0000	0.0321	0.0623
Over with Half	0.3638	0.3501	1.0000	0.0321	0.0623
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 96 ผลลัพธ์ที่ได้จากการใช้ AdaBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Decision Tree



ภาพประกอบ 97 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

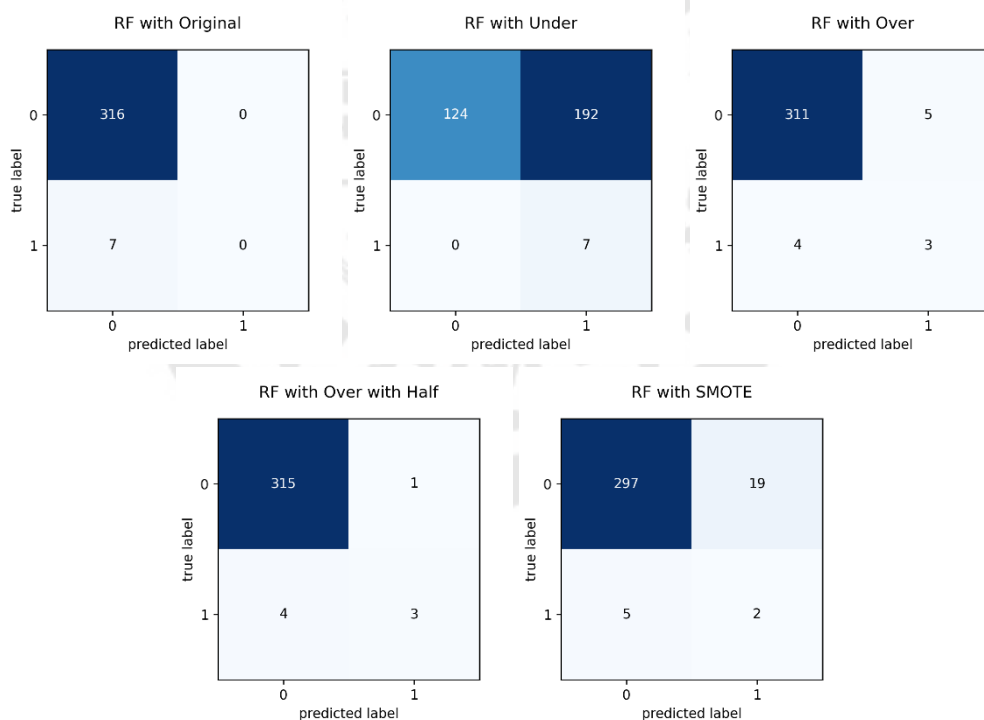
4.5 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters

4.5.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.5.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.4056	0.3924	1.0000	0.0352	0.0680
Over	0.9721	0.9842	0.4286	0.3750	0.4000
Over with Half	0.9845	0.9968	0.4286	0.7500	0.5455
SMOTE	0.9257	0.9399	0.2857	0.0952	0.1429

ภาพประกอบ 98 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

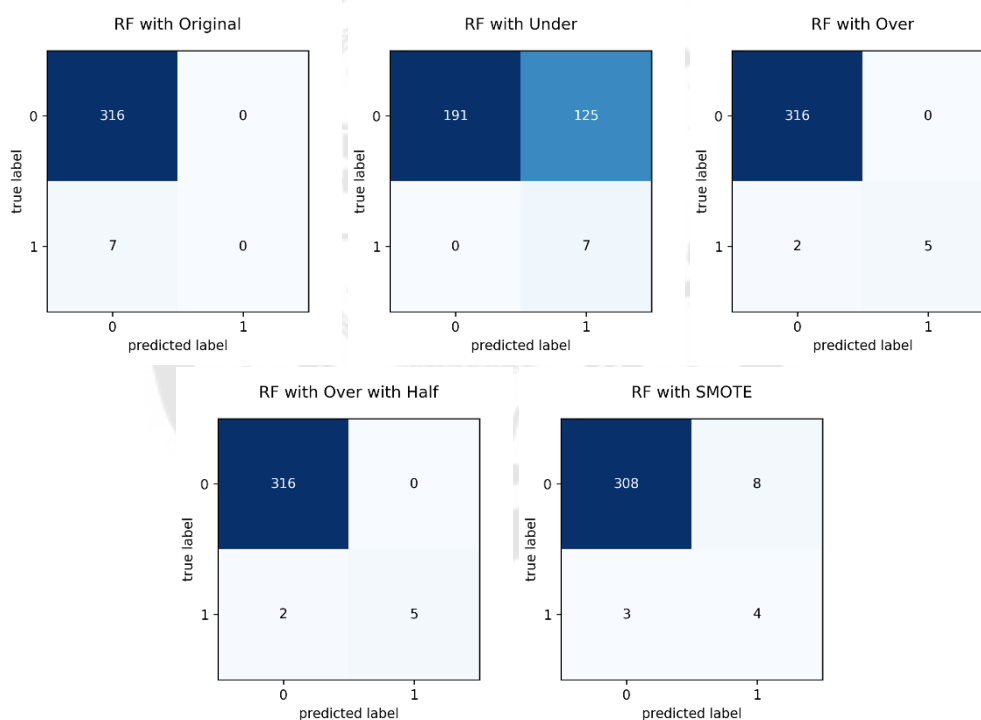


ภาพประกอบ 99 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.6130	0.6044	1.0000	0.0530	0.1007
Over	0.9938	1.0000	0.7143	1.0000	0.8333
Over with Half	0.9938	1.0000	0.7143	1.0000	0.8333
SMOTE	0.9659	0.9747	0.5714	0.3333	0.4211

ภาพประกอบ 100 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

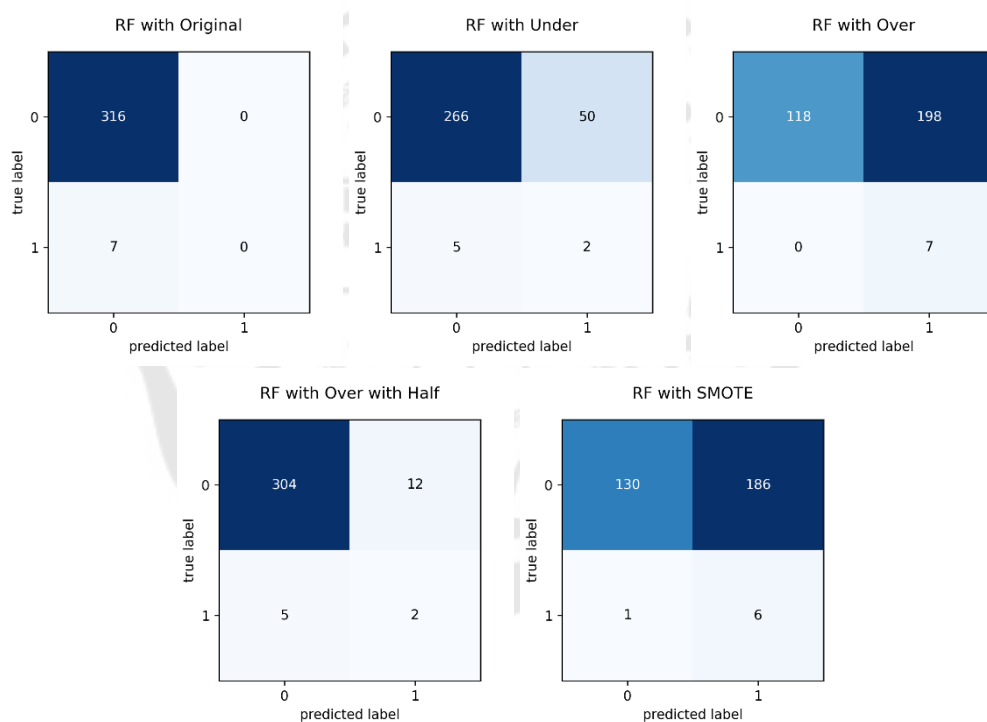


ภาพประกอบ 101 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.8297	0.8418	0.2857	0.0385	0.0678
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.9474	0.9620	0.2857	0.1429	0.1905
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 102 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



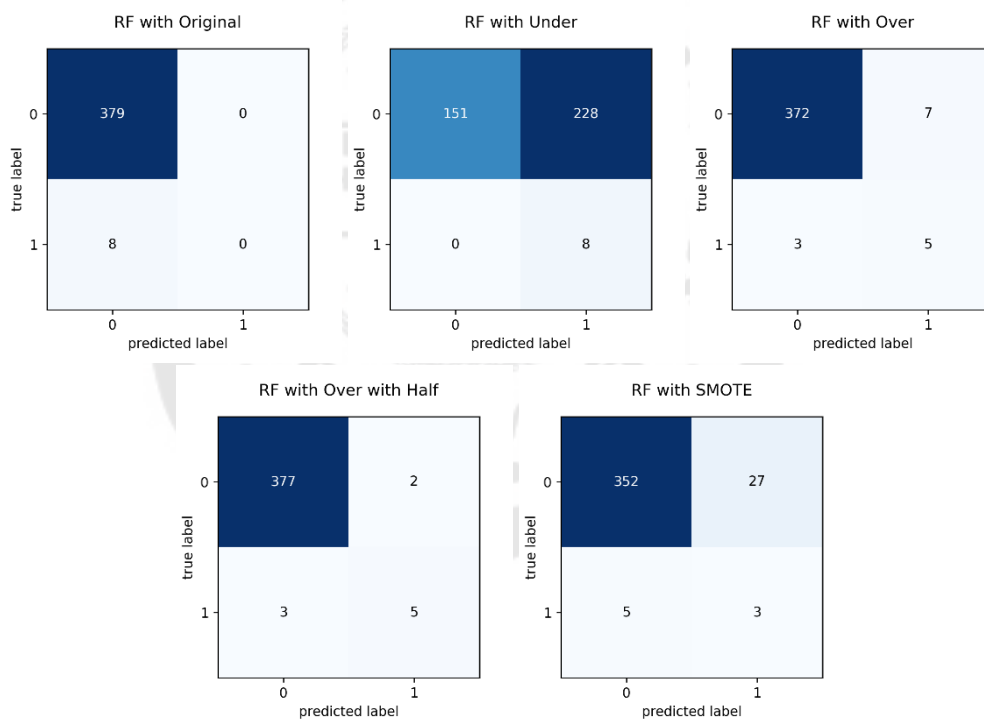
ภาพประกอบ 103 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.5.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4109	0.3984	1.0000	0.0339	0.0656
Over	0.9742	0.9815	0.6250	0.4167	0.5000
Over with Half	0.9871	0.9947	0.6250	0.7143	0.6667
SMOTE	0.9173	0.9288	0.3750	0.1000	0.1579

ภาพประกอบ 104 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

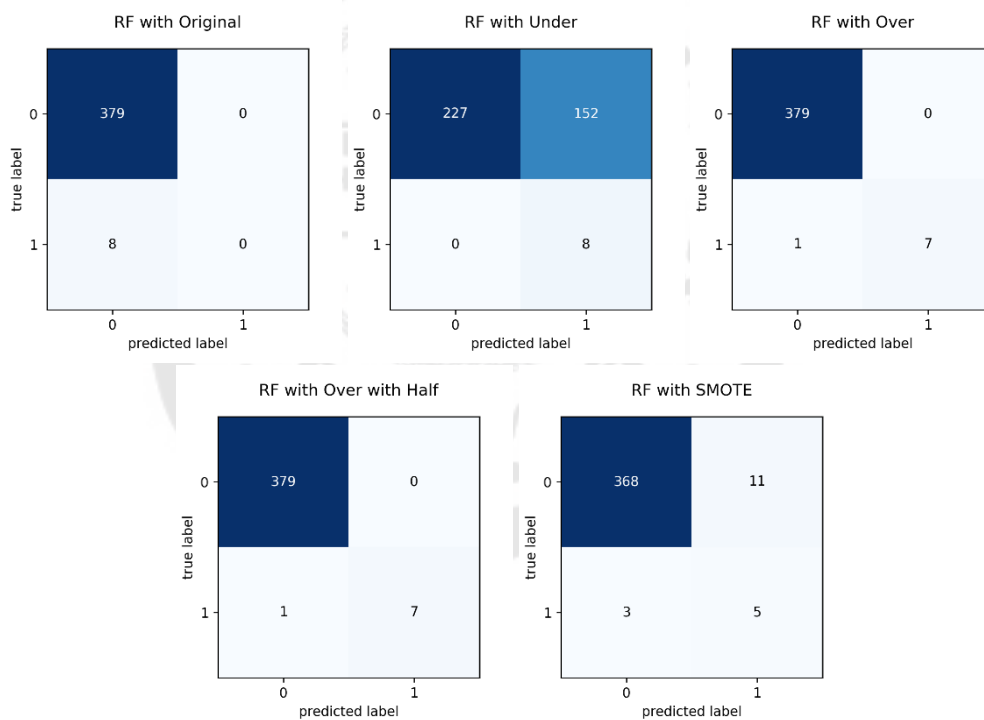


ภาพประกอบ 105 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.6072	0.5989	1.0000	0.0500	0.0952
Over	0.9974	1.0000	0.8750	1.0000	0.9333
Over with Half	0.9974	1.0000	0.8750	1.0000	0.9333
SMOTE	0.9638	0.9710	0.6250	0.3125	0.4167

ภาพประกอบ 106 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

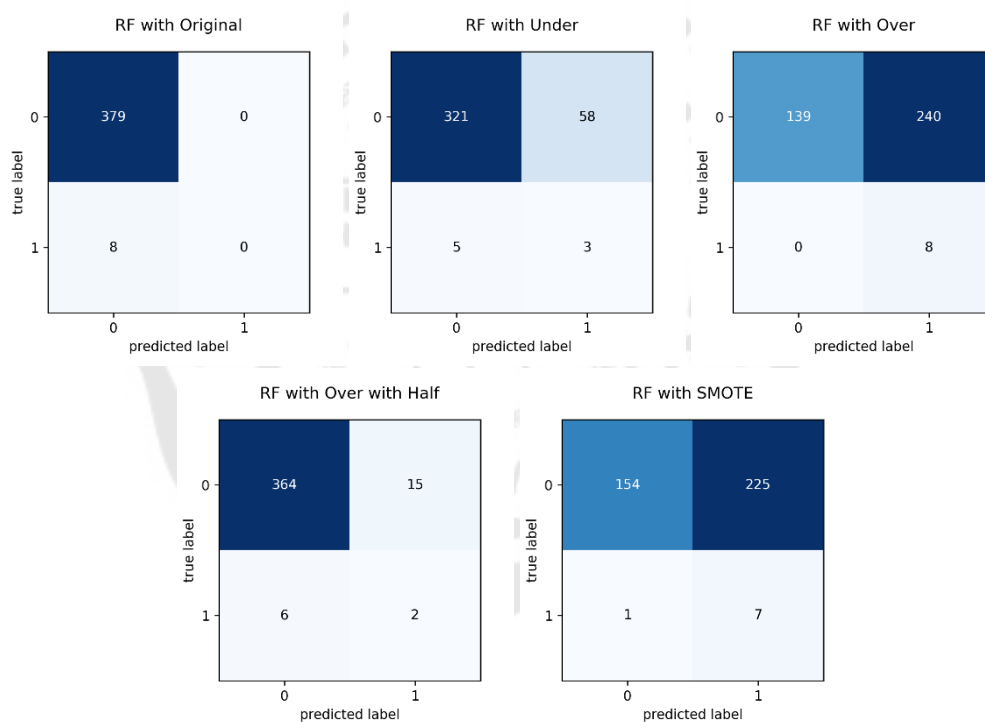


ภาพประกอบ 107 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.8372	0.8470	0.3750	0.0492	0.0870
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.9457	0.9604	0.2500	0.1176	0.1600
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 108 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



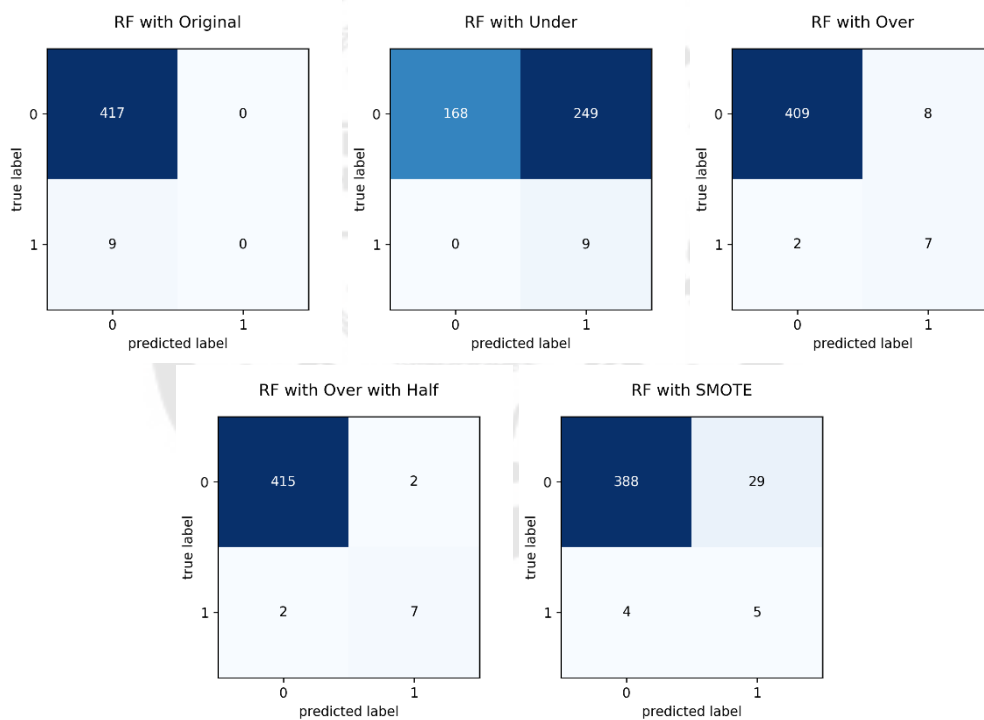
ภาพประกอบ 109 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.5.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.4155	0.4029	1.0000	0.0349	0.0674
Over	0.9765	0.9808	0.7778	0.4667	0.5833
Over with Half	0.9906	0.9952	0.7778	0.7778	0.7778
SMOTE	0.9225	0.9305	0.5556	0.1471	0.2326

ภาพประกอบ 110 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

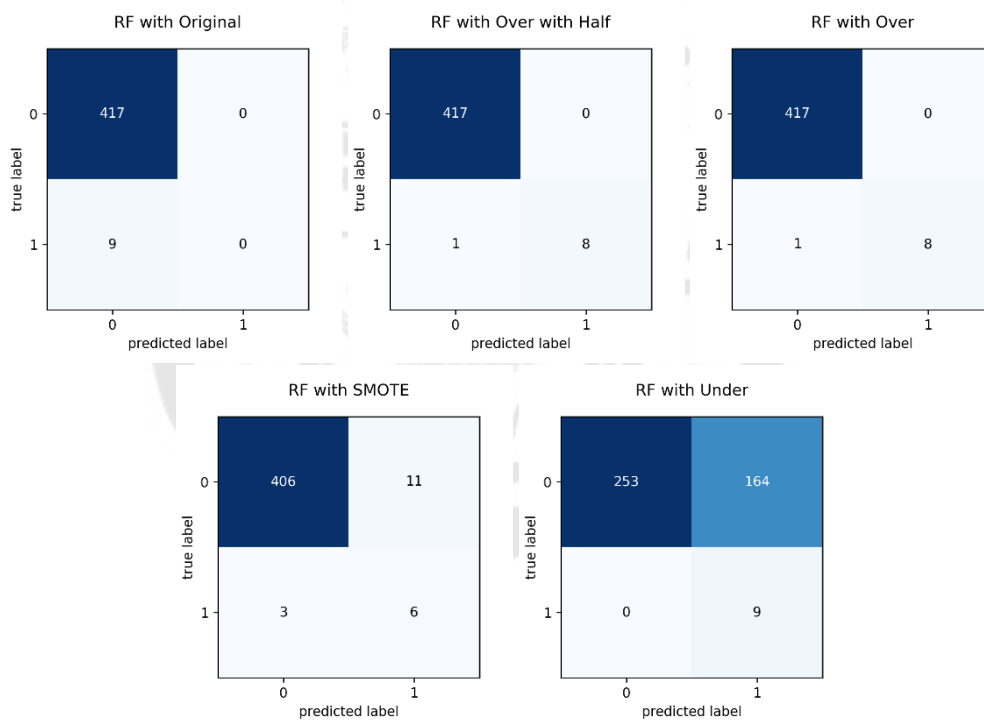


ภาพประกอบ 111 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.6150	0.6067	1.0000	0.0520	0.0989
Over	0.9977	1.0000	0.8889	1.0000	0.9412
Over with Half	0.9977	1.0000	0.8889	1.0000	0.9412
SMOTE	0.9671	0.9736	0.6667	0.3529	0.4615

ภาพประกอบ 112 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

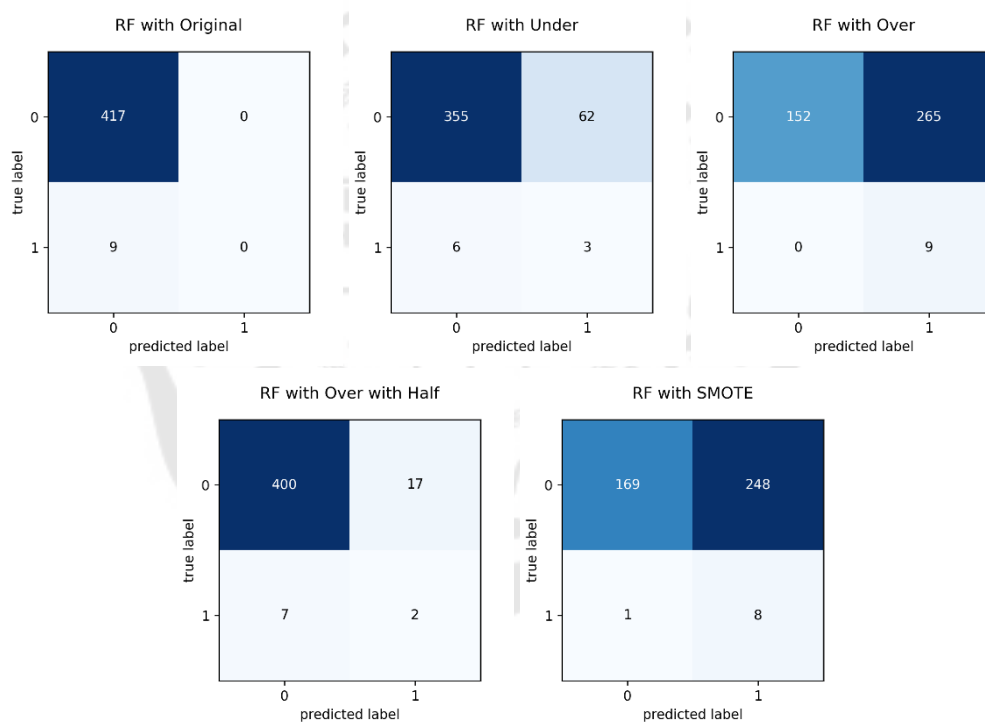


ภาพประกอบ 113 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.5.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.8404	0.8513	0.3333	0.0462	0.0811
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.9437	0.9592	0.2222	0.1053	0.1429
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 114 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 115 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Balanced Class Weight

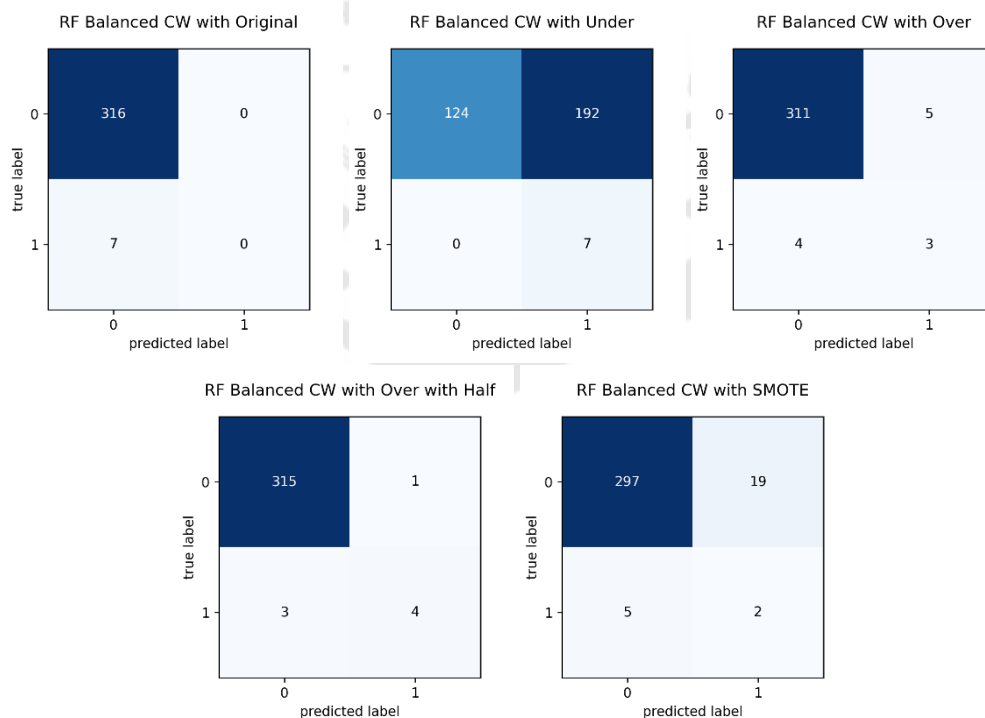
4.6.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.6.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.4056	0.3924	1.0000	0.0352	0.0680
Over	0.9721	0.9842	0.4286	0.3750	0.4000
Over with Half	0.9876	0.9968	0.5714	0.8000	0.6667
SMOTE	0.9257	0.9399	0.2857	0.0952	0.1429

ภาพประกอบ 116 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

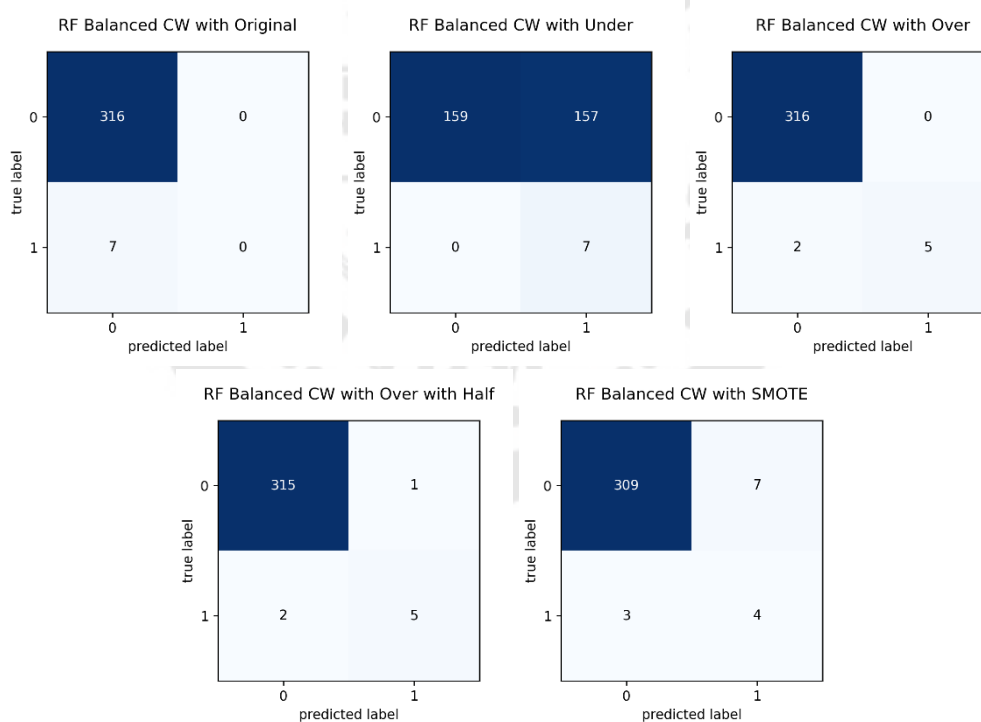


ภาพประกอบ 117 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.5139	0.5032	1.0000	0.0427	0.0819
Over	0.9938	1.0000	0.7143	1.0000	0.8333
Over with Half	0.9907	0.9968	0.7143	0.8333	0.7692
SMOTE	0.9690	0.9778	0.5714	0.3636	0.4444

ภาพประกอบ 118 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



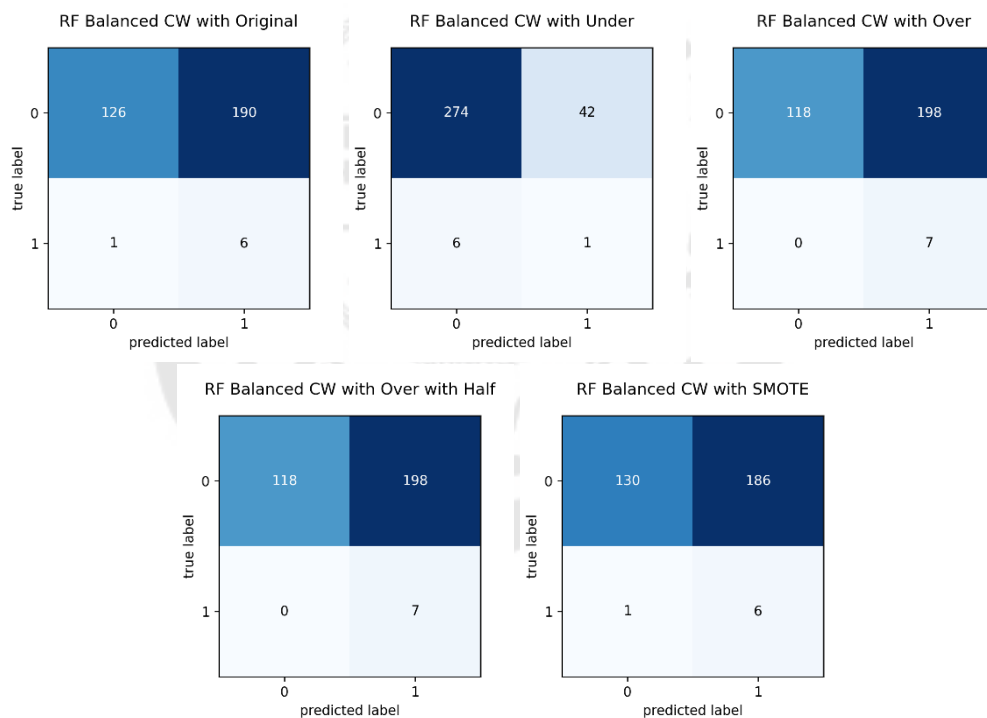
ภาพประกอบ 119 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4087	0.3987	0.8571	0.0306	0.0591
Under	0.8514	0.8671	0.1429	0.0233	0.0400
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3870	0.3734	1.0000	0.0341	0.0660
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 120 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Decision Tree



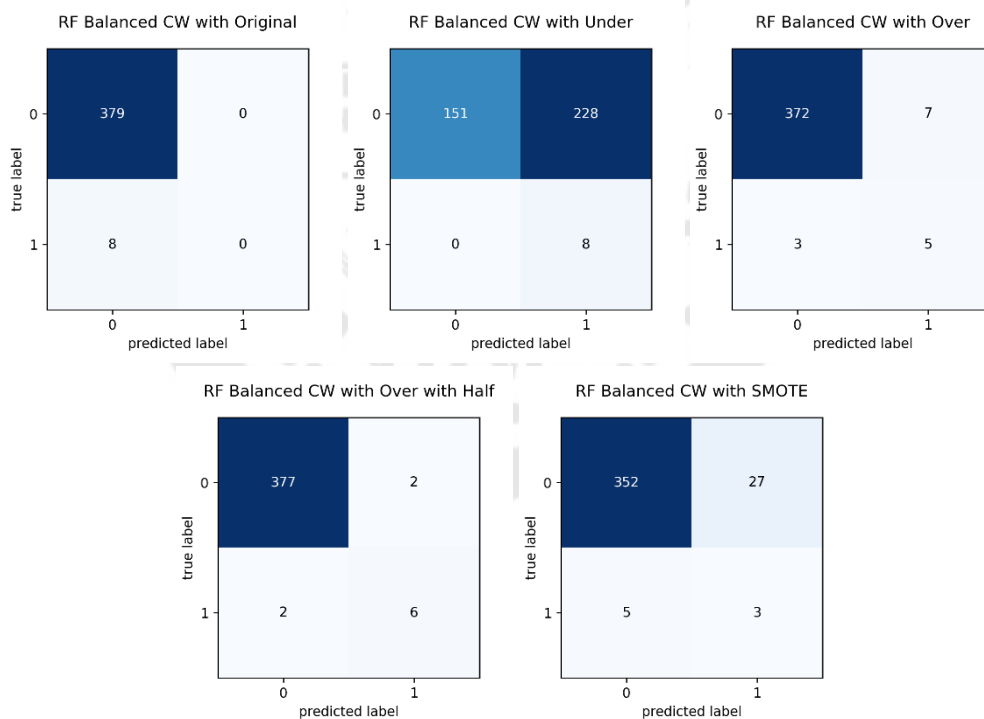
ภาพประกอบ 121 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.6.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4109	0.3984	1.0000	0.0339	0.0656
Over	0.9742	0.9815	0.6250	0.4167	0.5000
Over with Half	0.9897	0.9947	0.7500	0.7500	0.7500
SMOTE	0.9173	0.9288	0.3750	0.1000	0.1579

ภาพประกอบ 122 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

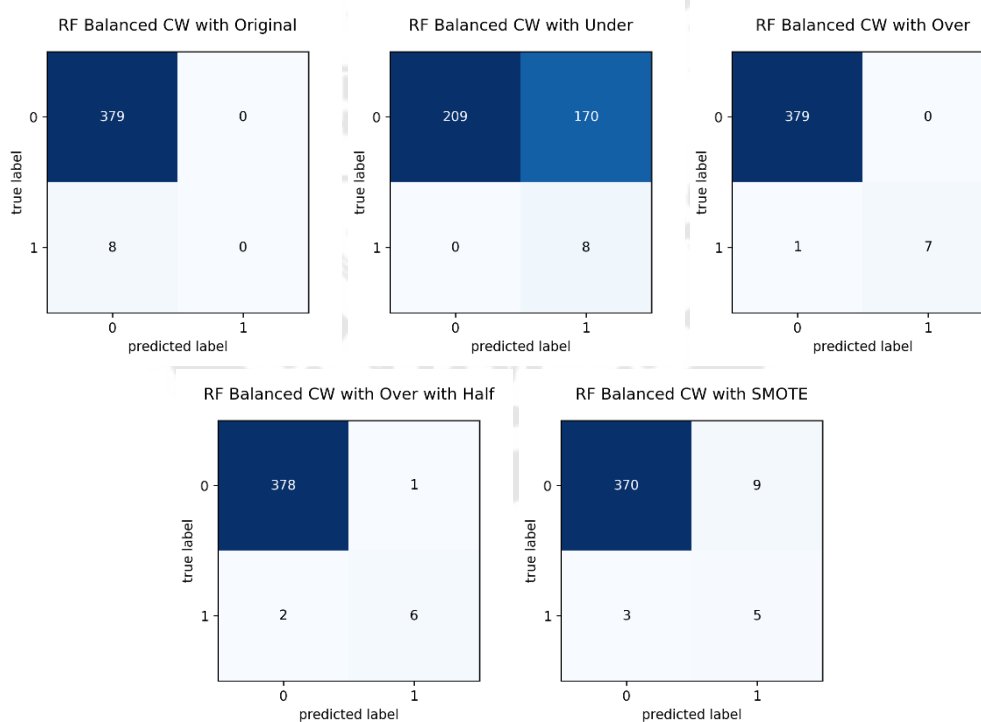


ภาพประกอบ 123 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.5607	0.5515	1.0000	0.0449	0.0860
Over	0.9974	1.0000	0.8750	1.0000	0.9333
Over with Half	0.9922	0.9974	0.7500	0.8571	0.8000
SMOTE	0.9690	0.9763	0.6250	0.3571	0.4545

ภาพประกอบ 124 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



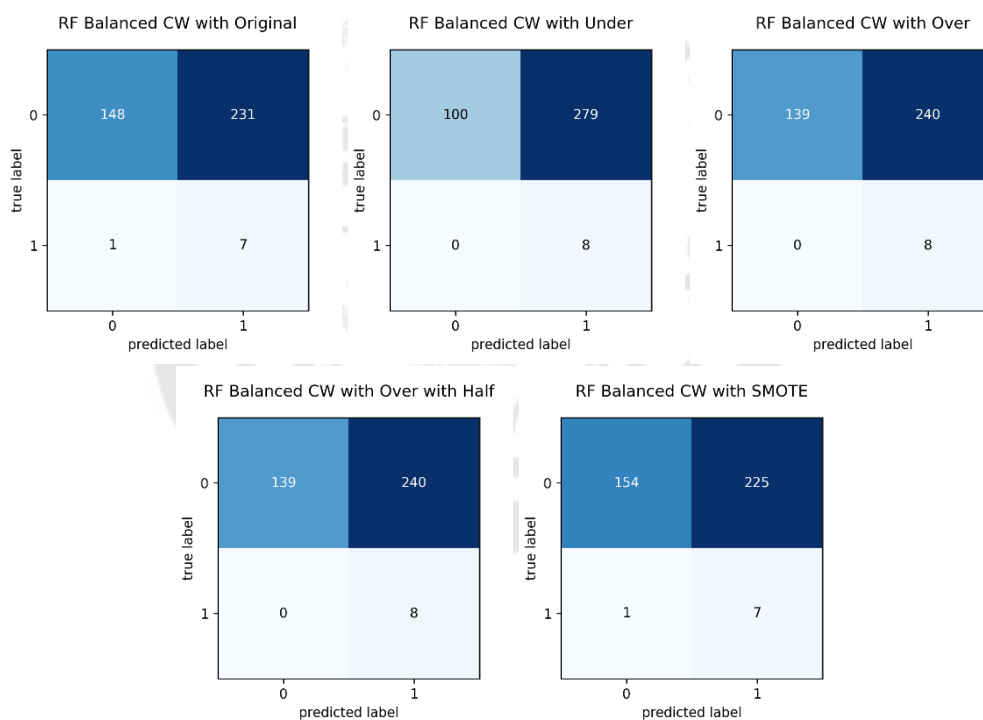
ภาพประกอบ 125 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4005	0.3905	0.8750	0.0294	0.0569
Under	0.2791	0.2639	1.0000	0.0279	0.0542
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 126 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



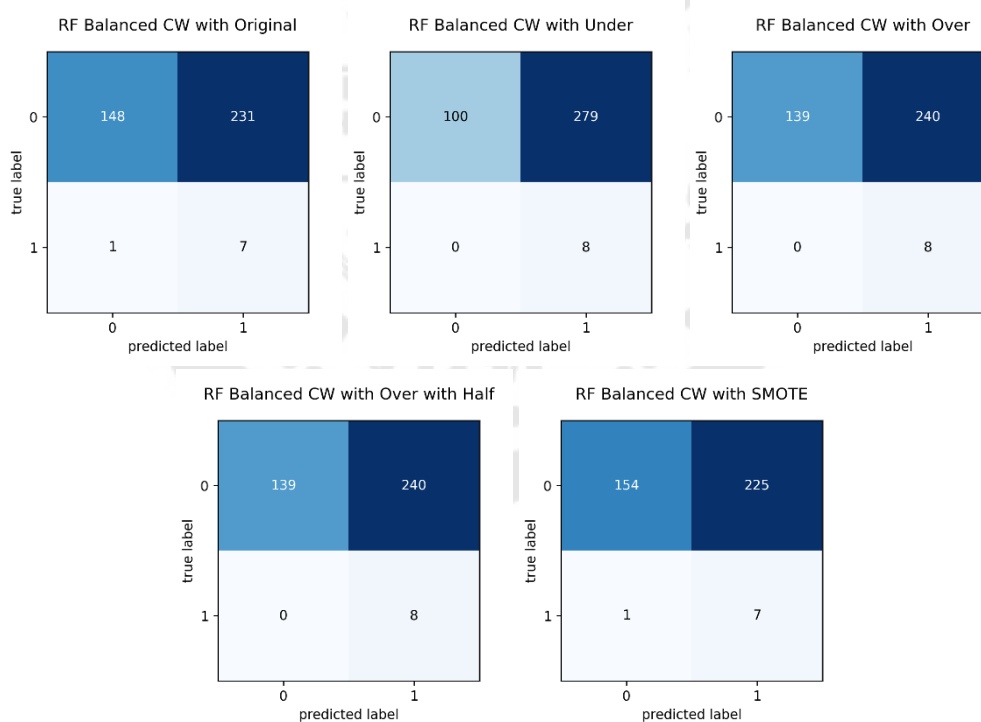
ภาพประกอบ 127 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.6.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4005	0.3905	0.8750	0.0294	0.0569
Under	0.2791	0.2639	1.0000	0.0279	0.0542
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 128 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

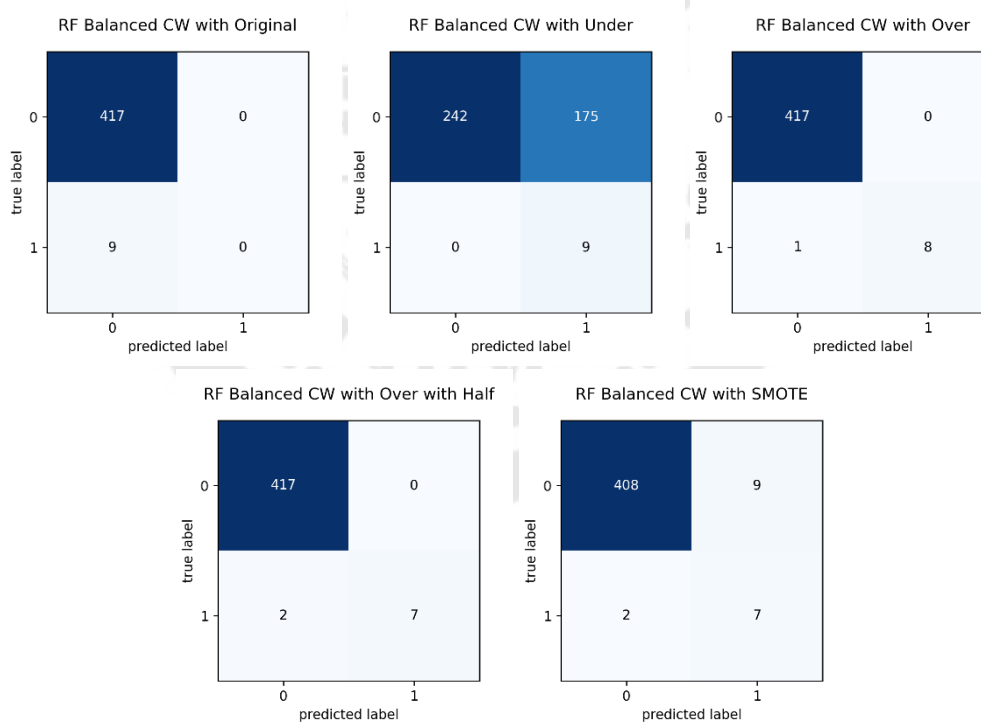


ภาพประกอบ 129 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.5892	0.5803	1.0000	0.0489	0.0933
Over	0.9977	1.0000	0.8889	1.0000	0.9412
Over with Half	0.9953	1.0000	0.7778	1.0000	0.8750
SMOTE	0.9742	0.9784	0.7778	0.4375	0.5600

ภาพประกอบ 130 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



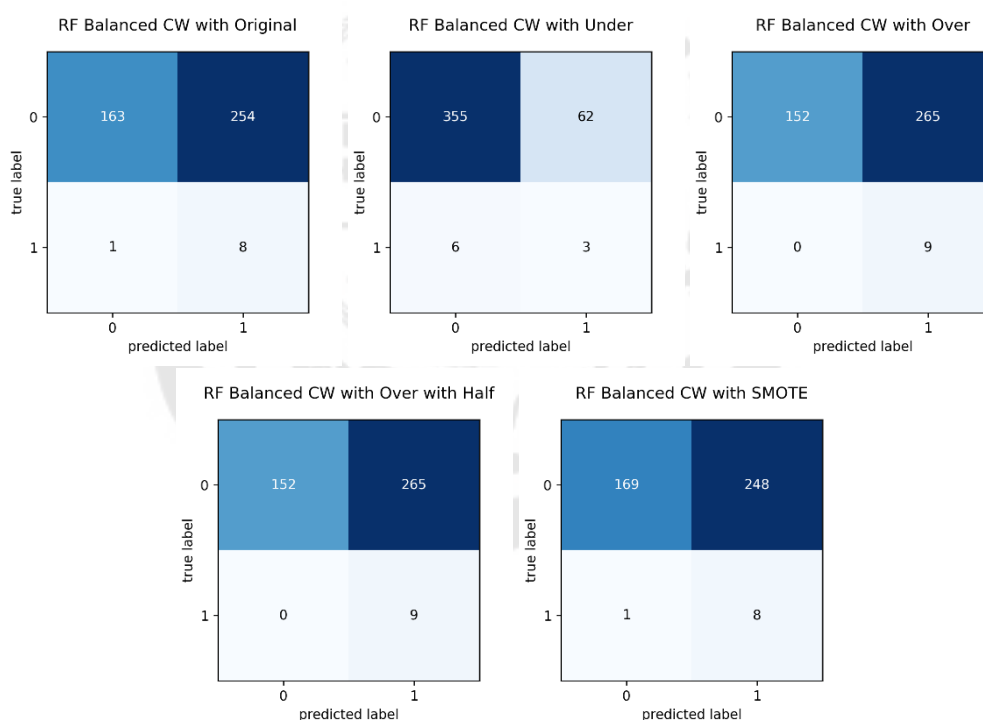
ภาพประกอบ 131 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.6.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4014	0.3909	0.8889	0.0305	0.0590
Under	0.8404	0.8513	0.3333	0.0462	0.0811
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 132 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Decision Tree



ภาพประกอบ 133 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

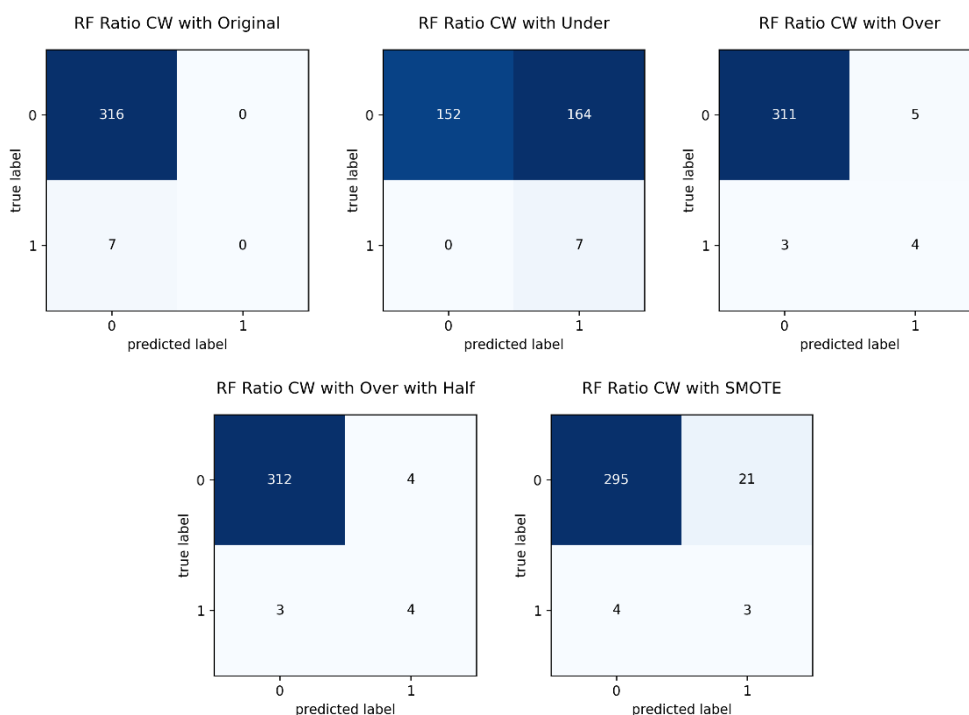
4.7 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Ratio Class Weight

4.7.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.7.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.4923	0.4810	1.0000	0.0409	0.0787
Over	0.9752	0.9842	0.5714	0.4444	0.5000
Over with Half	0.9783	0.9873	0.5714	0.5000	0.5333
SMOTE	0.9226	0.9335	0.4286	0.1250	0.1935

ภาพประกอบ 134 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

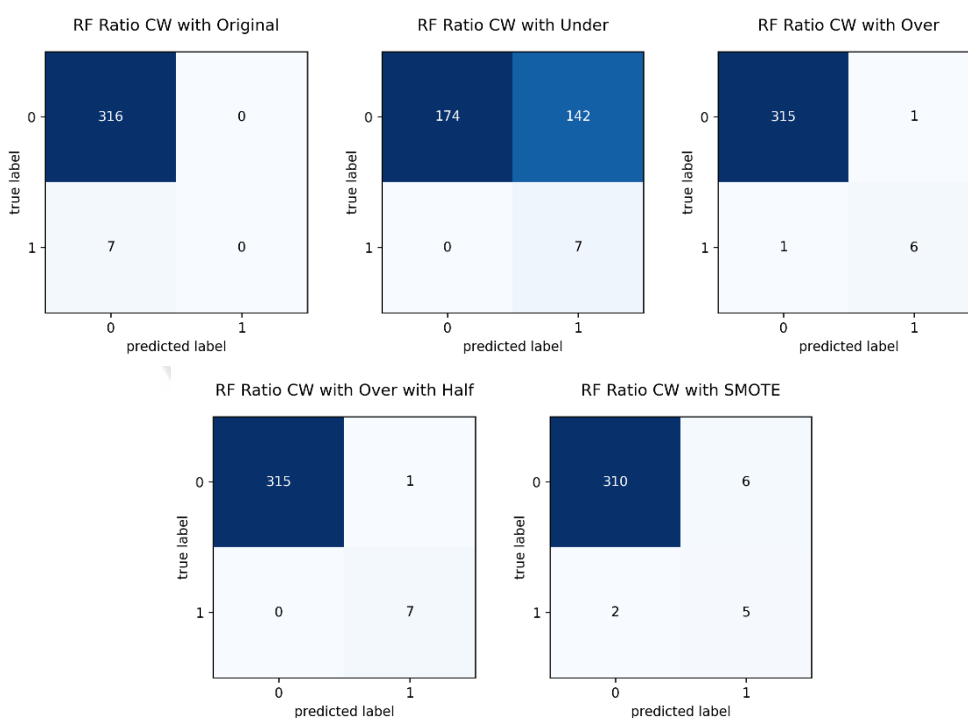


ภาพประกอบ 135 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.5604	0.5506	1.0000	0.0470	0.0897
Over	0.9938	0.9968	0.8571	0.8571	0.8571
Over with Half	0.9969	0.9968	1.0000	0.8750	0.9333
SMOTE	0.9752	0.9810	0.7143	0.4545	0.5556

ภาพประกอบ 136 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

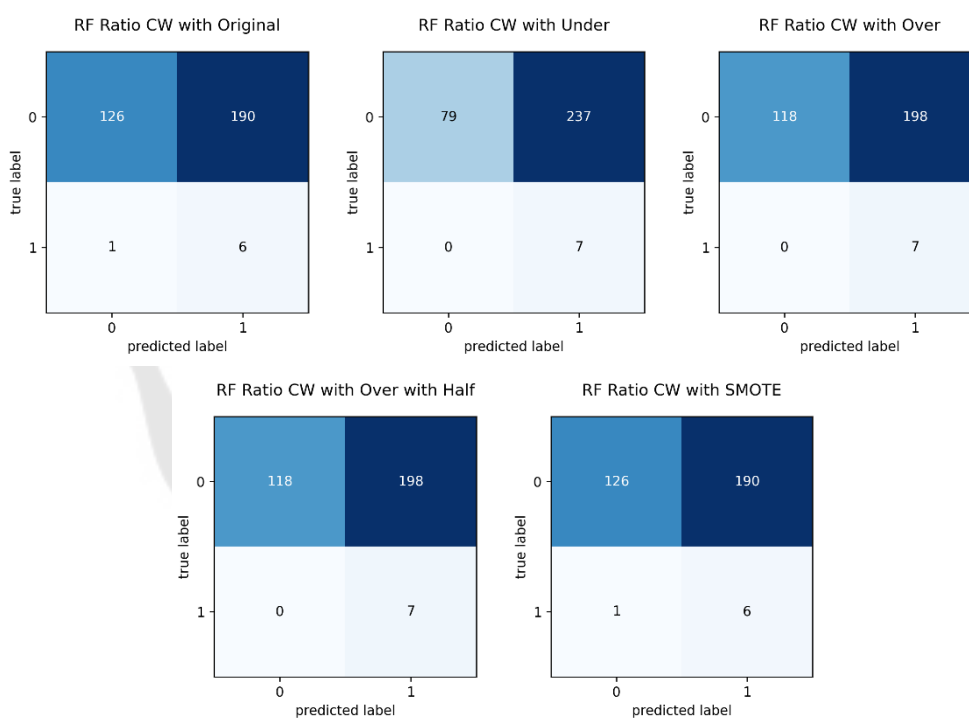


ภาพประกอบ 137 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4087	0.3987	0.8571	0.0306	0.0591
Under	0.2663	0.2500	1.0000	0.0287	0.0558
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3870	0.3734	1.0000	0.0341	0.0660
SMOTE	0.4087	0.3987	0.8571	0.0306	0.0591

ภาพประกอบ 138 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



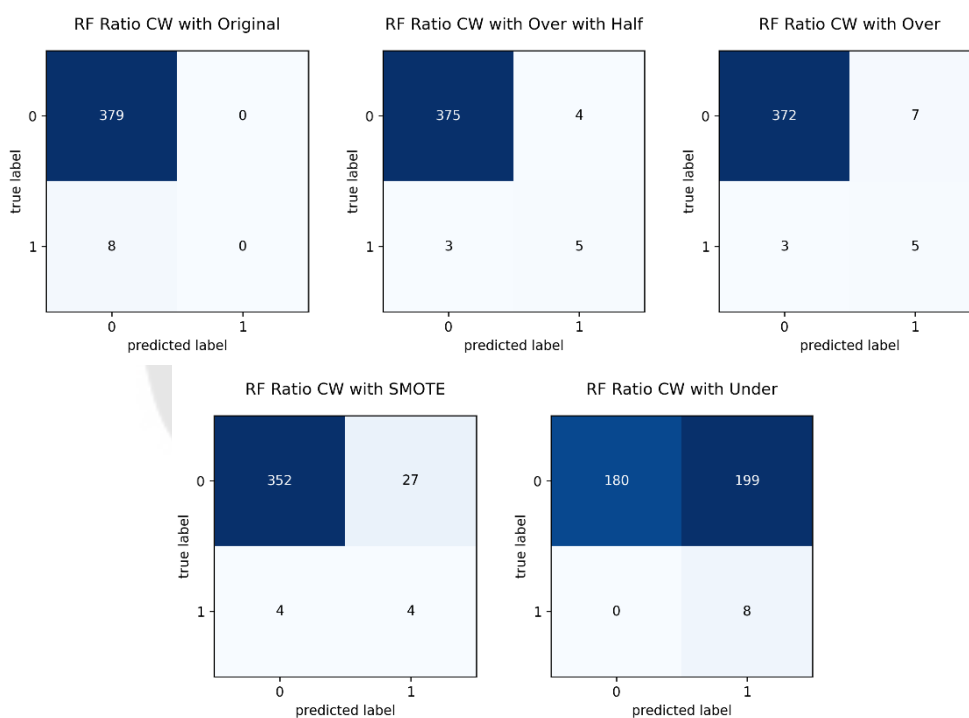
ภาพประกอบ 139 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.7.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4858	0.4749	1.0000	0.0386	0.0744
Over	0.9742	0.9815	0.6250	0.4167	0.5000
Over with Half	0.9819	0.9894	0.6250	0.5556	0.5882
SMOTE	0.9199	0.9288	0.5000	0.1290	0.2051

ภาพประกอบ 140 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

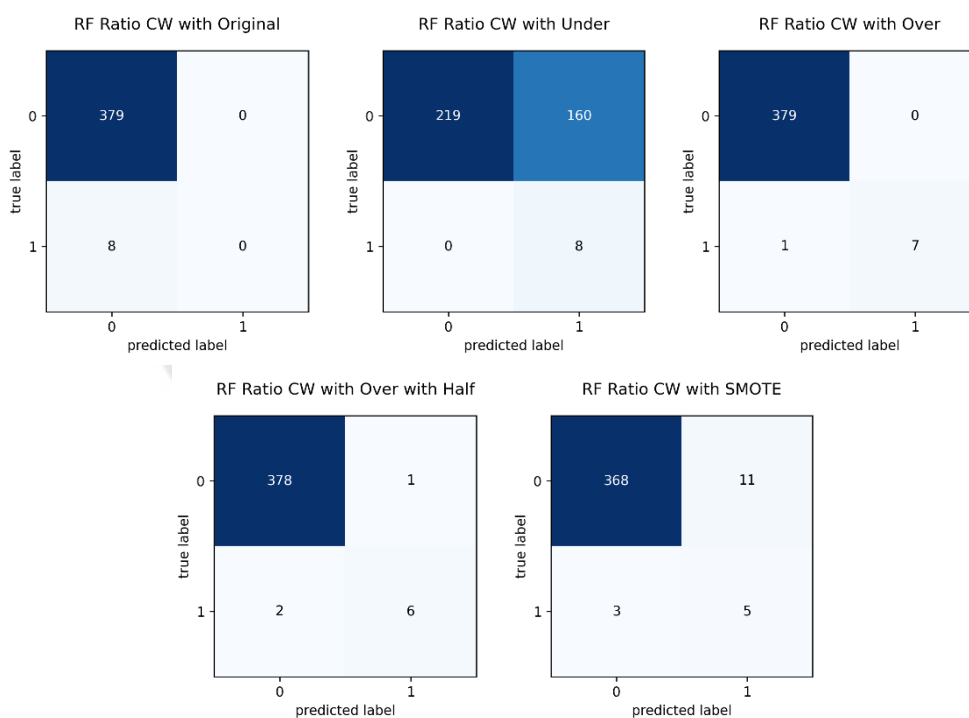


ภาพประกอบ 141 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.5866	0.5778	1.0000	0.0476	0.0909
Over	0.9974	1.0000	0.8750	1.0000	0.9333
Over with Half	0.9922	0.9974	0.7500	0.8571	0.8000
SMOTE	0.9638	0.9710	0.6250	0.3125	0.4167

