การตรวจสอบการทุจริตในงบการเงินโดยใช้การเรียนรู้แบบมีผู้สอน

FRAUD DETECTION IN FINANCIAL STATEMENTS WITH SUPERVISED LEARNING

ภาณุพงศ์ เมี้ยนกลาง

PANUPONG MEANKLANG

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2567

KMITL-2024-SC-M-017-065

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FRAUD DETECTION IN FINANCIAL STATEMENTS WITH SUPERVISED LEARNING

PANUPONG MEANKLANG

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-017-065

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2024
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABAN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การตรวจสอบการทุจริตในงบการเงินโดยใช้การเรียนรู้แบบมีผู้สอน
ชื่อนักศึกษา	ภาณุพงศ์ เมี้ยนกลาง
รหัสประจำตัว	65056070
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ.	2567
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์.ดร.บุษยมาส พิมพ์พรรณชาติ

บทคัดย่อ

งานวิจัยนี้นำเสนอวิธีการตรวจสอบการทุจริตในงบการเงิน ปัจจัยในงบการเงินที่มีผลต่อการเกิดการทุจริตในงบการเงิน รวมไปถึงการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้แบบมีผู้สอน โดยใช้ข้อมูลงบการเงินจากตลาดหลักทรัพย์ NYSE และ NASDAQ ในการฝึกฝนแบบจำลอง โดยงบการเงินที่ใช้จะประกอบไปด้วยงบแสดงฐานะทางการเงิน งบกำไรขาดทุน งบกระแสเงินสด และอัตราส่วนทางการเงิน โดยเลือกใช้การเรียนรู้แบบมีผู้สอนทั้งหมด 3 วิธี ได้แก่ การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และเกรเดียนต์ บูสติ้ง โดยการวิจัยจะเริ่มตั้งแต่การทำความเข้าใจเกี่ยวกับข้อมูล การจัดเตรียมข้อมูล และจัดการกับข้อมูลที่มีคลาสไม่สมดุลกัน จากนั้นจะทำการสร้างแบบจำลองและนำแบบจำลองแต่ละแบบมาทำการเปรียบเทียบประสิทธิภาพกัน โดยการวัดประสิทธิภาพจะดูจากค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ 1 รวมถึงยังมีการดูค่าความสำคัญของแต่ละคุณลักษณะ เพื่อที่จะศึกษาว่ามีปัจจัยใดที่ส่งผลต่อการทุจริตในงบการเงิน ผลการศึกษาพบว่าแบบจำลองเกรเดียนต์ บูสติ้ง เป็นแบบจำลองที่มีประสิทธิภาพมากที่สุดในการทำนายการทุจริตในงบการเงิน โดยรายการทางบัญชีที่เป็นปัจจัยในการเกิดทุจริตในงบการเงิน 3 อันดับแรกคือ บัญชีหนี้สินหมุนเวียน บัญชีค่าใช้จ่ายในการดำเนินงาน และ บัญชีดอกเบี้ยรับ

คำสำคัญ: การตรวจสอบการทุจริตในงบการเงิน การเรียนรู้แบบมีผู้สอน งบการเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Fraud Detection in Financial Statements with Supervised Learning
Student Name	Panupong Meanklang
Student ID	65056070
Degree	Master of Science (Data Science and Analytics)
	KMITL Digital Analytics and Intelligence Center
Year	2567
Independent Study Advisor	Asst. Prof. Dr. Busayamas Pimpunchat

Abstract

This research presents a method for detecting fraud in financial statements, examining the factors within financial reports that contribute to fraud, and comparing the performance of supervised machine learning models. Financial data from the NYSE and NASDAQ stock exchanges are used to train the models, which include balance sheets, income statements, cash flow statements, and financial ratios. There are 3 supervised machine learning techniques which are used in this research that are Multiple Logistic Regression, Random Forest, and Gradient Boosting. The study begins with data understanding, preparation, and handling of class imbalance. Following this, the models are built and their performance is compared using accuracy, precision, and F-1 Score. Additionally, the importance of each feature is analyzed to determine which factors influence financial fraud. The results indicate that Gradient Boosting is the most effective model for predicting financial fraud. The top three factors contributing to fraud are current liabilities, operating expenses, and interest income.

Keywords: fraud detection in financial statement, supervised learning, financial statements

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยการตรวจสอบการทุจริตในงบการเงินโดยใช้การเรียนรู้แบบมีผู้สอน สามารถสำเร็จลุล่วงไปด้วยดีด้วยความช่วยเหลือของ ผศ.ดร.บุษยมาส พิมพ์พรรณชาติ อาจารย์ที่ปรึกษางานการค้นคว้าอิสระที่ได้ให้คำปรึกษาการแนะนำ และเอื้อเฟื้อแนวทางในการจัดทำงานวิจัยครั้งนี้ ตลอดจนช่วยชี้แนะข้อบกพร่องต่าง ๆ และตรวจทานแก้ไขให้งานวิจัยครั้งนี้สำเร็จลุล่วงไปด้วยดี ผู้วิจัยขอขอบคุณเป็นอย่างยิ่ง ณ โอกาสนี้

ขอขอบคุณคณาจารย์สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้และช่วยเหลือ รวมถึงให้คำแนะนำในเรื่องต่าง ๆ มาโดยตลอด

ขอขอบคุณครอบครัวที่ให้ความช่วยเหลือ และการสนับสนุนในการศึกษาและการจัดทำการค้นคว้าอิสระครั้งนี้ รวมถึงขอบคุณเพื่อน ๆ คณะวิทยาศาสตร์ สาขาวิทยาการข้อมูลและการวิเคราะห์ที่คอยช่วยเหลือ และให้กำลังใจในการทำให้การค้นคว้าอิสระนี้ออกมาลุล่วงด้วยดี

ขอบคุณข้อมูลจากตลาดหลักทรัพย์อเมริกา และ ตลาดหลักทรัพย์แห่งประเทศไทย สำหรับแหล่งข้อมูลที่นำมาใช้ในงานวิจัยนี้

สุดท้ายนี้ ผู้วิจัยหวังว่างานวิจัยฉบับนี้จะเป็นประโยชน์สำหรับผู้ที่ต้องการจะศึกษา หากมีข้อผิดพลาดประการใดผู้วิจัยขออภัยมา ณ ที่นี้

นาย ภาณุพงศ์ เมี้ยนกลาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

บทคัดย่อ ก

Abstract ข

กิตติกรรมประกาศ ค

สารบัญ ง

สารบัญตาราง ฉ

สารบัญรูป ช

คำย่อ/สัญลักษณ์ ซ

1. บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 2

1.3 ขอบเขตของงานวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 2

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 3

2.1 งบการเงิน 3

2.2 การจัดเตรียมข้อมูล 7

2.2.1 การจัดการกับชุดข้อมูลที่ไม่สมดุล 7

2.2.2 การแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปของตัวเลข 7

2.2.3 การปรับช่วงข้อมูล 8

2.3 การเรียนรู้ของเครื่อง 8

2.3.1 การถดถอยลอจิสติกพหุคูณ 8

2.3.2 ป่าสุ่ม 9

2.3.3 เกรเดียนต์ บูสติ้ง 10

2.4 การวัดประสิทธิภาพของแบบจำลอง 12

2.4.1 ค่าความแม่น 13

2.4.2 ค่าความเที่ยง 13

2.4.3 ค่าการเรียกคืน 14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

2.4.4 ค่าคะแนนเอฟ1 14

2.5 การหาค่าตัวแปรที่สำคัญที่มีผลต่อแบบจำลอง 14

2.6 งานวิจัยที่เกี่ยวข้อง 15

3. วิธีดำเนินงานวิจัย 16

3.1 การเก็บรวบรวมข้อมูล 17

3.2 การทำความเข้าใจเกี่ยวกับข้อมูล 20

3.3 การจัดเตรียมข้อมูล 20

3.3.1 การตรวจสอบค่าสูญหายและแทนที่ค่าสูญหาย 20

3.3.2 การแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปของตัวเลข 21

3.3.3 การปรับช่วงข้อมูล 22

3.3.4 การแบ่งข้อมูล 23

3.3.3 การจัดการกับข้อมูลที่ไม่สมดุล 23

3.4 การสร้างแบบจำลอง 24

3.5 การวัดประสิทธิภาพของแบบจำลอง 24

3.6 การเลือกตัวแปรที่มีผลต่อแบบจำลอง 24

4. ผลการวิจัยและอภิปรายผล 25

4.1 ผลการวิจัย 25

4.2 การอภิปรายผล 35

5. สรุปผลการวิจัยและข้อเสนอแนะ 37

5.1 สรุปผลการวิจัย 37

5.2 ข้อเสนอแนะ 38

เอกสารอ้างอิง 39

ภาคผนวก 41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
ตารางที่ 2.1 เมทริกซ์ความสับสน	13
ตารางที่ 3.1 ตัวอย่างชุดข้อมูลของงบการเงิน	18
ตารางที่ 3.2 ประเภทของอุตสาหกรรม	19
ตารางที่ 3.3 ข้อมูลคุณลักษณะของงบการเงิน	20
ตารางที่ 3.4 การแปลงคุณลักษณะเชิงหมวดหมู่การให้อยู่ในรูปของตัวเลข	21
ตารางที่ 3.5 การปรับช่วงของข้อมูล ให้อยู่ในช่วงเดียวกัน	22
ตารางที่ 4.1 เมทริกซ์ความสับสนของการถดถอยลอจิสติกพหุคูณกับข้อมูลชุดฝึกฝน	25
ตารางที่ 4.2 เมทริกซ์ความสับสนของป่าสุ่มกับข้อมูลชุดฝึกฝน	26
ตารางที่ 4.3 เมทริกซ์ความสับสนของ XGBoost กับข้อมูลชุดฝึกฝน	26
ตารางที่ 4.4 การวัดประสิทธิภาพของแบบจำลองกับข้อมูลชุดฝึกฝน	26
ตารางที่ 4.5 เมทริกซ์ความสับสนของการถดถอยลอจิสติกพหุคูณกับข้อมูลชุดทดสอบ	27
ตารางที่ 4.6 เมทริกซ์ความสับสนของป่าสุ่มกับข้อมูลชุดทดสอบ	27
ตารางที่ 4.7 เมทริกซ์ความสับสนของ XGBoost กับข้อมูลชุดทดสอบ	28
ตารางที่ 4.8 การวัดประสิทธิภาพของแบบจำลองกับข้อมูลชุดทดสอบ	28
ตารางที่ 4.9 คุณลักษณะที่มีความสำคัญในแต่ละแบบจำลอง	29
ตารางที่ 4.11 ค่าความสำคัญของตัวแปรที่มีผลต่อการทุจริตของ XGBoost	29
ตารางที่ 4.10 การนำมาแบบจำลองมาใช้ตรวจสอบทุจริตในงบการเงินไทย	35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
รูปที่	2.1 ตัวอย่างของงบแสดงฐานะการเงิน	4
รูปที่	2.2 ตัวอย่างของงบกำไรขาดทุนเบ็ดเสร็จ	5
รูปที่	2.3 ตัวอย่างของงบกระแสเงินสด	6
รูปที่	2.4 ฟังก์ชันของการถดถอยลอจิสติก	8
รูปที่	2.5 หลักการทำงานของแบบจำลองป่าสุ่ม	9
รูปที่	2.6 ลักษณะการทำงานของ Bagging	10
รูปที่	2.7 ลักษณะการทำงานของ Boosting	11
รูปที่	2.8 ลักษณะการทำงานของเกรเดียนต์ บูสติ้ง	12
รูปที่	3.1 ผังขั้นตอนในการดำเนินงาน	16
รูปที่	3.2 เว็บไซต์ข้อมูลงบการเงินจากคณะกรรมการกำกับหลักทรัพย์ของสหรัฐอเมริกา	17
รูปที่	3.3 จำนวนบริษัทที่ทุจริต และไม่ทุจริต	19
รูปที่	3.4 ชุดข้อมูลหลังทำการการ Oversampling ด้วยวิธีการ SMOTE	23
รูปที่	4.1 อัตราส่วนหนี้สินหมุนเวียนอื่น ๆ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต	30
รูปที่	4.2 อัตราส่วนค่าใช้จ่ายในการดำเนินงาน ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต	31
รูปที่	4.3 อัตราส่วนดอกเบี้ยรับ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต	32
รูปที่	4.4 อัตราส่วนหุ้นสามัญ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต	33
รูปที่	4.5 อัตราส่วนหนี้สินหมุนเวียนอื่น ๆ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต	34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำย่อ/สัญลักษณ์

คำย่อ	ความหมาย
Interest Income	บัญชีดอกเบี้ยรับ
otherCurrentLiabilities	บัญชีหนี้สินหมุนเวียนอื่น ๆ
Total Current Liabilities	บัญชีหนี้สินหมุนเวียนสุทธิ
operatingExpenses	บัญชีค่าใช้จ่ายในการดำเนินงาน
Tax Payables	บัญชีภาษีค้างชำระ
otherExpenses	บัญชีค่าใช้จ่ายอื่น ๆ
Total Equity	บัญชีส่วนของเจ้าของ
propertyPlantEquipmentNet	บัญชีที่ดิน อาคาร และอุปกรณ์
commonStock	บัญชีทุนหุ้นสามัญ
Account Payables	บัญชีเจ้าหนี้การค้า
otherCurrentAssets	บัญชีสินทรัพย์หมุนเวียนอื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การดำเนินธุรกิจอย่างไม่โปร่งใสของบริษัทไม่ใช่เรื่องที่แปลกใหม่ ถึงแม้ว่าข้อมูลในอดีตจะบ่งบอกว่าการทุจริตไม่ส่งผลดีต่อบริษัทในระยะยาว แต่ก็ยังมีข่าวอื้อฉาวทางการเงินที่เกิดขึ้นมากมาย ไม่ว่าจะในประเทศ หรือต่างประเทศ อีกทั้งในปัจจุบันเทคโนโลยีเข้ามามีบทบาทในการดำเนินธุรกิจเพิ่มมากขึ้น ส่งผลให้การทุจริตในบริษัทสามารถทำได้ง่ายขึ้น ซึ่งการทุจริตทางการเงินสามารถเกิดขึ้นได้ในหลายรูปแบบ เช่น การบิดเบือนรายงานทางการเงิน การสร้างรายงานทางบัญชีที่ไม่ถูกต้อง หรือการปกปิดข้อมูลที่สำคัญ เพื่อทำให้ภาพลักษณ์ขององค์กรดูดีกว่าความเป็นจริง ปัญหาเหล่านี้ไม่เพียงแต่สร้างความเสียหายทางการเงินให้กับองค์กรเท่านั้น แต่ยังส่งผลกระทบต่อความเชื่อมั่นของนักลงทุนและผู้มีส่วนได้ส่วนเสียต่าง ๆ อีกด้วย การสูญเสียความเชื่อมั่นนี้อาจส่งผลกระทบระยะยาวต่อองค์กร ทั้งในแง่การดำเนินธุรกิจ การระดมทุน หรือแม้กระทั่งการรักษาความสัมพันธ์กับลูกค้าและพันธมิตรทางธุรกิจ (Wyrobek, 2020)

ในปัจจุบัน การตรวจสอบการทุจริตทางการเงินต้องใช้กระบวนการตรวจสอบแบบดั้งเดิม ซึ่งต้องใช้แรงงานคน โดยการตรวจสอบบัญชีอย่างละเอียดโดยผู้ตรวจสอบบัญชี ซึ่งมักจะใช้เวลานานและมีความเสี่ยงที่จะตรวจไม่พบความผิดปกติหรือการทุจริตที่ซ่อนอยู่ โดยเฉพาะในกรณีที่ข้อมูลมีปริมาณมากและมีความซับซ้อน ทำให้กระบวนการตรวจสอบแบบเดิมอาจไม่สามารถตอบสนองต่อความต้องการในสภาพแวดล้อมทางธุรกิจที่เปลี่ยนแปลงไปอย่างรวดเร็วได้ ดังนั้นการนำเทคโนโลยีเข้ามาประยุกต์ใช้ จึงสามารถที่จะช่วยลดระยะเวลาของการตรวจสอบบัญชีของผู้ตรวจสอบลงได้ อีกทั้งยังสามารถช่วยการลดการทำงานที่ซ้ำซ้อน และยังทำให้ผู้ตรวจสอบสามารถมุ่งเน้นในส่วนงานที่สำคัญได้มากยิ่งขึ้นอีกด้วย (Sattar, 2023)

ดังนั้น ผู้จัดทำจึงมีความสนใจที่จะศึกษาและพัฒนาแบบจำลองโดยใช้เทคโนโลยีการเรียนรู้แบบมีผู้สอน เพื่อนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลทางการเงินของบริษัท โดยมีเป้าหมายหลักเพื่อการตรวจจับและวิเคราะห์ความผิดปกติที่อาจเกิดขึ้นในข้อมูลทางการเงินได้อย่างรวดเร็วและมีประสิทธิภาพ ซึ่งเป็นการช่วยเพิ่มประสิทธิภาพในการตรวจสอบและลดความเสี่ยงในการทุจริตทางการเงิน ซึ่งช่วยให้ผู้มีส่วนได้ส่วนเสีย หรือผู้ลงทุนสามารถรับมือกับความเสี่ยงทางการเงินที่อาจเกิดขึ้นได้อย่างมีประสิทธิภาพและทันเวลา และสามารถสร้างความมั่นใจให้กับนักลงทุนให้มากขึ้นอีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อศึกษาแนวทางปฏิบัติในการนำการเรียนรู้แบบมีผู้สอน มาใช้ในการตรวจสอบทุจริตโดยใช้งบการเงิน
2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในการตรวจสอบการทุจริตในงบการเงิน
3) เพื่อศึกษาว่ามีปัจจัยทางบัญชีใดบ้าง ในงบการเงินที่มีส่วนสำคัญและส่งผลต่อการเกิดทุจริตในการเงิน

