

การตรวจจับการฟอกเงินโดยใช้การเรียนรู้ของเครื่อง

DETECTING MONEY LAUNDERING USING MACHINE LEARNING



ชนพัฒน์ เมืองเชียงหวาน

CHANAPAT MUEANGCHIANGWARN

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์  
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง  
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
พ.ศ. 2567

KMITL-2024-SC-M-017-029

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# DETECTING MONEY LAUNDERING USING MACHINE LEARNING



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE  
IN DATA SCIENCE AND ANALYTICS

KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2024

KMITL-2024-SC-M-017-029

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2024

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การตรวจจับการฟอกเงินโดยใช้การเรียนรู้ของเครื่อง
ชื่อนักศึกษา	ชนพัฒน์ เมืองเชียงหวาน
รหัสประจำตัว	64605020
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
พ.ศ.	2567
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ

### บทคัดย่อ

การค้นคว้าอิสระนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการตรวจจับธุรกรรมฟอกเงิน (Money Laundering) โดยใช้ชุดข้อมูลที่นำมาจากเว็บไซต์ Kaggle ซึ่งเป็นข้อมูลสังเคราะห์จากธุรกรรมจริงของธนาคารในประเทศไทยที่มีคุณลักษณะทั้งหมด 12 รายการ โดยจะใช้วิธีการจัดการข้อมูลไม่สมดุล 2 วิธี คือ การทำ Undersampling และ Oversampling แล้วทำการเปรียบเทียบประสิทธิภาพของแบบจำลองเอ็กซ์ตรีมกราเดียนต์บูสตีง (XGBoost) และไลต์กราเดียนต์บูสตีงแมชชีน (LightGBM) ผลการวิจัยพบว่าแบบจำลองที่จัดการข้อมูลด้วยวิธี Oversampling ให้ประสิทธิภาพดีกว่าวิธี Undersampling โดยผลลัพธ์จากการทำ Oversampling ของแบบจำลอง XGBoost นั้นให้ค่าความแม่นยำสูงถึง 99.93% โดยไม่เกิดปัญหา Overfitting และแบบจำลอง LightGBM มีประสิทธิภาพรองลงมา ซึ่งให้ค่าความแม่นยำที่ 89.40% ในขณะที่การทำ Undersampling ในการจัดการข้อมูลของแบบจำลอง XGBoost ให้ค่าความแม่นยำที่ 75.12% และแบบจำลอง LightGBM ให้ค่าความแม่นยำที่ 75.35% สรุปได้ว่าการทำ Oversampling เป็นวิธีที่เหมาะสมสำหรับการจัดการปัญหาข้อมูลไม่สมดุลในชุดข้อมูลสังเคราะห์นี้ อีกทั้งยังช่วยลดปัญหาผลบวกหลงได้อย่างมีประสิทธิภาพมากกว่าการทำ Undersampling

**คำสำคัญ :** การฟอกเงิน การเรียนรู้ของเครื่อง เอ็กซ์ตรีมกราเดียนต์บูสตีง ไลต์กราเดียนต์บูสตีงแมชชีน

Independent Study Title	Detecting Money Laundering Using Machine Learning
Student Name	Chanapat Mueangchiangwarn
Student ID	64605020
Degree	Master of Science (Data Science and Analytics) KMITL-Digital Analytics and Intelligence Center
Year	2024
Independent Study Advisor	Asst. Prof. Dr. Busayamas Pimpunchat

### Abstract

This independent study aims to develop machine learning models for detecting money laundering transactions using a dataset from the website Kaggle. The study utilized a synthetic dataset derived from real bank transactions in the UK, comprising 12 features. Then comparing the performance of XGBoost and LightGBM models using two methods for handling imbalanced data by undersampling and oversampling. The results indicate that models using oversampling outperform those using undersampling. The oversampling results of the XGBoost model achieves an accuracy of 99.93% without overfitting, and the LightGBM model achieves an accuracy of 89.40%. While the models using undersampling, the XGBoost model achieving an accuracy of 75.12% and the LightGBM model achieving an accuracy of 75.35%. Thus, oversampling is concluded to be the most appropriate method for handling imbalanced data in this synthetic dataset, as it effectively reduces false positives compared to undersampling.

**Keywords :** Money Laundering, Machine Learning, Extreme Gradient Boosting, Light Gradient Boosting Machine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

งานค้นคว้าอิสระนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองในการตรวจจับธุรกรรมการฟอกเงิน โดยใช้วิธีการเรียนรู้ของเครื่อง เพื่อมาเปรียบเทียบประสิทธิภาพทั้งหมด 2 วิธี ได้แก่ XGBoost และ LightGBM ซึ่งผู้วิจัยขอขอบพระคุณอาจารย์ที่ปรึกษา ผศ.ดร.บุษยมาศ พิมพ์พรรณชาติ ที่ได้ให้คำชี้แนะ แนวทางการแก้ไขปัญหา ความรู้เพิ่มเติมเกี่ยวกับการสร้างแบบจำลอง และการประยุกต์ใช้ความรู้ให้เกิดประโยชน์สูงสุด จนงานวิจัยนี้สำเร็จลุล่วงไปด้วยดี

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง และผู้ปกครองที่มอบโอกาสให้ได้เข้าศึกษาในสถาบันแห่งนี้ ซึ่งเป็นอีกหนึ่งปัจจัยสำคัญที่ทำให้ได้พบกับคณาจารย์ และวิทยากรที่มีศักยภาพในการช่วยพัฒนาทักษะและมอบความรู้ให้แก่นักศึกษา

ชนพัฒน์ เมืองเชียงหวาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

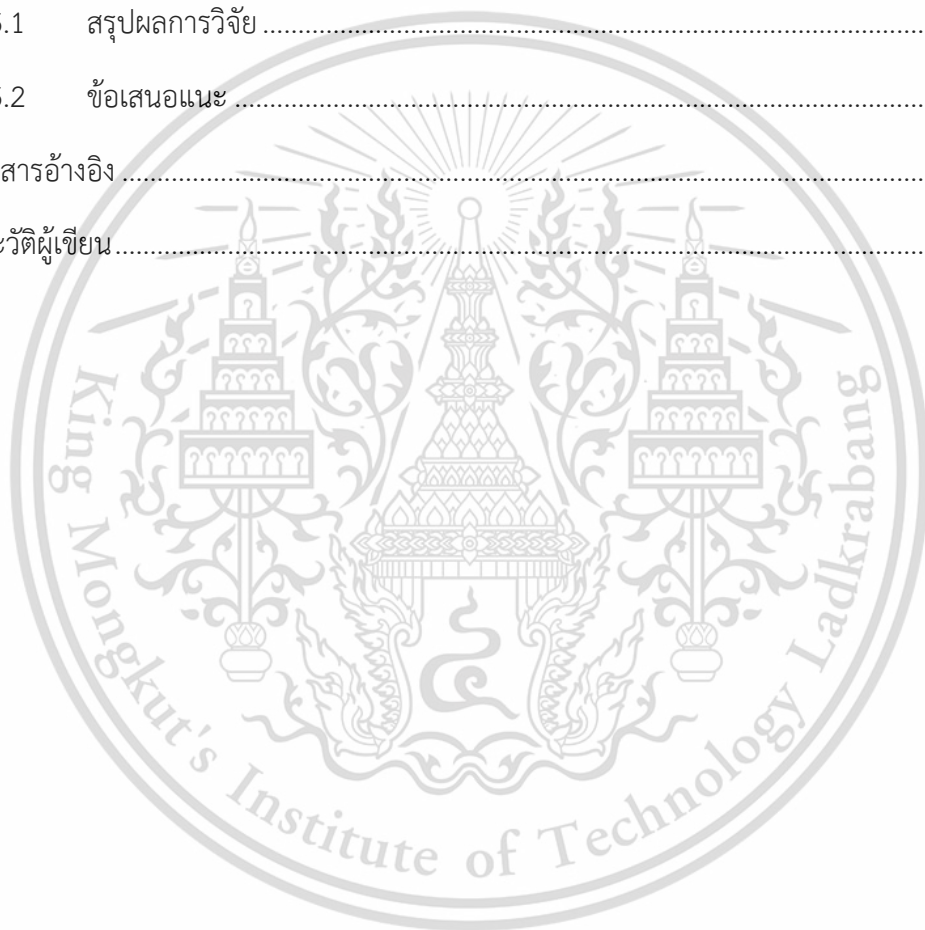
	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ช
สารบัญรูป.....	ฉ
<b>บทที่ 1 บทนำ.....</b>	<b>1</b>
1.1    ความเป็นมาและความสำคัญของปัญหา.....	1
1.2    วัตถุประสงค์ของงานวิจัย.....	2
1.3    ขอบเขตของงานวิจัย.....	2
1.4    วิธีการดำเนินงานวิจัย.....	3
1.5    ประโยชน์ที่คาดว่าจะได้รับ.....	3
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....</b>	<b>4</b>
2.1    แนวคิดและทฤษฎีที่เกี่ยวข้อง.....	4
2.1.1    การฟอกเงิน (Money Laundering).....	4
2.1.2    ขั้นตอนในการฟอกเงิน.....	5
2.1.3    องค์ประกอบความผิดฐานฟอกเงิน.....	5
2.1.4    Workflow Scenario.....	6
2.1.5    การเรียนรู้ของเครื่อง (Machine Learning).....	6
2.1.6    การทำ Cross validation.....	9
2.1.7    ข้อมูลไม่สมดุล (Imbalanced data).....	9
2.1.8    ขั้นตอนการสร้างแบบจำลอง.....	11
2.1.9    การปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning).....	11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.10	การประเมินประสิทธิภาพของแบบจำลอง .....	12
2.2	งานวิจัยที่เกี่ยวข้อง.....	13
2.2.1	Anti-Money Laundering Alert Optimization Using Machine Learning with Graphs .....	13
2.2.2	Automatic suppression of false positive alerts in anti-money laundering systems using machine learning.....	14
2.2.3	Improving Classification Performance of Money Laundering Transactions Using Typological Features .....	14
2.2.4	Machine Learning for Anti-Money Laundering and Fraud Detection .....	15
2.2.5	Money Laundering Detection using A Transaction-based Graph Learning Approach.....	15
2.2.6	SAML-D Exploratory Data Analysis & Classification .....	15
<b>บทที่ 3</b>	<b>วิธีการดำเนินงานวิจัย .....</b>	<b>16</b>
3.1	ขั้นตอนการดำเนินงานวิจัย.....	16
3.1.1	ศึกษาระบบตรวจจับการฟอกเงิน.....	17
3.1.2	จัดหาข้อมูลและทำการศึกษาค้นคว้าข้อมูล .....	17
3.1.3	เตรียมข้อมูลและวิเคราะห์ข้อมูล.....	18
3.1.4	ดำเนินการจัดการกับข้อมูลที่ไม่สมดุล .....	27
3.1.5	แบ่งชุดข้อมูล .....	28
3.1.6	สร้างแบบจำลองตรวจจับการฟอกเงิน .....	29
3.1.7	ประเมินประสิทธิภาพของแบบจำลอง.....	31
3.2	เครื่องมือที่ใช้ในการวิจัย.....	31
<b>บทที่ 4</b>	<b>ผลการวิจัยและการอภิปรายผล .....</b>	<b>32</b>
4.1	ผลลัพธ์ของแบบจำลอง XGBoost.....	32
4.1.1	ผลลัพธ์ของแบบจำลอง XGBoost จากการทำ Undersampling.....	33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2	ผลลัพธ์ของแบบจำลอง XGBoost จากการทำ Oversampling .....	37
4.2	ผลลัพธ์ของแบบจำลอง LightGBM .....	41
4.2.1	ผลลัพธ์ของแบบจำลอง LightGBM จากการทำ Undersampling .....	43
4.2.2	ผลลัพธ์ของแบบจำลอง LightGBM จากการทำ Oversampling .....	47
4.3	เปรียบเทียบประสิทธิภาพของทั้ง 2 แบบจำลอง.....	52
<b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....</b>		<b>54</b>
5.1	สรุปผลการวิจัย .....	54
5.2	ข้อเสนอแนะ .....	56
เอกสารอ้างอิง .....		58
ประวัติผู้เขียน .....		60



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
3.1 รายละเอียด ชนิด และจำนวนของข้อมูล .....	18
3.2 การแบ่งชุดข้อมูลที่ผ่านมาการทำ Undersampling .....	28
3.3 การแบ่งชุดข้อมูลที่ผ่านมาการทำ Oversampling.....	29
3.4 ค่าที่ใช้ปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost.....	30
3.5 ค่าที่ใช้ปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM .....	30
4.1 การจับคู่ของพารามิเตอร์ทั้งหมดในตาราง Grid Search ของแบบจำลอง XGBoost .....	32
4.2 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการ ทำ Undersampling .....	33
4.3 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling .....	34
4.4 อัตราส่วนการทำนายของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling	36
4.5 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง XGBoost โดยใช้ข้อมูลจาก การทำ Undersampling .....	36
4.6 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการ ทำ Oversampling .....	37
4.7 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling.....	38
4.8 อัตราส่วนการทำนายของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling..	40
4.9 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง XGBoost โดยใช้ข้อมูลจาก การทำ Oversampling.....	40
4.10 การจับคู่ของพารามิเตอร์ทั้งหมดในตาราง Grid Search ของแบบจำลอง LightGBM .....	42
4.11 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจาก การทำ Undersampling .....	43
4.12 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling .....	45
4.13 อัตราส่วนการทำนายของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling .....	46
4.14 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง LightGBM โดยใช้ข้อมูลจาก การทำ Undersampling .....	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

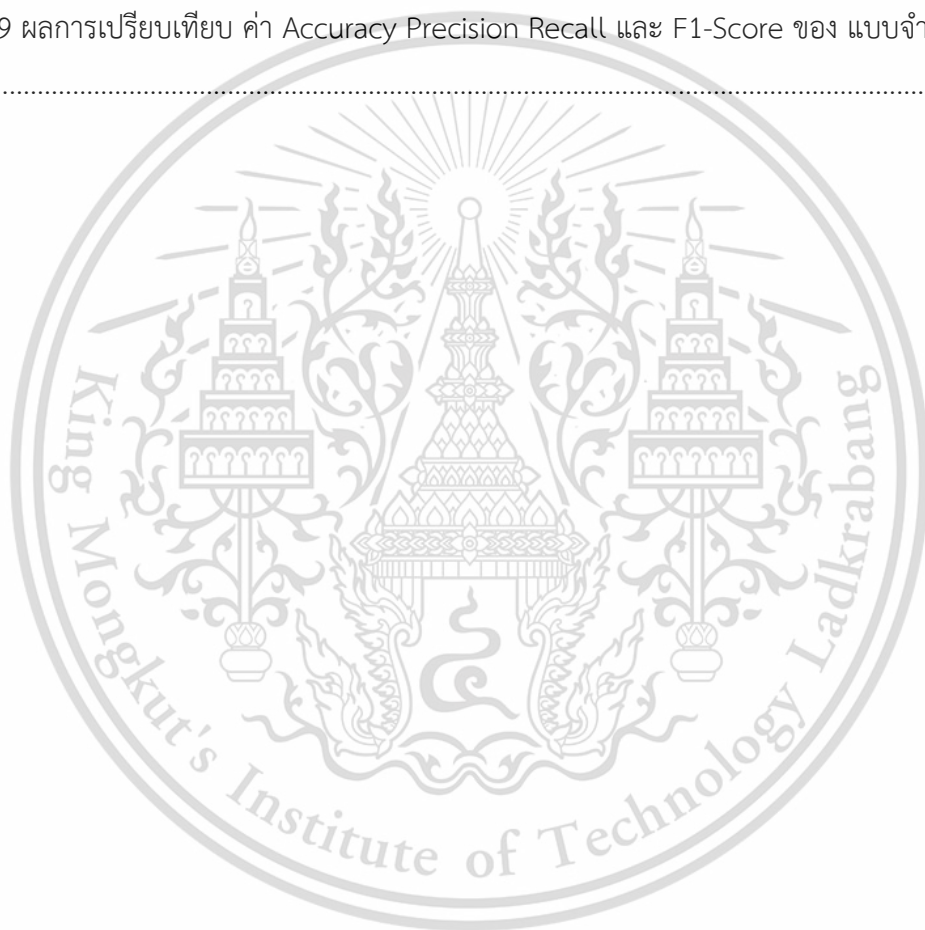
4.15 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling..... 48

4.16 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling ..... 50

4.17 อัตราส่วนการทำนายของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling ..... 51

4.18 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling..... 51

4.19 ผลการเปรียบเทียบ ค่า Accuracy Precision Recall และ F1-Score ของ แบบจำลองทั้งหมด ..... 53



## สารบัญรูป

รูปที่	หน้า
2.1 วิธีการแบ่งชุดข้อมูลในการทำ Cross validation .....	9
2.2 วิธีในการจัดการกับปัญหาข้อมูลไม่สมดุล .....	10
2.3 ตาราง Confusion Matrix .....	12
2.4 การคำนวณหาค่า Accuracy .....	12
2.5 การคำนวณหาค่า Precision .....	13
2.6 การคำนวณหาค่า Recall .....	13
2.7 การคำนวณหาค่า F - score.....	13
3.1 ขั้นตอนการดำเนินงานวิจัย .....	16
3.2 ตัวอย่างการทำงานของระบบตรวจจับการฟอกเงิน .....	17
3.3 จำนวนธุรกรรมทั้งหมด โดยแบ่งตามป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน .....	19
3.4 จำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นในแต่ละเดือน .....	19
3.5 จำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นในแต่ละวัน .....	20
3.6 จำนวนธุรกรรมทั้งหมดที่เกิดขึ้นในแต่ละชั่วโมง .....	20
3.7 จำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นในแต่ละชั่วโมง .....	21
3.8 การกระจายตัวของจำนวนเงิน .....	21
3.9 จำนวนของสกุลเงินในการชำระเงิน โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน .....	22
3.10 จำนวนของสกุลเงินของฝั่งปลายทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน ....	22
3.11 จำนวนที่ตั้งของธนาคารบัญชีต้นทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน ....	23
3.12 จำนวนที่ตั้งของธนาคารบัญชีปลายทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน	24
3.13 ประเภทของการชำระเงิน โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน .....	24
3.14 ประเภทของธุรกรรม โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน .....	25
3.15 ตารางเมทริกซ์สหสัมพันธ์ .....	26
3.16 ความไม่สมดุลของข้อมูล .....	27
3.17 ความสมดุลของข้อมูล หลังจากทำการ Undersampling .....	27
3.18 ความสมดุลของข้อมูล หลังจากทำการ Oversampling.....	28
4.1 ตาราง Confusion matrix ของแบบจำลอง XGBoost จากการทำ Undersampling .....	35
4.2 Feature Importance ของแบบจำลอง XGBoost จากการทำ Undersampling .....	37
4.3 ตาราง Confusion matrix ของแบบจำลอง XGBoost จากการทำ Oversampling.....	39
4.4 Feature Importance ของแบบจำลอง XGBoost จากการทำ Oversampling.....	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 ตาราง Confusion matrix ของแบบจำลอง LightGBM จากการทำ Undersampling.....	45
4.6 Feature Importance ของแบบจำลอง LightGBM จากการทำ Undersampling.....	47
4.7 ตาราง Confusion matrix ของแบบจำลอง LightGBM จากการทำ Oversampling .....	50
4.8 Feature Importance ของแบบจำลอง LightGBM จากการทำ Oversampling .....	52



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน การฟอกเงินเป็นปัญหาที่สำคัญที่เกิดขึ้นในระบบการเงินทั่วโลก การฟอกเงินนั้น เป็นกระบวนการที่มีวัตถุประสงค์ในการปกปิดแหล่งที่มาของรายได้จากกิจกรรมที่ผิดกฎหมาย โดยผ่านทางในการทำธุรกรรมทางการเงินที่มีการทำให้แหล่งที่มาของการเงินนั้นดูเหมือนถูกกฎหมายและมีการตรวจสอบอย่างถูกต้อง แต่ความจริงเป็นเพียงการสร้างธุรกรรมทางการเงินที่มีความซับซ้อน เพื่อให้แหล่งที่มาของเงินนั้นดูเป็นปกติ ซึ่งทำให้ผู้ที่ฟอกเงินนั้นสามารถนำเงินมาใช้ได้อย่างสบายใจ และเปิดเผย

ในประเทศไทย ปัญหาการฟอกเงินก็เป็นหนึ่งในปัญหาที่ต้องระวังอย่างมาก โดยประเทศไทย นั้นมีสำนักงานป้องกันและปราบปรามการฟอกเงิน (ปปง.) ซึ่งมีอำนาจหน้าที่ในการวางหลักเกณฑ์ และดูแลให้มีการปฏิบัติตามกฎหมายว่าด้วยการป้องกันและปราบปรามการฟอกเงิน รวมถึงทำหน้าที่ เป็นหน่วยงานกำกับดูแลการทำธุรกรรมทางการเงิน โดยมีการร่วมมือกับสถาบันการเงินต่างๆ ในการ สืบสวนและตรวจสอบการฟอกเงิน ซึ่งเป็นหน้าที่ของสถาบันการเงินที่จะต้องนำส่งบุคคล และธุรกรรม ที่เป็นที่ต้องสงสัยให้กับทาง ปปง.

โดยการตรวจสอบการฟอกเงินของบางสถาบันการเงินในประเทศไทยนั้น จะมี Workflow Scenario เป็นสถานการณ์ที่กำหนดไว้อย่างชัดเจนในการตรวจจับธุรกรรมต้องสงสัย เพื่อทำการแจ้ง เตือน (Alerts) ไปยังเจ้าหน้าที่ปฏิบัติตามข้อกำหนด (Compliance officer) ให้ทำการตรวจสอบ ธุรกรรมต้องสงสัยนั้น และทำการออกรายงานกิจกรรมที่น่าสงสัย (SARs) เพื่อนำส่งให้กับ ปปง. ไป ดำเนินการตรวจสอบต่อไป

ซึ่งจุดประสงค์ของการวิจัยในครั้งนี้ ผู้วิจัยได้เห็นปัญหาถึงการฟอกเงินในปัจจุบันที่ได้ถูก พัฒนาให้มีธุรกรรมทางการเงินที่ซับซ้อนมากขึ้น ทำให้เจ้าหน้าที่ปฏิบัติตามข้อกำหนดต้องเร่ง ตรวจสอบธุรกรรมที่มีการแจ้งเตือนในแต่ละวัน อีกทั้งการตรวจสอบทุกๆการแจ้งเตือนจะมีการใช้ เจ้าหน้าที่มากกว่า 1 คนในการตรวจสอบ เพื่อทำการยืนยันว่าธุรกรรมนั้นเป็นการฟอกเงินหรือไม่ ด้วยเหตุนี้ ผู้วิจัยจึงสร้างแบบจำลองโดยใช้การเรียนรู้ของเครื่องมาทำการตรวจจับธุรกรรมการฟอก เงิน โดยมุ่งเน้นไปที่ประสิทธิภาพในการตรวจสอบธุรกรรม และเพื่อช่วยลดการใช้ทรัพยากรมนุษย์ใน การตรวจสอบการฟอกเงิน

และเนื่องจากการวิจัยในครั้งนี้ จำเป็นต้องใช้ข้อมูลที่เป็นธุรกรรมการเงินของธนาคาร เพื่อมา ช่วยพัฒนาแบบจำลองให้มีประสิทธิภาพมากที่สุด แต่ข้อมูลทางการเงินนั้นเป็นข้อมูลที่ไม่สามารถ นำมาเปิดเผยได้เนื่องด้วยข้อจำกัดทางกฎหมาย ผู้วิจัยจึงเลือกใช้ข้อมูลจากเว็บไซต์ Kaggle ซึ่งเป็น

ข้อมูลธุรกรรมที่ถูกสังเคราะห์ขึ้นมาจากธนาคารในประเทศอังกฤษ และมีประเภทของธุรกรรมที่ระบุ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยผู้เชี่ยวชาญเรื่องการฟอกเงิน อย่างไรก็ตามการเลือกใช้ข้อมูลสังเคราะห์จากเว็บไซต์ Kaggle มาพัฒนาสร้างแบบจำลอง อาจทำให้แบบจำลองนี้ยังมีข้อจำกัดที่เกิดจากการไม่ได้ใช้ข้อมูลจริงของธนาคารในประเทศไทย ซึ่งทำให้แบบจำลองยังไม่สามารถนำไปใช้ได้จริงในธนาคาร โดยผู้วิจัยมีวัตถุประสงค์เพื่อใช้ชุดข้อมูลนี้ ไว้เป็นแนวทางในการทดลองพัฒนาแบบจำลองที่ใช้ตรวจจับธุรกรรม การฟอกเงินเท่านั้น

โดยมีผู้ทดลองท่านหนึ่ง Berkan Oztas (2024) นำชุดข้อมูลสังเคราะห์นี้ไปพัฒนาแบบจำลองด้วยวิธี XGBoost ผลการทดลองพบว่าแบบจำลอง XGBoost ให้ค่าความแม่นยำที่ 81.205% ซึ่งถือว่ามีความแม่นยำสูง แต่ในขณะเดียวกันผลการทำนายนั้นให้อัตราผลบวกหลงหรือ False Positive Rate สูงถึง 58% บ่งบอกว่าการทดลองนี้ยังคงไม่สามารถแยกธุรกรรมปกติกับธุรกรรม การฟอกเงินได้ดีเท่าที่ควร ดังนั้นผู้วิจัยจึงอยากพัฒนาแบบจำลองเพื่อตรวจจับธุรกรรม การฟอกเงินให้มีประสิทธิภาพในการตรวจจับธุรกรรมมากขึ้น และพัฒนาแบบจำลองที่เกิดผลบวกหลง น้อยที่สุด เพื่อช่วยลดต้นทุนที่ต้องใช้ในการตรวจสอบการฟอกเงิน