ภาพประกอบ 142 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

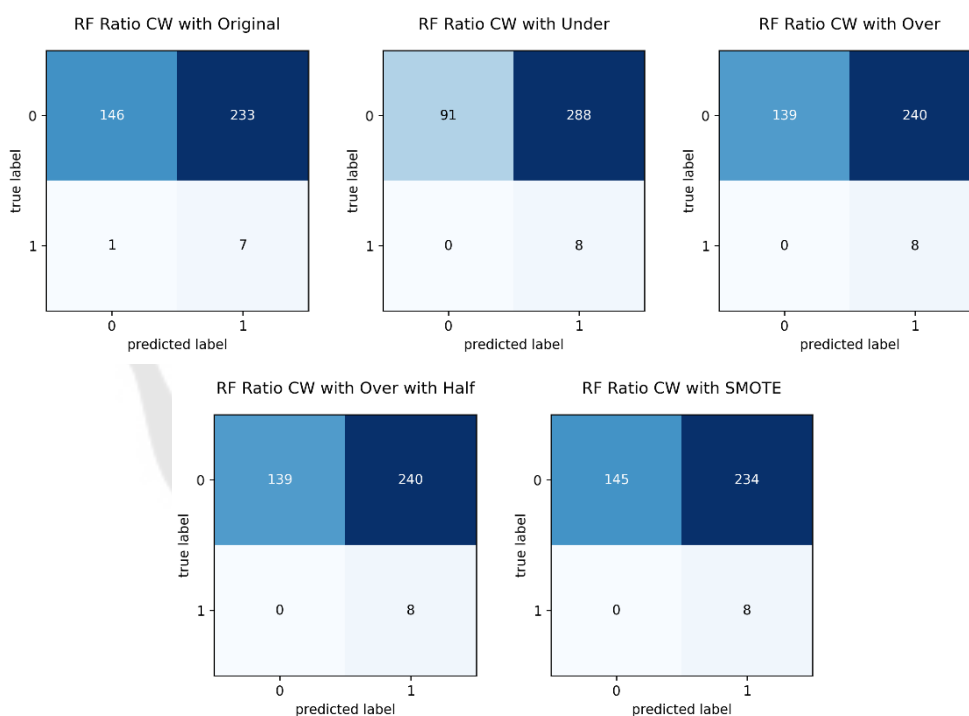


ภาพประกอบ 143 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3953	0.3852	0.8750	0.0292	0.0565
Under	0.2558	0.2401	1.0000	0.0270	0.0526
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.3953	0.3826	1.0000	0.0331	0.0640

ภาพประกอบ 144 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



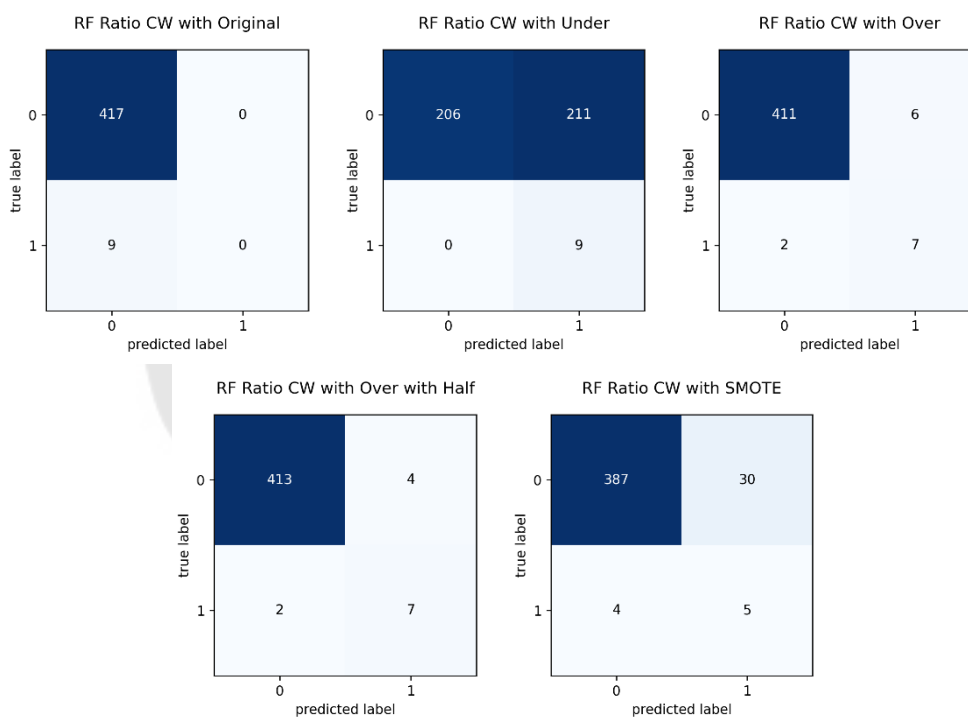
ภาพประกอบ 145 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.7.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.5047	0.4940	1.0000	0.0409	0.0786
Over	0.9812	0.9856	0.7778	0.5385	0.6364
Over with Half	0.9859	0.9904	0.7778	0.6364	0.7000
SMOTE	0.9202	0.9281	0.5556	0.1429	0.2273

ภาพประกอบ 146 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

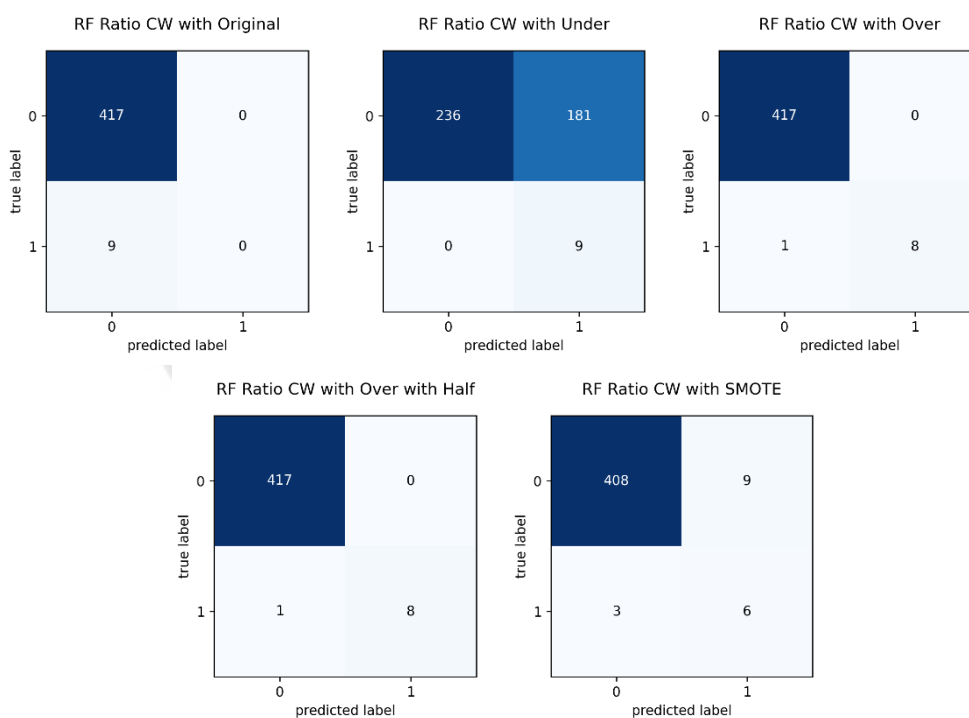


ภาพประกอบ 147 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.5751	0.5659	1.0000	0.0474	0.0905
Over	0.9977	1.0000	0.8889	1.0000	0.9412
Over with Half	0.9977	1.0000	0.8889	1.0000	0.9412
SMOTE	0.9718	0.9784	0.6667	0.4000	0.5000

ภาพประกอบ 148 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

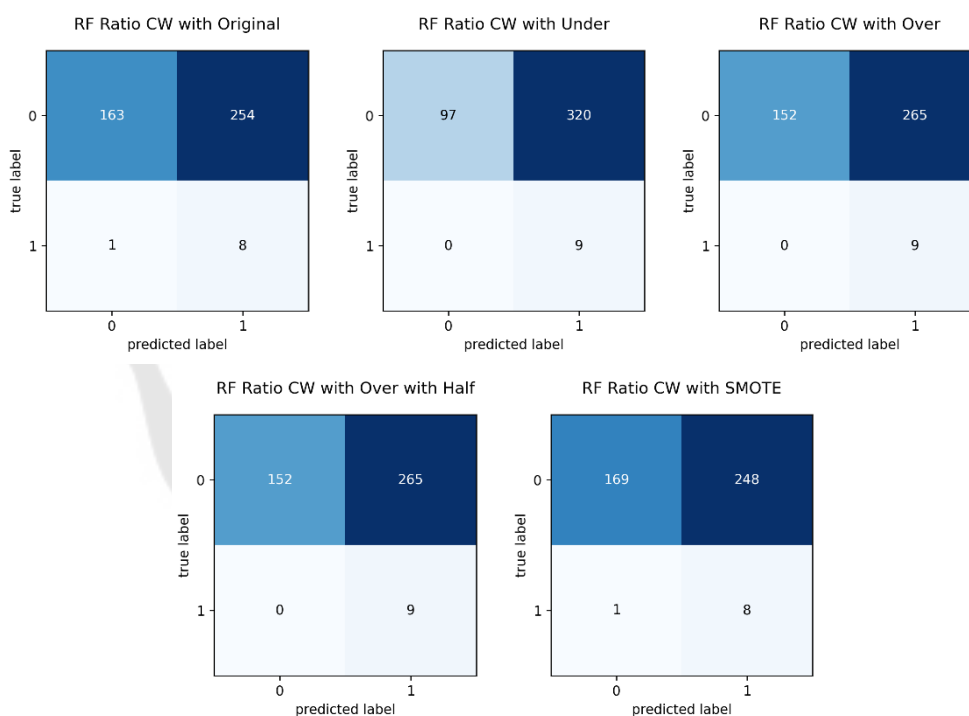


ภาพประกอบ 149 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.7.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4014	0.3909	0.8889	0.0305	0.0590
Under	0.2488	0.2326	1.0000	0.0274	0.0533
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 150 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 151 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

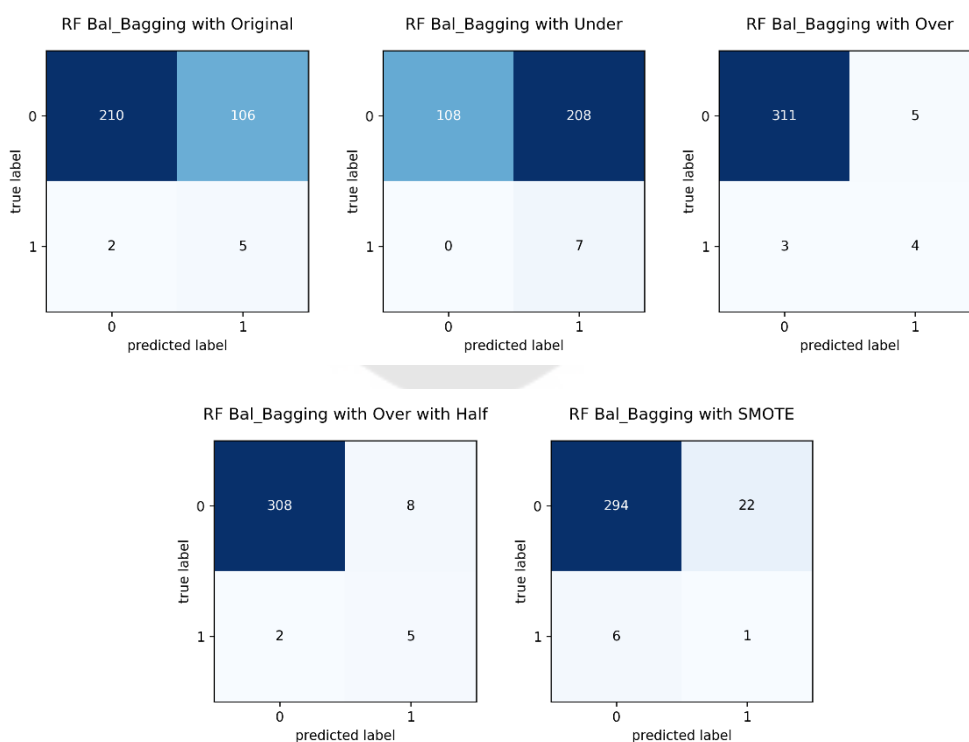
4.8 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Default Parameters และ Balanced Bagging Classifier

4.8.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.8.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6656	0.6646	0.7143	0.0450	0.0847
Under	0.3560	0.3418	1.0000	0.0326	0.0631
Over	0.9752	0.9842	0.5714	0.4444	0.5000
Over with Half	0.9690	0.9747	0.7143	0.3846	0.5000
SMOTE	0.9133	0.9304	0.1429	0.0435	0.0667

ภาพประกอบ 152 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

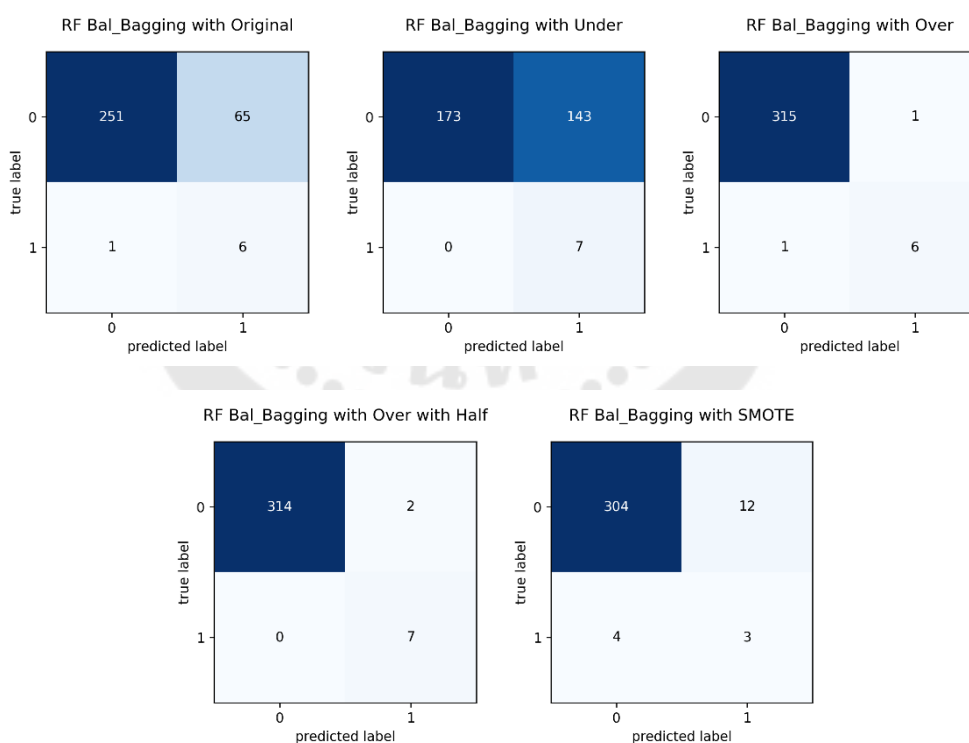


ภาพประกอบ 153 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7957	0.7943	0.8571	0.0845	0.1538
Under	0.5573	0.5475	1.0000	0.0467	0.0892
Over	0.9938	0.9968	0.8571	0.8571	0.8571
Over with Half	0.9938	0.9937	1.0000	0.7778	0.8750
SMOTE	0.9505	0.9620	0.4286	0.2000	0.2727

ภาพประกอบ 154 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



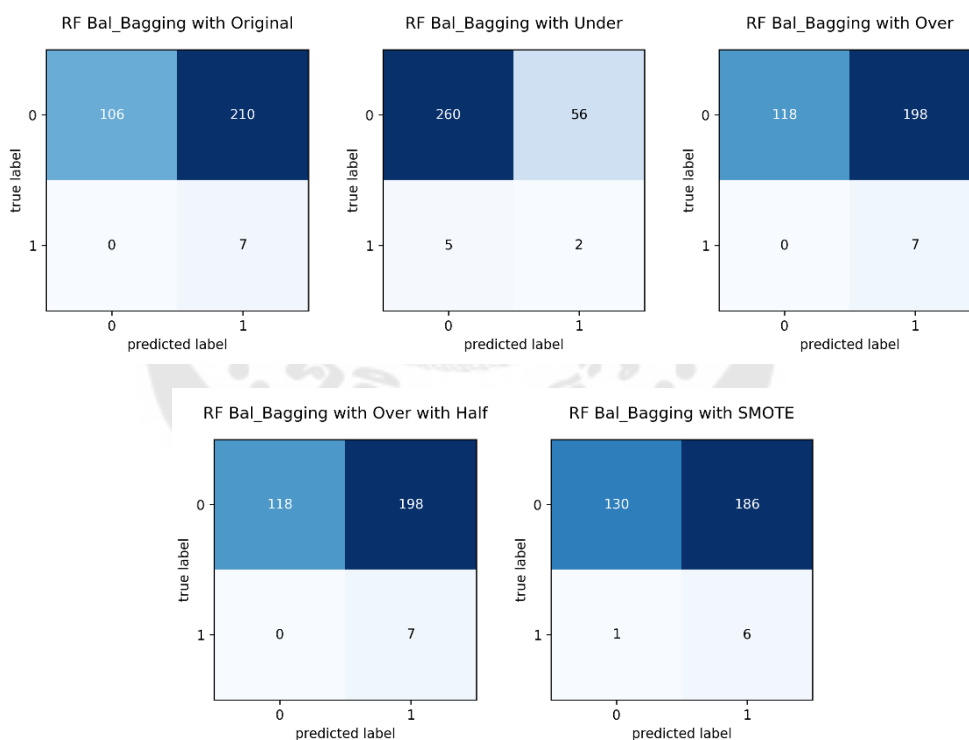
ภาพประกอบ 155 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3498	0.3354	1.0000	0.0323	0.0625
Under	0.8111	0.8228	0.2857	0.0345	0.0615
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3870	0.3734	1.0000	0.0341	0.0660
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 156 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Decision Tree



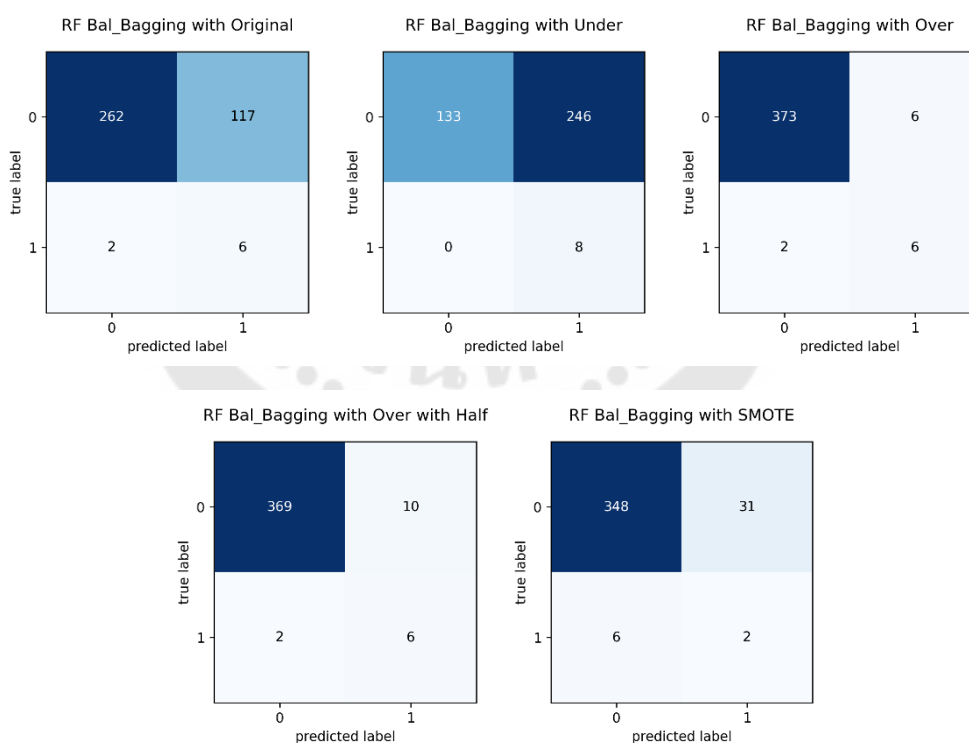
ภาพประกอบ 157 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.8.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6925	0.6913	0.7500	0.0488	0.0916
Under	0.3643	0.3509	1.0000	0.0315	0.0611
Over	0.9793	0.9842	0.7500	0.5000	0.6000
Over with Half	0.9690	0.9736	0.7500	0.3750	0.5000
SMOTE	0.9044	0.9182	0.2500	0.0606	0.0976

ภาพประกอบ 158 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

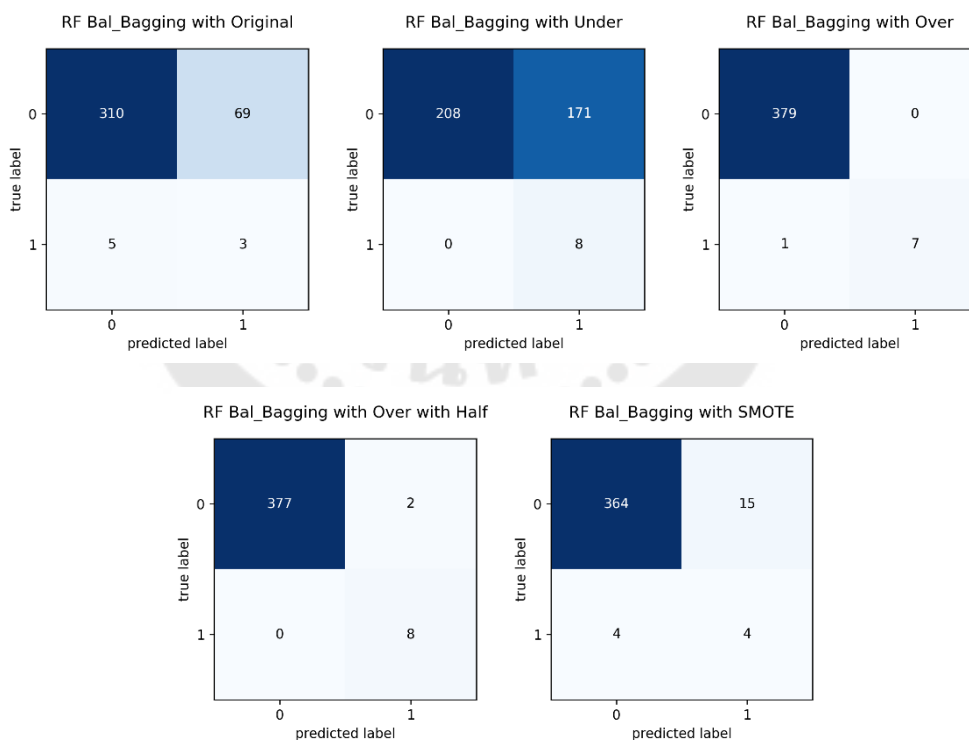


ภาพประกอบ 159 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8088	0.8179	0.3750	0.0417	0.0750
Under	0.5581	0.5488	1.0000	0.0447	0.0856
Over	0.9974	1.0000	0.8750	1.0000	0.9333
Over with Half	0.9948	0.9947	1.0000	0.8000	0.8889
SMOTE	0.9509	0.9604	0.5000	0.2105	0.2963

ภาพประกอบ 160 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



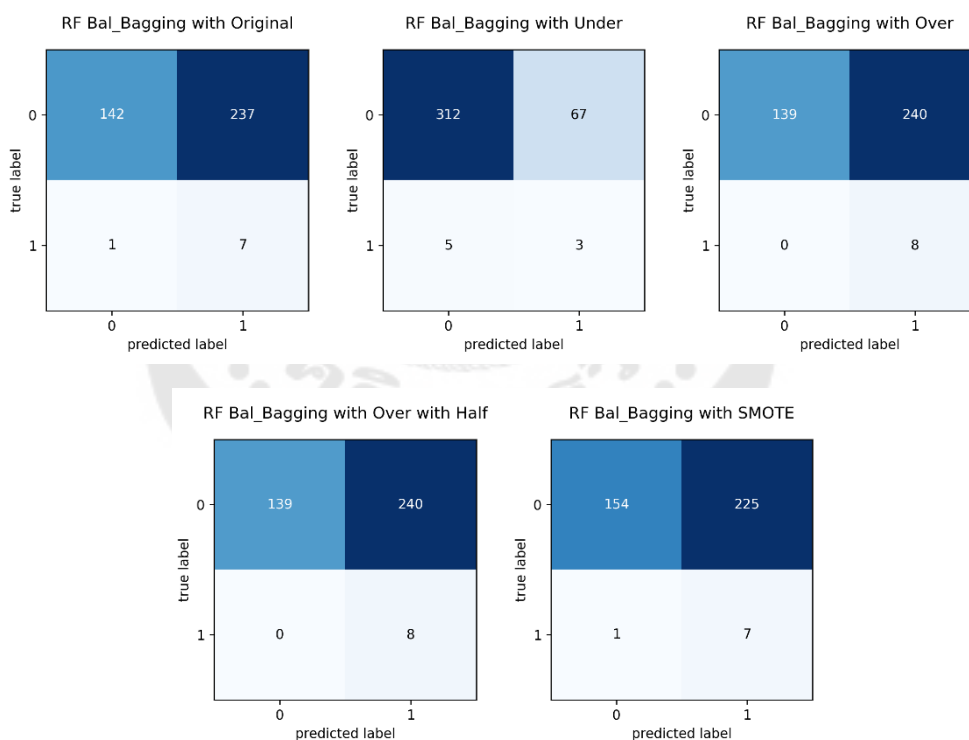
ภาพประกอบ 161 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3850	0.3747	0.8750	0.0287	0.0556
Under	0.8140	0.8232	0.3750	0.0429	0.0769
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 162 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



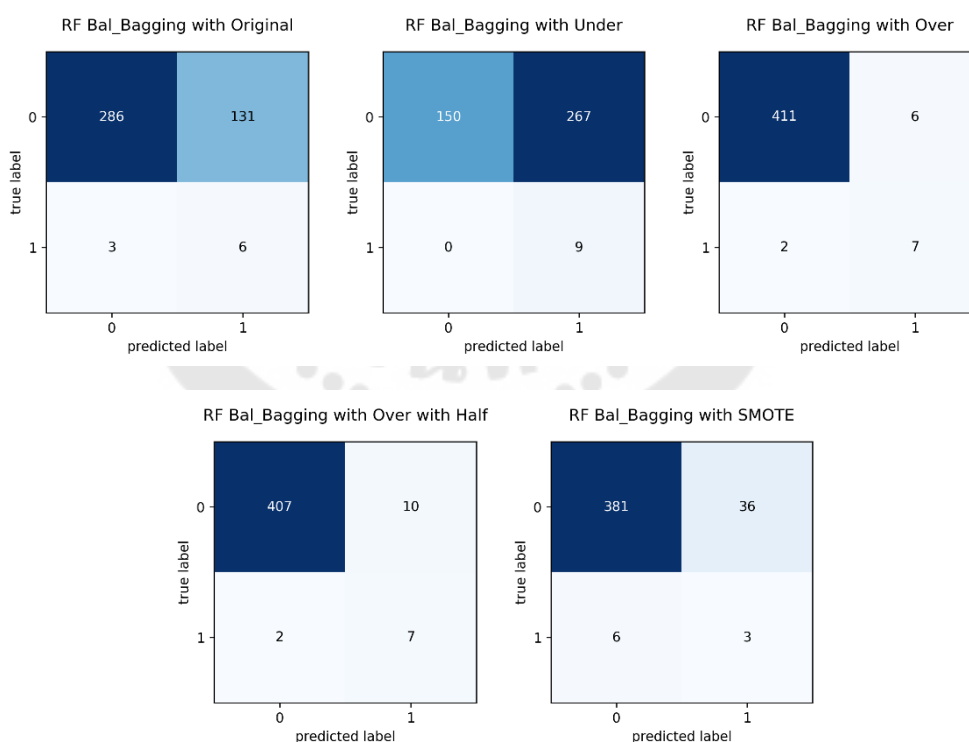
ภาพประกอบ 163 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.8.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6854	0.6859	0.6667	0.0438	0.0822
Under	0.3732	0.3597	1.0000	0.0326	0.0632
Over	0.9812	0.9856	0.7778	0.5385	0.6364
Over with Half	0.9718	0.9760	0.7778	0.4118	0.5385
SMOTE	0.9014	0.9137	0.3333	0.0769	0.1250

ภาพประกอบ 164 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

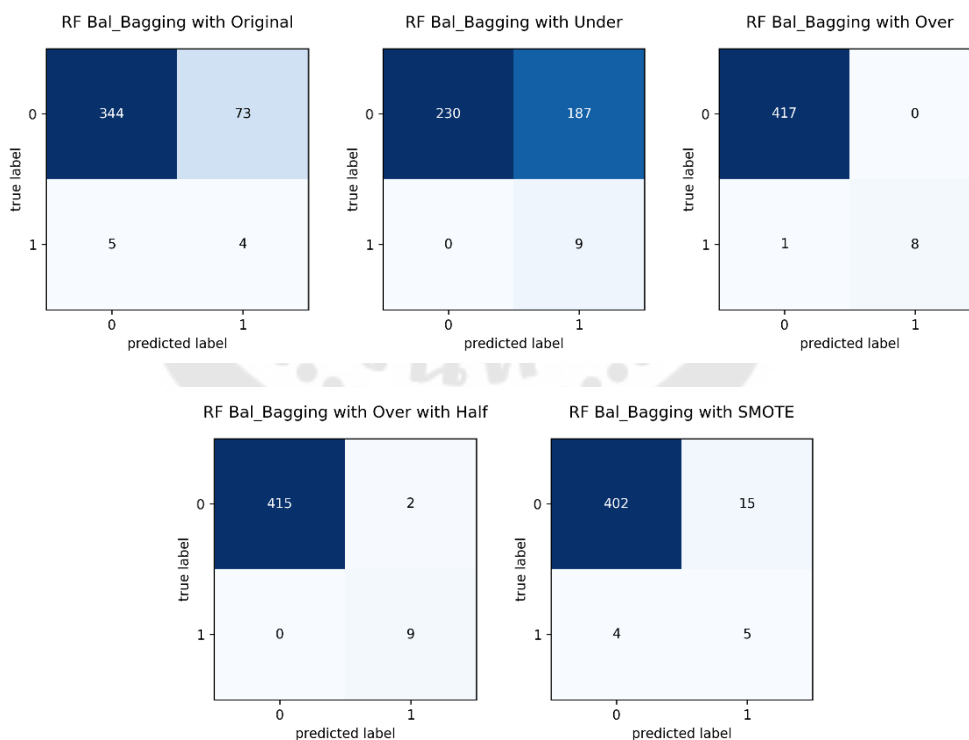


ภาพประกอบ 165 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8169	0.8249	0.4444	0.0519	0.0930
Under	0.5610	0.5516	1.0000	0.0459	0.0878
Over	0.9977	1.0000	0.8889	1.0000	0.9412
Over with Half	0.9953	0.9952	1.0000	0.8182	0.9000
SMOTE	0.9554	0.9640	0.5556	0.2500	0.3448

ภาพประกอบ 166 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



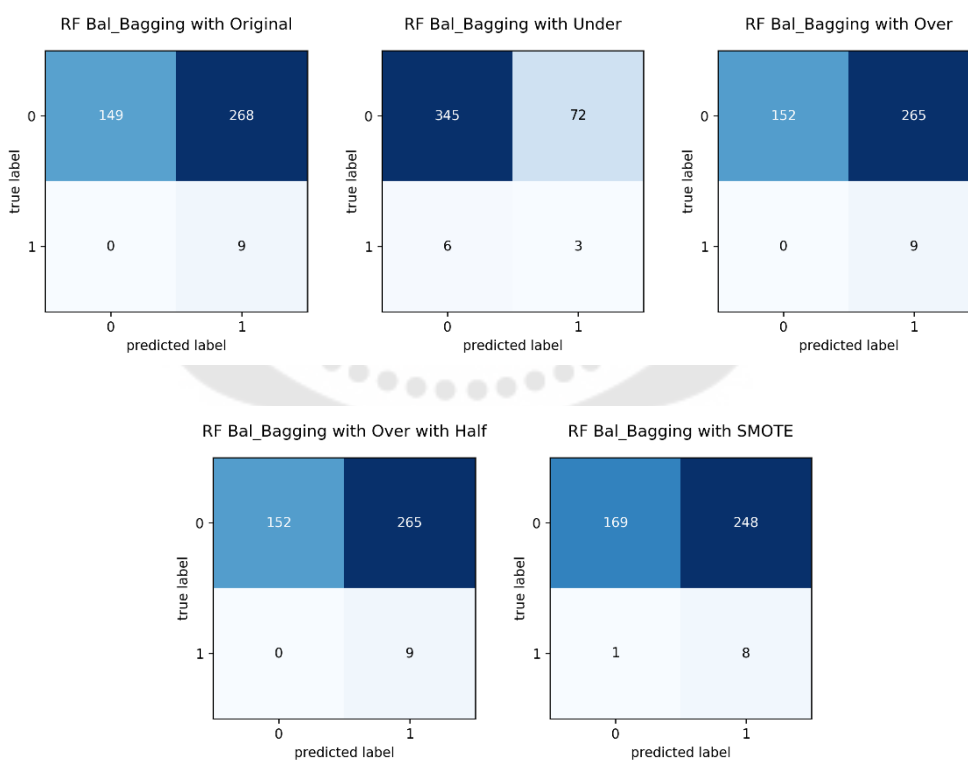
ภาพประกอบ 167 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.8.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3709	0.3573	1.0000	0.0325	0.0629
Under	0.8169	0.8273	0.3333	0.0400	0.0714
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 168 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Decision Tree



ภาพประกอบ 169 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

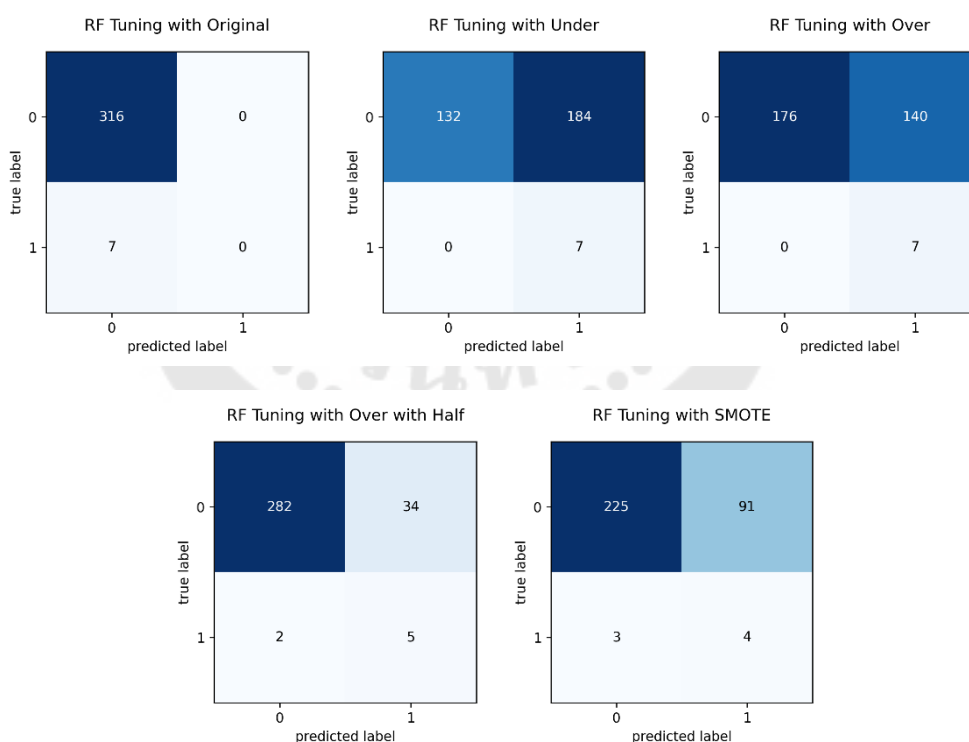
4.9 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters

4.9.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.9.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.4303	0.4177	1.0000	0.0366	0.0707
Over	0.5666	0.5570	1.0000	0.0476	0.0909
Over with Half	0.8885	0.8924	0.7143	0.1282	0.2174
SMOTE	0.7090	0.7120	0.5714	0.0421	0.0784

ภาพประกอบ 170 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

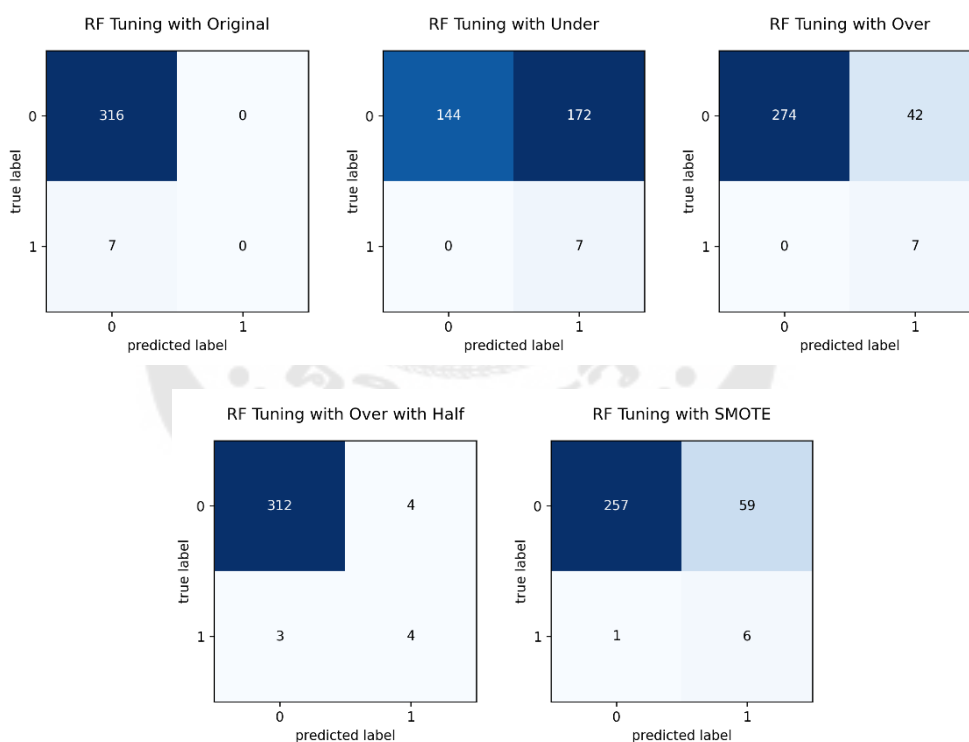


ภาพประกอบ 171 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.4675	0.4557	1.0000	0.0391	0.0753
Over	0.8700	0.8671	1.0000	0.1429	0.2500
Over with Half	0.9783	0.9873	0.5714	0.5000	0.5333
SMOTE	0.8142	0.8133	0.8571	0.0923	0.1667

ภาพประกอบ 172 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

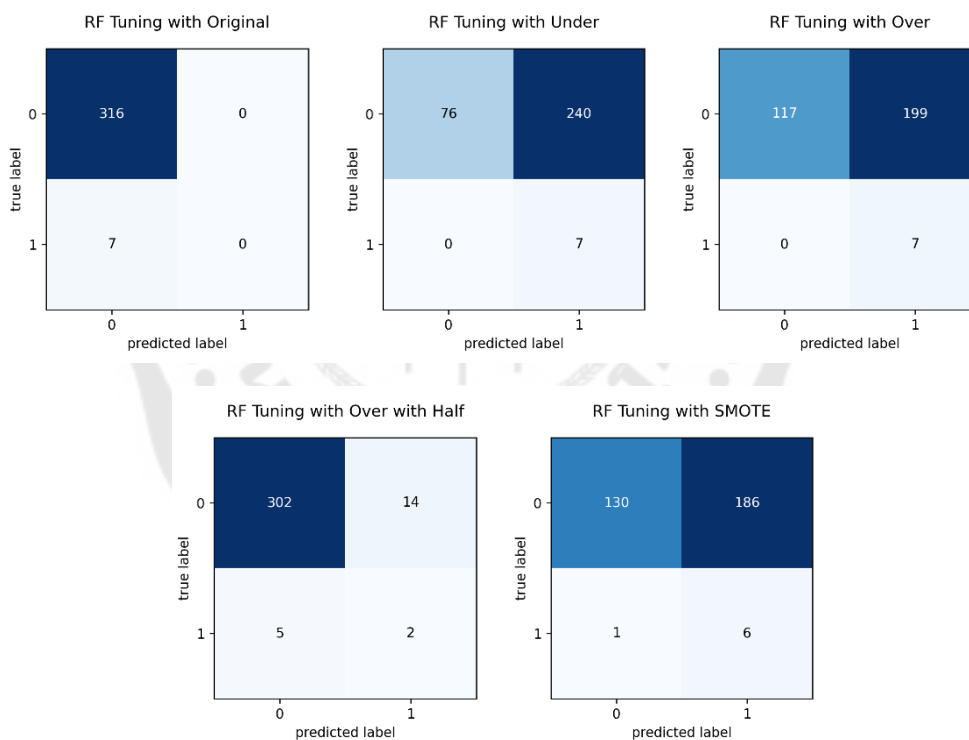


ภาพประกอบ 173 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.2570	0.2405	1.0000	0.0283	0.0551
Over	0.3839	0.3703	1.0000	0.0340	0.0657
Over with Half	0.9412	0.9557	0.2857	0.1250	0.1739
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 174 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



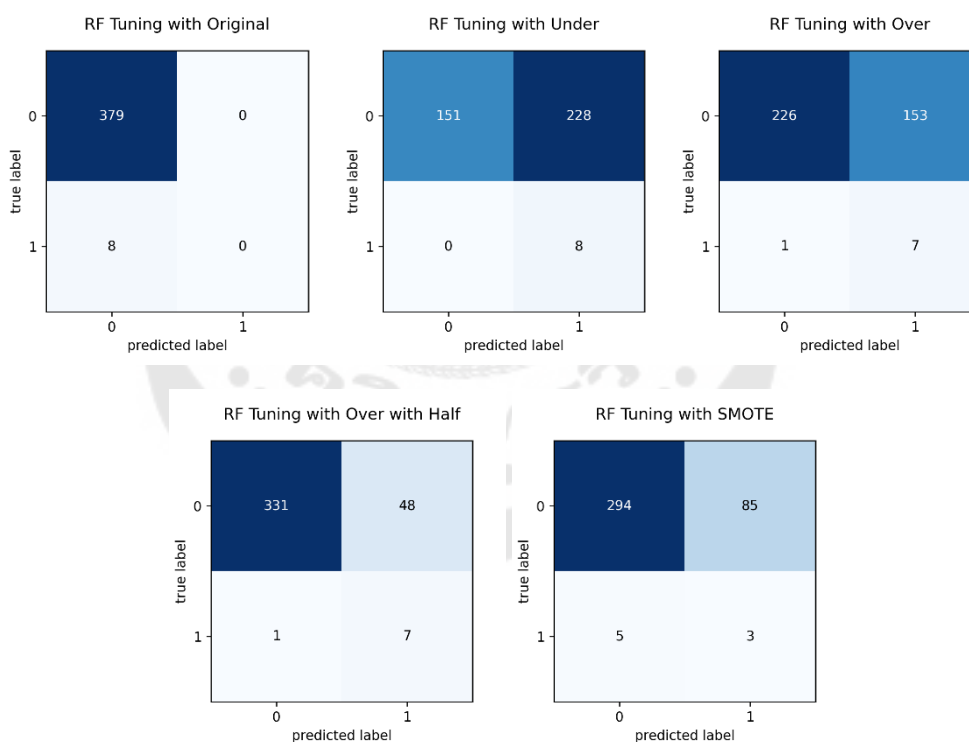
ภาพประกอบ 175 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.9.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4109	0.3984	1.0000	0.0339	0.0656
Over	0.6021	0.5963	0.8750	0.0437	0.0833
Over with Half	0.8734	0.8734	0.8750	0.1273	0.2222
SMOTE	0.7674	0.7757	0.3750	0.0341	0.0625

ภาพประกอบ 176 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

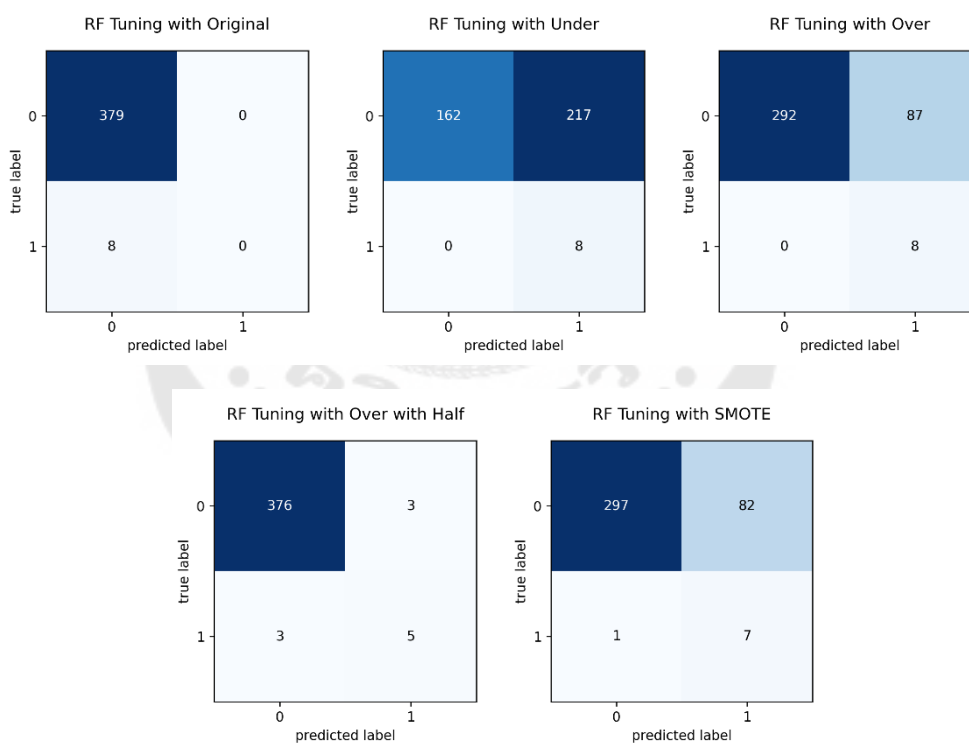


ภาพประกอบ 177 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.4393	0.4274	1.0000	0.0356	0.0687
Over	0.7752	0.7704	1.0000	0.0842	0.1553
Over with Half	0.9845	0.9921	0.6250	0.6250	0.6250
SMOTE	0.7855	0.7836	0.8750	0.0787	0.1443

ภาพประกอบ 178 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

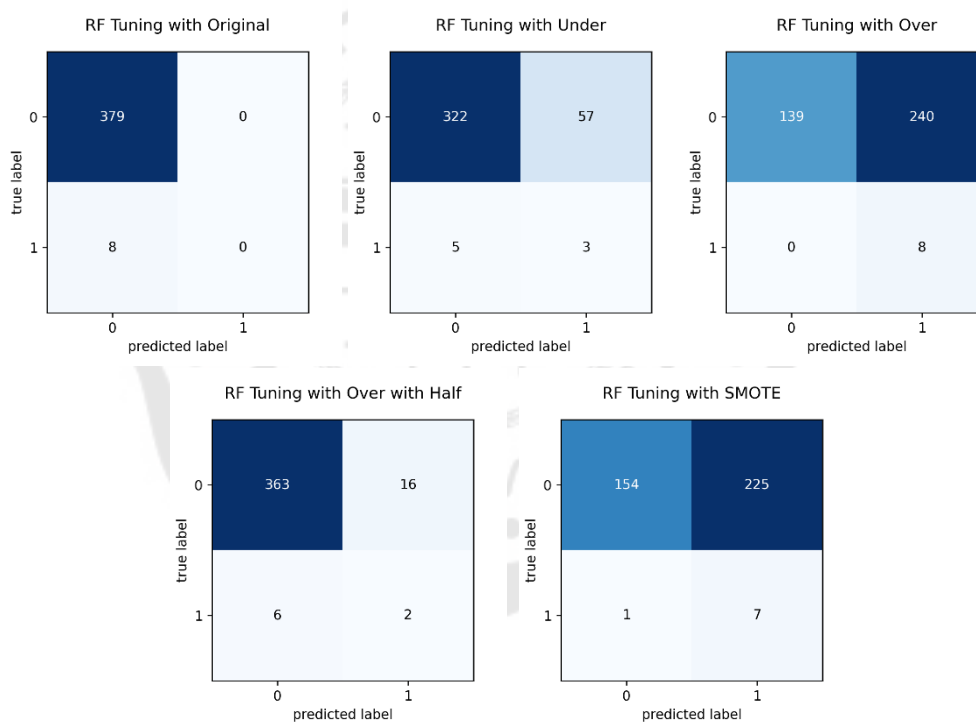


ภาพประกอบ 179 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.8398	0.8496	0.3750	0.0500	0.0882
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.9432	0.9578	0.2500	0.1111	0.1538
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 180 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



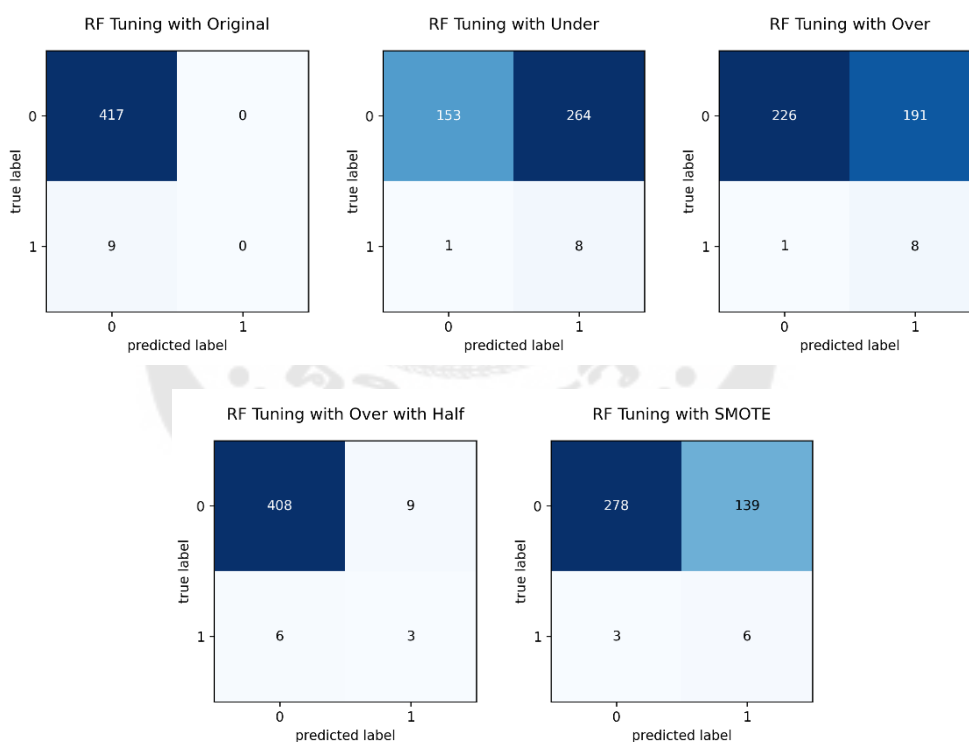
ภาพประกอบ 181 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.9.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.3779	0.3669	0.8889	0.0294	0.0569
Over	0.5493	0.5420	0.8889	0.0402	0.0769
Over with Half	0.9648	0.9784	0.3333	0.2500	0.2857
SMOTE	0.6667	0.6667	0.6667	0.0414	0.0779

ภาพประกอบ 182 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

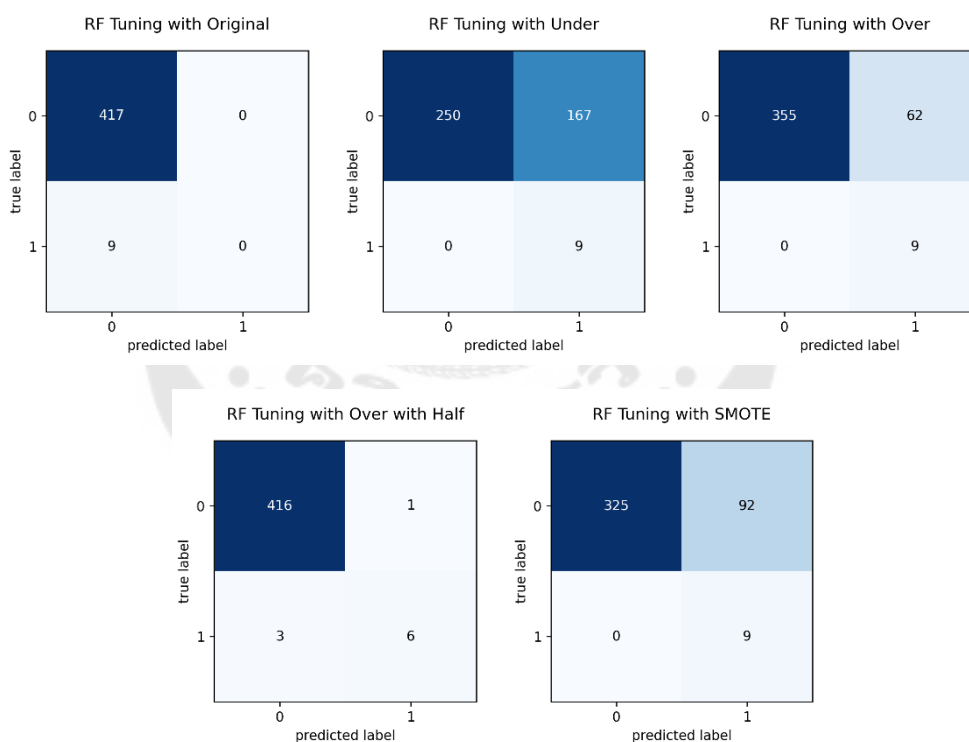


ภาพประกอบ 183 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.6080	0.5995	1.0000	0.0511	0.0973
Over	0.8545	0.8513	1.0000	0.1268	0.2250
Over with Half	0.9906	0.9976	0.6667	0.8571	0.7500
SMOTE	0.7840	0.7794	1.0000	0.0891	0.1636

ภาพประกอบ 184 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

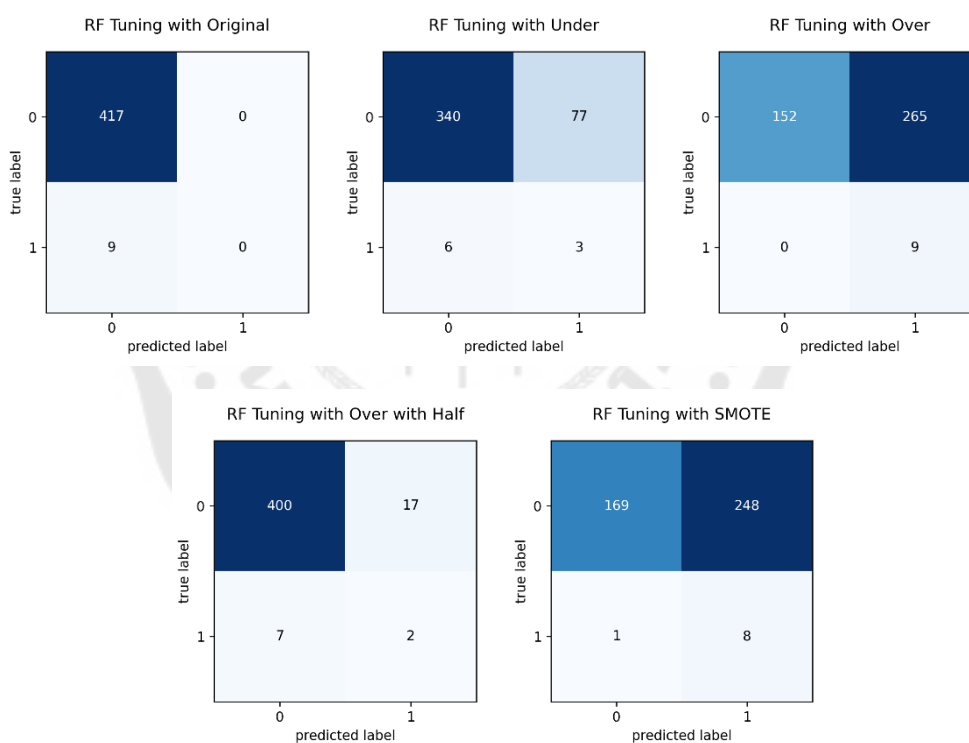


ภาพประกอบ 185 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.9.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.8052	0.8153	0.3333	0.0375	0.0674
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.9437	0.9592	0.2222	0.1053	0.1429
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 186 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 187 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight

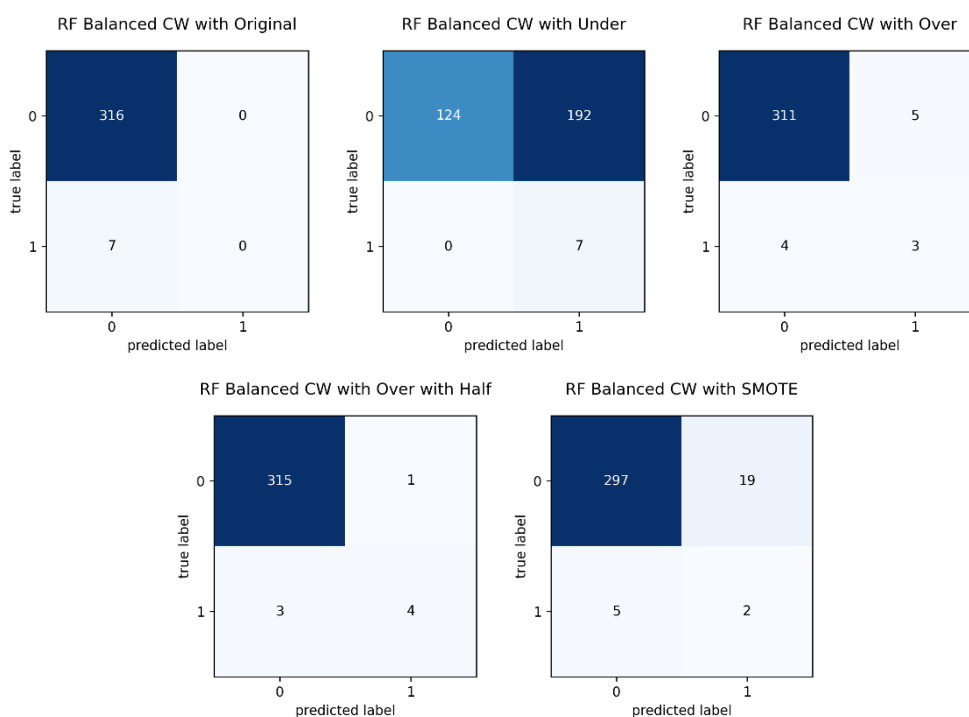
4.10.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.10.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7895	0.8006	0.2857	0.0308	0.0556
Under	0.3839	0.3703	1.0000	0.0340	0.0657
Over	0.5666	0.5601	0.8571	0.0414	0.0789
Over with Half	0.5944	0.5886	0.8571	0.0441	0.0839
SMOTE	0.7276	0.7310	0.5714	0.0449	0.0833