1.3 ขอบเขตของงานวิจัย

1) ขอบเขตในด้านข้อมูลที่ศึกษา เป็นข้อมูลของงบการเงิน ของบริษัทในตลาดหลักทรัพย์ NASDAQ และ NYSE ของประเทศสหรัฐอเมริกา โดยใช้สำหรับการฝึกฝนและการทดสอบแบบจำลอง และหุ้นไทยในตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้สำหรับการทดสอบแบบจำลอง ซึ่งงบการเงินทั้งหมดจะถูกเก็บรวบรวมในปี ค.ศ. 2007 - ค.ศ. 2022

2) งบเขตในด้านจำนวนข้อมูลที่ศึกษา เป็นข้อมูลของงบการเงินที่ถูกระบุว่าทุจริต จำนวน 349 งบ จากทั้งหมด 70 บริษัท และงบการเงินของบริษัทที่ถูกระบุว่าไม่ทุจริตทั้งหมด 1,013 งบการเงิน จากทั้งหมด 213 บริษัท

3) ขอบเขตของวิธีการสร้างแบบจำลอง โดยจะใช้แบบจำลองที่มีการเรียนรู้แบบมีผู้สอน โดยใช้แบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ช่วยให้สามารถนำการเรียนรู้แบบมีผู้สอนมาใช้ในการตรวจสอบความผิดปกติของงบการเงิน ก่อนที่จะออกงบการเงินได้

2) ช่วยเป็นแนวทางประกอบการตัดสินใจของผู้มีส่วนได้ส่วนเสีย ผู้ลงทุน และผู้ที่ศึกษางบการเงิน เพื่อช่วยลดความเสี่ยงในการใช้งบการเงินในการลงทุน

3) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองของการเรียนรู้แบบมีผู้สอน ซึ่งช่วยให้สามารถเลือกแบบจำลองที่เหมาะสมกับการตรวจสอบการทุจริตในงบการเงิน สามารถนำไปใช้งานได้

4) ช่วยให้ทราบถึงปัจจัยทางบัญชี ที่มีส่วนสำคัญต่อการเกิดทุจริตในงบการเงิน เพื่อให้ผู้ใช้งบการเงินทราบและตรวจสอบปัจจัยทางบัญชีที่ผิดปกติเหล่านี้อย่างถี่ถ้วนมากขึ้น ก่อนที่จะตัดสินใจลงทุนได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาการค้นคว้าอิสระนี้ มีทฤษฎีและงานวิจัยที่เกี่ยวข้อง หลักเกณฑ์ และงานวิจัยที่เกี่ยวข้องต่าง ๆ โดยแบ่งออกเป็นหัวข้อต่าง ๆ ได้แก่ ความหมายของงบการเงิน การจัดเตรียมข้อมูล การเรียนรู้ของเครื่อง การวัดประสิทธิภาพของของแบบจำลอง และงานวิจัยที่เกี่ยวข้อง

2.1 งบการเงิน

2.1.1 ความหมายของงบการเงิน

งบการเงิน เป็นสิ่งที่บ่งบอกถึงความสามารถในการดำเนินงาน และการเปลี่ยนแปลงของฐานะการเงินของกิจการ ซึ่งข้อมูลนี้จะมีประโยชน์ต่อผู้ใช้งบการเงินในการตัดสินใจทางธุรกิจ นอกจากนี้ งบการเงินยังแสดงข้อมูลของช่วงเวลาที่ผ่านมา เพื่อให้ผู้ใช้งบการเงินสามารถเปรียบเทียบฐานะการเงิน ผลการดำเนินงาน และการเปลี่ยนแปลงฐานะการเงินของกิจการในช่วงเวลาที่แตกต่างกัน โดยงบการเงินจะมีทั้งหมด 3 ประเภท ได้แก่ (คำนึง สา ริ สระ และ ธรรม นิติ เพรส, 2020)

1. งบดุล หรือที่เรียกว่างบแสดงฐานะทางการเงิน แสดงภาพรวมของสถานะทางการเงินของบริษัทใน ช่วงเวลาที่กำหนด โดยจะแสดงสินทรัพย์ หนี้สิน และส่วนของผู้ถือหุ้น สินทรัพย์เป็นตัวแทนของสิ่งที่บริษัทเป็นเจ้าของ ในขณะที่หนี้สินบ่งบอกถึงภาระผูกพันของบริษัท ส่วนของผู้ถือหุ้นสะท้อนถึงส่วนได้เสียคงเหลือในสินทรัพย์หลังหักหนี้สินแล้ว การวิเคราะห์งบดุลสามารถประเมินสภาพคล่อง ความสามารถในการชำระหนี้ และความมั่นคงทางการเงินของบริษัทได้

2. งบกำไรขาดทุนเบ็ดเสร็จ งบกำไรขาดทุนเบ็ดเสร็จ หรือที่เรียกว่างบกำไรขาดทุน สรุปรายได้ ค่าใช้จ่าย กำไร และขาดทุนของบริษัท แสดงความสามารถในการสร้างรายได้และการจัดการค่าใช้จ่าย ผู้มีส่วนได้ส่วนเสียสามารถใช้ข้อมูลนี้เพื่อประเมินความสามารถในการทำกำไร และผลการดำเนินงานของบริษัท นอกจากนี้ ยังช่วยในการเปรียบเทียบประสิทธิภาพในการดำเนินงานในช่วงเวลาต่าง ๆ ด้วย

3. งบกระแสเงินสด งบกระแสเงินสดแสดงการไหลเข้าและไหลออกของเงินสดและรายการเทียบเท่าเงินสด ช่วยให้เข้าใจความสามารถในการสร้างเงินสดและการจัดการสภาพคล่องของบริษัท โดยแบ่งกระแสเงินสดออกเป็น 3 ประเภท ได้แก่ กิจกรรมดำเนินงาน กิจกรรมการลงทุน กิจกรรมจัดหาเงิน

2.1.2 การทุจริตในงบการเงิน

การกระทำที่ไม่ถูกต้อง และมีเจตนาในการบิดเบือนข้อมูลในงบการเงิน โดยมีวัตถุประสงค์เพื่อให้ข้อมูลที่ไม่ตรงกับความเป็นจริง โดยมีจุดประสงค์เพื่อแสวงหาผลประโยชน์ ไม่ว่าจะเป็น ประโยชน์ส่วนตน การเพิ่มมูลค่าหุ้น การหลีกเลี่ยงภาษี หรือการได้รับเงินกู้ยืมจากธนาคาร โดยการทุจริตของงบการเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถเกิดขึ้นได้หลายรูปแบบเช่น การปลอมรายได้ การหลีกเลี่ยงค่าใช้จ่าย การปกปิดหนี้สิน หรือการบันทึกสินทรัพย์สูงเกินจริง (Beneish, 1999)

ตัวอย่างของบการเงินได้แก่ งบแสดงฐานะการเงิน งบกำไรขาดทุนเบ็ดเสร็จ และงบกระแสเงินสด ถูกแสดงดังรูปที่ 2.1 – 2.3

บริษัท ไทยเซ็นทรัลเคมี จำกัด (มหาชน) และบริษัทย่อย

งบแสดงฐานะการเงิน

ณ วันที่ 31 ธันวาคม 2566

สินทรัพย์	**งบการเงินรวม**	
	2566	2565
สินทรัพย์หมุนเวียน		
เงินสดและรายการเทียบเท่าเงินสด	4,854,665,297	2,408,742,857
เงินลงทุนชั่วคราว	125,015,226	138,187,464
ลูกหนี้การค้า	2,114,701,000	1,823,148,377
ลูกหนี้อื่น	3,842,904	41,418,606
เงินตราต่างประเทศล่วงหน้า	21,750,666	-
ค่าใช้จ่ายจ่ายล่วงหน้า	42,009,161	38,266,102
รายได้ค้างรับ	11,568,565	3,013,304
เงินทดรองจ่าย	4,417,221	2,515,097
สินค้าคงเหลือ	2,107,982,787	3,363,872,206
สินทรัพย์หมุนเวียนอื่น	11,832,279	19,759,800
รวมสินทรัพย์หมุนเวียน	9,297,787,662	7,838,926,368

รูปที่ 2.1 ตัวอย่างของงบแสดงฐานะการเงิน

จากรูปที่ 2.1 แสดงตัวอย่างของงบแสดงฐานะการเงิน โดยประกอบด้วยสินทรัพย์ หนี้สิน และส่วนของผู้ถือหุ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บริษัท ไทยเซ็นทรัลเคมี จำกัด (มหาชน) และบริษัทย่อย

งบกำไรขาดทุนเบ็ดเสร็จ

สำหรับปีสิ้นสุดวันที่ 31 ธันวาคม 2566

	งบการเงินรวม	
	2566	2565
รายได้จากการขาย	22,235,681,387	22,128,054,300
รายได้จากการให้บริการ	129,935,915	125,476,986
ต้นทุนขาย	(18,468,150,323)	(19,600,324,604)
ต้นทุนการให้บริการ	(69,303,240)	(66,507,084)
กำไรขั้นต้น	3,828,163,739	2,586,699,598
รายได้อื่น		
ดอกเบี้ยรับ	64,496,663	49,173,191
ค่าเช่ารับ	8,500,885	8,908,251
เงินปันผลรับ	-	-
กำไรจากการจำหน่ายสินทรัพย์ถาวร	6,544,118	34,754,195
กำไรจากอัตราแลกเปลี่ยน - สุทธิ	52,591,917	70,904,459
อื่นๆ	27,046,866	33,522,477
กำไรก่อนค่าใช้จ่าย	3,987,344,188	2,783,962,171

รูปที่ 2.2 ตัวอย่างของงบกำไรขาดทุนเบ็ดเสร็จ

จากรูปที่ 2.2 แสดงตัวอย่างของงบงบกำไรขาดทุนเบ็ดเสร็จ โดยประกอบบัญชีหลักคือ รายได้จากการขาย ต้นทุนขาย กำไรขั้นต้น กำไรก่อนหักค่าใช้จ่าย กำไรจากกิจกรรมการดำเนินงาน กำไรก่อนค่าใช้จ่ายภาษีเงินได้ และกำไรสุทธิ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บริษัท ไทยเซ็นทรัลเคมี จำกัด (มหาชน) และบริษัทย่อย

งบกระแสเงินสด

สำหรับปีสิ้นสุดวันที่ 31 ธันวาคม 2556

	งบการเงินรวม	
	2556	2555
กระแสเงินสดจากกิจกรรมดำเนินงาน		
กำไรก่อนค่าใช้จ่ายภาษีเงินได้	2,845,079,635	0
ปรับปรุงด้วย		
หนี้สงสัยจะสูญ (โอนกลับ)	1,510,233	854,153
ขาดทุนจากการลดมูลค่าของสินค้าคงเหลือ (โอนกลับ)	179,851,716	178,200,346
ค่าเสื่อมราคา	241,940	234,541
ค่าตัดจำหน่าย - สินทรัพย์ไม่มีตัวตน	(6,544,118)	(34,754,195)
กำไรจากการจำหน่ายสินทรัพย์ถาวร	15,903	17,582
ขาดทุนจากการตัดจำหน่ายสินทรัพย์ถาวร	-	(645,919)
ขาดทุนจากการด้อยค่าของสินทรัพย์ถาวร (โอนกลับ)	16,378,239	18,765,781
ค่าใช้จ่ายภาระผูกพันผลประโยชน์พนักงาน	(545,576)	(1,011,137)
ส่วนแบ่งกำไรจากเงินลงทุนในบริษัทร่วม	(428,913)	16,635,026
เงินสดสุทธิได้มาจากกิจกรรมดำเนินงาน	15,403,750	33,743,751

รูปที่ 2.3 ตัวอย่างของงบกระแสเงินสด

จากรูปที่ 2.3 แสดงตัวอย่างของงบงบกระแสเงินสด โดยงบกระแสเงินสดจะประกอบด้วย กระแสเงินสดจากกิจกรรมการดำเนินงาน กระแสเงินสดจากกิจกรรมการลงทุน และกระแสเงินสดจากกิจกรรมจัดหาเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 การจัดเตรียมข้อมูล (Data Preparation)

2.2.1 การจัดการกับชุดข้อมูลที่ไม่สมดุล

ข้อมูลที่ไม่สมดุล (Imbalance data) หมายถึงข้อมูลที่มีการกระจายของคลาสหรือกลุ่มไม่เท่าเทียมกัน ซึ่งเมื่อพิจรณาเกี่ยวกับชุดข้อมูลของการทุจริตในงบการซึ่งจะประกอบด้วยงบการเงินที่มีความโปร่งใสและงบการเงินที่มีการทุจริตซึ่งเป็นกรณีที่เกิดขึ้นเป็นส่วนน้อย ส่งผลให้ข้อมูลเกิดความไม่สมดุล โดยปัญหาที่จะเกิดขึ้นตามมาคือ แบบจำลองจะมีแนวโน้มที่เกิดความลำเอียง โดยที่แบบจำลองจะให้ความสำคัญกับคลาสที่มีจำนวนตัวอย่างมากกว่า และเพิกเฉยหรือให้ความสำคัญน้อยกว่ากับคลาสที่มีจำนวนตัวอย่างน้อย และการวัดผลด้วยค่าความแม่นอาจจะให้ผลที่สูงเกินจริง ถึงแม้ว่าแบบจำลองจะไม่สามารถจำแนกคลาสที่มีตัวอย่างจำนวนน้อยได้อย่างถูกต้อง ดังนั้นเพื่อแก้ไขปัญหาของชุดข้อมูลไม่สมดุลในการศึกษาครั้งนี้จึงใช้วิธี SMOTE เพื่อมาแก้ไขปัญหาข้อมูลที่ไม่สมดุล

SMOTE (Synthetic Minority Over-sampling Technique) คือ เทคนิคหนึ่งที่ใช้ในการแก้ปัญหาความไม่สมดุลของข้อมูลในการเรียนรู้ของเครื่อง โดยเฉพาะอย่างยิ่งในกรณีที่มีตัวอย่างของคลาส มีจำนวนน้อยมากเมื่อเปรียบเทียบกับคลาสอื่น ๆ โดยมีหลักการก็คือจะทำการเลือกตัวอย่างที่อยู่ใกล้เคียง (k-nearest neighbors) แล้วทำการสร้างตัวอย่างใหม่ระหว่างตัวอย่างเดิมและตัวอย่างที่เลือกขึ้นมา โดยตัวอย่างใหม่ที่สร้างขึ้นมานี้จะเป็นการผสมระหว่างตัวอย่างเดิมและตัวอย่างใกล้เคียง ทำให้ได้ตัวอย่างที่สมดุลมากขึ้น (Chawla et al., 2022)

2.2.2 การแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปของตัวเลข

การเรียนรู้ของเครื่องเป็นการทำงานที่อยู่บนพื้นฐานของตัวเลข ดังนั้นจึงต้องทำการแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปแบบของตัวเลข โดยคุณลักษณะที่มีลักษณะเรียงลำดับได้ (Ordinal) จะทำการแปลงตัวเลขให้มีลักษณะเป็นลำดับ สำหรับคุณลักษณะที่ไม่สามารถเรียงลำดับได้ (Nominal) จะทำการเปลี่ยนให้อยู่ในรูปแบบของค่าทวิภาค (Binary values) โดยจะมีค่าเป็น 0 หรือ 1 เท่านั้น ซึ่งการแปลงข้อมูลในลักษณะนี้จะเรียกว่า One-Hot Encoding (Chaiyadecha, 2020)

2.2.3 การปรับช่วงข้อมูล

การปรับขอบเขตของข้อมูล ให้อยู่ในช่วงเดียวกัน เพื่อช่วยลดความซับซ้อนของการคำนวณ เพิ่มประสิทธิภาพของอัลกอริธึม รวมถึงช่วยลดความเสี่ยงของการเกิด overfitting โดยในการศึกษาครั้งนี้จะทำการปรับช่วงของข้อมูล โดยใช้วิธี Min-max scaler (Bishop and Nasrabadi, 2006) โดยแสดงดังสมการที่ (2.1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$x_{scale} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.1)$$

โดยที่ x_{scale} ค่าของ x ที่ถูกปรับให้อยู่ในช่วง 0 ถึง 1

x ค่าจริงของข้อมูล

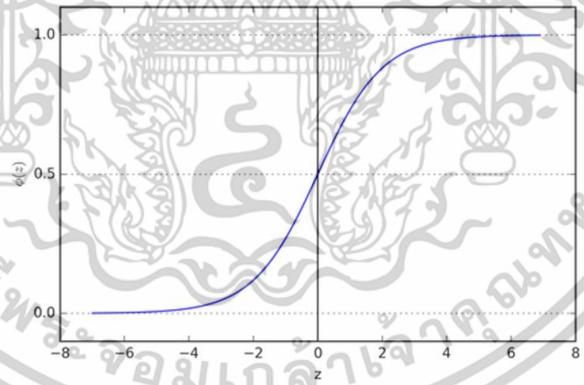
x_{min} ค่าที่ต่ำสุดในช่วงข้อมูล

x_{max} ค่าที่สูงสุดในช่วงข้อมูล

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

2.3.1 การถดถอยลอจิสติกพหุคูณ (Multiple Logistic Regression)