## 1.2 วัตถุประสงค์ของงานวิจัย

- 1.2.1 เพื่อสร้างแบบจำลองการเรียนรู้ของเครื่อง ที่สามารถใช้ตรวจจับการฟอกเงินได้อย่างมีประสิทธิภาพ
- 1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลองการเรียนรู้ของเครื่อง ในการตรวจจับ การฟอกเงิน
- 1.2.3 ลดการใช้ทรัพยากรมนุษย์ในการตรวจสอบธุรกรรม การฟอกเงิน

## 1.3 ขอบเขตของงานวิจัย

การวิจัยในครั้งนี้ได้นำข้อมูลจากเว็บไซต์ Kaggle มาทำการศึกษาเพื่อใช้ในการสร้างแบบจำลอง การเรียนรู้ของเครื่อง ซึ่งเป็นชุดข้อมูล SAML-D ที่ถูกสังเคราะห์ขึ้นมาจากธุรกรรมของธนาคารใน ประเทศอังกฤษ โดยมีเป้าหมายเพื่อช่วยให้นักวิจัยสามารถนำข้อมูลสังเคราะห์นี้ไปพัฒนาแบบจำลอง และวิธีการตรวจสอบธุรกรรมให้มีประสิทธิภาพมากขึ้น โดยชุดข้อมูลนี้ประกอบด้วยธุรกรรมทั้งหมด จำนวน 9,504,852 รายการ โดยมีคุณลักษณะทั้ง 12 รายการดังนี้

- 1) เวลาที่ทำธุรกรรม (Time)
- 2) วันที่ที่ทำธุรกรรม (Date)
- 3) บัญชีต้นทาง (Sender\_account)
- 4) บัญชีปลายทาง (Receiver\_account)
- 5) จำนวนเงิน (Amount)
- 6) สกุลเงินในการชำระเงิน (Payment\_currency)
- 7) สกุลเงินของฝั่งปลายทาง (Received\_currency)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 8) ที่ตั้งธนาคารบัญชีต้นทาง (Sender\_bank\_location)
- 9) ที่ตั้งธนาคารบัญชีปลายทาง (Receiver\_bank\_location)
- 10) ประเภทของการชำระเงิน (Payment\_type)
- 11) ป้ายกำกับที่บ่งชี้ธุรกรรมฟอกเงิน (Is\_laundering)
- 12) ประเภทของธุรกรรม (Laundering\_type)

#### 1.4 วิธีการดำเนินงานวิจัย

ในการสร้างแบบจำลองการเรียนรู้ของเครื่องเพื่อตรวจจับธุรกรรมฟอกเงิน ผู้ศึกษาได้วางแผนในการดำเนินการตามขั้นตอนการศึกษา เพื่อให้ตรงตามวัตถุประสงค์ดังต่อไปนี้

- 1.4.1 กำหนดหัวข้อสำหรับการวิจัยและวางแผนการดำเนินงานรวมถึงเขียนเค้าโครงของการดำเนินงาน
- 1.4.2 ศึกษาวิธีการตรวจจับธุรกรรมต้องสงสัยแบบดั้งเดิมและแบบการเรียนรู้ของเครื่อง
- 1.4.3 รวบรวมข้อมูล (Data collection) จากเว็บไซต์ Kaggle
- 1.4.4 ดำเนินการสำรวจข้อมูล (Data Exploratory) และเตรียมข้อมูล (Data Preparation) ก่อนนำเข้าแบบจำลอง
- 1.4.5 นำชุดข้อมูลธุรกรรมสังเคราะห์มาเข้ากระบวนการสร้างแบบจำลองการเรียนรู้ของเครื่อง
- 1.4.6 ประเมินและเปรียบเทียบแบบจำลองที่สามารถตรวจจับธุรกรรมฟอกเงินได้อย่างมีประสิทธิภาพ
- 1.4.7 สรุปผลตามวัตถุประสงค์ของการศึกษาและให้ข้อเสนอแนะในการวิจัยครั้งนี้

#### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 สามารถสร้างแบบจำลองการเรียนรู้ของเครื่องที่ใช้ตรวจจับการฟอกเงินได้อย่างมีประสิทธิภาพ
- 1.5.2 ทราบถึงความแตกต่างทางด้านประสิทธิภาพของแต่ละแบบจำลองการเรียนรู้ของเครื่องในการตรวจจับการฟอกเงิน
- 1.5.3 สามารถใช้แบบจำลองการเรียนรู้ของเครื่อง เพื่อลดการใช้ทรัพยากรมนุษย์ในการตรวจสอบธุรกรรมฟอกเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการค้นคว้าอิสระนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการเรียนรู้ของเครื่องสำหรับใช้ตรวจจับธุรกรรมการฟอกเงิน โดยผู้วิจัยได้ทำการค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้อง ทั้งในเรื่องของการฟอกเงิน ความผิดฐานฟอกเงิน Workflow Scenario การเรียนรู้ของเครื่อง และเทคนิคต่างๆ ในการสร้างแบบจำลองให้มีประสิทธิภาพ โดยจะนำเสนอหัวข้อตามลำดับต่อไปนี้

### 2.1 แนวคิดและทฤษฎีที่เกี่ยวข้อง

#### 2.1.1 การฟอกเงิน (Money Laundering)

การฟอกเงิน (Money Laundering) คือ การเปลี่ยนแปลงเงินหรือทรัพย์สินที่ได้มาโดยไม่ชอบด้วยกฎหมาย หรือไม่สุจริตให้กลายเป็นเงิน หรือทรัพย์สินที่ดูเหมือนว่าได้มาโดยชอบด้วยกฎหมาย หรือพิสูจน์ไม่ได้ว่าได้มาจากการกระทำความผิด ซึ่งสามารถทำได้หลายวิธี ยกตัวอย่างเช่น การนำเงินออกไปไว้นอกประเทศ การฝากเงินไว้กับสถาบันการเงิน การจัดตั้งบริษัทหรือกิจการขึ้นมาบังหน้า การซื้อขายที่ดินต่อกัน การแลกเปลี่ยนเงินไปเป็นเงินตราสกุลอื่น การให้ผู้อื่นถือเงิน หรือทรัพย์สินไว้แทน โดยสาเหตุที่ต้องมีการฟอกเงินนั้นเนื่องมาจาก เงินที่ได้มาจากรุทกกิจผิดกฎหมายและไม่สามารถอ้างอิงแหล่งที่มาได้ ซึ่งแหล่งที่มาของเงินนั้นมักจะมาจากแหล่งใหญ่ๆ ดังนี้ (สำนักงานป้องกันและปราบปรามการฟอกเงิน, 2559)

1. เงินจากการค้ายาเสพติด
2. เงินจากวงการเมือง
3. เงินที่ได้จากการฉ้อโกงประชาชน
4. เงินจากวงการพนัน
5. เงินสินบน

และเหตุผลของการฟอกเงิน มีจุดประสงค์ที่แตกต่างกันคือ (กรมสอบสวนคดีพิเศษ, 2567)

1. เพื่อปกปิดแหล่งที่มาของเงิน
2. เพื่อซุกซ่อนทรัพย์สิน
3. เพื่อปกปิดความเป็นเจ้าของที่แท้จริง
4. เพื่อปกปิดแหล่งที่มาที่ผิดกฎหมาย
5. เพื่อให้ดูเหมือนว่าเป็นเงิน หรือทรัพย์สินที่ถูกต้องตามกฎหมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.1.2 ขั้นตอนในการฟอกเงิน

แบ่งได้เป็น 3 ขั้นตอน คือ

### 1. Placement หรือ Conversion

หมายถึงการนำเงินที่ได้มาจากการกระทำที่ผิดกฎหมายเข้าไปอยู่ในระบบการเงินที่ถูกต้องตามกฎหมาย โดยอาจใช้วิธีการต่างๆ เช่น การนำเงินไปฝากธนาคาร หรือการนำไปซื้อทรัพย์สินอื่นๆ

### 2. Layering หรือ Concealment

หมายถึงการจัดทำรายการของเงินดังกล่าวเป็นหลายๆ ชั้น เพื่อปกปิดที่มาหรือร่องรอยที่แท้จริง ทำให้ธุรกรรมนั้นมีความสลับซับซ้อนและยากแก่การตรวจสอบติดตาม เช่น การฝากและถอนเงินระหว่างบัญชี หรือระหว่างธนาคารหลายๆ ครั้ง โดยใช้บัญชีม้าและชื่อปลอม การฝากถอนโดยการโอนเงินระหว่างประเทศซึ่งจะมีหลักการคือ เลือกประเทศที่เคร่งครัดในเรื่องของการปกปิดความลับของลูกค้า หรือโดยผ่านระบบธนาคารใต้ดิน การซื้อและขายที่ดินต่อกัน การซื้อขายทรัพย์สินที่ไม่มีหลักฐานทางทะเบียน อย่างเช่น ทองคำ เป็นต้น

### 3. Integration หรือ Creation of perception of legitimacy

หมายถึงการนำเงินที่ต้องการฟอก มาผสมผสานกับเงินที่ได้จากกิจการถูกกฎหมาย ซึ่งทำให้เงินที่ได้มาอย่างผิดกฎหมายนั้นดูเสมือนว่าเป็นเงินที่ได้รับจากการกระทำที่ถูกกฎหมาย ทำให้ผู้ฟอกเงินสามารถชี้แจงเรื่องและแหล่งที่มาของเงินดังกล่าวได้ เช่นการจัดตั้งบริษัทหรือกิจการไว้ข้างหน้าเพื่อใช้เป็นแหล่งผสมผสานเงิน การนำเงินไปซื้อกิจการที่มีรายรับเป็นเงินสด ยกตัวอย่างเช่น กิจการโรงแรม กิจการภัตตาคาร กิจการร้านค้าปลีก เป็นต้น

## 2.1.3 องค์ประกอบความผิดฐานฟอกเงิน

โดยความผิดตามกฎหมายจากการฟอกเงิน ได้กำหนดฐานความผิดไว้ ดังนี้

1. ความผิดเกี่ยวกับยาเสพติดติดตามกฎหมายว่าด้วยการป้องกันและปราบปรามยาเสพติด
2. ความผิดเกี่ยวกับเพศตามประมวลกฎหมายอาญา
3. ความผิดเกี่ยวกับการฉ้อโกงประชาชนตามประมวลกฎหมายอาญา หรือความผิดตามกฎหมายว่าด้วยการกักขังเงินที่เป็นการฉ้อโกงประชาชน
4. ความผิดเกี่ยวกับการยกยอก หรือฉ้อโกง หรือประทุษร้ายต่อทรัพย์สิน หรือกระทำโดยทุจริตตามกฎหมายว่าด้วยการธนาคารพาณิชย์
5. ความผิดต่อตำแหน่งหน้าที่ราชการ หรือความผิดต่อตำแหน่งหน้าที่ในการยุติธรรมตามประมวลกฎหมายอาญา
6. ความผิดเกี่ยวกับการกรรโชก หรือรีดเอาทรัพย์สินที่กระทำโดยอั้งอ้านาจอั้งยีหรือช่องโจรตามประมวลกฎหมายอาญา
7. ความผิดเกี่ยวกับการลักลอบหนีศุลกากรตามกฎหมายว่าด้วยศุลกากร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. ความผิดเกี่ยวกับการก่อการร้ายตามประมวลกฎหมายอาญา
9. ความผิดเกี่ยวกับการพนันตามกฎหมายว่าด้วยการพนัน เฉพาะความผิดเกี่ยวกับการเป็นผู้จัดให้มีการเล่นการพนันโดยไม่ได้รับอนุญาต

#### 2.1.4 Workflow Scenario

หมายถึงแผนการทำงาน หรือลำดับขั้นตอนที่เกี่ยวข้องกับการดำเนินงานหนึ่งๆ ที่สร้างขึ้น เพื่อให้การทำงานเป็นไปตามลำดับ โดยรวมถึงการแสดงผลแผนผังหรือขั้นตอนการทำงานที่ต้องทำ เพื่อให้ภาพรวมเกี่ยวกับการทำงานเป็นไปอย่างเรียบร้อยและชัดเจน การสร้าง Workflow Scenario มักเป็นกระบวนการที่ใช้ในการวางแผนและจัดการกับการดำเนินงานในองค์กรหรือธุรกิจ เพื่อให้ทราบถึงลำดับของกิจกรรม หน้าที่ และการตัดสินใจที่ต้องดำเนินการตามเวลา โดยมุ่งเน้นให้กระบวนการทำงานเป็นไปอย่างเป็นระเบียบและมีประสิทธิภาพ

ในการตรวจจับธุรกรรมการฟอกเงินของธนาคารนั้น Workflow Scenario จะเป็นสถานการณ์ต่างๆ ที่ถูกสร้างขึ้นมาเพื่อตรวจสอบกิจกรรมและเส้นทางการเงินของลูกค้า หากธุรกรรมใดที่เข้าเงื่อนไขของสถานการณ์ที่สงสัยว่าเป็นการฟอกเงิน ระบบของธนาคารจะทำการแจ้งเตือนสถานการณ์และลูกค้ารายนั้น ไปยังเจ้าหน้าที่กำกับดูแลการปฏิบัติตามกฎระเบียบของธนาคาร

#### 2.1.5 การเรียนรู้ของเครื่อง (Machine Learning)

หมายถึงกระบวนการเรียนรู้ที่พยายามทำความเข้าใจโครงสร้างของข้อมูล เพื่อมาเปรียบเทียบเข้ากับข้อมูลที่ผ่านมาทำความเข้าใจอย่างดีแล้ว โดยกระบวนการนั้นสามารถที่จะเรียนรู้และมีปฏิสัมพันธ์กับชุดข้อมูลต่างๆ รวมถึงสามารถระบุและทราบรูปแบบต่างๆ ที่เกิดขึ้น และนำไปสู่การทำนาย หรือตัดสินใจได้เองโดยไม่จำเป็นต้องพึ่งพาความสามารถของมนุษย์ โดยการเรียนรู้ของเครื่อง แบ่งเป็นประเภทต่างๆ ได้ดังนี้ (บริษัท SAS Software จำกัด, 2567)

##### 1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)

คือการเรียนรู้ด้วยการใช้ข้อมูลแบบมีผลเฉลย หรือป้ายกำกับ (Label) โดยแบบจำลองจะได้รับข้อมูลขาเข้าจำนวนหนึ่ง พร้อมกับผลลัพธ์ขาออกที่ถูกต้อง และระบบจะทำการเรียนรู้ด้วยการเปรียบเทียบผลลัพธ์ที่ทำนายออกมาได้ เทียบกับผลเฉลยที่มีอยู่ การเรียนรู้แบบมีผู้สอนนั้นแบ่งได้ 2 วิธี คือ Regression ใช้กับผลเฉลยที่เป็นค่าต่อเนื่อง (Continuous values) เช่น ราคาบ้าน อุณหภูมิ หรือระยะเวลาเดินทาง คุณสมบัติหลักของการใช้ Regression คือการหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม และอีกวิธีคือ Classification ใช้กับผลเฉลยที่เป็นค่าไม่ต่อเนื่อง คือข้อมูลที่เป็นลักษณะกลุ่ม หรือหมวดหมู่ (Categorical variable) ซึ่งค่าที่ได้จะเป็นจำนวนจำกัด เช่น ชนิดของสัตว์ การตัดสินใจอนุมัติสินเชื่อ หรือการวินิจฉัยโรค เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

คือการเรียนรู้กับข้อมูลที่ไม่มีผลเฉลย หรือไม่มีป้ายกำกับ แบบจำลองจะต้องทำการเรียนรู้หาคำตอบ โครงสร้าง และรูปแบบภายในข้อมูลด้วยตัวเอง โดยไม่มีการแทรกแซงหรือแนะนำโดยมนุษย์ ส่วนใหญ่จะใช้ในการทำ Clustering จัดกลุ่มข้อมูลที่มีลักษณะคล้ายกันเข้าด้วยกัน หรือ Association Rule มักจะใช้ในการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อค้นหาว่าสินค้าใดที่มักจะถูกซื้อพร้อมกันบ่อยๆ

## 3. การเรียนรู้แบบกึ่งมีผู้สอน (Semi-Supervised Learning)

คือการเรียนรู้ที่ใช้ในสถานการณ์เดียวกันกับ Supervised Learning แต่ความแตกต่างคือ มีการใช้ข้อมูลทั้งแบบมีผลเฉลยและไม่มีผลเฉลย โดยปกติจะใช้กับข้อมูลที่มีผลเฉลยจำนวนน้อย

## 4. Reinforcement Learning

คือการเรียนรู้ที่ใช้การลองผิดลองถูกและเรียนรู้ว่าเส้นทางการทำงานแบบใดที่ให้ผลตอบแทนที่ดีที่สุด การเรียนรู้ประเภทนี้จะแบ่งองค์ประกอบออกเป็นสามส่วน คือ agent, environment และ actions เป้าหมายคือการให้ agent เลือกหนทางการปฏิบัติที่ให้รางวัลหรือผลตอบแทนสูงสุด

### 2.1.5.1 เอ็กซ์ตรีมกราเดียนต์บูสตีง (XGBoost)

Extreme Gradient Boosting เป็นหนึ่งในแบบจำลองประเภท Ensemble learning ที่นำเอา Decision Tree มาฝึกต่อๆ กันหลายๆ tree โดยที่แต่ละ Decision Tree จะเรียนรู้จากข้อผิดพลาดของ Tree ก่อนหน้า ทำให้ความแม่นยำในการทำนายจะมากขึ้นเรื่อยๆ เมื่อมีการเรียนรู้ของ Tree ต่อเนื่องกันจนมีความลึกมากพอและแบบจำลองจะหยุดเรียนรู้เมื่อไม่เหลือรูปแบบของข้อผิดพลาดจาก Tree ก่อนหน้าให้เรียนรู้แล้ว (Daroontham, W., 2018)

ซึ่งแบบจำลอง XGBoost สามารถปรับแต่งพารามิเตอร์เพื่อปรับปรุงประสิทธิภาพของแบบจำลองให้เหมาะสมกับข้อมูลที่ใช้ โดยสามารถปรับค่าต่างๆ ได้ดังนี้ (Jain, A., 2024)

- n\_estimators: จำนวน tree ที่จะสร้าง
- learning\_rate (eta): อัตราการเรียนรู้ ช่วยในการป้องกัน Overfitting โดยการชะลออัตราการอัปเดตของน้ำหนักในแต่ละรอบ
- min\_child\_weight: น้ำหนักขั้นต่ำที่จำเป็นสำหรับการสร้างโหนดย่อยใหม่ใน tree
- max\_depth: ความลึกสูงสุดของแต่ละ tree
- gamma: ค่าการตัดสินใจในการแยก tree เพิ่มเติม ใช้เพื่อควบคุม Overfitting
- subsample: สัดส่วนของตัวอย่างที่ใช้สำหรับแต่ละ tree การลดค่านี้อาจช่วยป้องกัน Overfitting
- colsample\_bytree: สัดส่วนของคุณลักษณะที่ใช้สำหรับแต่ละ tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- `reg_alpha` (alpha): L1 regularization term ใช้เพื่อป้องกันการเกิด Overfitting
- `reg_lambda` (lambda): L2 regularization term ช่วยเสริมความแข็งแกร่งให้กับแบบจำลอง
- `scale_pos_weight`: สำหรับคลาสที่ไม่สมดุล ช่วยปรับสมดุลคลาสบวกและลบ
- `early_stopping_rounds`: หยุดการเรียนรู้ ถ้าไม่มีการปรับปรุงคะแนนการตรวจสอบในจำนวนรอบที่ระบุ
- `eval_metric`: เมตริกสำหรับการประเมินแบบจำลองในชุดการตรวจสอบ
- `seed`: ใช้สำหรับการสร้างตัวอย่างสุ่ม ช่วยให้การทดลองที่สามารถทำซ้ำได้

### 2.1.5.2 ไลต์กราเดียนต์บูสตีงแมชชีน (LightGBM)

Light Gradient Boosting Machine คือเฟรมเวิร์กการเรียนรู้ของเครื่อง (Machine Learning) ที่พัฒนาโดย Microsoft หลักการทำงานของ LightGBM จะใช้เทคนิค Gradient Boosting Decision Tree (GBDT) ซึ่งเป็นวิธีการที่สร้างชุดของต้นไม้การตัดสินใจ (decision trees) โดยการฝึกแต่ละ tree เพื่อแก้ไขข้อผิดพลาดของ tree ก่อนหน้า โดย LightGBM ออกแบบมาเพื่อให้สามารถจัดการกับข้อมูลขนาดใหญ่ และมีความเร็วในการฝึกสูงกว่า Gradient Boosting Machine อื่นๆ มีการใช้หน่วยความจำน้อยกว่า เนื่องจากการจัดเก็บข้อมูลในรูปแบบ histogram-based algorithm สามารถจัดการกับข้อมูลที่มีมิติสูง (high-dimensional data) และสามารถใช้กับข้อมูลทั้งเชิงตัวเลขและเชิงหมวดหมู่ได้

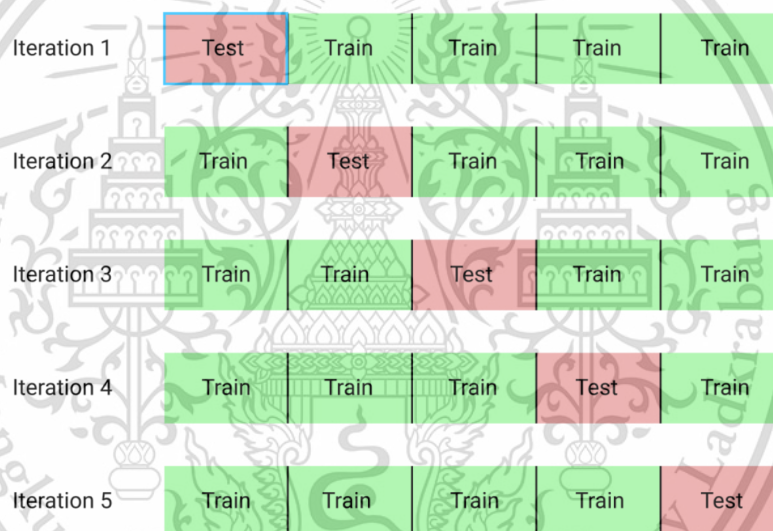
ซึ่งแบบจำลอง LightGBM สามารถปรับแต่งพารามิเตอร์เพื่อปรับปรุงประสิทธิภาพของแบบจำลองให้เหมาะสมกับข้อมูลที่ใช้ โดยสามารถปรับค่าต่างๆ ได้ดังนี้ (Singh, S., 2024)

- `boosting_type`: ชนิดของอัลกอริทึมที่ใช้สำหรับการฝึกแบบจำลอง เช่น `gbdt` (Gradient Boosting Decision Tree), `rf` (Random Forest), `dart` (Dropouts meet Multiple Additive Regression Trees) หรือ `goss` (Gradient-based One-Side Sampling)
- `num_leaves`: จำนวนใบสูงสุดบน tree
- `max_depth`: ความลึกสูงสุดของแต่ละ tree
- `learning_rate`: อัตราการเรียนรู้ ช่วยในการป้องกัน Overfitting
- `n_estimators`: จำนวน tree ที่จะสร้าง
- `max_bin`: จำนวนช่วงสูงสุดสำหรับการจัดเก็บข้อมูล
- `min_data_in_leaf`: จำนวนข้อมูลขั้นต่ำในใบไม้หนึ่งใบ
- `lambda_l1`: L1 regularization term
- `lambda_l2`: L2 regularization term

### 2.1.6 การทำ Cross validation

Cross validation เป็นเทคนิคในการทดสอบความแม่นยำของแบบจำลองการเรียนรู้ของเครื่อง โดยเป็นวิธีที่ช่วยให้สามารถประเมินประสิทธิภาพของแบบจำลองได้อย่างเป็นอิสระจากข้อมูลที่ใช้ฝึก ซึ่งเป็นวิธีที่ช่วยลดปัญหาการ Overfitting และให้การประเมินผลลัพธ์ที่มีความเชื่อถือได้มากขึ้นเมื่อแบบจำลองถูกนำไปใช้กับข้อมูลใหม่ที่ไม่ได้มีส่วนในการฝึกฝนแบบจำลอง

โดยหลักการคือการแบ่งชุดข้อมูลออกเป็นชุดย่อยๆ จำนวน  $k$  ชุด และในการทำงานจะทำงานทั้งหมด  $k$  รอบ โดยในแต่ละรอบ ชุดข้อมูลทดสอบจะมีการสับเปลี่ยนกัน ทำให้เมื่อทำการเรียนรู้ครบทั้ง  $k$  รอบ ชุดข้อมูลทุกส่วนจะมีโอกาสเป็นทั้งชุดข้อมูลสำหรับฝึกแบบจำลอง และสำหรับทดสอบ วิธีการนี้สามารถใช้เพื่อหลีกเลี่ยงสถานการณ์การ Overfitting ของแบบจำลอง ต่อข้อมูลส่วนใดส่วนหนึ่งได้ (AlgoAddict, 2021)



ภาพที่ 2.1 วิธีการแบ่งชุดข้อมูลในการทำ Cross validation  
(ที่มา <https://algoaddict.wordpress.com/2021/10/01/>)

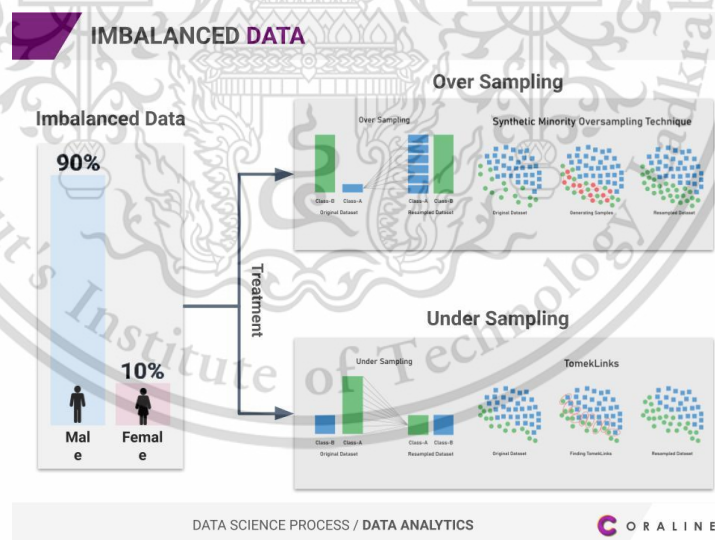
### 2.1.7 ข้อมูลไม่สมดุล (Imbalanced data)