ภาพประกอบ 188 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

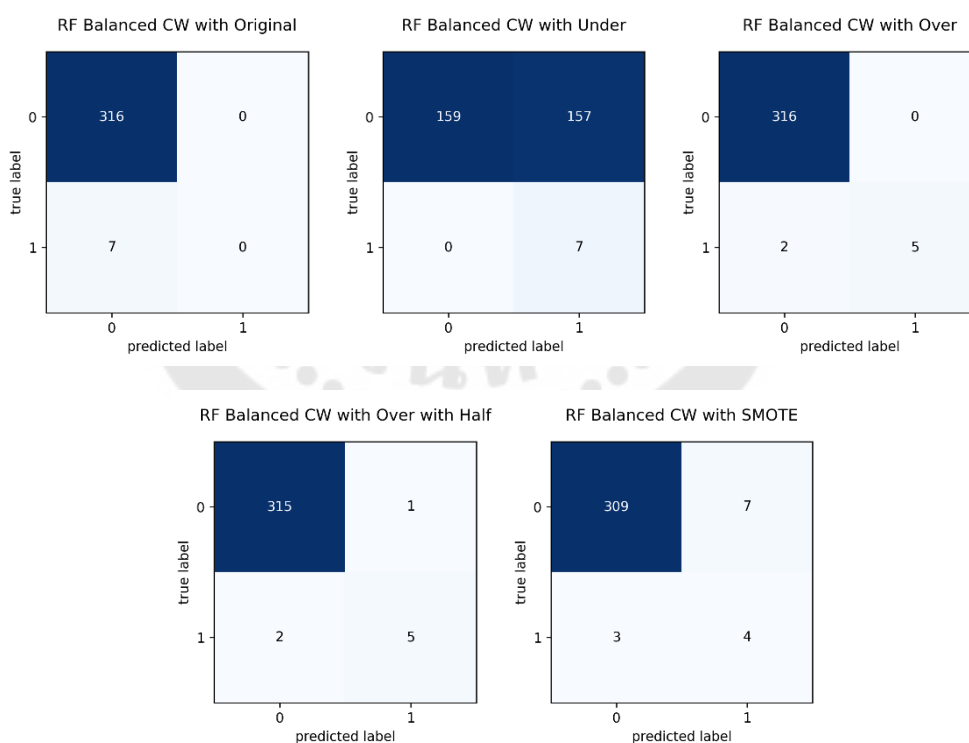


ภาพประกอบ 189 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9690	0.9905	0.0000	0.0000	0.0000
Under	0.4427	0.4304	1.0000	0.0374	0.0722
Over	0.8700	0.8671	1.0000	0.1429	0.2500
Over with Half	0.8700	0.8671	1.0000	0.1429	0.2500
SMOTE	0.8235	0.8259	0.7143	0.0833	0.1493

ภาพประกอบ 190 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



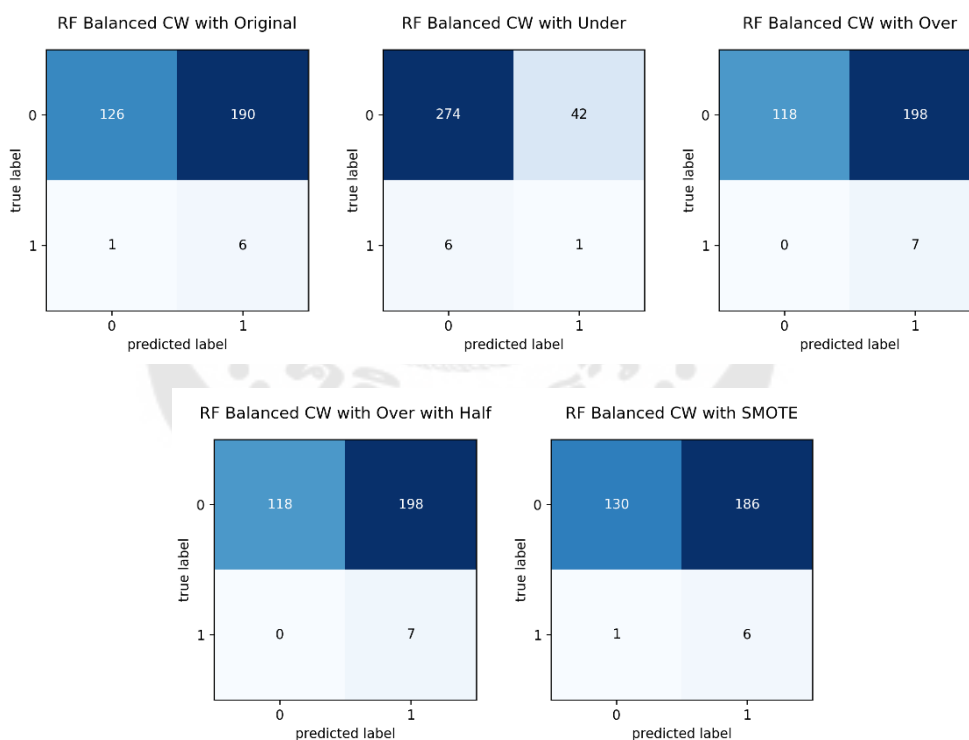
ภาพประกอบ 191 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4180	0.4082	0.8571	0.0311	0.0600
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3808	0.3671	1.0000	0.0338	0.0654
Over with Half	0.3839	0.3703	1.0000	0.0340	0.0657
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 192 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Decision Tree



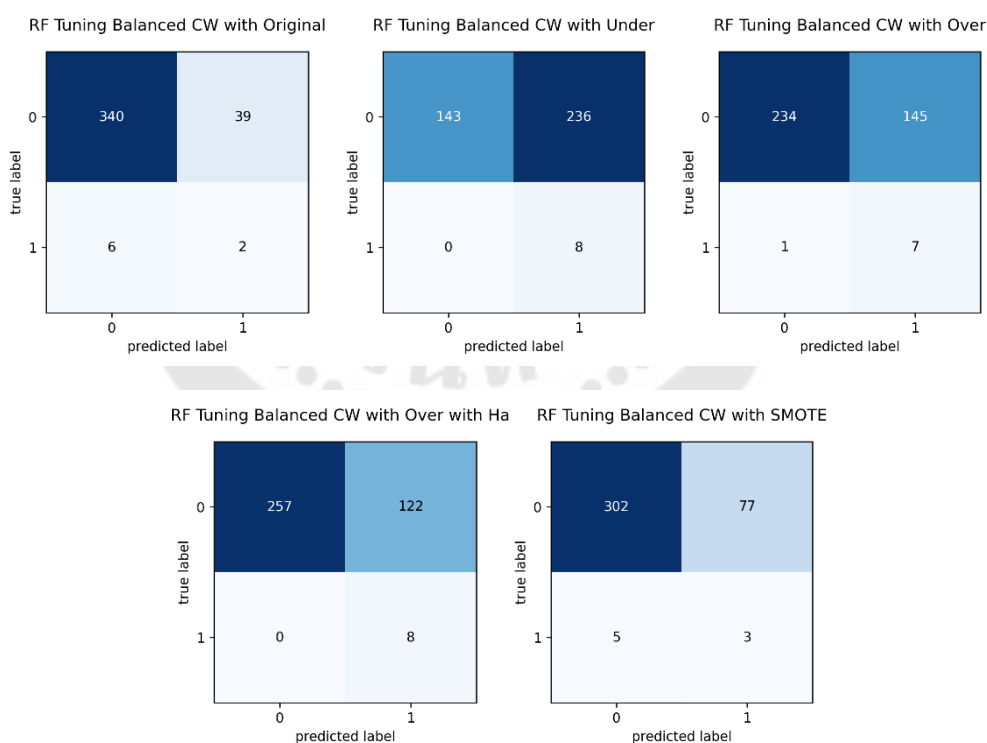
ภาพประกอบ 193 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.10.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8837	0.8971	0.2500	0.0488	0.0816
Under	0.3902	0.3773	1.0000	0.0328	0.0635
Over	0.6227	0.6174	0.8750	0.0461	0.0875
Over with Half	0.6848	0.6781	1.0000	0.0615	0.1159
SMOTE	0.7881	0.7968	0.3750	0.0375	0.0682

ภาพประกอบ 194 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

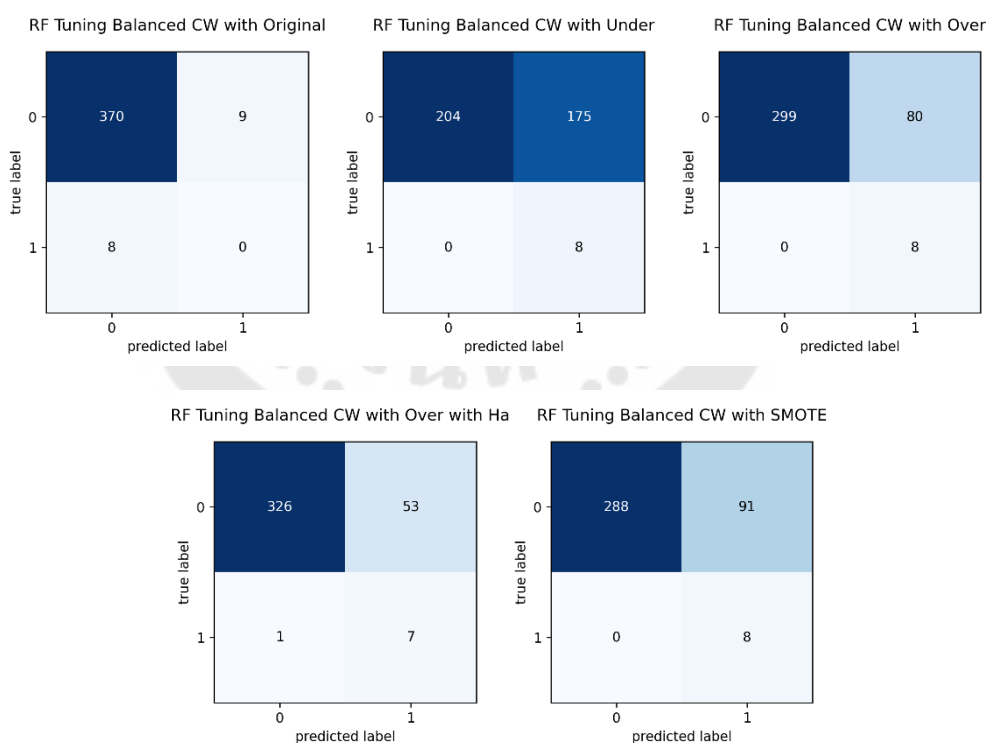


ภาพประกอบ 195 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9561	0.9763	0.0000	0.0000	0.0000
Under	0.5478	0.5383	1.0000	0.0437	0.0838
Over	0.7933	0.7889	1.0000	0.0909	0.1667
Over with Half	0.8605	0.8602	0.8750	0.1167	0.2059
SMOTE	0.7649	0.7599	1.0000	0.0808	0.1495

ภาพประกอบ 196 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



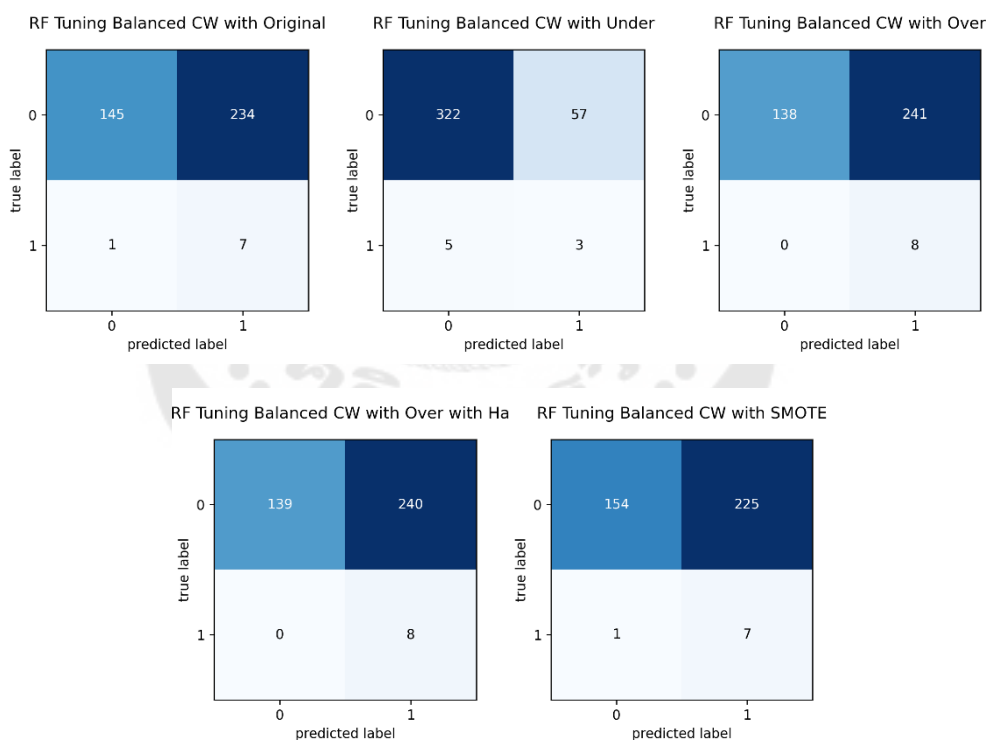
ภาพประกอบ 197 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3928	0.3826	0.8750	0.0290	0.0562
Under	0.8398	0.8496	0.3750	0.0500	0.0882
Over	0.3773	0.3641	1.0000	0.0321	0.0623
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 198 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



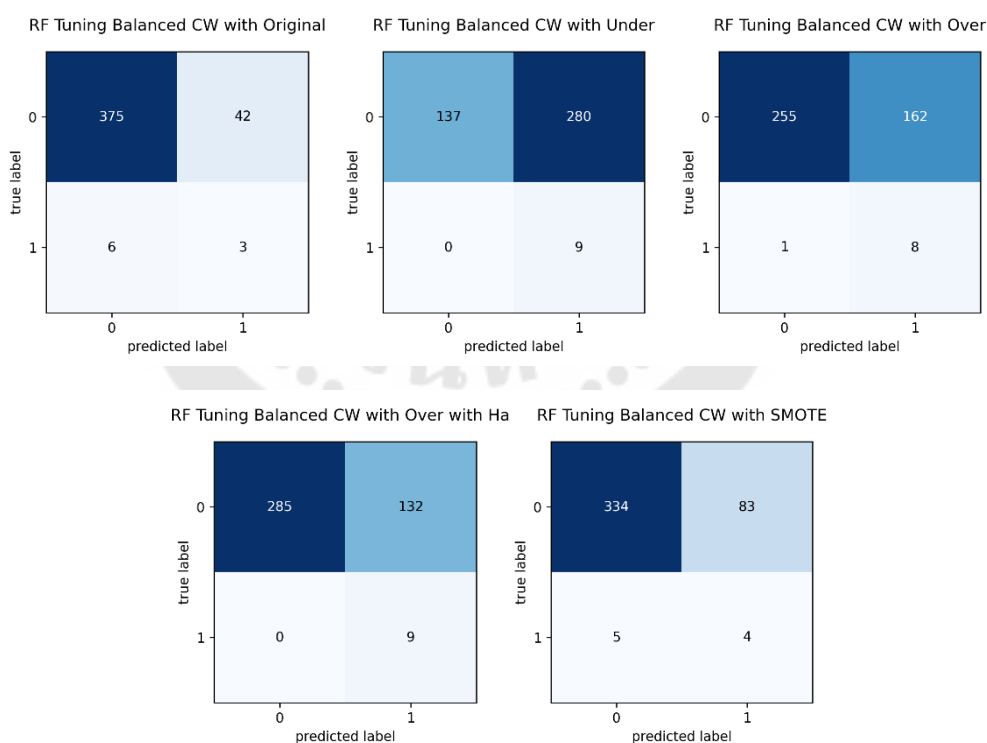
ภาพประกอบ 199 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.10.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8873	0.8993	0.3333	0.0667	0.1111
Under	0.3427	0.3285	1.0000	0.0311	0.0604
Over	0.6174	0.6115	0.8889	0.0471	0.0894
Over with Half	0.6901	0.6835	1.0000	0.0638	0.1200
SMOTE	0.7934	0.8010	0.4444	0.0460	0.0833

ภาพประกอบ 200 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

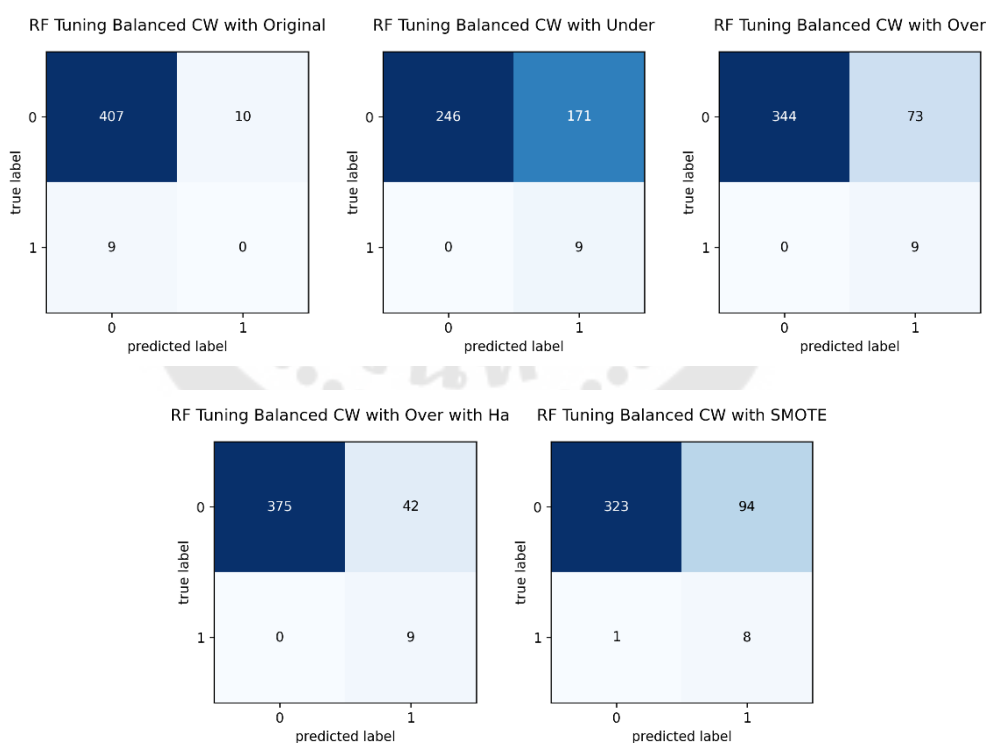


ภาพประกอบ 201 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9554	0.9760	0.0000	0.0000	0.0000
Under	0.5986	0.5899	1.0000	0.0500	0.0952
Over	0.8286	0.8249	1.0000	0.1098	0.1978
Over with Half	0.9014	0.8993	1.0000	0.1765	0.3000
SMOTE	0.7770	0.7746	0.8889	0.0784	0.1441

ภาพประกอบ 202 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



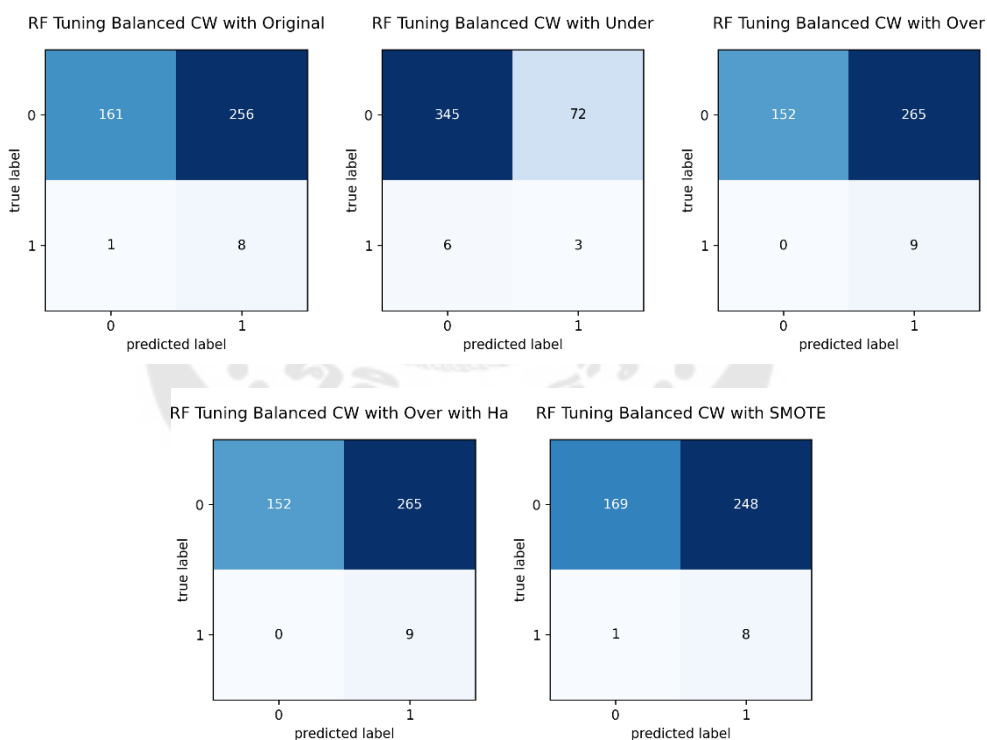
ภาพประกอบ 203 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.10.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3967	0.3861	0.8889	0.0303	0.0586
Under	0.8169	0.8273	0.3333	0.0400	0.0714
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 204 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Decision Tree



ภาพประกอบ 205 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

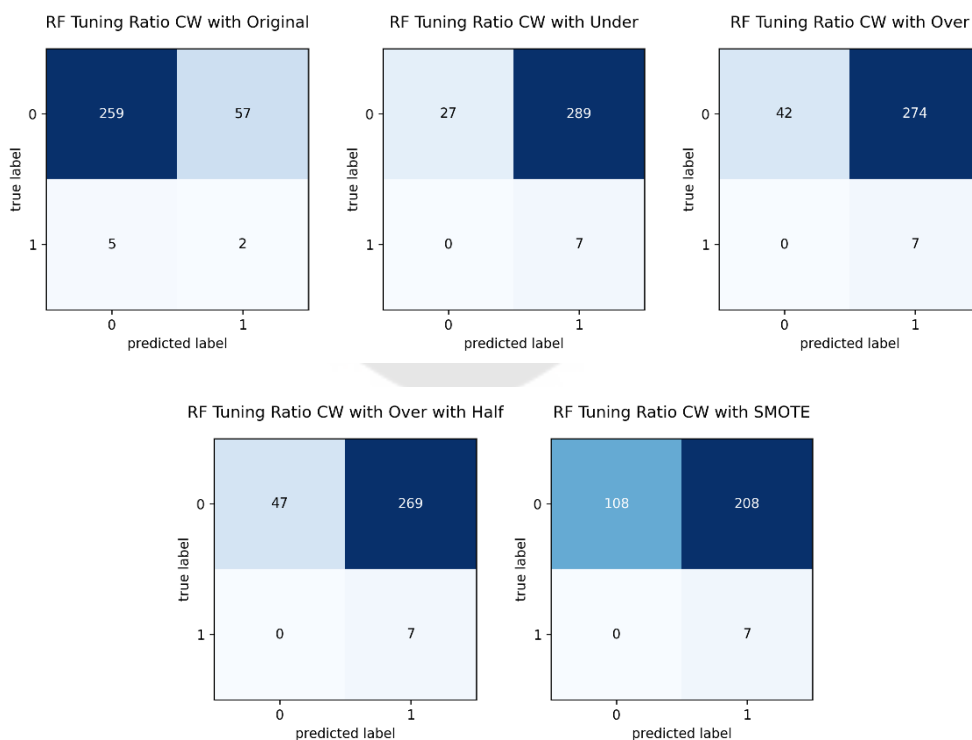
4.11 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight

4.11.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.11.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8080	0.8196	0.2857	0.0339	0.0606
Under	0.1053	0.0854	1.0000	0.0236	0.0462
Over	0.1517	0.1329	1.0000	0.0249	0.0486
Over with Half	0.1672	0.1487	1.0000	0.0254	0.0495
SMOTE	0.3560	0.3418	1.0000	0.0326	0.0631

ภาพประกอบ 206 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

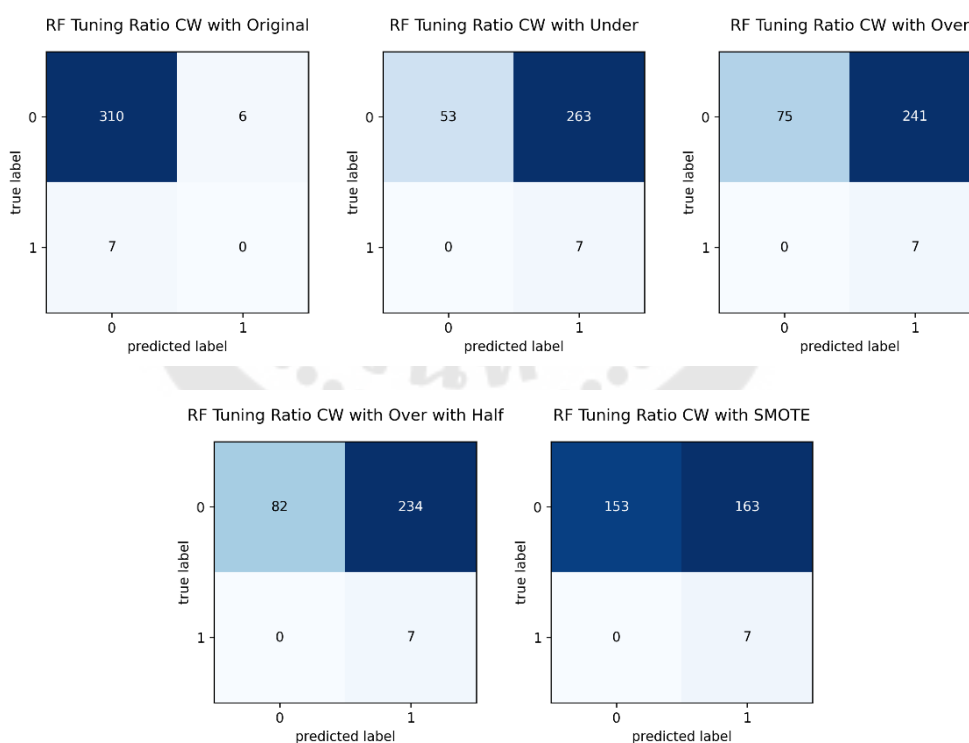


ภาพประกอบ 207 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9598	0.9810	0.0000	0.0000	0.0000
Under	0.1858	0.1677	1.0000	0.0259	0.0505
Over	0.2539	0.2373	1.0000	0.0282	0.0549
Over with Half	0.2755	0.2595	1.0000	0.0290	0.0565
SMOTE	0.4954	0.4842	1.0000	0.0412	0.0791

ภาพประกอบ 208 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

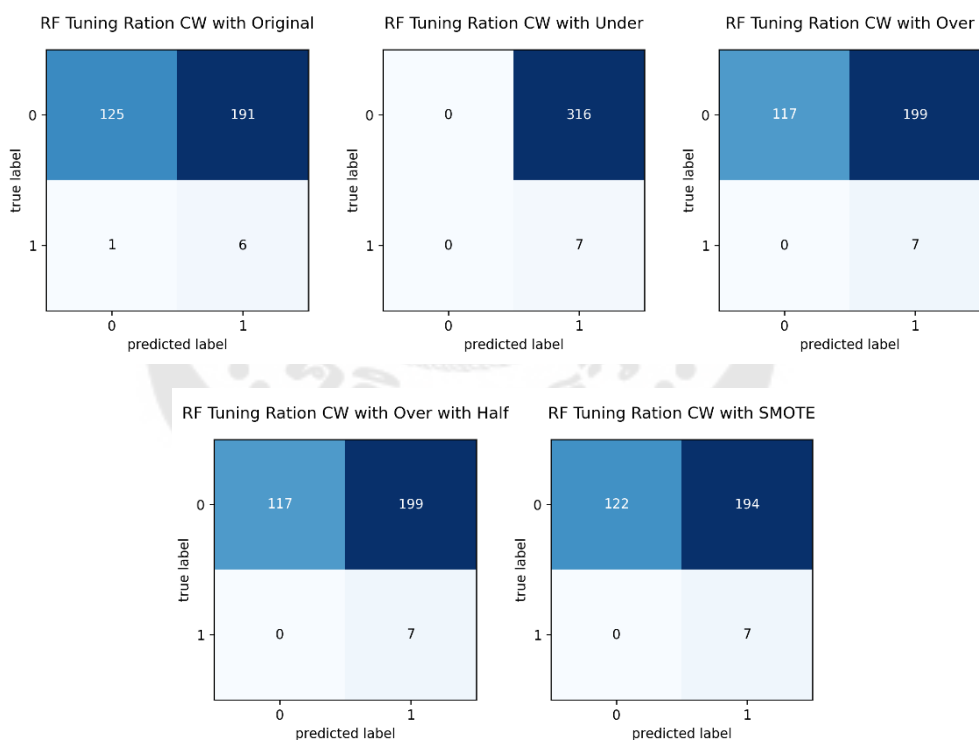


ภาพประกอบ 209 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4056	0.3956	0.8571	0.0305	0.0588
Under	0.0217	0.0000	1.0000	0.0217	0.0424
Over	0.3839	0.3703	1.0000	0.0340	0.0657
Over with Half	0.3839	0.3703	1.0000	0.0340	0.0657
SMOTE	0.3994	0.3861	1.0000	0.0348	0.0673

ภาพประกอบ 210 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



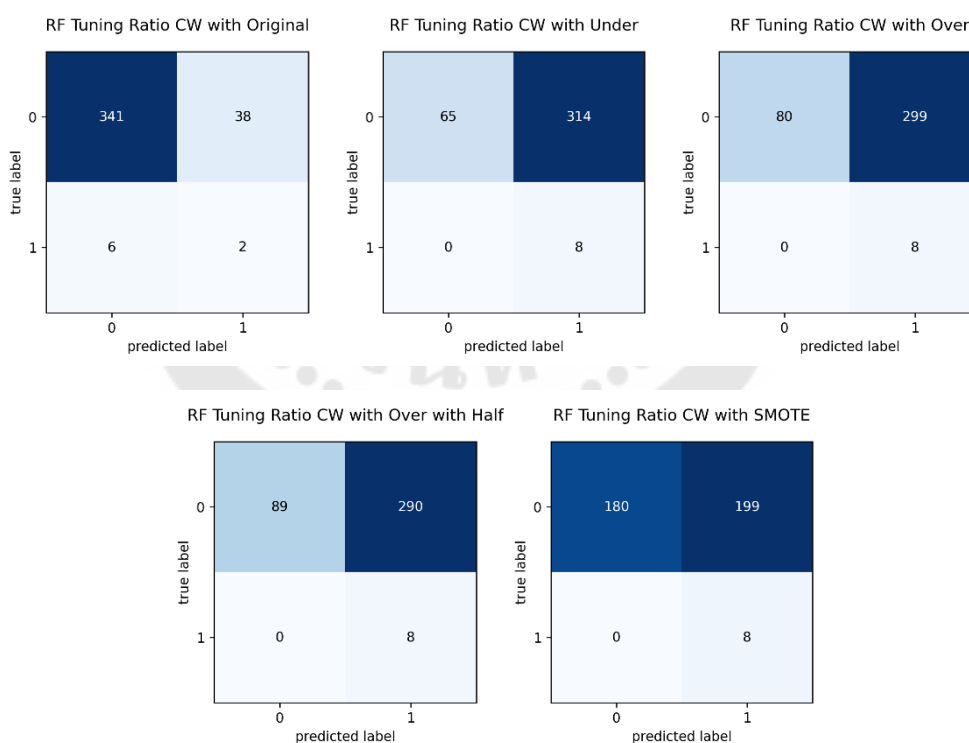
ภาพประกอบ 211 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.11.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8863	0.8997	0.2500	0.0500	0.0833
Under	0.1886	0.1715	1.0000	0.0248	0.0485
Over	0.2274	0.2111	1.0000	0.0261	0.0508
Over with Half	0.2506	0.2348	1.0000	0.0268	0.0523
SMOTE	0.4858	0.4749	1.0000	0.0386	0.0744

ภาพประกอบ 212 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

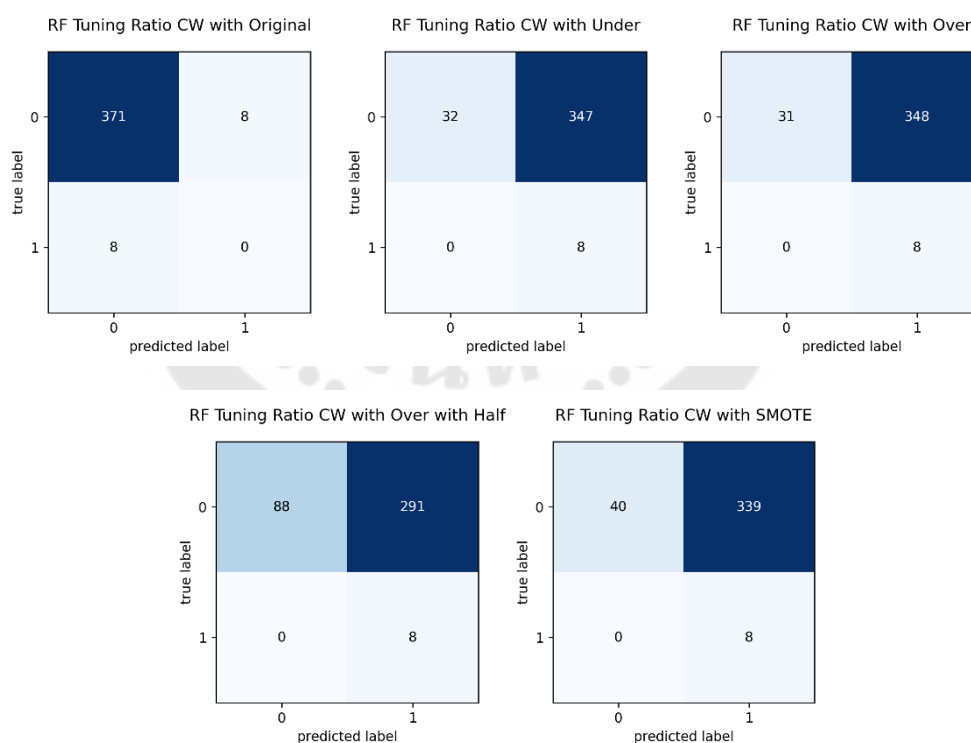


ภาพประกอบ 213 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9587	0.9789	0.0000	0.0000	0.0000
Under	0.1034	0.0844	1.0000	0.0225	0.0441
Over	0.1008	0.0818	1.0000	0.0225	0.0440
Over with Half	0.2481	0.2322	1.0000	0.0268	0.0521
SMOTE	0.1240	0.1055	1.0000	0.0231	0.0451

ภาพประกอบ 214 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

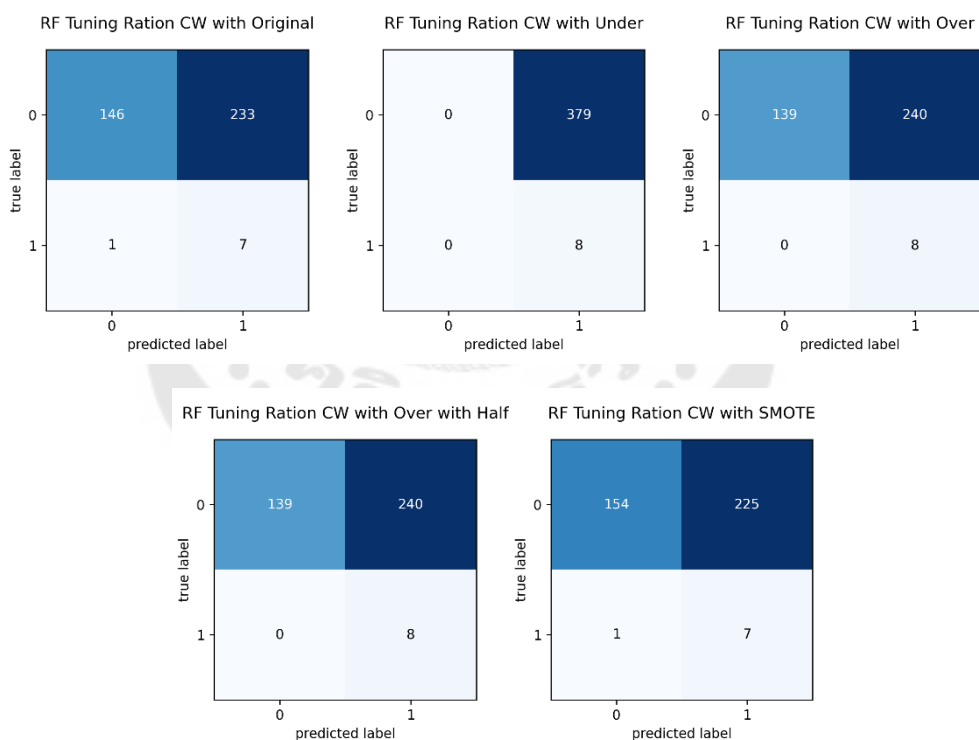


ภาพประกอบ 215 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3953	0.3852	0.8750	0.0292	0.0565
Under	0.0207	0.0000	1.0000	0.0207	0.0405
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 216 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



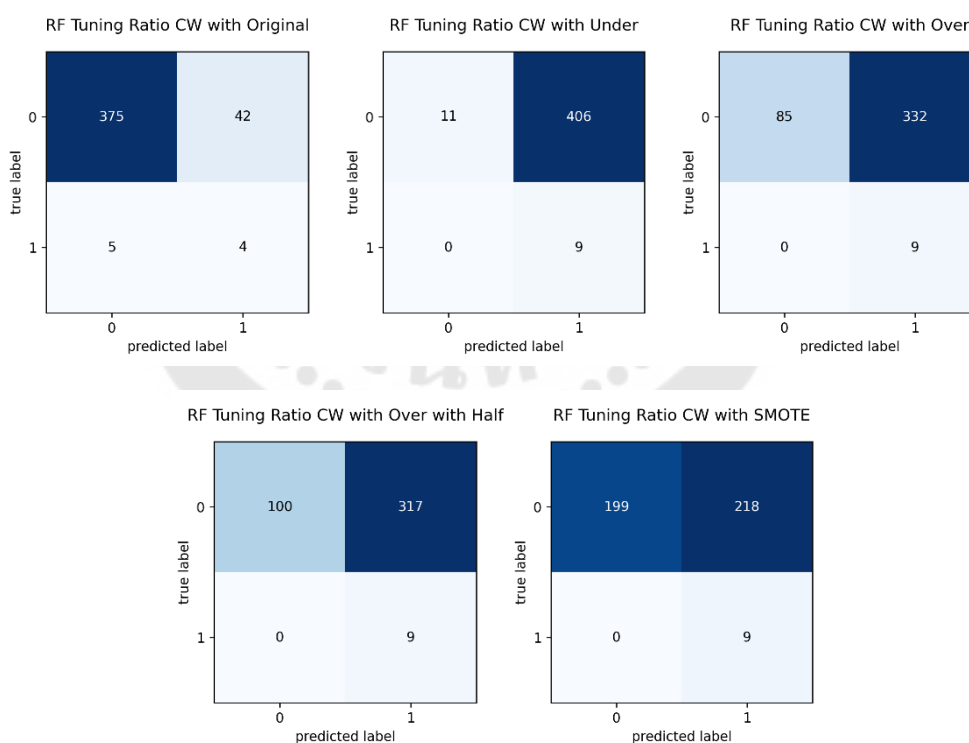
ภาพประกอบ 217 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.11.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8897	0.8993	0.4444	0.0870	0.1455
Under	0.0469	0.0264	1.0000	0.0217	0.0425
Over	0.2207	0.2038	1.0000	0.0264	0.0514
Over with Half	0.2559	0.2398	1.0000	0.0276	0.0537
SMOTE	0.4883	0.4772	1.0000	0.0396	0.0763

ภาพประกอบ 218 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

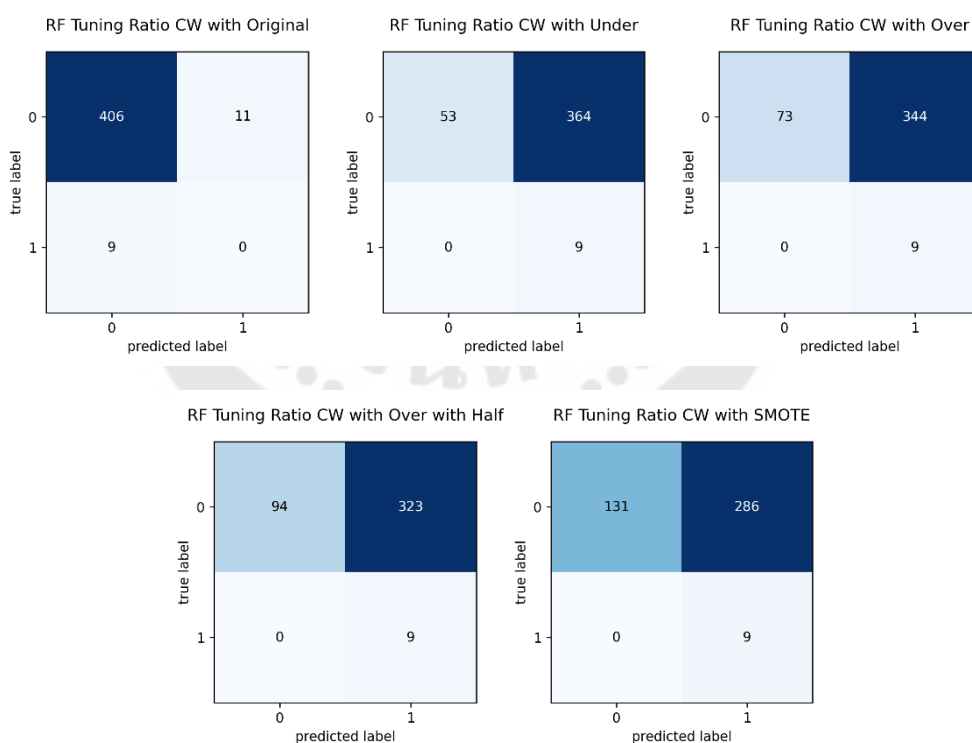


ภาพประกอบ 219 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9531	0.9736	0.0000	0.0000	0.0000
Under	0.1455	0.1271	1.0000	0.0241	0.0471
Over	0.1925	0.1751	1.0000	0.0255	0.0497
Over with Half	0.2418	0.2254	1.0000	0.0271	0.0528
SMOTE	0.3286	0.3141	1.0000	0.0305	0.0592

ภาพประกอบ 220 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

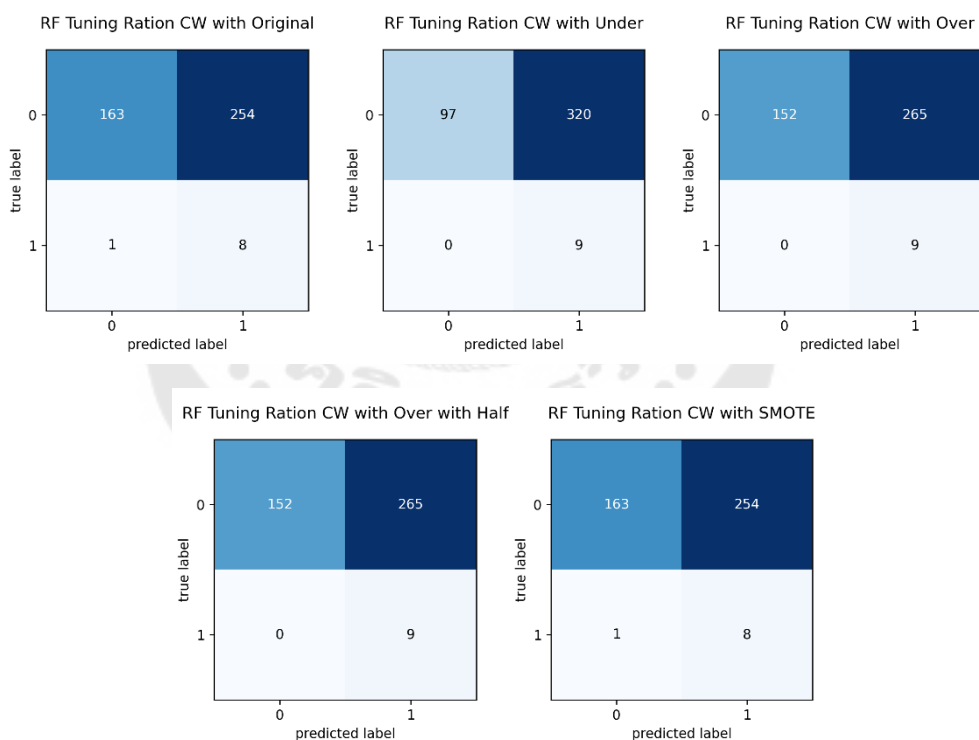


ภาพประกอบ 221 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.11.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4014	0.3909	0.8889	0.0305	0.0590
Under	0.2488	0.2326	1.0000	0.0274	0.0533
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4014	0.3909	0.8889	0.0305	0.0590

ภาพประกอบ 222 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Ratio Class Weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 223 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier

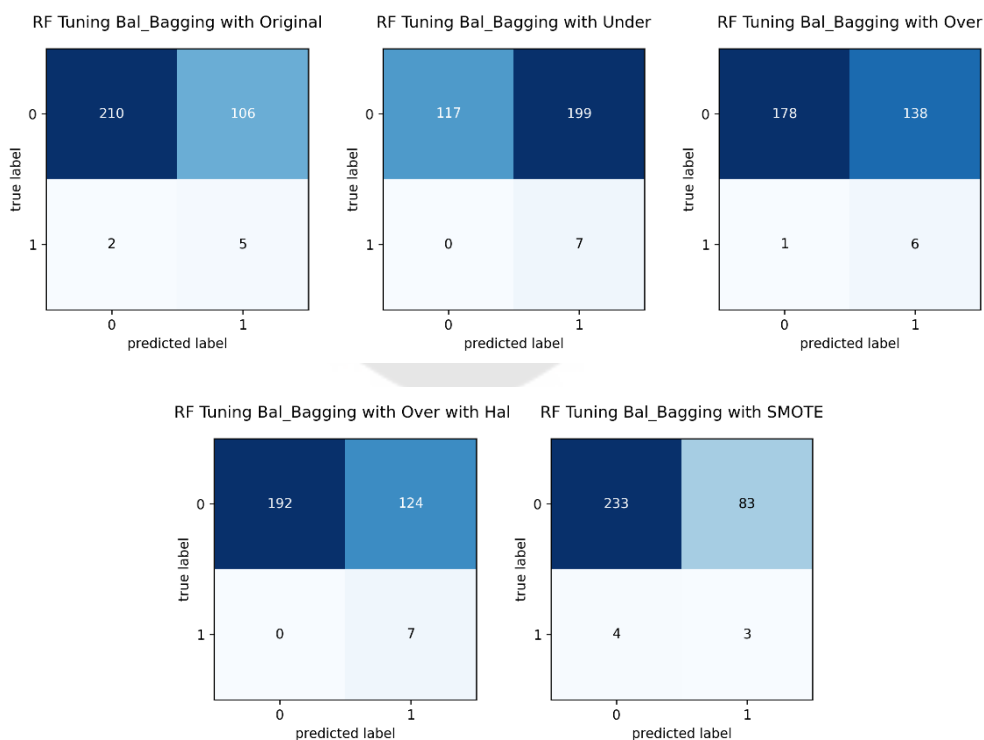
4.12.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.12.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6656	0.6646	0.7143	0.0450	0.0847
Under	0.3839	0.3703	1.0000	0.0340	0.0657
Over	0.5697	0.5633	0.8571	0.0417	0.0795
Over with Half	0.6161	0.6076	1.0000	0.0534	0.1014
SMOTE	0.7307	0.7373	0.4286	0.0349	0.0645

ภาพประกอบ 224 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

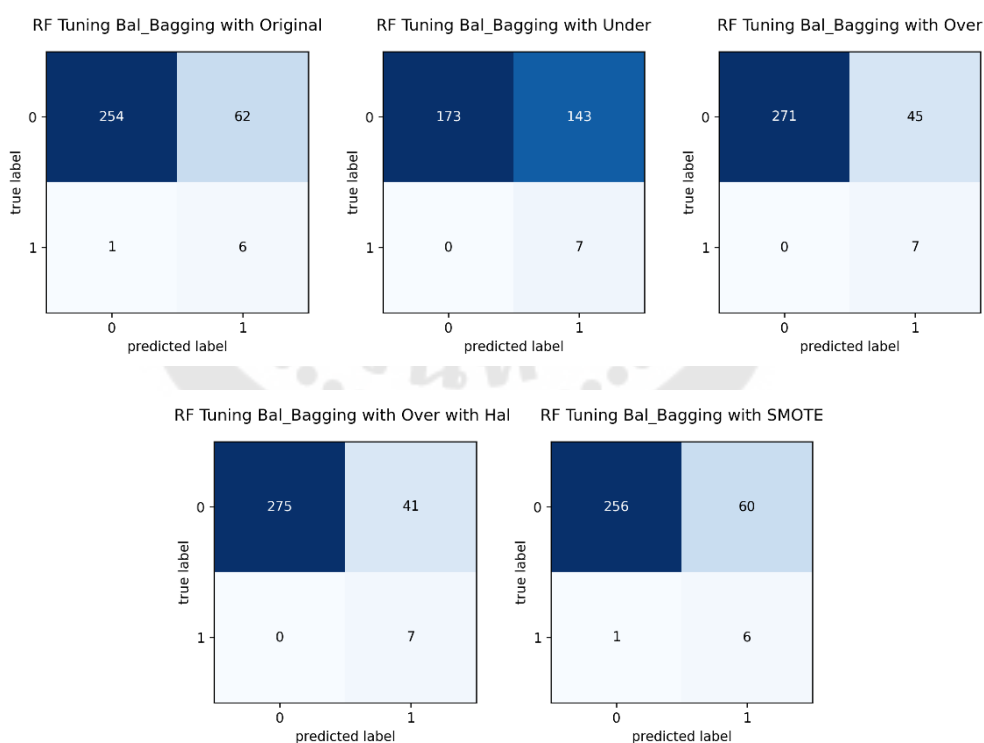


ภาพประกอบ 225 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8050	0.8038	0.8571	0.0882	0.1600
Under	0.5573	0.5475	1.0000	0.0467	0.0892
Over	0.8607	0.8576	1.0000	0.1346	0.2373
Over with Half	0.8731	0.8703	1.0000	0.1458	0.2545
SMOTE	0.8111	0.8101	0.8571	0.0909	0.1644

ภาพประกอบ 226 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



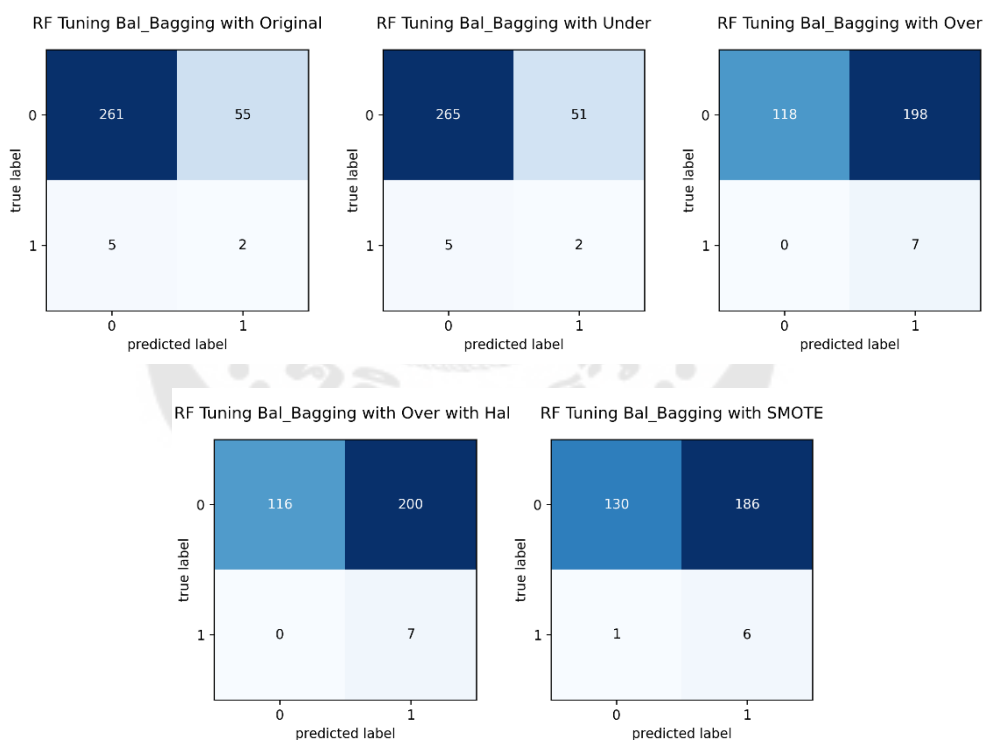
ภาพประกอบ 227 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8142	0.8259	0.2857	0.0351	0.0625
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3808	0.3671	1.0000	0.0338	0.0654
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 228 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Decision Tree



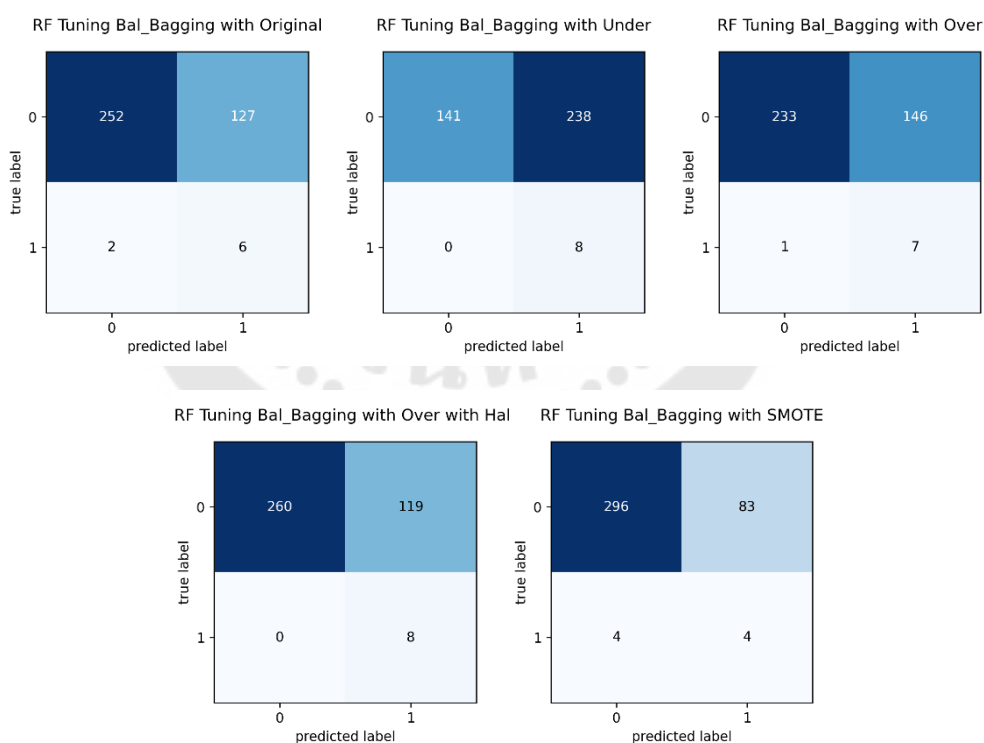
ภาพประกอบ 229 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.12.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6667	0.6649	0.7500	0.0451	0.0851
Under	0.3850	0.3720	1.0000	0.0325	0.0630
Over	0.6202	0.6148	0.8750	0.0458	0.0870
Over with Half	0.6925	0.6860	1.0000	0.0630	0.1185
SMOTE	0.7752	0.7810	0.5000	0.0460	0.0842

ภาพประกอบ 230 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

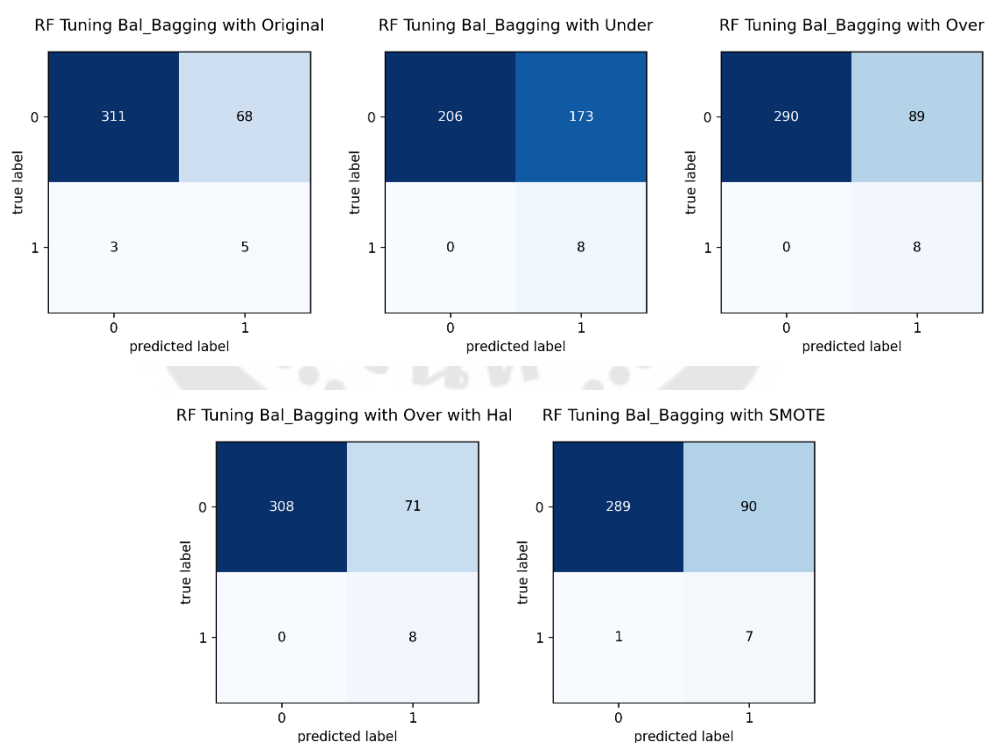


ภาพประกอบ 231 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8165	0.8206	0.6250	0.0685	0.1235
Under	0.5530	0.5435	1.0000	0.0442	0.0847
Over	0.7700	0.7652	1.0000	0.0825	0.1524
Over with Half	0.8165	0.8127	1.0000	0.1013	0.1839
SMOTE	0.7649	0.7625	0.8750	0.0722	0.1333