การถดถอยลอจิสติกพหุคูณ เป็นหนึ่งในเทคนิคทางสถิติและการเรียนรู้ของเครื่อง ที่ถูกพัฒนามาจากการถดถอยเชิงเส้น เพื่อประมาณค่าความน่าจะเป็นหรือทำนายผลลัพธ์ที่สนใจ คุณลักษณะของลอจิสติกฟังก์ชัน มีความแตกต่างกับการถดถอยเชิงเส้นตรง ตรงที่มีความน่าจะเป็นที่สนใจอยู่ในรูปของสมการที่เรียกว่าลอจิสติกฟังก์ชัน ดังรูปที่ 2.4 ซึ่งเป็นลักษณะของลอการิทึมธรรมชาติ ของอัตราส่วนของความน่าจะเป็นของผลลัพธ์ที่สนใจกับความน่าจะเป็นของผลลัพธ์ที่ไม่สนใจ (Saini, 2023)



รูปที่ 2.4 ฟังก์ชันของการถดถอยลอจิสติกพหุคูณ

รูปที่ 2.4 เป็นกราฟของการถดถอยลอจิสติกพหุคูณ โดยจะแสดงความสัมพันธ์ของตัวแปรอิสระและความน่าจะเป็นซึ่งจะอยู่ในช่วงระหว่าง 0 ถึง 1 ดังสมการที่ (2.2)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

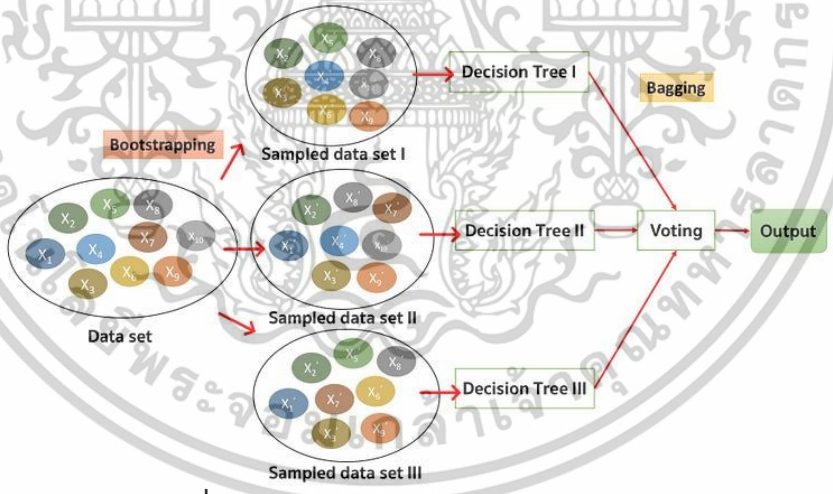
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การพิจารณาเหตุการณ์ที่สนใจสามารถใช้ Odd Ratio ในการอธิบายได้ โดย Odd Ratio คือ อัตราส่วนของโอกาสที่จะเกิดเหตุการณ์ที่สนใจ กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ดังสมการที่ (2.3)

$$\text{Odd Ratio} = \frac{P(\text{ผลลัพธ์ที่สนใจ})}{P(\text{ผลลัพธ์ที่ไม่สนใจ})} = \frac{P}{1-P} \tag{2.3}$$

2.3.2 ป่าสุ่ม (Random Forest)

แบบจำลองที่ถูกพัฒนาขึ้นมาจากต้นไม้ตัดสินใจ (Decision tree) ซึ่งจะมีการสร้างต้นไม้ตัดสินเป็นตัวจำแนกที่มีความแม่นต่ำ (Weak classifier) หลาย ๆ แบบจำลอง โดยเป็นการใช้ข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเทรนจะเลือกส่วนของข้อมูลที่แตกต่างกัน โดยการสุ่มคุณลักษณะของแต่ละแบบจำลองเพื่อให้เกิดความหลากหลายในขณะที่ทำการสอนแบบจำลอง โดยจุดประสงค์ของการสุ่มคุณลักษณะนั้น จะทำเพื่อลดค่าสหสัมพันธ์ (Correlation) ของแต่ละแบบจำลองเพื่อให้แต่ละแบบจำลองมีอิสระต่อกัน แล้วจากนั้นจึงนำการตัดสินใจของแบบจำลองเหล่านั้นมาทำการโหวตกันว่าคลาสไหนถูกเลือกมากที่สุด (International Business Machines, 2023)



รูปที่ 2.5 หลักการทำงานของแบบจำลองป่าสุ่ม

จากรูปที่ 2.5 แสดงตัวอย่างของการทำงานของแบบจำลองป่าสุ่ม โดยจะเริ่มโดยการทำการสุ่มข้อมูลที่ไม่เหมือนกันออกมาทั้งหมด ให้มีจำนวนชุดข้อมูลเป็นจำนวน n ชุดที่ไม่เหมือนกัน (Bootstrapping) ตามจำนวนต้นไม้ตัดสินใจในแบบจำลองป่าสุ่ม ซึ่งในแต่ละต้นไม้ตัดสินใจจะมีคุณลักษณะไม่เหมือนกัน อีกทั้งข้อมูลที่ได้ก็จะไม่ครบทุกระเบียน (Record) จากนั้นจะทำการสร้าง

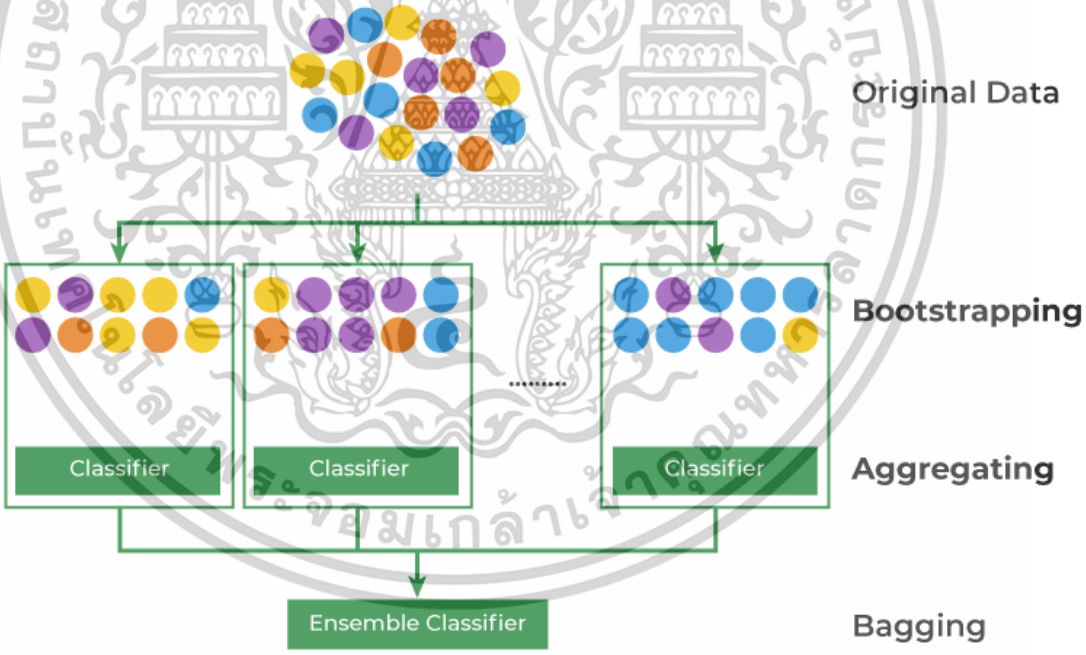
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองต้นไม้ตัดสินใจ แล้วจะนำผลลัพธ์จากการทำนายจากต้นไม้แต่ละต้นมาโหวต โดยผลการทำนายสุดท้ายจะถูกเลือกจากการทำนายที่ได้รับการเลือกโหวตมากที่สุด

2.3.3 เกรเดียนต์ บูสติ้ง (Gradient Boosting)

แบบจำลองที่เป็นหนึ่งในวิธีการเรียนรู้แบบกลุ่ม (Ensemble learning) โดยการนำแบบจำลองมาเรียนรู้หลาย ๆ ครั้ง เพื่อเพิ่มประสิทธิภาพของแบบจำลอง ซึ่งเทคนิคที่นำมาสร้างแบบจำลองจะประกอบด้วย เทคนิคการ Bagging และ Boosting ดังนี้ (Singh, 2023)

Bagging เป็นเทคนิคที่สร้างแบบจำลองหลายชุดขึ้นมา โดยสุ่มตัวอย่างข้อมูลจากชุดข้อมูลฝึกฝนหลาย ๆ ครั้ง การสุ่มนี้ใช้วิธีสุ่มแบบมีการแทนที่ ทำให้สามารถสุ่มตัวอย่างจากข้อมูลเดียวกันได้หลายครั้ง หลักการคือการเพิ่มจำนวนข้อมูลโดยการสุ่มตัวอย่างไปยังกลุ่มต่าง ๆ โดยแต่ละกลุ่มจะมีข้อมูลที่แตกต่างกัน ซึ่งจะทำให้เราได้ข้อมูลชุดตัวอย่างหลายชุดจากข้อมูลชุดเดิม

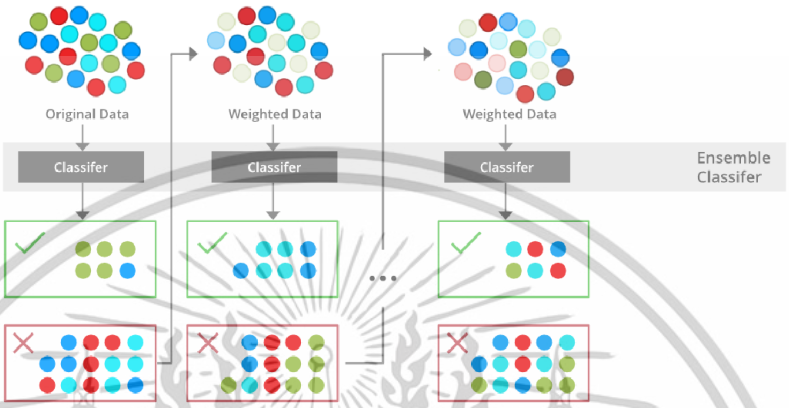


รูปที่ 2.6 ลักษณะการทำงานของ Bagging

จากรูปที่ 2.6 แสดงตัวอย่างของการทำงานของ Bagging โดยจะเริ่มโดยการทำการสุ่มข้อมูลออกมาทั้งหมด n ชุดที่ไม่เหมือนกัน โดยข้อมูลแต่ละชุดที่ได้จากการสุ่มข้อมูล จะมีจำนวนระเบียนและคุณลักษณะที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

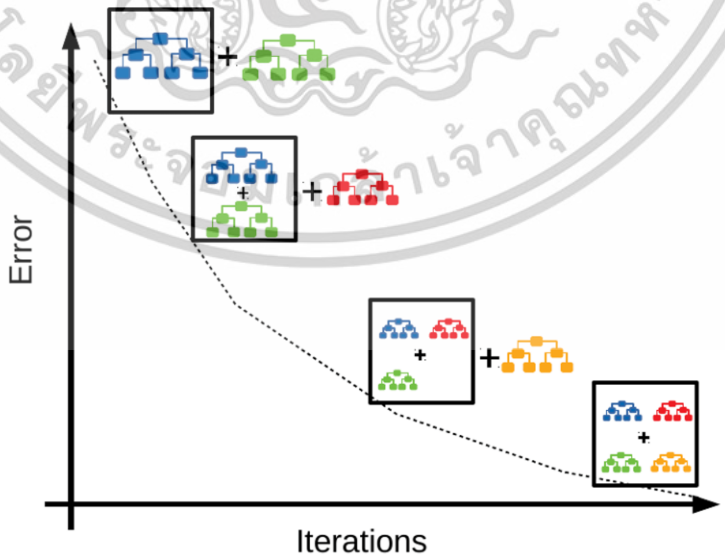
Boosting เป็นการใช้ตัวจำแนกที่มีความแม่นต่ำ (Weak classifier) มาทำนายข้อมูล จากนั้นจะให้ตัวจำแนกที่มีความแม่นต่ำอีกตัว มาแก้ไขข้อผิดพลาดของตัวจำแนกก่อนหน้า ซึ่งจะมีการทำแบบนี้ไปจนกว่าความคลาดเคลื่อนของแบบจำลองลดลงจนถึงน้อยที่สุด



รูปที่ 2.7 ลักษณะการทำงานของ Boosting

จากรูปที่ 2.7 แสดงตัวอย่างของการทำงานของ Boosting โดยจะเริ่มทำการทำนายข้อมูลโดยใช้แบบจำลองที่มีความแม่นต่ำ จากนั้นจะส่งข้อมูลของการทำนายไปยังแบบจำลองถัดไปเพื่อแก้ไขความผิดพลาดของแบบจำลองก่อนหน้า ทำให้ค่าความคลาดเคลื่อนที่ได้ลดลง

เกรเดียนต์ บูสติ้งเป็นเทคนิคการเรียนรู้ของเครื่องที่ใช้ในการแก้ปัญหาทั้งการถดถอย (Regression) และการจำแนกประเภท (Classification) โดยใช้หลักการของการเรียนรู้แบบกลุ่ม



รูปที่ 2.8 ลักษณะการทำงานของเกรเดียนต์ บูสติ้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.8 แสดงลักษณะการทำงานของเกรเดียนต์ บูสติ้ง โดยการสร้างแบบจำลองที่ประกอบด้วยตัวจำแนกหลายตัว (Classifiers) ที่มีความแม่นต่ำ จากนั้นใช้ตัวทำนายเหล่านี้แก้ไขข้อผิดพลาดที่เกิดขึ้นในขั้นตอนก่อนหน้า เพื่อสร้างตัวจำแนกใหม่ที่แม่นยำมากขึ้นเรื่อย ๆ

XGBoost เป็นเทคนิคการเรียนรู้ของเครื่อง (Machine learning) ที่ใช้สำหรับการสร้างแบบจำลองที่แม่นยำ โดยเฉพาะอย่างยิ่งในปัญหาของการจัดกลุ่ม (classification) และการทำนายตัวเลข (regression) โดย XGBoost เป็นการพัฒนาเพิ่มเติมจากวิธีการเกรเดียนต์ บูสติ้ง ซึ่งมีประสิทธิภาพสูงและความเร็วในการคำนวณที่ดีกว่า โดยคุณสมบัติที่โดดเด่นของ XGBoost คือความเร็วในการประมวลผลสูง ความแม่นสูง ความสามารถในการจัดการกับข้อมูลที่มีคุณลักษณะจำนวนมาก สามารถจัดการกับข้อมูลสูญหายและมีค่าว่างได้ดี (Tseng, 2018)

2.4 การวัดประสิทธิภาพของแบบจำลอง (Model Evaluation)

การวัดประสิทธิภาพของแบบจำลองเป็นขั้นตอนที่สำคัญในการประเมินว่าผลลัพธ์ที่ได้จากแบบจำลองนั้นมีความแม่นและเหมาะสมเพียงใด สำหรับการจำแนกประเภท (Classification) มักจะใช้เมทริกซ์ความสับสน (Confusion Matrix) เป็นเครื่องมือหลักในการประเมินประสิทธิภาพของแบบจำลอง ซึ่งเมทริกซ์ความสับสนจะแสดงจำนวนของการคาดการณ์ที่ถูกต้องและไม่ถูกต้องในแต่ละคลาส ดังตารางที่ 2.1 (Narkhede, 2018)

ตารางที่ 2.1 เมทริกซ์ความสับสน

ข้อมูลจริง	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	TP	FN
ไม่ทุจริต	FP	TN

โดยมีค่าความเป็นไปได้ทั้งหมด 4 ค่าดังนี้

1. True Positive (TP): จำนวนของตัวอย่างที่ถูกคาดการณ์ว่าเป็นคลาสบวก และเป็นคลาสบวกจริง หรือจำนวนบริษัทที่ถูกคาดการณ์ว่าทุจริต และทุจริตจริง
2. True Negative (TN): จำนวนของตัวอย่างที่ถูกคาดการณ์ว่าเป็นคลาสลบ และเป็นคลาสลบจริงหรือจำนวนบริษัทที่ถูกคาดการณ์ว่าไม่ทุจริต และไม่ทุจริตจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. False Positive (FP): จำนวนของตัวอย่างที่ถูกคาดการณ์ว่าเป็นคลาสบวก แต่จริงๆ แล้วเป็นคลาสลบ หรือจำนวนบริษัทที่ถูกคาดการณ์ว่าทุจริต แต่ไม่ได้ทุจริตจริง
4. False Negative (FN): จำนวนของตัวอย่างที่ถูกคาดการณ์ว่าเป็นคลาสลบ แต่จริงๆ แล้วเป็นคลาสบวก หรือจำนวนบริษัทที่ถูกคาดการณ์ว่าไม่ทุจริต แต่มีการทุจริตจริง

2.4.1 ค่าความแม่น (Accuracy)

การวัดค่าความแม่นของการทำนายในแบบจำลอง โดยวัดจากสัดส่วนของการทำนายที่ถูกต้อง ต่อจำนวนกลุ่มตัวอย่างทั้งหมด ดังสมการที่ (2.4) (Tharwat, 2021)