ข้อมูลไม่สมดุล หมายถึง ข้อมูลมีการกระจายที่ไม่สมดุลระหว่างกลุ่มหรือคลาสต่างๆ โดยมีจำนวนข้อมูลในแต่ละคลาสที่มีความแตกต่างกันอย่างมาก ในการแบ่งประเภท (classification) ข้อมูลไม่สมดุลสามารถเกิดขึ้นเมื่อมีคลาสหนึ่งมีจำนวนตัวอย่างมาก ในขณะที่คลาสอื่นๆ มีจำนวนตัวอย่างน้อยมาก ปัญหาของข้อมูลไม่สมดุลสามารถทำให้การสร้างแบบจำลองสำหรับการทำนายหรือการจำแนกประเภท มีความยากลำบาก เนื่องจากแบบจำลองมักจะมีแนวโน้มที่จะเรียนรู้และทำนายคลาสที่มีจำนวนตัวอย่างมากกว่า ซึ่งอาจทำให้มีปัญหาในการจำแนกคลาสที่มีจำนวนน้อย

มาก ทำให้คลาสที่เป็นส่วนน้อยอาจไม่ถูกจำแนกอย่างถูกต้องได้ โดยการจัดการกับข้อมูลไม่สมดุลสามารถทำได้โดยวิธีการต่างๆ

1. การลดข้อมูล (Undersampling) คือการลดจำนวนตัวอย่างของคลาสที่มีมากในชุดข้อมูล เพื่อให้มีความสมดุลกับคลาสที่มีน้อยกว่า โดยมีข้อดีคือลดเวลาการฝึกฝนและขนาดข้อมูล แต่ข้อเสียคืออาจสูญเสียข้อมูลสำคัญและอาจส่งผลกระทบต่อประสิทธิภาพของแบบจำลอง
2. การเพิ่มข้อมูล (Oversampling) คือเพิ่มจำนวนตัวอย่างของคลาสที่น้อยกว่า ให้มากขึ้น โดยการสร้างข้อมูลใหม่จากข้อมูลที่มีอยู่ ซึ่งข้อดีคือเพิ่มความสมดุลของชุดข้อมูลและช่วยให้แบบจำลองเรียนรู้คุณลักษณะจากคลาสน้อยได้ดีขึ้น แต่ข้อเสียคือเพิ่มความเสี่ยงต่อการเกิด Overfitting ต่อคลาสน้อย โดยเทคนิคที่นิยมในการเพิ่มข้อมูล ได้แก่

2.1 SMOTE (Synthetic Minority Over-sampling Technique) เป็นเทคนิคที่ใช้ในการจัดการกับข้อมูลไม่สมดุลในการเรียนรู้ของเครื่อง โดยเฉพาะในการจำแนกประเภทเพื่อเพิ่มจำนวนตัวอย่างของคลาสที่น้อยกว่า (minority class) โดยสร้างตัวอย่างใหม่จากตัวอย่างที่มีอยู่และทำการเพิ่มเข้าไปในชุดข้อมูล โดยวิธีการสร้างตัวอย่างใหม่นี้จะใช้การสร้างตัวอย่างแบบเชิงประสิทธิภาพเพื่อสร้างตัวอย่างที่เป็นสมมูลฐานและไม่เกิดการซ้ำซ้อนจากตัวอย่างที่มีอยู่ แต่อาจเกิดการสร้างตัวอย่างที่ไม่สมจริงได้เช่นกัน



ภาพที่ 2.2 วิธีในการจัดการกับปัญหาข้อมูลไม่สมดุล

(ที่มา Coraline 2021)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.1.8 ขั้นตอนการสร้างแบบจำลอง

1. Data Collection การรวบรวมข้อมูล
2. Data Preparation การเตรียมข้อมูล รวมถึงการทำความสะอาดข้อมูล (Data Cleaning) และการแปลงข้อมูล (Data Transformation) เพื่อให้อยู่ในรูปแบบที่เหมาะสมสำหรับการใช้งานในแบบจำลอง
3. Splitting data การแบ่งข้อมูลออกเป็นชุดข้อมูล Training สำหรับฝึกในแบบจำลอง และชุดข้อมูล Test ใช้สำหรับทดสอบในแบบจำลอง
4. Model Selection การเลือกแบบจำลอง โดยการเลือกนั้นจะต้องเหมาะสมกับชุดข้อมูลที่นำเข้าไปแบบจำลอง
5. Model Optimization การปรับปรุงแบบจำลอง จะใช้เทคนิคการปรับพารามิเตอร์ (parameter tuning) และการใช้เทคนิคการจัดการกับข้อมูลไม่สมดุล เพื่อเพิ่มประสิทธิภาพของแบบจำลอง
6. Model Evaluation การประเมินประสิทธิภาพของแบบจำลอง เป็นขั้นตอนสำคัญที่ใช้ในการวัดความสามารถในการทำนายหรือจำแนกประเภทของข้อมูล

### 2.1.9 การปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning)

การปรับแต่งไฮเปอร์พารามิเตอร์ คือการตั้งค่าคอนฟิกในการฝึกแบบจำลอง เป็นกระบวนการที่ใช้หาค่าที่เหมาะสมที่สุดสำหรับพารามิเตอร์ของแบบจำลองการเรียนรู้ของเครื่องที่ไม่ได้เรียนรู้โดยตรงจากข้อมูลการฝึก แต่เป็นค่าที่ถูกกำหนดก่อนการฝึก โดยพารามิเตอร์เหล่านี้มีผลอย่างมากต่อประสิทธิภาพของแบบจำลองและการปรับแต่งให้เหมาะสมสามารถช่วยเพิ่มความแม่นยำและประสิทธิภาพในการทำงานได้อย่างมีนัยสำคัญ โดยเทคนิคในการปรับแต่งไฮเปอร์พารามิเตอร์แบบอัตโนมัติ มีดังนี้ (ทินกร ม้าลายทอง และปพนธ์ ธรรมเจริญพร, 2564)

#### 1. Grid Search

หรือการค้นหาแบบกริด เป็นเทคนิคที่ใช้ในการหาค่า Hyperparameter ที่เข้าใจง่ายและตรงไปตรงมา ด้วยการวนใช้พารามิเตอร์ที่กำหนดไว้ทุกชุด และประเมินประสิทธิภาพหรือความแม่นยำของแบบจำลองแต่ละชุด จะเป็นการลองสร้างแบบจำลองจากค่าของ Hyperparameter ทุกชุด รูปแบบของการทำงานจะคล้ายกริด โดยค่าทั้งหมดจะอยู่ในรูปของเมทริกซ์ (Matrix) พารามิเตอร์แต่ละชุดจะถูกนำมาพิจารณาและวัดค่าความถูกต้อง เมื่อชุดของ Hyperparameter ทั้งหมดได้รับการประเมินแล้ว แบบจำลองที่มีชุดพารามิเตอร์ที่ให้ความแม่นยำสูงสุดจะถือว่าเป็นชุดที่ดีที่สุด

#### 2. Random Search

วิธีการทำงานของ Random Search คล้ายกันกับการทำ Grid Search แต่แทนที่จะลองใช้พารามิเตอร์ที่กำหนดไว้ล่วงหน้าในกริดทุกชุด Random Search จะทำการสุ่มเลือก

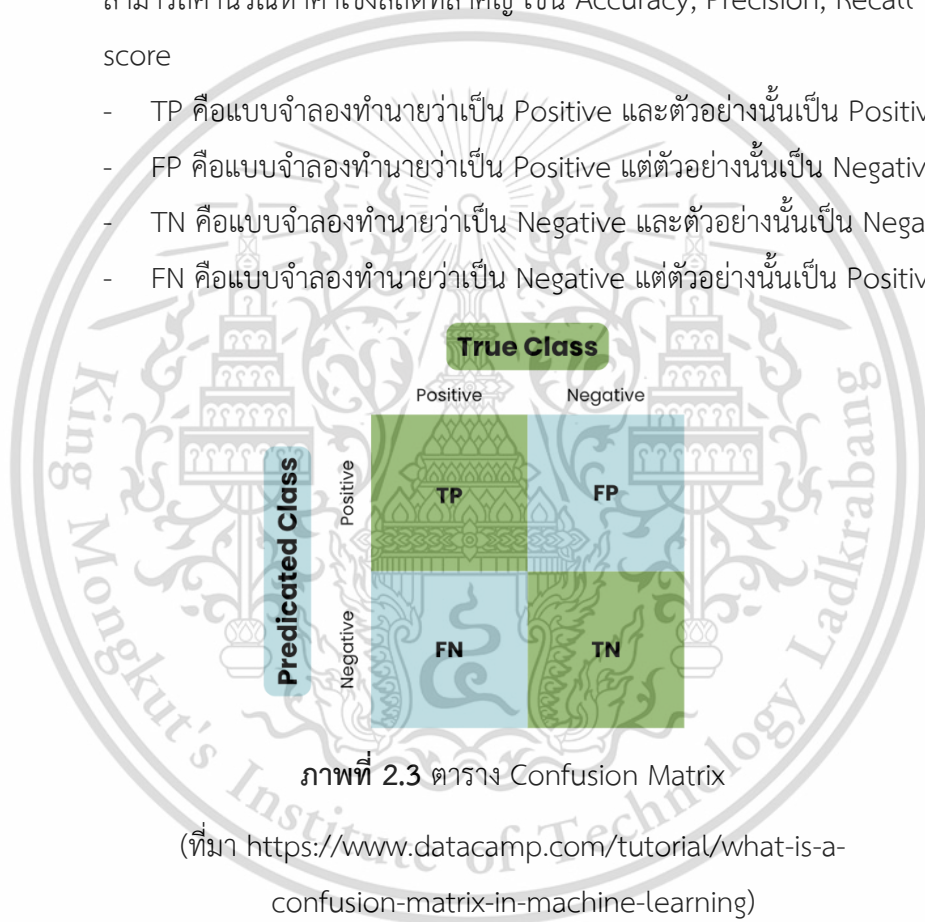
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าพารามิเตอร์จากกริดที่สร้างขึ้น ดังนั้นการทำ Random Search จะไม่รับประกันว่าจะได้แบบจำลองที่มีประสิทธิภาพที่สุดเหมือนกับ Grid Search

2.1.10 การประเมินประสิทธิภาพของแบบจำลอง

ในการประเมินประสิทธิภาพมีเทคนิคที่เป็นที่นิยม ดังนี้ (Ahmed, N. A., 2023)

1. Confusion Matrix เป็นตารางที่แสดงการจำแนกของข้อมูลจริงและการทำนายของแบบจำลอง ซึ่งจะประกอบด้วย 4 ช่องหลักคือ True Positive (TP), False Positive (FP), True Negative (TN) และ False Negative (FN) จาก confusion matrix สามารถคำนวณค่าเชิงสถิติที่สำคัญ เช่น Accuracy, Precision, Recall และ F1-score
  - TP คือแบบจำลองทำนายว่าเป็น Positive และตัวอย่างนั้นเป็น Positive จริง
  - FP คือแบบจำลองทำนายว่าเป็น Positive แต่ตัวอย่างนั้นเป็น Negative จริง
  - TN คือแบบจำลองทำนายว่าเป็น Negative และตัวอย่างนั้นเป็น Negative จริง
  - FN คือแบบจำลองทำนายว่าเป็น Negative แต่ตัวอย่างนั้นเป็น Positive จริง



ภาพที่ 2.3 ตาราง Confusion Matrix

(ที่มา <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>)

2. Accuracy ค่าความแม่นยำ หรืออัตราส่วนของการทำนายที่ถูกต้องโดยเปรียบเทียบกับจำนวนทั้งหมดของตัวอย่าง

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

ภาพที่ 2.4 การคำนวณค่า Accuracy

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. Precision ค่าความเที่ยง คืออัตราส่วนของตัวอย่างที่ทำนายถูกต้องว่าเป็น Positive จริงๆ เทียบกับตัวอย่างที่ทำนายว่าเป็น Positive ทั้งหมด

$$\text{Precision} = \frac{TP}{TP + FP} \text{ or } \frac{\text{True Positive}}{\text{Predictive Results}}$$

ภาพที่ 2.5 การคำนวณหาค่า Precision

4. Recall (Sensitivity) คืออัตราส่วนของตัวอย่างที่ทำนายถูกต้องว่าเป็น Positive จริงๆ เทียบกับจำนวนทั้งหมดของ Positive จริงๆ

$$\text{Recall} = \frac{TP}{TP + FN} \text{ or } \frac{\text{True Positive}}{\text{Actual Results}}$$

ภาพที่ 2.6 การคำนวณหาค่า Recall

5. F1-score คือตัวชี้วัดความถูกต้องของแบบจำลอง โดยการหาค่าเฉลี่ยฮาร์โมนิกของ Precision และ Recall

$$\text{F - score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

ภาพที่ 2.7 การคำนวณหาค่า F-score

## 2.2 งานวิจัยที่เกี่ยวข้อง

### 2.2.1 Anti-Money Laundering Alert Optimization Using Machine Learning with Graphs

งานวิจัยของ Eddin et al. (2022) มีจุดประสงค์หลัก คือการพัฒนาแบบจำลองการเรียนรู้ของเครื่องอย่างมีประสิทธิภาพ โดยมีเป้าหมายที่จะลดปริมาณการแจ้งเตือนที่เป็นผลบวกจาก ระบบตรวจจับการฟอกเงินที่ใช้ระบบอิงกฎ (Rule-based system) และให้ผลลัพธ์ในการตรวจจับ กิจกรรมการฟอกเงินที่มีประสิทธิภาพมากขึ้น ซึ่งการวิจัยนี้ได้พัฒนาแบบจำลองคัดกรอง (Triage model) เพื่อนำไปใช้หลังจากระบบมีการแจ้งเตือนธุรกรรมต้องสงสัยด้วยระบบอิงกฎ โดยมีการใช้ ข้อมูลจริงจากธนาคารและเสริมด้วยคุณสมบัติทางวิศวกรรมสองชุด คุณสมบัติที่เน้นเอนทิตีซึ่งจับ พฤติกรรมของเอนทิตีตามปกติ และคุณสมบัติใกล้เคียงโดยใช้กราฟซึ่งจับลักษณะของการโต้ตอบ ของเอนทิตี การรวมกันของคุณสมบัติทั้งสองนั้นทำให้สามารถแจ้งเตือนเพียง 20% ของธุรกรรม ทั้งหมด ในขณะที่ยังตรวจพบผลบวกจริงมากกว่า 90% แสดงให้เห็นว่าแบบจำลองคัดกรอง สามารถลดจำนวนผลบวกลงได้ถึง 80%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2.2 Automatic suppression of false positive alerts in anti-money laundering systems using machine learning

งานวิจัยของ Bakry et al. (2023) มีจุดประสงค์หลัก คือการพัฒนาเฟรมเวิร์กที่ใช้กับระบบตรวจจับการฟอกเงิน ที่เรียกว่า การปราบปรามอัตโนมัติด้วย XGBoost สำหรับการป้องกันการฟอกเงิน (ASXAML) เป้าหมายเพื่อลดจำนวนการแจ้งเตือนที่เป็นผลบวกวง โดย ASXAML นำข้อมูลและคุณสมบัติที่สร้างขึ้นโดยการสร้างสถานการณ์การฟอกเงินโดยผู้เชี่ยวชาญ และนำข้อมูลที่ได้ออกมาแบ่งกลุ่มตามลูกค้าและใช้วิธีการเลือกคุณสมบัติแบบ Recursive Feature Elimination with Cross-Validation (RFECV) เพื่อเลือกคุณสมบัติที่ดีที่สุด ก่อนที่จะส่งต่อไปยัง Optuna เพื่อปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost ให้เหมาะสม ผลลัพธ์ระบุว่า ASXAML บรรลุความสมดุลที่เหมาะสมที่สุดระหว่างการลดผลบวกวงและการหลีกเลี่ยงเหตุการณ์การฟอกเงินที่พลาดไป ด้วยคะแนน F-beta 86% และเหตุการณ์การฟอกเงินเพียง 11% เท่านั้นที่ถูกทำนายอย่างไม่ต้องสงสัย ในข้อมูลทดสอบ

## 2.2.3 Improving Classification Performance of Money Laundering Transactions Using Typological Features

งานวิจัยของ Phyu, T. H. and Uttama, S. (2023) มีจุดประสงค์หลัก คือเพื่อวิเคราะห์ประสิทธิภาพการจำแนกประเภทของชุดข้อมูล IBM Anti-Money Laundering โดยมุ่งเน้นไปที่การระบุคุณลักษณะที่มีประสิทธิภาพและประสิทธิผล รวมถึงการระบุประเภทของธุรกรรมต่างๆ ที่เกี่ยวข้องกับการฟอกเงิน เช่น กระจายเข้า (fan-in) กระจายออก (fan-out) ร่วมกัน (mutual) และวงจร (cycle) นอกจากนี้งานวิจัยนี้ยังได้เปรียบเทียบแบบจำลองการเรียนรู้ของเครื่องเหล่านี้เพื่อนบ้านใกล้สุด k ตัว (KNN) ต้นไม้การตัดสินใจ (DT) Support Vector Machine (SVM) และ Naïve Bayes (NB) โดยผลการทดลองแสดงให้เห็นว่าแบบจำลองที่มีความแม่นยำมากที่สุดคือ DT ความแม่นยำอยู่ที่ 87.6% ซึ่งสามารถจำแนกธุรกรรมฟอกเงินได้อย่างมีประสิทธิภาพ ในขณะที่แบบจำลองอื่นๆ มีความแม่นยำ 87.2% (KNN), 84.4% (SVM) และ 71.3% (NB) โดยใช้ชุดข้อมูลที่มีประเภทของธุรกรรมแบบเต็ม ในขณะที่การจำแนกประเภทแบบเดี่ยวที่ได้ทำการตรวจสอบ การจำแนกประเภท mutual มีประสิทธิภาพเหนือกว่าประเภทอื่นๆ อย่างต่อเนื่องในแง่ของความแม่นยำ โดยการค้นพบนี้เน้นย้ำถึงความสำคัญของการใช้ประเภทของธุรกรรมในการเพิ่มประสิทธิภาพของการจำแนกประเภทการฟอกเงิน

#### 2.2.4 Machine Learning for Anti-Money Laundering and Fraud Detection

งานวิจัยของ Pathiyaparambath, T. S. and Gopika P., Professor. (2024) มีจุดประสงค์หลัก คือการสร้างระบบป้องกันการฟอกเงินที่มีประสิทธิภาพ ซึ่งสามารถระบุเส้นทางทางการเงินที่ถูกฟอกโดยใช้แนวทาง Hash-based Association รวมถึงประสบความสำเร็จในการระบุตัวแทนและผู้เกี่ยวข้องในขั้นตอนของการฟอกเงินโดยวิธี Graph Theoretic และยังสามารถตรวจจับการฉ้อโกงบัตรเครดิตได้ด้วย นอกจากนี้ยังช่วยลดระยะเวลาดำเนินการตรวจสอบการฟอกเงินของเจ้าหน้าที่กำกับดูแลการปฏิบัติตามกฎระเบียบ งานวิจัยนี้ได้ทดลองใช้แบบจำลองการเรียนรู้ของเครื่องสองแบบ ได้แก่ Logistic Regression และ Random Forest โดยผลการทดลองพบว่าแบบจำลอง Random Forest ให้ผลลัพธ์ที่ดีกว่า Logistic Regression

#### 2.2.5 Money Laundering Detection using A Transaction-based Graph Learning Approach

งานวิจัยของ Huong et al. (2024) มีจุดประสงค์หลัก คือเนื่องจากวิธีการป้องกันการฟอกเงินแบบดั้งเดิมนั้นอาศัยระบบอิงกฎและวิธีการทางสถิติเป็นส่วนใหญ่ จึงมีข้อจำกัดในการจับภาพความสัมพันธ์ที่ซับซ้อนและเชื่อมโยงถึงกันที่ซ่อนอยู่ในเครือข่ายการฟอกเงิน งานวิจัยนี้จึงสร้างกราฟเครือข่ายจากชุดข้อมูลธุรกรรมธนาคารขนาดใหญ่ โดยดึงข้อมูลเชิงลึกจากการสร้างแบบจำลองภาษาและการเรียนรู้แบบมีผู้สอน จากนั้นใช้แบบจำลอง Random Forest (RF) เพื่อคาดการณ์พฤติกรรมที่น่าสงสัยที่เกี่ยวข้องกับการฟอกเงิน นอกจากนี้ยังจัดการกับความไม่สมดุลของคลาสในบริบทของการตรวจจับการฟอกเงินผ่านการทดลองทั้งการสุ่มตัวอย่างเกินและการสุ่มตัวอย่างต่ำกว่า ผลการทดลองพบว่าประสิทธิภาพการทำนายของแบบจำลอง RF ที่มีการสุ่มตัวอย่างเกินนั้นให้ความแม่นยำ 86% ในขณะที่เมื่อใช้การสุ่มตัวอย่างต่ำกว่านั้น ความแม่นยำจะเพิ่มขึ้นเป็น 92%

#### 2.2.6 SAML-D Exploratory Data Analysis & Classification

การทดลองของ Berkan Oztas (2024) มีการใช้ชุดข้อมูลสังเคราะห์ชุดเดียวกับงานวิจัยนี้ โดยการทดลองดังกล่าวมีจุดประสงค์หลัก คือการทำความเข้าใจและการใช้งานชุดข้อมูล SAML-D โดยเจาะลึกไปที่การวิเคราะห์ข้อมูลเชิงสำรวจ ซึ่งเป็นการตรวจสอบคุณลักษณะของชุดข้อมูลอย่างละเอียด จากนั้นทำการจัดการข้อมูลเพื่อให้แน่ใจว่าข้อมูลได้รับการปรับให้เหมาะสมสำหรับการสร้างแบบจำลอง โดยใช้เทคนิค Log Transform กับข้อมูลจำนวนเงิน เพื่อให้จำนวนเงินมีการแจกแจงแบบปกติ จากนั้นจึงทำการสร้างแบบจำลองด้วย XGBoost ผลการทดลองพบว่าแบบจำลอง XGBoost ให้ความแม่นยำที่ 81.205% แต่สร้างผลบวกคลวง (False Positive Rate) มากถึง 58%

### บทที่ 3

## วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานของงานวิจัยนี้ มุ่งเน้นไปที่การสร้างแบบจำลองที่สามารถตรวจจับธุรกรรม การฟอกเงินได้ และทำการเปรียบเทียบวิธีการเรียนรู้ของเครื่องแบบต่างๆ เพื่อหาแบบจำลองที่มี ประสิทธิภาพที่สุด

### 3.1 ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง ตั้งแต่การศึกษาระบบตรวจจับการฟอกเงิน ของธนาคาร การจัดการข้อมูลและเตรียมข้อมูลให้พร้อมก่อนนำเข้าสู่แบบจำลอง วิเคราะห์และแก้ไข ปัญหาที่พบเจอของข้อมูล แบ่งชุดข้อมูลเพื่อใช้ฝึกแบบจำลองและเพื่อทดสอบกับแบบจำลอง จากนั้น เริ่มขั้นตอนการสร้างแบบจำลอง ปรับแต่งไฮเปอร์พารามิเตอร์เพื่อหาพารามิเตอร์ชุดที่เหมาะสมกับ แบบจำลองแต่ละวิธีมากที่สุด รวมไปถึงเปรียบเทียบและประเมินประสิทธิภาพของแบบจำลองแต่ละ ตัว เพื่อให้ได้แบบจำลองที่ตรวจจับธุรกรรมการฟอกเงินได้มีประสิทธิภาพสูงสุด แสดงดัง ภาพที่ 3.1

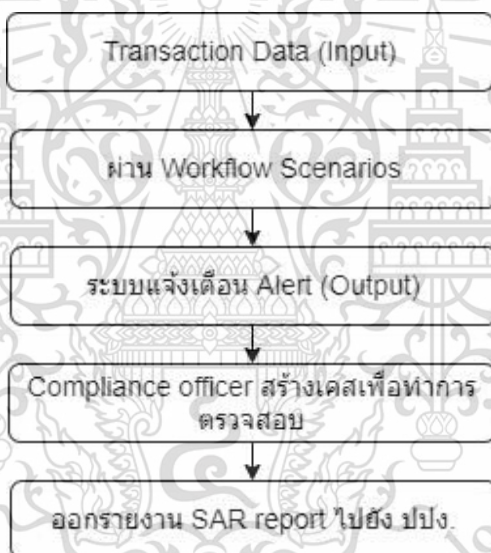


ภาพที่ 3.1 ขั้นตอนการดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.1 ศึกษาและทำความเข้าใจในระบบตรวจจับการฟอกเงิน