ภาพประกอบ 232 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



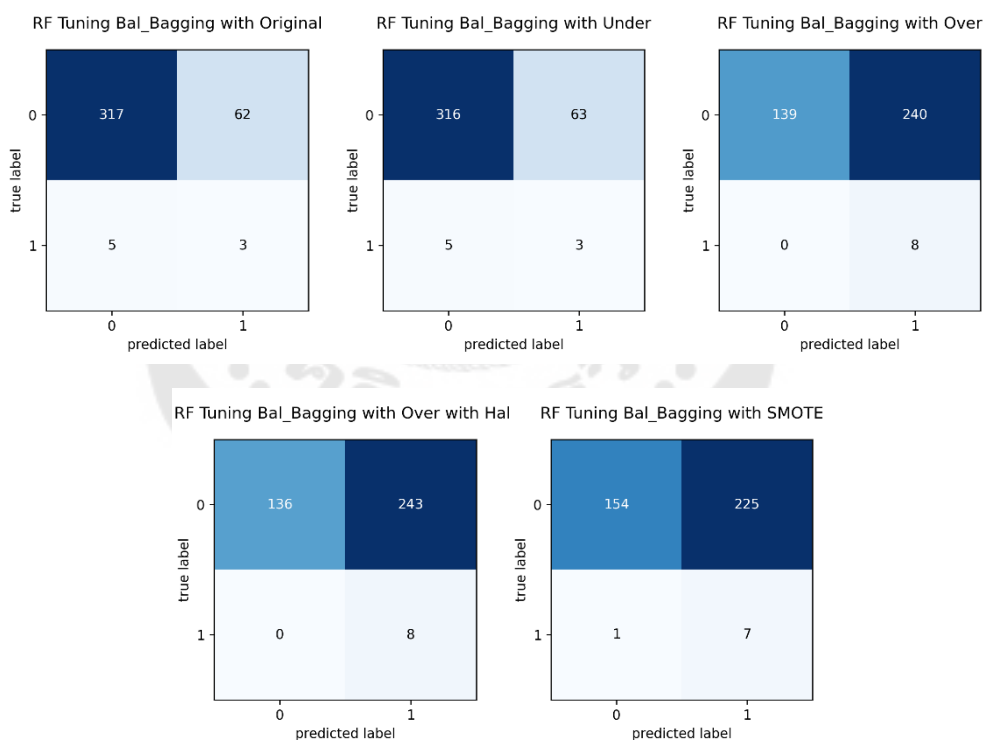
ภาพประกอบ 233 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8269	0.8364	0.3750	0.0462	0.0822
Under	0.8243	0.8338	0.3750	0.0455	0.0811
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3721	0.3588	1.0000	0.0319	0.0618
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 234 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



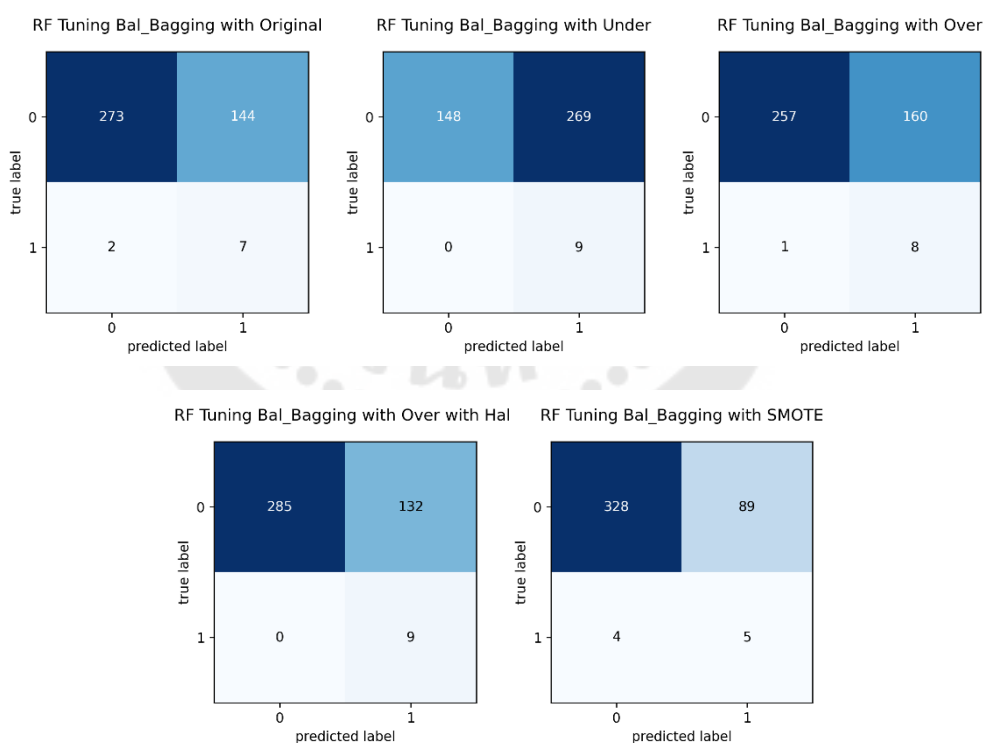
ภาพประกอบ 235 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12. ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.12.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6573	0.6547	0.7778	0.0464	0.0875
Under	0.3685	0.3549	1.0000	0.0324	0.0627
Over	0.6221	0.6163	0.8889	0.0476	0.0904
Over with Half	0.6901	0.6835	1.0000	0.0638	0.1200
SMOTE	0.7817	0.7866	0.5556	0.0532	0.0971

ภาพประกอบ 236 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold



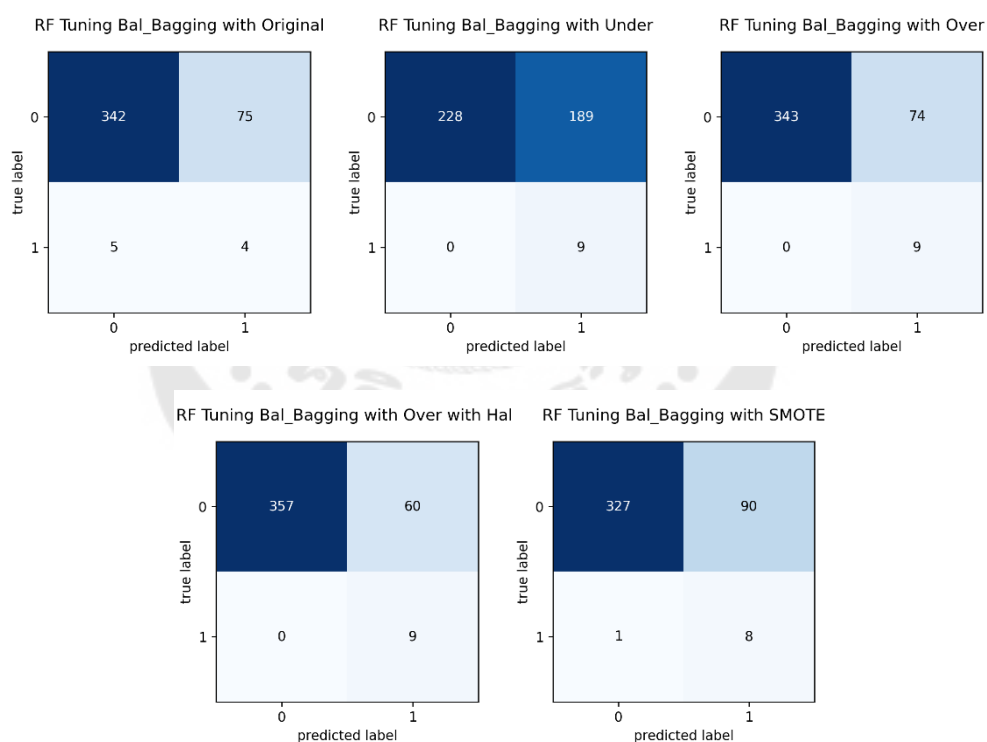
ภาพประกอบ 237 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8122	0.8201	0.4444	0.0506	0.0909
Under	0.5563	0.5468	1.0000	0.0455	0.0870
Over	0.8263	0.8225	1.0000	0.1084	0.1957
Over with Half	0.8592	0.8561	1.0000	0.1304	0.2308
SMOTE	0.7864	0.7842	0.8889	0.0816	0.1495

ภาพประกอบ 238 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Mutual Information Classification



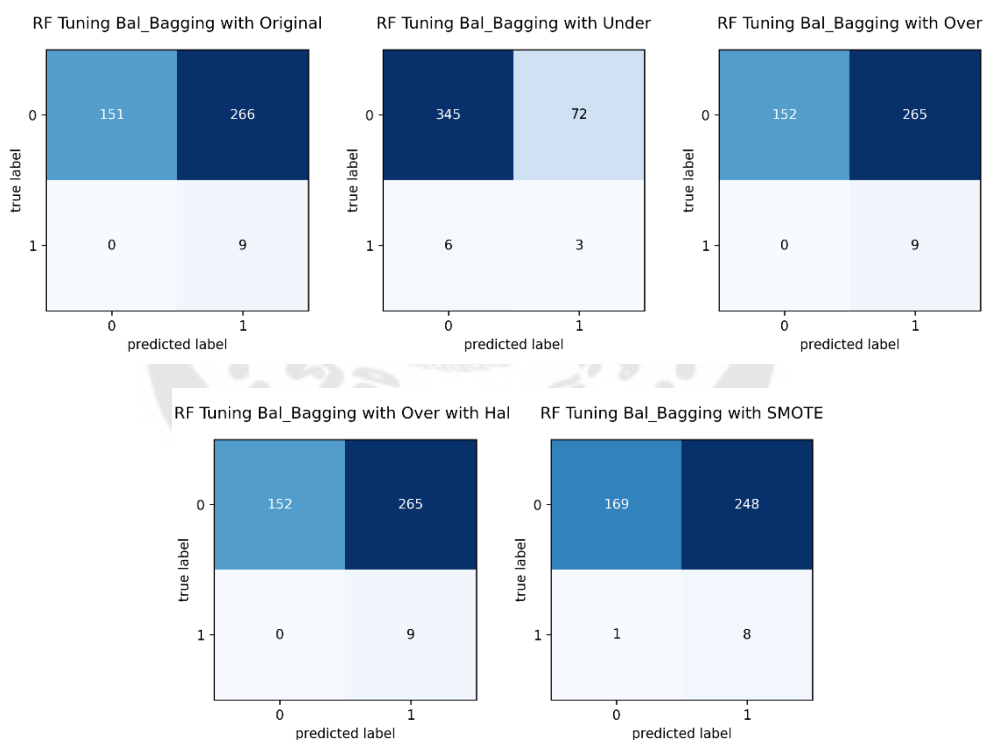
ภาพประกอบ 239 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.12.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3756	0.3621	1.0000	0.0327	0.0634
Under	0.8169	0.8273	0.3333	0.0400	0.0714
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 240 ผลลัพธ์ที่ได้จากการใช้ Random Forest กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

Decision Tree



ภาพประกอบ 241 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

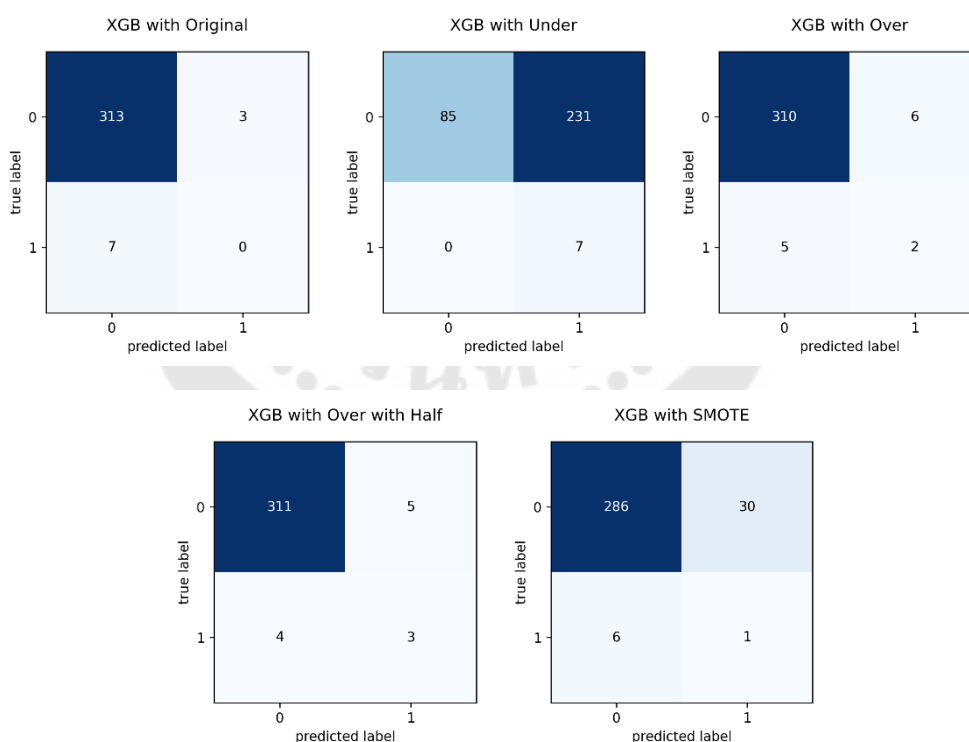
4.13 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters

4.13.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.13.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9690	0.9905	0.0000	0.0000	0.0000
Under	0.2848	0.2690	1.0000	0.0294	0.0571
Over	0.9659	0.9810	0.2857	0.2500	0.2667
Over with Half	0.9721	0.9842	0.4286	0.3750	0.4000
SMOTE	0.8885	0.9051	0.1429	0.0323	0.0526

ภาพประกอบ 242 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

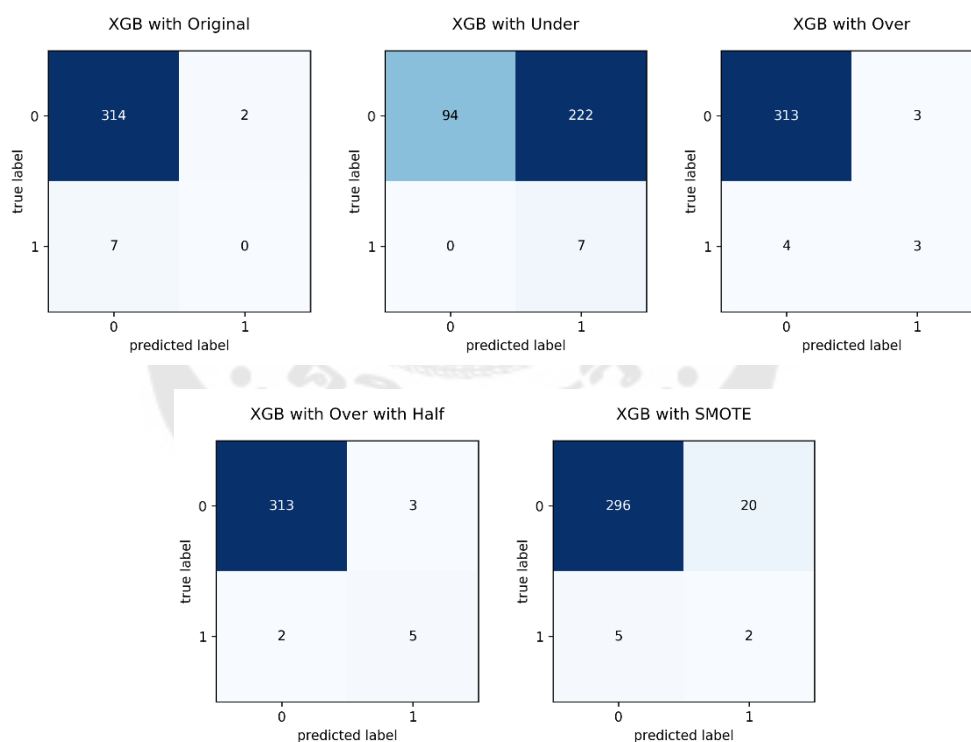


ภาพประกอบ 243 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9721	0.9937	0.0000	0.0000	0.0000
Under	0.3127	0.2975	1.0000	0.0306	0.0593
Over	0.9783	0.9905	0.4286	0.5000	0.4615
Over with Half	0.9845	0.9905	0.7143	0.6250	0.6667
SMOTE	0.9226	0.9367	0.2857	0.0909	0.1379

ภาพประกอบ 244 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

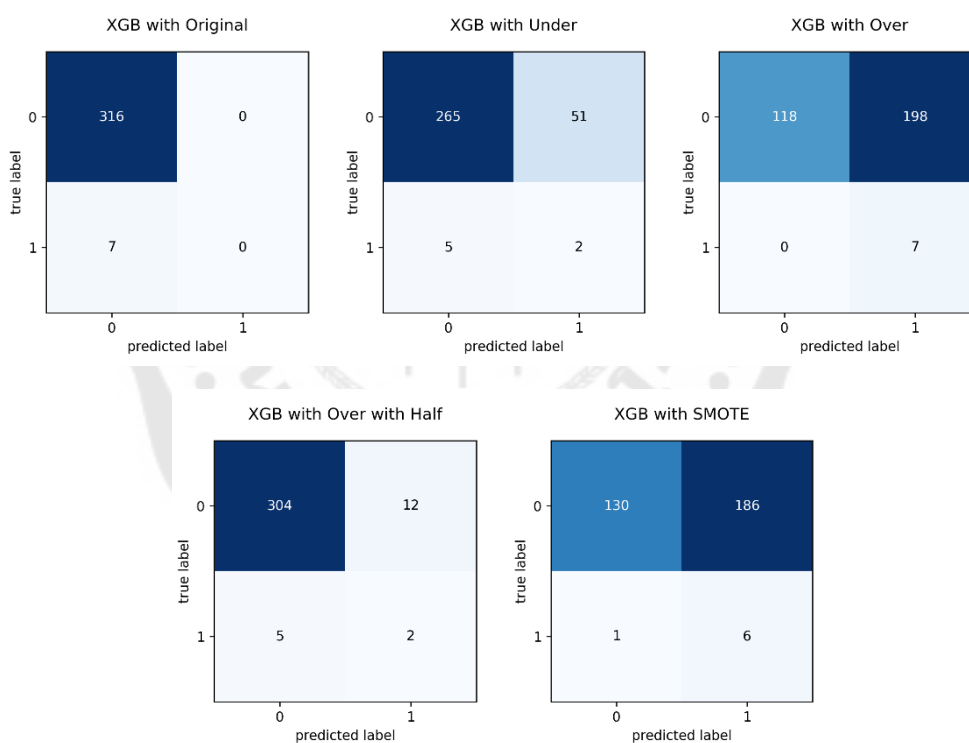


ภาพประกอบ 245 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.9474	0.9620	0.2857	0.1429	0.1905
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 246 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



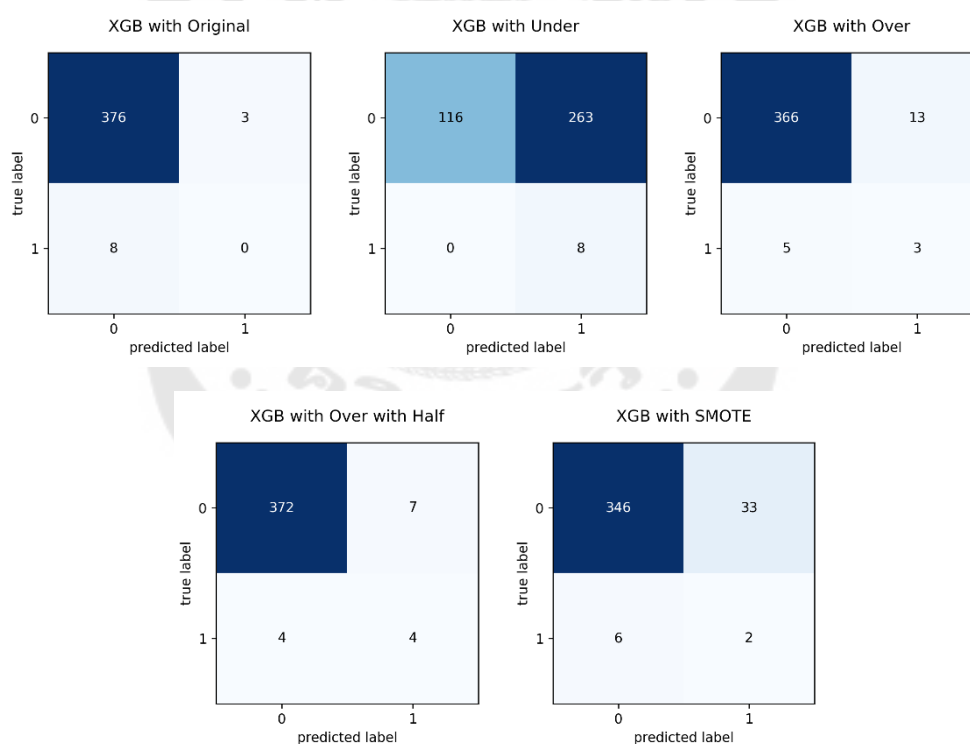
ภาพประกอบ 247 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.13.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9716	0.9921	0.0000	0.0000	0.0000
Under	0.3204	0.3061	1.0000	0.0295	0.0573
Over	0.9535	0.9657	0.3750	0.1875	0.2500
Over with Half	0.9716	0.9815	0.5000	0.3636	0.4211
SMOTE	0.8992	0.9129	0.2500	0.0571	0.0930

ภาพประกอบ 248 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

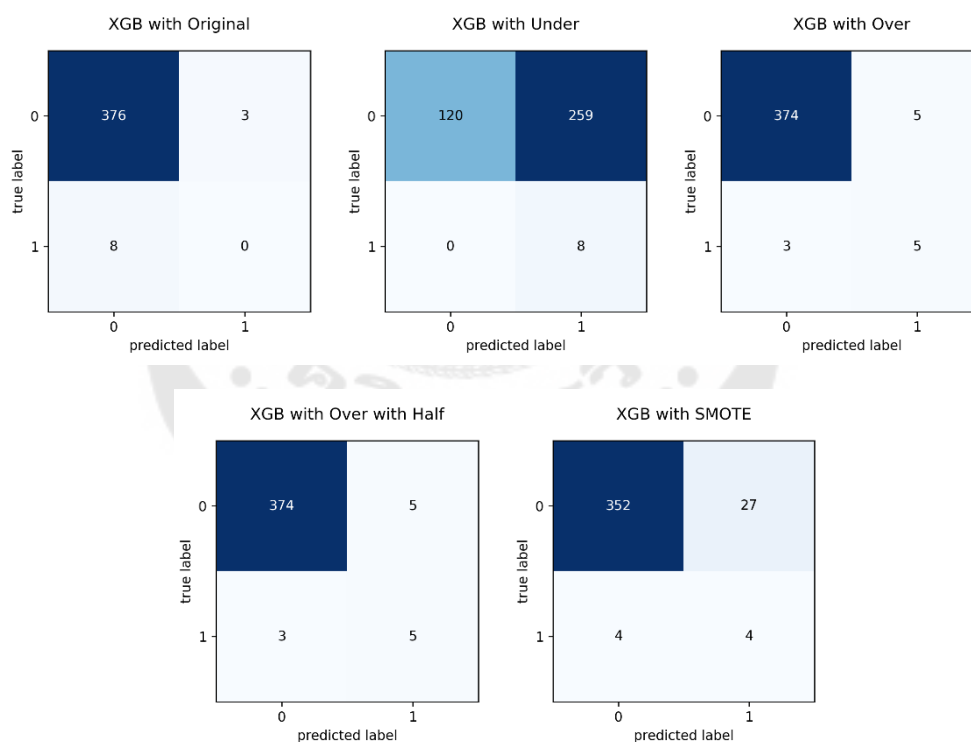


ภาพประกอบ 249 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9716	0.9921	0.0000	0.0000	0.0000
Under	0.3307	0.3166	1.0000	0.0300	0.0582
Over	0.9793	0.9868	0.6250	0.5000	0.5556
Over with Half	0.9793	0.9868	0.6250	0.5000	0.5556
SMOTE	0.9199	0.9288	0.5000	0.1290	0.2051

ภาพประกอบ 250 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

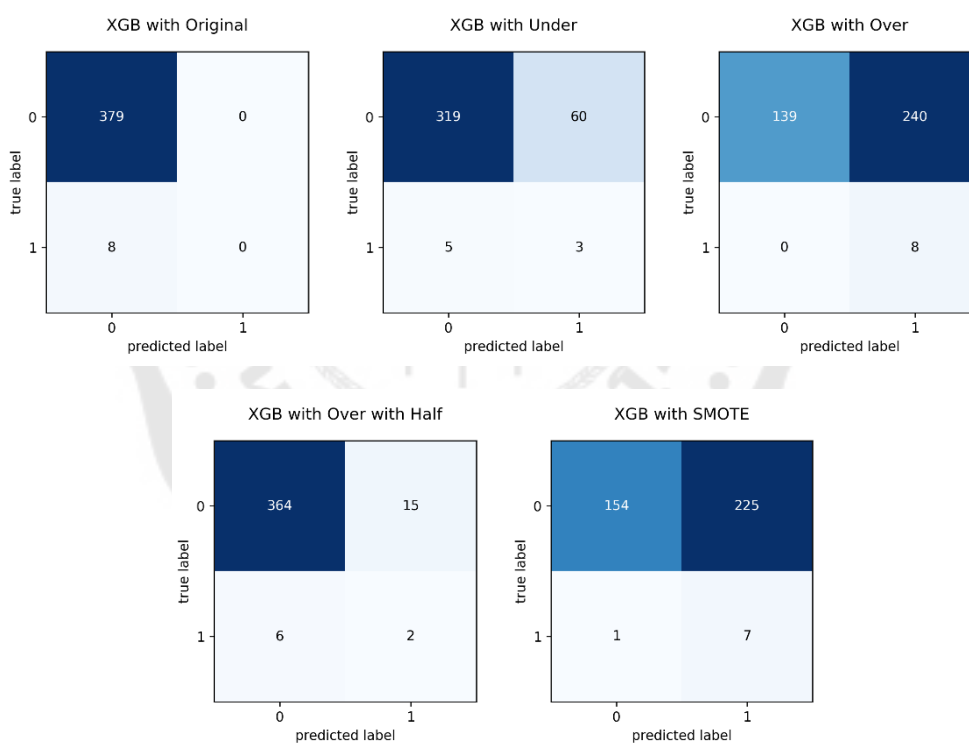


ภาพประกอบ 251 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.9457	0.9604	0.2500	0.1176	0.1600
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 252 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



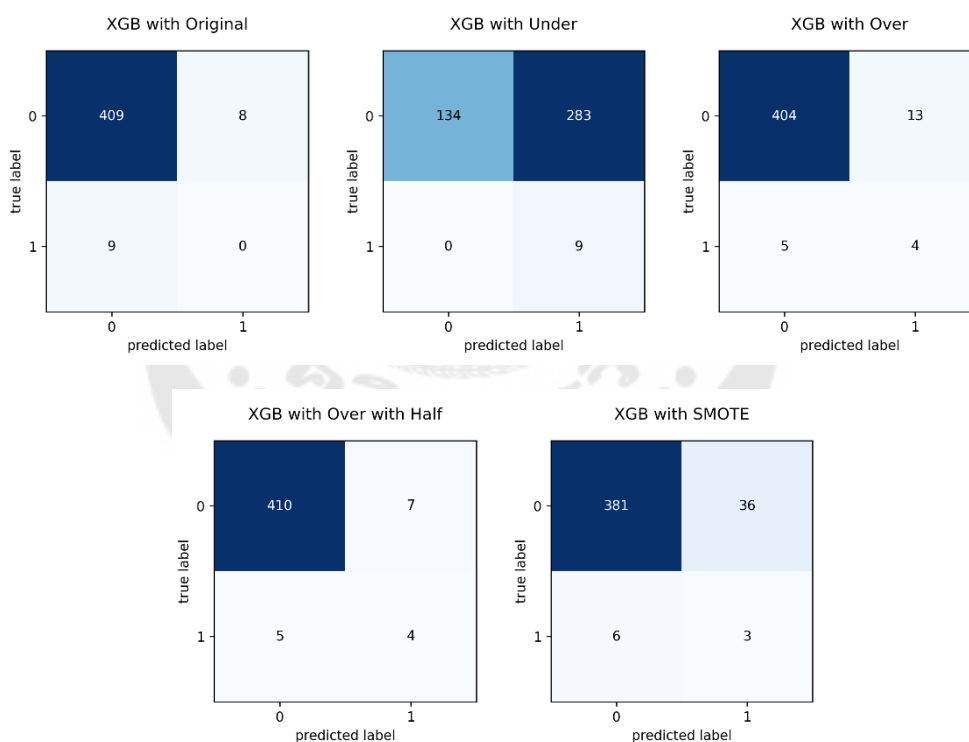
ภาพประกอบ 253 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.13.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9601	0.9808	0.0000	0.0000	0.0000
Under	0.3357	0.3213	1.0000	0.0308	0.0598
Over	0.9577	0.9688	0.4444	0.2353	0.3077
Over with Half	0.9718	0.9832	0.4444	0.3636	0.4000
SMOTE	0.9014	0.9137	0.3333	0.0769	0.1250

ภาพประกอบ 254 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

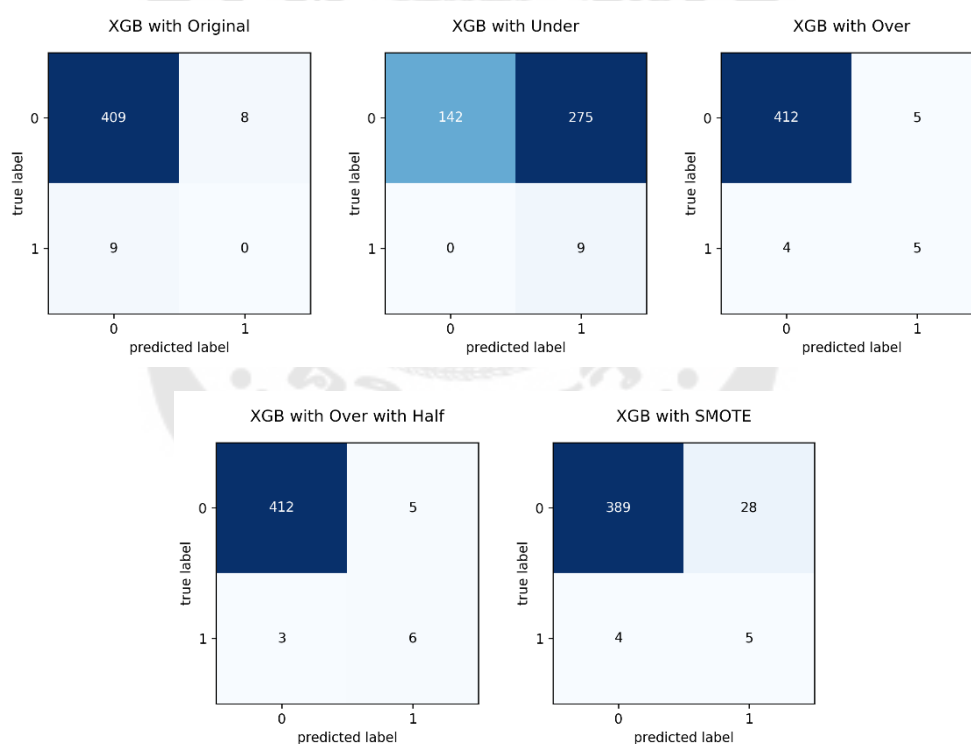


ภาพประกอบ 255 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9601	0.9808	0.0000	0.0000	0.0000
Under	0.3545	0.3405	1.0000	0.0317	0.0614
Over	0.9789	0.9880	0.5556	0.5000	0.5263
Over with Half	0.9812	0.9880	0.6667	0.5455	0.6000
SMOTE	0.9249	0.9329	0.5556	0.1515	0.2381

ภาพประกอบ 256 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

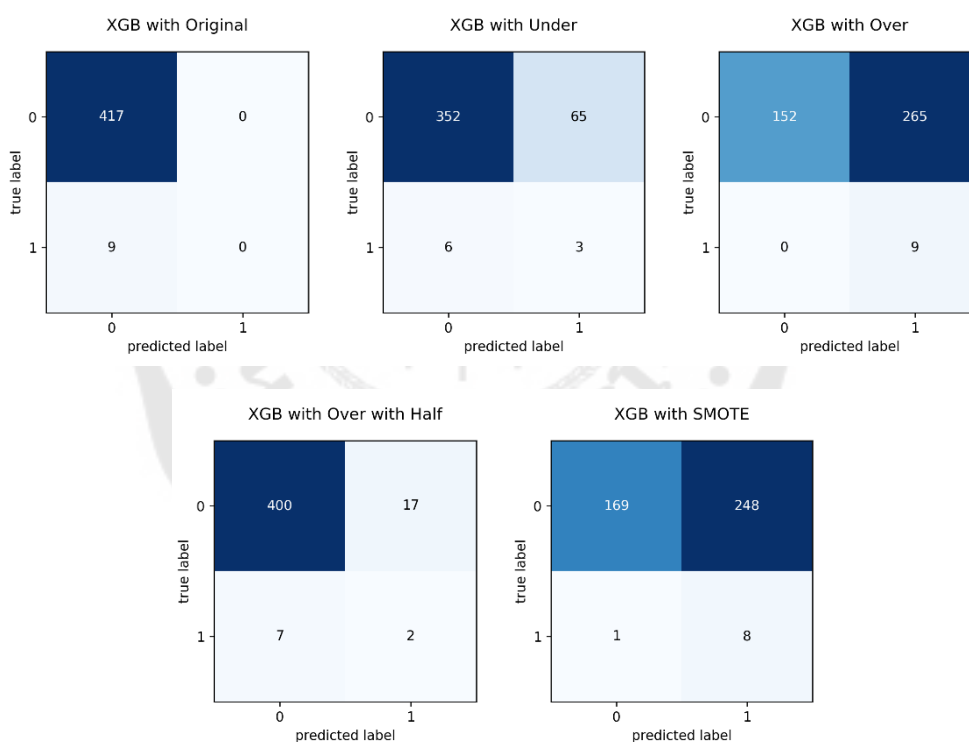


ภาพประกอบ 257 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.13.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.9437	0.9592	0.2222	0.1053	0.1429
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 258 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 259 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

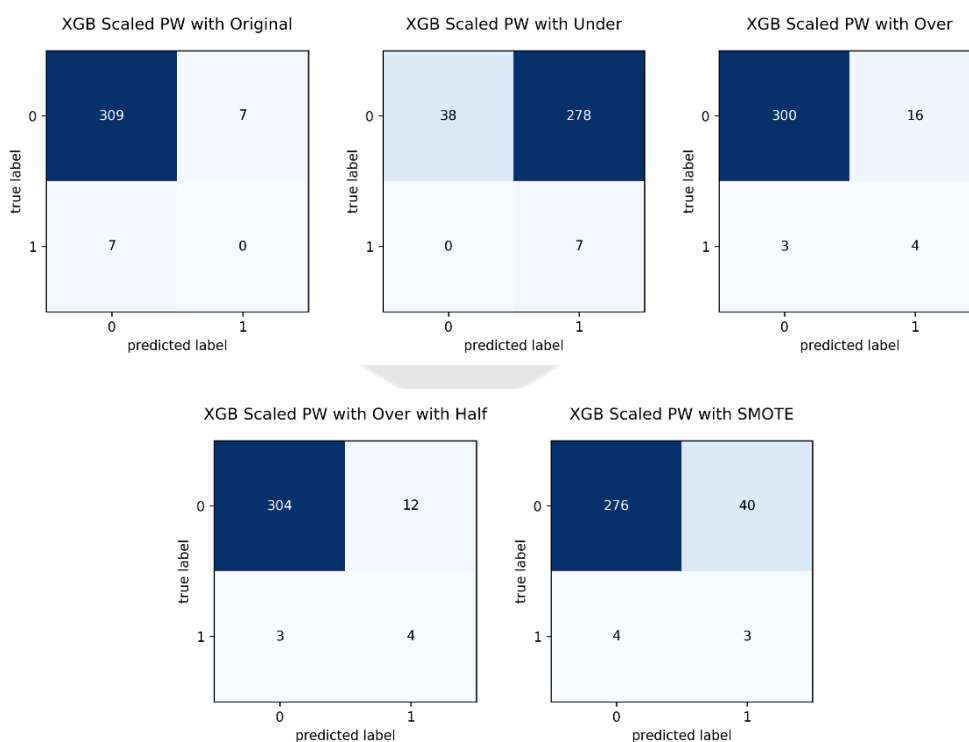
4.14 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ scaled_pos_weight

4.14.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.14.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9567	0.9778	0.0000	0.0000	0.0000
Under	0.1393	0.1203	1.0000	0.0246	0.0479
Over	0.9412	0.9494	0.5714	0.2000	0.2963
Over with Half	0.9536	0.9620	0.5714	0.2500	0.3478
SMOTE	0.8638	0.8734	0.4286	0.0698	0.1200

ภาพประกอบ 260 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

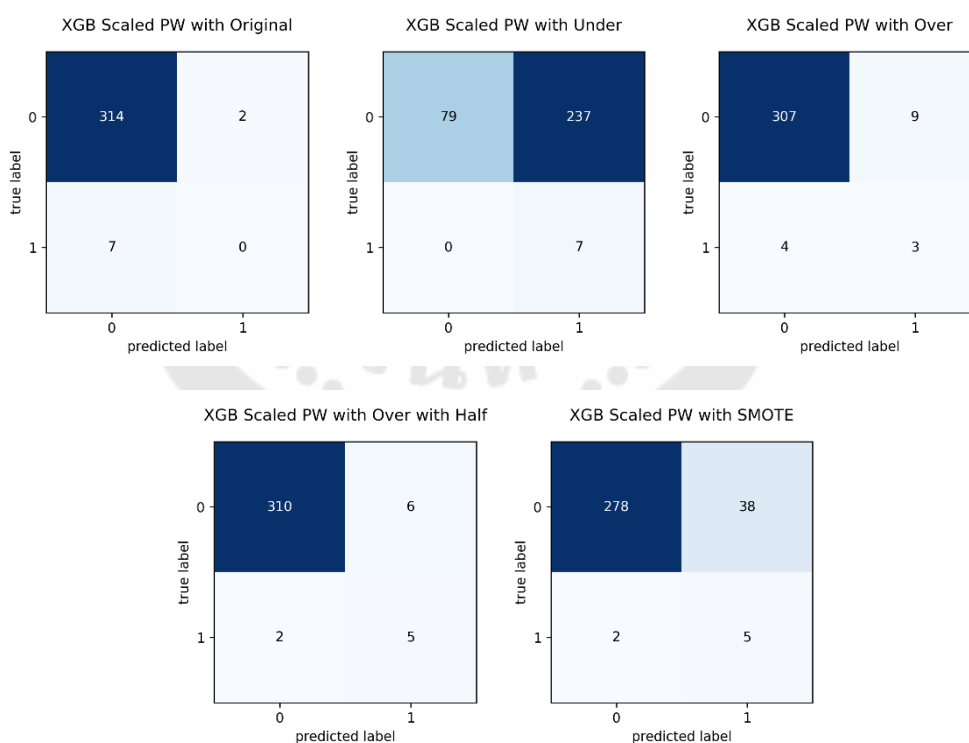


ภาพประกอบ 261 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9721	0.9937	0.0000	0.0000	0.0000
Under	0.2663	0.2500	1.0000	0.0287	0.0558
Over	0.9598	0.9715	0.4286	0.2500	0.3158
Over with Half	0.9752	0.9810	0.7143	0.4545	0.5556
SMOTE	0.8762	0.8797	0.7143	0.1163	0.2000

ภาพประกอบ 262 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

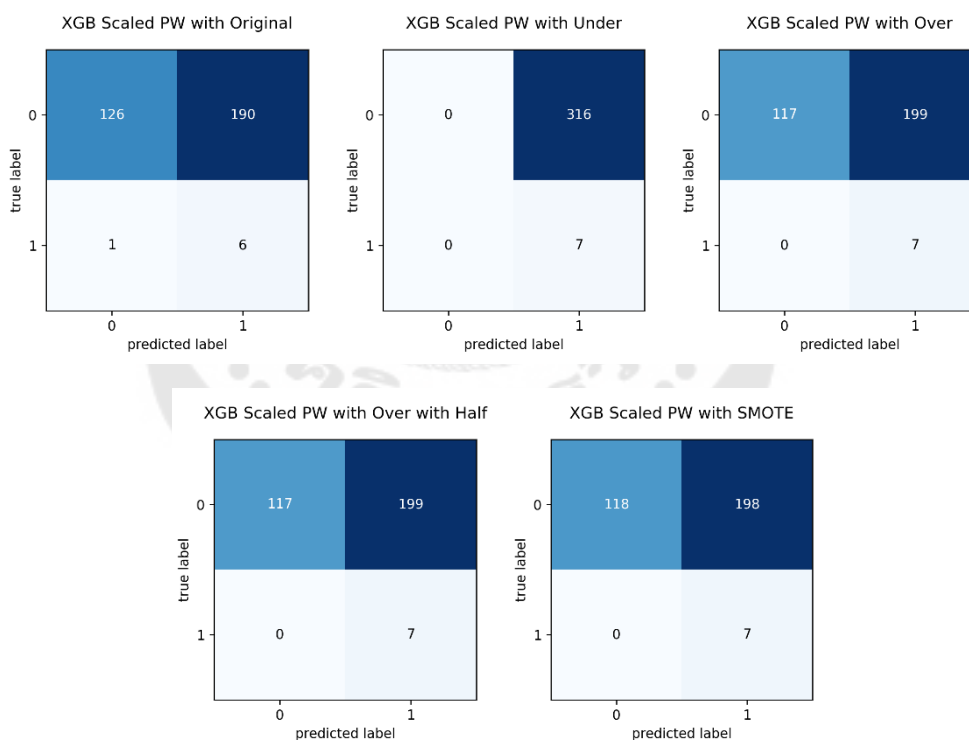


ภาพประกอบ 263 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4087	0.3987	0.8571	0.0306	0.0591
Under	0.0217	0.0000	1.0000	0.0217	0.0424
Over	0.3839	0.3703	1.0000	0.0340	0.0657
Over with Half	0.3839	0.3703	1.0000	0.0340	0.0657
SMOTE	0.3870	0.3734	1.0000	0.0341	0.0660

ภาพประกอบ 264 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



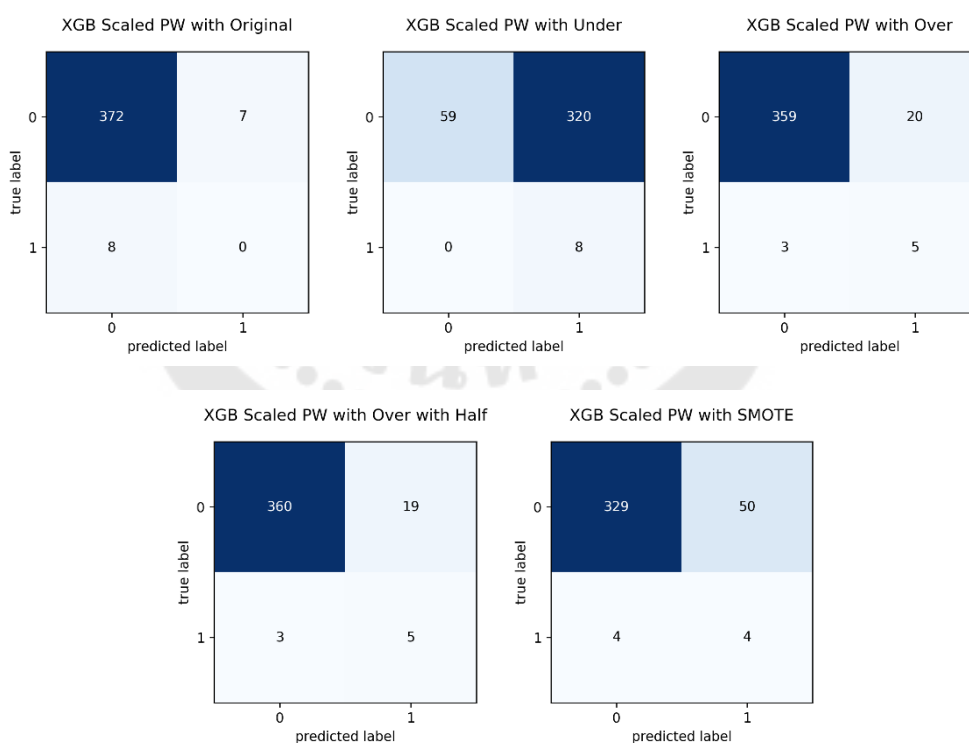
ภาพประกอบ 265 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.14.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9612	0.9815	0.0000	0.0000	0.0000
Under	0.1731	0.1557	1.0000	0.0244	0.0476
Over	0.9406	0.9472	0.6250	0.2000	0.3030
Over with Half	0.9432	0.9499	0.6250	0.2083	0.3125
SMOTE	0.8605	0.8681	0.5000	0.0741	0.1290

ภาพประกอบ 266 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

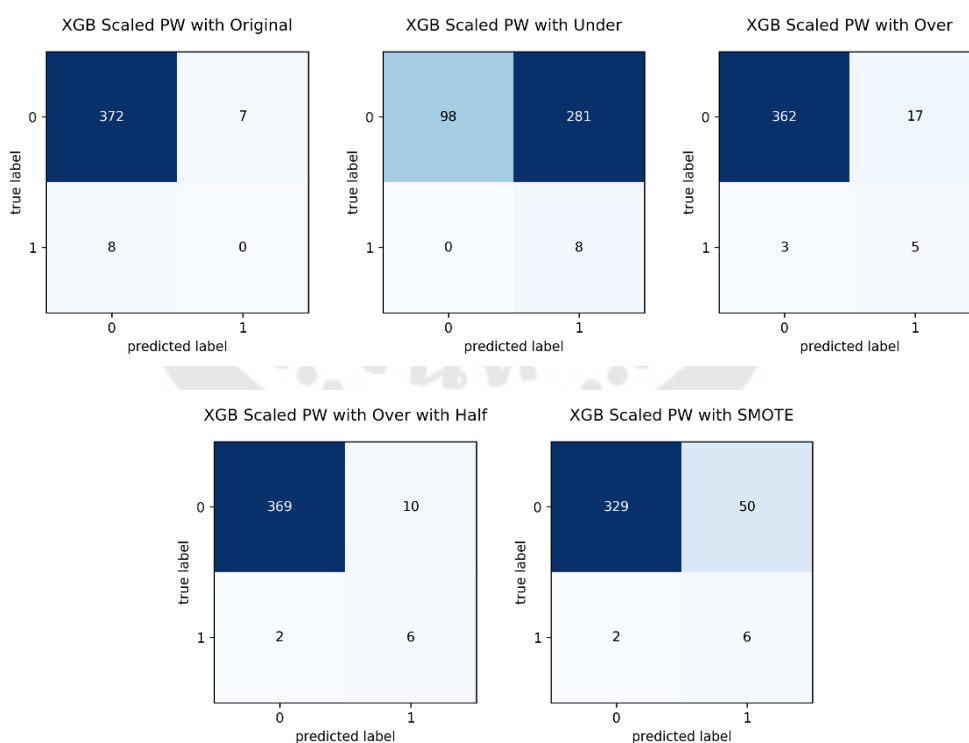


ภาพประกอบ 267 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9612	0.9815	0.0000	0.0000	0.0000
Under	0.2739	0.2586	1.0000	0.0277	0.0539
Over	0.9483	0.9551	0.6250	0.2273	0.3333
Over with Half	0.9690	0.9736	0.7500	0.3750	0.5000
SMOTE	0.8656	0.8681	0.7500	0.1071	0.1875

ภาพประกอบ 268 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

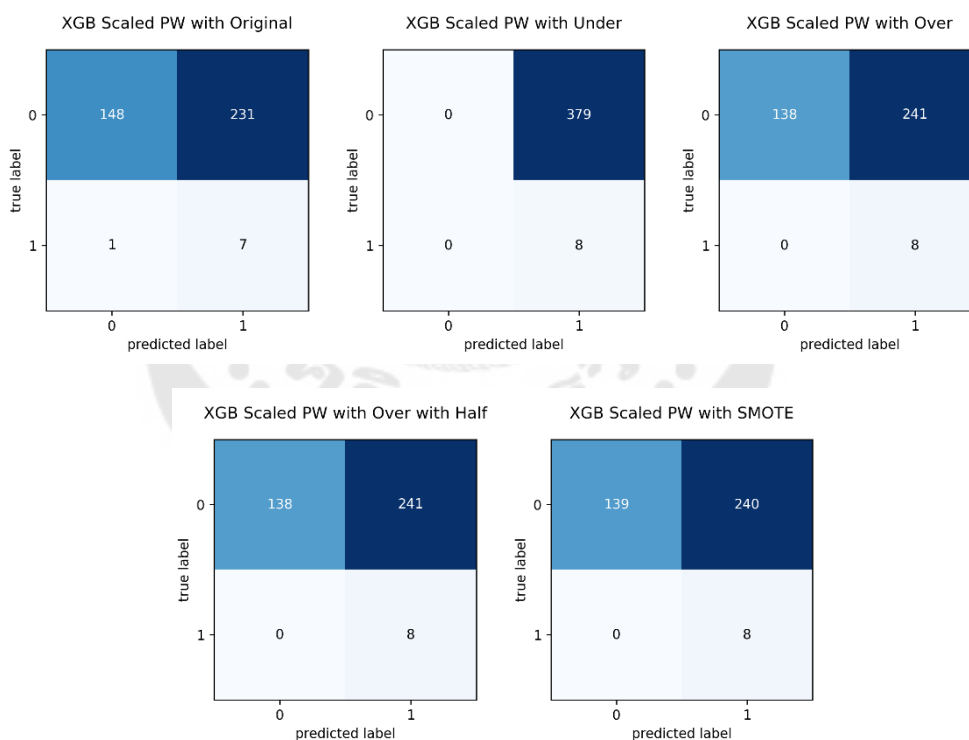


ภาพประกอบ 269 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4005	0.3905	0.8750	0.0294	0.0569
Under	0.0207	0.0000	1.0000	0.0207	0.0405
Over	0.3773	0.3641	1.0000	0.0321	0.0623
Over with Half	0.3773	0.3641	1.0000	0.0321	0.0623
SMOTE	0.3798	0.3668	1.0000	0.0323	0.0625

ภาพประกอบ 270 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



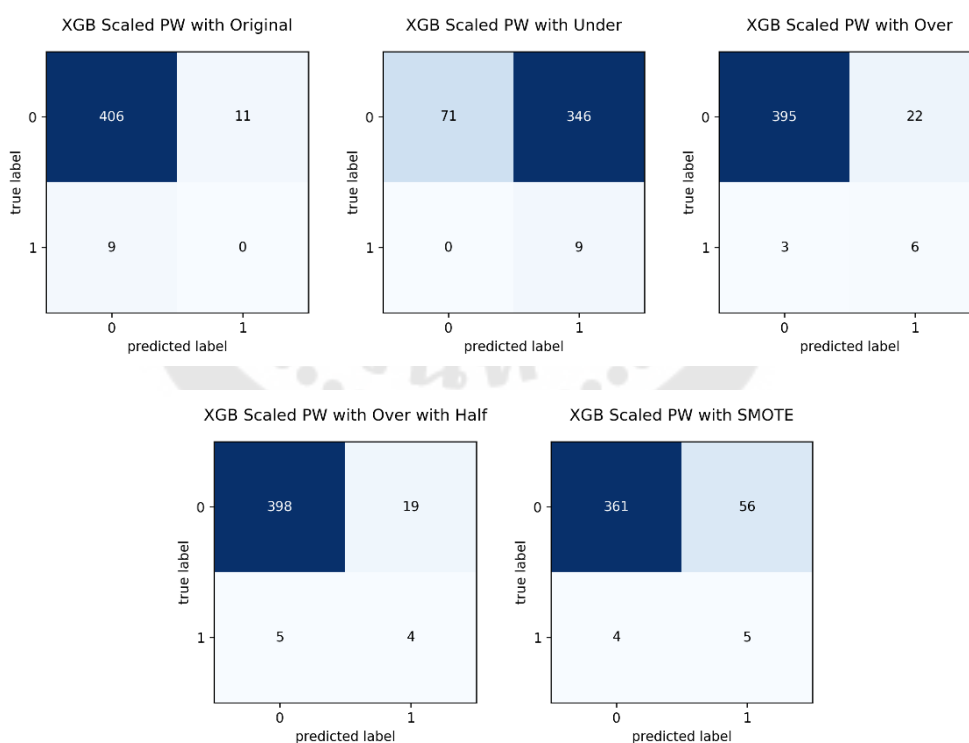
ภาพประกอบ 271 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.14.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9531	0.9736	0.0000	0.0000	0.0000
Under	0.1878	0.1703	1.0000	0.0254	0.0495
Over	0.9413	0.9472	0.6667	0.2143	0.3243
Over with Half	0.9437	0.9544	0.4444	0.1739	0.2500
SMOTE	0.8592	0.8657	0.5556	0.0820	0.1429

ภาพประกอบ 272 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

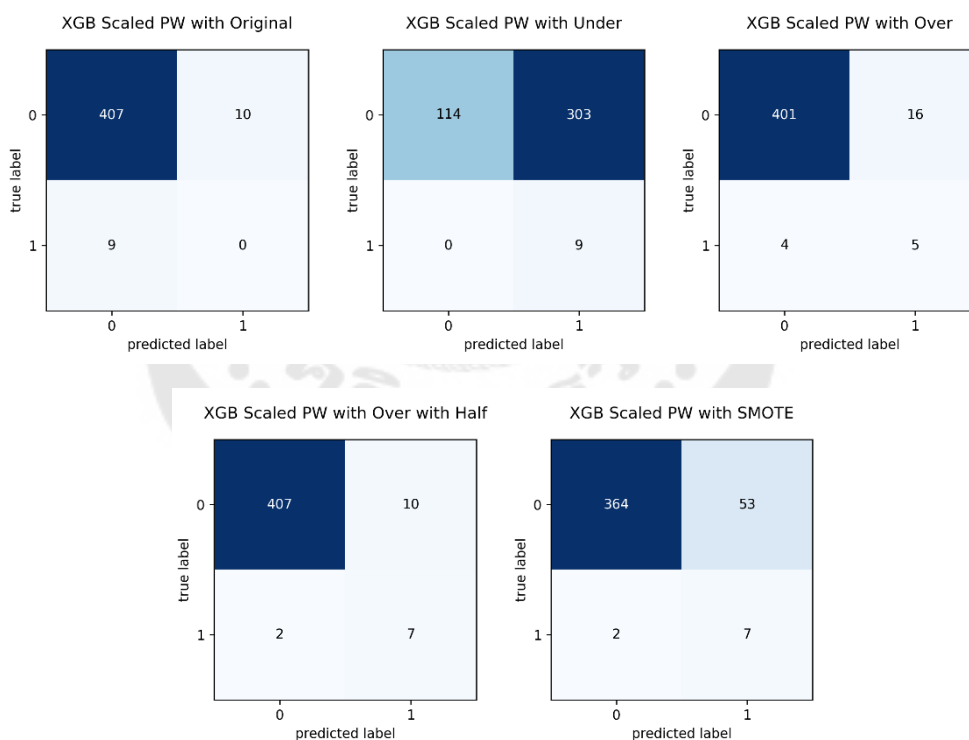


ภาพประกอบ 273 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9554	0.9760	0.0000	0.0000	0.0000
Under	0.2887	0.2734	1.0000	0.0288	0.0561
Over	0.9531	0.9616	0.5556	0.2381	0.3333
Over with Half	0.9718	0.9760	0.7778	0.4118	0.5385
SMOTE	0.8709	0.8729	0.7778	0.1167	0.2029

ภาพประกอบ 274 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

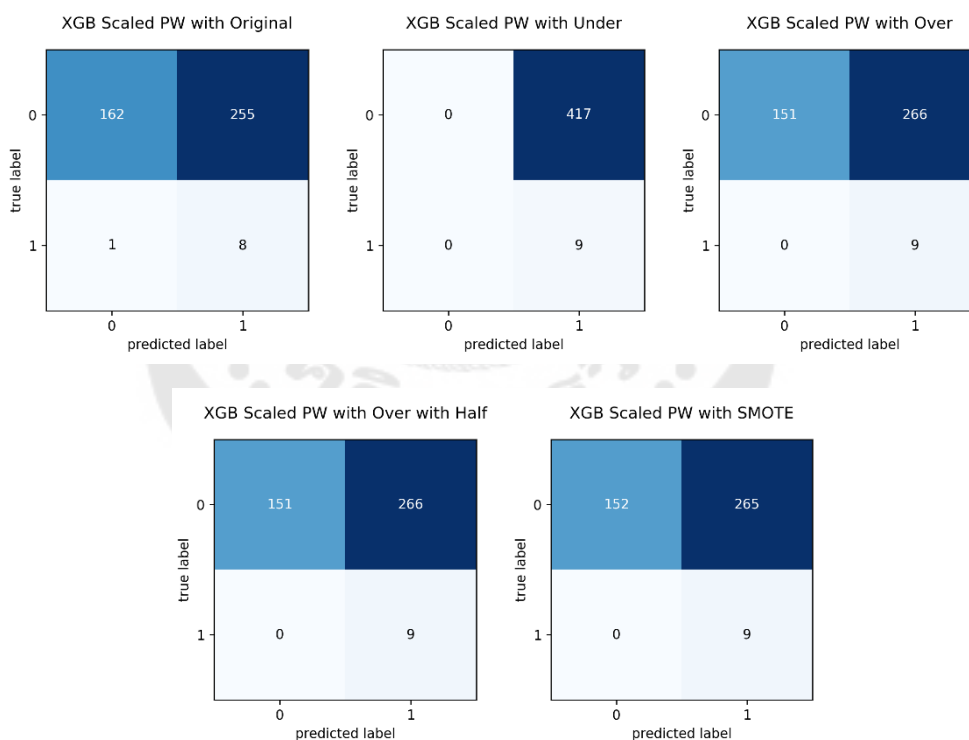


ภาพประกอบ 275 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.14.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3991	0.3885	0.8889	0.0304	0.0588
Under	0.0211	0.0000	1.0000	0.0211	0.0414
Over	0.3756	0.3621	1.0000	0.0327	0.0634
Over with Half	0.3756	0.3621	1.0000	0.0327	0.0634
SMOTE	0.3779	0.3645	1.0000	0.0328	0.0636