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{2.4}$$

2.4.2 ค่าความเที่ยง (Precision)

การวัดค่าความเที่ยงของแบบจำลองเมื่อทำนายว่าเป็นบวก โดยวัดจากสัดส่วนของการทำนายที่ถูกต้องจากจำนวนที่ทำนายว่าเป็นบวก ดังสมการที่ (2.5) (Tharwat, 2021)

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2.5}$$

2.4.3 ค่าการเรียกคืน (Recall)

การวัดค่าว่าแบบจำลองสามารถตรวจจับกรณีที่เป็นบวกได้มากน้อยเพียงใด โดยวัดจากสัดส่วนของจำนวนตัวอย่างที่ทำนายเป็นบวกและถูกต้องจากจำนวนตัวอย่างที่เป็นบวกทั้งหมด ดังสมการที่ (2.6) (Tharwat, 2021)

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{2.6}$$

2.4.4 ค่าคะแนนเอฟ1 (F1-Score)

มาตรวัดที่ใช้สำหรับการประเมินเมื่อมีความไม่สมดุลระหว่างกลุ่มตัวอย่าง โดยใช้ค่าเฉลี่ยเชิงฮาร์โมนิกของค่าความแม่น และ ค่าการเรียกคืน ดังสมการที่ (2.7) (Tharwat, 2021)

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.7}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 การหาค่าตัวแปรที่สำคัญที่มีผลต่อแบบจำลอง

การหาค่าตัวแปรที่มีผลต่อแบบจำลอง สามารถวัดโดยใช้ค่าเพอร์มิวเทชัน อิมพอร์แทนซ์ (Permutation importance) ซึ่งเป็นวิธีการวัดค่าความสำคัญของตัวแปรของแบบจำลอง โดยวิธีนี้จะวัดผลกระทบว่าตัวแปรใดที่มีผลต่อการทำงานของแบบจำลอง โดยหลักการทำงานจะเริ่มจาก การวัดประสิทธิภาพเบื้องต้นของแบบจำลอง เพื่อใช้เป็นค่าอ้างอิง จากนั้นจะทำการสุ่มสลับค่าของตัวแปรที่ต้องการวัด จากนั้นทำการวัดค่าประสิทธิภาพของแบบจำลองใหม่ ซึ่งสามารถคำนวณได้ด้วยมาตรวัด เช่น ค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน หรือค่าคะแนนเอฟ1 จากนั้นจะทำการเปรียบเทียบประสิทธิภาพก่อนและหลังจากการสลับค่าของตัวแปร ถ้าประสิทธิภาพมีความแตกต่างจากเดิมสูง แสดงให้เห็นว่าตัวแปรนั้นมีความสำคัญที่สูง (Terence, 2023)

2.6 งานวิจัยที่เกี่ยวข้อง

Wyrobek (2020) ได้ศึกษาเกี่ยวกับประสิทธิภาพของแบบจำลองสำหรับการตรวจสอบความผิดปกติของงบการเงินที่สำคัญในองค์กร โดยใช้งบการเงินประจำปีเพื่อใช้ในการตรวจสอบการทุจริตทางการเงินในรูปแบบต่าง ๆ โดยทำการใช้งบการเงินของบริษัททั้งหมด 54 บริษัทที่ถูกระบุว่าทุจริต และอีก 58 บริษัทที่ถูกระบุว่าโปร่งใส ของตลาดหลักทรัพย์ NYSE และ NASDAQ ซึ่งมีจำนวนงบรวมทั้งหมด 1317 งบการเงิน และได้สร้างแบบจำลองโดยทำการแบ่งสัดส่วนชุดข้อมูลโดยชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบในสัดส่วน 70:30 โดยแบบจำลองที่นำมาใช้ประกอบด้วยการถดถอยลอจิสติกพหุคูณ โครงข่ายประสาทเทียม ซัพพอร์ตเวกเตอร์แมชชีน และป่าสุ่ม เพื่อเปรียบเทียบว่าแบบจำลองที่สามารถตรวจจับทุจริตได้ดีที่สุด โดยวัดผลด้วยค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1 ผลการศึกษาพบว่าวิธีการป่าสุ่มให้ผลลัพธ์ดีที่สุด ซึ่งมีความแม่นอยู่ที่ร้อยละ 0.94 ค่าความเที่ยงร้อยละ 0.80 ค่าการเรียกคืน 0.76 และค่าคะแนนเอฟ1 0.78 และรายการทางบัญชีที่เป็นปัจจัย 3 อันดับแรกที่ทำให้เกิดความผิดปกติในงบการเงิน คือ ค่าเสื่อมราคาและค่าตัดจำหน่าย กระแสเงินสดจากกิจกรรมการจัดหาเงิน และกระแสเงินสดจากกิจกรรมการดำเนินงาน

Liu et al. (2021) ได้ศึกษาปัจจัยที่ทำให้เกิดการทุจริตในงบการเงินโดยใช้การเรียนรู้ของเครื่อง ในการวิเคราะห์งบการเงินประจำปี รวมถึงอัตราส่วนทางการเงิน โดยได้ทำการเก็บรวบรวมงบการเงินของตลาดหลักทรัพย์ในประเทศจีนตั้งแต่ปี ค.ศ. 2012 – 2017 และได้ใช้ตัวแปรที่เกี่ยวข้องกับงบการเงินทั้งหมด 363 ตัวแปรในการวิเคราะห์ โดยแบบจำลองที่นำมาใช้ประกอบด้วย LightGBM, Support vector machine, ป่าสุ่ม และการถดถอยลอจิสติกพหุคูณ ซึ่งผลการศึกษาพบว่า LightGBM สามารถ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงผลลัพธ์ออกมาได้ดีที่สุด โดยมีค่า AUC เท่ากับ 0.86 รองลงมาคือ XGBoost เท่ากับ 0.83 และนอกจากนี้ยังมีการศึกษาตัวแปรที่มีผลกับการทุจริตในงบการเงิน ซึ่ง 3 อันดับแรก ซึ่งก็คือยอดดอกเบี้ยค้างจ่าย กระแสเงินสดจากกิจการลงทุน และค่าใช้จ่ายที่เกิดจากการดำเนินงาน

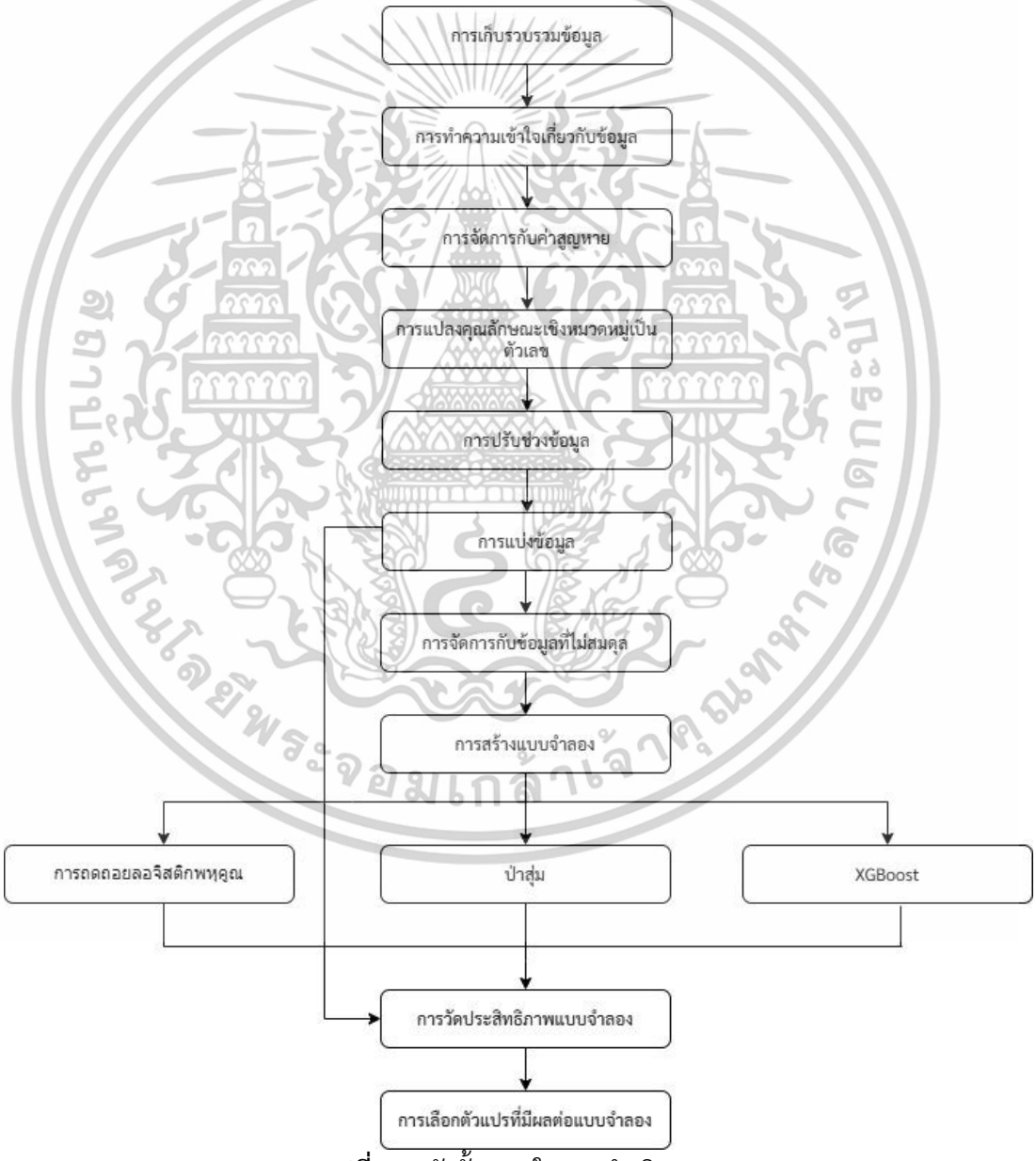
Aftabi et al. (2023) ได้ศึกษาแบบจำลองที่ใช้ในการทำนายการบิดเบือนและการทุจริตในงบการเงิน โดยได้มีการเก็บรวบรวมงบการประจำปี ทั้งหมด 490 งบการเงิน จาก 10 ธนาคารที่แตกต่างกันในประเทศอิหร่าน โดยเก็บรวมรวมงบการเงินระหว่างปี ค.ศ. 2014-2019 โดยวิเคราะห์แยกตามประเภทของ รายได้ เงินกู้ เงินฝาก และค่าใช้จ่าย โดยใช้แบบจำลอง XGBoost ซัพพอร์ตเวกเตอร์แมชชีน เพื่อนบ้านที่ใกล้ที่สุด และการถดถอยลอจิสติกพหุคูณ ในการทำนาย โดยวัดผลด้วยค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน โดยผลการศึกษาได้แสดงให้เห็นว่า XGBoost สามารถแสดงผลลัพธ์ออกมาได้ดีที่สุด โดยมีค่าความแม่นเท่ากับ 0.97 ค่าความเที่ยง 0.96 และค่าการเรียกคืน 0.94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานวิจัย

การศึกษาเรื่อง “การตรวจสอบการทุจริตในงบการเงินโดยใช้การเรียนรู้แบบมีผู้สอน” ในส่วนของการดำเนินการวิจัย เพื่อให้ได้มาซึ่งกลุ่มตัวอย่าง และได้นำไปวิเคราะห์ตามหลักสถิติต่อไป ในการวิจัยครั้งผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอนดังรูปที่ 3.1



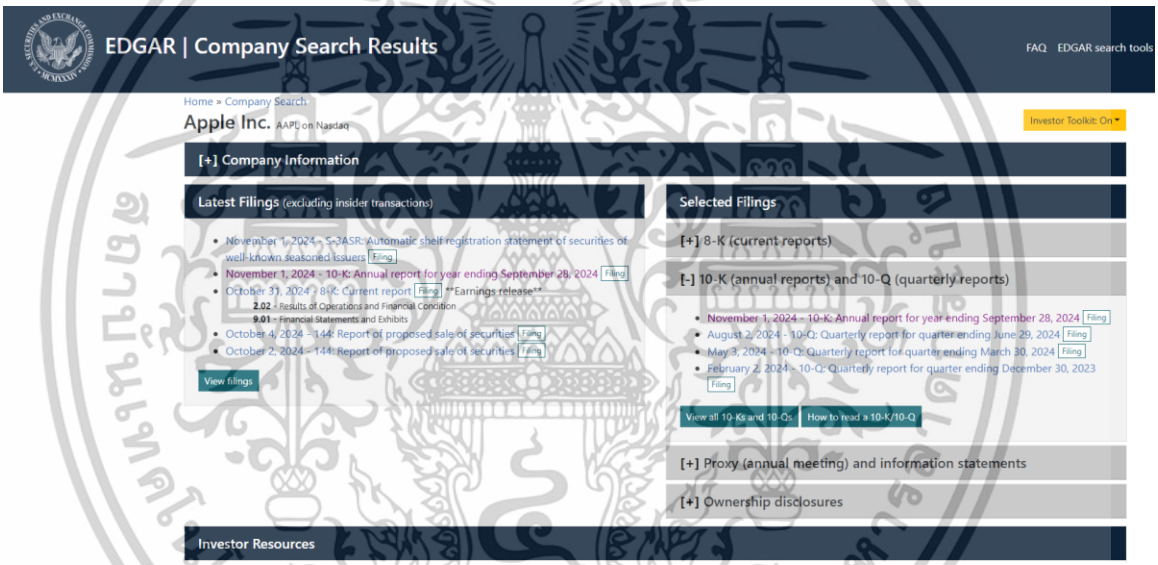
รูปที่ 3.1 ผังขั้นตอนในการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.1 แสดงแผนขั้นตอนการดำเนินงานของการวิจัยซึ่งมีขั้นตอนคือ การเก็บรวบรวมข้อมูล การทำความเข้าใจเกี่ยวกับข้อมูล การจัดการกับค่าสูญหาย การแปลงคุณลักษณะเชิงหมวดหมู่เป็นตัวเลข การปรับช่วงข้อมูล การแบ่งข้อมูล การจัดการกับข้อมูลที่ไม่สมดุล การสร้างแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost การวัดประสิทธิภาพของแบบจำลอง และการเลือกตัวแปรที่มีผลต่อแบบจำลอง

3.1 การเก็บรวบรวมข้อมูล

การตรวจสอบการทุจริตในงบการเงิน เราได้ทำการเก็บรวบรวมข้อมูลของงบการเงินของบริษัทในตลาดหลักทรัพย์ โดยมีการดำเนินการดังนี้



รูปที่ 3.2 เว็บไซต์ข้อมูลงบการเงินจากคณะกรรมการกำกับหลักทรัพย์ของสหรัฐอเมริกา

จากรูปที่ 3.2 การเก็บข้อมูลของงบการเงินจากคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ของสหรัฐอเมริกา (U.S. Securities and Exchange Commission) จาก https://www.sec.gov/edgar/browse โดยใช้งบการเงินทั้งหมด 349 งบ จาก 70 บริษัท ที่ถูกระบุว่าเกิดการทุจริต โดยจะใช้งบการเงินย้อนหลังนับจากปีที่เกิดการทุจริตไม่เกิน 5 ปี ต่อจากนั้นผู้วิจัยได้ทำการนำบริษัทเหล่านี้ มาจับคู่กับงบการเงินของบริษัทที่ถูกระบุว่ามีการดำเนินงานอย่างโปร่งใส ทั้งหมด 1,013 งบ จาก 215 บริษัท โดยบริษัทเหล่านี้ ผู้วิจัยได้เลือกมาจากตลาดหลักทรัพย์ NASDAQ และ NYSE ของประเทศสหรัฐอเมริกา โดยบริษัทเหล่านี้ผู้วิจัยจะเลือกโดยอิงจากประเภทของอุตสาหกรรม และขนาดของอุตสาหกรรม โดยแบ่งตามขนาด เล็ก กลาง ใหญ่ และ ใหญ่พิเศษ ที่ใกล้เคียงกัน โดยตัวอย่างของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวอย่างชุดข้อมูลของงบการเงิน

Fraud	industry	symbol	calendarYear	cashAndCashEquivalents
0	energy	BORR	2022	108,000,000
0	energy	BORR	2021	34,900,000
0	energy	BORR	2020	19,200,000
0	energy	BORR	2019	59,100,000
0	energy	BORR	2018	27,900,000
0	energy	CAPL	2022	16,054,000
0	energy	CAPL	2021	7,648,000
0	energy	CAPL	2020	513,000
0	energy	CAPL	2019	1,780,000
0	energy	CAPL	2018	3,191,000
0	energy	DK	2022	841,300,000
0	energy	DK	2021	856,500,000
0	energy	DK	2020	787,500,000
0	energy	DK	2019	955,300,000
0	energy	DK	2018	1,079,300,000
0	energy	DKL	2022	7,970,000
0	energy	DKL	2021	4,292,000
0	energy	DKL	2020	4,243,000
0	energy	DKL	2019	554,5000
0	energy	DKL	2018	4,522,000
0	energy	DO	2022	63,041,000