ศึกษาและทำความเข้าใจในระบบตรวจจับการฟอกเงินแบบดั้งเดิมของธนาคาร ตั้งแต่ข้อมูลขาเข้าจนผ่านกระบวนการตรวจสอบธุรกรรมต้องสงสัย โดยเริ่มจากศึกษาการสร้างสถานการณ์ หรือ Workflow Scenario ที่อาจเกิดขึ้นได้ของธุรกรรมที่บ่งชี้ว่าอาจจะเป็นการฟอกเงิน ซึ่งระบบจะทำการแจ้งเตือนธุรกรรมต้องสงสัยที่เข้าเงื่อนไขของสถานการณ์ที่ถูกสร้างไว้ ไปยังธนาคาร จากนั้นเจ้าหน้าที่ปฏิบัติตามข้อกำหนดจะทำการสร้างเคสเพื่อเข้าไปตรวจสอบธุรกรรมและบุคคลที่ทำธุรกรรมนั้น รวมถึงตรวจสอบบุคคลที่มีความเกี่ยวข้องกันกับบัญชีที่ทำธุรกรรม เพื่อตรวจหาเครือข่ายที่ร่วมขบวนการฟอกเงินนี้ หากการแจ้งเตือนนั้นเป็นข้อผิดพลาดที่ไม่ใช่การฟอกเงิน เจ้าหน้าที่ปฏิบัติตามข้อกำหนดจะปิดเคสนั้นไป แต่หากการแจ้งเตือนนั้นถูกตรวจสอบแล้วว่าเป็นธุรกรรมฟอกเงินจริง เจ้าหน้าที่ปฏิบัติตามข้อกำหนดจะทำการส่งรายชื่อบุคคลที่เกี่ยวข้องทั้งหมดไปให้กับสำนักงานป้องกันและปราบปรามการฟอกเงิน แสดงดัง ภาพที่ 3.2



ภาพที่ 3.2 ตัวอย่างการทำงานของระบบตรวจจับการฟอกเงิน

### 3.1.2 จัดหาข้อมูลและทำการศึกษาข้อมูล

จากข้อจำกัดทางกฎหมายและปัญหาในการเปิดเผยข้อมูลของธนาคาร ทำให้ไม่สามารถนำข้อมูลจริงที่เป็นธุรกรรมการเงินของธนาคารมาใช้ในการสร้างแบบจำลองได้ และจากการค้นคว้าข้อมูลที่จะสามารถนำมาสร้างแบบจำลองเพื่อตรวจจับธุรกรรมฟอกเงินได้นั้น ผู้วิจัยจึงเลือกใช้ข้อมูลจากเว็บไซต์ Kaggle ซึ่งเป็นข้อมูลธุรกรรมที่ถูกสังเคราะห์ขึ้นมาจากข้อมูลจริงของธนาคารในประเทศอังกฤษ จำนวนทั้งหมด 9,504,852 ธุรกรรม โดยมีคุณลักษณะ 12 ประการ ที่รวมถึงผลเฉลยของธุรกรรมฟอกเงิน คือ ป้ายกำกับที่บ่งชี้ธุรกรรมฟอกเงิน มีค่าเป็น 0 และ 1 โดยที่ 0 มีความหมายว่า ธุรกรรมนั้นเป็นธุรกรรมปกติ ในทางกลับกัน 1 มีความหมายว่า ธุรกรรมนั้นเป็นธุรกรรมฟอกเงิน และประเภทของธุรกรรม ที่ถูกระบุจากผู้เชี่ยวชาญ ซึ่งมีข้อมูลเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทของธุรกรรมปกติ ระบุไว้ด้วยคำว่า Normal นำหน้าชื่อของประเภทนั้นๆ จึงทำให้คุณลักษณะดังกล่าวเป็นเหมือนผลเฉลยของข้อมูลชุดนี้ด้วย รายละเอียดดัง ตารางที่ 3.1

**ตารางที่ 3.1** รายละเอียด ชนิด และจำนวนของข้อมูล

คุณลักษณะ	ความหมาย	ชนิดของข้อมูล	ข้อมูลทั้งหมด
Time	เวลาที่ทำธุรกรรม	object	9504852
Date	วันที่ที่ทำธุรกรรม	object	9504852
Sender_account	บัญชีต้นทาง	int64	9504852
Receiver_account	บัญชีปลายทาง	int64	9504852
Amount	จำนวนเงิน	float64	9504852
Payment_currency	สกุลเงินในการชำระเงิน	object	9504852
Received_currency	สกุลเงินของฝั่งปลายทาง	object	9504852
Sender_bank_location	ที่ตั้งธนาคารบัญชีต้นทาง	object	9504852
Receiver_bank_location	ที่ตั้งธนาคารบัญชีปลายทาง	object	9504852
Payment_type	ประเภทของการชำระเงิน	object	9504852
Is_laundring	ป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน	int64	9504852
Laundring_type	ประเภทของธุรกรรม	object	9504852

### 3.1.3 เตรียมข้อมูลและวิเคราะห์ข้อมูล

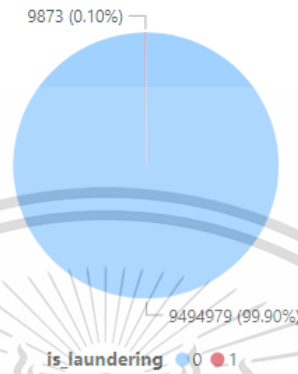
ขั้นตอนสำคัญก่อนสร้างแบบจำลอง คือการเตรียมข้อมูลให้พร้อมก่อนจะนำไปฝึกแบบจำลอง ซึ่งผู้ทำวิจัย ได้สำรวจชุดข้อมูลและทำการจัดเตรียมข้อมูลดังนี้ เนื่องจากการทำธุรกรรมทางธนาคาร วันที่และเวลา เป็นปัจจัยสำคัญไม่ว่าจะเป็นธุรกรรมปกติหรือธุรกรรมการฟอกเงิน เนื่องจากเราอาจจะพบกับรูปแบบของการทำธุรกรรมบางประเภทได้ โดยชุดข้อมูลนี้มีคุณลักษณะ วันที่และเวลาในการทำธุรกรรม เพื่อให้พบรูปแบบของการทำธุรกรรมได้ชัดเจน จึงทำการนำคุณลักษณะ Date มาแยกเป็น เดือน จะได้ข้อมูลตัวเลขตั้งแต่ 1 ถึง 12 และแยกเป็น วัน จะได้ข้อมูลตัวเลขตั้งแต่ 1 ถึง 31 ต่อมานำคุณลักษณะ Time มาแยกเป็น ชั่วโมง จะได้ข้อมูลตัวเลขตั้งแต่ 1 ถึง 24 จากนั้นทำการตัดคุณลักษณะ Date และ Time ที่ทิ้ง ซึ่งการแยกคุณลักษณะเพิ่มเติมนั้นจะสามารถช่วยให้การศึกษา และการวิเคราะห์คุณลักษณะต่างๆ ของข้อมูลนั้นมีประสิทธิภาพมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต่อมาจะเป็นขั้นตอนที่สำคัญอีกหนึ่งอย่างเพื่อให้การสร้างแบบจำลองมีประสิทธิภาพที่ดีเยี่ยมคือการสำรวจและวิเคราะห์ ข้อมูล คุณลักษณะต่างๆประกอบกัน ดังนี้

### 3.1.3.1 ป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน (Is\_laundering)

Laundering and Non-Laundering Distribution

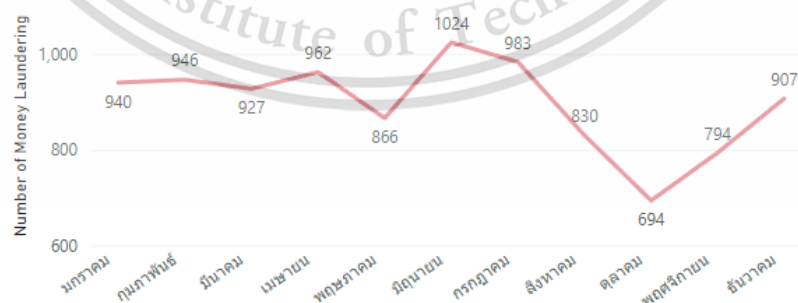


ภาพที่ 3.3 จำนวนธุรกรรมทั้งหมด โดยแบ่งตามป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน

จากข้อมูลชุดนี้ พบว่าจำนวนธุรกรรมทั้งหมดมี 9,504,852 รายการ โดยแบ่งเป็นธุรกรรมที่เป็นการฟอกเงินอยู่เพียง 9,873 รายการ ซึ่งคิดเป็นแค่ 0.1% ของจำนวนธุรกรรมทั้งหมด ในขณะที่ธุรกรรมที่ไม่ใช่การฟอกเงินมีอยู่มากถึง 9,494,979 รายการ และนับเป็นกว่า 99.9% ของข้อมูลทั้งหมด เนื่องจากการฟอกเงินนั้นไม่ได้เกิดเหตุบ่อยมากนัก เหมือนกับจำนวนธุรกรรมปกติทั่วไป ทำให้ผลเฉลยที่มีอาจพบปัญหาความไม่สมดุลของข้อมูล ด้วยเหตุนี้จึงจะมีขั้นตอนในการจัดการปัญหาข้อมูลไม่สมดุลในภายหลัง ก่อนที่จะนำเข้าแบบจำลอง

### 3.1.3.2 เดือนที่ทำธุรกรรม (Month)

Number of Money Laundering per month



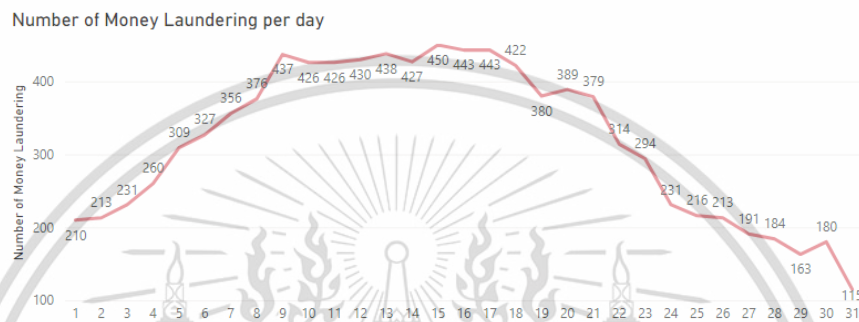
ภาพที่ 3.4 จำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นในแต่ละเดือน

จากการสำรวจธุรกรรมที่เป็นการฟอกเงิน โดยดูจากจำนวนธุรกรรมต่อเดือนที่เกิดขึ้นในแต่ละปี ซึ่งจะพบว่าตั้งแต่เดือนมกราคม จนถึง ธันวาคม จำนวนธุรกรรมที่เกิดขึ้นในแต่ละ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เดือนมีจำนวนที่ไม่ต่างกันมากนัก โดยในเดือนมิถุนายนจะเกิดธุรกรรมการฟอกเงินสูงที่สุดอยู่ที่ 1,024 ธุรกรรม และเดือนตุลาคมจะเกิดธุรกรรมการฟอกเงินต่ำที่สุดอยู่ที่ 694 ธุรกรรม และเนื่องจากชุดข้อมูลนี้มีธุรกรรมตั้งแต่เดือน ตุลาคม ของปี 2022 จนถึง สิงหาคม ปี 2023 ทำให้ในเดือนกันยายนนั้นไม่มีข้อมูลธุรกรรมอยู่เลย สรุปได้ว่าจำนวนธุรกรรมต่อเดือนไม่พบรูปแบบที่มีนัยสำคัญสำหรับการฟอกเงินได้

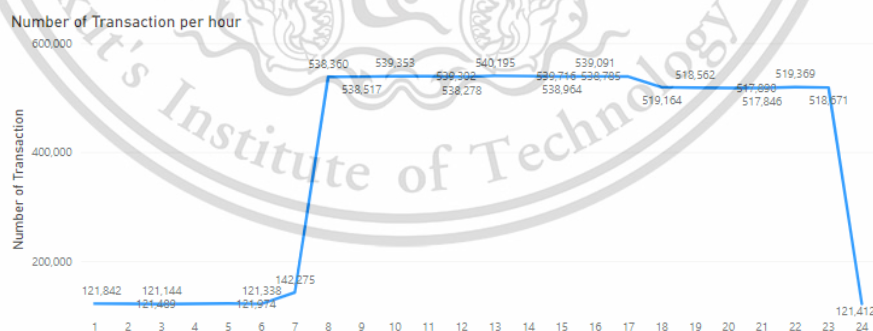
### 3.1.3.3 วันที่ทำธุรกรรม (Day)



ภาพที่ 3.5 จำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นในแต่ละวัน

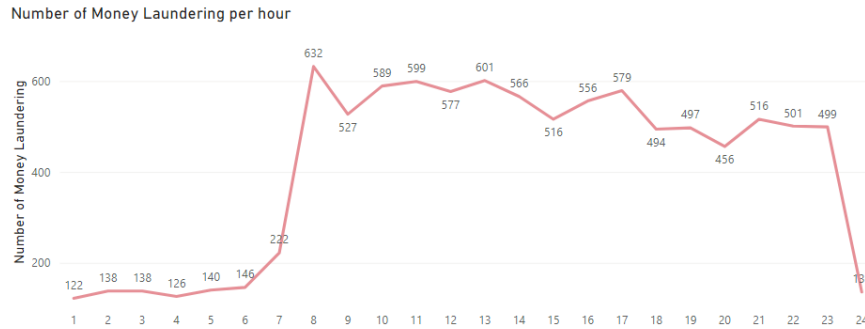
จากการสำรวจธุรกรรมที่เป็นการฟอกเงิน โดยดูจากจำนวนธุรกรรมต่อวันที่เกิดขึ้นในแต่ละเดือน พบว่าจำนวนธุรกรรมการฟอกเงินที่เกิดขึ้นนั้น มีแนวโน้มมากขึ้นเรื่อยๆ และสูงที่สุดอยู่ในช่วงกลางเดือนตั้งแต่ วันที่ 9 ถึง 18 และจากนั้นจำนวนจะลดลงเรื่อยๆ จนถึงสิ้นเดือน ซึ่งสรุปได้ว่า วันที่ในการฟอกเงินมีรูปแบบที่เป็นนัยสำคัญสำหรับการฟอกเงิน

### 3.1.3.4 เวลาที่ทำธุรกรรม (Hour)



ภาพที่ 3.6 จำนวนธุรกรรมทั้งหมดที่เกิดขึ้นในแต่ละชั่วโมง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

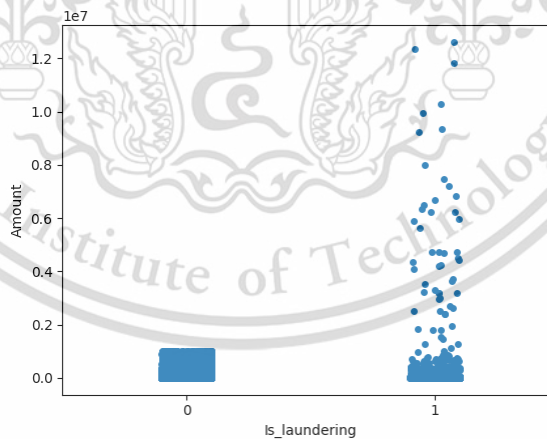


ภาพที่ 3.7 จำนวนธุรกรรมฟอกเงินที่เกิดขึ้นในแต่ละชั่วโมง

จากการสำรวจธุรกรรมทั้งหมด โดยดูจากจำนวนธุรกรรมรายชั่วโมงที่เกิดขึ้นในแต่ละวัน พบว่าธุรกรรมส่วนใหญ่มักจะเกิดขึ้นในช่วง 8 โมงเช้าเป็นต้นไป จนถึงเที่ยงคืน และจะมีจำนวนลดลงอย่างเห็นได้ชัดในช่วง เที่ยงคืน ไปจนถึง 7 โมงเช้า ซึ่งรูปแบบที่เกิดขึ้นนี้ ดูเหมือนเป็นรูปแบบทั่วไปสำหรับการทำธุรกรรมปกติ

และเมื่อทำการสำรวจธุรกรรมที่เป็นการฟอกเงิน โดยดูจากจำนวนธุรกรรมรายชั่วโมงที่เกิดขึ้นในแต่ละวัน ตรวจสอบว่ามีรูปแบบคล้ายกันกับธุรกรรมปกติ คือ ธุรกรรมฟอกเงินส่วนใหญ่จะเกิดขึ้นในช่วง 8 โมงเช้า จนถึงเที่ยงคืน และจะมีจำนวนลดลงช่วงเที่ยงคืน จนถึง 7 โมงเช้า ดังนั้นได้ข้อสรุปว่า ชั่วโมงในการทำธุรกรรมไม่มีรูปแบบที่เกี่ยวกับการฟอกเงิน

### 3.1.3.5 จำนวนเงิน (Amount)



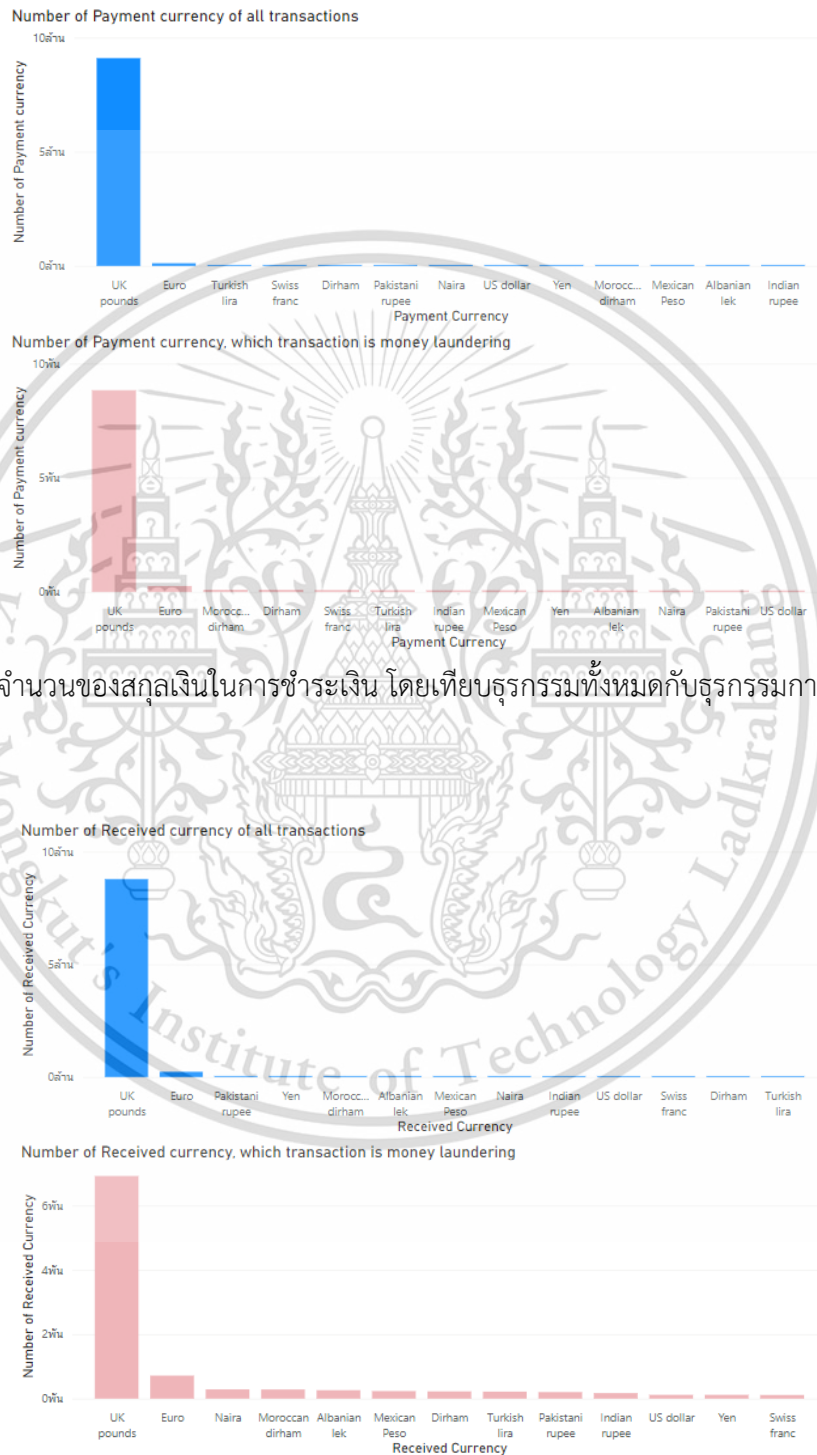
ภาพที่ 3.8 การกระจายตัวของจำนวนเงิน

จากการสำรวจจำนวนเงินที่ทำธุรกรรม โดยแยกกลุ่มระหว่างธุรกรรมที่เป็นการฟอกเงิน และธุรกรรมที่เป็นปกติทั่วไป จากภาพที่ 3.8 พบว่าจำนวนเงินในกลุ่มธุรกรรมที่เป็นปกตินั้น จะมีการกระจุกตัวของจำนวนเงิน มากกว่ากลุ่มของธุรกรรมฟอกเงิน ที่ดูมีการกระจายตัวกันมากกว่าธุรกรรมปกติ และเห็นได้ว่าการกระจุกตัวของธุรกรรมปกติ เป็นการกระจุกตัว

ในช่วงจำนวนเงินที่ต่ำกว่าจำนวนเงินของธุรกรรมฟอกเงินมาก เนื่องมาจากเหตุผลของการเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟอกเงินนั้น มีจำนวนเงินมหาศาลมาเกี่ยวข้อง ซึ่งบ่งบอกถึงการโอนย้ายเงินจำนวนมากในการทำธุรกรรมฟอกเงิน

3.1.3.6 สกุลเงินในการชำระเงิน และสกุลเงินของฝั่งปลายทาง



ภาพที่ 3.9 จำนวนของสกุลเงินในการชำระเงิน โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมฟอกเงิน

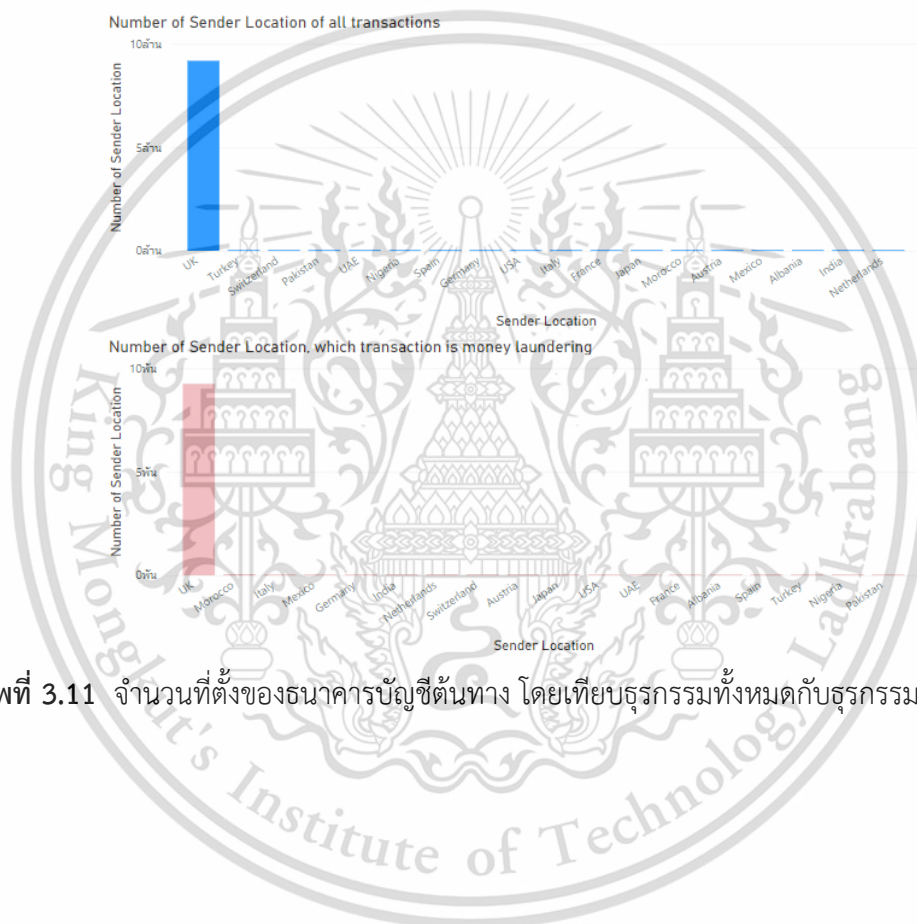
ภาพที่ 3.10 จำนวนของสกุลเงินของฝั่งปลายทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมฟอกเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

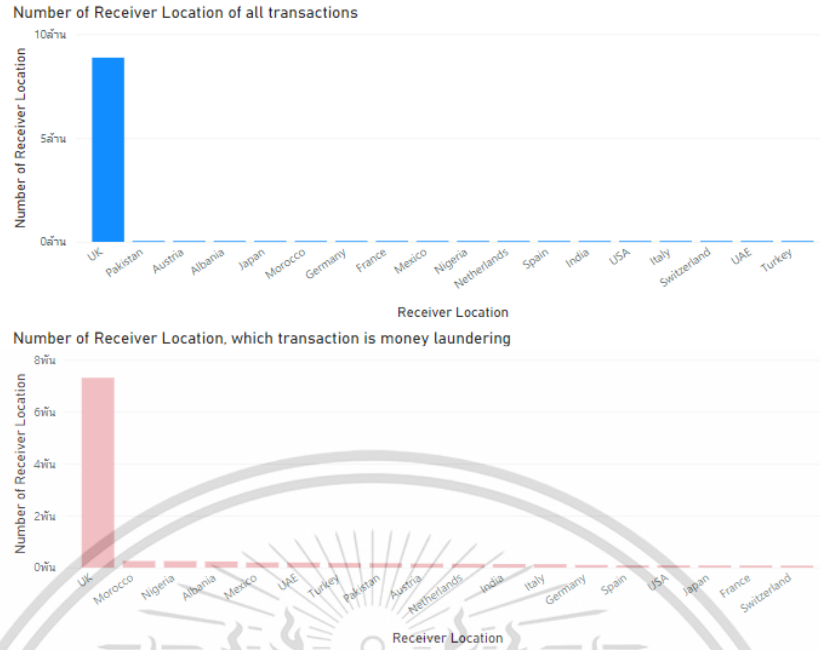
จากการสำรวจสกุลเงินในการชำระเงิน โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน พบว่าธุรกรรมส่วนใหญ่ใช้สกุลเงิน ปอนด์ ของประเทศอังกฤษ (UK pounds) เนื่องจากธุรกรรมสังเคราะห์ชุดนี้ มาจากธนาคารในประเทศอังกฤษ