ภาพประกอบ 276 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 277 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

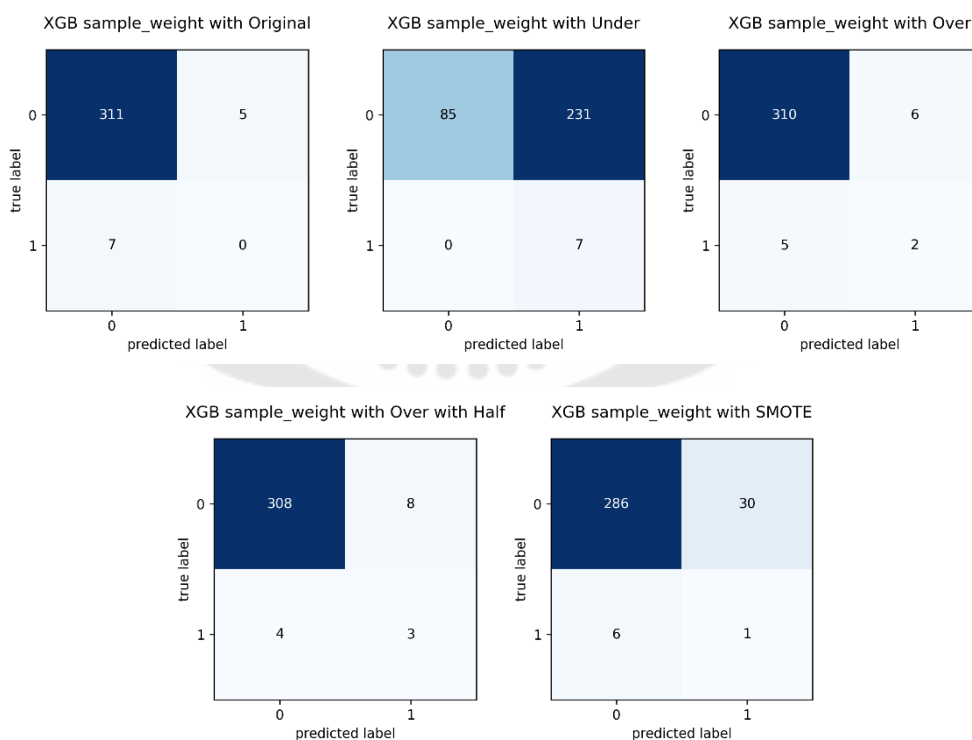
4.15 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ sample_weight

4.15.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.15.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9628	0.9842	0.0000	0.0000	0.0000
Under	0.2848	0.2690	1.0000	0.0294	0.0571
Over	0.9659	0.9810	0.2857	0.2500	0.2667
Over with Half	0.9628	0.9747	0.4286	0.2727	0.3333
SMOTE	0.8885	0.9051	0.1429	0.0323	0.0526

ภาพประกอบ 278 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

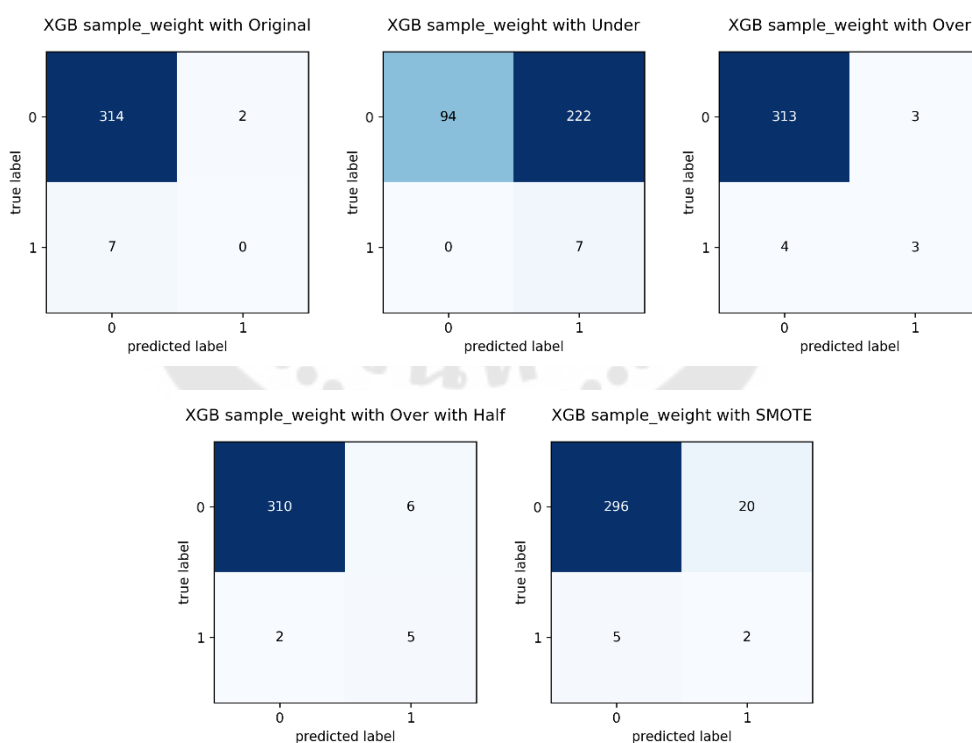


ภาพประกอบ 279 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9721	0.9937	0.0000	0.0000	0.0000
Under	0.3127	0.2975	1.0000	0.0306	0.0593
Over	0.9783	0.9905	0.4286	0.5000	0.4615
Over with Half	0.9752	0.9810	0.7143	0.4545	0.5556
SMOTE	0.9226	0.9367	0.2857	0.0909	0.1379

ภาพประกอบ 280 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

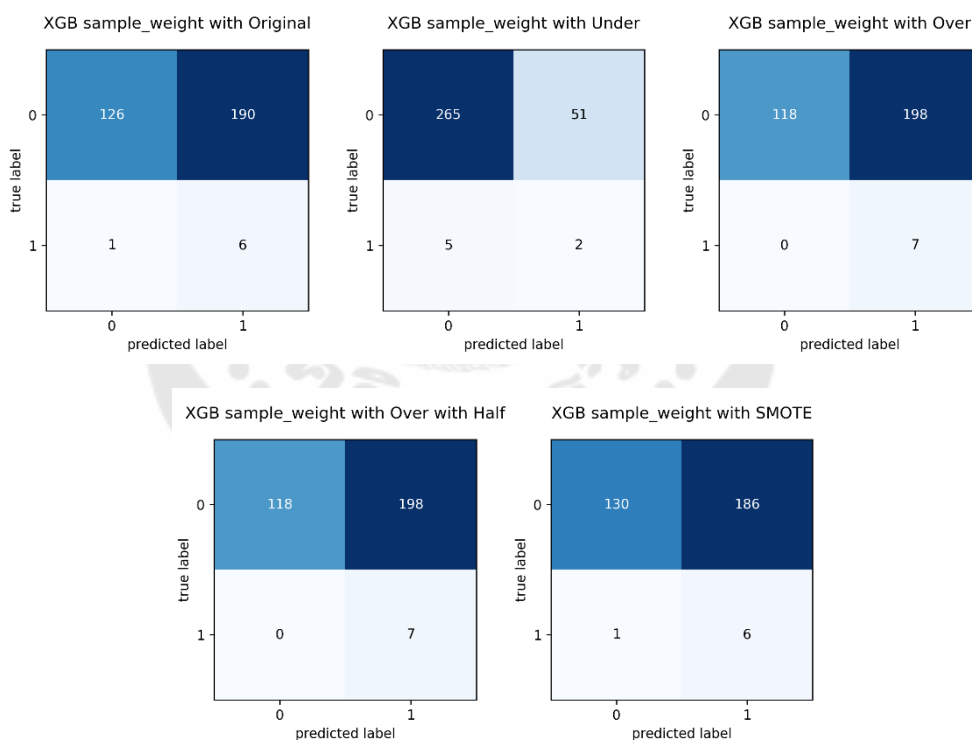


ภาพประกอบ 281 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.4087	0.3987	0.8571	0.0306	0.0591
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3870	0.3734	1.0000	0.0341	0.0660
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 282 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



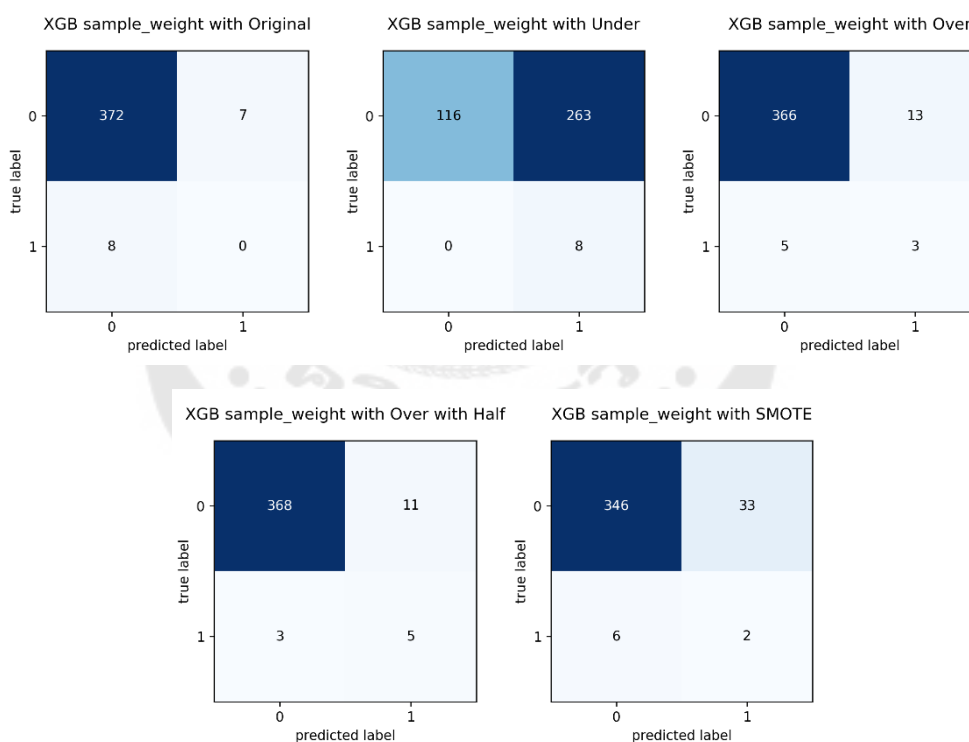
ภาพประกอบ 283 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.15.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9612	0.9815	0.0000	0.0000	0.0000
Under	0.3204	0.3061	1.0000	0.0295	0.0573
Over	0.9535	0.9657	0.3750	0.1875	0.2500
Over with Half	0.9638	0.9710	0.6250	0.3125	0.4167
SMOTE	0.8992	0.9129	0.2500	0.0571	0.0930

ภาพประกอบ 284 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

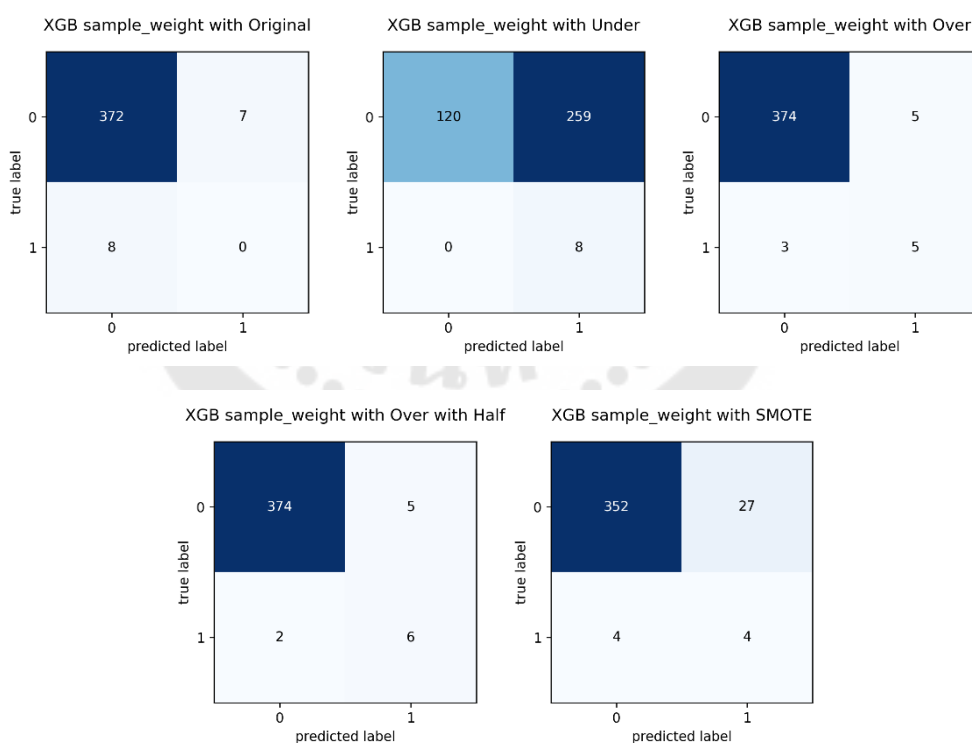


ภาพประกอบ 285 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9612	0.9815	0.0000	0.0000	0.0000
Under	0.3307	0.3166	1.0000	0.0300	0.0582
Over	0.9793	0.9868	0.6250	0.5000	0.5556
Over with Half	0.9819	0.9868	0.7500	0.5455	0.6316
SMOTE	0.9199	0.9288	0.5000	0.1290	0.2051

ภาพประกอบ 286 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

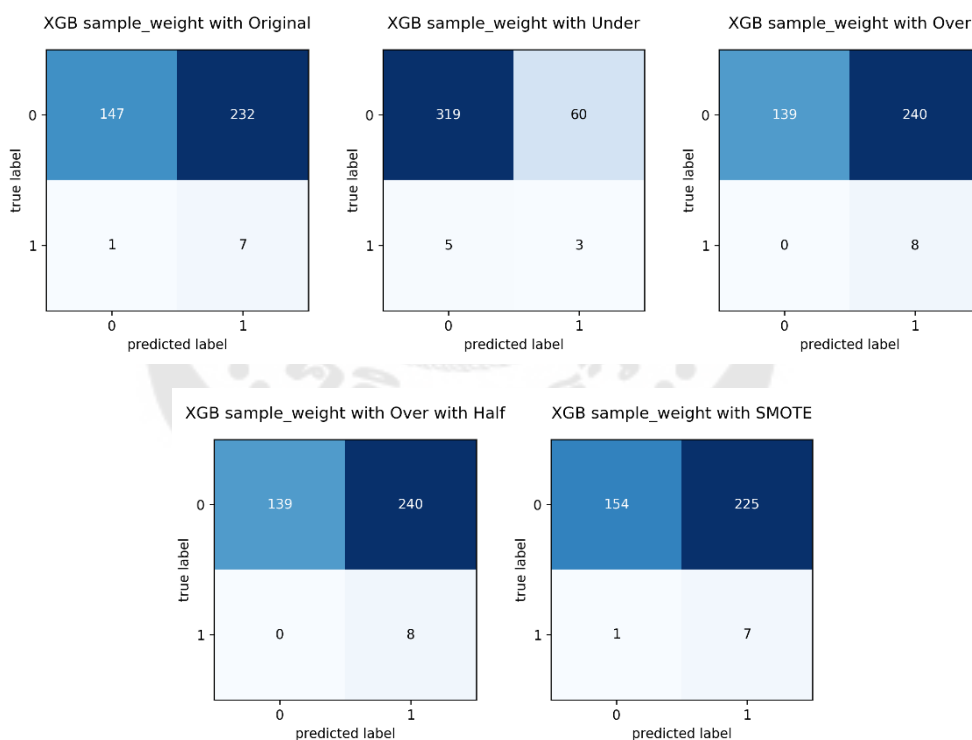


ภาพประกอบ 287 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3979	0.3879	0.8750	0.0293	0.0567
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3798	0.3668	1.0000	0.0323	0.0625
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 288 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



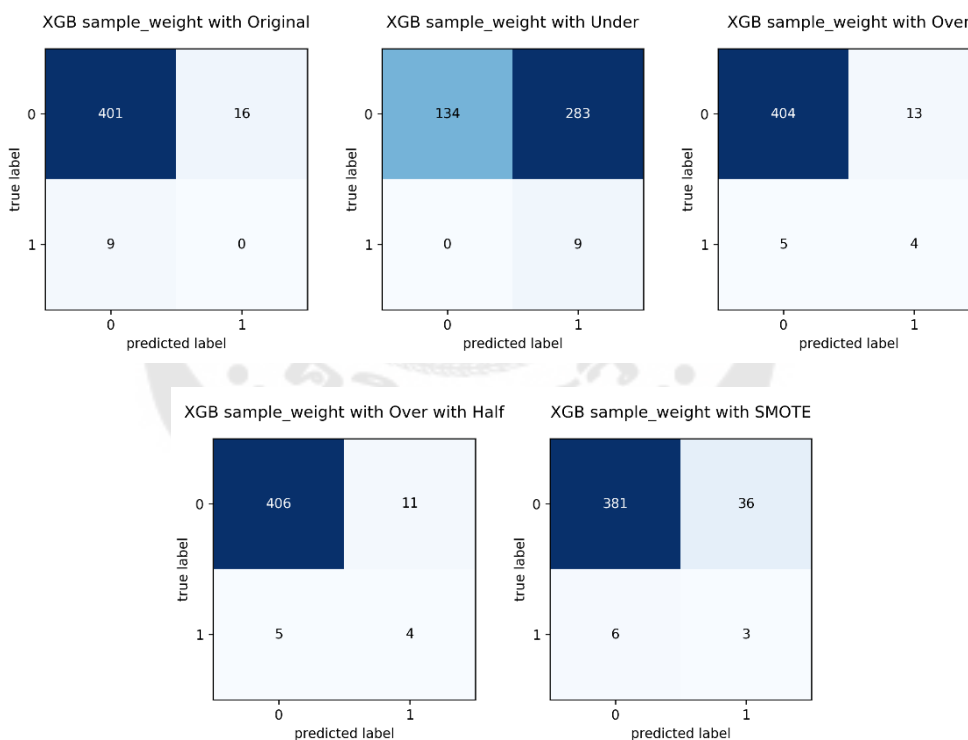
ภาพประกอบ 289 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.15.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9413	0.9616	0.0000	0.0000	0.0000
Under	0.3357	0.3213	1.0000	0.0308	0.0598
Over	0.9577	0.9688	0.4444	0.2353	0.3077
Over with Half	0.9624	0.9736	0.4444	0.2667	0.3333
SMOTE	0.9014	0.9137	0.3333	0.0769	0.1250

ภาพประกอบ 290 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

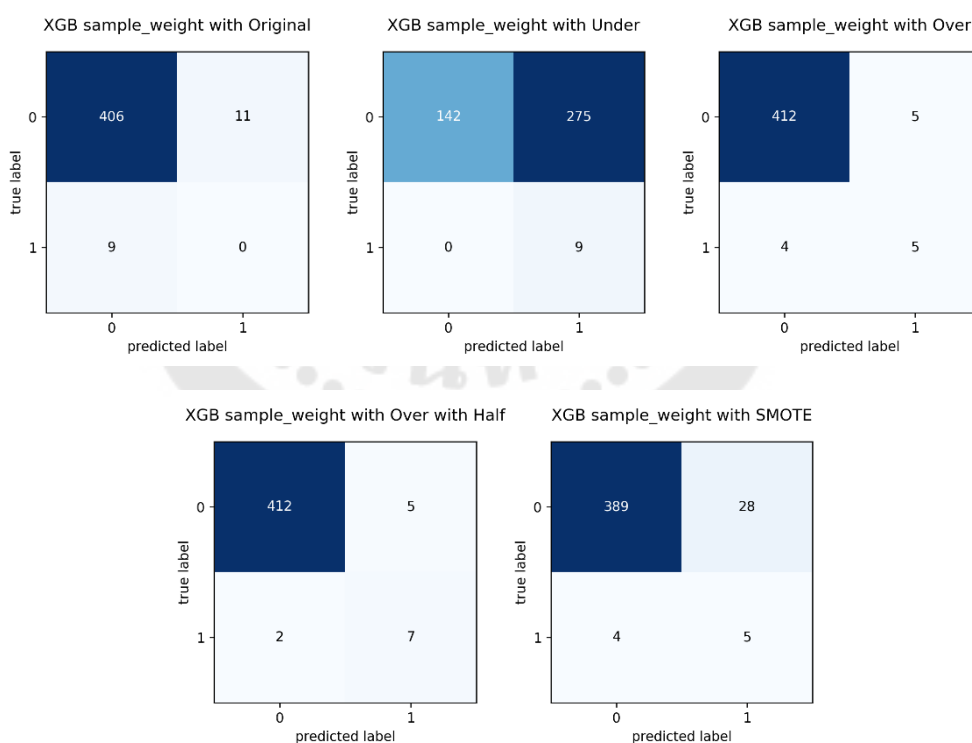


ภาพประกอบ 291 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9531	0.9736	0.0000	0.0000	0.0000
Under	0.3545	0.3405	1.0000	0.0317	0.0614
Over	0.9789	0.9880	0.5556	0.5000	0.5263
Over with Half	0.9836	0.9880	0.7778	0.5833	0.6667
SMOTE	0.9249	0.9329	0.5556	0.1515	0.2381

ภาพประกอบ 292 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

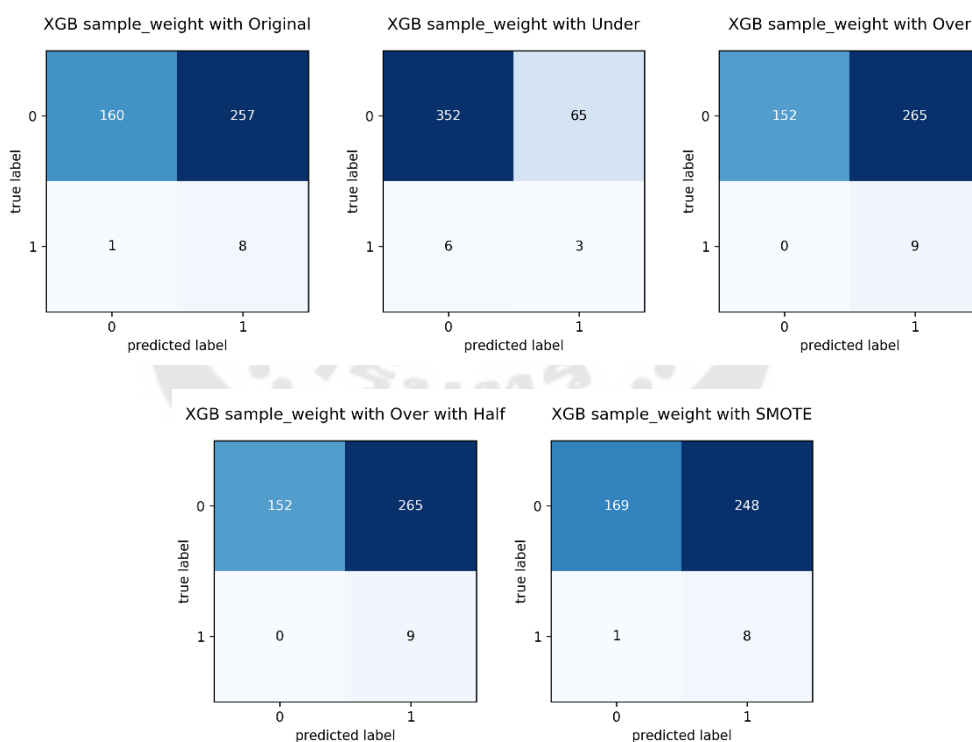


ภาพประกอบ 293 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.15.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3944	0.3837	0.8889	0.0302	0.0584
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 294 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 295 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

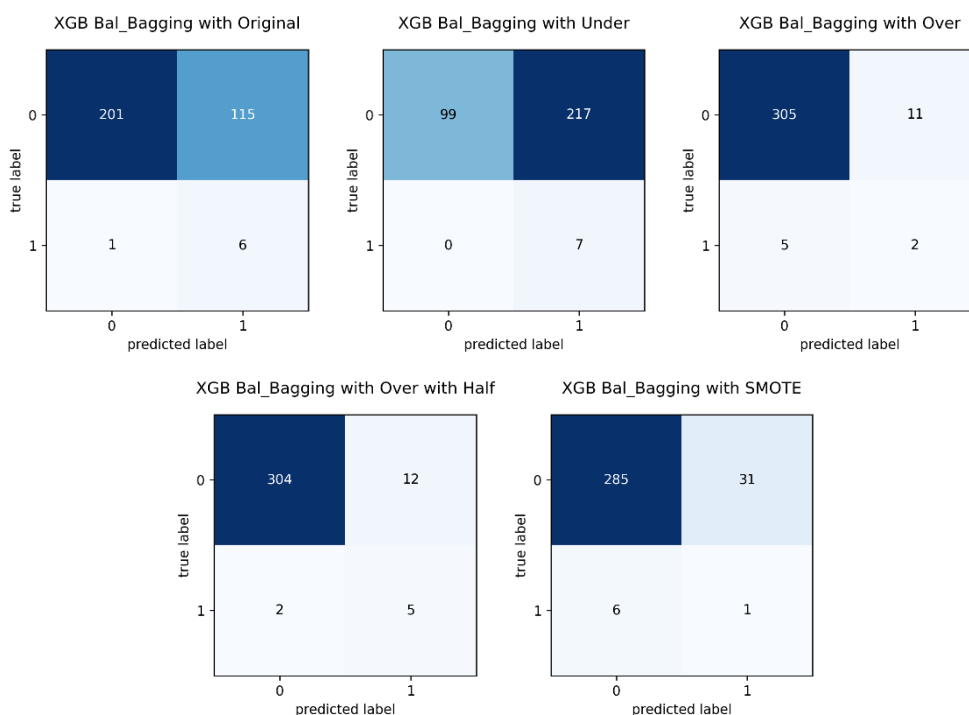
4.16 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Default Parameters และ Balanced Bagging Classifier

4.16.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.16.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6409	0.6361	0.8571	0.0496	0.0938
Under	0.3282	0.3133	1.0000	0.0312	0.0606
Over	0.9505	0.9652	0.2857	0.1538	0.2000
Over with Half	0.9567	0.9620	0.7143	0.2941	0.4167
SMOTE	0.8854	0.9019	0.1429	0.0312	0.0513

ภาพประกอบ 296 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

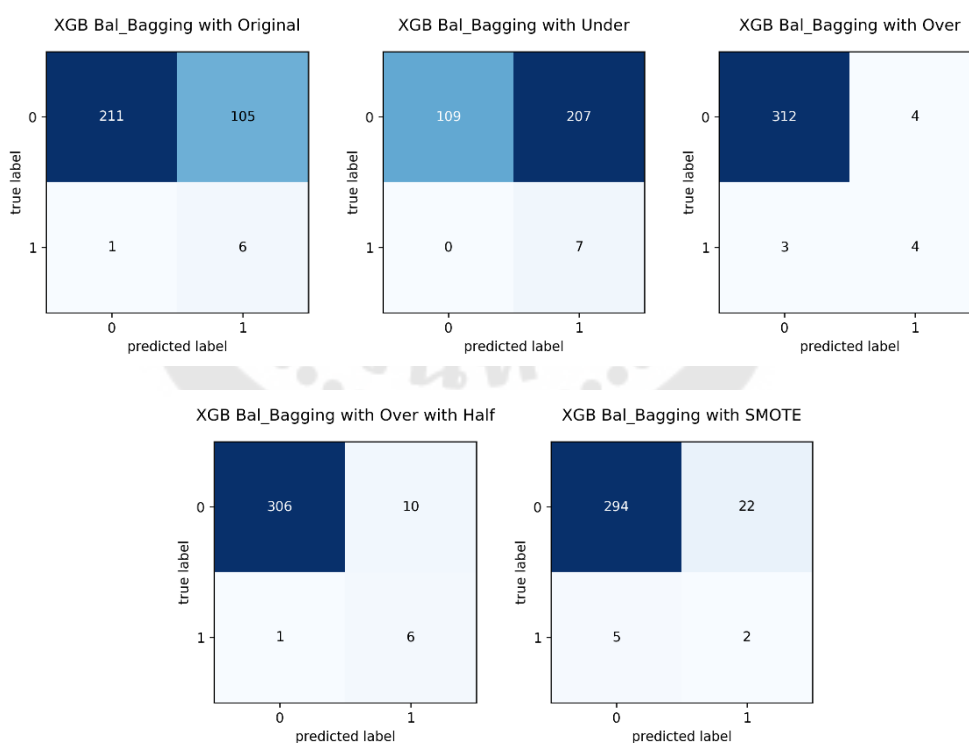


ภาพประกอบ 297 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6718	0.6677	0.8571	0.0541	0.1017
Under	0.3591	0.3449	1.0000	0.0327	0.0633
Over	0.9783	0.9873	0.5714	0.5000	0.5333
Over with Half	0.9659	0.9684	0.8571	0.3750	0.5217
SMOTE	0.9164	0.9304	0.2857	0.0833	0.1290

ภาพประกอบ 298 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

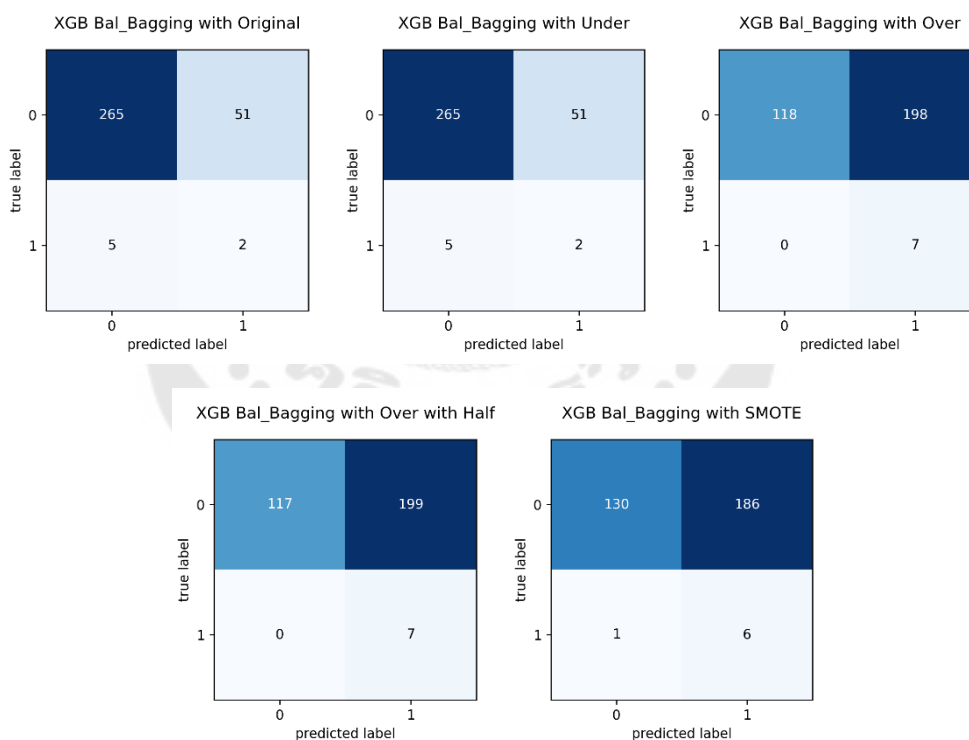


ภาพประกอบ 299 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8266	0.8386	0.2857	0.0377	0.0667
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3870	0.3734	1.0000	0.0341	0.0660
Over with Half	0.3839	0.3703	1.0000	0.0340	0.0657
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 300 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



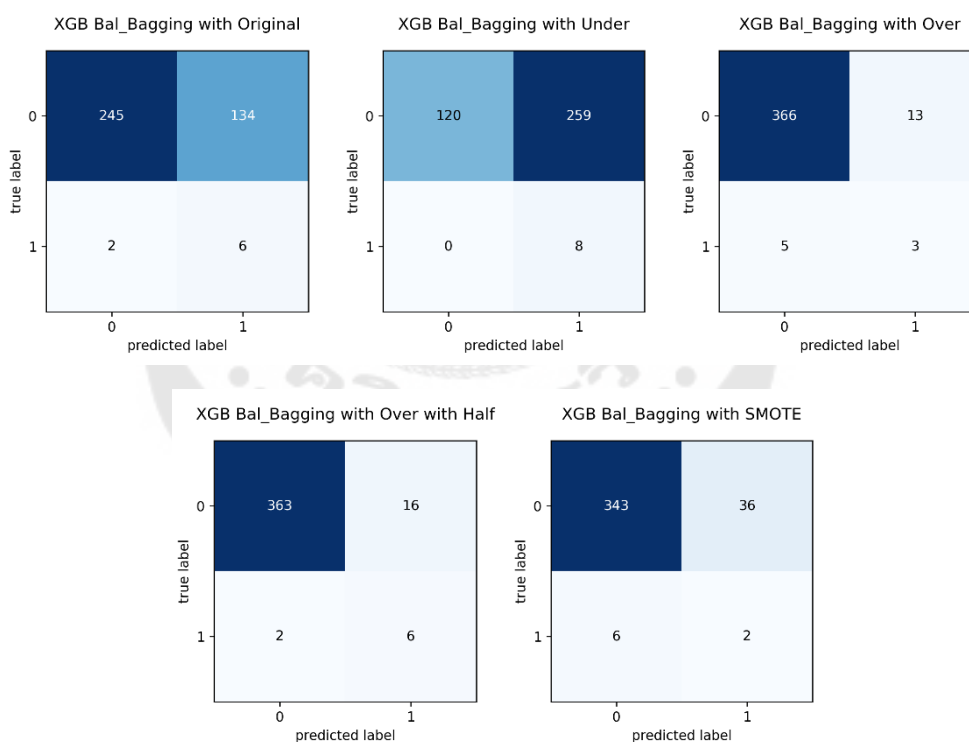
ภาพประกอบ 301 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.16.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6486	0.6464	0.7500	0.0429	0.0811
Under	0.3307	0.3166	1.0000	0.0300	0.0582
Over	0.9535	0.9657	0.3750	0.1875	0.2500
Over with Half	0.9535	0.9578	0.7500	0.2727	0.4000
SMOTE	0.8915	0.9050	0.2500	0.0526	0.0870

ภาพประกอบ 302 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

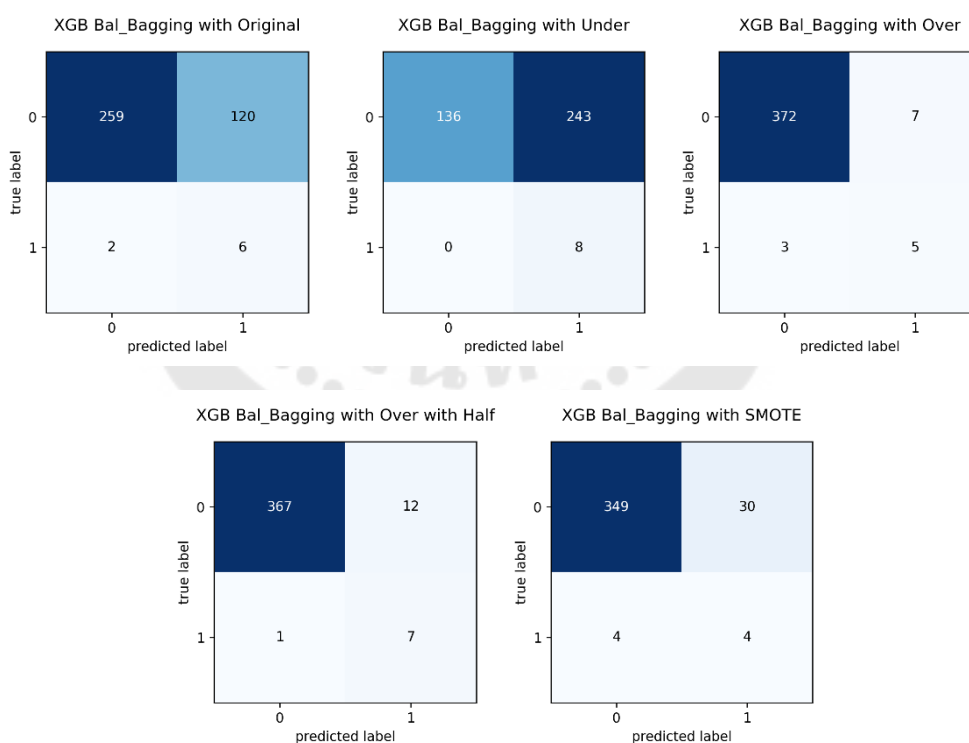


ภาพประกอบ 303 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6848	0.6834	0.7500	0.0476	0.0896
Under	0.3721	0.3588	1.0000	0.0319	0.0618
Over	0.9742	0.9815	0.6250	0.4167	0.5000
Over with Half	0.9664	0.9683	0.8750	0.3684	0.5185
SMOTE	0.9121	0.9208	0.5000	0.1176	0.1905

ภาพประกอบ 304 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

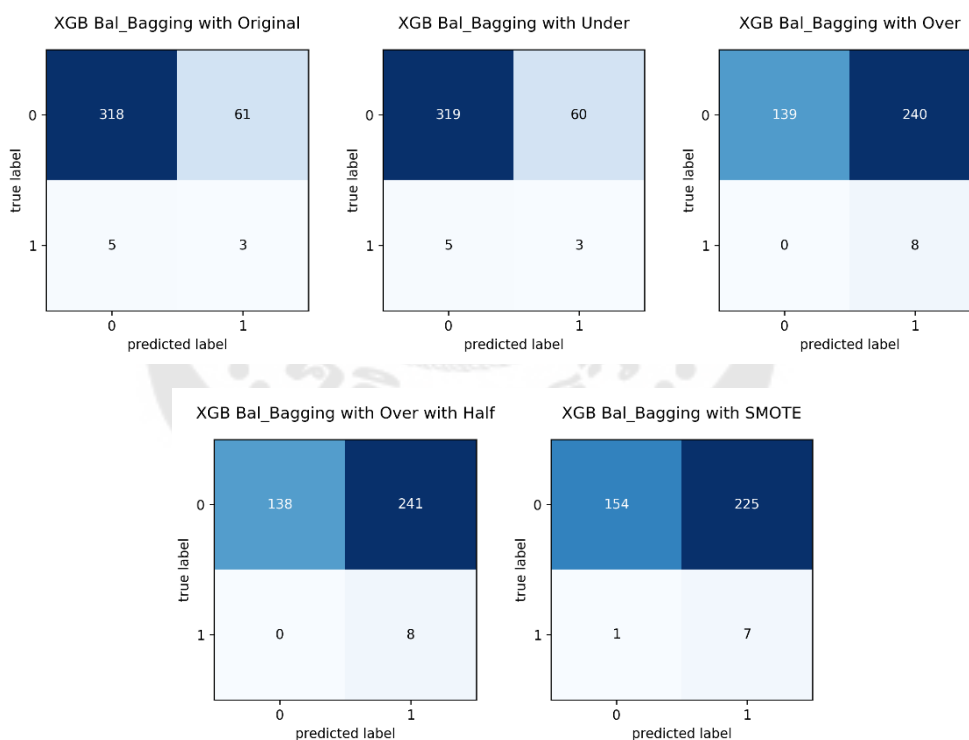


ภาพประกอบ 305 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8295	0.8391	0.3750	0.0469	0.0833
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3798	0.3668	1.0000	0.0323	0.0625
Over with Half	0.3773	0.3641	1.0000	0.0321	0.0623
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 306 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



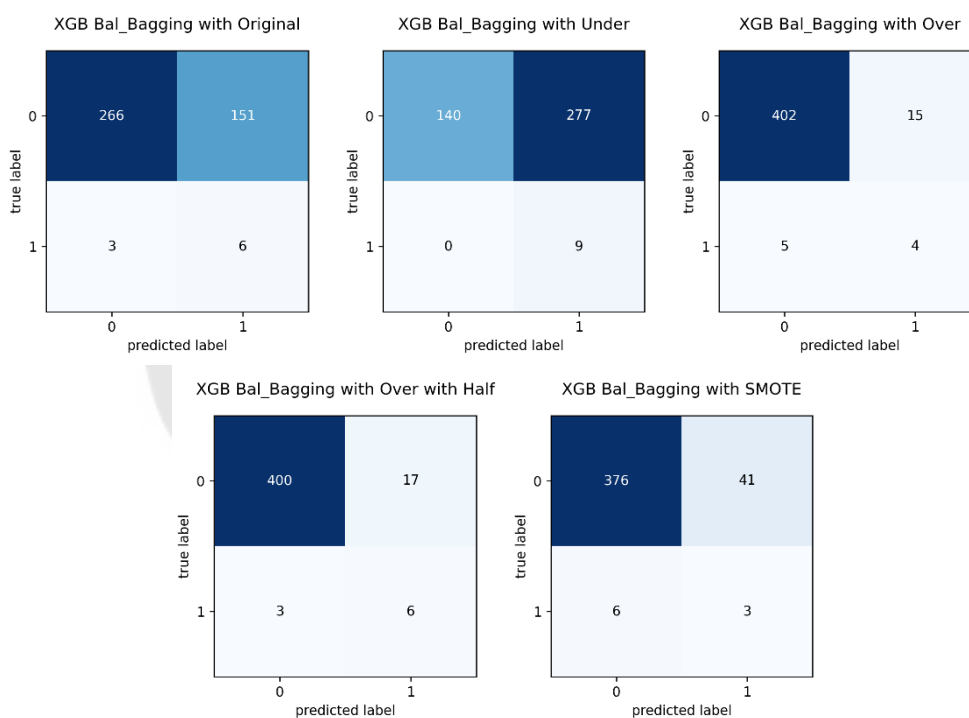
ภาพประกอบ 307 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.16.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6385	0.6379	0.6667	0.0382	0.0723
Under	0.3498	0.3357	1.0000	0.0315	0.0610
Over	0.9531	0.9640	0.4444	0.2105	0.2857
Over with Half	0.9531	0.9592	0.6667	0.2609	0.3750
SMOTE	0.8897	0.9017	0.3333	0.0682	0.1132

ภาพประกอบ 308 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

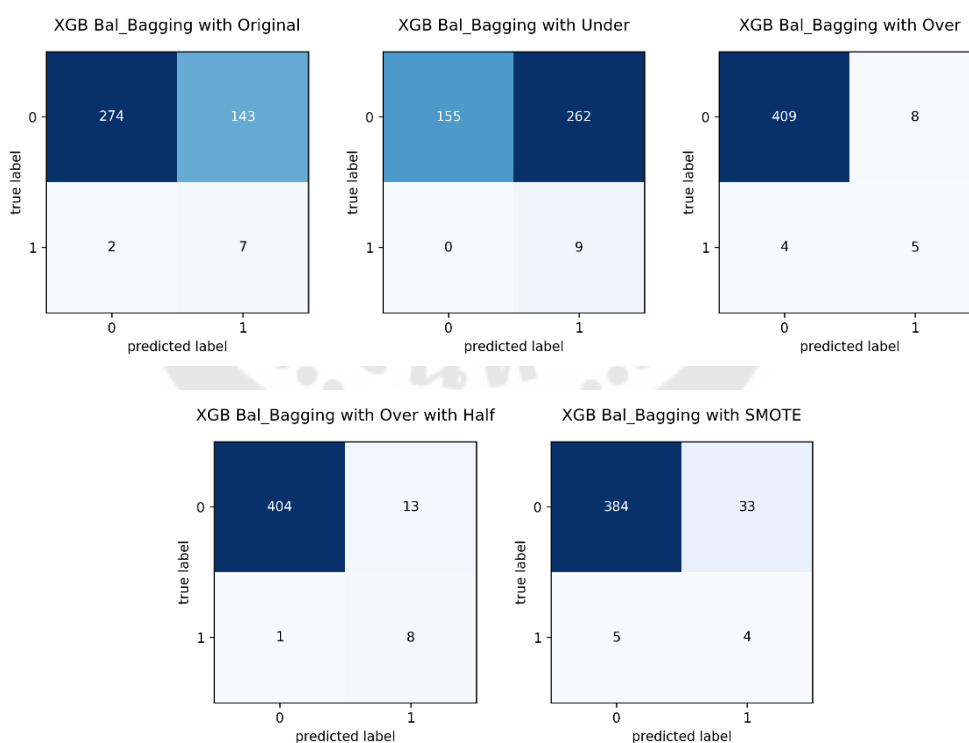


ภาพประกอบ 309 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6596	0.6571	0.7778	0.0467	0.0881
Under	0.3850	0.3717	1.0000	0.0332	0.0643
Over	0.9718	0.9808	0.5556	0.3846	0.4545
Over with Half	0.9671	0.9688	0.8889	0.3810	0.5333
SMOTE	0.9108	0.9209	0.4444	0.1081	0.1739

ภาพประกอบ 310 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

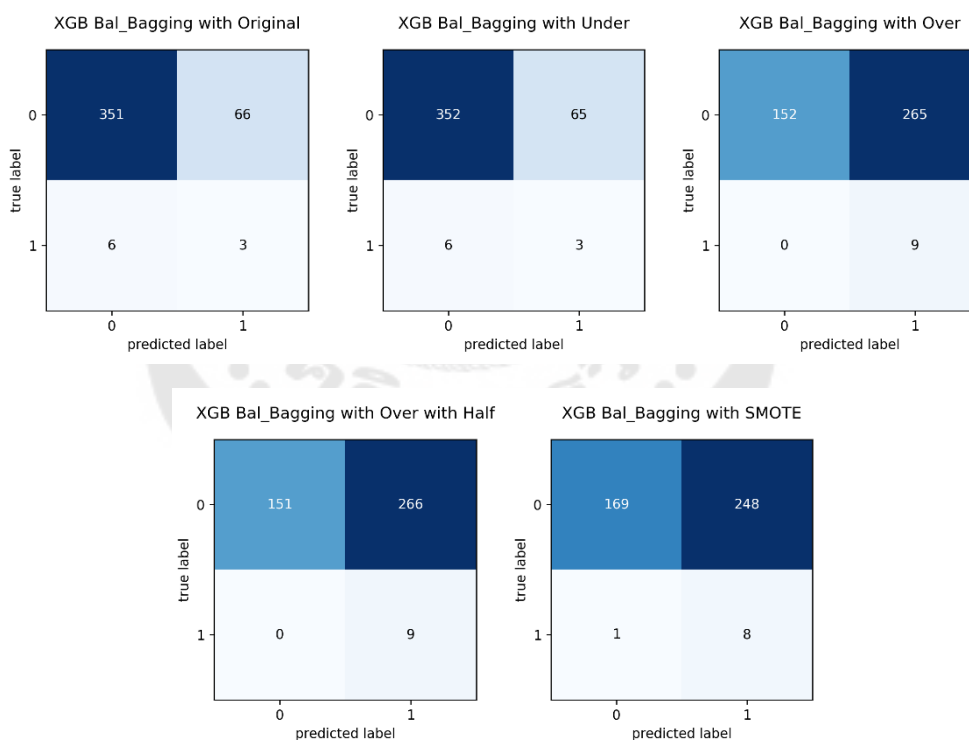


ภาพประกอบ 311 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.16.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8310	0.8417	0.3333	0.0435	0.0769
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3756	0.3621	1.0000	0.0327	0.0634
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 312 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Default Parameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 313 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

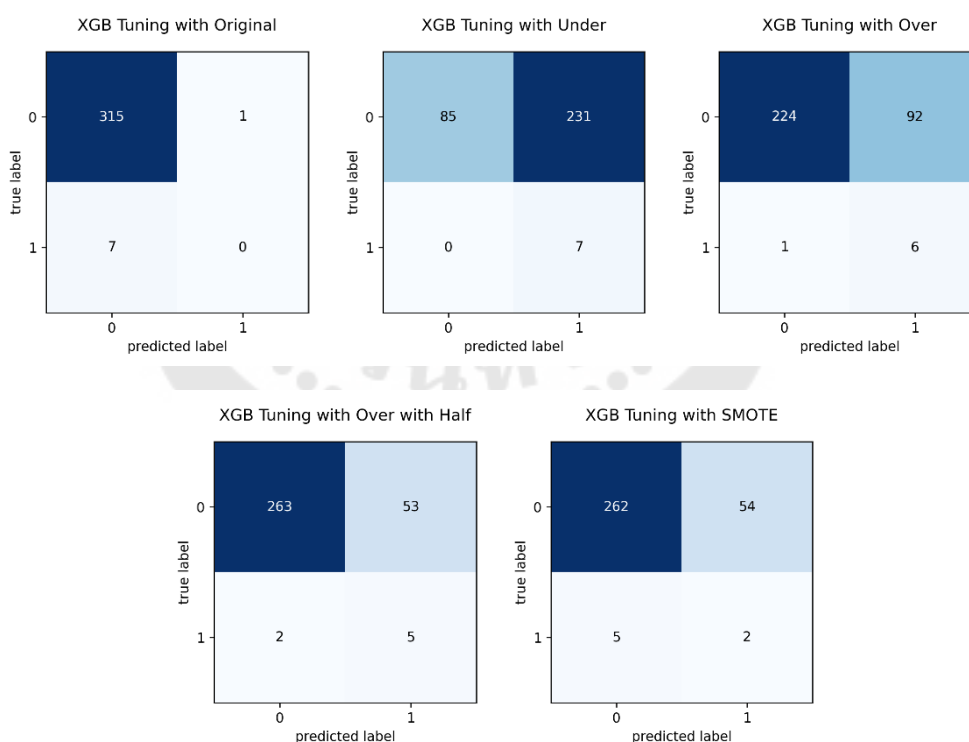
4.17 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters

4.17.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.17.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.2848	0.2690	1.0000	0.0294	0.0571
Over	0.7121	0.7089	0.8571	0.0612	0.1143
Over with Half	0.8297	0.8323	0.7143	0.0862	0.1538
SMOTE	0.8173	0.8291	0.2857	0.0357	0.0635

ภาพประกอบ 314 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

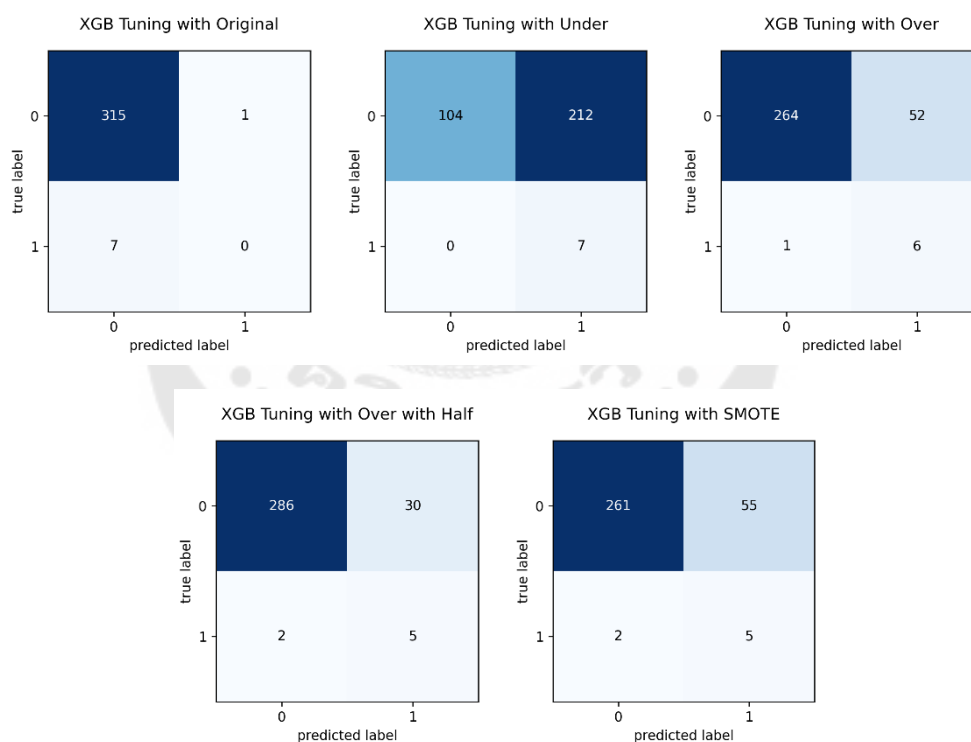


ภาพประกอบ 315 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9752	0.9968	0.0000	0.0000	0.0000
Under	0.3437	0.3291	1.0000	0.0320	0.0619
Over	0.8359	0.8354	0.8571	0.1034	0.1846
Over with Half	0.9009	0.9051	0.7143	0.1429	0.2381
SMOTE	0.8235	0.8259	0.7143	0.0833	0.1493

ภาพประกอบ 316 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

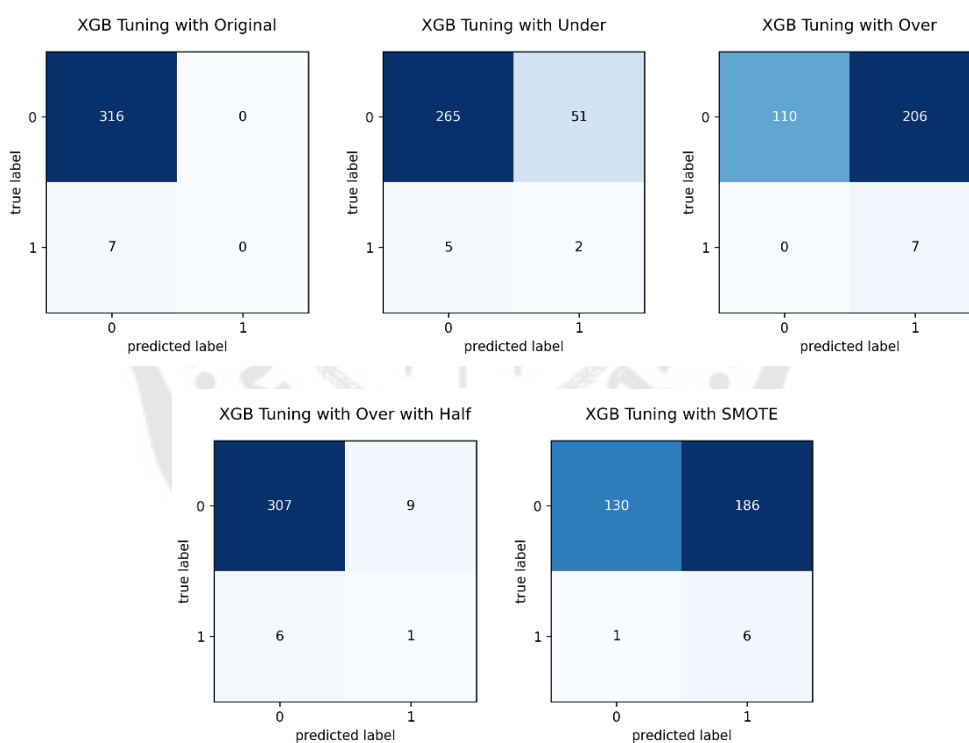


ภาพประกอบ 317 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9783	1.0000	0.0000	nan	0.0000
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3622	0.3481	1.0000	0.0329	0.0636
Over with Half	0.9536	0.9715	0.1429	0.1000	0.1176
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 318 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



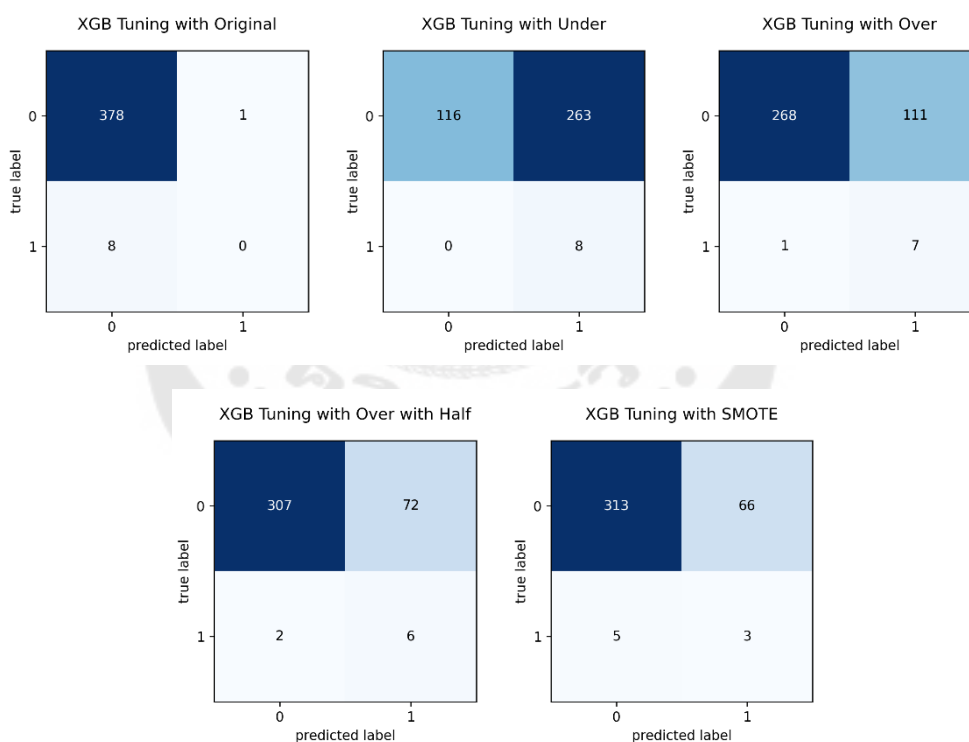
ภาพประกอบ 319 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.17.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9767	0.9974	0.0000	0.0000	0.0000
Under	0.3204	0.3061	1.0000	0.0295	0.0573
Over	0.7106	0.7071	0.8750	0.0593	0.1111
Over with Half	0.8088	0.8100	0.7500	0.0769	0.1395
SMOTE	0.8165	0.8259	0.3750	0.0435	0.0779