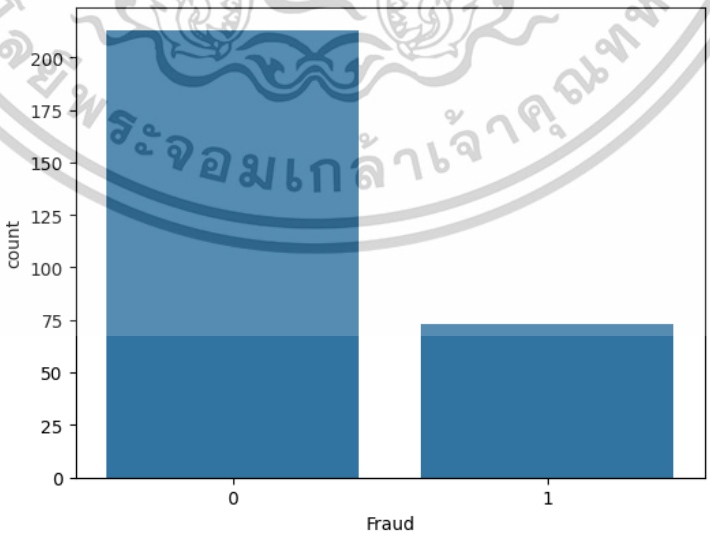
จากตารางที่ 3.1 แสดงตัวอย่างชุดข้อมูลที่ทำการเก็บมาจากจากคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ของสหรัฐอเมริกา โดยงบการเงินของบริษัทที่ไม่เกิดการทุจริตจะถูกแทนค่าด้วย 0 และงบการเงินของบริษัทที่เกิดการทุจริตจะถูกแทนค่าด้วย 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ประเภทของอุตสาหกรรม

ประเภทอุตสาหกรรม	จำนวนบริษัท	จำนวนงบการเงิน
โทรคมนาคม (Telecommunications)	13	54
เทคโนโลยี (Technology)	41	203
การเงิน (Financial)	59	289
สินค้าอุปโภคบริโภค (Consumer Staples)	27	134
พลังงาน (Energy)	19	85
การดูแลสุขภาพ (Healthcare)	33	162
อสังหาริมทรัพย์ (Real Estate)	29	144
สินค้าฟุ่มเฟือย (Consumer Discretionary)	23	112
วัสดุพื้นฐาน (Basic Materials)	21	100
สาธารณูปโภค (Utilities)	20	79

จากตารางที่ 3.2 แสดงการแบ่งข้อมูลตามประเภทของของอุตสาหกรรมที่บริษัทดำเนินกิจการอยู่ เช่น โทรคมนาคม เทคโนโลยี การเงิน และอื่นๆ จำนวนบริษัทที่อยู่ในแต่ละอุตสาหกรรม และจำนวนงบการเงินที่บริษัททั้งหมดในแต่ละอุตสาหกรรมได้จัดทำขึ้น



รูปที่ 3.3 จำนวนบริษัทที่ทุจริต และไม่ทุจริต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.3 แสดงสัดส่วนของจำนวนงบของบริษัทที่เกิดการทุจริต และบริษัทที่ไม่เกิดการทุจริต โดยจำนวนงบที่เกิดการทุจริตมีจำนวน 329 งบการเงิน และงบของบริษัทที่ไม่เกิดการทุจริตมีจำนวน 1,013 งบการเงิน

3.2 การทำความเข้าใจเกี่ยวกับข้อมูล

การทำความเข้าใจเกี่ยวกับข้อมูลที่จะนำมาใช้ ประกอบด้วย ข้อมูลของงบแสดงฐานะทางการเงิน งบกำไร (ขาดทุน) งบกระแสเงินสด และ อัตราส่วนทางการเงิน โดยข้อมูลทั้งหมดจะมี 122 คุณลักษณะ 1,362 ระเบียน และมีป้ายกำกับ โดยบริษัทที่มีการทุจริตจะมีการกำหนดให้เป็น “1” สำหรับบริษัทที่ไม่มีการทุจริตจะถูกกำหนดให้เป็น “0” โดยมีรายละเอียดของข้อมูลบางส่วน ดังที่แสดงในตารางที่ 3.2

ตารางที่ 3.3 ข้อมูลคุณลักษณะของงบการเงิน

กลุ่มข้อมูล	จำนวนคุณลักษณะ	ชนิดของข้อมูล
ประเภทอุตสาหกรรม	1	จำแนกประเภท
ขนาดของอุตสาหกรรม	1	จำแนกประเภท
งบแสดงฐานะทางการเงิน	43	จำนวนจริง
งบกำไรขาดทุน	24	จำนวนจริง
งบกระแสเงินสด	25	จำนวนจริง
อัตราส่วนทางการเงิน	28	จำนวนจริง

จากตารางที่ 3.3 แสดงจำนวนนวนคุณลักษณะของข้อมูล โดยแบ่งตามประเภทกลุ่มข้อมูล ได้แก่ ประเภทของอุตสาหกรรม ขนาดของอุตสาหกรรม งบแสดงฐานะทางการเงิน งบกำไรขาดทุน งบกระแสเงินสด และอัตราส่วนทางการเงิน

3.3 การจัดเตรียมข้อมูล

3.3.1 การจัดการกับค่าสูญหาย

การจัดการกับข้อมูลที่มีค่าสูญหายในคุณลักษณะบัญชีเฉพาะบริษัท ซึ่งเป็นคุณลักษณะที่ไม่พบในบริษัทอื่น ๆ จะมีการดำเนินการโดยการตัดคุณลักษณะดังกล่าวออกจากชุดข้อมูลทั้งหมด ทำให้จำนวนคุณลักษณะที่นำมาวิเคราะห์ลดลงจาก 122 เหลือ 86 คุณลักษณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 การแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปของตัวเลข (One-Hot Encoding)

การจัดการข้อมูลที่มีลักษณะเชิงหมวดหมู่โดยประเภทของอุตสาหกรรม (industry) และ ขนาดของอุตสาหกรรม (Market_cap) ไม่ได้อยู่ในรูปของตัวเลขจึงต้องทำการแปลงเป็นตัวเลข โดยจะทำการแปลงคุณลักษณะเชิงหมวดหมู่ให้อยู่ในรูปแบบของตัวเลข

ตารางที่ 3.4 การแปลงคุณลักษณะเชิงหมวดหมู่การให้อยู่ในรูปของตัวเลข

สัญลักษณ์	ปีงบการเงิน	อุตสาหกรรมการเงิน	อุตสาหกรรมการเงิน
BORR	2022	1	0
BORR	2021	1	0
BORR	2020	1	0
BORR	2019	1	0
BORR	2018	1	0
CAPL	2022	1	0
CAPL	2021	1	0
CAPL	2020	1	0
CAPL	2019	1	0
CAPL	2018	1	0
DK	2022	1	0
DK	2021	1	0
DK	2020	1	0
DK	2019	1	0
DK	2018	1	0

จากตารางที่ 3.4 แสดงการเปลี่ยนข้อมูลที่มีลักษณะเชิงหมวดหมู่ให้อยู่ในรูปของตัวเลข โดยในรูปจะแสดงตัวอย่างของข้อมูลที่อยู่ในอุตสาหกรรมพลังงาน จะทำการแทนค่าในคุณลักษณะอุตสาหกรรมพลังงาน (industry_energy) ด้วย 1 และข้อมูลที่ไม่อยู่ในอุตสาหกรรมพลังงานจะถูกแทนด้วย 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3 การปรับช่วงของข้อมูล

การปรับช่วงของข้อมูลที่เป็นตัวเลข ให้อยู่ในช่วงเดียวกัน โดยทำการปรับช่วงโดยแบ่งตามกลุ่มของข้อมูลได้แก่ กลุ่มของงบแสดงฐานะทางการเงิน งบกำไรขาดทุน และงบกระแสเงินสด โดยใช้วิธี Min-max scaler

ตารางที่ 3.5 การปรับช่วงของข้อมูล ให้อยู่ในช่วงเดียวกัน

inventory	otherCurrentAssets	totalCurrentAssets
0.281743	0.190797	0.494888
0.066342	0	0.066342
0.07	0.062857	0.132857
0.485844	0.022051	0.507895
0.007258	0	0.007258
0.004823	0	0.004823
0.18161	0.077028	0.258638
0.158364	0.0026	0.160964
0.07059	0	0.07059
0.243441	0.011676	0.255117
870.8558	-869.985	0.870856
0.079599	0.094071	0.17367
0.037664	0	0.037664
0.147237	0	0.147237
0.036548	0	0.036548
0.025599	0.010689	0.036288
0.013328	0.065941	0.079269

จากตารางที่ 3.5 แสดงตัวอย่างของการปรับช่วงข้อมูลของแต่ละบัญชีในงบการเงิน ให้อยู่ในช่วง 0 ถึง 1 โดยใช้วิธี Min-max scaler

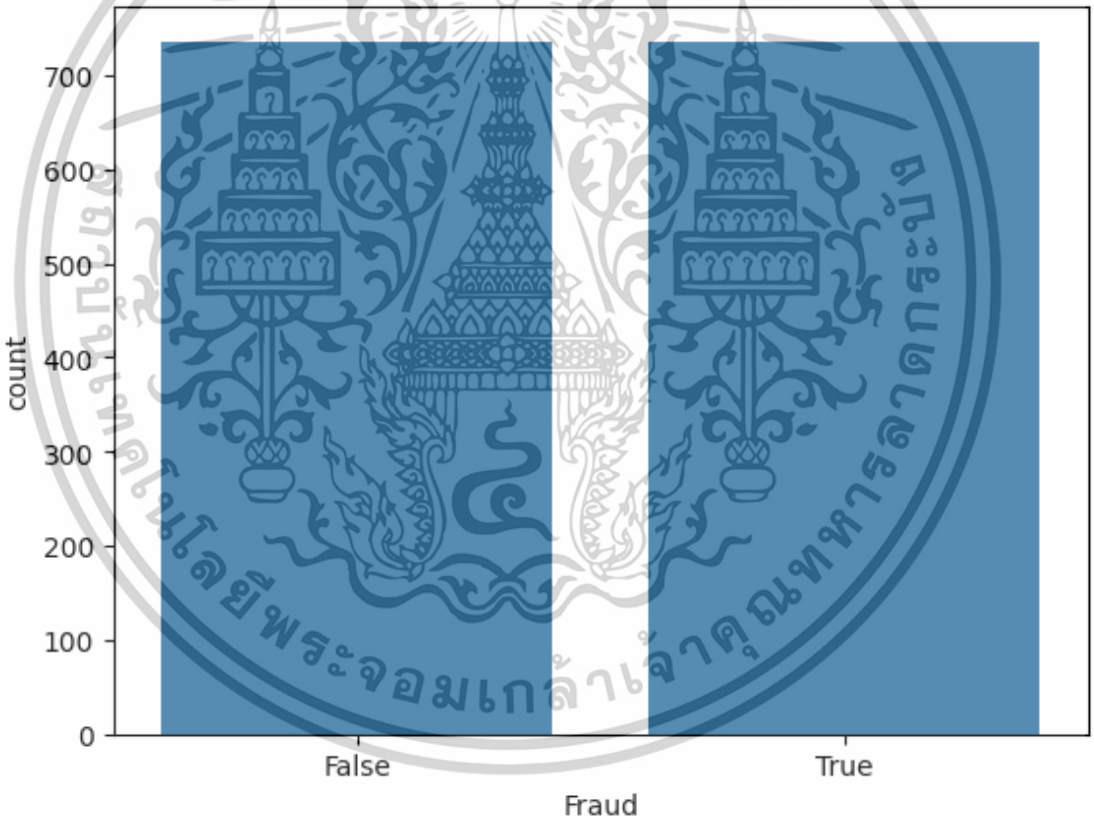
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.4 การแบ่งข้อมูล

การแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ โดยทำการแบ่งในสัดส่วน 70:30 ตามงานวิจัยของ Wyrobek (2021) โดยแบ่งข้อมูลเป็นข้อมูลชุดฝึกฝน 989 ระเบียน และข้อมูลชุดทดสอบจำนวน 423 ระเบียน

3.3.5 การจัดการกับข้อมูลที่ไม่สมดุล

เนื่องจากข้อมูลที่ใช้ในการทำนายมีความไม่สมดุล โดยที่มีจำนวนบริษัทที่ไม่ทุจริตมากกว่าบริษัทที่เกิดการทุจริต ซึ่งส่งผลให้แบบจำลอง ให้ความสำคัญกับบริษัทที่ไม่เกิดการทุจริต มากกว่าบริษัทที่เกิดการทุจริต ส่งผลให้ประสิทธิภาพในการทำนายลดลง ดังนั้นจึงทำการเพิ่มข้อมูลโดยใช้วิธีการเพื่อนบ้านที่ใกล้ที่สุดในการเพิ่มข้อมูล ดังรูปที่ 3.4



รูปที่ 3.4 ชุดข้อมูลหลังทำการการ Oversampling ด้วยวิธีการ SMOTE

จากรูปที่ 3.4 แสดงการเพิ่มจำนวนข้อมูลของบริษัทที่เกิดการทุจริตจาก 253 ระเบียน เป็น 736 ระเบียน ด้วยวิธีการเพื่อนบ้านที่ใกล้ที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การสร้างแบบจำลอง

การนำชุดข้อมูลฝึกฝนที่ได้มาสร้างแบบจำลอง โดยแบบจำลองที่นำมาใช้ ผู้วิจัยได้เลือกมาจากการศึกษาวิจัยที่เกี่ยวข้องและเลือกวิธีที่มีประสิทธิภาพที่ดีที่สุดมาใช้ทั้งหมด 2 แบบจำลอง ได้แก่ ป่าสุ่ม และ XGBoost และได้ทำการเปรียบเทียบกับแบบจำลองการถดถอยลอจิสติกพหุคูณซึ่งเป็นแบบจำลองพื้นฐานที่ใช้ในการจำแนกประเภท โดยกำหนดพารามิเตอร์ต่าง ๆ ดังตารางที่ 3.23

ตารางที่ 3.6 การกำหนดพารามิเตอร์ในแต่ละแบบจำลอง

แบบจำลอง	พารามิเตอร์
การถดถอยลอจิสติกพหุคูณ	Random state = 0, max_iter = 100
ป่าสุ่ม	Random state = 0, n_estimators = 100
XGBoost	Random state = 0, n_estimators = 100

จากตารางที่ 3.6 แสดงการกำหนดค่าของแบบจำลองการถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost

3.5 การวัดประสิทธิภาพของแบบจำลอง

การเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งหมด 3 จำลอง จะทำการเปรียบเทียบด้วยมาตรวัดประสิทธิภาพทั้งหมด 4 ค่า ได้แก่ ค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1

3.6 การเลือกตัวแปรที่มีผลต่อแบบจำลอง

การเลือกตัวแปรที่มีผลต่อการทำนายของแบบจำลอง เป็นการหาเพื่อตรวจสอบว่าตัวแปรใดที่ส่งผลต่อการเกิดทุจริตในงบการเงินมากที่สุด โดยดูจากค่าเพอร์มิวเทชัน อิมพอร์แทนซ์ โดยค่าดังกล่าวจะแสดงถึงความสำคัญของตัวแปรที่มีอิทธิพลต่อประสิทธิภาพของแบบจำลอง โดยทำการสุ่มสลับค่าตัวแปรที่สนใจออกจากแบบจำลอง ถ้าแบบจำลองมีประสิทธิภาพที่ลดลงแสดงว่าตัวแปรนั้นมีความสำคัญที่สูง โดยค่าความสำคัญที่มีผลกระทบเกินร้อยละ 10 ของประสิทธิภาพของแบบจำลองจะถือว่าเป็นตัวแปรที่มีความสำคัญ (Terence, 2023)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

4.1 ผลการวิจัย

การเปรียบเทียบประสิทธิภาพของแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost กับชุดข้อมูลฝึกฝนจำนวน 989 ระเบียน และ ข้อมูลชุดทดสอบจำนวน 423 ระเบียน ทำการเปรียบเทียบโดยใช้ ค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1

4.1.1 การเปรียบเทียบประสิทธิภาพของแบบจำลองกับข้อมูลชุดฝึกฝน

การวัดประสิทธิภาพของแบบจำลอง กับชุดข้อมูลฝึกฝนกับแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost โดยทดสอบกับข้อมูลชุดฝึกฝน 989 ระเบียน เพื่อทำนายการทุจริตและไม่ทุจริต โดยแบบจำลองที่มีค่าความแม่นสูงสุดคือ XGBoost ซึ่งมีความแม่นอยู่ที่ร้อยละ 94.50 รองลงมาเป็น ป่าสุ่ม ซึ่งมีความแม่นอยู่ที่ร้อยละ 94.02 และความแม่นน้อยสุดคือ การถดถอยลอจิสติกพหุคูณ ซึ่งมีความแม่นอยู่ที่ร้อยละ 74.05

การวัดประสิทธิภาพของแบบจำลอง กับชุดข้อมูลฝึกฝนกับแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost โดยวัดผลด้วยเมทริกซ์ความสับสนได้ผลลัพธ์ดังนี้

ตารางที่ 4.1 เมทริกซ์ความสับสนของการถดถอยลอจิสติกพหุคูณกับข้อมูลชุดฝึกฝน

การถดถอยลอจิสติกพหุคูณ	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	513	223
ไม่ทุจริต	159	577

จากตารางที่ 4.1 แสดงการวัดผลโดยใช้แบบจำลองการถดถอยลอจิสติกพหุคูณสามารถทำถูกต้อง ว่างบการเงินมีการทุจริตทั้งหมด 557 กรณี ทำนายว่าถูกว่าไม่ทุจริต 513 กรณี ทำนายว่าทุจริตแต่ผิด 159 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 223 กรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 เมทริกซ์ความสับสนของป่าสุ่มกับข้อมูลชุดฝึกฝน