และจากการสำรวจสกุลเงินของฝั่งปลายทาง โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน ก็พบว่าธุรกรรมส่วนใหญ่ใช้สกุลเงิน ปอนด์ ของประเทศอังกฤษ (UK pounds) เช่นเดียวกันกับสกุลเงินในการชำระเงิน

### 3.1.3.7 ที่ตั้งธนาคารบัญชีต้นทาง และปลายทาง



ภาพที่ 3.11 จำนวนที่ตั้งของธนาคารบัญชีต้นทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมฟอกเงิน

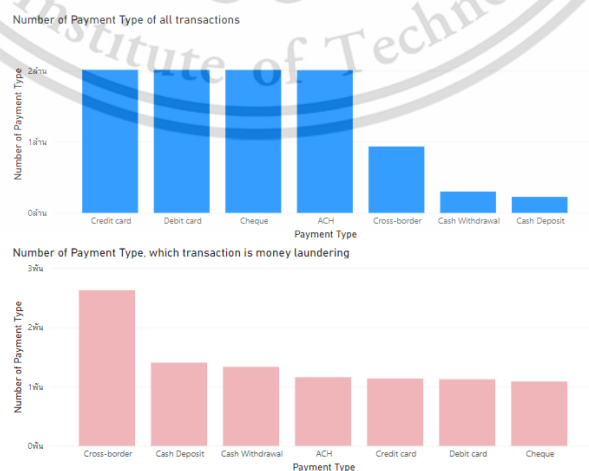


ภาพที่ 3.12 จำนวนที่ตั้งของธนาคารบัญชีปลายทาง โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน

จากการสำรวจที่ตั้งของธนาคารบัญชีต้นทาง โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน พบว่าบัญชีต้นทางส่วนใหญ่มีที่ตั้งอยู่ที่ประเทศอังกฤษ เนื่องจากข้อมูลชุดนี้มาจากธนาคารแห่งหนึ่งในประเทศอังกฤษ

จากการสำรวจที่ตั้งของธนาคารบัญชีปลายทาง โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน พบว่าบัญชีปลายทางส่วนใหญ่มีที่ตั้งอยู่ที่ประเทศอังกฤษ เช่นเดียวกับบัญชีต้นทาง

### 3.1.3.8 ประเภทของการชำระเงิน

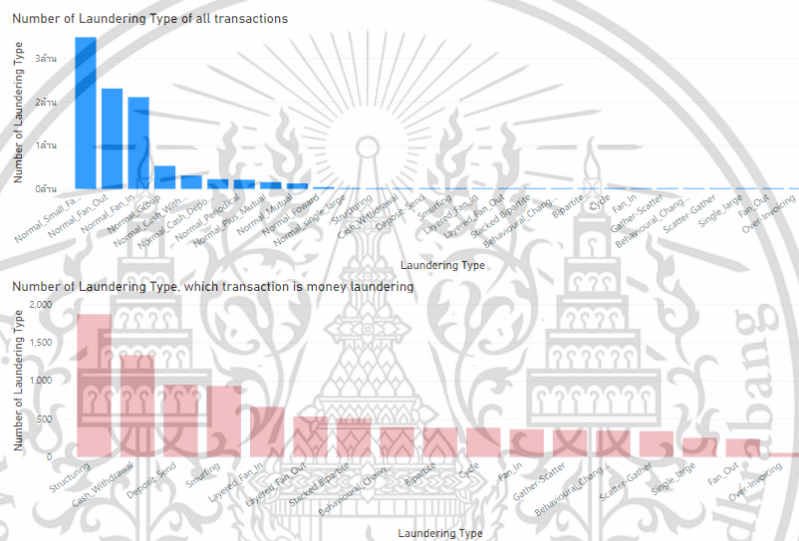


ภาพที่ 3.13 ประเภทของการชำระเงิน โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการสำรวจประเภทของการชำระเงิน โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน พบว่าธุรกรรมปกติมักจะมีรูปแบบการชำระเงินส่วนใหญ่เป็น บัตรเครดิต (Credit card) บัตรเดบิต (Debit card) เช็ค (Cheque) และสำนักหักบัญชีอัตโนมัติ (ACH: Automated Clearing House) แต่ธุรกรรมกลุ่มที่เป็นการฟอกเงิน จะมีรูปแบบการชำระเงินหลักที่เห็นได้ชัดที่สุดคือ การทำธุรกรรมข้ามประเทศ (Cross-border) ซึ่งถือเป็นหนึ่งในวิธีที่นิยมใช้สำหรับผู้ที่ต้องการฟอกเงิน รองลงมาคือวิธีการฝากเงินสด (Cash Deposit) และการถอนเงินสด (Cash Withdrawal)

3.1.3.9 ประเภทของธุรกรรม



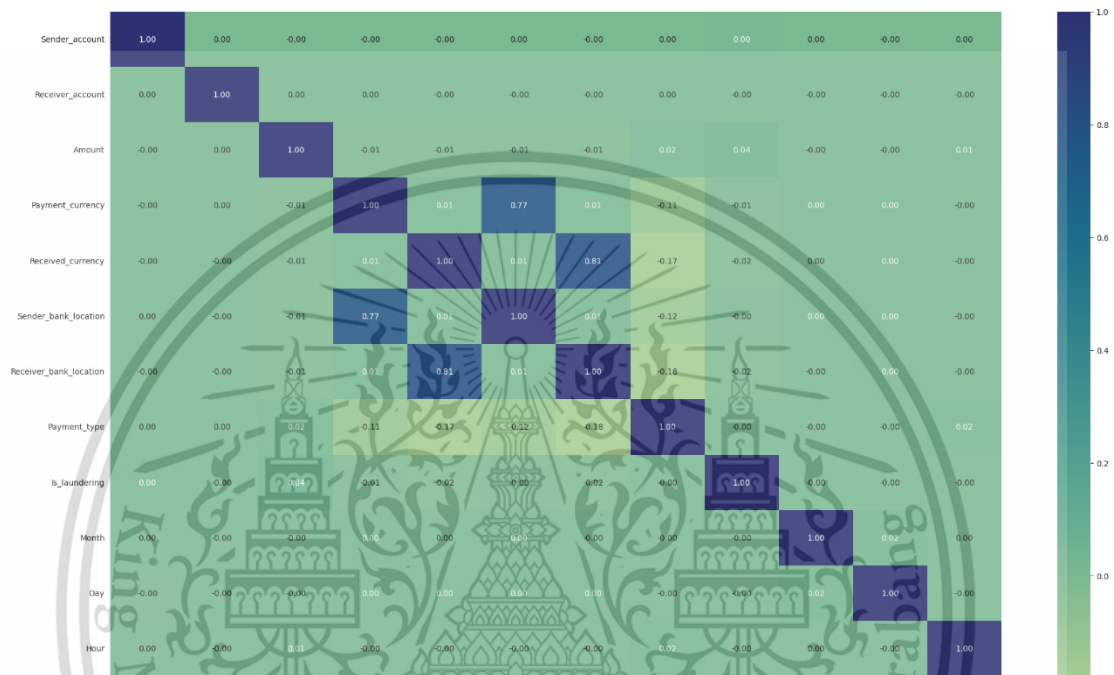
ภาพที่ 3.14 ประเภทของธุรกรรม โดยเทียบธุรกรรมทั้งหมดกับธุรกรรมการฟอกเงิน

จากการสำรวจประเภทของธุรกรรม ที่มาจากการระบุโดยผู้เชี่ยวชาญ โดยเทียบจากธุรกรรมทั้งหมดกับธุรกรรมกลุ่มที่เป็นการฟอกเงิน พบว่ากลุ่มที่เป็นการฟอกเงิน จะมีประเภทของธุรกรรมที่เป็นแบบมีโครงสร้าง (Structuring) มากที่สุด รองลงมาคือ การถอนเงินสด (Cash Withdrawal) และการฝากส่งเงิน (Deposit – send) ตามลำดับ

เนื่องจากประเภทของธุรกรรม มีการแยกประเภทชัดเจนสำหรับธุรกรรมปกติกับธุรกรรมกลุ่มฟอกเงิน จึงทำการตัดคุณลักษณะประเภทของธุรกรรมทิ้ง เพื่อไม่ให้แบบจำลองเรียนรู้ผู้ผลเฉลยจากข้อมูลดังกล่าว

### 3.1.3.10 ความสัมพันธ์ของคุณลักษณะ

การหาความสัมพันธ์ของแต่ละคุณลักษณะ มีจุดประสงค์เพื่อตัดคุณลักษณะบางตัวที่มีความสัมพันธ์กันมากเกินไปออก เป็นการลดคุณลักษณะที่ไม่จำเป็นก่อนสร้างแบบจำลอง โดยจะทำการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร และแสดงผลด้วยตารางเมทริกซ์สหสัมพันธ์ (Correlation matrix) ดังภาพที่ 3.15



ภาพที่ 3.15 ตารางเมทริกซ์สหสัมพันธ์

จากเมทริกซ์สหสัมพันธ์ ภาพที่ 3.15 จะเห็นว่าคุณลักษณะที่ตั้งธนาคารของบัญชีต้นทางมีความสัมพันธ์กับคุณลักษณะสกุลเงินในการชำระเงินในทิศทางเดียวกันค่อนข้างสูง โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.77 และคุณลักษณะที่ตั้งธนาคารของบัญชีปลายทางมีความสัมพันธ์กับคุณลักษณะสกุลเงินของฝั่งปลายทางในทิศทางเดียวกันค่อนข้างสูงเช่นกัน โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.81

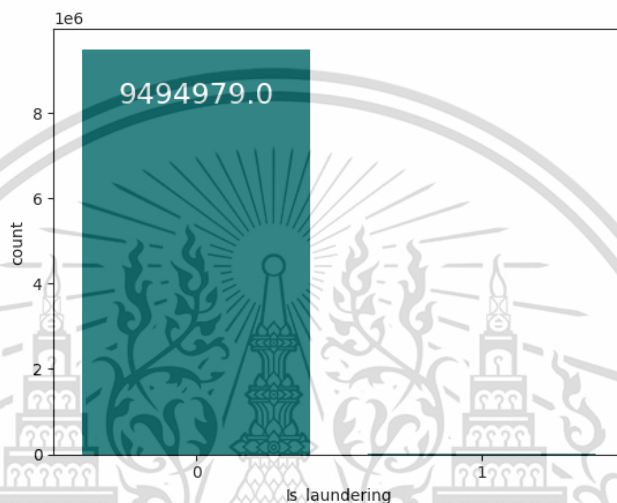
เนื่องมาจากที่ตั้งธนาคารและสกุลเงินเป็นผลโดยตรงจากการที่ข้อมูลชุดนี้สังเคราะห์มาจากธนาคารในประเทศไทย และธุรกรรมที่เกิดขึ้นในข้อมูลชุดนี้ส่วนมากใช้สกุลเงิน ปอนด์ ในการทำธุรกรรม เลยจะเห็นความสัมพันธ์กันของคุณลักษณะที่ตั้งธนาคารและสกุลเงินค่อนข้างสูง

ดังนั้นเพื่อไม่ให้แบบจำลองเรียนรู้ข้อมูลที่ซ้ำกันหรือมีความสัมพันธ์ไปในทางเดียวกันนี้ จึงทำการเลือกตัดหนึ่งในคุณลักษณะดังกล่าวทิ้ง คือ สกุลเงินในการชำระเงิน (Payment currency) และสกุลเงินของฝั่งปลายทาง (Received currency)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

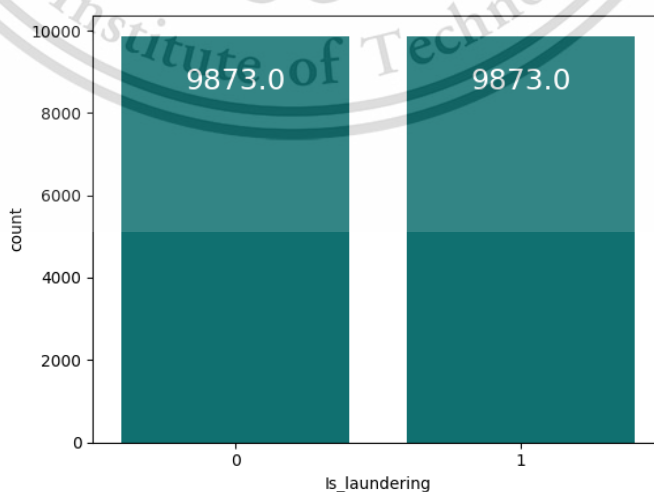
### 3.1.4 ดำเนินการจัดการกับข้อมูลที่ไม่สมดุล

จากการสำรวจข้อมูลป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน จะพบว่าข้อมูลที่เป็นธุรกรรมการฟอกเงินมีเพียง 0.1% จากธุรกรรมทั้งหมด ในขณะที่ข้อมูลธุรกรรมปกติมีจำนวนกว่า 99.9% ของข้อมูลทั้งหมด ซึ่งอาจทำให้แบบจำลองเรียนรู้จากข้อมูลที่มีได้ไม่มีประสิทธิภาพมากนักจากปัญหาข้อมูลที่ไม่สมดุลนี้ เนื่องจากว่าแบบจำลองจะสามารถเรียนรู้ธุรกรรมปกติได้ดีมาก ในขณะที่จะไม่สามารถเรียนรู้ธุรกรรมที่เป็นการฟอกเงินได้ดีเท่าที่ควร เพราะข้อมูลที่มีให้แบบจำลองได้ฝึกนั้นมีน้อยเกินไป ดังภาพที่ 3.16



ภาพที่ 3.16 ความไม่สมดุลของข้อมูล

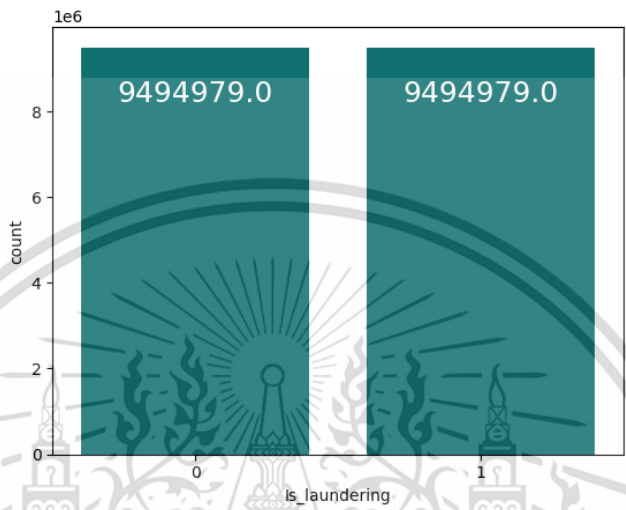
ผู้วิจัยได้ทำการทดลองจัดการกับปัญหาข้อมูลไม่สมดุล 2 วิธี คือการทำ Undersampling โดยตัดจำนวนข้อมูลธุรกรรมปกติ ให้มีขนาดเท่ากับจำนวนข้อมูลธุรกรรมการฟอกเงิน ทำให้ข้อมูลทั้งสองคลาสมีขนาดเท่ากัน คือคลาสละ 9,873 ธุรกรรม รวมเป็น 19,746 ธุรกรรม ดังภาพที่ 3.17 ซึ่งวิธีนี้จะช่วยลดปัญหาที่แบบจำลองจะฝึกกับข้อมูลคลาสมากกว่าได้ดีเกินไป



ภาพที่ 3.17 ความสมดุลของข้อมูล หลังจากทำการ Undersampling

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อีกหนึ่งวิธีที่ผู้วิจัยเลือกใช้จัดการกับปัญหาข้อมูลไม่สมดุล คือการทำ Oversampling ด้วยวิธี SMOTE โดยมีหลักการคือ การสังเคราะห์ตัวอย่างข้อมูลธุรกรรมการฟอกเงินขึ้นมาจากข้อมูลที่มีอยู่ โดยไม่ให้เกิดความซ้ำซ้อนของข้อมูล จากนั้นสังเคราะห์ตัวอย่างให้มีขนาดเท่ากับข้อมูลธุรกรรมปกติ ซึ่งจะทำให้ข้อมูลที่เป็นธุรกรรมการฟอกเงินนั้นมีจำนวนเท่ากับธุรกรรมปกติ คือ 9,494,979 ธุรกรรม รวมธุรกรรมทั้งหมดเป็นจำนวน 18,989,958 ธุรกรรม ดังภาพที่ 3.18



ภาพที่ 3.18 ความสมดุลของข้อมูล หลังจากทำการ Oversampling

### 3.1.5 แบ่งชุดข้อมูล

ขั้นตอนสำคัญก่อนทำการสร้างแบบจำลอง คือการแบ่งข้อมูลออกเป็น 2 ชุด ชุดข้อมูลที่ใช้สำหรับฝึกแบบจำลอง (Training dataset) และชุดข้อมูลที่ใช้สำหรับทดสอบการทำนายของแบบจำลอง (Test dataset) โดยจะทำการแบ่งข้อมูลแบบ 70% สำหรับชุดข้อมูลที่ใช้ฝึกแบบจำลอง และแบ่ง 30% สำหรับชุดข้อมูลที่ใช้ทดสอบแบบจำลอง

โดยชุดข้อมูลที่มีการจัดการกับปัญหาข้อมูลไม่สมดุล ด้วยการทำ Undersampling จะแบ่งข้อมูลได้เป็น ชุดข้อมูลที่ใช้ฝึกแบบจำลองจำนวน 13,822 ธุรกรรม และแบ่งสำหรับชุดข้อมูลที่ใช้ทดสอบแบบจำลอง จำนวน 5,924 ธุรกรรมดังตารางที่ 3.2

ตารางที่ 3.2 การแบ่งชุดข้อมูลที่ผ่านมาการทำ Undersampling

ชุดข้อมูล	การแบ่ง	จำนวนข้อมูล
Training dataset	70 %	13,822
Test dataset	30 %	5,924

และในส่วนของคุณข้อมูลที่มีการจัดการกับปัญหาข้อมูลไม่สมดุล ด้วยการทำ Oversampling จะแบ่งข้อมูลได้เป็น ชุดข้อมูลที่ใช้ฝึกแบบจำลองจำนวน 13,292,970 รายการ และแบ่งสำหรับชุดข้อมูลที่ใช้ทดสอบแบบจำลอง จำนวน 5,696,988 รายการดังตารางที่ 3.3

ตารางที่ 3.3 การแบ่งชุดข้อมูลผ่านการทำ Oversampling

ชุดข้อมูล	การแบ่ง	จำนวนข้อมูล
Training dataset	70 %	13,292,970
Test dataset	30 %	5,696,988

### 3.1.6 สร้างแบบจำลองตรวจจับการฟอกเงิน

งานวิจัยนี้ ทำการสร้างแบบจำลองการเรียนรู้ของเครื่องเพื่อใช้ตรวจจับธุรกรรมการฟอกเงิน โดยมีการเลือกใช้แบบจำลองทั้งหมด 2 แบบ และมีการทดลองใช้วิธีการจัดการกับปัญหาข้อมูลไม่สมดุลทั้งการทำ Undersampling และ Oversampling เพื่อนำมาเปรียบเทียบประสิทธิภาพในการตรวจจับการฟอกเงิน โดยแบบจำลองที่เลือกใช้ มีดังนี้

1. XGBoost
2. LightGBM

#### 3.1.6.1 ปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning)

การปรับแต่งไฮเปอร์พารามิเตอร์นั้น มีจุดประสงค์เพื่อหาค่าที่เหมาะสมที่สุดของพารามิเตอร์ต่างๆ ในแบบจำลอง เพื่อให้แบบจำลองมีประสิทธิภาพสูงสุดในการจำแนกธุรกรรมปกติกับธุรกรรมที่เป็นการฟอกเงิน โดยในแต่ละแบบจำลองจะใช้พารามิเตอร์ในการปรับต่างกัน ขึ้นอยู่กับประเภทของแบบจำลองนั้น และมีการใช้ค่าหลายตัวสำหรับแต่ละพารามิเตอร์เช่นกัน เพื่อให้การปรับแต่งนี้ได้ทดลองใช้พารามิเตอร์หลายชุด เพื่อหาชุดที่ดีที่สุด

ซึ่งงานวิจัยนี้ จะเลือกใช้วิธี Grid Search ในการปรับแต่งไฮเปอร์พารามิเตอร์ คือการค้นหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมโดยการแบ่งช่วงของแต่ละพารามิเตอร์เป็นตาราง (Grid) หรือเป็นการใส่ค่าต่างๆ ในพารามิเตอร์แต่ละตัว จากนั้นทำการวนทดลองค่าทั้งหมดที่กำหนดไว้ในตาราง ดังตารางที่ 3.4 ถึง 3.5 เพื่อหาค่าที่ดีที่สุดที่ทำให้แบบจำลองมีประสิทธิภาพที่สุด

ตารางที่ 3.4 ค่าที่ใช้ปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost

param_grid	
max_depth	3, 5, 10, 20
eta	0.01, 0.1, 0.3, 0.5
grid_search	
estimator	XGBClassifier(eval_metric='logloss')
param_grid	param_grid
scoring	accuracy
cv	3
verbose	2
n_jobs	-1

ตารางที่ 3.5 ค่าที่ใช้ปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM

param_grid	
num_leaves	5, 10, 20
learning_rate	0.3, 0.5
max_depth	3, 5, 10, 20
n_estimators	50
grid_search	
estimator	LGBMClassifier(boosting_type='gbdt', objective='binary')
param_grid	param_grid
scoring	accuracy
cv	3
verbose	2
n_jobs	-1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.7 ประเมินประสิทธิภาพของแบบจำลอง

วิธีการประเมินและเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลองนั้น จะใช้ตัววัด คือ ตาราง Confusion matrix เพื่อเปรียบเทียบข้อมูลจริงและข้อมูลที่เป็นผลลัพธ์จากการทำนายของแบบจำลอง โดยจะนำผลการทำนายมาคำนวณหา ค่าความแม่นยำ (Accuracy), ค่าความเที่ยง (Precision) ค่าความไว (Recall หรือ Sensitivity) และ ค่า F1-Score จากนั้นจึงนำผลการคำนวณมาวิเคราะห์และทำการเปรียบเทียบประสิทธิภาพของแบบจำลอง เพื่อหาแบบจำลองที่ดีที่สุด

### 3.2 เครื่องมือที่ใช้ในการวิจัย

1. Google Colab Pro+ ใช้สำหรับเขียน python
2. Power BI Desktop ใช้สำหรับแสดงกราฟเพื่อวิเคราะห์ข้อมูล
3. Python ใช้สำหรับจัดการข้อมูล และสร้างแบบจำลอง



## บทที่ 4

### ผลการวิจัยและการอภิปรายผล

ในบทนี้จะกล่าวถึงผลการดำเนินงานจากการสร้างแบบจำลองการเรียนรู้ของเครื่องเพื่อตรวจจับธุรกรรมการฟอกเงิน ซึ่งมีการจัดการปัญหาข้อมูลไม่สมดุลทั้งการทำ Undersampling กับ Oversampling และทำการอภิปรายผลลัพธ์ที่ได้จากการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 2 วิธี XGBoost และ LightGBM โดยจะนำผลจากตาราง Confusion Matrix อัตราการทำนายของแบบจำลอง ค่า Accuracy Precision Recall และ F1-Score ที่ได้จากการคำนวณด้วยผลในตาราง Confusion Matrix มาเป็นเครื่องมือในการประเมินประสิทธิภาพของแบบจำลอง

#### 4.1 ผลลัพธ์ของแบบจำลอง XGBoost

จากการปรับแต่งไฮเปอร์พารามิเตอร์ ด้วยการทดลองใส่ค่าพารามิเตอร์ต่างๆ ใน Grid Search เพื่อให้แบบจำลองทำการวนลูปหาชุดของพารามิเตอร์ที่เหมาะสม และมีประสิทธิภาพมากที่สุด โดยแบบจำลอง XGBoost มีการปรับแต่งไฮเปอร์พารามิเตอร์ทั้งหมด 2 ตัว คือ eta และ max\_depth นอกจากนี้เพื่อเพิ่มประสิทธิภาพของการเรียนรู้ จึงทำการตั้งค่าให้แบบจำลองทำ Cross validation จากนั้น วัดผลการเรียนรู้ด้วยค่า Accuracy ซึ่งจากการวนลูปของ Grid Search นั้น แบบจำลองจะทำการวนลูปพารามิเตอร์ ทั้งหมด 16 ชุด ชุดละ 3 รอบ รวมเป็นทั้งหมด 48 รอบ ดังตารางที่ 4.1

ตารางที่ 4.1 การจับคู่ของพารามิเตอร์ทั้งหมดในตาราง Grid Search ของแบบจำลอง XGBoost

การจับคู่พารามิเตอร์ใน Grid Search ของแบบจำลอง XGBoost		
eta	max_depth	จำนวนรอบ (cv)
0.01	3	3
0.01	5	3
0.01	10	3
0.01	20	3
0.1	3	3
0.1	5	3
0.1	10	3
0.1	20	3
0.3	3	3
0.3	5	3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้เฉพาะในการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำหรือเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 การจับคู่ของพารามิเตอร์ทั้งหมดในตาราง Grid Search ของแบบจำลอง XGBoost (ต่อ)

การจับคู่พารามิเตอร์ใน Grid Search ของแบบจำลอง XGBoost		
eta	max_depth	จำนวนรอบ (cv)
0.3	10	3
0.3	20	3
0.5	3	3
0.5	5	3
0.5	10	3
0.5	20	3