ภาพประกอบ 320 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

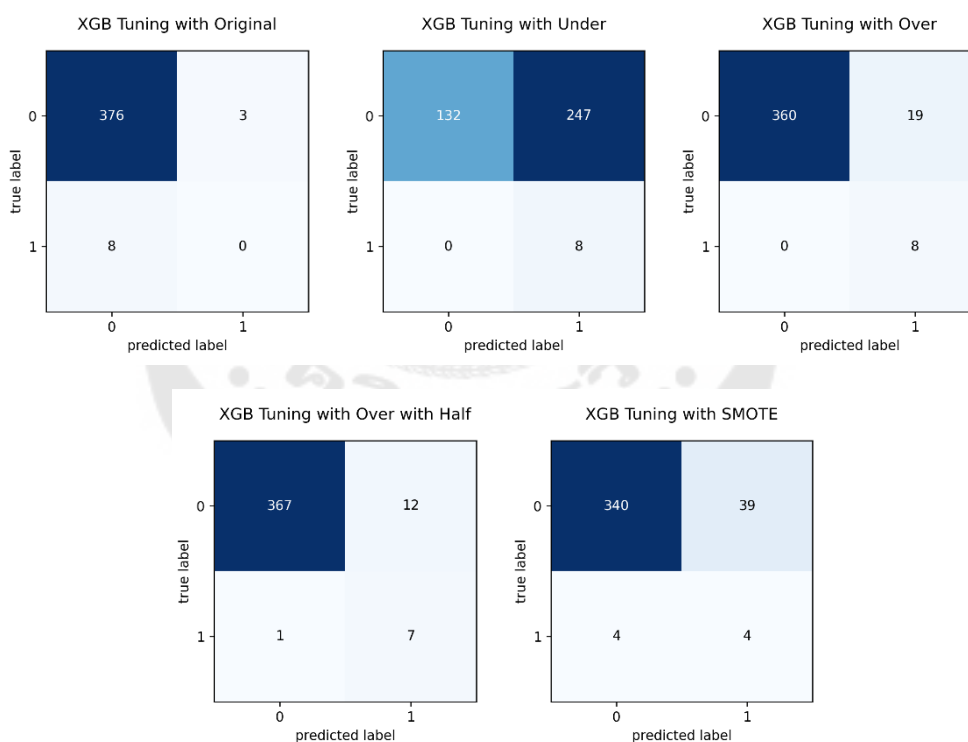


ภาพประกอบ 321 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9716	0.9921	0.0000	0.0000	0.0000
Under	0.3618	0.3483	1.0000	0.0314	0.0608
Over	0.9509	0.9499	1.0000	0.2963	0.4571
Over with Half	0.9664	0.9683	0.8750	0.3684	0.5185
SMOTE	0.8889	0.8971	0.5000	0.0930	0.1569

ภาพประกอบ 322 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

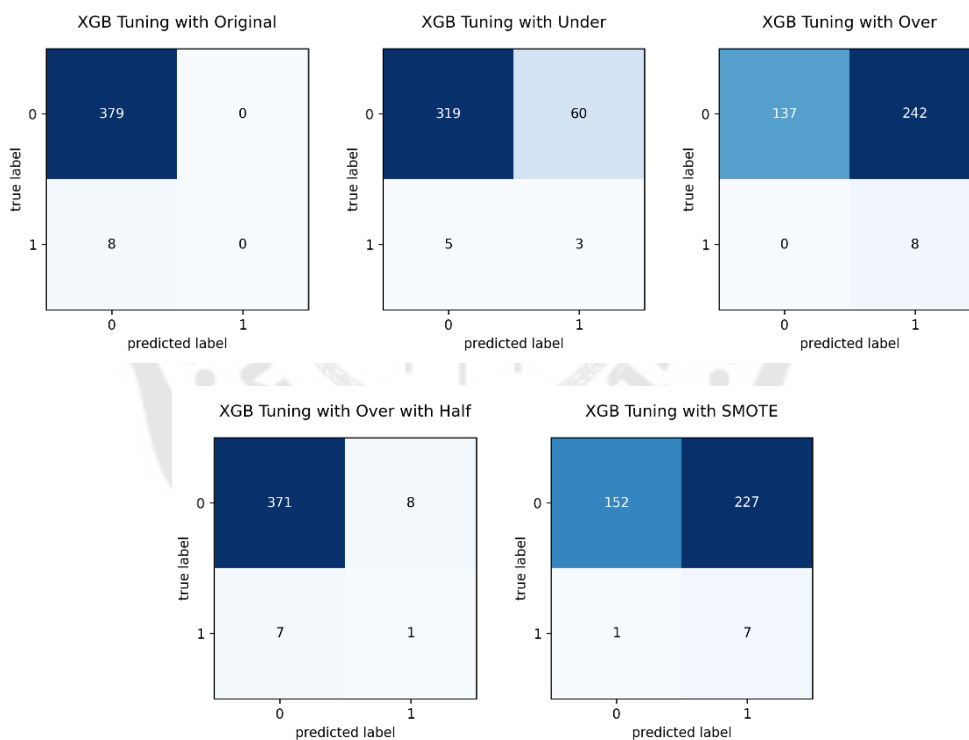


ภาพประกอบ 323 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9793	1.0000	0.0000	nan	0.0000
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3747	0.3615	1.0000	0.0320	0.0620
Over with Half	0.9612	0.9789	0.1250	0.1111	0.1176
SMOTE	0.4109	0.4011	0.8750	0.0299	0.0579

ภาพประกอบ 324 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



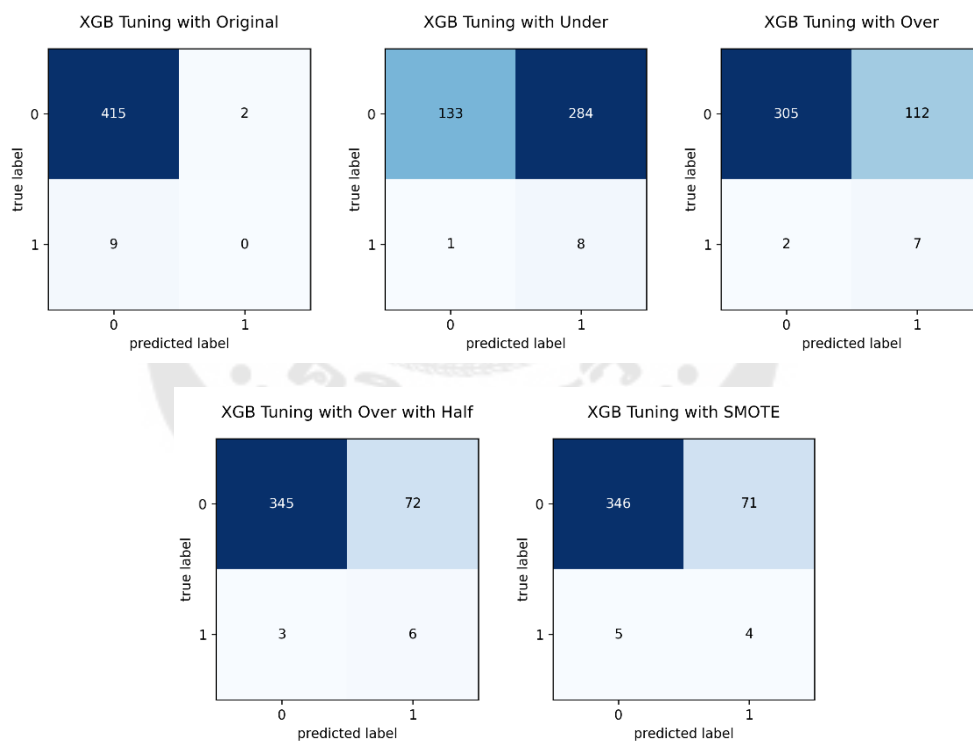
ภาพประกอบ 325 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.17.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9742	0.9952	0.0000	0.0000	0.0000
Under	0.3310	0.3189	0.8889	0.0274	0.0532
Over	0.7324	0.7314	0.7778	0.0588	0.1094
Over with Half	0.8239	0.8273	0.6667	0.0769	0.1379
SMOTE	0.8216	0.8297	0.4444	0.0533	0.0952

ภาพประกอบ 326 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

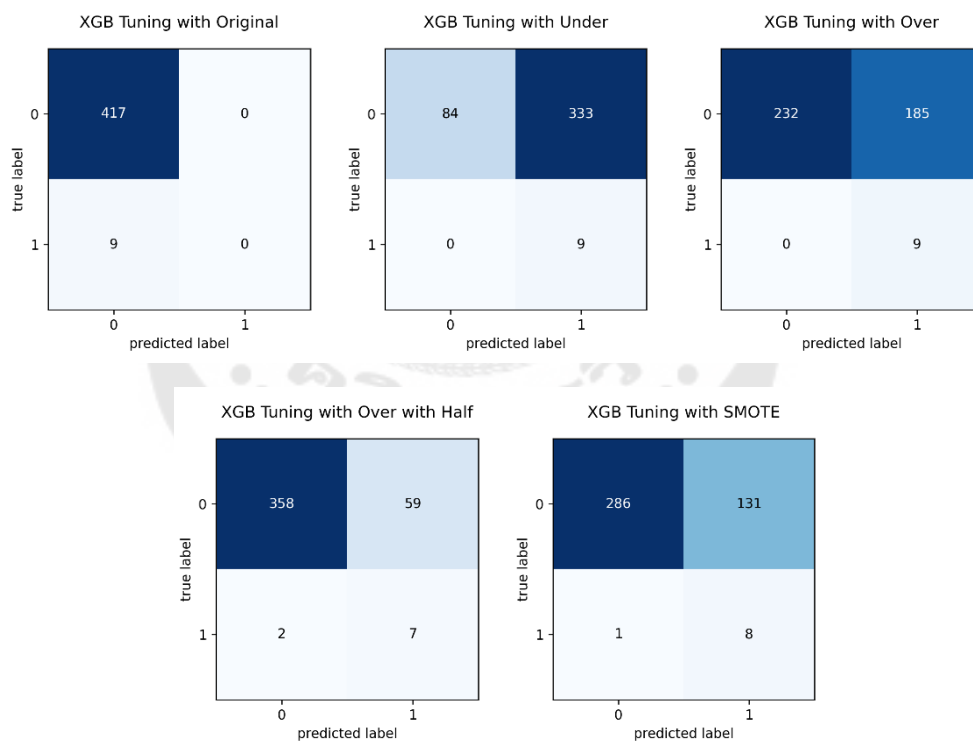


ภาพประกอบ 327 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.2183	0.2014	1.0000	0.0263	0.0513
Over	0.5657	0.5564	1.0000	0.0464	0.0887
Over with Half	0.8568	0.8585	0.7778	0.1061	0.1867
SMOTE	0.6901	0.6859	0.8889	0.0576	0.1081

ภาพประกอบ 328 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

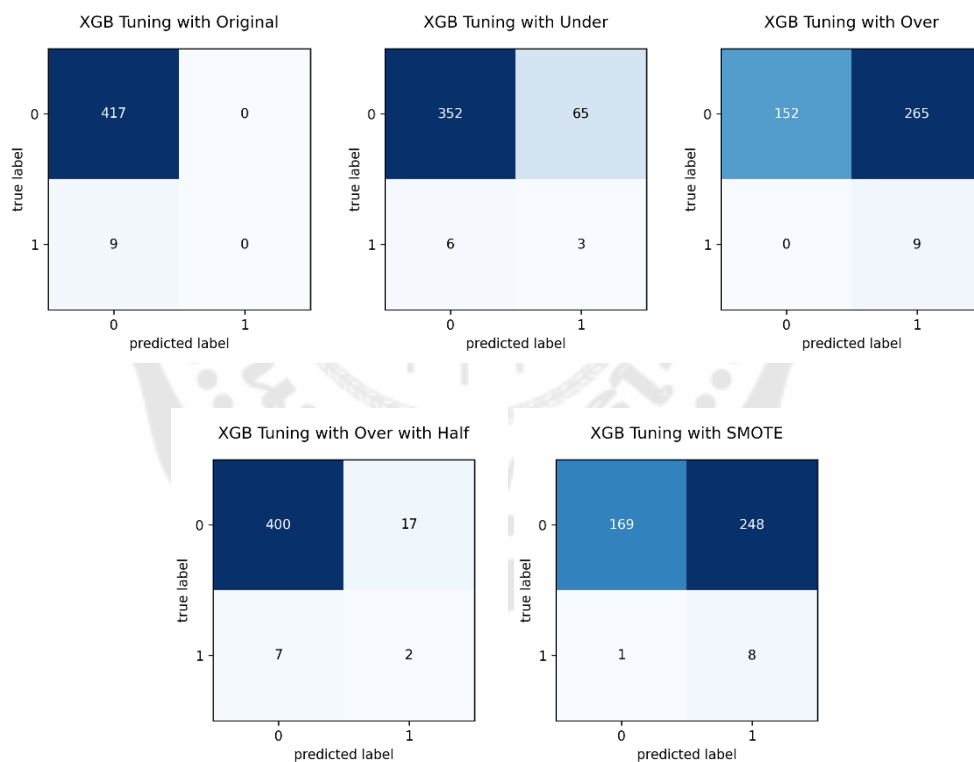


ภาพประกอบ 329 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.17.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9789	1.0000	0.0000	nan	0.0000
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.9437	0.9592	0.2222	0.1053	0.1429
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 330 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 331 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight

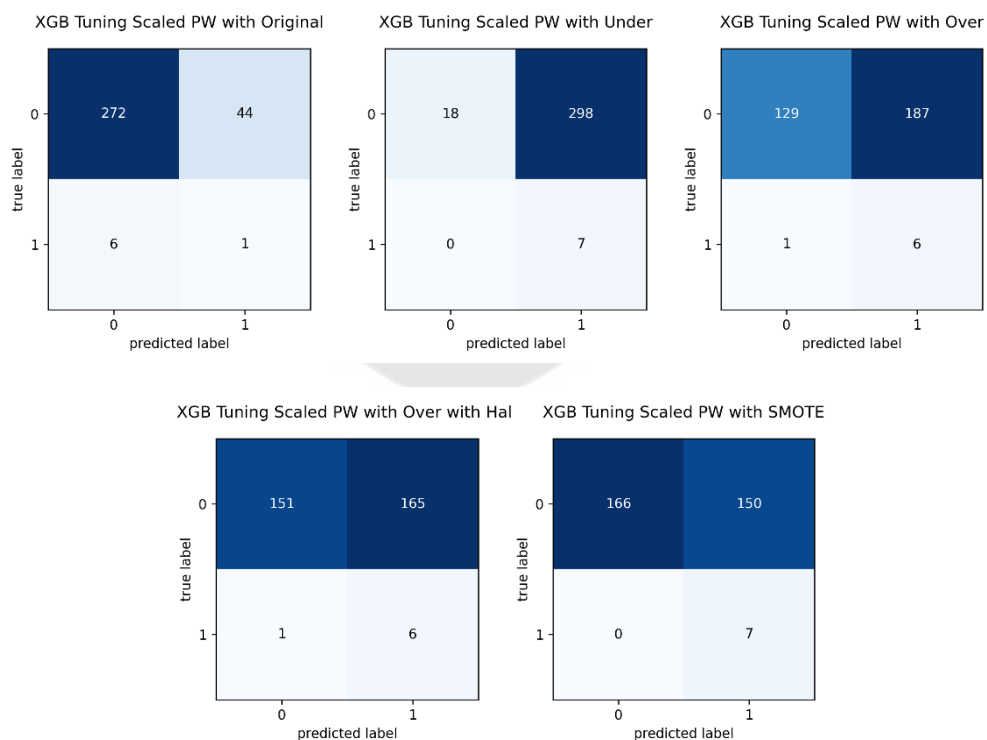
4.18.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.18.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8452	0.8608	0.1429	0.0222	0.0385
Under	0.0774	0.0570	1.0000	0.0230	0.0449
Over	0.4180	0.4082	0.8571	0.0311	0.0600
Over with Half	0.4861	0.4778	0.8571	0.0351	0.0674
SMOTE	0.5356	0.5253	1.0000	0.0446	0.0854

ภาพประกอบ 332 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

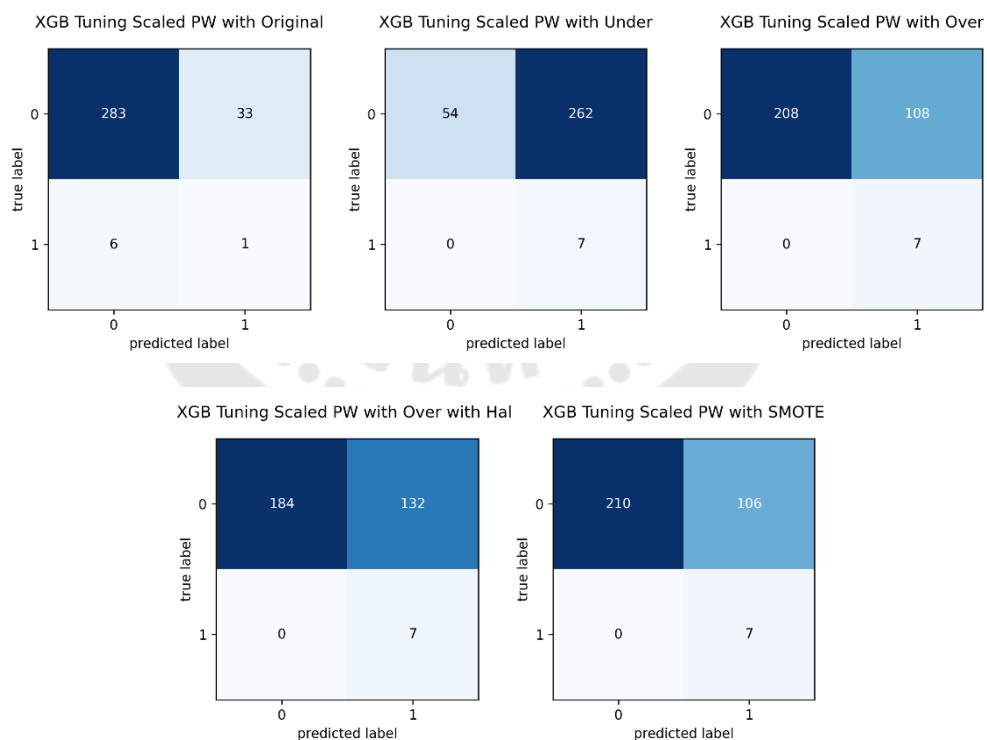


ภาพประกอบ 333 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8793	0.8956	0.1429	0.0294	0.0488
Under	0.1889	0.1709	1.0000	0.0260	0.0507
Over	0.6656	0.6582	1.0000	0.0609	0.1148
Over with Half	0.5913	0.5823	1.0000	0.0504	0.0959
SMOTE	0.6718	0.6646	1.0000	0.0619	0.1167

ภาพประกอบ 334 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

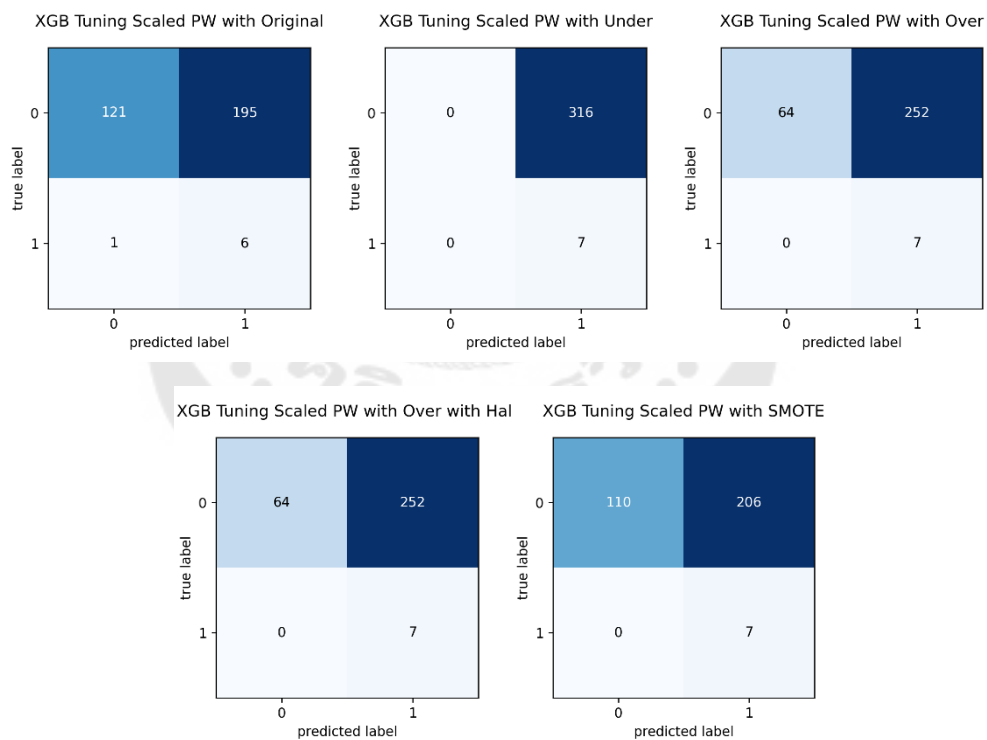


ภาพประกอบ 335 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3932	0.3829	0.8571	0.0299	0.0577
Under	0.0217	0.0000	1.0000	0.0217	0.0424
Over	0.2198	0.2025	1.0000	0.0270	0.0526
Over with Half	0.2198	0.2025	1.0000	0.0270	0.0526
SMOTE	0.3622	0.3481	1.0000	0.0329	0.0636

ภาพประกอบ 336 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



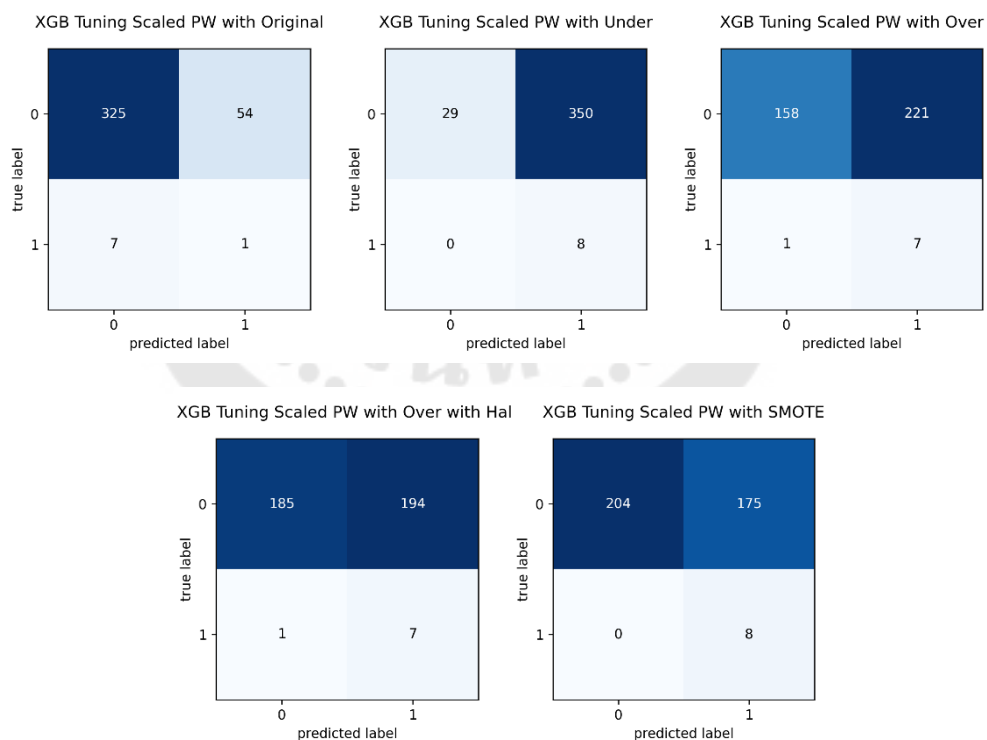
ภาพประกอบ 337 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.18.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8424	0.8575	0.1250	0.0182	0.0317
Under	0.0956	0.0765	1.0000	0.0223	0.0437
Over	0.4264	0.4169	0.8750	0.0307	0.0593
Over with Half	0.4961	0.4881	0.8750	0.0348	0.0670
SMOTE	0.5478	0.5383	1.0000	0.0437	0.0838

ภาพประกอบ 338 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

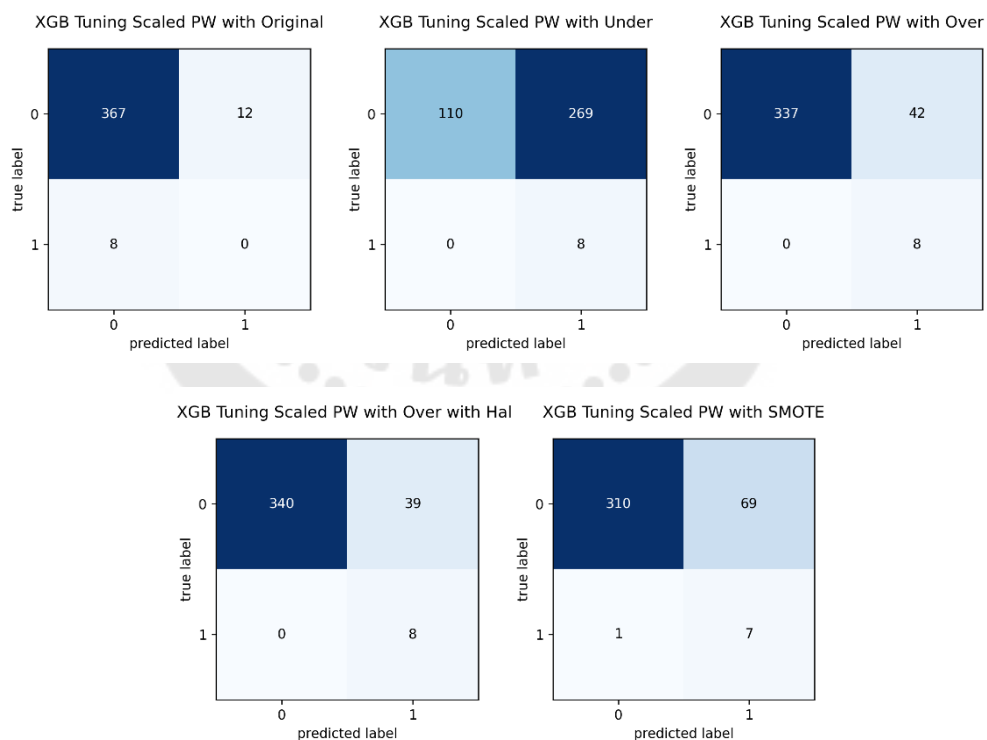


ภาพประกอบ 339 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9483	0.9683	0.0000	0.0000	0.0000
Under	0.3049	0.2902	1.0000	0.0289	0.0561
Over	0.8915	0.8892	1.0000	0.1600	0.2759
Over with Half	0.8992	0.8971	1.0000	0.1702	0.2909
SMOTE	0.8191	0.8179	0.8750	0.0921	0.1667

ภาพประกอบ 340 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

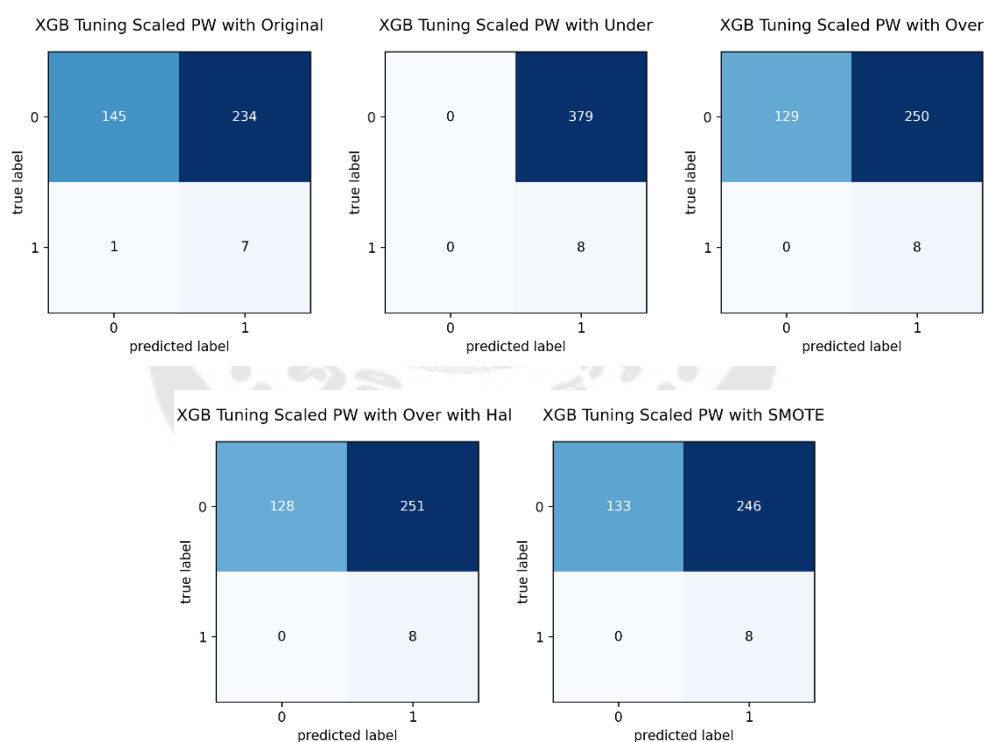


ภาพประกอบ 341 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3928	0.3826	0.8750	0.0290	0.0562
Under	0.0207	0.0000	1.0000	0.0207	0.0405
Over	0.3540	0.3404	1.0000	0.0310	0.0602
Over with Half	0.3514	0.3377	1.0000	0.0309	0.0599
SMOTE	0.3643	0.3509	1.0000	0.0315	0.0611

ภาพประกอบ 342 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



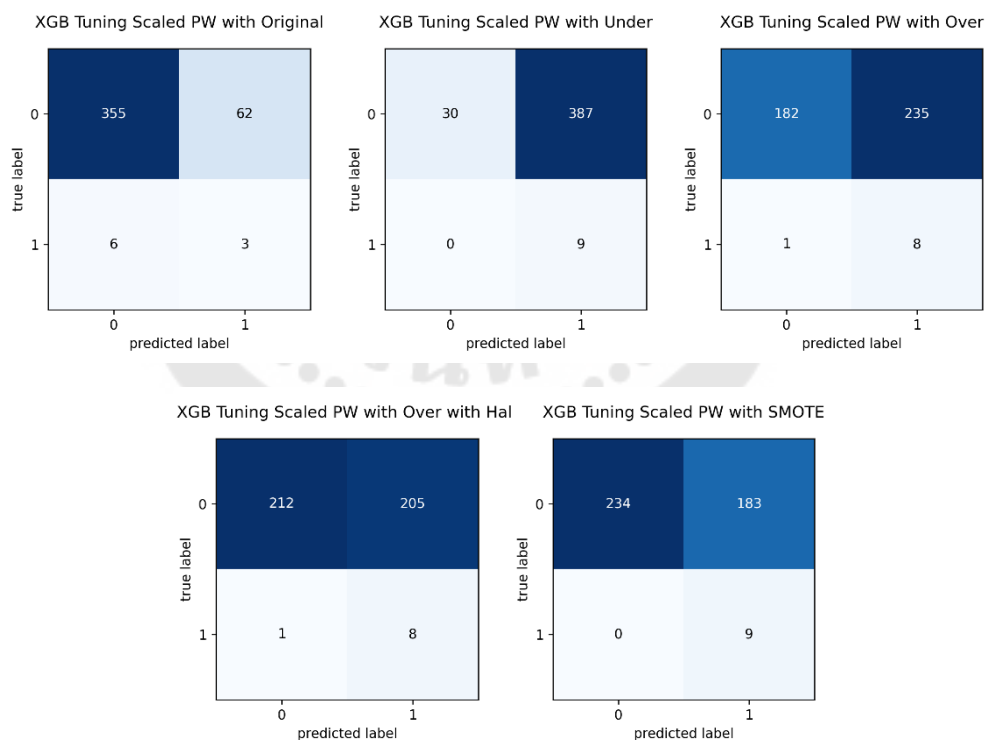
ภาพประกอบ 343 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.18.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8404	0.8513	0.3333	0.0462	0.0811
Under	0.0915	0.0719	1.0000	0.0227	0.0444
Over	0.4460	0.4365	0.8889	0.0329	0.0635
Over with Half	0.5164	0.5084	0.8889	0.0376	0.0721
SMOTE	0.5704	0.5612	1.0000	0.0469	0.0896

ภาพประกอบ 344 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

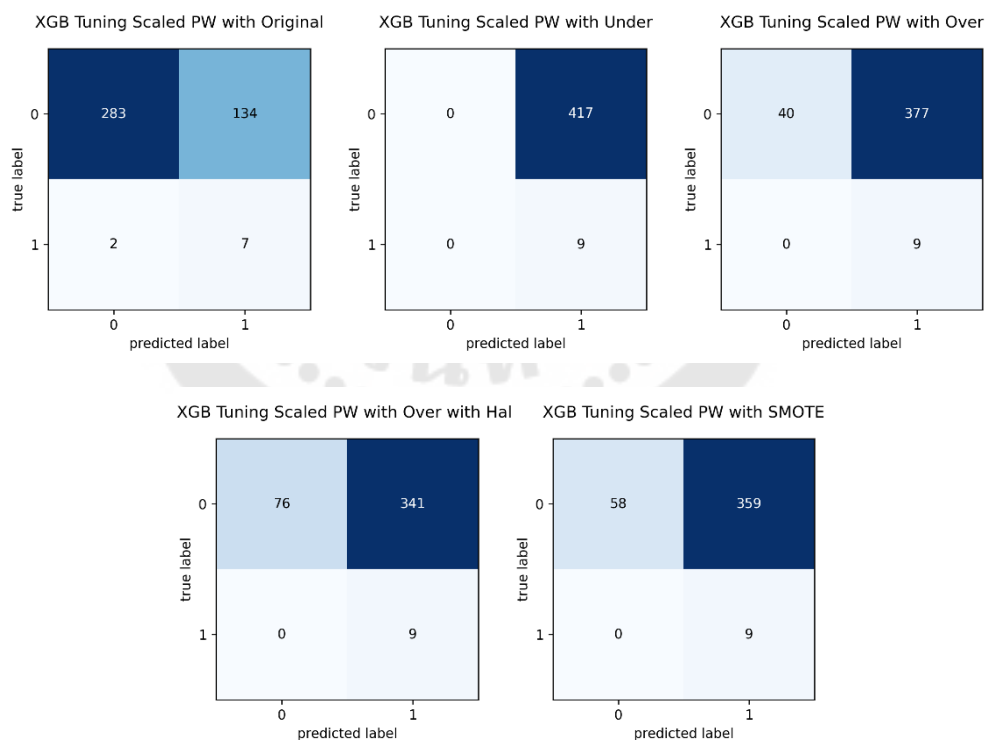


ภาพประกอบ 345 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6808	0.6787	0.7778	0.0496	0.0933
Under	0.0211	0.0000	1.0000	0.0211	0.0414
Over	0.1150	0.0959	1.0000	0.0233	0.0456
Over with Half	0.1995	0.1823	1.0000	0.0257	0.0501
SMOTE	0.1573	0.1391	1.0000	0.0245	0.0477

ภาพประกอบ 346 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

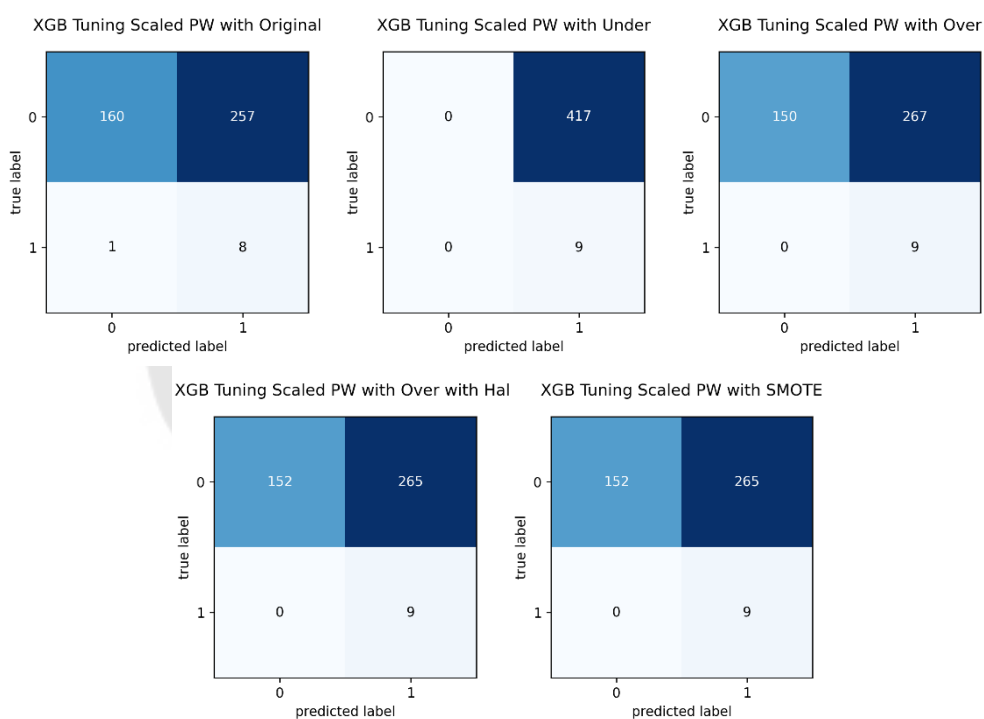


ภาพประกอบ 347 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.18.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3944	0.3837	0.8889	0.0302	0.0584
Under	0.0211	0.0000	1.0000	0.0211	0.0414
Over	0.3732	0.3597	1.0000	0.0326	0.0632
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.3779	0.3645	1.0000	0.0328	0.0636

ภาพประกอบ 348 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ scaled_pos_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 349 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

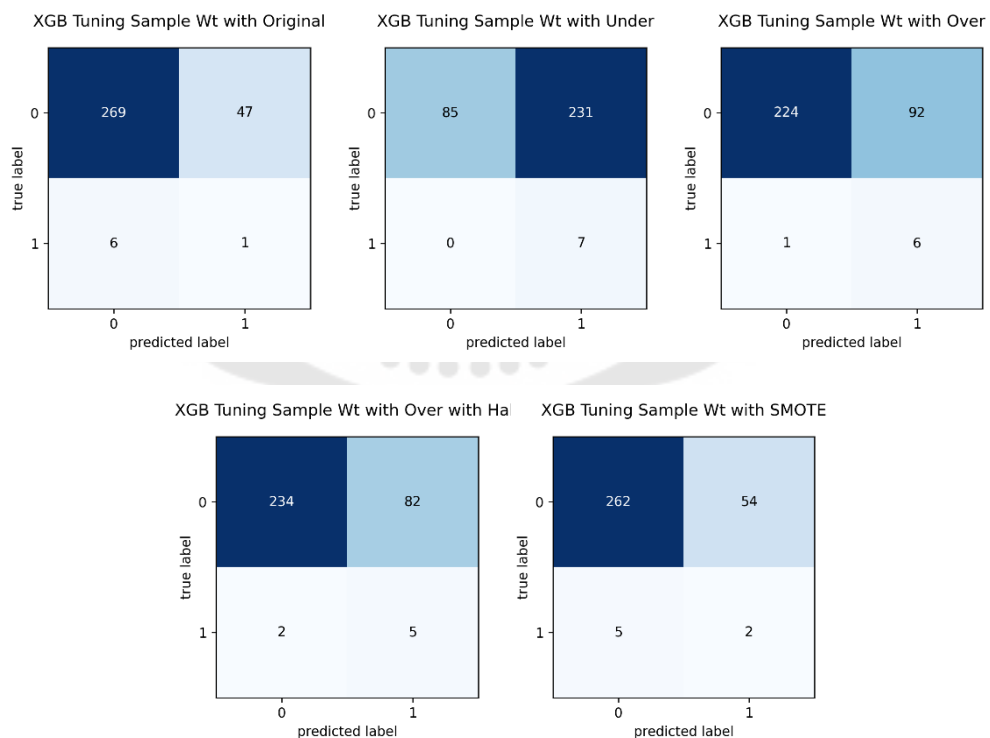
4.19 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ sample_weight

4.19.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.19.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8359	0.8513	0.1429	0.0208	0.0364
Under	0.2848	0.2690	1.0000	0.0294	0.0571
Over	0.7121	0.7089	0.8571	0.0612	0.1143
Over with Half	0.7399	0.7405	0.7143	0.0575	0.1064
SMOTE	0.8173	0.8291	0.2857	0.0357	0.0635

ภาพประกอบ 350 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Threshold

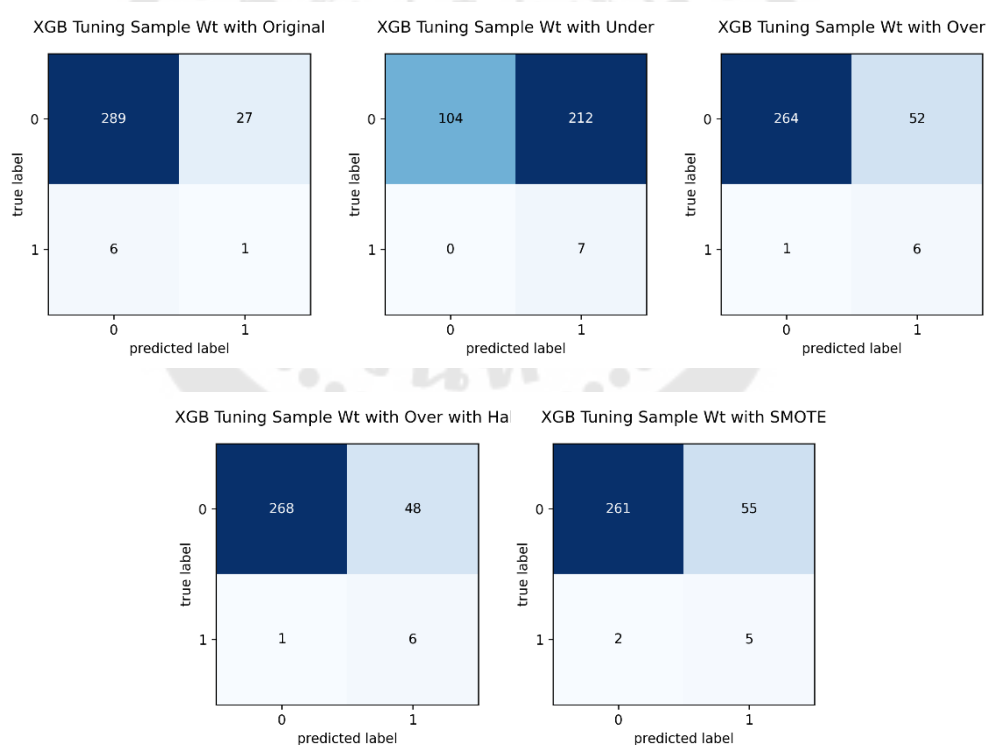


ภาพประกอบ 351 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8978	0.9146	0.1429	0.0357	0.0571
Under	0.3437	0.3291	1.0000	0.0320	0.0619
Over	0.8359	0.8354	0.8571	0.1034	0.1846
Over with Half	0.8483	0.8481	0.8571	0.1111	0.1967
SMOTE	0.8235	0.8259	0.7143	0.0833	0.1493

ภาพประกอบ 352 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

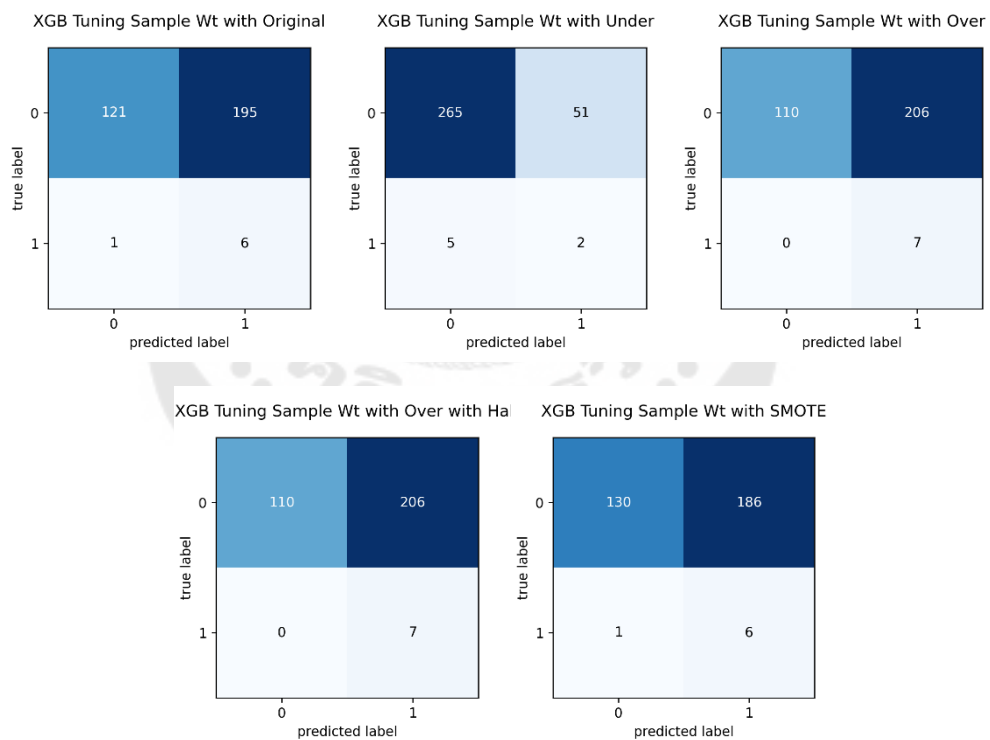


ภาพประกอบ 353 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3932	0.3829	0.8571	0.0299	0.0577
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3622	0.3481	1.0000	0.0329	0.0636
Over with Half	0.3622	0.3481	1.0000	0.0329	0.0636
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 354 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Decision Tree



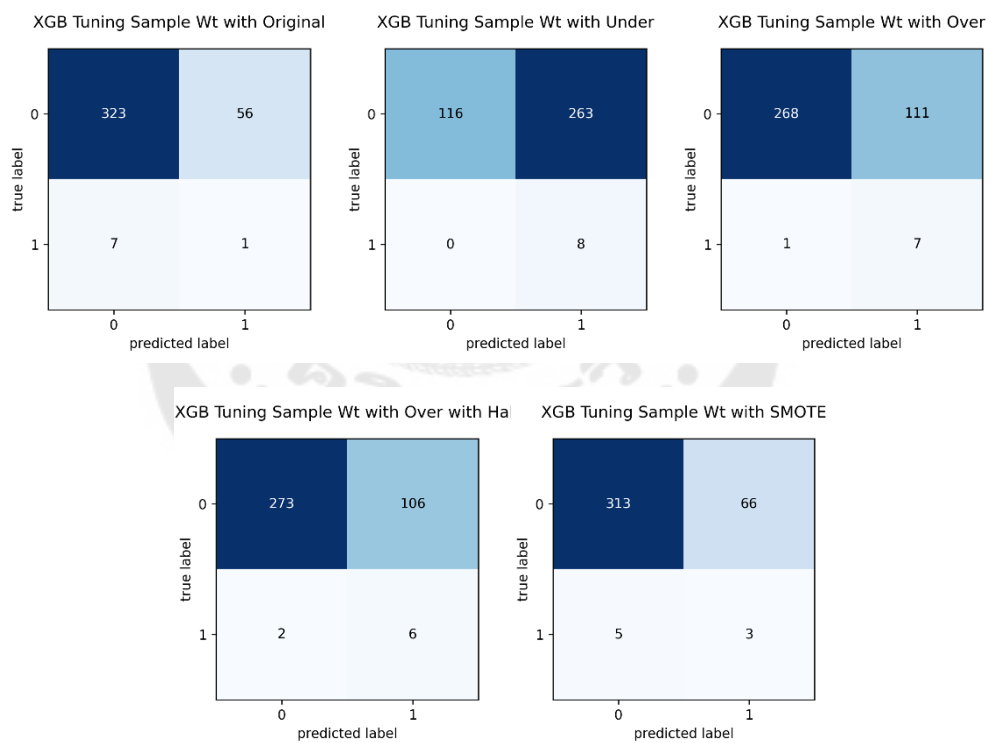
ภาพประกอบ 355 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.19.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8372	0.8522	0.1250	0.0175	0.0308
Under	0.3204	0.3061	1.0000	0.0295	0.0573
Over	0.7106	0.7071	0.8750	0.0593	0.1111
Over with Half	0.7209	0.7203	0.7500	0.0536	0.1000
SMOTE	0.8165	0.8259	0.3750	0.0435	0.0779

ภาพประกอบ 356 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

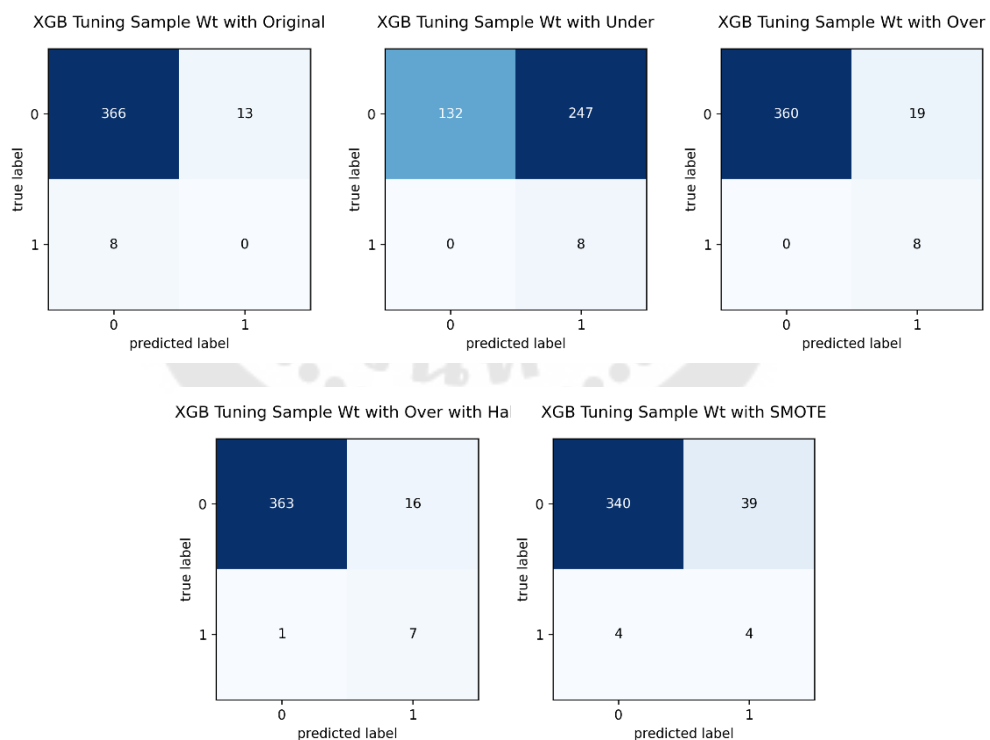


ภาพประกอบ 357 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.9457	0.9657	0.0000	0.0000	0.0000
Under	0.3618	0.3483	1.0000	0.0314	0.0608
Over	0.9509	0.9499	1.0000	0.2963	0.4571
Over with Half	0.9561	0.9578	0.8750	0.3043	0.4516
SMOTE	0.8889	0.8971	0.5000	0.0930	0.1569

ภาพประกอบ 358 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

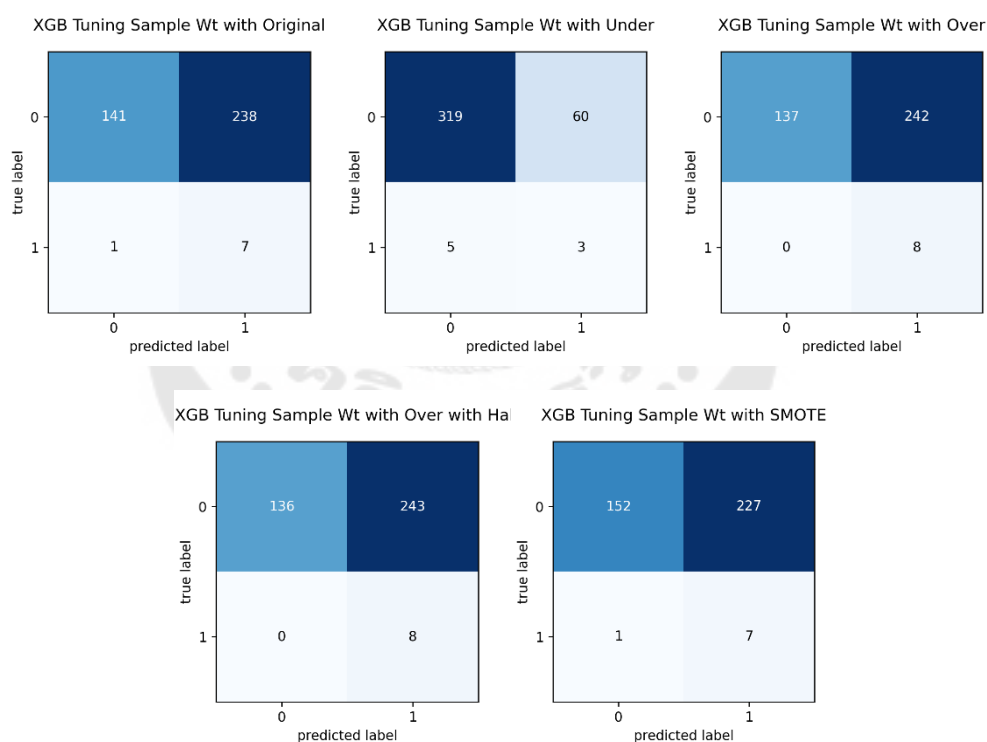


ภาพประกอบ 359 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3824	0.3720	0.8750	0.0286	0.0553
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3747	0.3615	1.0000	0.0320	0.0620
Over with Half	0.3721	0.3588	1.0000	0.0319	0.0618
SMOTE	0.4109	0.4011	0.8750	0.0299	0.0579

ภาพประกอบ 360 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Decision Tree



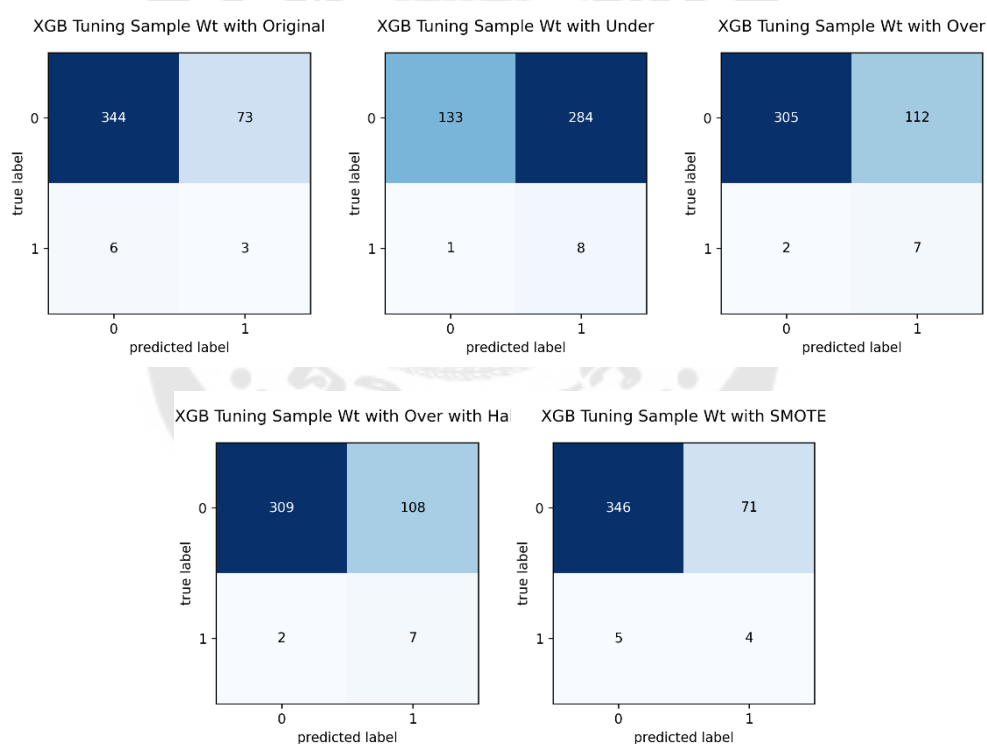
ภาพประกอบ 361 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.19.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8146	0.8249	0.3333	0.0395	0.0706
Under	0.3310	0.3189	0.8889	0.0274	0.0532
Over	0.7324	0.7314	0.7778	0.0588	0.1094
Over with Half	0.7418	0.7410	0.7778	0.0609	0.1129
SMOTE	0.8216	0.8297	0.4444	0.0533	0.0952

ภาพประกอบ 362 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

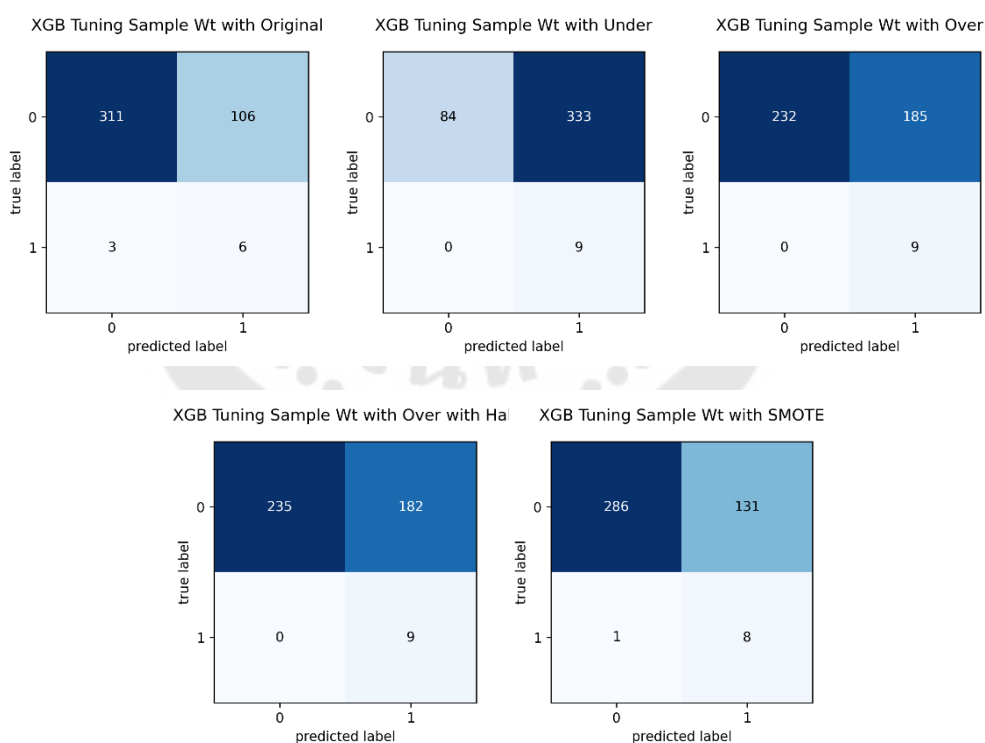


ภาพประกอบ 363 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.7441	0.7458	0.6667	0.0536	0.0992
Under	0.2183	0.2014	1.0000	0.0263	0.0513
Over	0.5657	0.5564	1.0000	0.0464	0.0887
Over with Half	0.5728	0.5635	1.0000	0.0471	0.0900
SMOTE	0.6901	0.6859	0.8889	0.0576	0.1081