ป่าสุ่ม	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	682	54
ไม่ทุจริต	34	702

จากตารางที่ 4.2 แสดงการวัดผลโดยใช้แบบจำลองป่าสุ่มสามารถทำถูกต้องว่างบการเงินมีการทุจริตทั้งหมด 682 กรณี ทำนายว่าถูกว่าไม่ทุจริต 702 กรณี ทำนายว่าทุจริตแต่ผิด 159 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 223 กรณี

ตารางที่ 4.3 เมทริกซ์ความสับสนของ XGBoost กับข้อมูลชุดฝึกฝน

XGBoost	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	687	49
ไม่ทุจริต	32	704

จากตารางที่ 4.3 แสดงการวัดผลโดยใช้แบบจำลอง XGBoost สามารถทำถูกต้องว่างบการเงินมีการทุจริตทั้งหมด 687 กรณี ทำนายว่าถูกว่าไม่ทุจริต 704 กรณี ทำนายว่าทุจริตแต่ผิด 32 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 49 กรณี

การเปรียบเทียบสิทธิภาพของแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost กับชุดข้อมูลทดสอบ โดยทำการวัดผลด้วยค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1

ตารางที่ 4.4 การวัดประสิทธิภาพของแบบจำลองกับข้อมูลชุดฝึกฝน

แบบจำลอง	**ค่าทางสถิติ**			
	ค่าความแม่น	ค่าความเที่ยง	ค่าการเรียกคืน	ค่าคะแนนเอฟ1
การถดถอยลอจิสติกพหุคูณ	74.05%	76.34%	69.70%	72.87%
ป่าสุ่ม	94.02%	92.66%	95.25%	93.94%
XGBoost	**94.49%**	**93.34%**	**95.55%**	**94.43%**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.4 แบบจำลองที่มีประสิทธิภาพมากที่สุดคือ XGBoost และ ป่าสุ่ม โดยทั้งสองแบบจำลองจะมีประสิทธิภาพที่ใกล้เคียงกัน โดยมีค่าความแม่นอยู่ที่ร้อยละ 94.49 และ 94.02 ตามลำดับ ในขณะที่ค่าความเที่ยงอยู่ที่ร้อยละ 93.34 และ 92.66 ตามลำดับ ค่าการเรียกคืนอยู่ที่ร้อยละ 95.55 และ 95.25 ตามลำดับ และค่าค่าคะแนนเอฟ1อยู่ที่ 94.43 และ 93.94 ตามลำดับ

4.1.2 การเปรียบเทียบประสิทธิภาพของแบบจำลองกับข้อมูลชุดทดสอบ

การวัดประสิทธิภาพของแบบจำลอง กับชุดข้อมูลฝึกฝนกับแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost โดยทดสอบกับข้อมูลชุดฝึกฝน 423 ระเบียน เพื่อทำนายการทุจริตและไม่ทุจริต โดยแบบจำลองที่มีค่าความแม่นสูงสุดคือ XGBoost ซึ่งมีความแม่นอยู่ที่ร้อยละ 88.18 รองลงมาเป็น ป่าสุ่ม ซึ่งมีความแม่นอยู่ที่ร้อยละ 87.94 และความแม่นน้อยสุดคือ การถดถอยลอจิสติกพหุคูณ ซึ่งมีความแม่นอยู่ที่ร้อยละ 65.72

การวัดประสิทธิภาพของแบบจำลอง กับชุดข้อมูลฝึกฝนกับแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost โดยวัดผลด้วยเมทริกซ์ความสับสนได้ผลลัพธ์ดังนี้

ตารางที่ 4.5 เมทริกซ์ความสับสนของการถดถอยลอจิสติกพหุคูณกับข้อมูลชุดทดสอบ

การถดถอยลอจิสติกพหุคูณ	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	75	103
ไม่ทุจริต	36	209

จากตารางที่ 4.5 การวัดผลโดยใช้แบบจำลองการถดถอยลอจิสติกพหุคูณสามารถทำถูกต้องว่างบการเงินมีการทุจริตทั้งหมด 75 กรณี ทำนายว่าถูกว่าไม่ทุจริต 209 กรณี ทำนายว่าทุจริตแต่ผิด 36 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 103 กรณี

ตารางที่ 4.6 เมทริกซ์ความสับสนของป่าสุ่มกับข้อมูลชุดทดสอบ

ป่าสุ่ม	**ทำนาย**	
	ทุจริต	ไม่ทุจริต
ทุจริต	85	25
ไม่ทุจริต	26	287

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.6 การวัดผลโดยใช้แบบจำลองป่าสุ่มสามารถทำถูกต้องว่างบการเงินมีการทุจริตทั้งหมด 85 กรณี ทำนายว่าถูกว่าไม่ทุจริต 287 กรณี ทำนายว่าทุจริตแต่ผิด 26 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 25 กรณี

ตารางที่ 4.7 เมทริกซ์ความสับสนของ XGBoost กับข้อมูลชุดทดสอบ

XGBoost	ทำนาย	
	ทุจริต	ไม่ทุจริต
ทุจริต	85	26
ไม่ทุจริต	24	288

จากตารางที่ 4.7 การวัดผลโดยใช้แบบจำลอง XGBoost สามารถทำถูกต้องว่างบการเงินมีการทุจริตทั้งหมด 85 กรณี ทำนายว่าถูกว่าไม่ทุจริต 288 กรณี ทำนายว่าทุจริตแต่ผิด 24 กรณี ทำนายว่าไม่ทุจริตแต่ผิด 26 กรณี

การเปรียบเทียบสิทธิภาพของแบบจำลอง การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost กับชุดข้อมูลทดสอบ โดยทำการวัดผลด้วยค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1

ตารางที่ 4.8 การวัดประสิทธิภาพของแบบจำลองกับข้อมูลชุดทดสอบ

แบบจำลอง	ค่าทางสถิติ			
	ค่าความแม่น	ค่าความเที่ยง	ค่าการเรียกคืน	ค่าคะแนนเอฟ1
การถดถอยลอจิสติกพหุคูณ	65.72%	67.57%	40.76%	50.85%
ป่าสุ่ม	87.94%	**76.58%**	77.27%	76.92%
XGBoost	**88.18%**	**76.58%**	**77.98%**	**77.27%**

จากตารางที่ 4.8 แบบจำลองที่มีประสิทธิภาพมากที่สุดคือ XGBoost และ ป่าสุ่ม โดยทั้งสองแบบจำลองจะมีประสิทธิภาพที่ใกล้เคียงกัน โดยมีค่าความแม่นอยู่ที่ร้อยละ 88.18 และ 87.94 ตามลำดับ ในขณะที่ค่าความเที่ยงอยู่ที่ร้อยละ 76.58 และ 76.58 ตามลำดับ ค่าการเรียกคืนอยู่ที่ร้อยละ 77.98 และ 77.27 ตามลำดับ และค่าค่าคะแนนเอฟ1อยู่ที่ 77.27 และ 76.92 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 คุณลักษณะที่มีความสำคัญในแต่ละแบบจำลอง

ลำดับ	การถดถอยลอจิสติกพหุคูณ	ป่าสุ่ม	XGBoost
1	Interest Income	otherCurrentLiabilities	operatingExpenses
2	Total Current Liabilities	operatingExpenses	otherCurrentLiabilities
3	Tax Payables	otherExpenses	interestIncome
4	Total Equity	propertyPlantEquipment	commonStock
5	Account Payables	interestIncome	otherCurrentAssets

ตารางที่ 4.9 แสดงให้เห็นถึงคุณลักษณะ 5 อันดับแรกที่มีความสำคัญต่อแบบจำลองโดยเรียงลำดับจากมากไปน้อย โดยเมื่อพิจารณาการเลือกคุณลักษณะของแบบจำลองที่มีประสิทธิภาพสูงสุดอย่างป่าสุ่ม และ XGBoost ซึ่งมีคุณลักษณะในสองอันดับแรกที่เหมือนกันซึ่งก็คือ บัญชีหนี้สินหมุนเวียนอื่น ๆ (otherCurrentLiabilities) และ บัญชีค่าใช้จ่ายในการดำเนินงาน (operatingExpenses) ซึ่งแสดงให้เห็นว่าการทุจริตในงบการเงินเกิดขึ้นที่สองบัญชีที่กล่าวไปข้างต้นมากที่สุด

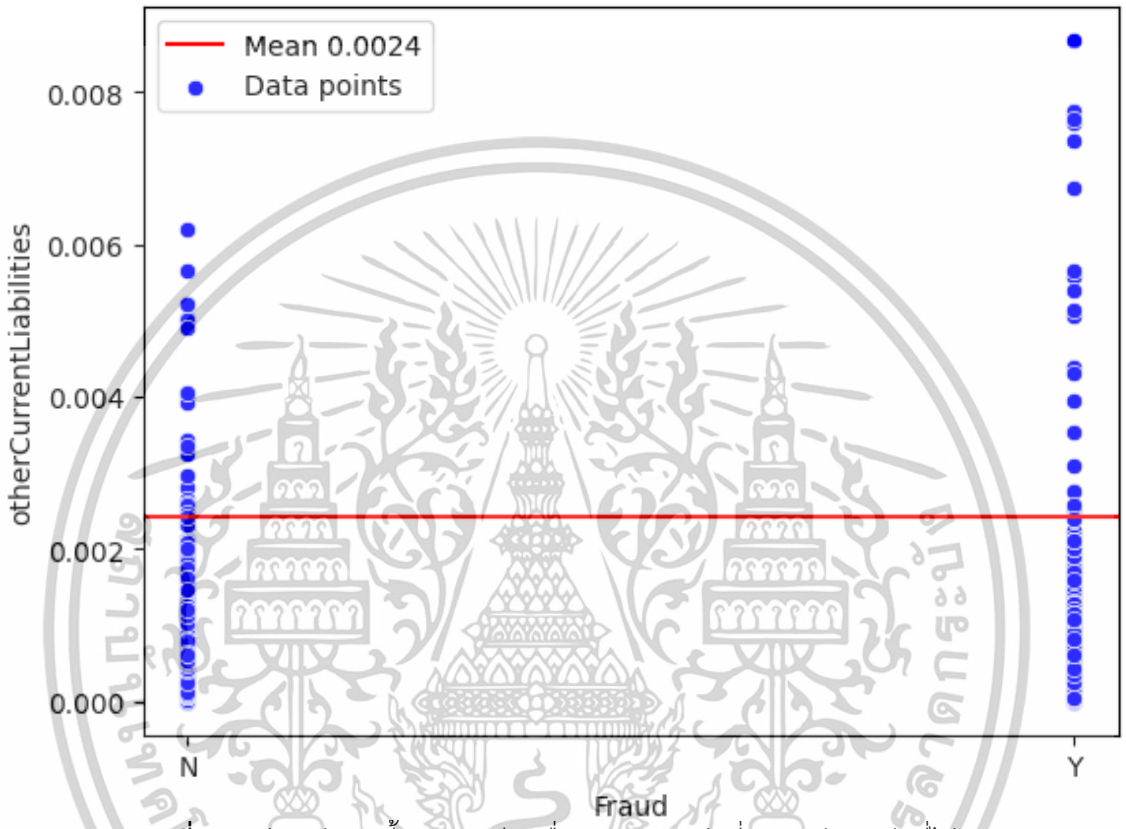
ตารางที่ 4.10 ค่าความสำคัญของตัวแปรที่มีผลต่อการทุจริตของ XGBoost

คุณลักษณะ	ค่าความสำคัญ
operatingExpenses	0.816177
otherCurrentLiabilities	0.746205
otherCurrentAssets	0.537029
interestIncome	0.479327
commonStock	0.373784

จากรูปที่ 4.1 แสดงถึงค่าความสำคัญของตัวแปรที่มีผลต่อการทุจริตของแบบจำลอง XGBoost ยิ่งมีค่าสูงแสดงว่าตัวแปรมีความสำคัญต่อแบบจำลองในการทำนายว่าจะเกิดการทุจริตในงบการเงินโดยตัวแปรที่มีค่าที่สูงที่สุด 5 อันดับแรกคือ ค่าใช้จ่ายในการดำเนินงาน (operatingExpenses) คือ บัญชีหนี้สินหมุนเวียนอื่น ๆ (otherCurrentLiabilities) บัญชีสินทรัพย์หมุนเวียนอื่น ๆ (otherCurrentAssets) บัญชีดอกเบี้ยรับ (InterestIncome) บัญชีหุ้นสามัญ (commonStock)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

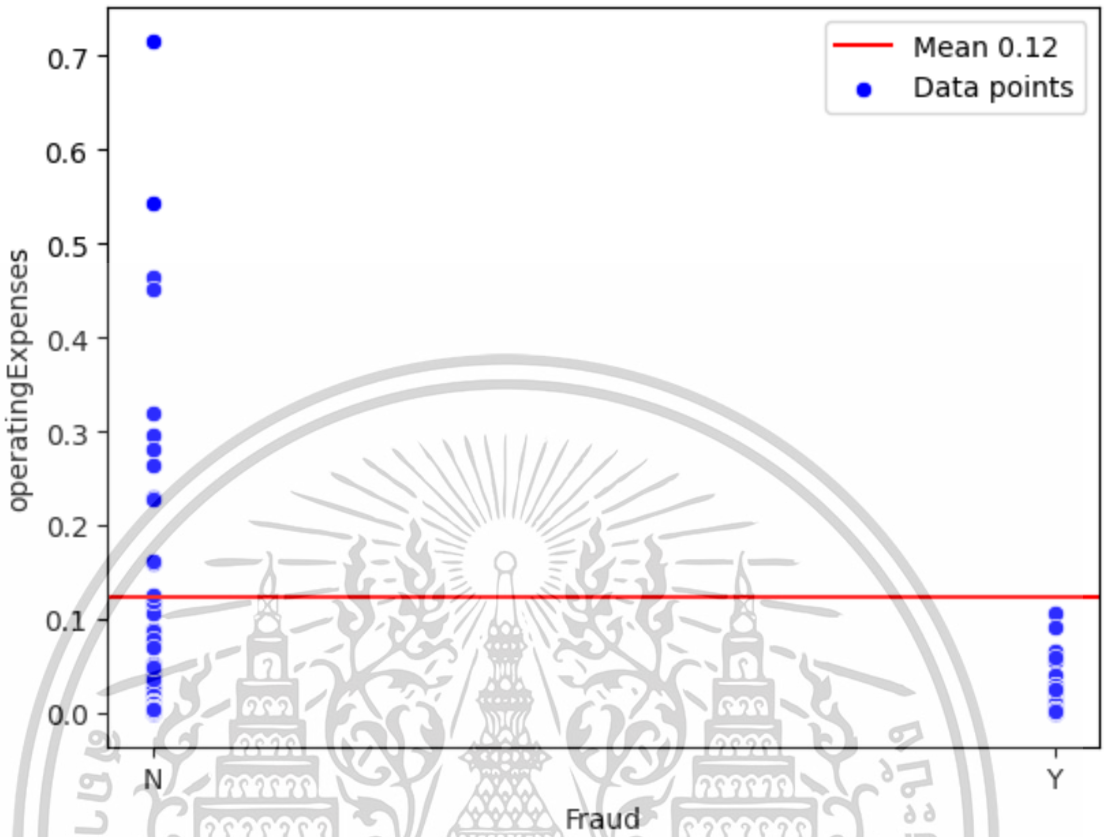
รูปที่ 4.1 – 4.5 แสดงอัตราส่วนของปัจจัยสำคัญที่มีผลต่อการเกิดการทุจริตในงบการเงิน ของแบบจำลอง XGBoost ซึ่งได้แก่ หนี้สินหมุนเวียนอื่น ๆ ค่าใช้จ่ายในการดำเนินงาน ดอกเบี้ยรับ หุ้นสามัญ และหนี้สินหมุนเวียนอื่น ๆ



รูปที่ 4.1 อัตราส่วนหนี้สินหมุนเวียนอื่น ๆ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต

จากรูปที่ 4.1 แสดงให้เห็นความสัมพันธ์ระหว่างอัตราส่วนของบัญชีหนี้สินหมุนเวียนอื่น ๆ กับความเสี่ยงของการกระทำที่อาจเป็นการทุจริต โดยพบว่างบการเงินที่มีค่าอัตราส่วนสูงกว่าค่าเฉลี่ย 0.0024 มีแนวโน้มที่จะมีความเสี่ยงในการทุจริตสูงกว่า เนื่องจากอัตราส่วนนี้สะท้อนถึงระดับหนี้สินที่องค์กรถือครองในระยะสั้น โดยทั่วไป หากบัญชีหนี้สินหมุนเวียนอื่น ๆ มีค่าสูงมากกว่าปกติ อาจแสดงให้เห็นถึงการจัดการเงินทุนที่ไม่โปร่งใส การเพิ่มหนี้สินอย่างรวดเร็ว หรือการสร้างบัญชีหนี้สินที่ไม่เป็นไปตามจริง ดังนั้น การที่ค่าอัตราบัญชีหนี้สินหมุนเวียนอื่น ๆ สูงกว่าค่าเฉลี่ย อาจเป็นตัวบ่งชี้ว่ามีโอกาสที่องค์กรจะทำการปกปิดข้อมูลทางการเงิน หรือมีกิจกรรมที่ผิดปกติซึ่งอาจเข้าข่ายการทุจริต