#### 4.1.1 ผลลัพธ์ของแบบจำลอง XGBoost จากการทำ Undersampling

จากการจัดการข้อมูลไม่สมดุลด้วยการทำ Undersampling ผลลัพธ์จากการปรับแต่งไฮเปอร์พารามิเตอร์ พบว่าชุดของพารามิเตอร์ที่มีประสิทธิภาพที่สุดของแบบจำลอง XGBoost นี้ คือการใช้ eta หรือ learning rate เท่ากับ 0.1 และกำหนดความลึกของ tree ใน XGBoost ให้มีความลึกไม่เกิน 10 ระดับ โดยการปรับพารามิเตอร์ชุดนี้ ให้ค่า Accuracy จากการทดสอบรอบที่ 1 เท่ากับ 0.741753 การทดสอบรอบที่ 2 เท่ากับ 0.73041 และการทดสอบรอบที่ 3 เท่ากับ 0.739093 จากนั้นทำการหาค่าเฉลี่ยออกมาจะได้ค่า Accuracy เฉลี่ยจากทั้ง 3 รอบ เท่ากับ 0.737085 ซึ่งเป็นชุดของพารามิเตอร์ที่ให้ประสิทธิภาพในการตรวจจับธุรกรรมการฟอกเงินได้ดีที่สุดในพารามิเตอร์ทั้ง 16 ชุด ดังตารางที่ 4.2

ตารางที่ 4.2 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling

ผลลัพธ์จากแบบจำลอง XGBoost					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'eta': 0.1, 'max_depth': 10}	0.741753	0.73041	0.739093	0.737085	1
{'eta': 0.3, 'max_depth': 5}	0.740885	0.729108	0.736488	0.735494	2
{'eta': 0.3, 'max_depth': 20}	0.739149	0.724984	0.73931	0.734481	3
{'eta': 0.3, 'max_depth': 10}	0.741102	0.723681	0.738007	0.734264	4
{'eta': 0.1, 'max_depth': 20}	0.738064	0.725852	0.736922	0.733613	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานี้เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านอื่นๆ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling (ต่อ)

ผลลัพธ์จากแบบจำลอง XGBoost					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'eta': 0.5, 'max_depth': 3}	0.730252	0.730193	0.730844	0.73043	6
{'eta': 0.3, 'max_depth': 3}	0.732205	0.724115	0.732581	0.729634	7
{'eta': 0.1, 'max_depth': 5}	0.727865	0.726286	0.731496	0.728549	8
{'eta': 0.5, 'max_depth': 10}	0.72678	0.721945	0.733883	0.727536	9
{'eta': 0.5, 'max_depth': 20}	0.730252	0.721077	0.729325	0.726884	10
{'eta': 0.5, 'max_depth': 5}	0.725911	0.718038	0.728023	0.723991	11
{'eta': 0.01, 'max_depth': 10}	0.72092	0.71934	0.730193	0.723484	12
{'eta': 0.1, 'max_depth': 3}	0.722222	0.717387	0.726503	0.722037	13
{'eta': 0.01, 'max_depth': 20}	0.718533	0.709355	0.72303	0.716973	14
{'eta': 0.01, 'max_depth': 5}	0.705295	0.710224	0.72086	0.712126	15
{'eta': 0.01, 'max_depth': 3}	0.694444	0.697417	0.701107	0.697656	16

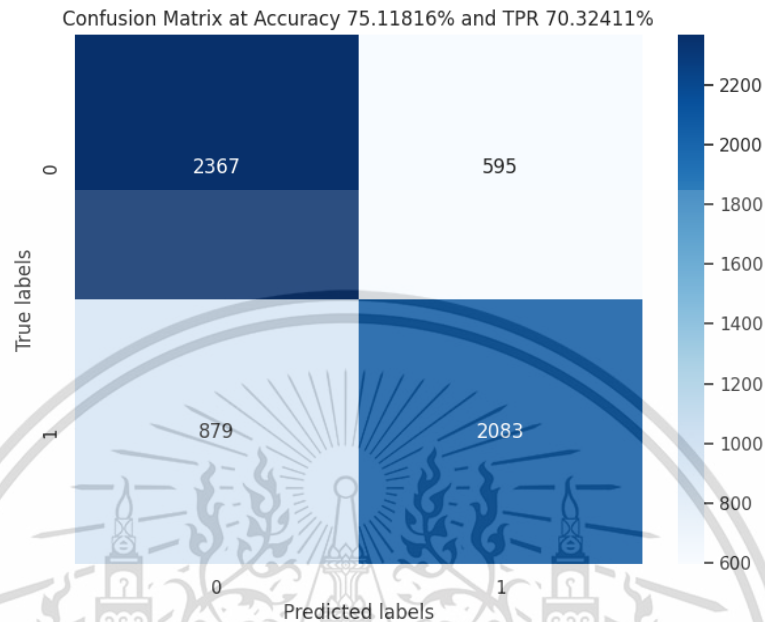
จากนั้นจะทำการตรวจสอบว่าแบบจำลองไม่พบปัญหา Overfitting โดยเปรียบเทียบค่า Accuracy ที่ได้จากชุดข้อมูลที่แบ่งไว้ 2 ชุด คือ ชุดข้อมูลสำหรับฝึกแบบจำลอง กับชุดข้อมูลสำหรับทดสอบ ซึ่งได้ผลดังนี้ ผลจากการทำนายด้วยชุดข้อมูลสำหรับฝึกแบบจำลองให้ค่า Accuracy เท่ากับ 0.9356 ส่วนผลจากการทำนายด้วยชุดข้อมูลทดสอบได้ค่า Accuracy เท่ากับ 0.7512 ดังตารางที่ 4.3 สรุปได้ว่าแบบจำลอง XGBoost พบปัญหา Overfitting อยู่เล็กน้อยเนื่องจากผลลัพธ์จากการทำนายด้วยชุดข้อมูลทั้ง 2 นั้น มีค่าที่ต่างกันถึง 18.44%

ตารางที่ 4.3 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling

ชุดข้อมูล	Accuracy
Training dataset	0.9356
Test dataset	0.7512

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากนำพารามิเตอร์ชุดที่ให้ประสิทธิภาพมากที่สุดมาตรวจจับธุรกรรมการฟอกเงิน และทำการวัดผลทำนายของแบบจำลอง XGBoost ด้วยตาราง Confusion Matrix ดังภาพที่ 4.1



ภาพที่ 4.1 ตาราง Confusion matrix ของแบบจำลอง XGBoost จากการทำ Undersampling

จากภาพที่ 4.1 แสดงผลการทำนายของแบบจำลอง XGBoost โดยการใช้การเทียบผลลัพธ์จากแบบจำลองกับผลเฉลยของชุดข้อมูลจริง โดยข้อมูลสำหรับใช้ทดสอบนี้มีจำนวน 5,924 ธุรกรรม ซึ่งแบ่งเป็นธุรกรรมปกติกับธุรกรรมที่เป็นการฟอกเงินในจำนวนเท่าๆ กัน ผลการทำนายนั้นพบว่าแบบจำลอง XGBoost ทำนายธุรกรรมปกติจริง ว่าเป็นธุรกรรมปกติได้ถูกต้องจำนวน 2,367 ธุรกรรม คิดเป็น 79.912% (TNR) และตรวจจับธุรกรรมที่เป็นการฟอกเงินได้ 2,083 ธุรกรรม ซึ่งคิดเป็น 70.324% (TPR) และผลการทำนายที่ไม่ถูกต้องของแบบจำลอง พบว่าธุรกรรมที่เป็นปกติแต่ถูกแบบจำลองทำนายว่าเป็นธุรกรรมการฟอกเงิน มีจำนวน 595 ธุรกรรม คิดเป็น 20.088% (FPR) และธุรกรรมการฟอกเงินที่แบบจำลองไม่สามารถตรวจจับได้มี 879 ธุรกรรม คิดเป็น 29.676% (FNR) ซึ่งถือเป็นจำนวนที่ไม่น้อย เมื่อเทียบกับจำนวนธุรกรรมทั้งหมดที่เป็นธุรกรรมฟอกเงิน ดังตารางที่ 4.4

**ตารางที่ 4.4** อัตราส่วนการทำนายของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ

Undersampling

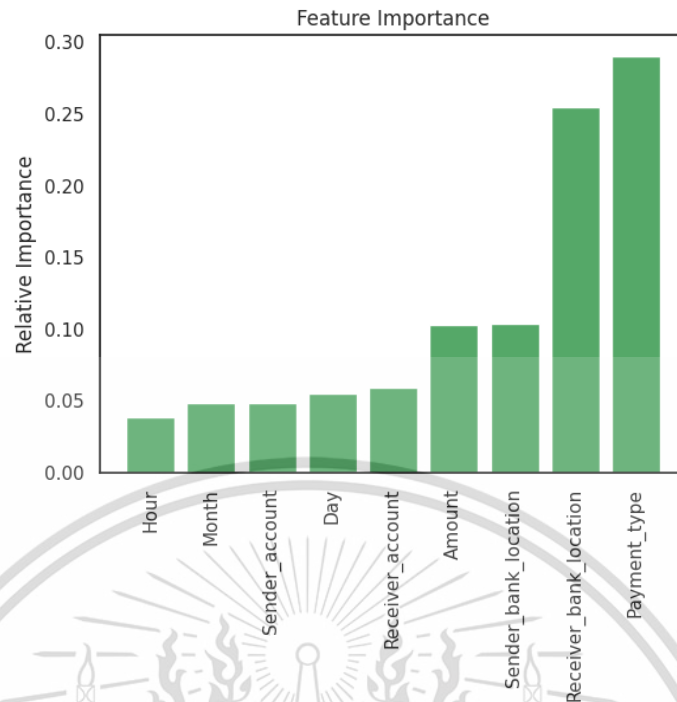
อัตราส่วนการทำนายของแบบจำลอง XGBoost	
False Positive Rate (FPR)	0.20088
True Positive Rate (TPR)	0.70324
True Negative Rate (TNR)	0.79912
False Negative Rate (FNR)	0.29676

จากการคำนวณค่าทางสถิติเพื่อประเมินประสิทธิภาพของแบบจำลอง XGBoost โดยคำนวณจากการทำนายของแบบจำลอง ได้ผลการประเมินดังนี้ ค่า Accuracy เท่ากับ 75.118% ค่า Precision เท่ากับ 77.781% ค่า Recall เท่ากับ 70.324% และค่า F1-Score เท่ากับ 73.865% ดังตารางที่ 4.5 ซึ่งจากค่าทางสถิติดังกล่าวจึงได้ข้อสรุปว่า แบบจำลอง XGBoost ที่มีการใช้ข้อมูลจากการทำ Undersampling สามารถตรวจจับธุรกรรมการฟอกเงินได้ค่อนข้างมีประสิทธิภาพ แต่ยังไม่มากเท่าที่ควร เนื่องจากยังคงสร้างผลบวกสูงกว่า 20% และพลาดการตรวจจับการฟอกเงินกว่า 29%

**ตารางที่ 4.5** ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Undersampling

ผลการประเมินประสิทธิภาพของแบบจำลอง XGBoost	
Accuracy	0.75118
Precision	0.77781
Recall	0.70324
F1-Score	0.73865

จากนั้นผู้วิจัยได้ทำการหาว่าคุณลักษณะใดของชุดข้อมูลจากการทำ Undersampling นี้ ที่มีผลต่อแบบจำลอง XGBoost พบว่าคุณลักษณะด้าน ประเภทของการชำระเงิน มีผลต่อแบบจำลองสูงที่สุด รองลงมาเป็นคุณลักษณะด้าน ที่ตั้งของธนาคารบัญชีปลายทาง และที่ตั้งของธนาคารบัญชีต้นทาง ตามลำดับ ดังภาพที่ 4.2



ภาพที่ 4.2 Feature Importance ของแบบจำลอง XGBoost จากการทำ Undersampling

#### 4.1.2 ผลลัพธ์ของแบบจำลอง XGBoost จากการทำ Oversampling

จากการจัดการข้อมูลไม่สมดุลด้วยการทำ Oversampling ผลลัพธ์จากการปรับแต่งไฮเปอร์พารามิเตอร์ พบว่าชุดของพารามิเตอร์ที่มีประสิทธิภาพที่สุดของแบบจำลอง XGBoost นี้ คือการใช้ eta หรือ learning rate เท่ากับ 0.5 และกำหนดความลึกของ tree ใน XGBoost ให้มีความลึกไม่เกิน 20 ระดับ โดยการปรับพารามิเตอร์ชุดนี้ ให้ค่า Accuracy จากการทดสอบรอบที่ 1 เท่ากับ 0.99898 การทดสอบรอบที่ 2 เท่ากับ 0.99897 และการทดสอบรอบที่ 3 เท่ากับ 0.99899 จากนั้นทำการหาค่าเฉลี่ยออกมาจะได้ค่า Accuracy เฉลี่ยจากทั้ง 3 รอบ เท่ากับ 0.99898 ซึ่งเป็นชุดของพารามิเตอร์ที่ให้ประสิทธิภาพในการตรวจจับธุรกรรมการฟอกเงินได้ดีที่สุดในพารามิเตอร์ทั้ง 16 ชุด ดังตารางที่ 4.6

ตารางที่ 4.6 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling

ผลลัพธ์จากแบบจำลอง XGBoost					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'eta': 0.5, 'max_depth': 20}	0.99898	0.99897	0.99899	0.99898	1
{'eta': 0.3, 'max_depth': 20}	0.99887	0.9989	0.99885	0.99887	2
{'eta': 0.1, 'max_depth': 20}	0.99359	0.99266	0.99287	0.99304	3

ตารางที่ 4.6 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling (ต่อ)

ผลลัพธ์จากแบบจำลอง XGBoost					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'eta': 0.5, 'max_depth': 10}	0.97523	0.97597	0.97563	0.97561	4
{'eta': 0.3, 'max_depth': 10}	0.95768	0.95813	0.95752	0.95778	5
{'eta': 0.01, 'max_depth': 20}	0.92683	0.92647	0.92682	0.92671	6
{'eta': 0.1, 'max_depth': 10}	0.90294	0.90639	0.90505	0.90479	7
{'eta': 0.5, 'max_depth': 5}	0.89292	0.89281	0.8933	0.89301	8
{'eta': 0.3, 'max_depth': 5}	0.88279	0.88047	0.88224	0.88183	9
{'eta': 0.5, 'max_depth': 3}	0.86913	0.86789	0.86773	0.86825	10
{'eta': 0.3, 'max_depth': 3}	0.86355	0.86156	0.86254	0.86255	11
{'eta': 0.1, 'max_depth': 5}	0.86326	0.86161	0.86118	0.86202	12
{'eta': 0.01, 'max_depth': 10}	0.85427	0.85367	0.85408	0.85401	13
{'eta': 0.1, 'max_depth': 3}	0.84528	0.84272	0.84263	0.84354	14
{'eta': 0.01, 'max_depth': 5}	0.8279	0.82716	0.82754	0.82753	15
{'eta': 0.01, 'max_depth': 3}	0.78061	0.77975	0.77987	0.78008	16

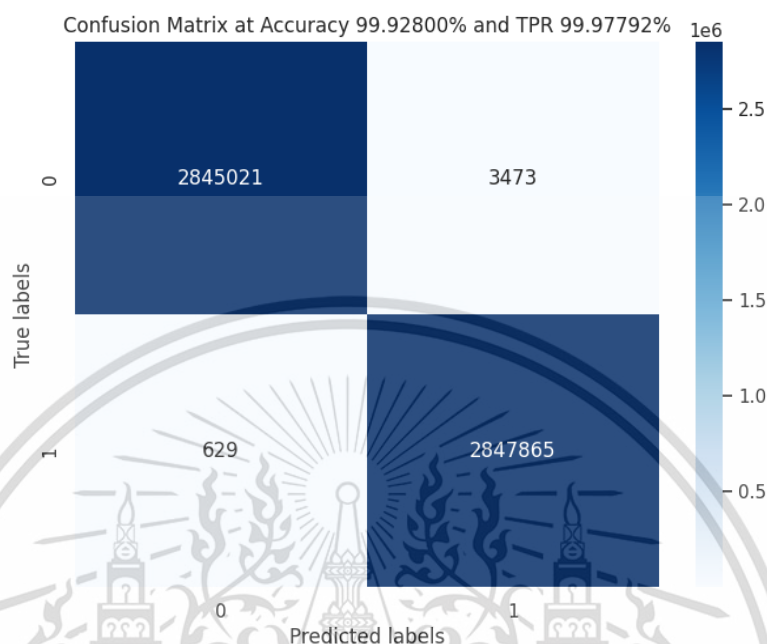
จากนั้นจะทำการตรวจสอบว่าแบบจำลองไม่พบปัญหา Overfitting โดยเปรียบเทียบค่า Accuracy ที่ได้จากชุดข้อมูลที่แบ่งไว้ 2 ชุด คือ ชุดข้อมูลสำหรับฝึกแบบจำลอง กับชุดข้อมูลสำหรับทดสอบ ซึ่งได้ผลดังนี้ ผลจากการทำนายด้วยชุดข้อมูลสำหรับฝึกแบบจำลองให้ค่า Accuracy เท่ากับ 0.9999 ส่วนผลจากการทำนายด้วยชุดข้อมูลทดสอบได้ค่า Accuracy เท่ากับ 0.9993 ดังตารางที่ 4.7 สรุปได้ว่าแบบจำลอง XGBoost ไม่พบปัญหา Overfitting เนื่องจากผลลัพธ์จากการทำนายด้วยชุดข้อมูลทั้ง 2 นั้น มีค่าที่ใกล้เคียงกันมาก ซึ่งให้ผลที่ต่างกันเพียง 0.0006

ตารางที่ 4.7 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling

ชุดข้อมูล	Accuracy
Training dataset	0.9999
Test dataset	0.9993

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากนำพารามิเตอร์ชุดที่ให้ประสิทธิภาพมากที่สุดมาตรวจจับธุรกรรมการฟอกเงิน และทำการวัดผลทำนายของแบบจำลอง XGBoost ด้วยตาราง Confusion Matrix ดังภาพที่ 4.3



ภาพที่ 4.3 ตาราง Confusion matrix ของแบบจำลอง XGBoost จากการทำ Oversampling

จากภาพที่ 4.3 แสดงผลการทำนายของแบบจำลอง XGBoost โดยใช้การเทียบผลลัพธ์จากแบบจำลองกับผลเฉลยของชุดข้อมูลจริง โดยข้อมูลสำหรับใช้ทดสอบนี้มีจำนวน 5,696,988 ธุรกรรม ซึ่งแบ่งเป็นธุรกรรมปกติกับธุรกรรมที่เป็นการฟอกเงินในจำนวนเท่าๆ กัน ผลการทำนายนั้นพบว่าแบบจำลอง XGBoost ทำนายธุรกรรมปกติจริง ว่าเป็นธุรกรรมปกติได้ถูกต้องจำนวน 2,845,021 ธุรกรรม คิดเป็น 99.878% (TNR) และตรวจจับธุรกรรมที่เป็นการฟอกเงินได้ถึง 2,847,865 ธุรกรรม ซึ่งคิดเป็น 99.978% (TPR) และผลการทำนายที่ไม่ถูกต้องของแบบจำลอง พบว่าธุรกรรมที่เป็นปกติ แต่ถูกแบบจำลองทำนายว่าเป็นธุรกรรมการฟอกเงิน มีจำนวน 3,473 ธุรกรรม คิดเป็น 0.122% (FPR) และธุรกรรมการฟอกเงินที่แบบจำลองไม่สามารถตรวจจับได้มีเพียง 629 ธุรกรรม คิดเป็น 0.022% (FNR) ซึ่งเป็นจำนวนที่น้อยมากๆ เมื่อเทียบกับจำนวนธุรกรรมทั้งหมดที่เป็นธุรกรรมฟอกเงิน ดังตารางที่ 4.8

**ตารางที่ 4.8** อัตราส่วนการทำนายของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling

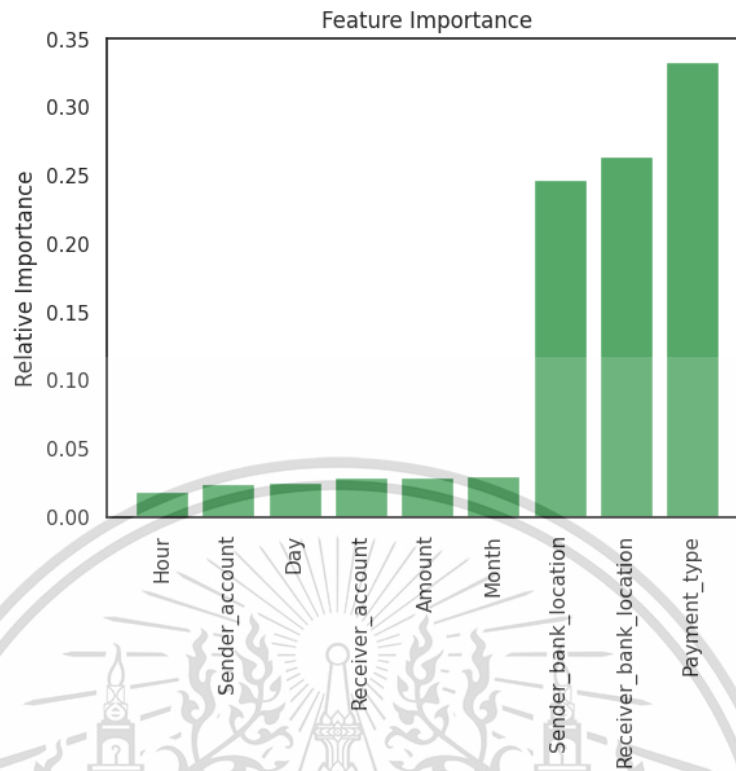
อัตราส่วนการทำนายของแบบจำลอง XGBoost	
False Positive Rate (FPR)	0.00122
True Positive Rate (TPR)	0.99978
True Negative Rate (TNR)	0.99878
False Negative Rate (FNR)	0.00022

จากการคำนวณค่าทางสถิติเพื่อประเมินประสิทธิภาพของแบบจำลอง XGBoost โดยคำนวณจากการทำนายของแบบจำลอง ได้ผลการประเมินดังนี้ ค่า Accuracy เท่ากับ 99.927% ค่า Precision เท่ากับ 99.878% ค่า Recall เท่ากับ 99.977% และค่า F1-Score เท่ากับ 99.928% ดังตารางที่ 4.9 ซึ่งจากค่าทางสถิติดังกล่าวจึงได้ข้อสรุปว่า แบบจำลอง XGBoost มีการใช้ข้อมูลจากการทำ Oversampling สามารถตรวจจับธุรกรรมการฟอกเงิน ได้อย่างมีประสิทธิภาพมาก และให้ผลการทำนายที่แม่นยำกว่าการใช้ข้อมูลจากการทำ Undersampling

**ตารางที่ 4.9** ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง XGBoost โดยใช้ข้อมูลจากการทำ Oversampling

ผลการประเมินประสิทธิภาพของแบบจำลอง XGBoost	
Accuracy	0.99927
Precision	0.99878
Recall	0.99977
F1-Score	0.99928

จากนั้นผู้วิจัยได้ทำการหาว่าคุณลักษณะใดของชุดข้อมูลจากการทำ Oversampling นี้ ที่มีผลต่อแบบจำลอง XGBoost พบว่าคุณลักษณะด้าน ประเภทของการชำระเงิน มีผลต่อแบบจำลองสูงที่สุด รองลงมาเป็นคุณลักษณะด้าน ที่ตั้งของธนาคารบัญชีปลายทาง และที่ตั้งของธนาคารบัญชีต้นทาง ตามลำดับ โดยจะเห็นว่าคุณลักษณะ 3 รายการแรกของทั้งแบบจำลอง XGBoost ที่ใช้ข้อมูลจากการทำ Undersampling และ Oversampling นั้นมีแนวโน้มไปในทางเดียวกัน ดังภาพที่ 4.4



ภาพที่ 4.4 Feature Importance ของแบบจำลอง XGBoost จากการทำ Oversampling

## 4.2 ผลลัพธ์ของแบบจำลอง LightGBM

จากการปรับแต่งไฮเปอร์พารามิเตอร์ ด้วยการทดลองใส่ค่าพารามิเตอร์ต่างๆ ใน Grid Search เพื่อให้แบบจำลองทำการวนลูปหาชุดของพารามิเตอร์ที่เหมาะสม และมีประสิทธิภาพมากที่สุด โดยแบบจำลอง LightGBM มีการปรับแต่งไฮเปอร์พารามิเตอร์ทั้งหมด 4 ตัว คือ learning\_rate max\_depth n\_estimators และ num\_leaves นอกจากนั้นเพื่อเพิ่มประสิทธิภาพของการเรียนรู้ จึงทำการตั้งค่าให้แบบจำลองทำ Cross validation เพื่อหิยชุดข้อมูลคนละส่วนกัน มาฝึกซ้ำด้วยพารามิเตอร์ชุดเดียวกัน แล้วทำการวัดผลด้วยค่า Accuracy ที่ได้จากการฝึก ซึ่งจากการวนจับคู่ของ Grid Search นั้น แบบจำลองจะทำการวนลูปพารามิเตอร์ทั้งหมด 24 ชุด ชุดละ 3 รอบ รวมเป็นทั้งหมด 72 รอบ ดังตารางที่ 4.10

ตารางที่ 4.10 การจับคู่ของพารามิเตอร์ทั้งหมดในตาราง Grid Search ของแบบจำลอง LightGBM

การจับคู่พารามิเตอร์ใน Grid Search ของแบบจำลอง LightGBM				
learning_rate	max_depth	n_estimators	num_leaves	จำนวนรอบ (cv)
0.01	3	50	31	3
0.01	3	50	70	3
0.01	5	50	31	3
0.01	5	50	70	3
0.01	10	50	31	3
0.01	10	50	70	3
0.01	20	50	31	3
0.01	20	50	70	3
0.1	3	50	31	3
0.1	3	50	70	3
0.1	5	50	31	3
0.1	5	50	70	3
0.1	10	50	31	3
0.1	10	50	70	3
0.1	20	50	31	3
0.1	20	50	70	3
0.3	3	50	31	3
0.3	3	50	70	3
0.3	5	50	31	3
0.3	5	50	70	3
0.3	10	50	31	3
0.3	10	50	70	3
0.3	20	50	31	3
0.3	20	50	70	3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.2.1 ผลลัพธ์ของแบบจำลอง LightGBM จากการทำ Undersampling