ภาพประกอบ 364 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification

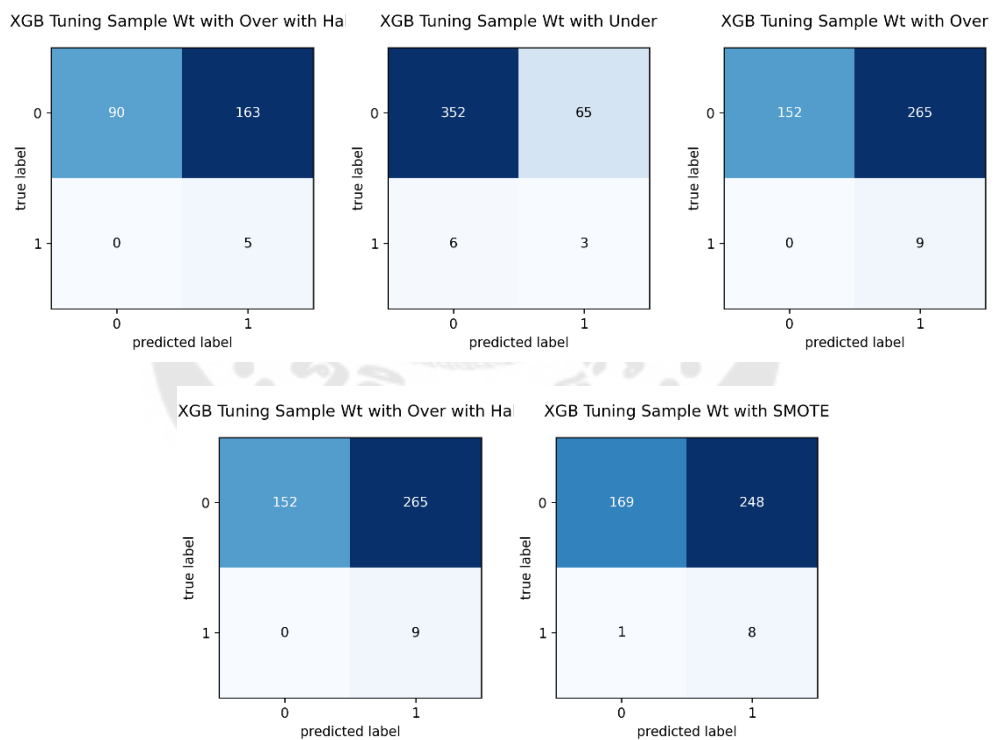


ภาพประกอบ 365 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.19.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.3944	0.3837	0.8889	0.0302	0.0584
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3779	0.3645	1.0000	0.0328	0.0636
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 366 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ sample_weight โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Decision Tree



ภาพประกอบ 367 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20 ผลลัพธ์ที่ได้จากการใช้อัลกอริทึม XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier

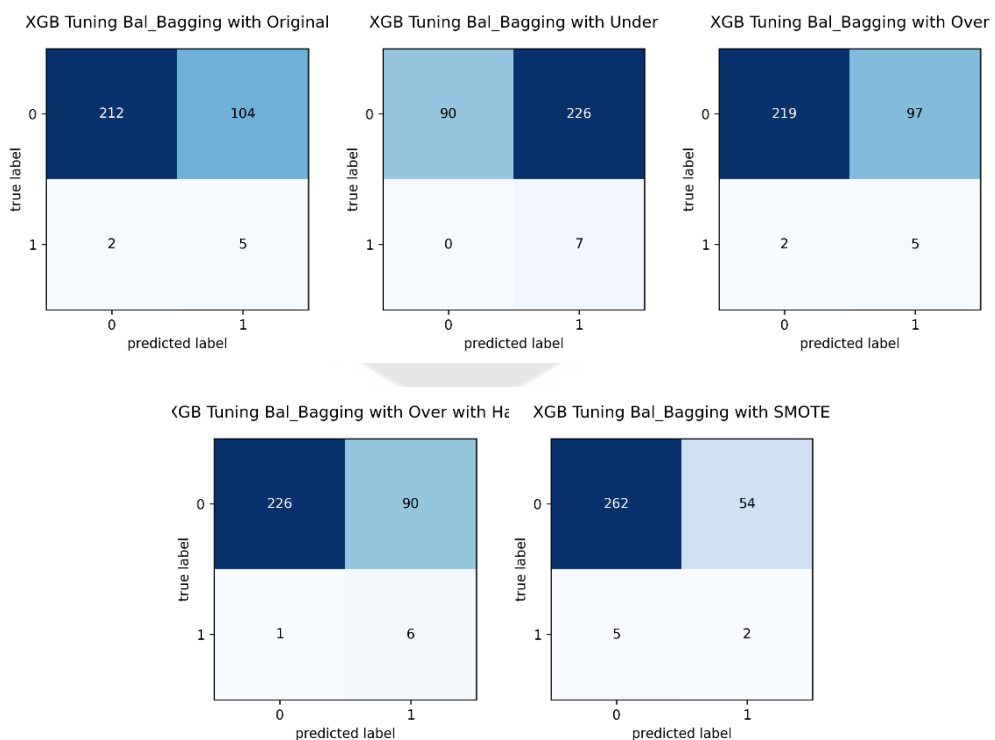
4.20.1 ผลลัพธ์จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 75:25

4.20.1.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6718	0.6709	0.7143	0.0459	0.0862
Under	0.3003	0.2848	1.0000	0.0300	0.0583
Over	0.6935	0.6930	0.7143	0.0490	0.0917
Over with Half	0.7183	0.7152	0.8571	0.0625	0.1165
SMOTE	0.8173	0.8291	0.2857	0.0357	0.0635

ภาพประกอบ 368 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Threshold

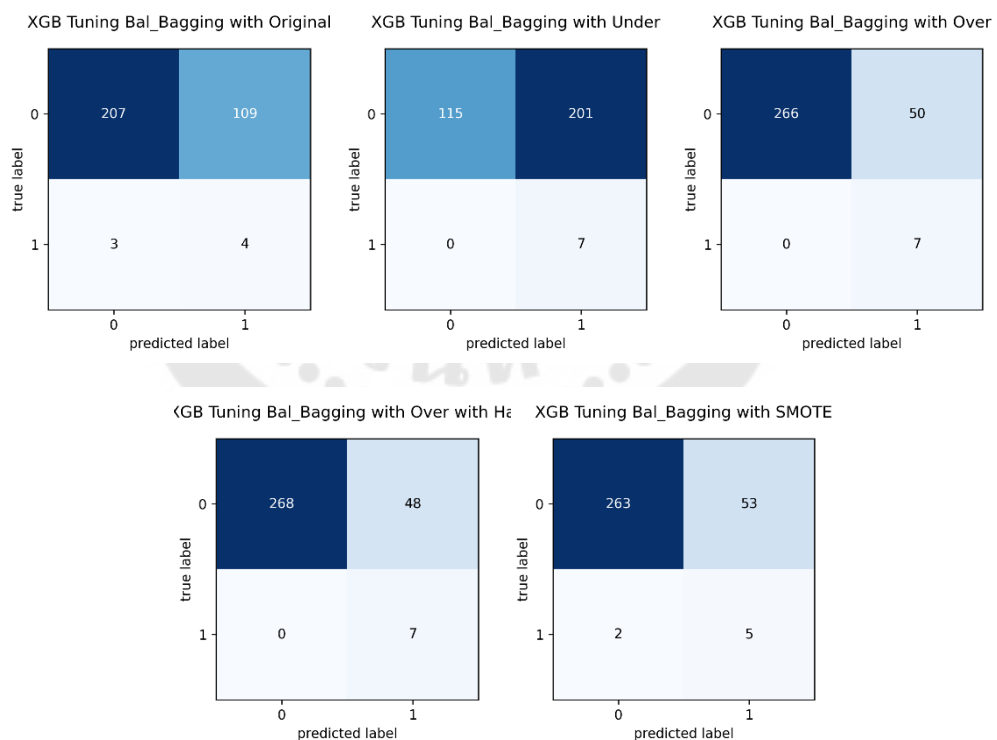


ภาพประกอบ 369 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.1.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6533	0.6551	0.5714	0.0354	0.0667
Under	0.3777	0.3639	1.0000	0.0337	0.0651
Over	0.8452	0.8418	1.0000	0.1228	0.2188
Over with Half	0.8514	0.8481	1.0000	0.1273	0.2258
SMOTE	0.8297	0.8323	0.7143	0.0862	0.1538

ภาพประกอบ 370 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



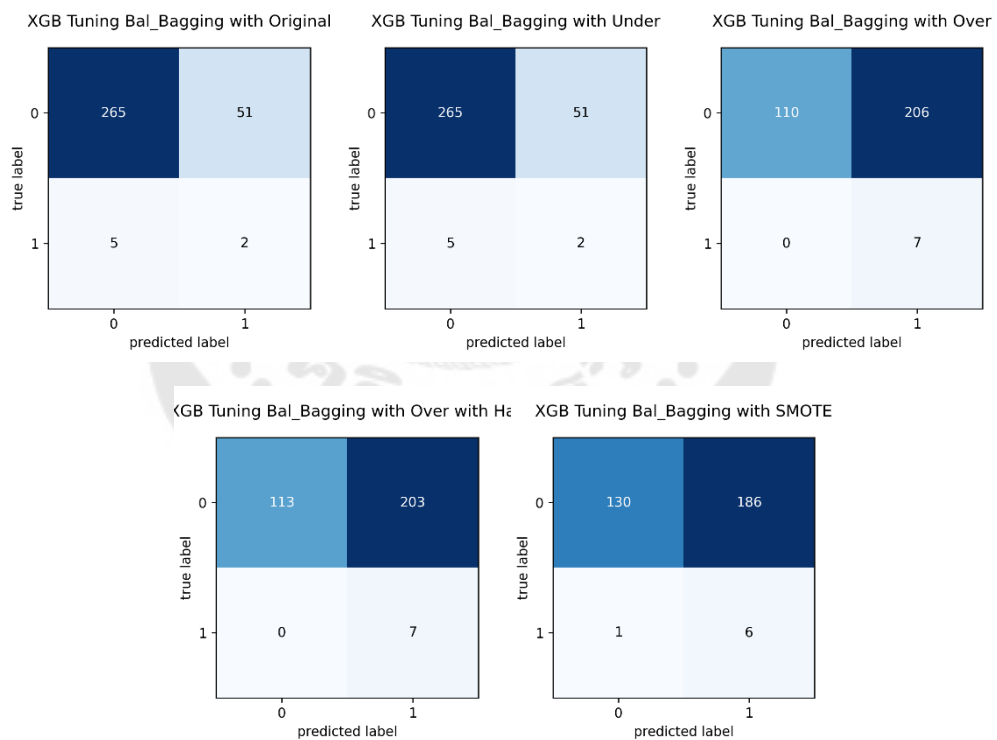
ภาพประกอบ 371 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.1.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8266	0.8386	0.2857	0.0377	0.0667
Under	0.8266	0.8386	0.2857	0.0377	0.0667
Over	0.3622	0.3481	1.0000	0.0329	0.0636
Over with Half	0.3715	0.3576	1.0000	0.0333	0.0645
SMOTE	0.4211	0.4114	0.8571	0.0312	0.0603

ภาพประกอบ 372 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 75:25 ใช้ Feature Selection ด้วยวิธี

Decision Tree



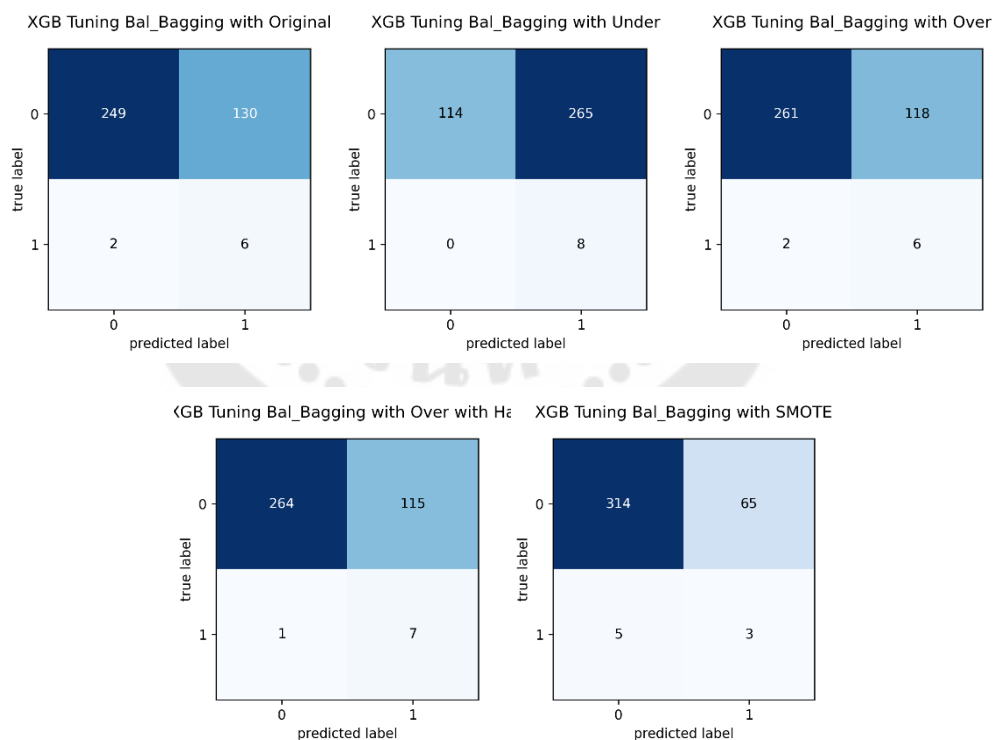
ภาพประกอบ 373 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.2 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 70:30

4.20.2.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6589	0.6570	0.7500	0.0441	0.0833
Under	0.3152	0.3008	1.0000	0.0293	0.0569
Over	0.6899	0.6887	0.7500	0.0484	0.0909
Over with Half	0.7003	0.6966	0.8750	0.0574	0.1077
SMOTE	0.8191	0.8285	0.3750	0.0441	0.0789

ภาพประกอบ 374 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Threshold

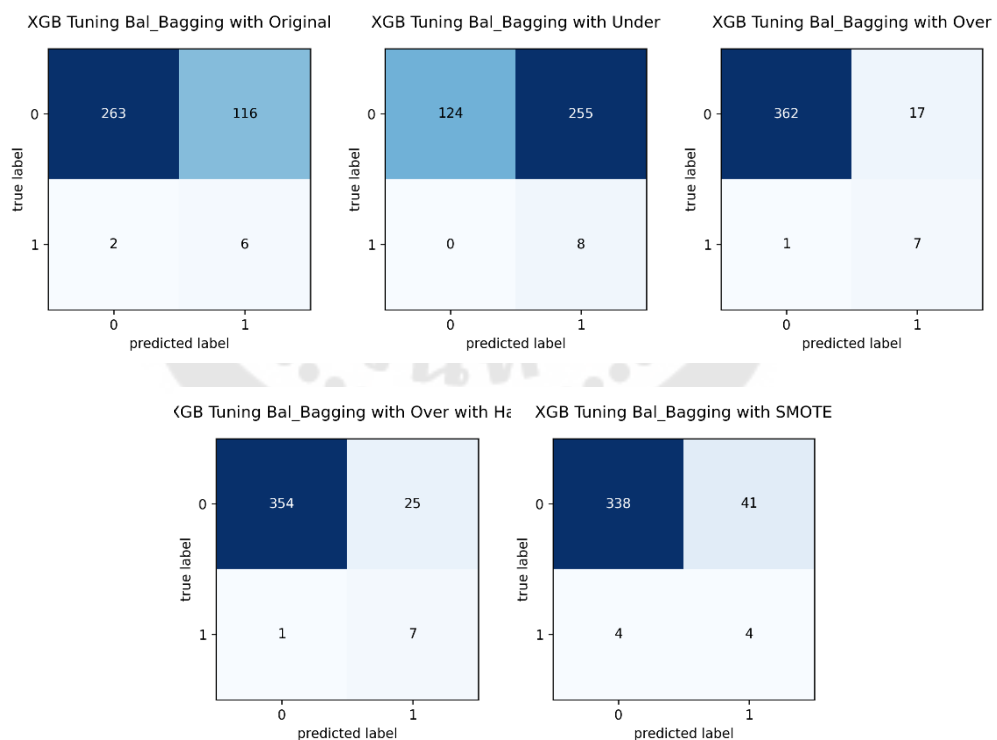


ภาพประกอบ 375 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.2.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.6951	0.6939	0.7500	0.0492	0.0923
Under	0.3411	0.3272	1.0000	0.0304	0.0590
Over	0.9535	0.9551	0.8750	0.2917	0.4375
Over with Half	0.9328	0.9340	0.8750	0.2188	0.3500
SMOTE	0.8837	0.8918	0.5000	0.0889	0.1509

ภาพประกอบ 376 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



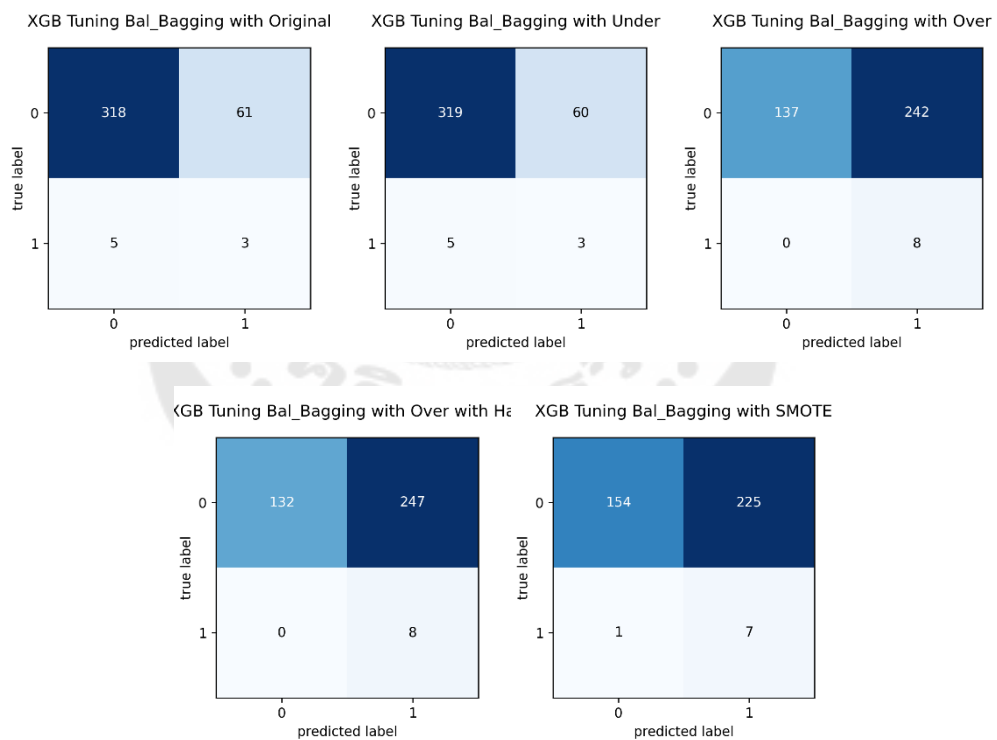
ภาพประกอบ 377 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.2.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8295	0.8391	0.3750	0.0469	0.0833
Under	0.8320	0.8417	0.3750	0.0476	0.0845
Over	0.3747	0.3615	1.0000	0.0320	0.0620
Over with Half	0.3618	0.3483	1.0000	0.0314	0.0608
SMOTE	0.4160	0.4063	0.8750	0.0302	0.0583

ภาพประกอบ 378 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 70:30 ใช้ Feature Selection ด้วยวิธี

Decision Tree



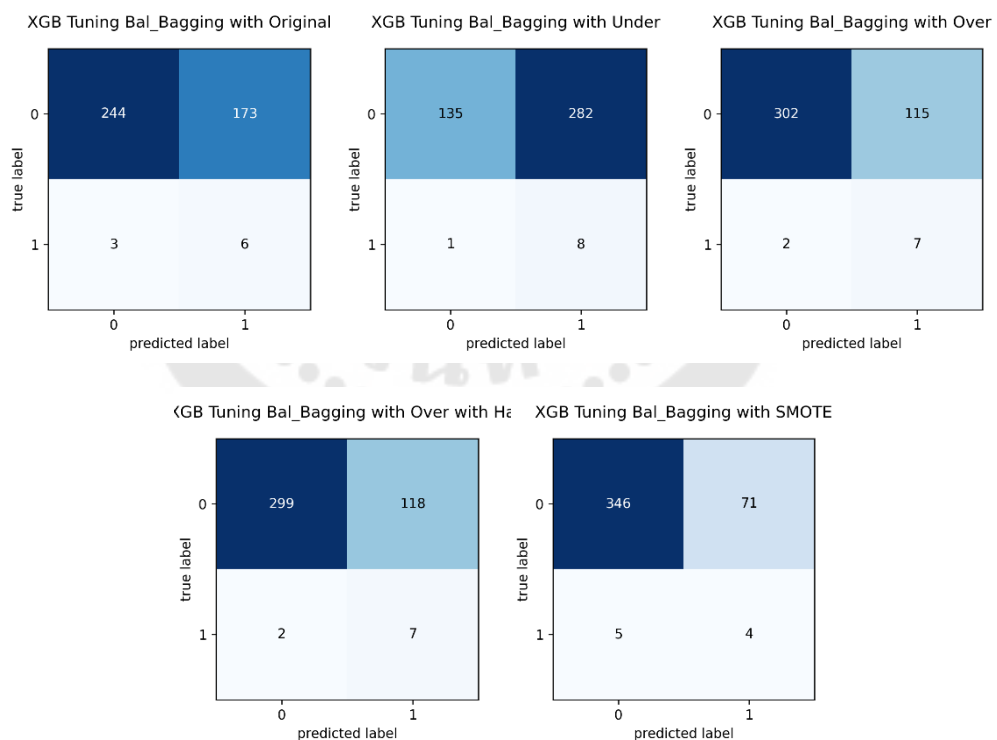
ภาพประกอบ 379 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.3 ผลลัพธ์ที่ได้จากการแบ่งอัตราส่วนของข้อมูลเท่ากับ 67:33

4.20.3.1 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Threshold

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.5869	0.5851	0.6667	0.0335	0.0638
Under	0.3357	0.3237	0.8889	0.0276	0.0535
Over	0.7254	0.7242	0.7778	0.0574	0.1069
Over with Half	0.7183	0.7170	0.7778	0.0560	0.1045
SMOTE	0.8216	0.8297	0.4444	0.0533	0.0952

ภาพประกอบ 380 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Threshold

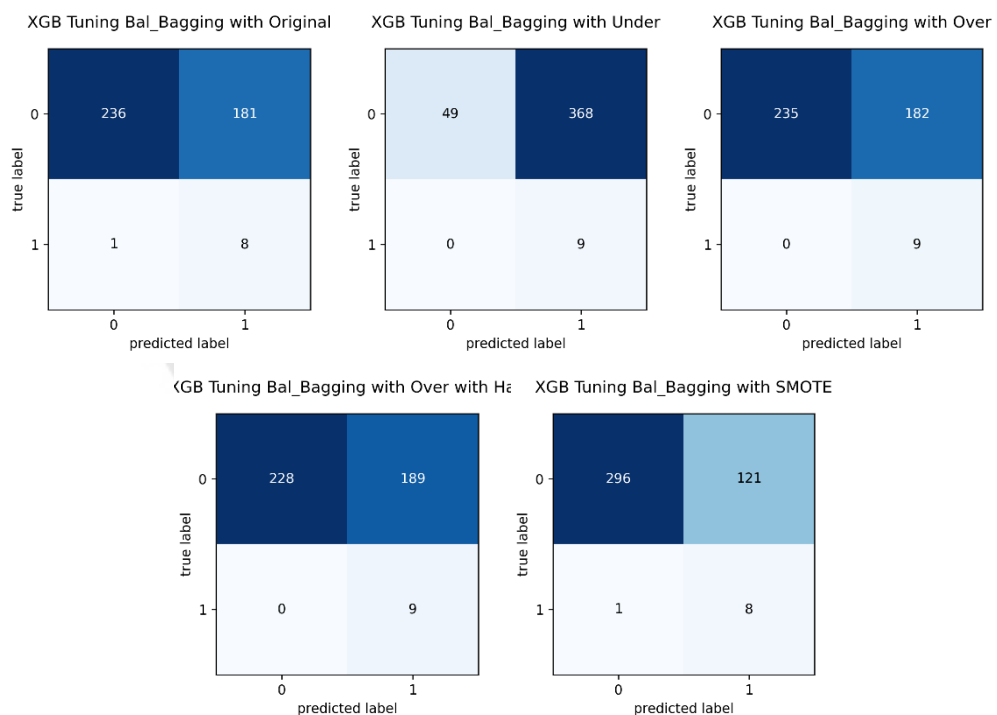


ภาพประกอบ 381 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.3.2 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Mutual Information Classification

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.5728	0.5659	0.8889	0.0423	0.0808
Under	0.1362	0.1175	1.0000	0.0239	0.0466
Over	0.5728	0.5635	1.0000	0.0471	0.0900
Over with Half	0.5563	0.5468	1.0000	0.0455	0.0870
SMOTE	0.7136	0.7098	0.8889	0.0620	0.1159

ภาพประกอบ 382 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี Mutual Information Classification



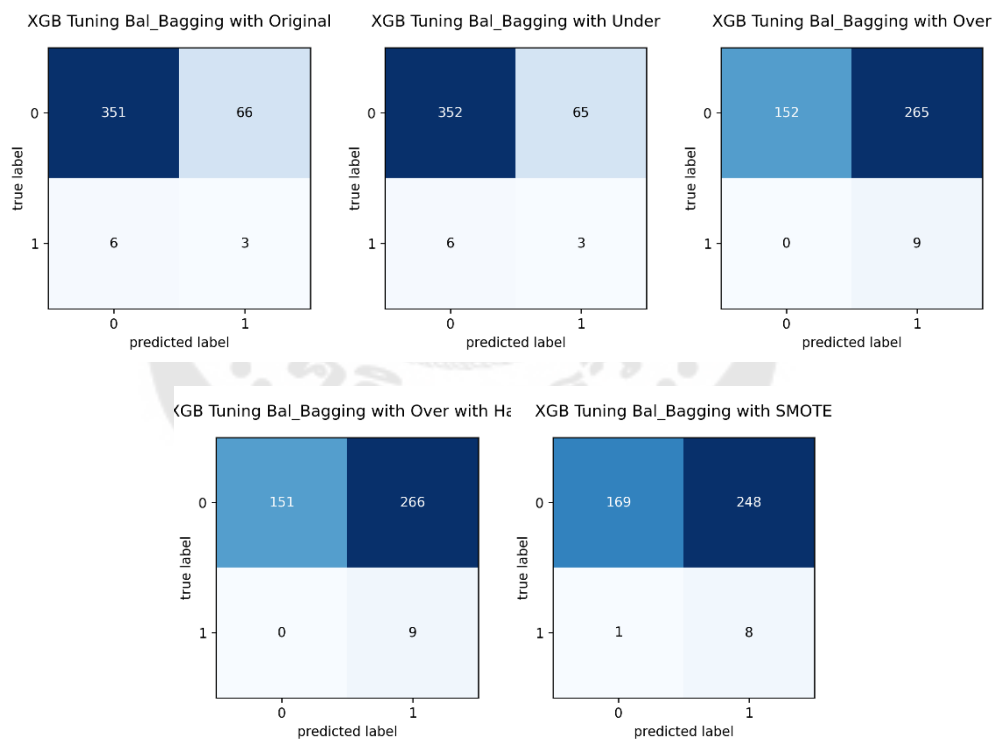
ภาพประกอบ 383 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

4.20.3.3 ผลลัพธ์ที่ได้จากการใช้ Feature Selection โดยวิธี Decision Tree

	Accuracy	Specificity	Sensitivity	Precision	F1 Score
Original	0.8310	0.8417	0.3333	0.0435	0.0769
Under	0.8333	0.8441	0.3333	0.0441	0.0779
Over	0.3779	0.3645	1.0000	0.0328	0.0636
Over with Half	0.3756	0.3621	1.0000	0.0327	0.0634
SMOTE	0.4155	0.4053	0.8889	0.0312	0.0604

ภาพประกอบ 384 ผลลัพธ์ที่ได้จากการใช้ XGBoost กับ Tuning Hyperparameters และ Balanced Bagging Classifier โดยแบ่งข้อมูลในอัตราส่วน 67:33 ใช้ Feature Selection ด้วยวิธี

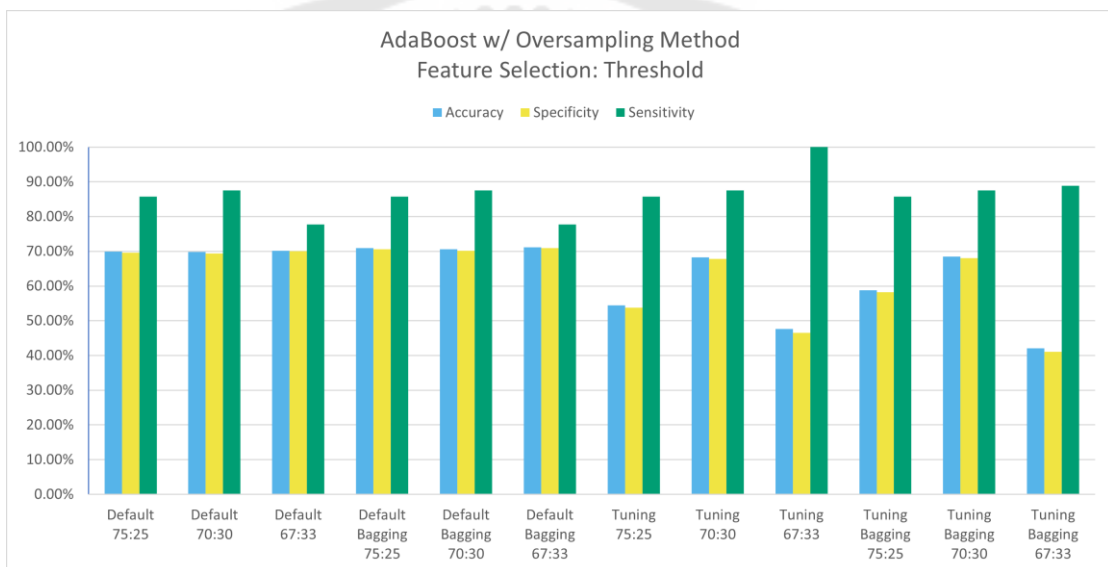
Decision Tree



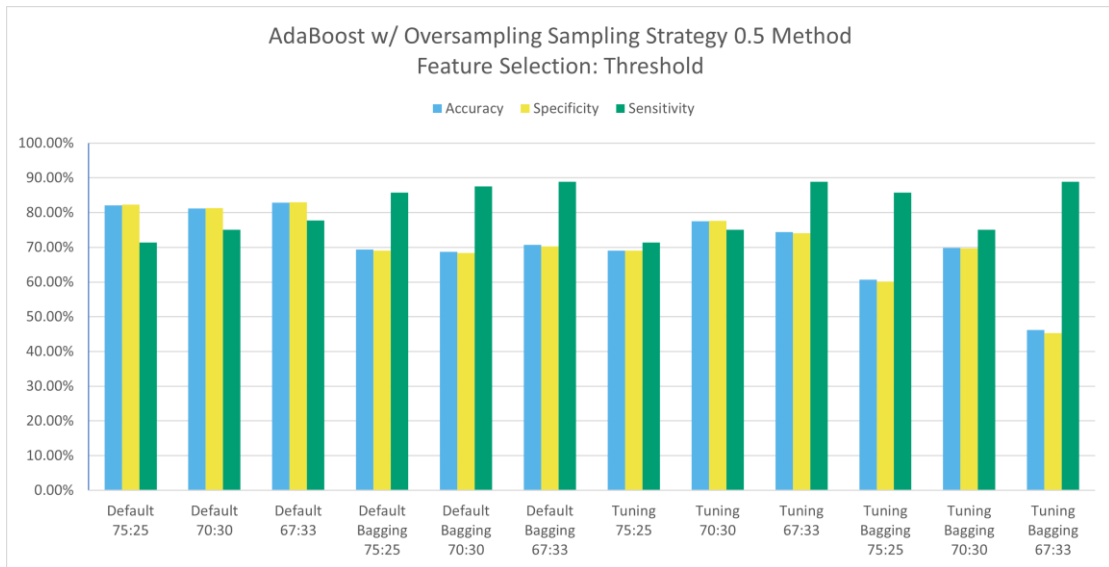
ภาพประกอบ 385 ผลลัพธ์ที่ได้จาก Confusion Matrix กับวิธีการจัดการกับปัญหาความไม่สมดุลของข้อมูลทั้งหมด

จากผลการทดลองของภาพประกอบ 26 - 385 แสดงให้เห็นว่า แบบจำลองที่ใช้วิธีการแก้ไขปัญหาความไม่สมดุลของข้อมูล (Resampling Data Method) ด้วยวิธี Oversampling และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 นั้น สามารถให้ผลลัพธ์ของ Accuracy, Sensitivity และ Specificity ที่ดีกว่า Resampling Data Method วิธีอื่นๆ

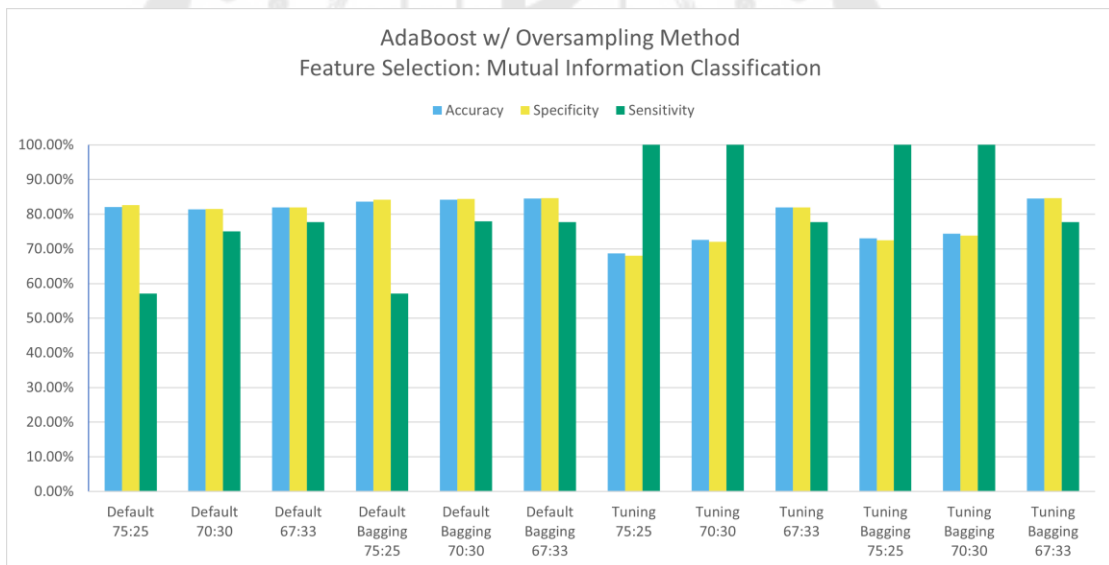
ผู้วิจัยจึงได้นำผลการทดลองของแบบจำลอง ที่ใช้วิธีการแก้ไขปัญหาความไม่สมดุลของข้อมูล (Resampling Data Method) ด้วยวิธี Oversampling และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 ทั้งหมด มานำเสนอเป็นแผนภูมิแท่งเพื่อใช้เปรียบเทียบผลลัพธ์ของแต่ละแบบจำลอง เพื่อหาแบบจำลองที่ให้ผลลัพธ์ที่ดีที่สุด ดังภาพประกอบ 386 - 415



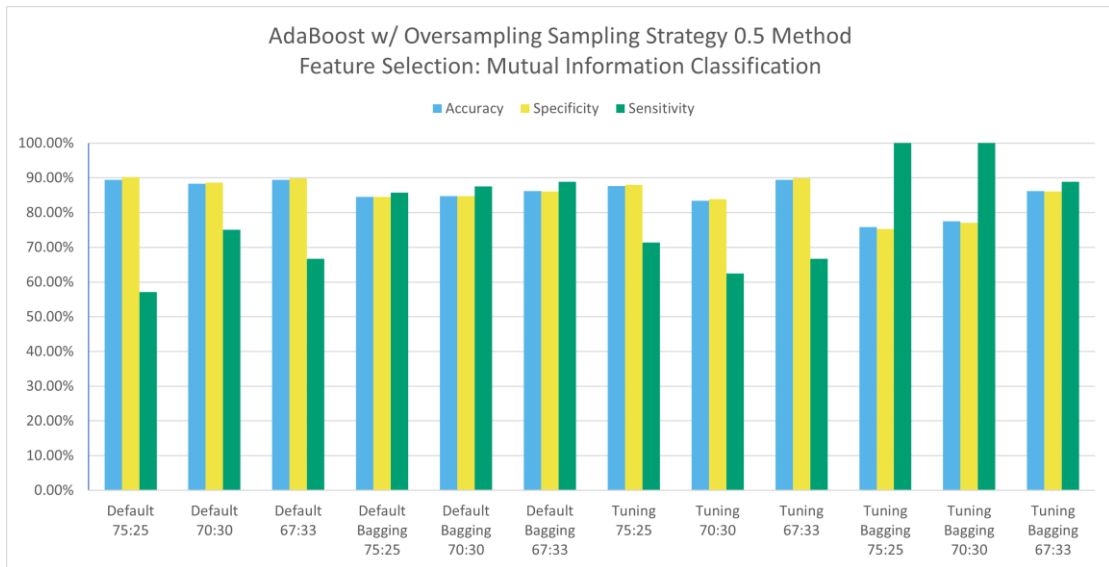
ภาพประกอบ 386 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling



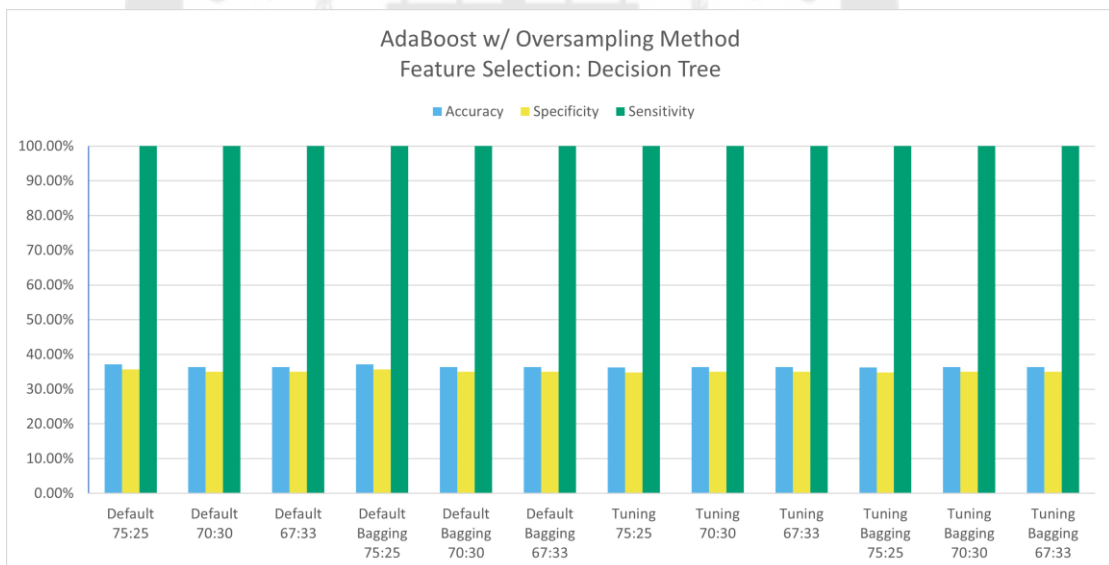
ภาพประกอบ 387 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.5



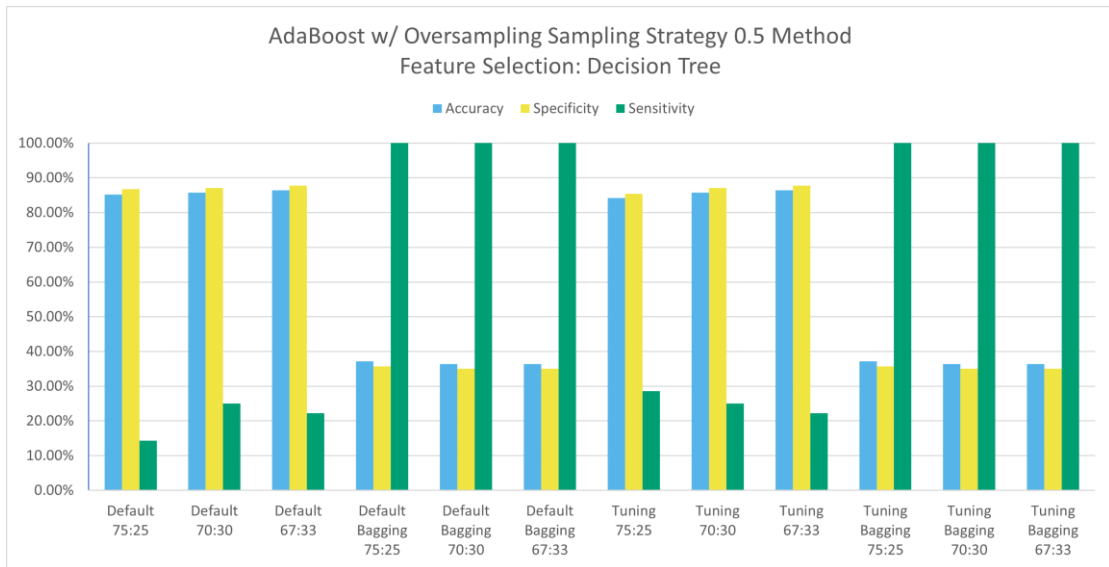
ภาพประกอบ 388 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling



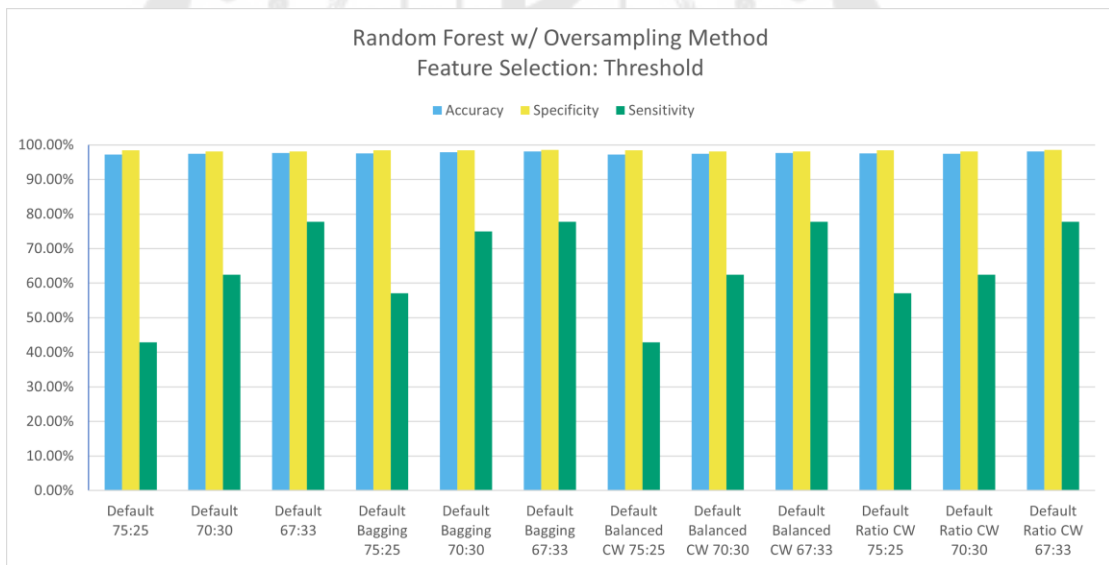
ภาพประกอบ 389 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



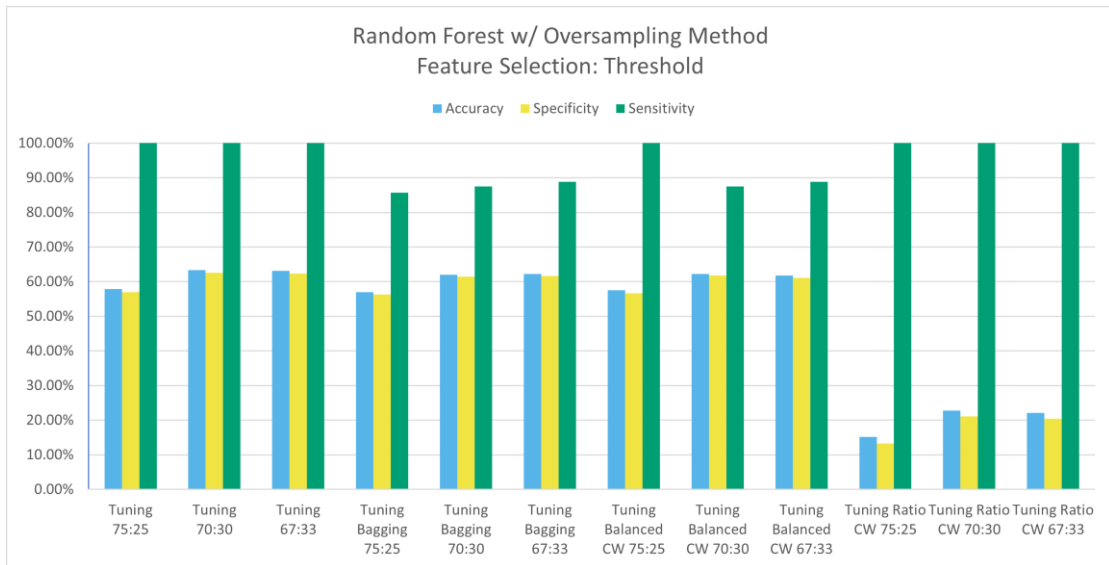
ภาพประกอบ 390 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling



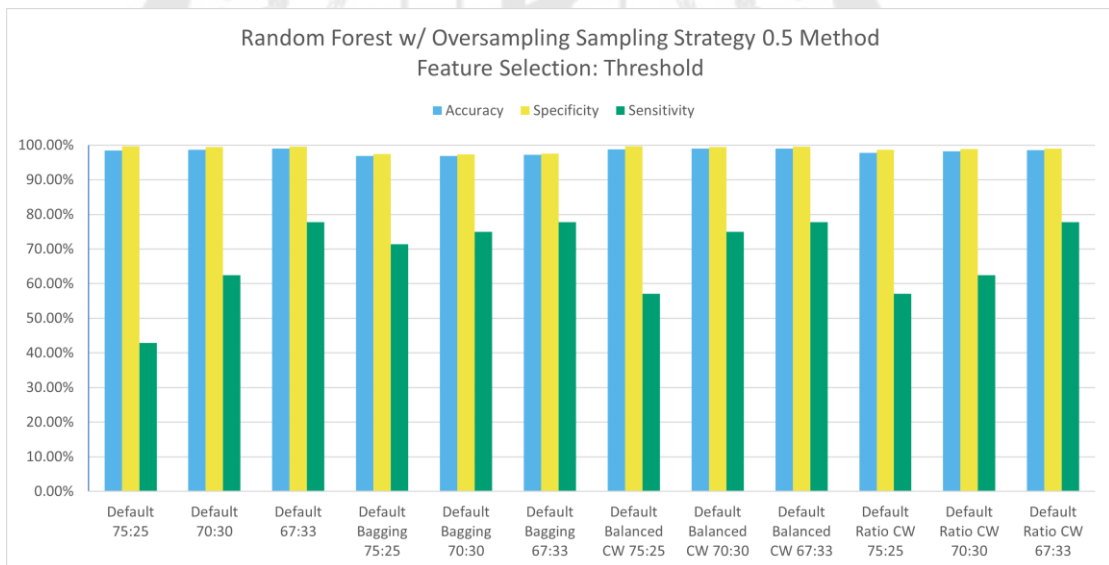
ภาพประกอบ 391 แผนภูมิแท่งของแบบจำลอง AdaBoost ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



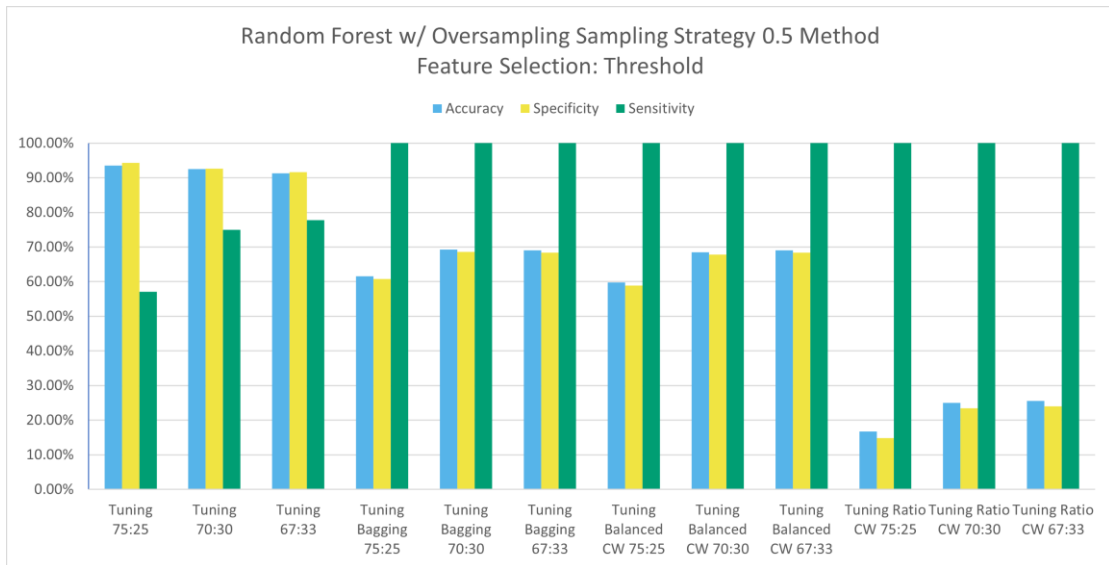
ภาพประกอบ 392 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling



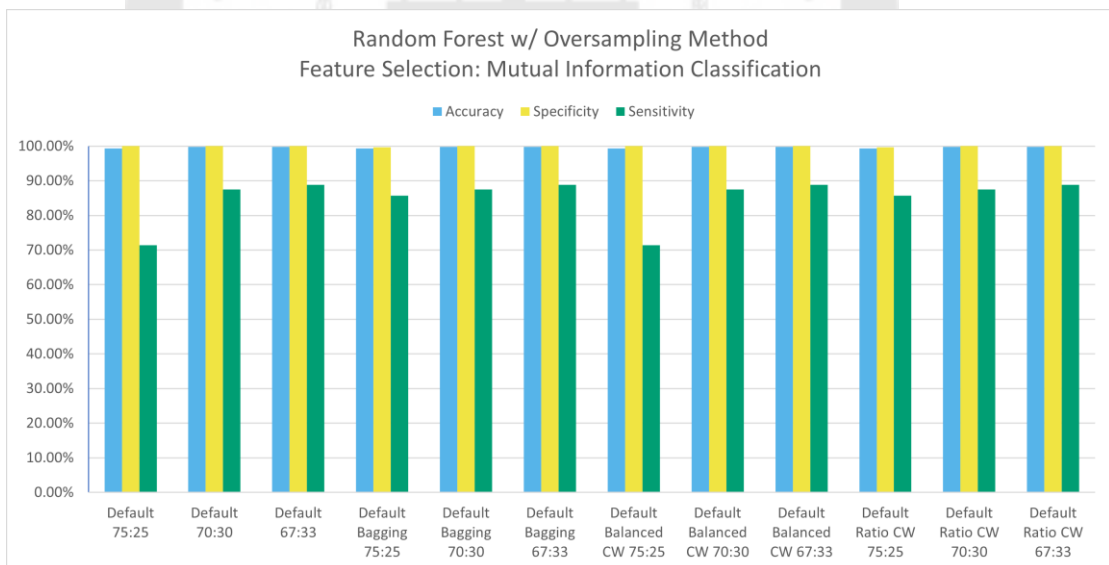
ภาพประกอบ 393 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling



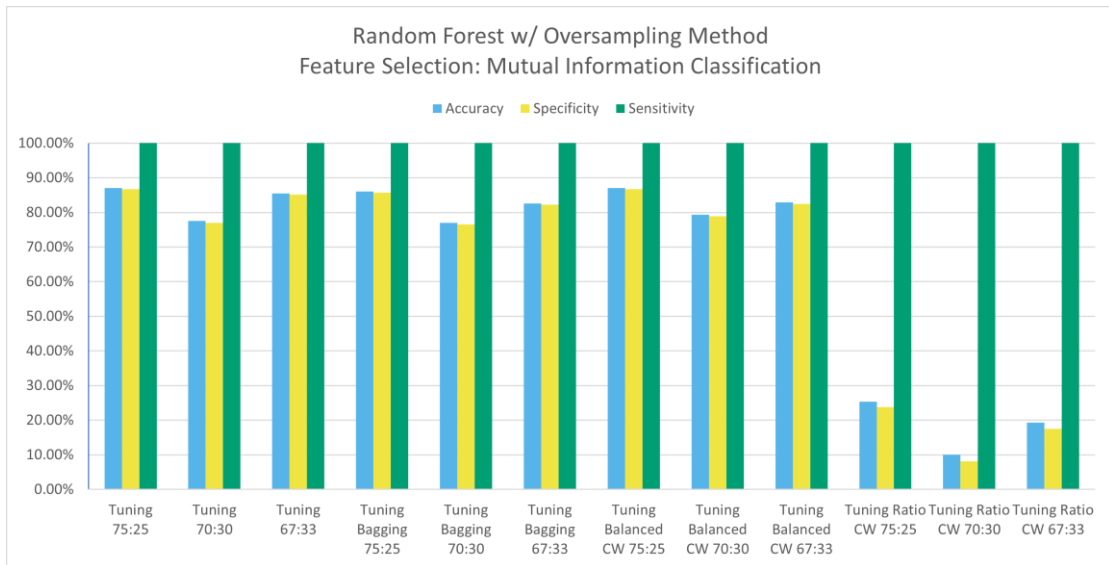
ภาพประกอบ 394 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



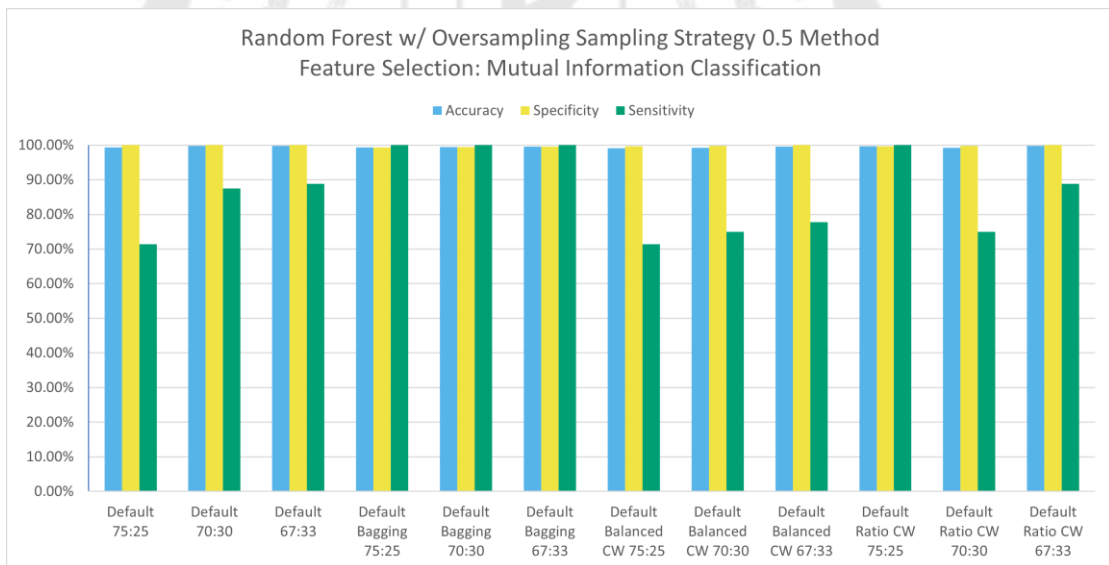
ภาพประกอบ 395 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



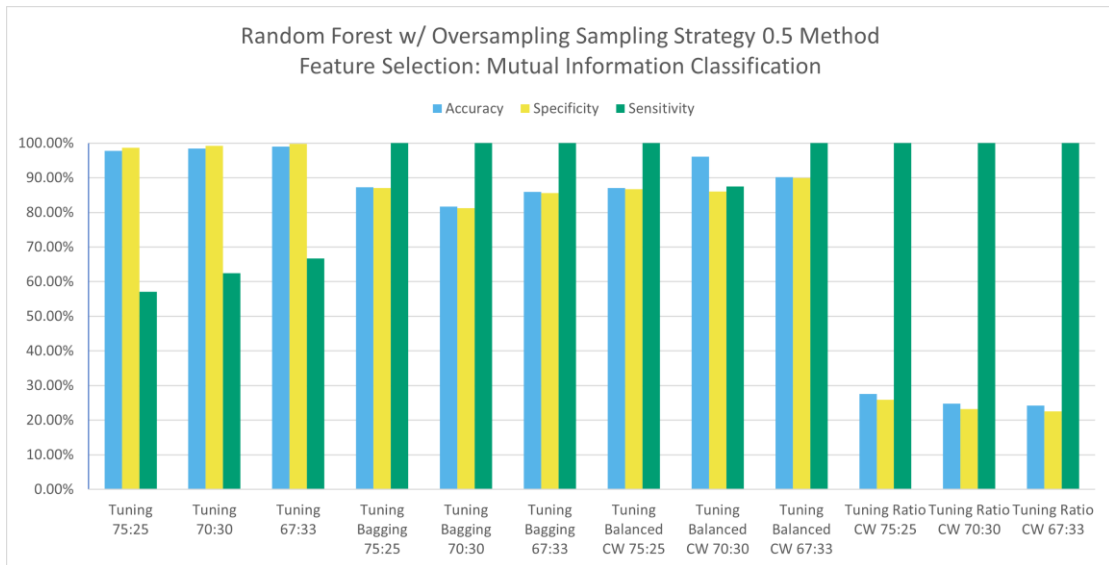
ภาพประกอบ 396 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling