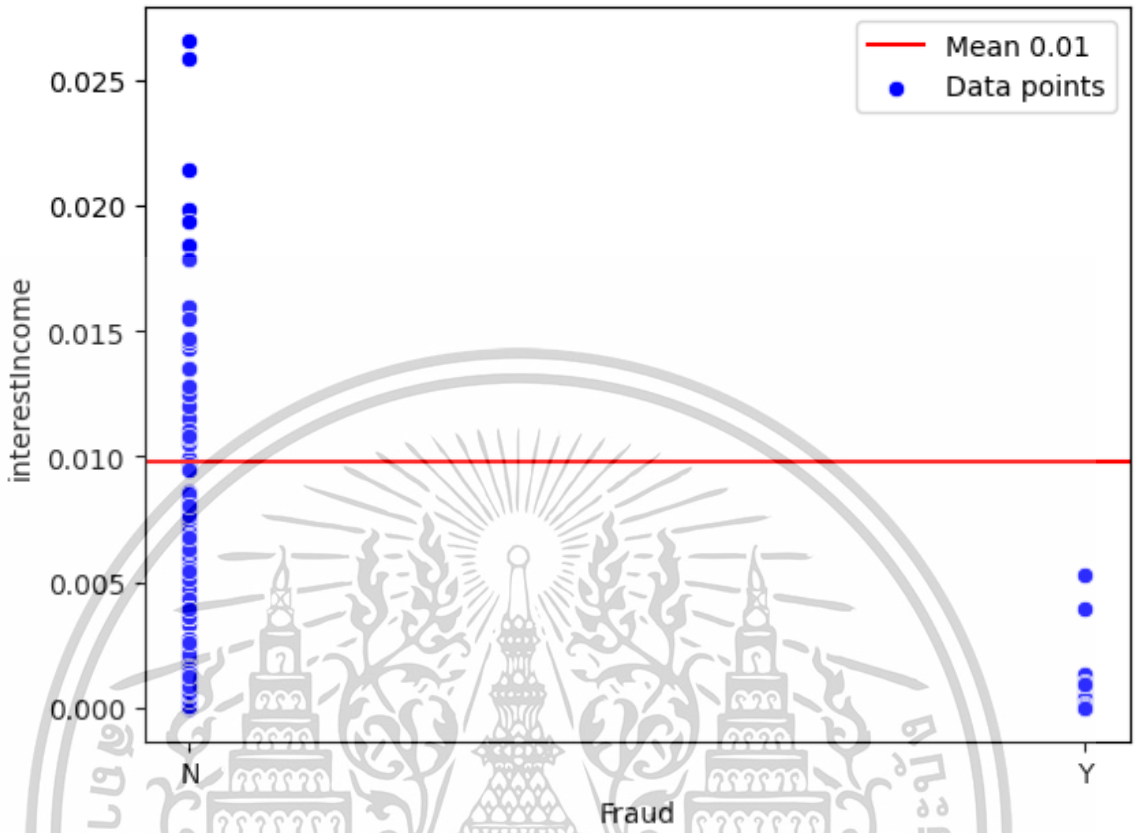
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 อัตราส่วนค่าใช้จ่ายในการดำเนินงาน ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต

จากรูปที่ 4.2 แสดงให้เห็นความสัมพันธ์ระหว่างอัตราส่วนของบัญชีค่าใช้จ่ายในการดำเนินงานกับความเสี่ยงของการกระทำที่อาจเป็นการทุจริต โดยพบว่างบการเงินที่มีค่าอัตราต่ำกว่าค่าเฉลี่ย 0.12 มีแนวโน้มที่จะมีความเสี่ยงในการทุจริตสูงกว่า เนื่องจากเมื่อองค์กรมีอัตราส่วนค่าใช้จ่ายในการดำเนินงานต่ำกว่าค่าเฉลี่ย อาจสะท้อนถึงการรายงานข้อมูลที่ไม่ครบถ้วนหรือการลดค่าใช้จ่ายที่ควรจะเกิดขึ้นในช่วงเวลานั้น ซึ่งอาจเป็นผลมาจากการพยายามปรับลดตัวเลขค่าใช้จ่ายที่แสดงอยู่ในงบการเงินเพื่อสร้างภาพลักษณ์ที่ดีกว่าความเป็นจริง ตัวอย่างเช่น อาจเกิดจากการบันทึกค่าใช้จ่ายต่ำกว่าที่ควร หรือการเลื่อนการรับรู้ค่าใช้จ่ายไปในช่วงถัดไป ซึ่งเป็นเทคนิคที่องค์กรใช้เพื่อเพิ่มกำไรสุทธิหรือแสดงให้เห็นถึงผลประกอบการที่ดีเกินกว่าความเป็นจริง ดังนั้นการที่ค่าใช้จ่ายในการดำเนินงานมีอัตราส่วนที่ต่ำกว่าค่าเฉลี่ยอาจเป็นตัวบ่งชี้ว่ามีโอกาสที่จะเกิดการทุจริตในงบการเงินขึ้นได้

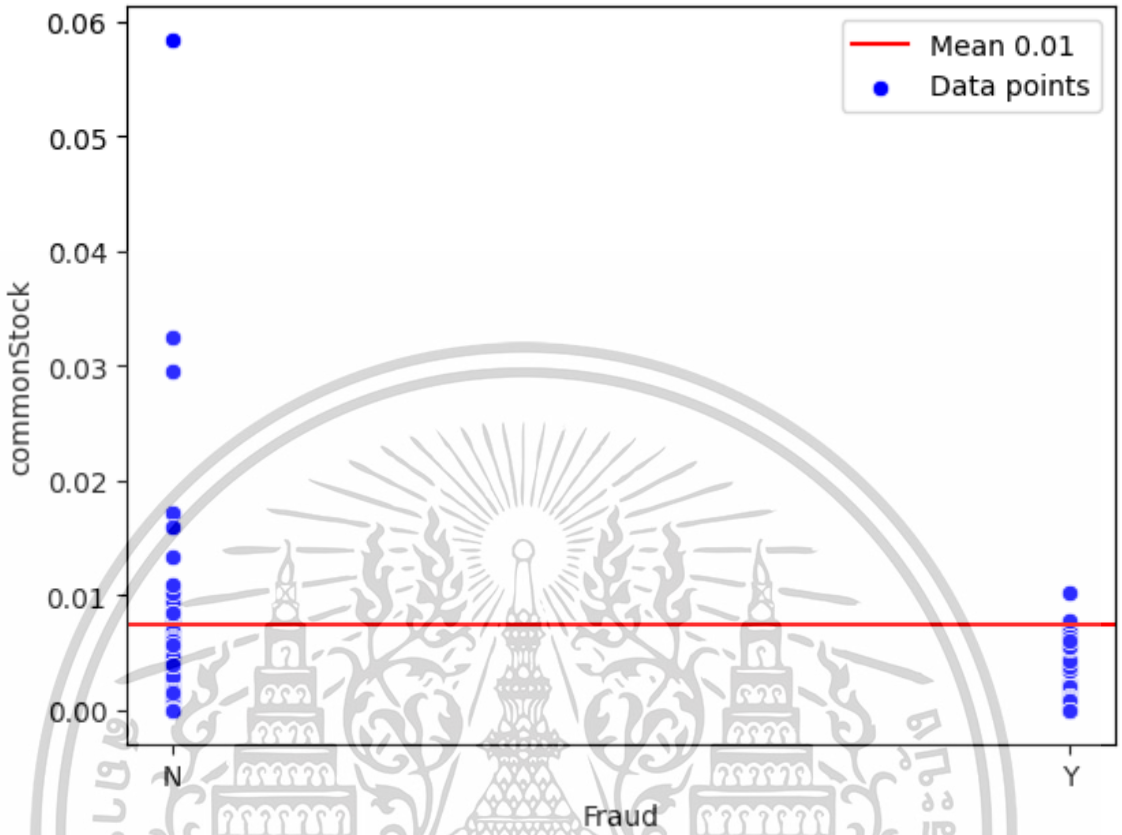
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 อัตราส่วนดอกเบี้ยรับ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต

จากรูปที่ 4.3 แสดงให้เห็นความสัมพันธ์ระหว่างอัตราส่วนของบัญชีดอกเบี้ยรับกับความเสี่ยงของการกระทำที่อาจเป็นการทุจริต โดยพบว่างบการเงินที่มีค่าอัตราต่ำกว่าค่าเฉลี่ย 0.01 มีแนวโน้มที่จะมีความเสี่ยงในการทุจริตสูงกว่า เนื่องจากการมีอัตราส่วนดอกเบี้ยรับต่ำกว่าค่าเฉลี่ยอาจสะท้อนถึงการขาดสภาพคล่องหรือการบริหารจัดการสินทรัพย์ที่ไม่เป็นไปตามปกติ ทั้งนี้ดอกเบี้ยรับเกิดจากรายได้ที่องค์กรได้รับจากการลงทุนหรือสินทรัพย์ที่สร้างผลตอบแทน เมื่ออัตราส่วนดอกเบี้ยรับอยู่ในระดับต่ำ อาจบ่งชี้ถึงการที่องค์กรไม่ได้มีการบริหารจัดการเงินทุนอย่างเหมาะสม หรือมีการพยายามปกปิดรายได้ที่แท้จริงจากการลงทุนเพื่อหลีกเลี่ยงการแสดงข้อมูลที่อาจส่งผลกระทบต่อการรับรู้ของผู้มีส่วนได้ส่วนเสีย ดังนั้นการที่ดอกเบี้ยรับมีอัตราส่วนที่ต่ำกว่าค่าเฉลี่ยอาจเป็นตัวบ่งชี้ว่ามีโอกาสที่จะเกิดการทุจริตในงบการเงินขึ้นได้

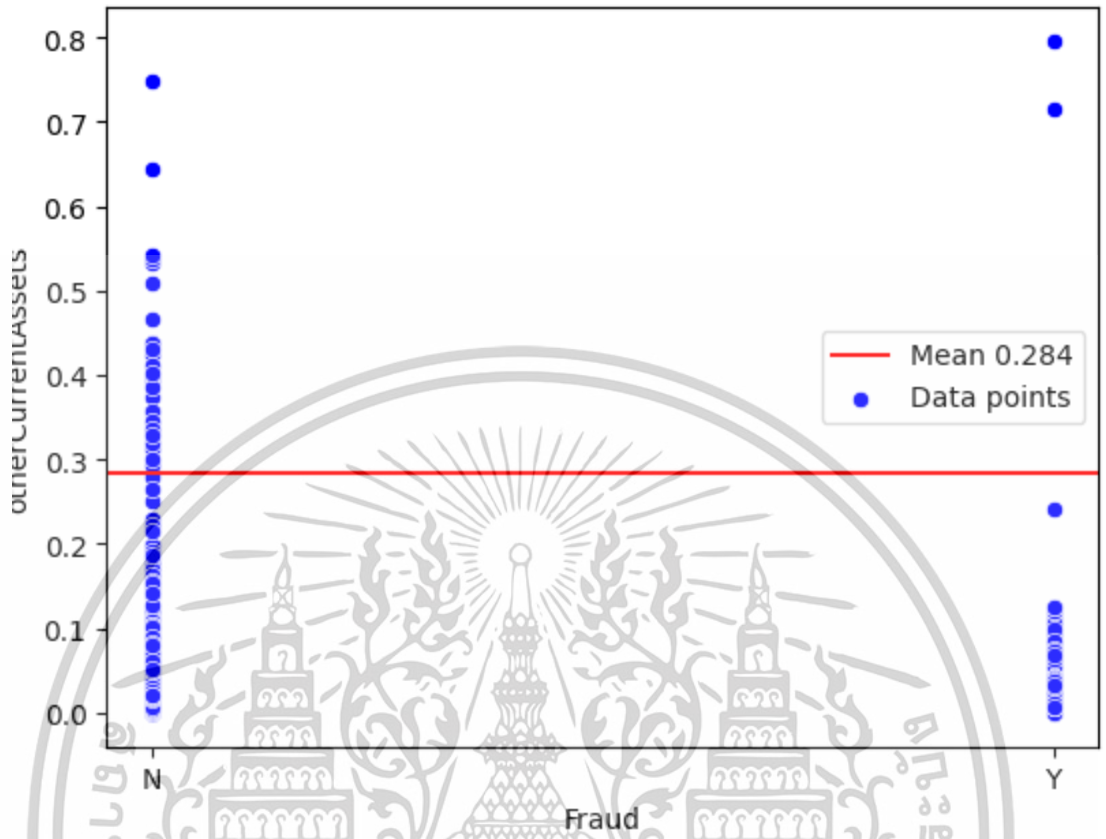
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 อัตราส่วนหุ้นสามัญ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต

จากรูปที่ 4.4 แสดงให้เห็นความสัมพันธ์ระหว่างอัตราส่วนของบัญชีหุ้นสามัญกับความเสี่ยงของการกระทำที่อาจเป็นการทุจริต โดยพบว่างบการเงินที่มีค่าอัตราต่ำกว่าค่าเฉลี่ย 0.01 มีแนวโน้มที่จะมีความเสี่ยงในการทุจริตสูงกว่า เนื่องจากการมีอัตราส่วนของบัญชีหุ้นสามัญต่ำกว่าค่าเฉลี่ยอาจบ่งบอกถึงความพยายามที่จะลดทอนสัดส่วนการถือครองหรือมูลค่าทรัพย์สินที่ต้องแสดงในงบการเงิน ซึ่งเป็นเทคนิคที่องค์กรอาจใช้เพื่อปรับแต่งงบการเงินให้ดูมีผลประกอบการดีขึ้นหรือเพื่อบิดเบือนสภาพทางการเงินให้แตกต่างจากความเป็นจริง ดังนั้นการที่หุ้นสามัญมีอัตราส่วนที่ต่ำกว่าค่าเฉลี่ยอาจเป็นตัวบ่งชี้ว่ามีโอกาสที่จะเกิดการทุจริตในงบการเงินขึ้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.5 อัตราส่วนสินทรัพย์หมุนเวียนอื่น ๆ ของบริษัทที่ทุจริตกับบริษัทที่ไม่ทุจริต

จากรูปที่ 4.5 แสดงให้เห็นว่าการมีอัตราส่วนสินทรัพย์หมุนเวียนอื่น ๆ ต่ำกว่าค่าเฉลี่ย อาจเกิดจากการพยายามซ่อนสินทรัพย์บางส่วนหรือการลดมูลค่าสินทรัพย์ลงเพื่อให้ภาพรวมทางการเงินดูเป็นไปในทิศทางที่ต้องการ อาทิ เพื่อปกปิดภาระผูกพันทางการเงินหรือหลีกเลี่ยงการเปิดเผยทรัพย์สินที่แท้จริง ซึ่งเป็นสัญญาณที่องค์กรอาจใช้ในการบิดเบือนข้อมูลทางการเงินเพื่อหลีกเลี่ยงการตรวจสอบหรือเพื่อผลประโยชน์อื่น ๆ ดังนั้นการที่สินทรัพย์หมุนเวียนอื่น ๆ มีอัตราส่วนที่ต่ำกว่าค่าเฉลี่ยอาจเป็นตัวบ่งชี้ว่ามีโอกาสที่จะเกิดการทุจริตในงบการเงินขึ้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 การนำมาแบบจำลองมาใช้ตรวจสอบทุจริตในงบการเงินไทย

อุตสาหกรรม	จำนวน	ปี	XGBoost
Basic Materials	3	2564	ทุจริต
		2565	ทุจริต
		2566	ทุจริต
Media	1	2566	ไม่ทุจริต
Utilities	1	2565	ไม่ทุจริต

ตารางที่ 4.11 แสดงให้เห็นถึงการนำแบบจำลอง XGBoost มาใช้สำหรับงบการเงินในประเทศไทย โดยได้นำงบการเงินของบริษัทที่ถูกระบุว่าทุจริตมาทั้งหมด 3 บริษัท ในอุตสาหกรรมที่แตกต่างกัน ซึ่งประกอบด้วยงบการเงินทั้งหมด 5 งบการเงิน ผลการทดลองพบว่า แบบจำลอง XGBoost สามารถทำนายบริษัทที่ทุจริตได้ 3 งบ จากงบการเงินทั้งหมด 5 งบการเงิน

4.2 การอภิปรายผล

การวัดประสิทธิภาพของแบบจำลอง โดยใช้วิธีการเรียนรู้แบบมีผู้สอน 3 วิธี ได้แก่ การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost พบว่าวิธีที่มีประสิทธิภาพในการทำนายงบการเงินที่เกิดการทุจริตได้ดีที่สุดคือวิธี XGBoost โดยมีค่าความแม่น ค่าความเที่ยง ค่าการเรียนกคืน และค่าคะแนนเอฟ1 สูงที่สุดที่ร้อยละ 0.88, 0.77, 0.78 และ 0.77 ตามลำดับ สอดคล้องกับงานวิจัยของ Aftabi et al. (2023) ที่ใช้ XGBoost ทำนายการทุจริตของงบการเงินในธนาคารประเทศอิหร่าน