จากการจัดการข้อมูลไม่สมดุลด้วยการทำ Undersampling ผลลัพธ์จากการปรับแต่งไฮเปอร์พารามิเตอร์ พบว่าชุดของพารามิเตอร์ที่มีประสิทธิภาพที่สุดของแบบจำลอง LightGBM นี้ คือการใช้ learning rate เท่ากับ 0.1 กำหนดความลึกของ tree ใน LightGBM ให้มีความลึกไม่เกิน 20 ระดับ กำหนดจำนวน tree ที่ 50 และกำหนดจำนวนใบสูงสุดที่ 70 โดยการปรับพารามิเตอร์ชุดนี้ ให้ค่า Accuracy จากการทดสอบรอบที่ 1 เท่ากับ 0.73828 การทดสอบรอบที่ 2 เท่ากับ 0.73345 และการทดสอบรอบที่ 3 เท่ากับ 0.7417 จากนั้นทำการหาค่าเฉลี่ยออกมาจะได้ค่า Accuracy เฉลี่ยจากทั้ง 3 รอบ เท่ากับ 0.73781 ซึ่งเป็นชุดของพารามิเตอร์ที่ให้ประสิทธิภาพในการตรวจจับธุรกรรมการฟอกเงินได้ดีที่สุดในพารามิเตอร์ทั้ง 24 ชุด ดังตารางที่ 4.11

ตารางที่ 4.11 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling

ผลลัพธ์จากแบบจำลอง LightGBM					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'learning_rate': 0.1, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.73828	0.73345	0.7417	0.73781	1
{'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.73633	0.72889	0.73627	0.73383	2
{'learning_rate': 0.1, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.73416	0.72542	0.74105	0.73354	3
{'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.73307	0.72737	0.73931	0.73325	4
{'learning_rate': 0.3, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.73481	0.72498	0.73236	0.73072	5
{'learning_rate': 0.3, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.7296	0.73106	0.72889	0.72985	6
{'learning_rate': 0.01, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.73177	0.71739	0.73128	0.72681	7
{'learning_rate': 0.3, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.72418	0.72194	0.73323	0.72645	8
{'learning_rate': 0.3, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.7309	0.72021	0.7265	0.72587	9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling (ต่อ)

ผลลัพธ์จากแบบจำลอง LightGBM					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'learning_rate': 0.3, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.72786	0.72477	0.72368	0.72544	10
{'learning_rate': 0.3, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.72786	0.72477	0.72368	0.72544	10
{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.72786	0.72129	0.72694	0.72537	12
{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 70}	0.72786	0.72129	0.72694	0.72537	12
{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.72135	0.72346	0.72759	0.72414	14
{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.72135	0.72346	0.72759	0.72414	14
{'learning_rate': 0.01, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.72027	0.71608	0.72954	0.72197	16
{'learning_rate': 0.01, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.72266	0.71305	0.72824	0.72131	17
{'learning_rate': 0.01, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.71875	0.71413	0.72737	0.72008	18
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.71072	0.71608	0.72346	0.71676	19
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 70}	0.71072	0.71608	0.72346	0.71676	19
{'learning_rate': 0.01, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.70378	0.70957	0.71522	0.70952	21
{'learning_rate': 0.01, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.70378	0.70957	0.71522	0.70952	21
{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.68034	0.69308	0.68678	0.68673	23

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้กับโรงเรียนเพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่หรือใช้ประโยชน์ใด ๆ

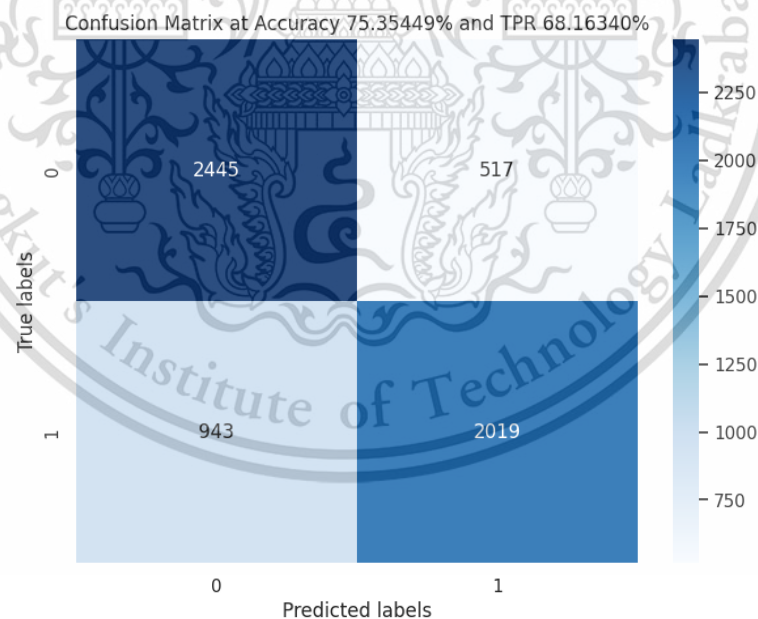
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นจะทำการตรวจสอบว่าแบบจำลองไม่พบปัญหา Overfitting โดยเปรียบเทียบค่า Accuracy ที่ได้จากชุดข้อมูลที่แบ่งไว้ 2 ชุด คือ ชุดข้อมูลสำหรับฝึกแบบจำลอง กับชุดข้อมูลสำหรับทดสอบ ซึ่งได้ผลดังนี้ ผลจากการทำนายด้วยชุดข้อมูลสำหรับฝึกแบบจำลองให้ค่า Accuracy เท่ากับ 0.8155 ส่วนผลจากการทำนายด้วยชุดข้อมูลทดสอบได้ค่า Accuracy เท่ากับ 0.7535 ดังตารางที่ 4.12 สรุปได้ว่าแบบจำลอง LightGBM ไม่พบปัญหา Overfitting เนื่องจากผลลัพธ์จากการทำนายด้วยชุดข้อมูลทั้ง 2 นั้น มีค่าที่ใกล้เคียงกัน ซึ่งให้ผลที่ต่างกันเพียง 6%

ตารางที่ 4.12 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling

ชุดข้อมูล	Accuracy
Training dataset	0.8155
Test dataset	0.7535

หลังจากนำพารามิเตอร์ชุดที่ให้ประสิทธิภาพมากที่สุดมาตรวจจับธุรกรรมการฟอกเงิน และทำการวัดผลทำนายของแบบจำลอง LightGBM ด้วยตาราง Confusion Matrix ดังภาพที่ 4.5



ภาพที่ 4.5 ตาราง Confusion matrix ของแบบจำลอง LightGBM จากการทำ Undersampling

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากภาพที่ 4.5 แสดงผลการทำนายของแบบจำลอง LightGBM โดยใช้การเทียบผลลัพธ์จากแบบจำลองกับผลเฉลยของชุดข้อมูลจริง โดยข้อมูลสำหรับใช้ทดสอบนี้มีจำนวน 5,924 ธุรกรรม ผลการทำนายนั้นพบว่า แบบจำลอง LightGBM ทำนายธุรกรรมปกติจริง ว่าเป็นธุรกรรมปกติได้ถูกต้องจำนวน 2,445 ธุรกรรม คิดเป็น 82.546% (TNR) และตรวจจับธุรกรรมที่เป็นการฟอกเงินได้ 2,019 ธุรกรรม ซึ่งคิดเป็น 68.163% (TPR) และผลการทำนายที่ไม่ถูกต้องของแบบจำลอง พบว่าธุรกรรมที่เป็นปกติแต่ถูกแบบจำลองทำนายพลาดว่าเป็นการฟอกเงิน มีจำนวน 517 ธุรกรรม คิดเป็น 17.454% (FPR) และธุรกรรมการฟอกเงินที่แบบจำลองไม่สามารถตรวจพบได้มี 943 ธุรกรรม คิดเป็น 31.837% (FNR) ถือว่าเป็นตรวจจับการฟอกเงินผิดพลาดเป็นจำนวนที่ค่อนข้างเยอะ เมื่อเทียบกับจำนวนธุรกรรมทั้งหมดที่เป็นธุรกรรมฟอกเงิน ดังตารางที่ 4.13

ตารางที่ 4.13 อัตราส่วนการทำนายของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ

Undersampling

อัตราส่วนการทำนายของแบบจำลอง LightGBM	
False Positive Rate (FPR)	0.17454
True Positive Rate (TPR)	0.68163
True Negative Rate (TNR)	0.82546
False Negative Rate (FNR)	0.31837

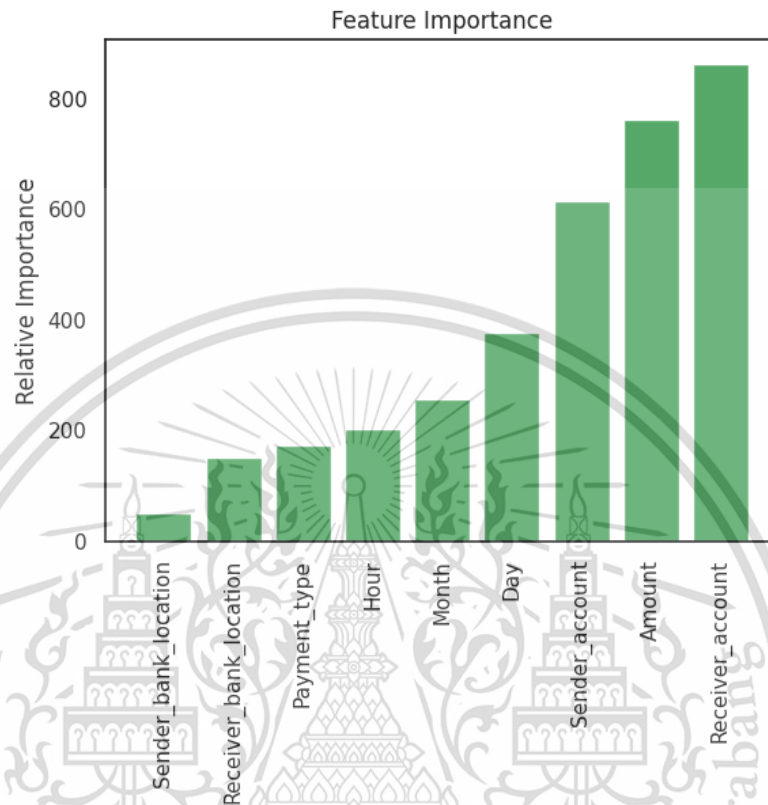
จากการคำนวณค่าทางสถิติเพื่อประเมินประสิทธิภาพของแบบจำลอง LightGBM โดยคำนวณจากการทำนายของแบบจำลอง ได้ผลการประเมินดังนี้ ค่า Accuracy เท่ากับ 75.354% ค่า Precision เท่ากับ 79.613% ค่า Recall เท่ากับ 68.163% และค่า F1-Score เท่ากับ 73.444% ดังตารางที่ 4.14 ซึ่งจากค่าทางสถิติดังกล่าวจึงได้ข้อสรุปว่า แบบจำลอง LightGBM ที่มีการใช้ข้อมูลจากการทำ Undersampling สามารถตรวจจับธุรกรรมการฟอกเงินได้ค่อนข้างมีประสิทธิภาพใกล้เคียงกับแบบจำลอง XGBoost ที่มีการใช้ข้อมูลจากการทำ Undersampling แต่ก็ยังคงไม่มีประสิทธิภาพมากเท่าที่ควร เนื่องจากพลาดการตรวจจับการฟอกเงินกว่า 31%

ตารางที่ 4.14 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Undersampling

ผลการประเมินประสิทธิภาพของแบบจำลอง LightGBM	
Accuracy	0.75354
Precision	0.79613
Recall	0.68163
F1-Score	0.73444

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นผู้วิจัยได้ทำการหาว่าคุณลักษณะใดของชุดข้อมูลจากการทำ Undersampling นี้ ที่มีผลต่อแบบจำลอง LightGBM ซึ่งพบว่าคุณลักษณะด้าน บัญชีปลายทาง จำนวนเงิน และ บัญชีต้นทาง มีผลต่อแบบจำลองสูงที่สุดตามลำดับ ดังภาพที่ 4.6



ภาพที่ 4.6 Feature Importance ของแบบจำลอง LightGBM จากการทำ Undersampling

#### 4.2.2 ผลลัพธ์ของแบบจำลอง LightGBM จากการทำ Oversampling

จากการจัดการข้อมูลไม่สมดุลด้วยการทำ Oversampling ผลลัพธ์จากการปรับแต่งไฮเปอร์พารามิเตอร์ พบว่าชุดของพารามิเตอร์ที่มีประสิทธิภาพที่สุดของแบบจำลอง LightGBM นี้ คือการใช้ learning rate เท่ากับ 0.3 กำหนดความลึกของ tree ใน LightGBM ให้มีความลึกไม่เกิน 20 ระดับ กำหนดจำนวน tree ที่ 50 และกำหนดจำนวนใบสูงสุดที่ 70 โดยการปรับพารามิเตอร์ชุดนี้ ให้ค่า Accuracy จากการทดสอบรอบที่ 1 เท่ากับ 0.89623 การทดสอบรอบที่ 2 เท่ากับ 0.89645 และการทดสอบรอบที่ 3 เท่ากับ 0.89615 จากนั้นทำการหาค่าเฉลี่ยออกมาจะได้ค่า Accuracy เฉลี่ยจากทั้ง 3 รอบ เท่ากับ 0.89627 ซึ่งเป็นชุดของพารามิเตอร์ที่ให้ประสิทธิภาพในการตรวจจับธุรกรรมการฟอกเงินได้ดีที่สุดในพารามิเตอร์ทั้ง 24 ชุด ดังตารางที่ 4.15

ตารางที่ 4.15 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling

ผลลัพธ์จากแบบจำลอง LightGBM					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'learning_rate': 0.3, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.89623	0.89645	0.89615	0.89627	1
{'learning_rate': 0.3, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.89239	0.89288	0.89361	0.89296	2
{'learning_rate': 0.3, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.88212	0.8809	0.88092	0.88131	3
{'learning_rate': 0.3, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.88008	0.87828	0.87901	0.87912	4
{'learning_rate': 0.1, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.87625	0.87536	0.87508	0.87557	5
{'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.87049	0.87064	0.87054	0.87056	6
{'learning_rate': 0.3, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.86986	0.87061	0.86861	0.8697	7
{'learning_rate': 0.3, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.86951	0.87051	0.8685	0.86951	8
{'learning_rate': 0.1, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.86322	0.8632	0.86312	0.86318	9
{'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.86295	0.86236	0.86129	0.8622	10
{'learning_rate': 0.01, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 70}	0.85224	0.85175	0.85203	0.85201	10
{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.85212	0.85135	0.85072	0.8514	12
{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 70}	0.85212	0.85135	0.85072	0.8514	12
{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.85215	0.84789	0.85051	0.85018	14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์อื่นใด

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 ผลลัพธ์ที่ได้จากการปรับแต่งไฮเปอร์พารามิเตอร์ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling (ต่อ)

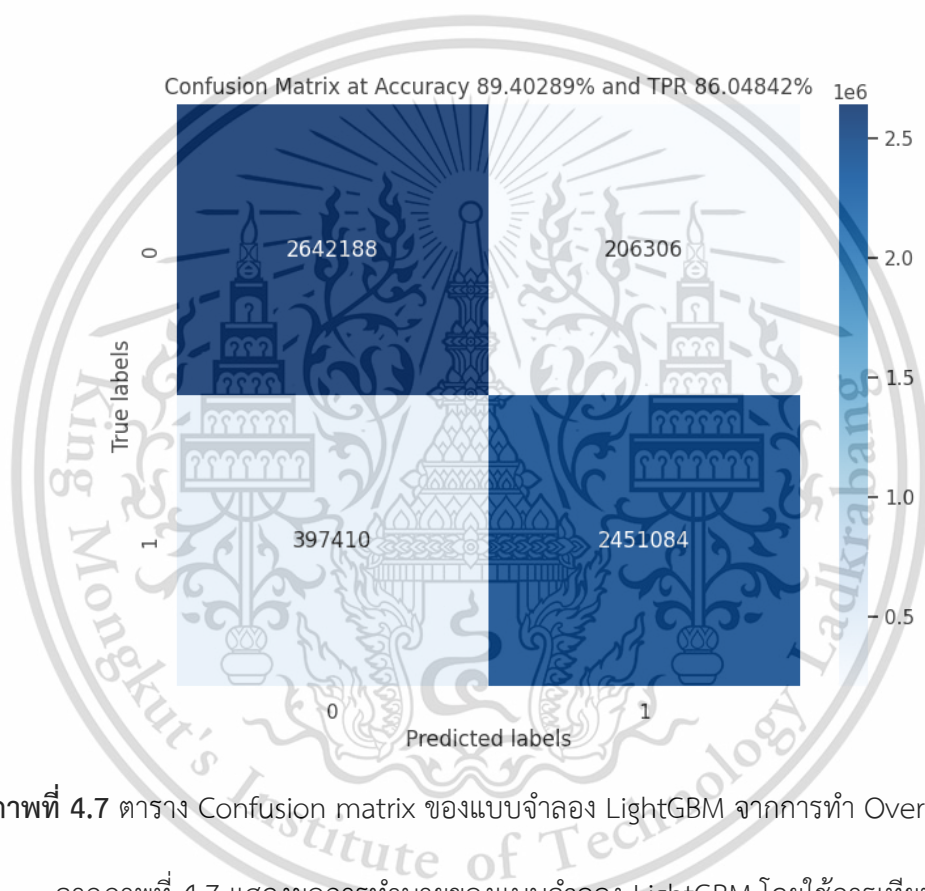
ผลลัพธ์จากแบบจำลอง LightGBM					
พารามิเตอร์	รอบที่ 1	รอบที่ 2	รอบที่ 3	ค่าเฉลี่ย	Rank
{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.85077	0.84785	0.85051	0.84971	14
{'learning_rate': 0.01, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 70}	0.84643	0.84521	0.84544	0.8457	16
{'learning_rate': 0.01, 'max_depth': 20, 'n_estimators': 50, 'num_leaves': 31}	0.83884	0.83844	0.8385	0.83859	17
{'learning_rate': 0.01, 'max_depth': 10, 'n_estimators': 50, 'num_leaves': 31}	0.83853	0.83786	0.83809	0.83816	18
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.83119	0.82566	0.82357	0.82681	19
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 70}	0.83119	0.82566	0.82357	0.82681	19
{'learning_rate': 0.01, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 31}	0.8248	0.82467	0.82507	0.82485	21
{'learning_rate': 0.01, 'max_depth': 5, 'n_estimators': 50, 'num_leaves': 70}	0.8248	0.82467	0.82507	0.82485	21
{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 31}	0.78069	0.77942	0.77945	0.77985	23
{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 50, 'num_leaves': 70}	0.78069	0.77942	0.77945	0.77985	23

จากนั้นจะทำการตรวจสอบว่าแบบจำลองไม่พบปัญหา Overfitting โดยเปรียบเทียบค่า Accuracy ที่ได้จากชุดข้อมูลที่แบ่งไว้ 2 ชุด คือ ชุดข้อมูลสำหรับฝึกแบบจำลอง กับชุดข้อมูลสำหรับทดสอบ ซึ่งได้ผลดังนี้ ผลจากการทำนายด้วยชุดข้อมูลสำหรับฝึกแบบจำลองให้ค่า Accuracy เท่ากับ 0.8944 ส่วนผลจากการทำนายด้วยชุดข้อมูลทดสอบได้ค่า Accuracy เท่ากับ 0.894 ดังตารางที่ 4.16 สรุปได้ว่าแบบจำลอง LightGBM ไม่พบปัญหา Overfitting เนื่องจากผลลัพธ์จากการทำนายด้วยชุดข้อมูลทั้ง 2 นั้น มีค่าที่ใกล้เคียงกันมาก โดยต่างกันเพียง 0.04%

ตารางที่ 4.16 ค่า Accuracy ของชุดข้อมูลสำหรับฝึกหัด เทียบกับชุดข้อมูลทดสอบของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling

ชุดข้อมูล	Accuracy
Training dataset	0.8944
Test dataset	0.894

หลังจากนำพารามิเตอร์ชุดที่ให้ประสิทธิภาพมากที่สุดมาตรวจจับธุรกรรมการฟอกเงิน และทำการวัดผลทำนายของแบบจำลอง LightGBM ด้วยตาราง Confusion Matrix ดังภาพที่ 4.7



ภาพที่ 4.7 ตาราง Confusion matrix ของแบบจำลอง LightGBM จากการทำ Oversampling

จากภาพที่ 4.7 แสดงผลการทำนายของแบบจำลอง LightGBM โดยการใช้การเทียบผลลัพธ์จากแบบจำลองกับผลเฉลยของชุดข้อมูลจริง โดยข้อมูลสำหรับใช้ทดสอบนี้มีจำนวน 5,696,988 ธุรกรรม ผลการทำนายนั้นพบว่า แบบจำลอง LightGBM ทำนายธุรกรรมปกติจริง ว่าเป็นธุรกรรมปกติได้ถูกต้องจำนวน 2,642,188 ธุรกรรม คิดเป็น 92.757% (TNR) และตรวจจับธุรกรรมที่เป็นการฟอกเงินได้ถึง 2,451,084 ธุรกรรม ซึ่งคิดเป็น 86.048% (TPR) และผลการทำนายที่ไม่ถูกต้องของแบบจำลองพบว่าธุรกรรมที่เป็นปกติแต่ถูกแบบจำลองทำนายว่าเป็นธุรกรรมการฟอกเงิน มีจำนวน 206,306 ธุรกรรม คิดเป็นเพียง 7.243% (FPR) และธุรกรรมการฟอกเงินที่แบบจำลองไม่สามารถตรวจจับได้มี 397,410 ธุรกรรม คิดเป็น 13.952% (FNR) ดังตารางที่ 4.17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 อัตราส่วนการทำนายของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling

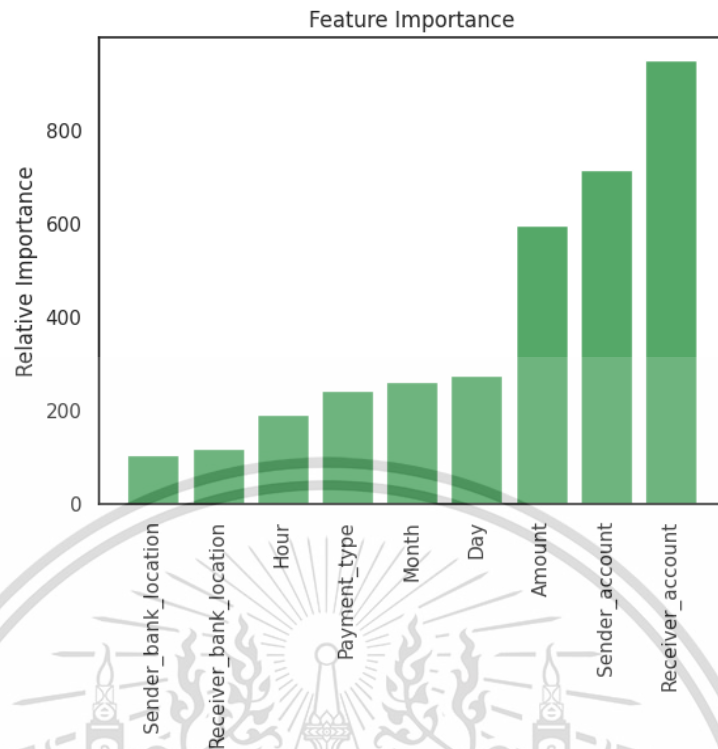
อัตราส่วนการทำนายของแบบจำลอง LightGBM	
False Positive Rate (FPR)	0.07243
True Positive Rate (TPR)	0.86048
True Negative Rate (TNR)	0.92757
False Negative Rate (FNR)	0.13952

จากการคำนวณค่าทางสถิติเพื่อประเมินประสิทธิภาพของแบบจำลอง LightGBM โดยคำนวณจากการทำนายของแบบจำลอง ได้ผลการประเมินดังนี้ ค่า Accuracy เท่ากับ 89.402% ค่า Precision เท่ากับ 92.236% ค่า Recall เท่ากับ 86.048% และค่า F1-Score เท่ากับ 89.035% ดังตารางที่ 4.18 ซึ่งจากค่าทางสถิติดังกล่าวจึงได้ข้อสรุปว่า แบบจำลอง LightGBM มีการใช้ข้อมูลจากการทำ Oversampling สามารถตรวจจับธุรกรรมการฟอกเงินได้อย่างมีประสิทธิภาพ มากกว่าการใช้ข้อมูลจากการทำ Undersampling

ตารางที่ 4.18 ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลอง LightGBM โดยใช้ข้อมูลจากการทำ Oversampling

ผลการประเมินประสิทธิภาพของแบบจำลอง LightGBM	
Accuracy	0.89402
Precision	0.92236
Recall	0.86048
F1-Score	0.89035

จากนั้นผู้วิจัยได้ทำการหาว่าคุณลักษณะใดของชุดข้อมูลจากการทำ Oversampling นี้ ที่มีผลต่อแบบจำลอง LightGBM พบว่าคุณลักษณะด้าน บัญชีปลายทาง บัญชีต้นทาง และ จำนวนเงิน มีผลต่อแบบจำลองสูงที่สุดตามลำดับ โดยจะเห็นว่าคุณลักษณะ 3 รายการแรกของทั้งแบบจำลอง LightGBM ที่ใช้ข้อมูลจากการทำ Undersampling และ Oversampling นั้นมีแนวโน้มไปในทางเดียวกัน ดังภาพที่ 4.8