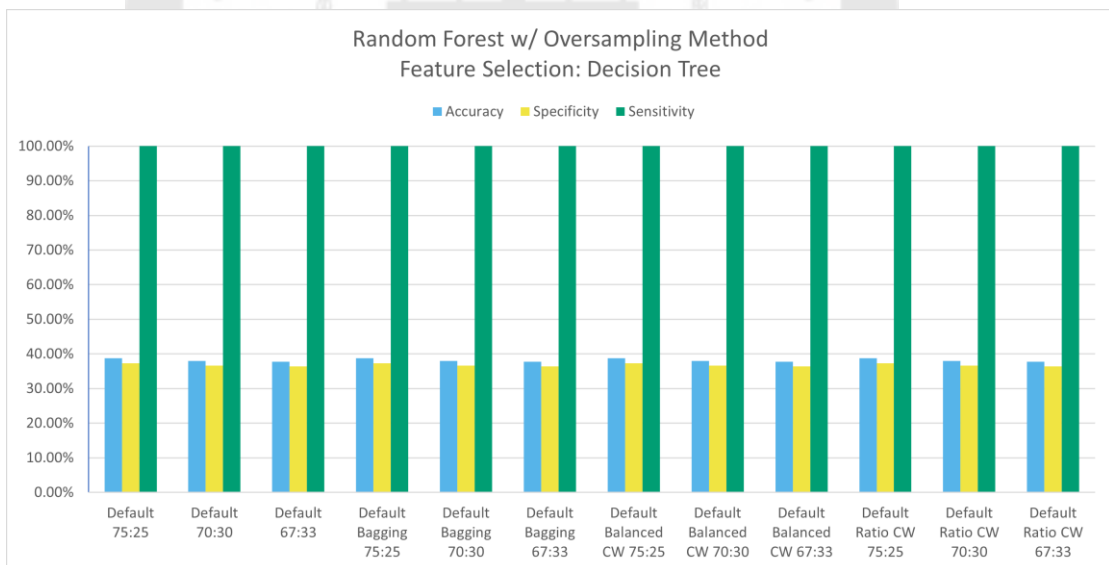
ภาพประกอบ 397 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling



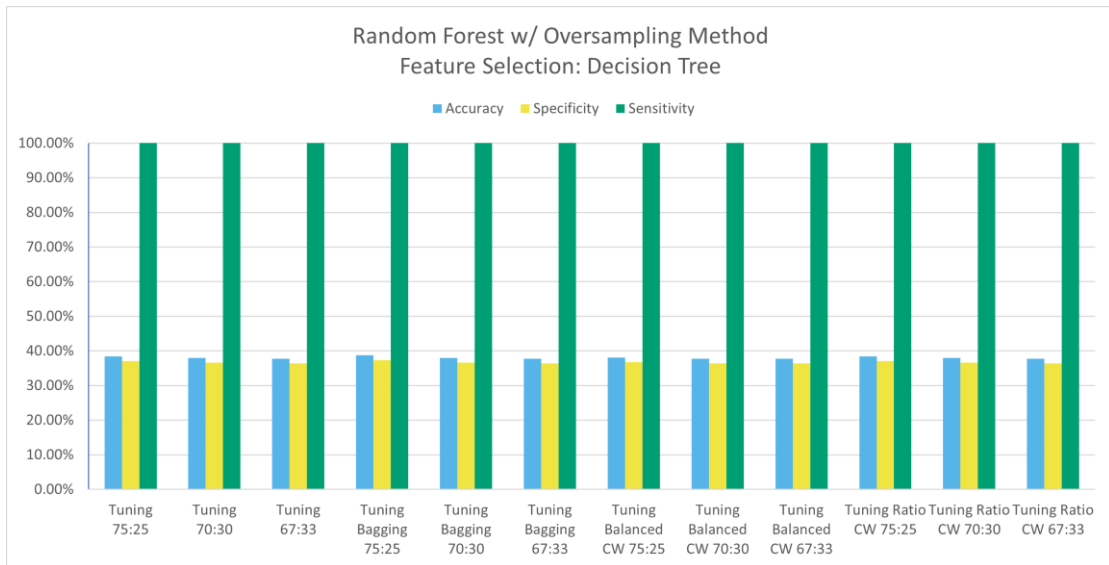
ภาพประกอบ 398 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับ พารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



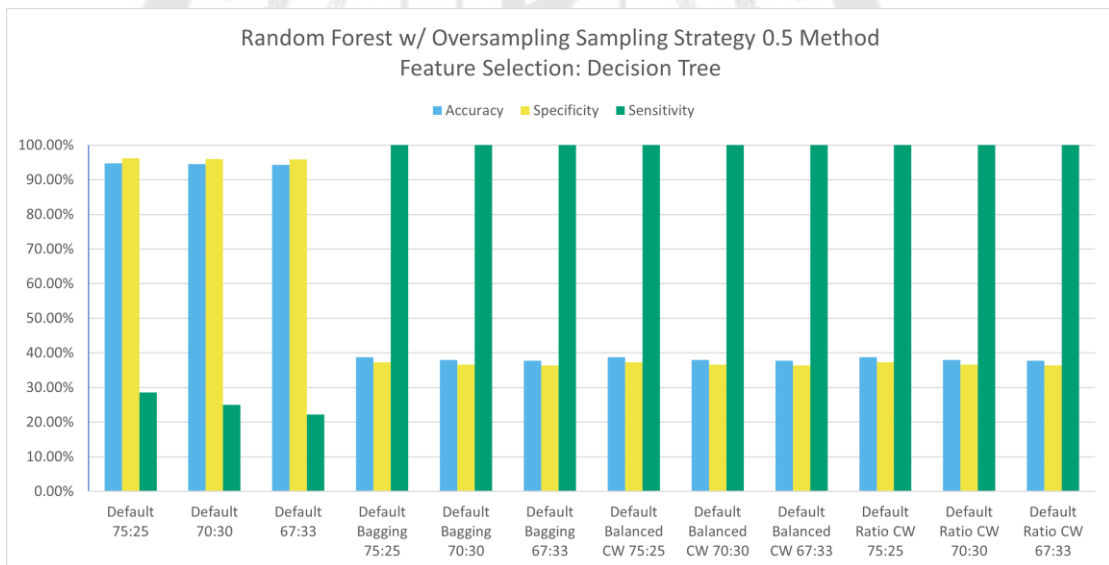
ภาพประกอบ 399 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับ พารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.5



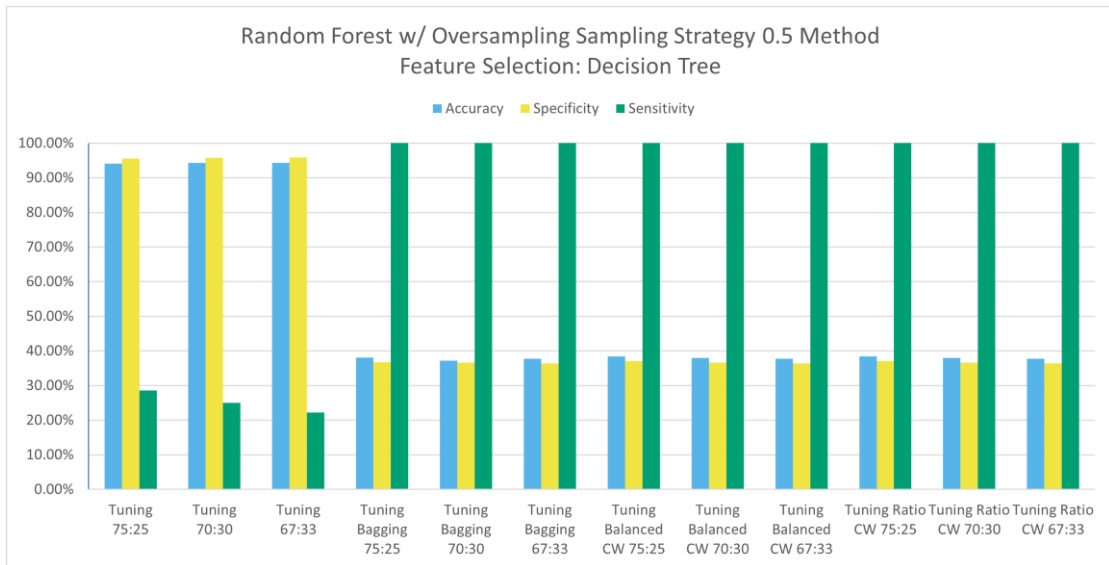
ภาพประกอบ 400 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling



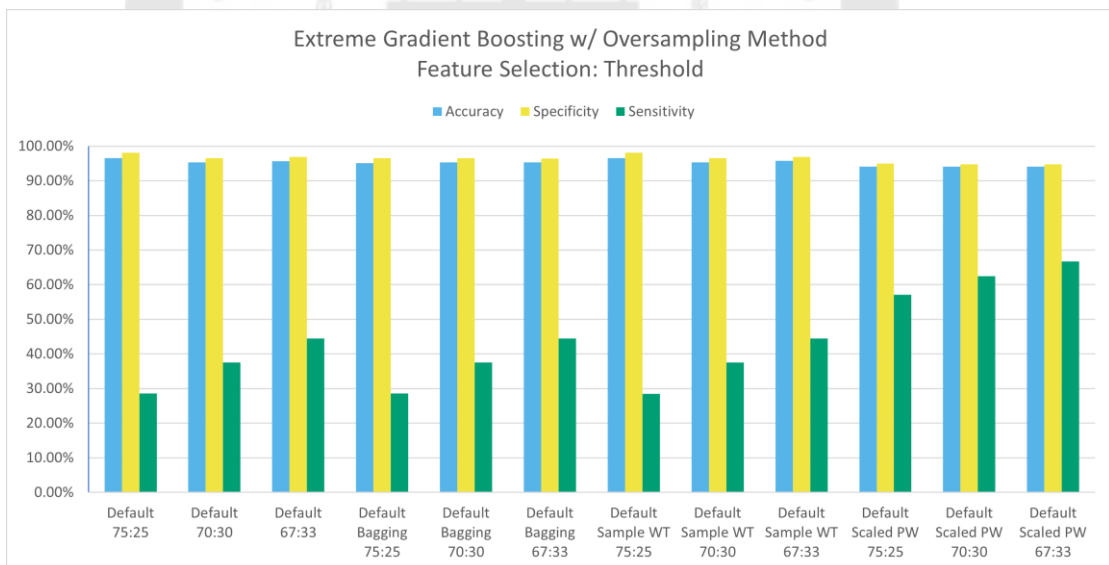
ภาพประกอบ 401 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling



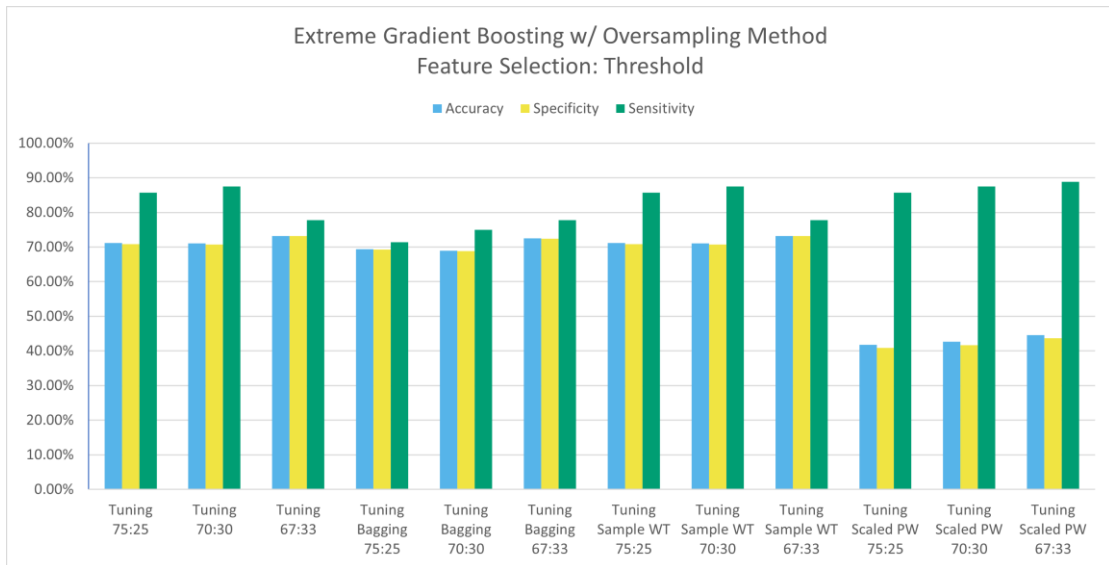
ภาพประกอบ 402 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



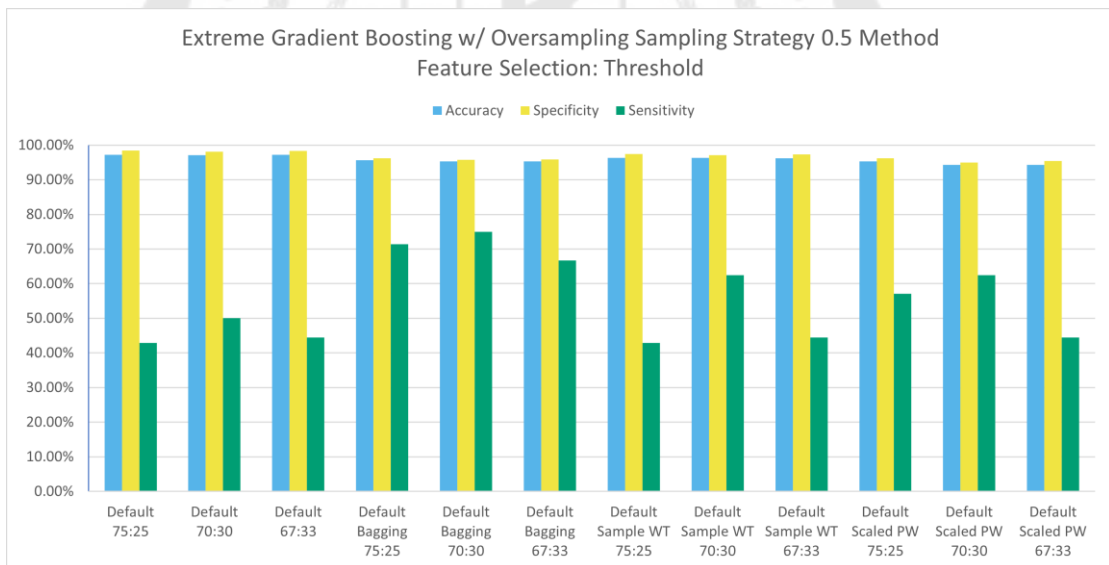
ภาพประกอบ 403 แผนภูมิแท่งของแบบจำลอง Random Forest กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.5



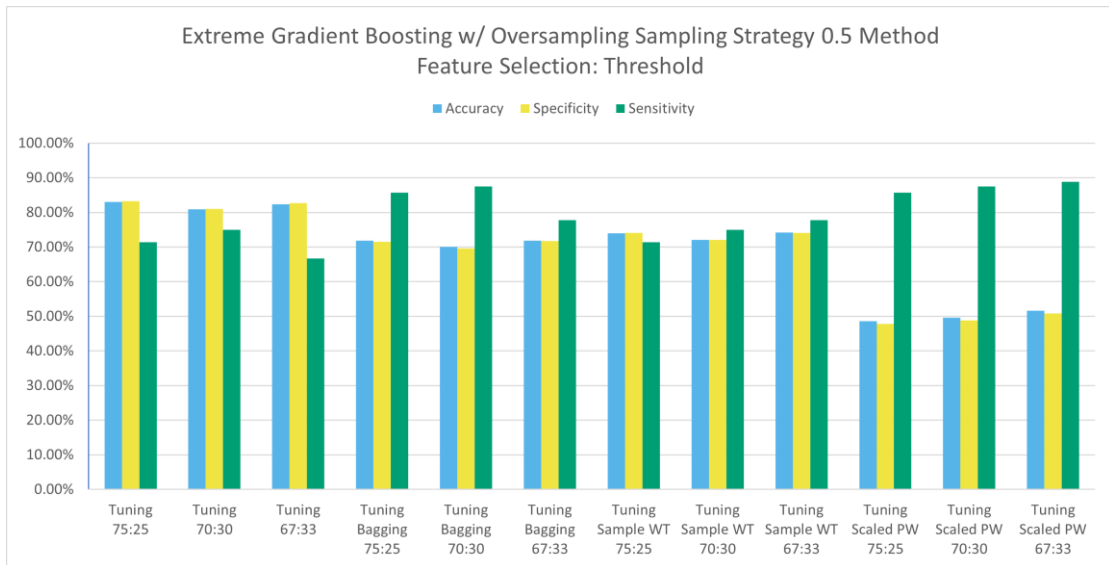
ภาพประกอบ 404 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling



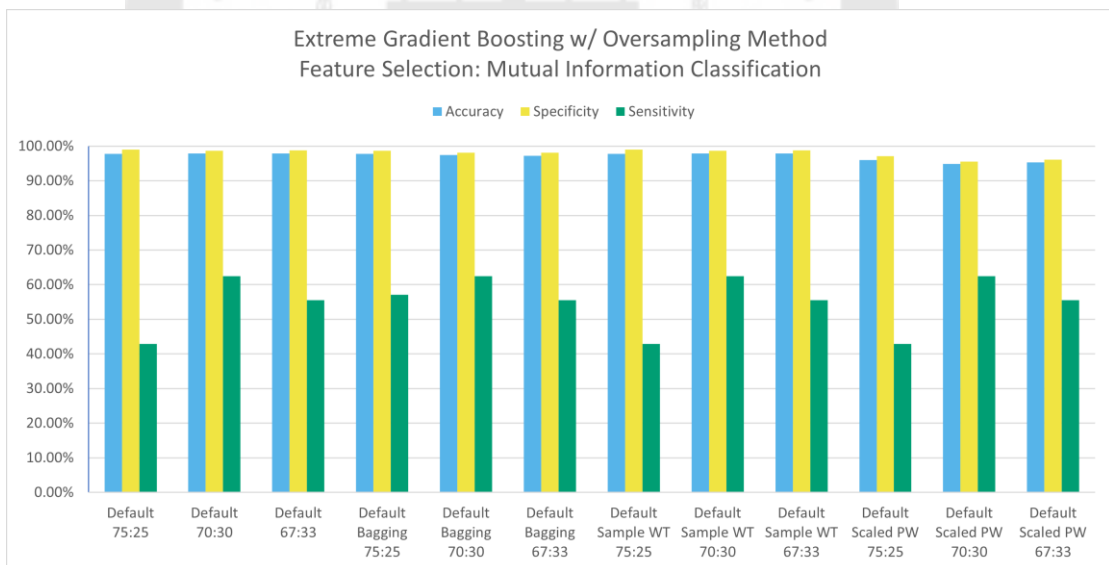
ภาพประกอบ 405 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling



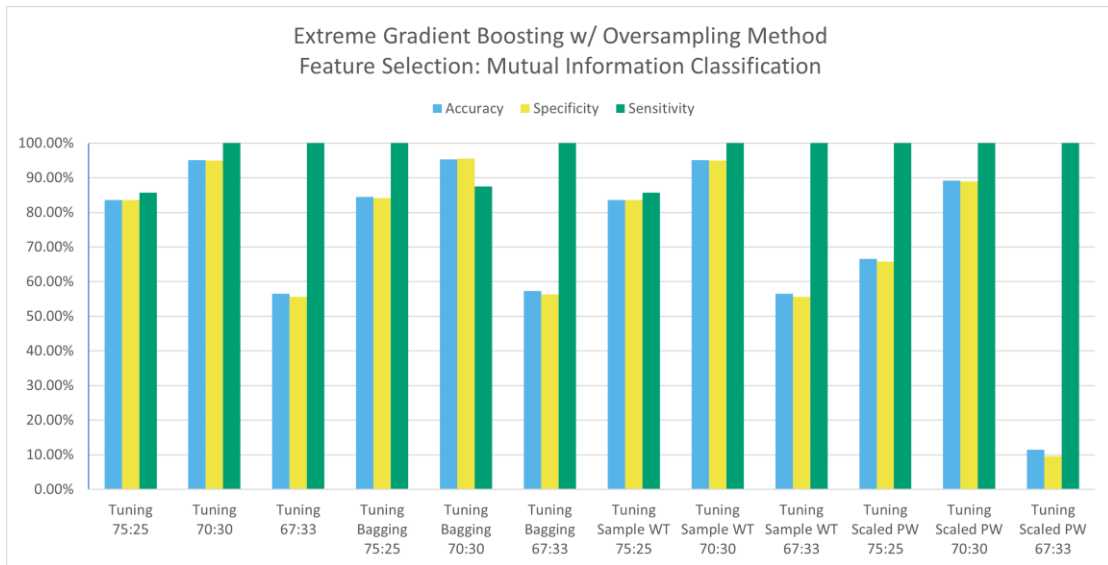
ภาพประกอบ 406 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



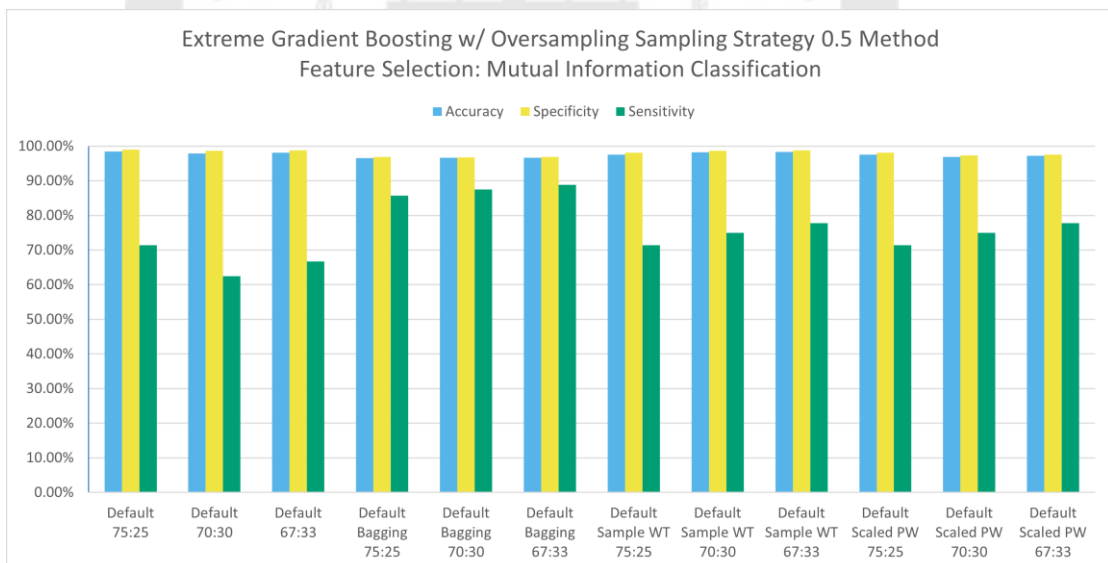
ภาพประกอบ 407 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Threshold และ Oversampling กับ พารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



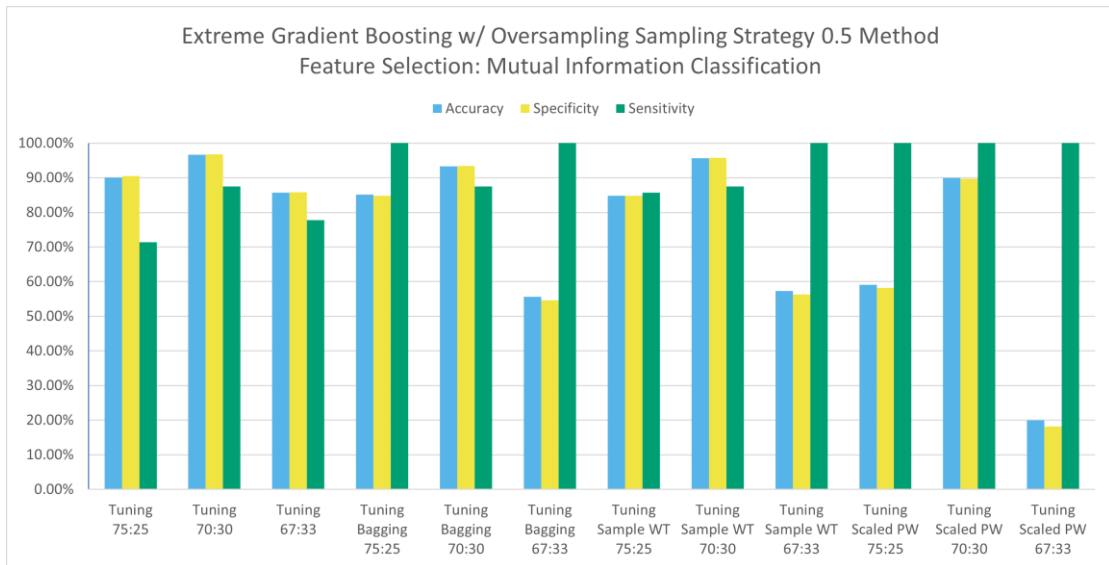
ภาพประกอบ 408 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling



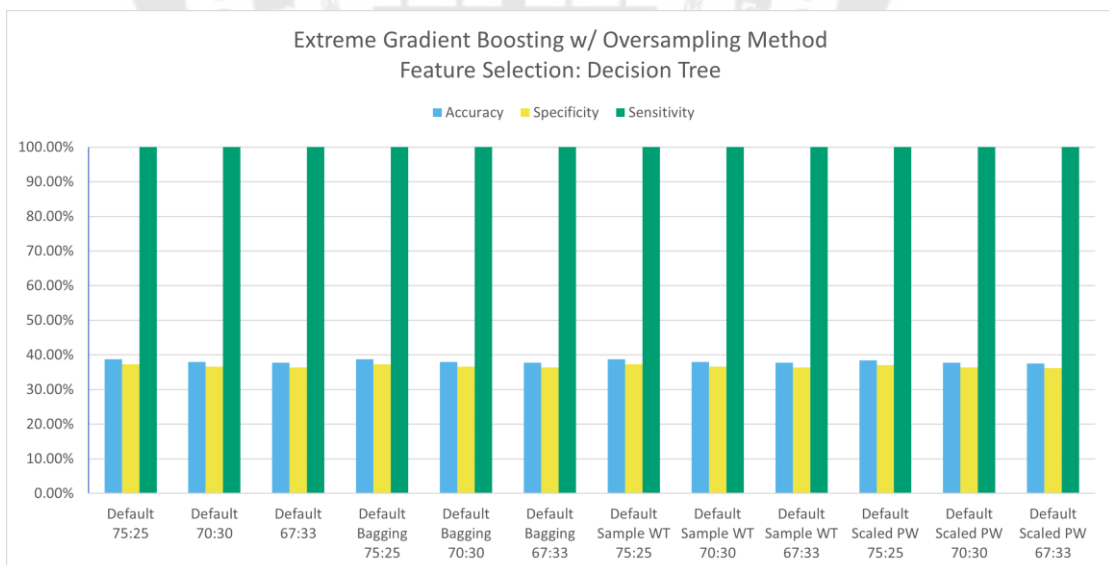
ภาพประกอบ 409 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling



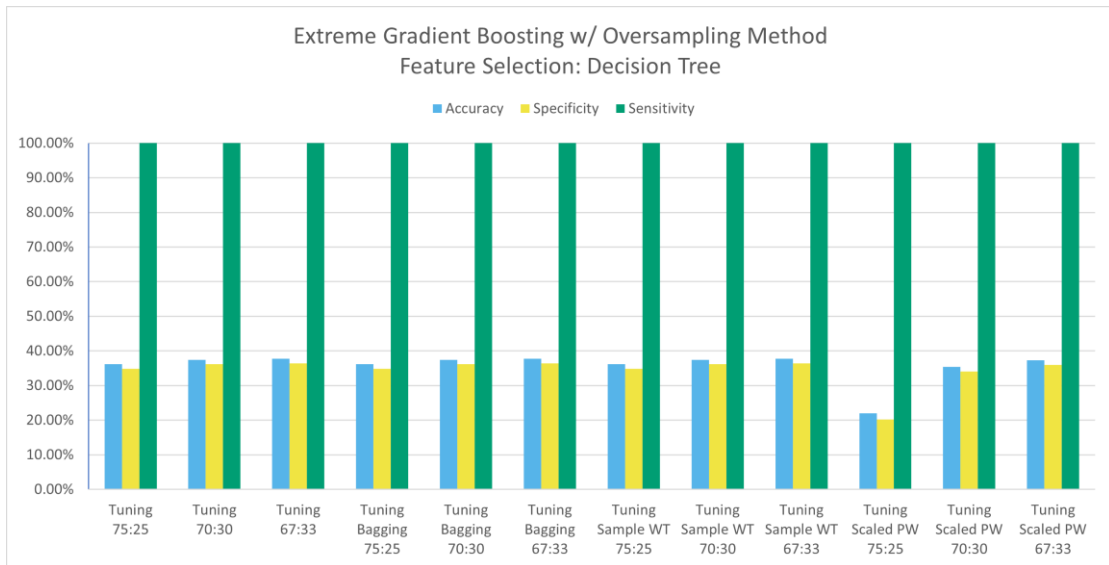
ภาพประกอบ 410 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



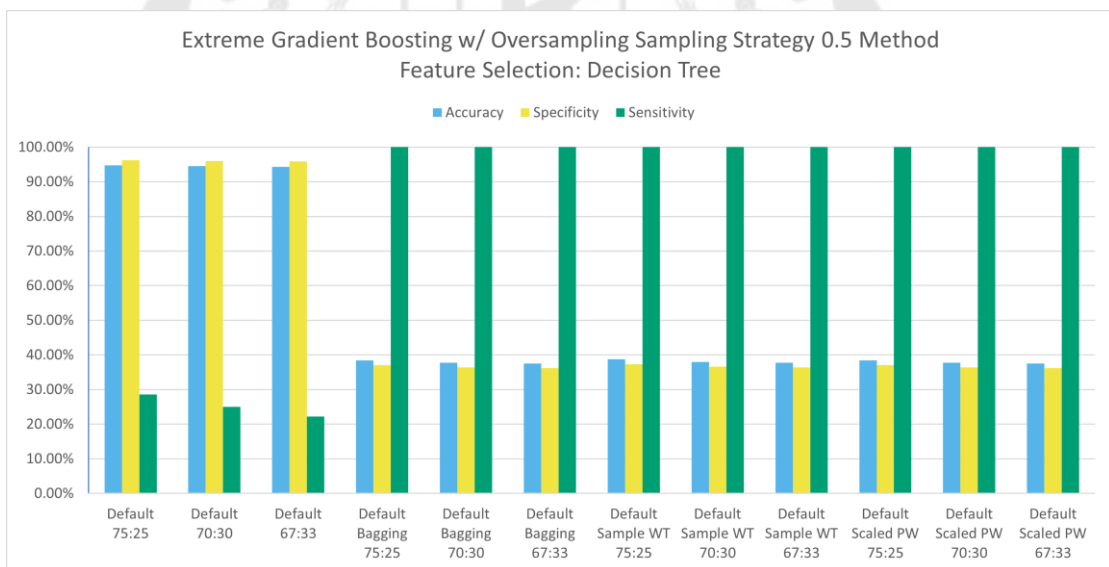
ภาพประกอบ 411 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Mutual Information Classification และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5



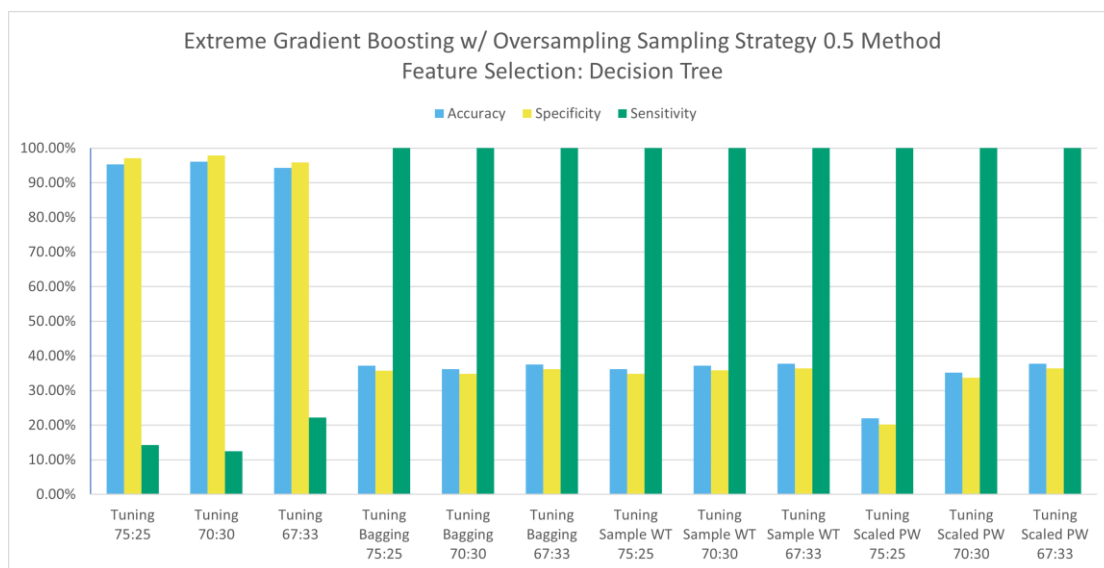
ภาพประกอบ 412 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling



ภาพประกอบ 413 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling



ภาพประกอบ 414 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Default Parameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5

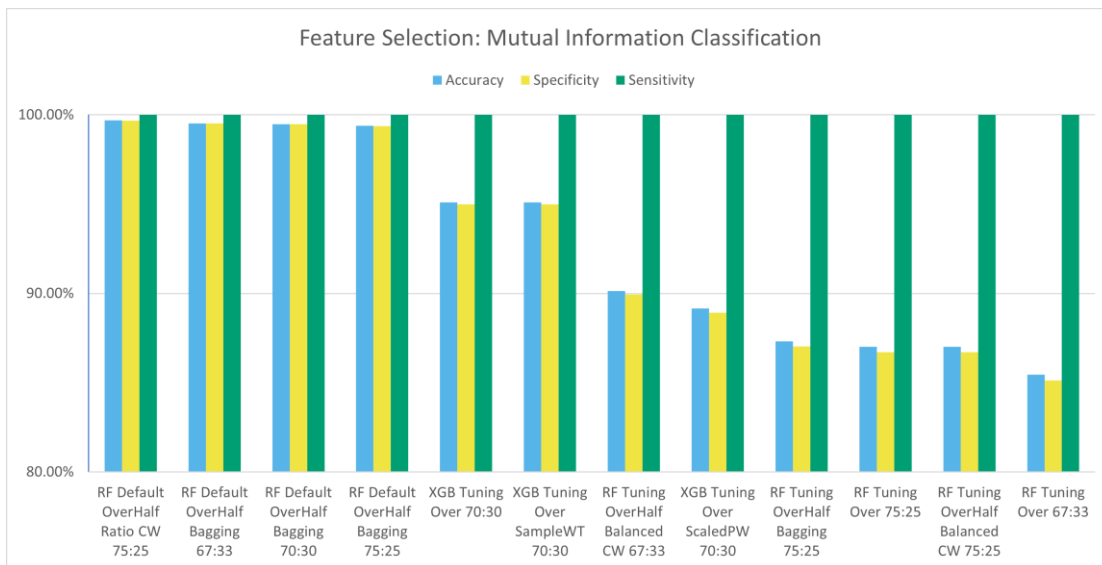


ภาพประกอบ 415 แผนภูมิแท่งของแบบจำลอง Extreme Gradient Boosting กับ Tuning Hyperparameters ที่ใช้ Feature Selection โดยวิธี Decision Tree และ Oversampling กับ พารามิเตอร์ `sampling_strategy` ที่มีค่าเท่ากับ 0.5

เนื่องจากรายงานวิจัยชิ้นนี้เกี่ยวกับทางการแพทย์ ดังนั้นการประเมินผลจะใช้ Sensitivity และ Specificity ที่มีค่าสูงๆ โดย Sensitivity จะเป็นเครื่องมือสำหรับการคัดกรองโรค หรือสำหรับเป็นเครื่องมือ Rule Out โรคที่สงสัย ซึ่งหากแบบจำลองทำนายผิดพลาด ทำให้ผู้ป่วยที่เป็นโรคจริงๆ จะถูกวินิจฉัยว่าไม่เป็นโรค ซึ่งทำให้ผู้ป่วยอาจจะไม่ได้เข้ารับการรักษา ถ้าหากผู้ป่วยเป็นโรคร้ายแรงก็อาจทำให้ผู้ป่วยถึงแก่ชีวิตได้ แบบจำลองที่ดีควรมีค่า Sensitivity สูงๆ เพื่อให้แบบจำลองสามารถทำนายผู้ป่วยที่เป็นโรคจริงๆ ได้อย่างแม่นยำมากที่สุด และทำนายผิดพลาดน้อยที่สุด

สำหรับ Specificity จะเป็นเครื่องมือสำหรับวินิจฉัยโรค (Diagnostic Test) หรือเป็นเครื่องมือ Rule In โรคที่สงสัย ซึ่งจะตรงกันข้ามกับ Sensitivity กล่าวคือ เป็นการวินิจฉัยผลการตรวจของคนปกติ ว่ามีโอกาสที่ไม่เป็นโรคน้อยเท่าไร ถ้าค่า Specificity มีค่าสูงก็จะแปลว่ามีโอกาสที่ไม่เป็นโรคสูง แบบจำลองที่ดีก็ควรมีค่า Specificity สูงๆ อีกเช่นกัน

จากภาพประกอบ 386 – 415 ผู้วิจัยจึงได้คัดเลือกแบบจำลองที่มีค่า Sensitivity เท่ากับ 100% นำมาเปรียบเทียบเพื่อหาแบบจำลองที่ให้ผลลัพธ์ที่ดีที่สุด ดังภาพประกอบ 416



ภาพประกอบ 416 แผนภูมิแท่งของแบบจำลอง 12 ตัวที่ให้ผลลัพธ์ที่ดีที่สุด

จากภาพประกอบ 416 นั้นได้เรียงผลลัพธ์ของแบบจำลองที่ดีที่สุดจากมากไปหาน้อย พบว่าแบบจำลองที่ดีที่สุดคือ Random Forest ที่ใช้พารามิเตอร์ Ratio Class Weight กับ Default Parameters โดยใช้ข้อมูลในอัตราส่วน 75:25 และทำการปรับความสมดุลของข้อมูลด้วยวิธี Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากงานวิจัยนี้ ผู้วิจัยได้ประเมินประสิทธิภาพของแบบจำลองแต่ละเทคนิคเพื่อนำมาเปรียบเทียบและสรุปผล โดยสามารถแบ่งหัวข้อในการสรุปผลได้ดังต่อไปนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ข้อเสนอแนะ

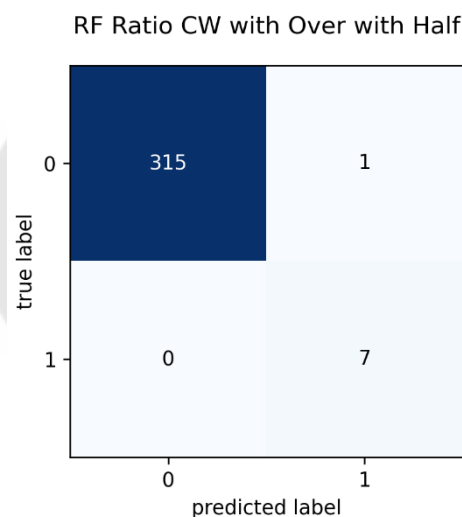
5.1 สรุปผลการวิจัย

ภาวะลิ้มเลือดอุดตันในหลอดเลือดดำเกิดจากปัจจัยเสี่ยงหลายๆ ปัจจัยด้วยกัน เมื่อเกิดภาวะลิ้มเลือดอุดตันในหลอดเลือดดำขึ้นแล้ว อาจต้องใช้ระยะเวลาในการรักษาตัวที่โรงพยาบาล ถึงแม้ว่าโรคลิ้มเลือดอุดตันในหลอดเลือดดำเป็นโรคที่พบได้ไม่บ่อยมากนัก แต่อาการที่เกิดขึ้นของโรคนี้มีความรุนแรงมาก ซึ่งอาจนำไปสู่เสียชีวิตได้อย่างรวดเร็ว จึงควรดูแลร่างกายให้แข็งแรง และตรวจสุขภาพเป็นประจำทุกปีอยู่เสมอ ดังนั้นภาวะการเกิดลิ้มเลือดอุดตันในหลอดเลือดดำจึงเป็นโรคที่ควรศึกษาและทำความเข้าใจเกี่ยวกับปัจจัยต่างๆ ที่จะนำไปสู่การเกิดโรคให้มากขึ้น นอกจากนี้แพทย์ รวมถึงบุคลากรทางการแพทย์ที่เกี่ยวข้องในการรักษา จะต้องตระหนักถึงการวินิจฉัยและการใช้เครื่องมือสำหรับประเมินการเกิดโรคให้เหมาะสม

งานวิจัยนี้ ได้สร้างแบบจำลองการวินิจฉัยภาวะลิ้มเลือดอุดตันในหลอดเลือดดำด้วยหลักการเรียนรู้ของเครื่อง โดยข้อมูลที่มาจากการเก็บรวบรวมข้อมูลโดยแพทย์ผู้เชี่ยวชาญในหอผู้ป่วยแผนกอายุรศาสตร์ โรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย ซึ่งมีข้อมูลทั้งหมด 1,290 แถว 65 คอลัมน์ จากนั้นผู้วิจัยได้ทำการเตรียมข้อมูลเพื่อวิเคราะห์ ตรวจสอบค่าที่ขาดหายไปของข้อมูล พร้อมทั้งแปลงข้อมูลที่เป็นตัวอักษรให้เป็นตัวเลข พร้อมทั้งนำไปสร้างแบบจำลองการทำนายในลำดับต่อไป จากนั้นแบ่งข้อมูลสำหรับการฝึกสอนและข้อมูลสำหรับการทดสอบด้วยอัตราส่วนตามที่กำหนดไว้ในข้างต้น และเนื่องจากข้อมูลงานวิจัยนี้มีปัญหาข้อมูลที่ไม่สมดุล เพราะมีจำนวนคนที่ไม่เป็นโรคลิ้มเลือดอุดตันทั้งหมด 1,263 คน และมีจำนวนคนที่เป็นโรคลิ้มเลือดอุดตันทั้งหมด 27 คน จะเห็นได้ว่ากลุ่มของคนที่เป็นโรคลิ้มเลือดอุดตันนั้นมีจำนวนมากเมื่อเทียบกับกลุ่มคนที่ไม่เป็นโรคลิ้มเลือดอุดตัน ซึ่งจะเป็นกระทบต่อการสร้างแบบจำลองการทำนาย ผู้วิจัยจึงจัดการกับปัญหาข้อมูลที่ไม่สมดุล จากนั้นทำการจูนไฮเปอร์พารามิเตอร์เพื่อหาพารามิเตอร์ที่ดีที่สุดของแต่ละอัลกอริทึมให้สามารถทำนายได้มีประสิทธิภาพที่ดีที่สุด โดยจะทำการประเมินประสิทธิภาพของ

แบบจำลองด้วยตาราง Confusion Matrix ซึ่งแบบจำลองการทำนาย 3 อัลกอริทึมที่เลือกใช้ได้แก่ Adaptive Boosting (AdaBoost), Random Forest และ Extreme Gradient Boosting (XGBoost)

ในการประเมินผลทางการแพทย์ค่า Sensitivity และ Specificity จะเลือกใช้การตรวจที่มีค่าดังกล่าวสูงๆ ในการตรวจคัดกรองผู้ป่วยที่เป็นโรคล้มเลือดอุดตัน เพราะเป็นการทำนายที่ถูกต้อง จากผลการทดลองพบว่าแบบจำลอง Random Forest ที่ใช้พารามิเตอร์ Ratio Class Weight กับ Default Parameters โดยใช้ข้อมูลในอัตราส่วน 75:25 และทำการปรับความสมดุลของข้อมูลด้วยวิธี Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.5 ซึ่งจะให้ค่า Sensitivity เท่ากับ 1.00 หรือ 100 เปอร์เซ็นต์ และค่า Specificity เท่ากับ 0.9968 หรือ 99.68 เปอร์เซ็นต์ ดังภาพประกอบ 417



ภาพประกอบ 417 ผลลัพธ์ Confusion Matrix ที่ได้จาก Random Forest กับ Default Parameters กับพารามิเตอร์ Ratio Class Weight โดยใช้อัตราส่วนของข้อมูลเท่ากับ 75:25 และใช้ Feature Selection ด้วยวิธี Mutual Information Classification

จากภาพประกอบ 421 ผลลัพธ์ของ Confusion Matrix สามารถอธิบายการประเมินผลทางการแพทย์ด้วยค่า Sensitivity และ Specificity ได้เป็นดังนี้

Sensitivity 100% คือ ผู้ป่วยที่เป็นโรคล้มเลือดอุดตันในหลอดเลือดดำมีทั้งหมด 7 ราย และแบบจำลองสามารถทำนายผู้ป่วยที่เป็นโรคล้มเลือดอุดตันในหลอดเลือดดำได้ถูกต้องทั้งหมด 7 ราย

Specificity 99.68% คือ ผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดดำมีทั้งหมด 316 ราย แต่แบบจำลองสามารถทำนายผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันในหลอดเลือดได้ 315 ราย และทำนายผิด 1 ราย

5.2 อภิปรายผลการวิจัย

งานวิจัยนี้ ได้สร้างแบบจำลองการวินิจฉัยอัตโนมัติสำหรับความเสี่ยงต่อการเกิดลิ่มเลือดอุดตันในหลอดเลือดดำตามอาการ โดยอาศัยการเรียนรู้ของเครื่อง เป็นการจำแนกประเภทและทำนายข้อมูลแบบ Classification ปัจจัยที่อาจมีผลต่อการทดลองของงานวิจัย คือ จากข้อมูลที่น่ามาวิเคราะห์นั้นเป็นข้อมูลทางการแพทย์ที่มีจำนวนน้อยมาก มีจำนวนแค่ 1,290 แถว พบว่าข้อมูลที่มีความหลากหลายของแต่ละ Feature และยังเป็นปัญหาข้อมูลที่ไม่สมดุลกัน ซึ่งพบผู้ป่วยที่ไม่เป็นโรคลิ่มเลือดอุดตันมีจำนวนมากกว่าผู้ป่วยที่เป็นโรคลิ่มเลือดอุดตัน จึงทำให้เกิดการ Sensitive กับข้อมูลคลาสที่มากกว่า และทำให้เกิด Bias ขึ้นกับข้อมูล เนื่องจากอัลกอริทึมจะไม่สามารถตรวจพบข้อมูลในคลาสที่น้อยกว่าได้ การจัดการกับปัญหาข้อมูลที่ไม่สมดุลนั้น เป็นวิธีการเปลี่ยนแปลงโครงสร้างและคุณสมบัติของข้อมูลเดิม โดยการเพิ่มเข้าหรือเอาออก ซึ่งอาจทำให้เกิดความสูญเสียคุณสมบัติของข้อมูลเดิมก็เป็นไปได้ ด้วยความจำเป็นของงานวิจัยนี้จึงจำเป็นต้องจัดการและแก้ปัญหาค่าข้อมูลให้สมดุลขึ้น เพื่อที่จะเพิ่มประสิทธิภาพในการทำนายของแบบจำลองให้มีประสิทธิภาพที่ดีมากยิ่งขึ้น

จากผลการทดลองพบว่า Resampling Data Method ด้วยวิธี Oversampling และ Oversampling กับพารามิเตอร์ sampling_strategy ที่มีค่าเท่ากับ 0.50 จะได้ผลลัพธ์ที่ดี แต่วิธี Synthetic Minority Oversampling Technique (SMOTE) นั้นจะได้ผลลัพธ์ที่ไม่ดีเท่าที่ควร เนื่องจาก SMOTE เป็นเทคนิคการสุ่มตัวอย่างแบบสุ่มเพิ่ม มิใช่เป็นการสุ่มเพิ่มจากข้อมูลเดิมเหมือน Oversampling ของทั้งสองวิธีนั้น ตามเอกสารอ้างอิงของ (Chawla, 2002) นั้น SMOTE จะเหมาะกับข้อมูลที่เป็นประเภท Continuous Data แต่ข้อมูลในงานวิจัยนี้เป็นประเภท Discrete Data จึงทำให้ได้ผลลัพธ์ที่ไม่ดี

จากปัจจัยเสี่ยงที่มีนัยสำคัญในงานวิจัยนี้ พบว่ายังสอดคล้องกับผลลัพธ์ที่ได้จากงานวิจัย Predictive and associative models to identify hospitalized medical patients at risk for VTE (Spyropoulos et al., 2011) ได้แก่ อายุ, การไม่เคลื่อนไหว, ภาวะหัวใจล้มเหลว, มะเร็ง และ ประวัติการเป็น VTE

ในการประเมินผลการทดลอง เพื่อให้แบบจำลองสามารถทำนายคนที่เป็นโรคลิ่มเลือดอุดตันให้ถูกต้องที่สุดนั้น จะเน้นดูผลลัพธ์ที่ค่า Sensitivity และ Specificity เนื่องจากข้อมูลไม่มีความ

สมดุล ทำให้แบบจำลองที่ทำนายผู้ป่วยที่ไม่เป็นโรค VTE ได้ดี แต่ทำนายผู้ป่วยที่เป็นโรค VTE ไม่ดีเท่าที่ควร จึงให้ผล Accuracy ที่สูง

ถึงแม้ว่างานวิจัยนี้จะให้ประสิทธิภาพของแบบจำลองที่ดี และสามารถจัดการกับปัญหาข้อมูลที่ไม่สมดุลได้แล้ว แต่จำนวนข้อมูลที่นำมาใช้ในการวิเคราะห์นั้นยังมีจำนวนน้อย ไม่สามารถนำมาใช้แทนประชากรทั้งหมดได้ และที่ผ่านมาก็เกิดโรคอุบัติใหม่ ๆ เกิดขึ้นมากมายทั่วโลก ซึ่งอาจทำให้แบบจำลองที่ได้จากงานวิจัยนี้ไม่สามารถทำนายข้อมูลที่เกี่ยวข้องกับโรคอุบัติใหม่ ๆ ได้

5.3 ข้อเสนอแนะ

5.3.1 ข้อมูลผู้ป่วยที่เป็นโรคล้มเลือดอุดตันในหลอดเลือดดำที่นำมาวิเคราะห์ในการวิจัยนี้มีจำนวนน้อยเกินไป เพราะใช้ระยะเวลาในการเก็บข้อมูลเพียง 1 ปีเท่านั้น ซึ่งทำให้เกิดปัญหาความไม่สมดุลของข้อมูล จึงจำเป็นต้องจัดการปัญหาความไม่สมดุลของข้อมูลดังกล่าวด้วยอัลกอริทึม ซึ่งอาจจะทำให้เกิดความสูญเสียคุณสมบัติเดิมของข้อมูลไป ในอนาคตอาจจะต้องเก็บข้อมูลที่จะใช้ในการวิเคราะห์มากขึ้น เพื่อให้ผลการทำนายของแบบจำลองมีประสิทธิภาพและแม่นยำมากยิ่งขึ้น

5.3.2 การจัดเก็บข้อมูลของผู้ป่วยควรให้มีความหลากหลาย และมีจำนวนมากพอที่มีนัยสำคัญทางสถิติ

5.3.3 เพิ่มการทำ Feature Selection ด้วยวิธีการอื่น ๆ ที่หลากหลายขึ้น เพื่อลด Dimension ของข้อมูล ทำให้สามารถสร้างแบบจำลองที่มีประสิทธิภาพได้ดียิ่งขึ้น

5.3.4 ในอนาคตอาจจะมีอัลกอริทึมใหม่ ๆ ที่สามารถจัดการแก้ปัญหาความไม่สมดุลของข้อมูลที่ไม่ทำให้ข้อมูลที่มีอยู่เดิมไม่สูญเสียคุณสมบัติของข้อมูลไป รวมถึงอาจจะมีอัลกอริทึมใหม่ ๆ ที่เหมาะสมสามารถใช้ได้กับข้อมูลที่นำมาวิเคราะห์ในงานวิจัยนี้ได้

บรรณานุกรม

boom626. (2019). Confusion Matrix.

https://medium.com/@mirthful_sunset_cattle_231/confusion-matrix-48cc396b1b58

Chawla, N. V. (2002). SMOTE Synthetic Minority Over sampling Technique. *Artificial Intelligence Research*, 16(1).

Crous-Bou, M., Harrington, L. B., & Kabrhel, C. (2016). Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. *Semin Thromb Hemost*, 42(8), 808-820.

Days, A. (2020). Deep-Vein Thrombosis Awareness Month 2020.

<https://www.awarenessdays.com/awareness-days-calendar/category/2020/2020-03-01/>

Drugs.com. (2022, May 2, 2022). Pulmonary Embolism.

<https://www.drugs.com/cg/pulmonary-embolism-discharge-care.html>

Eakasit Pacharawongsakda, P. D. (2015). การคัดเลือก feature (feature selection) ด้วยวิธี Information Gain. <https://th.linkedin.com/pulse/การคัดเลือก-feature-selection-ด้วยวิธี-information-gain-pacharawongsakda>

Haibo, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.

Health, U. Pulmonary Embolism (PE).

<https://www.uclahealth.org/pulmonary/pvd/pulmonary-embolism>

Mahapatra, A. (2022). A Comprehensive Guide to Ensemble Learning Methods.

Nafee, T., Gibson, C. M., Travis, R., Yee, M. K., Kerneis, M., Chi, G., . . . Goldhaber, S. Z. (2020). Machine learning to predict venous thrombosis in acutely ill medical patients. *Res Pract Thromb Haemost*, 4(2), 230-237.

Palta, S., Saroa, R., & Palta, A. (2014). Overview of the coagulation system. *Indian J Anaesth*, 58(5), 515-523.

Pimpim, P. (2019). Feature selection การเลือกคุณลักษณะ.

<https://medium.com/@doohpim/03-feature-selection-การเลือกคุณลักษณะ->

446f8f0c094b

Praharsha, V.). One hot encoding in TensorFlow.

Rojnuckarin, P., Uaprasert, N., Vajragupta, L., Numkarunrunrote, N., Tanpowpong, N., & Sutcharitchan, P. (2011). Risk factors for symptomatic venous thromboembolism in Thai hospitalised medical patients. *Thromb Haemost*, 106(6), 1103-1108.

Ryan, L., Mataraso, S., Siefkas, A., Pellegrini, E., Barnes, G., Green-Saxena, A., . . . Das, R. (2021). A Machine Learning Approach to Predict Deep Venous Thrombosis Among Hospitalized Patients. *Clin Appl Thromb Hemost*, 27, 1076029621991185.

Spyropoulos, A. C., Anderson, F. A., Jr., FitzGerald, G., Decousus, H., Pini, M., Chong, B. H., . . . Investigators, I. (2011). Predictive and associative models to identify hospitalized medical patients at risk for VTE. *Chest*, 140(3), 706-714.

Sukperm, A., Rojnuckarin, P., Akkawat, B., & Sa-Ing, V. (2021). *Automatic Diagnosis of Venous Thromboembolism Risk based on Machine Learning*. Paper presented at the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS).

Tapson, V. F., Decousus, H., Pini, M., Chong, B. H., Froehlich, J. B., Monreal, M., . . . Investigators, I. (2007). Venous thromboembolism prophylaxis in acutely ill hospitalized medical patients: findings from the International Medical Prevention Registry on Venous Thromboembolism. *Chest*, 132(3), 936-945.

TITIPATA.). Bagging (Bootstrap Aggregation).

Tseng, G. (2018). Gradient Boosting and XGBoost. Gradient Boosting and XGBoost

โรงพยาบาลบำรุงราษฎร์. ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำ.

[https://www.bumrungrad.com/th/conditions/venous-thromboembolism-](https://www.bumrungrad.com/th/conditions/venous-thromboembolism-vte#:~:text=pulmonary%20embolism%3A%20PE)

[vte#:~:text=pulmonary%20embolism%3A%20PE](https://www.bumrungrad.com/th/conditions/venous-thromboembolism-vte#:~:text=pulmonary%20embolism%3A%20PE));ภาวะลิ่มเลือดอุดตันในหลอดเลือดดำส่วนลึก,(pulmonary%20embolism%3A%20PE)%20ได้

โรงพยาบาลพญาไท. (2021, 18 พ.ค. 64). ลิ่มเลือดอุดตัน ภัยร้ายที่หลายคนคาดไม่ถึง. 18 พ.ค. 64.

https://www.phyathai.com/article_detail/3523/th/ลิ่มเลือดอุดตัน_ภัยร้ายที่หลายคนคาดไม่ถึง?branch=PYT3

คำอินทร์, ส. ระบบอวัยวะในร่างกายของเรา.

<https://sites.google.com/site/suwaneescience/bth-thi-2-rangkay-khxng-rea/1-rabb-xwaywa-ni-rangkay-khxng-rea>

พบแพทย์. (2016). ภาวะลิ่มเลือดในหลอดเลือดดำอุดตัน (VTE). [https://www.pobpad.com/vte-ภาวะลิ่มเลือดในหลอดเลือด#:~:text=VTE%20\(Venous%20Thrombosis\)%20คือภาวะ,ว่ามีอาการผิดปกติ](https://www.pobpad.com/vte-ภาวะลิ่มเลือดในหลอดเลือด#:~:text=VTE%20(Venous%20Thrombosis)%20คือภาวะ,ว่ามีอาการผิดปกติ)

มหาวิทยาลัยมหิดล, ศ. ค. (2020). ประวัติความเป็นมา.

<https://www.si.mahidol.ac.th/th/division/SiStroke/history.asp>

วงการแพทย์. ลิ่มเลือดอุดตันของหลอดเลือดดำโรคที่คนรู้จักน้อย แต่เสี่ยงให้เกิดภาวะทุพพลภาพ
เรื้อรัง. <http://www.wongkarnpat.com/viewya.php?id=2485>

สำนักงานสถิติแห่งชาติ. ประเภทของข้อมูลสถิติ.

http://service.nso.go.th/nso/knowledge/estat/esta1_3.html



ประวัติผู้เขียน

ชื่อ-สกุล	ภาคพล อนุภูมิชัยยา
วัน เดือน ปี เกิด	7 ตุลาคม 2521
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	พ.ศ. 2542 วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัย รังสิต
ที่อยู่ปัจจุบัน	2/102 ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพมหานคร 10400