จากการศึกษาพบว่า XGBoost มีประสิทธิภาพสูงในการทำนายความเสี่ยงของการทุจริตในงบการเงิน โดยมีความแม่นเหนือกว่าแบบจำลองอื่น ๆ เนื่องจาก XGBoost ใช้หลักการของเกรเดียนต์บูสติ้งในการสร้างตัวจำแนกหลายชุด โดยตัวจำแนกแต่ละชุดได้รับข้อมูลที่แตกต่างกันและพยายามเรียนรู้จากข้อผิดพลาดของตัวจำแนกก่อนหน้าเพื่อปรับปรุงการทำนายในรอบถัดไป กระบวนการนี้ช่วยลดความคลาดเคลื่อนของแบบจำลองอย่างมีนัยสำคัญ การค้นคว้าอิสระนี้ใช้คุณลักษณะทั้งหมด 86 คุณลักษณะในการสร้างแบบจำลอง ซึ่ง XGBoost มีประสิทธิภาพสูงในการจัดการกับข้อมูลที่มีคุณลักษณะจำนวนมาก ส่งผลให้แบบจำลองที่พัฒนาขึ้นมีความสามารถในการทำนายที่แม่นยำและเชื่อถือได้

จากการศึกษาตัวแปรที่มีผลต่อแบบจำลองของแบบจำลองที่มีประสิทธิภาพในการทำนายการทุจริตในงบการเงินสูงที่สุด คือ XGBoost โดยตัวแปร 5 อันดับแรกที่มีผลต่อแบบจำลองคือ บัญชี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าใช้จ่ายในการดำเนินงาน บัญชีหนี้สินหมุนเวียนอื่น ๆ บัญชีดอกเบี้ยรับ บัญชีหุ้นสามัญ โดยจากการศึกษาพบว่าสำหรับตัวแปรข้างต้นยิ่งตัวแปรมีค่าที่ลดลงหรือเพิ่มขึ้นมากเท่าไหร่ โอกาสในการที่จะเกิดการทุจริตในงบการเงินยิ่งมีมากขึ้นเท่านั้น โดยที่มี 1 ตัวแปรที่สอดคล้องกับงานวิจัยของ Zhongzhu et al. (2021) ซึ่งได้แก่ บัญชีค่าใช้จ่ายในการดำเนินงาน โดยสาเหตุที่บัญชีอื่น ๆ มีความแตกต่างกัน เป็นผลมาจากการเลือกใช้ข้อมูลชุดฝึกฝน และวิธีการจัดการกับข้อมูลที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การศึกษาการตรวจสอบการทุจริตในงบการเงินโดยการสร้างแบบจำลองโดยใช้การเรียนรู้แบบมีผู้สอน ทั้งหมด 3 วิธี ได้แก่ การถดถอยลอจิสติกพหุคูณ ป่าสุ่ม และ XGBoost โดยใช้ข้อมูลจากตลาดหลักทรัพย์ NASDAQ และ NYSE ของประเทศสหรัฐอเมริกา ซึ่งชุดข้อมูลมีจำนวนคุณลักษณะทั้งหมด 122 คุณลักษณะ และมีข้อมูล 1,362 ระเบียน โดยใช้วิธีการวัดประสิทธิภาพของแบบจำลอง ได้แก่ ค่าความแม่น ค่าความเที่ยง ค่าการเรียกคืน และค่าคะแนนเอฟ1 และทำการเลือกตัวแปรที่มีค่าความสำคัญสูงที่สุด 5 ลำดับแรก เพื่อศึกษาตัวแปรที่มีผลต่อการเกิดทุจริตในงบการเงิน

5.1 สรุปผลการวิจัย

การศึกษาในงานวิจัยนี้ ซึ่งมุ่งเน้นการสร้างแบบจำลองเพื่อใช้ทำนายความผิดปกติของงบการเงิน โดยอาศัยข้อมูลจากงบการเงินประจำปีของบริษัทและการใช้เทคนิคการเรียนรู้แบบมีผู้สอนนำมาสร้างแบบจำลอง และได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลอง ทั้งหมด 3 แบบจำลอง โดยแบบจำลองที่มีประสิทธิภาพในการทำนายความผิดปกติในงบการเงิน ซึ่งมีความแม่นสูงที่สุด คือ XGBoost ซึ่งมีค่าความแม่นอยู่ที่ ร้อยละ 88.18 ค่าความเที่ยง ร้อยละ 76.58 ค่าการเรียกคืน ร้อยละ 77.98 และค่าคะแนนเอฟ1 ร้อยละ 77.27 ซึ่งถือว่ามีประสิทธิภาพในการทำนายมากที่สุดเมื่อเทียบกับแบบจำลองป่าสุ่ม และ การถดถอยลอจิสติกพหุคูณ

การศึกษาพบว่าปัจจัยทางการเงินที่ส่งผลมากที่สุดต่อการทำนายความผิดปกติทางการเงิน ได้แก่ หนี้สินหมุนเวียนอื่น ๆ ค่าใช้จ่ายในการดำเนินงาน ดอกเบี้ยรับ ส่วนของผู้ถือหุ้น และสินทรัพย์หมุนเวียนอื่น ๆ ซึ่งจากการวิเคราะห์ความสัมพันธ์ของตัวแปรเหล่านี้ช่วยให้สามารถที่จะระบุได้ว่ายิ่งตัวแปรที่ได้กล่าวมาข้างต้นมีความผิดปกติมากเท่าไหร่ ยิ่งส่งผลต่อโอกาสที่เกิดความผิดปกติในงบการเงินมากขึ้นเท่านั้น โดยมีตัวแปรของบัญชีค่าใช้จ่ายในการดำเนินงาน มีความสอดคล้องกับงานวิจัยของ Zhongzhu et al. (2021) ซึ่งได้ทำการศึกษาตัวแปรที่มีผลต่อการทุจริตแล้วพบว่าบัญชีค่าใช้จ่ายในการดำเนินงานเป็นปัจจัยในการเกิดทุจริตในงบการเงิน อย่างไรก็ตามตัวแปรที่กล่าวมาข้างต้น สามารถใช้ในการตรวจสอบความผิดปกติของงบการเงินในเบื้องต้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

1. เนื่องจากงบการเงินที่ถูกว่าระบุว่ามีการทุจริตทางงบการเงิน มีค่อนข้างน้อย เมื่อเทียบกับงบการเงินที่ถูกระบุว่ามีการดำเนินงานอย่างโปร่งใส ดังนั้นถ้าสามารถเพิ่มคลาสของข้อมูลเหล่านี้มีความสมดุลมากยิ่งขึ้น จะสามารถเพิ่มประสิทธิของแบบจำลองให้ดียิ่งขึ้นได้

2. ข้อมูลงบการเงินในประเทศไทย ที่ระบุว่ามีการทุจริตมีจำนวนน้อย ทำให้ไม่สามารถใช้ในการสอนแบบจำลองให้มีประสิทธิภาพได้ ซึ่งถ้าข้อมูลข้างต้นนี้มีเพียงพอต่อการสอนแบบจำลอง จะสามารถนำมาปรับใช้กับงบการเงินในประเทศไทยได้ดียิ่งขึ้น

3. ศึกษาการปรับค่าพารามิเตอร์ของแต่ละแบบจำลองเพิ่มเติม เช่น จำนวนของต้นไม้ ความลึกสูงสุด เพื่อเพิ่มประสิทธิภาพของแบบจำลองให้สามารถตรวจจับความผิดปกติในงบการเงินได้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

คำนึง สา ริ สระ และ ธรรม นิติ เพรส. 2020. *101 งบ การเงิน นัก บัญชี มือใหม่ ต้อง รู้*. DHARMNITI PRESS CO., LTD.

Aftabi, S. Z., Ahmadi, A., and Farzi, S. 2023. Fraud detection in financial statements using data mining and GAN models. *Expert Systems with Applications*, *227*, 120144. https://doi.org/10.1016/j.eswa.2023.120144

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, *55*(5), 24-36.

Bishop, C. M., and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*. New York springer, 738.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. 2010. Detecting management fraud in public companies. *Management Science*, *56*(7), 1146-1160.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Chaiyadecha, S. 2020. One-Hot Encoding สร้างตัวแปร Dummies สำหรับ Classification Model. [Online]. Available: https://lengyi.medium.com/one-hot encoding-737c66e5b1bd.

Corporate Finance Institute. 2021. Financial statement manipulation: Overview, reasons, and factors. Corporate Finance Institute. [Online]. Available: https://corporatefinanceinstitute.com/resources/accounting/financial-statement-manipulation/

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367-378.

International Business Machines. 2023. What is random forest? IBM. [Online]. Available: https://www.ibm.com/topics/random-forest

Liu, Z., Ye, R., and Ye, R. 2021. Detecting Financial Statement Fraud with Interpretable Machine Learning.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Maklin, C. 2022. Synthetic minority over-sampling technique (SMOTE). Medium. [Online]. Available: https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c

Narkhede, S. 2018. Understanding confusion matrix. Towards Data Science. [Online]. Available: https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c

Saini, A. 2023. What is logistic regression: A complete guide. Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/08/conceptual-understanding-of-logistic-regression-for-data-science-beginners/

Sattar, G. 2023. Financial Statement Manipulation of Listed Companies and Prevention. *Academic Journal of Business & Management*, *5*(20), 119-124.

Singh, A. 2023. Ensemble learning: Ensemble techniques. Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

Tharwat, A. 2021. Classification assessment methods. *Applied computing and informatics*, *17*(1), 168-192.

Terence, S. 2023. Understanding Feature Importance in Machine Learning. [Online]. Available: https://builtin.com/data-science/feature-importance

Tseng, G. 2018. Gradient boosting and XGBoost. Medium. [Online]. Available: https://medium.com/@gabrieltseng/gradient-boosting-and-xgboost-c306c1bcfaf5

Vrigazova, B. 2021. The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, *12*(1), 228-242. https://doi.org/10.2478/bsrj-2021-0015

Wyrobek, J. 2020. Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture. *Procedia Computer Science*, *176*, 3037-3046.DOI: https://doi.org/10.1016/j.procs.2020.09.335

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

1. คุณลักษณะทั้งหมดในงบแสดงฐานะการเงิน

ตารางที่ 1 คุณลักษณะทั้งหมดในงบแสดงฐานะการเงิน

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
1	cashAndCashEquivalents	จำนวนจริง
2	shortTermInvestments	จำนวนจริง
3	cashAndShortTermInvestments	จำนวนจริง
4	netReceivables	จำนวนจริง
5	inventory	จำนวนจริง
6	otherCurrentAssets	จำนวนจริง
7	totalCurrentAssets	จำนวนจริง
8	propertyPlantEquipmentNet	จำนวนจริง
9	goodwill	จำนวนจริง
10	intangibleAssets	จำนวนจริง
11	goodwillAndIntangibleAssets	จำนวนจริง
12	longTermInvestments	จำนวนจริง
13	taxAssets	จำนวนจริง
14	otherNonCurrentAssets	จำนวนจริง
15	totalNonCurrentAssets	จำนวนจริง
16	totalAssets	จำนวนจริง
17	accountPayables	จำนวนจริง
18	shortTermDebt	จำนวนจริง
19	taxPayables	จำนวนจริง
20	deferredRevenue	จำนวนจริง
21	otherCurrentLiabilities	จำนวนจริง
22	totalCurrentLiabilities	จำนวนจริง
23	longTermDebt	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1 คุณลักษณะทั้งหมดในงบแสดงฐานะการเงิน (ต่อ)

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
24	deferredRevenueNonCurrent	จำนวนจริง
25	deferredTaxLiabilitiesNonCurrent	จำนวนจริง
26	otherNonCurrentLiabilities	จำนวนจริง
27	totalNonCurrentLiabilities	จำนวนจริง
28	otherLiabilities	จำนวนจริง
29	capitalLeaseObligations	จำนวนจริง
30	totalLiabilities	จำนวนจริง
31	preferredStock	จำนวนจริง
32	commonStock	จำนวนจริง
33	retainedEarnings	จำนวนจริง
34	othertotalStockholdersEquity	จำนวนจริง
35	totalStockholdersEquity	จำนวนจริง
36	totalEquity	จำนวนจริง
37	totalLiabilitiesAndStockholdersEquity	จำนวนจริง
38	minorityInterest	จำนวนจริง
39	longTermLiabilities	จำนวนจริง
40	totalLiabilitiesAndTotalEquity	จำนวนจริง
41	totalInvestments	จำนวนจริง
42	totalDebt	จำนวนจริง
43	netDebt	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. คุณลักษณะทั้งหมดในงบกำไรขาดทุนเบ็ดเสร็จ

ตารางที่ 2 คุณลักษณะทั้งหมดในงบกำไรขาดทุนเบ็ดเสร็จ

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
1	revenue	จำนวนจริง
2	costOfRevenue	จำนวนจริง
3	grossProfit	จำนวนจริง
4	grossProfitRatio	จำนวนจริง
5	researchAndDevelopmentExpenses	จำนวนจริง
6	generalAndAdministrativeExpenses	จำนวนจริง
7	sellingAndMarketingExpenses	จำนวนจริง
8	sellingGeneralAndAdministrativeExpenses	จำนวนจริง
9	otherExpenses	จำนวนจริง
10	operatingExpenses	จำนวนจริง
11	costAndExpenses	จำนวนจริง
12	interestIncome	จำนวนจริง
13	interestExpense	จำนวนจริง
14	depreciationAndAmortization	จำนวนจริง
15	ebit	จำนวนจริง
16	ebitda	จำนวนจริง
17	ebitdaratio	จำนวนจริง
18	operatingIncome	จำนวนจริง
19	operatingIncomeRatio	จำนวนจริง
20	totalOtherIncomeExpensesNet	จำนวนจริง
21	incomeBeforeTax	จำนวนจริง
22	incomeBeforeTaxRatio	จำนวนจริง
23	incomeTaxExpense	จำนวนจริง
24	netIncome	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. คุณลักษณะทั้งหมดในงบกำไรขาดทุนเบ็ดเสร็จ

ตารางที่ 3 คุณลักษณะทั้งหมดในงบกำไรขาดทุนเบ็ดเสร็จ

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
1	netCashProvidedByOperatingActivities	จำนวนจริง
2	investmentsInPropertyPlantAndEquipment	จำนวนจริง
3	acquisitionsNet	จำนวนจริง
4	purchasesOfInvestments	จำนวนจริง
5	salesMaturitiesOfInvestments	จำนวนจริง
6	otherInvestingActivites	จำนวนจริง
7	netCashUsedForInvestingActivites	จำนวนจริง
8	debtRepayment	จำนวนจริง
9	commonStockIssued	จำนวนจริง
10	commonStockRepurchased	จำนวนจริง
11	dividendsPaid	จำนวนจริง
12	otherFinancingActivites	จำนวนจริง
13	netCashUsedProvidedByFinancingActivities	จำนวนจริง
14	effectOfForexChangesOnCash	จำนวนจริง
15	netChangeInCash	จำนวนจริง
16	cashAtEndOfPeriod	จำนวนจริง
17	cashAtBeginningOfPeriod	จำนวนจริง
18	operatingCashFlow	จำนวนจริง
19	capitalExpenditure	จำนวนจริง
20	freeCashFlow	จำนวนจริง
21	revenuePerShare	จำนวนจริง
22	netIncomePerShare	จำนวนจริง
23	operatingCashFlowPerShare	จำนวนจริง
24	freeCashFlowPerShare	จำนวนจริง
25	cashPerShare	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. คุณลักษณะทั้งหมดของอัตราส่วนทางการเงิน

ตารางที่ 4 คุณลักษณะทั้งหมดของอัตราส่วนทางการเงิน

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
1	bookValuePerShare	ชนิดข้อมูล
2	tangibleBookValuePerShare	จำนวนจริง
3	shareholdersEquityPerShare	จำนวนจริง
4	interestDebtPerShare	จำนวนจริง
5	marketCap	จำนวนจริง
6	enterpriseValue	จำนวนจริง
7	peRatio	จำนวนจริง
8	priceToSalesRatio	จำนวนจริง
9	pocfratio	จำนวนจริง
10	pfcfRatio	จำนวนจริง
11	pbRatio	จำนวนจริง
12	ptbRatio	จำนวนจริง
13	evToSales	จำนวนจริง
14	enterpriseValueOverEBITDA	จำนวนจริง
15	evToOperatingCashFlow	จำนวนจริง
16	evToFreeCashFlow	จำนวนจริง
17	earningsYield	จำนวนจริง
18	freeCashFlowYield	จำนวนจริง
19	debtToEquity	จำนวนจริง
20	debtToAssets	จำนวนจริง
21	netDebtToEBITDA	จำนวนจริง
22	currentRatio	จำนวนจริง
23	interestCoverage	จำนวนจริง
24	incomeQuality	จำนวนจริง
25	dividendYield	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4 คุณลักษณะทั้งหมดของอัตราส่วนทางการเงิน (ต่อ)

ลำดับที่	คุณลักษณะ	ชนิดข้อมูล
26	payoutRatio	จำนวนจริง
27	capexToRevenue	จำนวนจริง
28	capexToDepreciation	จำนวนจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นาย ภาณุพงศ์ เมี้ยนกลาง
วัน เดือน ปีเกิด	4 มิถุนายน 2541
ที่อยู่ปัจจุบัน	308/5 ถ.เทศา 2 ต.ในเมือง อ.เมืองกำแพงเพชร จ.กำแพงเพชร 62000
ประวัติการศึกษา	(2564) บัญชีบัณฑิต สาขาวิชา การบัญชี (จุฬาลงกรณ์มหาวิทยาลัย)
	(2566) วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูลและการวิเคราะห์ (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้