ภาพที่ 4.8 Feature Importance ของแบบจำลอง LightGBM จากการทำ Oversampling

#### 4.3 เปรียบเทียบประสิทธิภาพของทั้ง 2 แบบจำลอง

จากการสร้างแบบจำลองที่ใช้ตรวจจับธุรกรรมการฟอกเงิน ด้วยแบบจำลองการเรียนรู้ของเครื่อง 2 วิธี ได้แก่ XGBoost และ LightGBM ซึ่งมีการใช้วิธีการจัดการกับข้อมูลไม่สมดุล 2 แบบเพื่อเปรียบเทียบกัน รวมถึงการทำ Hyperparameter Tuning ด้วยวิธี Grid Search เพื่อหาพารามิเตอร์ที่ทำให้แบบจำลองมีประสิทธิภาพมากที่สุด โดยขอสรุปผลแต่ละแบบจำลองดังนี้

- แบบจำลองที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Undersampling มีประสิทธิภาพที่ต่ำที่สุด โดยแบบจำลอง XGBoost มีค่า Accuracy อยู่ที่ 75.11% Precision 77.78% Recall 70.32% และ F1-Score 73.86% ในขณะที่แบบจำลอง LightGBM มีค่า Accuracy อยู่ที่ 75.35% Precision 79.61% Recall 68.16% และ F1-Score 73.44%
- แบบจำลอง LightGBM ที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Oversampling แสดงผลการทำนายที่ค่อนข้างดี โดยมีค่า Accuracy 89.40% Precision 92.24% Recall 86.05% และ F1-Score 89.03%
- แบบจำลอง XGBoost ที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Oversampling ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า Accuracy สูงมากถึง 99.93% Precision 99.88% Recall ประมาณ 99.98% และ F1-Score ประมาณ 99.93%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งหากเทียบด้วยตาราง แบบจำลองที่มีค่า Accuracy Precision Recall และ F1-Score สูงที่สุดคือ แบบจำลอง XGBoost ที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Oversampling ในขณะที่ทั้งสองแบบจำลองที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Undersampling มีประสิทธิภาพต่ำที่สุด ดังตารางที่ 4.19

**ตารางที่ 4.19** ผลการเปรียบเทียบ ค่า Accuracy Precision Recall และ F1-Score ของแบบจำลองทั้งหมด

Model	Method	Accuracy	Precision	Recall	F1-Score
XGBoost	Undersampling	75.1181634	77.78192681	70.32410533	73.86524823
LightGBM		75.3544902	79.61356467	68.16340311	73.44488905
XGBoost	Oversampling	99.927997	99.87819753	99.97791816	99.92803297
LightGBM		89.4028915	92.23651779	86.04841716	89.03507593

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

การสรุปผลวิจัยในบทนี้ จะทำการสรุปผลการดำเนินงานทั้งหมดที่ใช้ในการสร้างแบบจำลอง การเรียนรู้ของเครื่องเพื่อทำการตรวจจับธุรกรรมการฟอกเงิน พร้อมทั้งสรุปผลจากการเปรียบเทียบ ประสิทธิภาพของแบบจำลองทั้ง 2 ดังนี้ XGBoost และ LightGBM ซึ่งแบบจำลองทั้งสองนี้เป็น แบบจำลองประเภท Ensemble learning คือการนำแบบจำลอง Decision Tree มาฝึกต่อๆ กัน หลายรอบ โดยที่แต่ละ Decision Tree จะเรียนรู้จากข้อผิดพลาดของ Tree ก่อนหน้า จากนั้นนำ ข้อผิดพลาดมาปรับปรุง Tree ถัดๆ ไป และแบบจำลองจะทำการโหวตเลือกผลเฉลยที่แต่ละ Tree ได้ ผลลัพธ์ตรงกันมากที่สุด ซึ่งทำให้ XGBoost และ LightGBM นั้นเป็นแบบจำลองที่มีประสิทธิภาพกว่า แบบจำลองทั่วไป

โดยเริ่มจากผลการศึกษาระบบตรวจจับการฟอกเงินของธนาคาร พบว่าปกติแล้วธนาคารจะมีการสร้าง Workflow Scenario เพื่อใช้ตรวจจับธุรกรรมที่บ่งชี้ว่าอาจจะเป็นการฟอกเงิน ซึ่งระบบจะทำการแจ้งเตือนให้กับธนาคาร และจากนั้นเจ้าหน้าที่ปฏิบัติตามข้อกำหนดจะทำการตรวจสอบ ธุรกรรม บุคคลที่ทำธุรกรรมนั้น และบุคคลที่มีความเกี่ยวข้อง เพื่อส่งรายงานให้กับสำนักงานป้องกัน และปราบปรามการฟอกเงิน ด้วยเหตุนี้ผู้วิจัยจึงเห็นประโยชน์จากการนำเทคโนโลยีการเรียนรู้ของ เครื่องมาประยุกต์ใช้ในการตรวจจับธุรกรรมการฟอกเงิน เพื่อเพิ่มประสิทธิภาพในการตรวจจับ และ ลดการทำงานของเจ้าหน้าที่ปฏิบัติตามข้อกำหนด

ซึ่งจากข้อจำกัดทางด้านกฎหมายและด้านการเปิดเผยข้อมูลทางการเงินของธนาคาร ผู้วิจัย ได้ใช้ชุดข้อมูลที่มาจากการสังเคราะห์จากธุรกรรมจริงของธนาคารในประเทศไทย โดยมี คุณลักษณะทั้งหมด 12 รายการ ดังนี้ เวลาที่ทำธุรกรรม วันที่ที่ทำธุรกรรม บัญชีต้นทาง บัญชี ปลายทาง จำนวนเงิน สกุลเงินในการชำระเงิน สกุลเงินของฝั่งปลายทาง ที่ตั้งธนาคารบัญชีต้นทาง ที่ตั้งธนาคารบัญชีปลายทาง ประเภทของการชำระเงิน ป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงิน และ ประเภทของธุรกรรม โดยมีป้ายกำกับที่บ่งชี้ธุรกรรมการฟอกเงินและประเภทของธุรกรรมเป็นผลเฉลย ในการตรวจจับการฟอกเงิน จากนั้นผู้วิจัยได้ทำขั้นตอนการเตรียมข้อมูลดังนี้ การแยกเดือน วัน ชั่วโมง ออกมาจากคุณลักษณะวันที่และเวลา เพื่อทำการหารูปแบบเฉพาะสำหรับการฟอกเงิน โดยพบ รูปแบบการทำธุรกรรมการฟอกเงินส่วนใหญ่มักจะเป็นช่วงวันที่กลางเดือน และในชั่วโมงปกติช่วงเวลา 8 โมงเช้า จนถึง 5 โมง จากนั้นทำการวิเคราะห์ที่ตั้งธนาคาร ก็พบว่าที่ตั้งธนาคารทั้งของบัญชีต้นทาง และปลายทาง ส่วนใหญ่อยู่ที่ประเทศไทย โดยมีสกุลเงินที่ชำระและสกุลเงินฝั่งปลายเป็นเงิน ปอนด์ ซึ่งแน่นอนว่าเป็นเพราะข้อมูลชุดนี้สังเคราะห์มาจากธนาคารในประเทศไทย และหาก สสำรวจประเภทในการชำระเงินของธุรกรรมฟอกเงิน ก็พบว่าผู้ที่ฟอกเงินมักจะใช้การทำธุรกรรมข้าม ประเทศในการฟอกเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้ผู้วิจัยได้ทำการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคุณลักษณะด้วย Correlation Matrix เพื่อใช้ในการตัดคุณลักษณะที่มีความสัมพันธ์กันมากเกินไปทิ้ง พบว่าคุณลักษณะทางด้านที่ตั้งของธนาคารกับสกุลเงินนั้น มีความสัมพันธ์กันสูง จึงเลือกตัดคุณลักษณะด้านสกุลเงินทิ้ง และจากการแยกคุณลักษณะเดือน วัน ชั่วโมง จึงสามารถตัดคุณลักษณะทางด้านวันที่และเวลาที่ทิ้งด้วย รวมถึงการสำรวจประเภทของธุรกรรม ก็พบว่า เป็นคุณลักษณะที่เป็นผลเฉลยของธุรกรรมการฝากเงิน เช่นเดียวกันกับป้ายกำกับที่บ่งชี้ธุรกรรมการฝากเงิน จึงทำการตัดประเภทของธุรกรรมออกไปด้วย

สุดท้ายนี้ผู้วิจัยได้ตรวจสอบป้ายกำกับที่บ่งชี้ธุรกรรมการฝากเงินเพิ่มเติม พบว่าผลเฉลยนี้มีความไม่สมดุลกันสูง เนื่องจากชุดข้อมูลนี้มีธุรกรรมที่ไม่ใช่การฝากเงินอยู่ 9,494,979 รายการ แต่มีธุรกรรมที่เป็นการฝากเงินอยู่เพียง 9,873 รายการ จึงดำเนินการจัดการข้อมูลด้วยการทำ Undersampling และ Oversampling ให้ธุรกรรมการฝากเงินมีขนาดเท่ากับธุรกรรมปกติ เพื่อให้แบบจำลองสามารถเรียนข้อมูลทั้ง 2 คลาสนี้ได้อย่างมีประสิทธิภาพ โดยจะนำข้อมูลจากทั้ง 2 วิธี ไปสร้างแบบจำลอง เพื่อเปรียบเทียบหาวิธีที่เหมาะสมในการจัดการปัญหาข้อมูลไม่สมดุลของชุดข้อมูลสังเคราะห์นี้

หลังจากเตรียมข้อมูลเรียบร้อยแล้ว จึงเริ่มสร้างแบบจำลองสำหรับตรวจจับธุรกรรมการฝากเงิน โดยมีขั้นตอนดังนี้ เริ่มจากการทดลองใช้ Grid Search เพื่อทำ Hyperparameter Tuning ในการวนลูบพารามิเตอร์ จากนั้นทำการฝึกแบบจำลองแล้วทดลองซ้ำด้วยชุดข้อมูลคนละส่วน 3 รอบ เพื่อให้ได้ชุดของพารามิเตอร์ที่เหมาะสมที่สุดในแต่ละแบบจำลอง

โดยผลลัพธ์ที่ได้ของทั้ง 2 แบบจำลอง ที่ใช้วิธีการจัดการข้อมูลไม่สมดุล 2 แบบ จึงสรุปผลการเปรียบเทียบประสิทธิภาพดังนี้ แบบจำลอง XGBoost และ LightGBM ที่มีการจัดการกับข้อมูลไม่สมดุลด้วยการทำ Undersampling มีประสิทธิภาพต่ำที่สุด โดยแบบจำลอง XGBoost ให้ผล Accuracy 75.12% ส่วนแบบจำลอง LightGBM ให้ผล Accuracy 75.35% ซึ่งถือว่าแบบจำลองตรวจจับธุรกรรมการฝากเงินได้มีประสิทธิภาพ แต่ยังคงมีความแม่นยำไม่มากเท่าที่ควร

ในขณะเดียวกันแบบจำลองที่จัดการกับข้อมูลไม่สมดุลด้วยการทำ Oversampling นั้นให้ประสิทธิภาพที่ดีกว่า โดยที่แบบจำลอง LightGBM ให้ผล Accuracy ที่ 89.40% ซึ่งถือว่ามีประสิทธิภาพมากกว่าการใช้ข้อมูลที่ผ่านการทำ Undersampling และแบบจำลองที่โดดเด่นที่สุด คือ XGBoost เนื่องจากให้ค่า Accuracy สูงที่สุด ซึ่งสูงถึง 99.93% โดยไม่เกิดปัญหา Overfitting จึงสรุปได้ว่าการทำ Oversampling นั้น ช่วยให้ผลลัพธ์ของแบบจำลอง ดีกว่าวิธี Undersampling โดยมีแบบจำลอง XGBoost เป็นแบบจำลองที่ตรวจจับธุรกรรมการฝากเงินได้อย่างมีประสิทธิภาพมากที่สุด

อย่างไรก็ตาม การใช้วิธี Oversampling ข้อมูลเพื่อสร้างความสมดุลนี้ มีข้อเสียคือ ข้อมูลธุรกรรมการฝากเงินที่ถูกสังเคราะห์ด้วยวิธี SMOTE อาจทำให้ข้อมูลที่ได้เพิ่มมาเป็นข้อมูลตัวอย่างที่ไม่สมจริง ซึ่งอาจเป็นผลให้แบบจำลอง XGBoost สามารถทำนายได้แม่นยำมากถึง 99.93% ในขณะที่

ที่ การทำ Undersampling นั้น แบบจำลอง XGBoost สามารถทำนายได้แม่นยำเพียงแค่ 75.12% ซึ่งให้ผลที่ต่างกันถึง 24.81%

และจากการนำข้อมูลมาจากเว็บไซต์ Kaggle โดยมีนักพัฒนาท่านอื่นได้ทดลองสร้างแบบจำลองด้วยข้อมูลชุดเดียวกัน และใช้ข้อมูลที่ไม่สมดุลชุดนี้ในการฝึกแบบจำลอง XGBoost ผลพบว่า แบบจำลอง XGBoost สามารถทำนายได้แม่นยำถึง 81.205% แต่ผลการตรวจจับธุรกรรมปกติว่าเป็นธุรกรรมการฟอกเงิน หรือผลบวกวงนั้นีมีมากถึง 58% ซึ่งหากเปรียบเทียบกับการพัฒนาแบบจำลอง XGBoost ของงานวิจัยในครั้งนี้ พบว่าการจัดการข้อมูลไม่สมดุลด้วย Oversampling นั้นทำให้แบบจำลองมีประสิทธิภาพมากขึ้น ด้วยผลการทำนายแม่นยำกว่า 99.93% และให้ผลบวกวงที่น้อยมากเพียงแค่ 0.122% (Berkan Oztas. 2024)

สุดท้ายนี้ จากการหาคุณลักษณะของชุดข้อมูลที่มีผลต่อแบบจำลอง พบว่าคุณลักษณะที่สำคัญต่อแบบจำลอง XGBoost มากที่สุด คือคุณลักษณะด้าน ประเภทของการชำระเงิน ที่ตั้งของธนาคารบัญชี ปลายทาง และที่ตั้งของธนาคารบัญชีต้นทาง ซึ่งการใช้ข้อมูลจากการทำ Undersampling และ Oversampling ได้ให้ผลลัพธ์ของการหา Feature Importance ที่ตรงกัน แต่ในขณะเดียวกัน แบบจำลอง LightGBM พบว่าคุณลักษณะด้าน บัญชีปลายทาง จำนวนเงิน และบัญชีต้นทาง เป็นคุณลักษณะที่สำคัญต่อแบบจำลองมากที่สุด โดยการใช้ข้อมูลจากการทำ Undersampling และ Oversampling ได้ให้ผลลัพธ์ที่ใกล้เคียงกัน จะเห็นว่าอัลกอริทึมของทั้ง 2 แบบจำลอง มีการเลือกใช้และให้ความสำคัญกับคุณลักษณะเพื่อฝึกฝนและทำนายไม่เหมือนกัน

## 5.2 ข้อเสนอแนะ

จากการวิจัยนี้ ผู้วิจัยมีข้อเสนอแนะที่สามารถนำไปปรับปรุงหรือพัฒนาต่อในอนาคตได้ ดังนี้

### 1. การนำแบบจำลองไปใช้งานกับข้อมูลจริงของธนาคาร

เนื่องด้วยปัญหาด้านความปลอดภัยของข้อมูลธนาคาร ซึ่งทำให้ผู้วิจัยเลือกใช้ข้อมูลสังเคราะห์จากเว็บไซต์ Kaggle มาพัฒนาสร้างแบบจำลองเพื่อตรวจจับธุรกรรมการฟอกเงิน ทำให้แบบจำลองนี้ยังมีข้อจำกัดที่เกิดจากการไม่ได้ใช้ข้อมูลจริงของธนาคารในประเทศไทย ซึ่งอาจทำให้แบบจำลองยังไม่สามารถนำไปใช้ได้จริงในธนาคาร ดังนั้นผู้วิจัยมีข้อเสนอแนะว่า หากมีการนำแบบจำลองไปพัฒนาต่อ ต้องคำนึงถึงชุดข้อมูลที่จะนำมาใช้ฝึกแบบจำลอง ซึ่งควรมีรูปแบบและโครงสร้างของข้อมูล ที่ใกล้เคียงกับข้อมูลจริงของทางธนาคารมากที่สุด เพื่อให้แบบจำลองมีประสิทธิภาพมากพอสำหรับการนำไปใช้งานจริง

### 2. เพิ่มคุณลักษณะของชุดข้อมูล

จากการศึกษาชุดข้อมูลที่ใช้ในการสร้างแบบจำลองการตรวจจับธุรกรรมฟอกเงินนี้ ที่มีเพียง 12 คุณลักษณะ บวกกับการศึกษาโครงสร้างของข้อมูลที่ใช้ในระบบการตรวจจับธุรกรรมฟอกเงินของธนาคาร ผู้วิจัยมีข้อเสนอแนะว่าควรเพิ่มคุณลักษณะในแบบจำลองให้มีความครบถ้วนมากขึ้น เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่ในทางวิชาการไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่น วันที่ในการเปิดบัญชีเพื่อเปรียบเทียบกับ วันที่เริ่มทำธุรกรรมและจำนวนเงินที่ทำธุรกรรมครั้งแรก เพิ่มข้อมูลจุดประสงค์ในการเปิดบัญชี ทั้งรูปแบบบุคคลหรือนิติบุคคล รูปแบบของธุรกิจ เพื่อสามารถนำไปตรวจสอบเพิ่มเติมได้ และเพิ่มข้อมูลส่วนบุคคลของเจ้าของบัญชี เพื่อตรวจสอบไปถึงบัญชีอื่นๆ ภายใต้เจ้าของบัญชีบุคคลเดียวกันหรือบัญชีร่วม รวมถึงบุคคลที่อาจเกี่ยวข้องกันในครอบครัว เป็นต้น

### 3. พัฒนาการสร้างแบบจำลองที่ตรวจจับเครือข่าย

จากการศึกษาระบบตรวจจับการฟอกเงินของธนาคาร พบว่าทางธนาคารจะมีการเตรียมข้อมูลที่เป็นขั้นตอนสำคัญอย่างมาก เพื่อให้เจ้าหน้าที่ปฏิบัติตามข้อกำหนดได้ตรวจสอบธุรกรรมอย่างครอบคลุม โดยหนึ่งในการเตรียมข้อมูลนั้นคือ การหารายชื่อบุคคล รายชื่อนิติบุคคล รวมถึงบัญชีที่เกี่ยวข้องกันทั้งหมด เพื่อตรวจหาเครือข่ายที่สำคัญในการฟอกเงิน เนื่องจากการทำธุรกรรม การฟอกเงินส่วนใหญ่จะทำงานเป็นขบวนการ และส่วนใหญ่สามารถตรวจจับได้จริงจากการเชื่อมโยงไปถึงบัญชีที่มีการโอนจ่ายเป็นทอดๆ ต่อกัน การมีชื่ออยู่ภายใต้ธุรกิจเดียวกัน หรือการแลกเปลี่ยนทรัพย์สินต่อกัน ซึ่งหากสามารถพัฒนาแบบจำลองที่เรียนรู้พฤติกรรมแบบเครือข่ายได้ จะสามารถเชื่อมโยงความเกี่ยวข้องและตรวจจับขบวนการฟอกเงินได้ครอบคลุมมากขึ้น

### 4. เครื่องมือที่ใช้สร้างแบบจำลองที่มีข้อมูลขนาดใหญ่

เนื่องด้วยข้อมูลที่นำมาใช้สร้างแบบจำลองในงานวิจัยนี้ มีจำนวนกว่า 10 ล้านธุรกรรม ผู้วิจัยขอเสนอแนะเครื่องมือ Colab Pro+ ในการพัฒนาแบบจำลองด้วยภาษา Python ซึ่งเป็นบริการจาก Google ที่ให้ทรัพยากรคอมพิวเตอร์ที่มีประสิทธิภาพสูงสำหรับการประมวลผลข้อมูลและการสร้างแบบจำลองการเรียนรู้ของเครื่อง โดยข้อดีของ Colab Pro+ คือสามารถเข้าถึง GPU และ TPU ระดับสูง ซึ่งมีประสิทธิภาพสูงกว่า GPU มาตรฐาน ทำให้การฝึกและทำนายแบบจำลองเร็วขึ้นมาก และเข้าถึงการใช้งาน CPU และ RAM ที่สูงขึ้น ทำให้ความสามารถในการประมวลผลสูงขึ้นอีกด้วย ขยายเวลาการประมวลผล โดยสามารถประมวลผลงานที่ต้องใช้เวลานานได้โดยไม่ถูกตัดการเชื่อมต่อบ่อยๆ เหมือน Colab แบบธรรมดา เหมาะสำหรับคนที่ไม่มีข้อจำกัดในการใช้ทรัพยากรส่วนตัวเพื่อสร้างแบบจำลองด้วยข้อมูลขนาดใหญ่ ซึ่งการสมัครใช้งาน Colab Pro+ จะมีค่าใช้จ่ายอยู่ที่ 1,677.76 บาทต่อเดือน

## เอกสารอ้างอิง

- Ahmed, N. A. 2023. **What is A Confusion Matrix in Machine Learning? The Model Evaluation Tool Explained.** [Online]. Available : <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>
- Algoaddict. 2021. **8 เทคนิคง่ายๆ ป้องกัน Overfitting เพื่อโมเดล Machine Learning ที่มีประสิทธิภาพ.** [Online]. Available : <https://algoaddict.wordpress.com/2021/10/01/>
- Bakry, A. N. Alsharkawy, A. S. Farag, M. S. and Raslan, K. R. 2023. “Automatic suppression of false positive alerts in anti-money laundering systems using machine learning.” Volume 80, pages 6264–6284, **The Journal of Supercomputing.** <https://doi.org/10.1007/s11227-023-05708-z>
- Daroontham, W. 2018. **รู้จัก Decision Tree, Random Forest, และ XGBoost!!! — PART1.** [Online]. เข้าถึงได้จาก : <https://medium.com/@witchapongdaroontham/>
- Eddin, A. N. Bono, J. Aparicio, D. Polido, D. Ascensão, J. T. Bizarro, P. and Ribeiro, P. 2022. “Anti-Money Laundering Alert Optimization Using Machine Learning with Graphs.” **DCC-FCUP, University of Porto, Portugal.** <https://doi.org/10.48550/arXiv.2112.07508>
- Huong, H. Nguyen, X. Dang, T. K. and Tran-Truong, P. T. 2024. “Money Laundering Detection Using A Transaction-Based Graph Learning Approach.” **2024 18th International Conference** <https://doi.org/10.1109/IMCOM60618.2024.10418307>
- Jain, A. 2024. **Mastering XGBoost Parameters Tuning: A Complete Guide with Python Codes.** [Online]. Available : <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>
- Oztas, B. 2024. **SAML-D Exploratory Data Analysis & Classification.** [Online]. Available : <https://www.kaggle.com/code/berkanoztas/saml-d-exploratory-data-analysis-classification>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Pathiyaparambath, T. S. and Gopika P., Professor. 2024. “Machine Learning for Anti-Money Laundering and Fraud Detection.” **International Research Journal of Innovations in Engineering and Technology (IRJIET)**

DOI : 10.47001/IRJIET/2024.802021

Phyu, T. H. and Uttama, S. 2023. “Improving Classification Performance of Money Laundering Transactions Using Typological Features.” **2023 7th International Conference on Information Technology (InCIT)**

<https://doi.org/10.1109/InCIT60207.2023.10413155>

Singh, S. 2024. **LightGBM (Light Gradient Boosting Machine)** [Online]. Available : <https://www.geeksforgeeks.org/lightgbm-light-gradient-boosting-machine/>

กรมสอบสวนคดีพิเศษ. 2567. **แนวทางการดำเนินความผิดฐานฟอกเงิน.** [Online].

เข้าถึงได้จาก : [https://www.dsi.go.th/Files/20140813/F20140813163938-ptum1%20\(1\).pdf](https://www.dsi.go.th/Files/20140813/F20140813163938-ptum1%20(1).pdf)

ทินกร ม้าลายทอง และปพจน์ ธรรมเจริญพร. 2564. **เพิ่มประสิทธิภาพของ Machine Learning Model ด้วย Hyperparameter Optimization.** [Online]. เข้าถึงได้จาก : <https://bdi.or.th/big-data-101/machine-learning-model-hyperparameter-optimization/>

บริษัท SAS Software จำกัด. 2567. **การเรียนรู้ของเครื่อง (machine learning).** [Online]. เข้าถึงได้จาก : [https://www.sas.com/th\\_th/insights/analytics/machine-learning.html](https://www.sas.com/th_th/insights/analytics/machine-learning.html)

สถาบันนวัตกรรมและสรรมาภิบาลข้อมูล DIGI. 2567. **what-is-classification-model.**

[Online]. เข้าถึงได้จาก : <https://digi.data.go.th/blog/what-is-classification-model/>

สำนักงานป้องกันและปราบปรามการฟอกเงิน. 2559. **การฟอกเงิน (Money Laundering).**

[Online]. เข้าถึงได้จาก : [https://eportal.amlo.go.th/AmloPortal/knowledge-detail/Amlo-Article-\(8\).aspx](https://eportal.amlo.go.th/AmloPortal/knowledge-detail/Amlo-Article-(8).aspx)

## ประวัติผู้เขียน

ชื่อ นางสาวชนพัฒน์ เมืองเชียงหวาน  
วัน เดือน ปีเกิด วันที่ 11 สิงหาคม พ.ศ.2541  
ประวัติการศึกษา พ.ศ. 2563 เศรษฐศาสตรบัณฑิต สาขาเศรษฐศาสตร์  
มหาวิทยาลัยเกษตรศาสตร์  
พ.ศ. 2567 วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูลและการวิเคราะห์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้